

CAMILE SOTHE

**CLASSIFICAÇÃO DO ESTÁDIO SUCESSIONAL DA
VEGETAÇÃO EM ÁREAS DE FLORESTA OMBRÓFILA
MISTA EMPREGANDO ANÁLISE BASEADA EM OBJETO E
ORTOIMAGENS**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Florestal, da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Engenharia Florestal.

Orientador: Marcos Benedito
Schimalski

LAGES, SC
2015

SOTHE, CAMILE

CLASSIFICAÇÃO DO ESTÁDIO SUCESSIONAL DA VEGETAÇÃO
EM ÁREAS DE FLORESTA OMBRÓFILA MISTA EMPREGANDO
ANÁLISE BASEADA EM OBJETO E ORTOIMAGENS / CAMILE
SOTHE. - LAGES, 2015

250 p. : il. ; 21 cm

Orientador: Marcos Benedito Schimalski

Inclui bibliografia

Dissertação (mestrado) - Universidade do Estado de
Santa Catarina, Centro de Ciências Agroveterinárias,
Programa de Pós-Graduação em Engenharia Florestal,
Lages, 2015.

1. Florestas secundárias a. 2. Levantamento
aerofotogramétrico b. 3. Processamento digital de
imagens c. 4. Mineração de dados d. 5. Máquinas de
vetor de suporte e. I. SOTHE, CAMILE. II Schimalski,
Marcos Benedito. III. Universidade do Estado de Santa
Catarina. Programa de Pós-Graduação em Engenharia
Florestal. IV. CLASSIFICAÇÃO DO ESTÁDIO SUCESSIONAL
DA VEGETAÇÃO EM ÁREAS DE FLORESTA OMBRÓFILA MISTA
EMPREGANDO ANÁLISE BASEADA EM OBJETO E ORTOIMAGENS

CAMILE SOTHE

**CLASSIFICAÇÃO DO ESTÁDIO SUCESSIONAL DA
VEGETAÇÃO EM ÁREAS DE FLORESTA OMBRÓFILA
MISTA EMPREGANDO ANÁLISE BASEADA EM OBJETO E
ORTOIMAGENS**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Florestal, da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Engenharia Florestal.

Banca Examinadora:

Orientador: _____

Prof. Dr. Marcos Benedito Schimalski
Universidade do Estado de Santa Catarina

Membro: _____

Prof. Dr. Sílvio Luís Rafaeli Neto
Universidade do Estado de Santa Catarina

Membro: _____

Prof. Dra. Cláudia Maria de Almeida
Instituto Nacional de Pesquisas Espaciais

Lages, 31/07/2015

AGRADECIMENTOS

Agradeço a Deus por estar concluindo mais esta etapa da minha vida.

Ao meu orientador Prof. Dr. Marcos Benedito Schimalski, por ter acreditado no meu potencial desde o início, acompanhando-me e sempre respondendo prontamente os e-mails, possibilitando que, mesmo a distância, eu pudesse ter um bom aprendizado.

Aos meus pais, Vilmar e Ceci, pelo apoio, dedicação e incentivo sempre. Agradeço em especial ao meu pai que, preocupado com minha segurança, me levou tantas vezes para Lages.

Agradeço aos meus amigos e colegas de Mestrado, pelos dois anos que dividimos as experiências, e que de certa forma contribuíram para a minha conquista. Em especial agradeço a Carla, minha colega e amiga desde o tempo da graduação, por ter me fornecido algumas estadias em Lages e por sua companhia.

Agradeço à Polícia Militar Ambiental por incentivar e não criar nenhum empecilho para que eu estudasse. Agradeço em especial a minha comandante, Tenente Alessandra, pela compreensão nessa etapa, sempre possibilitando a flexibilização do meu horário de trabalho para que eu pudesse viajar para Lages.

Agradeço ao meu amigo e colega de trabalho Luís Cláudio, por ter escutado todas as minhas preocupações e angústias durante esse período, me apoiando e dando ideias, sempre com uma dose de bom humor.

Agradeço aos bolsistas de Graduação, Camila e João, por terem me auxiliado em alguns procedimentos, e ao professor Veraldo pelas dicas e correções.

Enfim, agradeço a todas as pessoas que diretamente ou indiretamente contribuíram para a realização deste sonho.

“[...] acho que você deveria realmente promover uma mudança radical em seu estilo de vida e começar a fazer corajosamente coisas em que talvez nunca tenha pensado, ou que fosse hesitante demais para tentar. Tanta gente vive em circunstâncias infelizes e, contudo, não toma a iniciativa de mudar a situação porque estão condicionados a uma vida de segurança, conformismo e conservadorismo. Tudo isso que parece dar paz de espírito, mas na realidade nada é mais maléfico para o espírito aventureiro do homem que um futuro seguro. A coisa mais essencial do espírito vivo de um homem é sua paixão pela aventura. A alegria da vida vem de nossos encontros com novas experiências, e, portanto, não há alegria maior que ter um horizonte sempre cambiante, cada dia com um novo e diferente Sol”.

Cristopher McCandless

RESUMO

SOTHE, C. CLASSIFICAÇÃO DO ESTÁDIO SUCESSIONAL DA VEGETAÇÃO EM ÁREAS DE FLORESTA OMBRÓFILA MISTA EMPREGANDO ANÁLISE BASEADA EM OBJETO E ORTOIMAGENS. 2015. 250 f. Dissertação (Mestrado em Engenharia Florestal) – Universidade do Estado de Santa Catarina. Programa de Pós-Graduação em Engenharia Florestal, Lages, SC. 2015.

Nas últimas décadas, observou-se uma notável evolução das tecnologias espaciais destinadas ao monitoramento dos recursos florestais. Um significativo avanço nos dados de sensoriamento é o refinamento da resolução espacial. Com a crescente disponibilidade de imagens de alta resolução espacial, a classificação por regiões e a análise baseada em objeto (*Object Based Image Analysis*- OBIA) apresentam-se como abordagens mais adequadas para extrair informações dessas imagens. No entanto, os custos envolvidos na aquisição de licenças de aplicativos comerciais disponíveis para este propósito costuma ser demasiadamente alto. Assim, faz-se necessário avaliar o uso de aplicativos na modalidade *open source*. Este trabalho teve como objetivo avaliar aplicativos *open source* relacionadas à classificação baseada em regiões e à mineração de dados, para classificar estádios sucessionais de florestas secundárias da Floresta Ombrófila Mista (FOM) em três áreas-teste situadas na região serrana de Santa Catarina (SC). No processamento, utilizaram-se ortofotogramas do Sistema Aerotransportado de Aquisição e Pós-processamento de Imagens (*Airborne System for Acquisition and Post-processing of Images*- SAAPI) com alta resolução espacial (0,39 m) obtidas no levantamento aerofotogramétrico de SC. Os dados consistem de três bandas no visível (0,38 - 0,70 μm), três bandas no infravermelho próximo (0,76 - 0,78 μm) e o Modelo Digital de Superfície. Três metodologias foram desenvolvidas

utilizando mineração de dados com algoritmos de árvore de decisão, nos aplicativos InterIMAGE, GeoDMA, WEKA e QGIS. O algoritmo de máquina de vetor suporte (*Support Vector Machine- SVM*) foi selecionado para a classificação baseada em regiões e por *pixel* no *software* Orfeo ToolBox Monteverdi. Testou-se também o *software* SPT para avaliação e escolha automática dos parâmetros da segmentação das imagens. Os resultados se mostraram satisfatórios para classificar estádios sucessionais da FOM, assim como outras classes de uso e cobertura da terra. As classificações apresentaram um índice Kappa variando entre 0,6 e 0,89. A avaliação condicional das classes referentes aos estádios sucessionais (exatidão e Kappa condicional do produtor e usuário), no geral, foram superiores a 0,5, sendo os melhores resultados obtidos na identificação do estádio inicial e os piores para o estádio médio. A classificação baseada em regiões e a OBIA foram significativamente superiores à classificação *pixel* a *pixel*. Estes resultados demonstram o potencial dessas abordagens na extração de informações de imagens de alta resolução espacial, como os provenientes do recobrimento aéreo estadual, bem como, a possibilidade de fornecer subsídios para a implementação de políticas públicas e no monitoramento dos recursos florestais em nível estadual.

Palavras-chave: Florestas secundárias. Levantamento aerofotogramétrico. Processamento digital de imagens. Mineração de dados. Máquinas de vetor de suporte.

ABSTRACT

SOTHE, C. CLASSIFYING SUCESSIONAL FOREST STAGES IN MIXED OMBROPHILOUS FOREST ENVIRONMENTS USING OBJECT BASED IMAGE ANALYSIS AND ORTHOIMAGES. 2015. 250 f. Dissertation (Master in Forest Engineering) - Santa Catarina State University (UDESC), Forest Engineering Graduate Program, Lages. 2015.

Over the last decades, advances in Earth Observation technology have been playing an important role for forest monitoring worldwide. A remarkable improvement in remotely sensed data is the refinement of the spatial resolution. The region-based classification and object-based image analysis (OBIA) appears to be the most appropriate approaches to extract information coming from high spatial resolution images. Such information is becoming even more frequently. However, the costs involved in acquiring propriety OBIA software licenses is often too high. Therefore, the use of open source softwares is envisaged. This study aimed to evaluate open source softwares in order to classify secondary successional forest stages of Ombrophilous Mixed Forest environments in Southern Brazil. Three test sites were selected in the mountainous region of Santa Catarina State (SC). We used scenes from the airborne system for acquisition and post-processing of images (SAAPI) with a spatial resolution of 0.39m. The dataset consists of orthorectified images containing three spectral bands in the visible range (i.e. 0.38–0.70 μ m), three spectral bands in the near infrared (i.e. 0.76–0.78 μ m) and a digital surface model. Three methodologies were developed using decision tree algorithms available at the following open source softwares: InterIMAGE, GeoDMA, WEKA and QGIS. We selected the support vector machine algorithm (SVM) available in Orfeo

ToolBox Monteverdi software for both region-based and pixel-based classification. We also evaluate the SPT software in order to obtain the ideal set for segmentation parameters. Results show that the classification of secondary successional forest stages as well as other land use classes performed well. Kappa index ranged from 0.60 to 0.89. Conditional accuracy and both producer's and user's accuracy were higher than 0.5. The best overall accuracy results were found for initial forest stages while the worst performance was observed in the intermediate successional forest stage. Region based classification and OBIA performed better than *pixel by pixel* based classification. The generated classified maps reveal the applicability of this approaches for extracting useful information from the SAAPI images. Such information can also provide useful information for forest resources monitoring practices at the state level.

Keywords: Secondary forests. Aerophotogrammetric survey. Digital image processing. Data mining. Support vector machines.

LISTA DE ILUSTRAÇÕES

| | | |
|-----------|--|----|
| Figura 1 | - Distribuição das tipologias florestais do Bioma Mata Atlântica. | 38 |
| Figura 2 | - Mapa fitogeográfico de Santa Catarina. | 39 |
| Figura 3 | - Perfil esquemático de sucessão florestal, com predomínio de pioneiras (A); secundárias iniciais (B); secundárias tardias (C); e climácicas (D). | 41 |
| Figura 4 | - Exemplo de vegetação em estágio inicial de regeneração. | 44 |
| Figura 5 | - Exemplo de vegetação em estágio médio de regeneração. | 45 |
| Figura 6 | - Exemplo de vegetação em estágio avançado de regeneração. | 46 |
| Figura 7 | - Curva de reflectância típica de uma folha verde. | 52 |
| Figura 8 | - Comparação na diminuição resolução espacial de GeoEye (esquerda) para MODIS (à direita) e SPOT-5 (centro) em uma floresta decídua mista no Sul de Camarões. | 58 |
| Figura 9 | - Exemplo de imagem de alta resolução espacial classificada por <i>pixel</i> (a) e baseada em regiões (b). | 60 |
| Figura 10 | - Relacionamento entre objetos em relação à resolução espacial. | 61 |
| Figura 11 | - Exemplo de resultados satisfatório (a) e não satisfatórios (b,c) da segmentação. | 65 |
| Figura 12 | - Representação da função objetivo. | 68 |
| Figura 13 | - Avaliação de aptidão de uma segmentação por AG. | 69 |
| Figura 14 | - Representação de uma árvore de decisão. | 77 |
| Figura 15 | - Esquema de classificação por meio do SVM, em que os pontos e quadrados representam duas classes no espaço de atributos. | 80 |
| Figura 16 | - Localização das áreas de estudo. | 83 |

| | | |
|-----------|--|-----|
| Figura 17 | - Localização da área-teste A..... | 84 |
| Figura 18 | - Localização da área-teste B..... | 85 |
| Figura 19 | - Localização da área-teste C..... | 86 |
| Figura 20 | - Fluxograma metodológico de todas as etapas desenvolvidas..... | 91 |
| Figura 21 | - Fluxograma das etapas envolvendo classificação por árvores de decisão..... | 96 |
| Figura 22 | - Metodologia implementada no SPT..... | 99 |
| Figura 23 | - Redes semânticas elaboradas para cada área-teste no InterIMAGE..... | 102 |
| Figura 24 | - Implementação do modelo WEKA no QGIS. ... | 110 |
| Figura 25 | - Funcionamento do <i>plug-in</i> GeoDMA..... | 112 |
| Figura 26 | - Resultado da segmentação pelo GeoDMA. Visualizam-se os limites entre os blocos. | 113 |
| Figura 27 | - Fluxograma das etapas desenvolvidas no OTB Monteverdi..... | 116 |
| Figura 28 | - Parâmetros do algoritmo SVM na classificação por regiões..... | 119 |
| Figura 29 | - Segmentos de referência; resultados da segmentação com escolha visual e automática dos parâmetros do algoritmo Baatz e Schäpe..... | 127 |
| Figura 30 | - Áreas-teste A e B (esquerda) e resultados da segmentação com parâmetros fornecidos pelo SPT (direita)..... | 130 |
| Figura 31 | - Resultado de quatro cenários de segmentação avaliados para as áreas-teste A e B..... | 131 |
| Figura 32 | - Perfil espectral das classes de uso e cobertura da terra da área-teste B. | 135 |
| Figura 33 | - Espacialização da B1 (a) e B4 (b) na área-teste B..... | 137 |
| Figura 34 | - Comportamento espectral das classes de uso e cobertura da terra em função do MDS na área-teste B..... | 138 |

| | | |
|-----------|---|-----|
| Figura 35 | - Demonstração gráfica da árvore de decisão gerada pelo algoritmo SimpleCart para a área-teste B no WEKA..... | 140 |
| Figura 36 | - Mapa de uso e cobertura da terra da área-teste A gerado no <i>software</i> InterIMAGE com algoritmo C4.5. | 144 |
| Figura 37 | - Mapa de uso e cobertura da terra da área-teste A gerado a partir do modelo do WEKA com o algoritmo SimpleCart no <i>software</i> QGIS. | 145 |
| Figura 38 | - Mapa de uso e cobertura da terra da área-teste A gerado no <i>software</i> GeoDMA com o algoritmo C4.5. | 145 |
| Figura 39 | - Principais confusões ocorridas na classificação da área-teste A, sendo: (a) imagem original; (b) classificação pelo GeoDMA; (c) classificação pelo InterIMAGE. | 149 |
| Figura 40 | - Mapa de uso e cobertura da terra da área-teste B gerado no <i>software</i> InterIMAGE com o algoritmo C4.5. | 150 |
| Figura 41 | - Mapa de uso e cobertura da terra da área-teste B gerado a partir do modelo do WEKA com o algoritmo SimpleCart no <i>software</i> QGIS. | 150 |
| Figura 42 | - Mapa de uso e cobertura da terra da área-teste B gerado no <i>software</i> GeoDMA com o algoritmo C4.5. | 151 |
| Figura 43 | - Principais confusões das classificações da área-teste B, sendo: (a) imagem original; (b) classificação pelo modelo do WEKA; (c) classificação pelo GeoDMA..... | 154 |
| Figura 44 | - Mapa de uso e cobertura da terra da área-teste C gerado no <i>software</i> InterIMAGE com o algoritmo C4.5. | 155 |
| Figura 45 | - Mapa de uso e cobertura da terra da área-teste C gerado no <i>software</i> InterIMAGE, sem o MDS com o algoritmo C4.5..... | 155 |

| | | |
|-----------|--|-----|
| Figura 46 | - Mapa de uso e cobertura da terra da área-teste C gerado a partir do modelo do WEKA com o algoritmo SimpleCart no <i>software</i> QGIS..... | 156 |
| Figura 47 | - Mapa de uso e cobertura da terra da área-teste C gerado no <i>software</i> GeoDMA com o algoritmo C4.5..... | 156 |
| Figura 48 | - Principais confusões ocorridas na classificação da área-teste C; (a) imagem original; (b) classificação com modelo do WEKA; (c) classificação pelo InterIMAGE..... | 159 |
| Figura 49 | - Mapa de uso e cobertura da terra da área-teste A gerado no <i>software</i> OTB Monteverdi com o algoritmo SVM. | 163 |
| Figura 50 | - Principais confusões da classificação por regiões da área-teste A no OTB Monteverdi..... | 165 |
| Figura 51 | - Mapa de uso e cobertura da terra da área-teste B gerado no <i>software</i> OTB Monteverdi com o algoritmo SVM. | 166 |
| Figura 52 | - Mapa de uso e cobertura da terra da área-teste C gerado no <i>software</i> OTB Monteverdi com o algoritmo SVM. | 167 |
| Figura 53 | - Mapa de uso e cobertura da terra da área-teste B gerado no <i>software</i> OTB Monteverdi através de metodologia <i>pixel-a-pixel</i> com o algoritmo SVM..... | 168 |
| Figura 54 | - Gráfico com o Kappa condicional do usuário da classificação pelo algoritmo SVM por objeto e por <i>pixel</i> realizada no OTB Monteverdi na área-teste B..... | 170 |
| Figura 55 | - Principais erros ocorridos na classificação da área-teste B pelo algoritmo SVM, sendo: (a) imagem original; (b) classificação por regiões; (c) classificação <i>pixel-a-pixel</i> | 171 |
| Figura 56 | - Kappa global obtido nas metodologias de acordo com a área-teste. | 179 |

LISTA DE TABELAS

| | | |
|-----------|--|-----|
| Tabela 1 | - Legislações federais que dispõem sobre os estádios sucessionais da vegetação..... | 42 |
| Tabela 2 | - Características das ortoimagens utilizadas..... | 87 |
| Tabela 3 | - Dados de entrada das áreas-teste e respectivas bandas..... | 88 |
| Tabela 4 | - Programas utilizados e finalidade..... | 89 |
| Tabela 5 | - Parâmetros selecionados visualmente para a criação dos diferentes cenários de segmentação com algoritmo de Baatz e Schäpe..... | 98 |
| Tabela 6 | - Número de amostras (objetos) e respectiva classe coletadas no InterIMAGE..... | 103 |
| Tabela 7 | - Amostras (objetos) e respectiva classe coletadas no GeoDMA, separadas automaticamente entre treinamento e validação..... | 114 |
| Tabela 8 | - Parâmetros avaliados visualmente para segmentação pela média-móvel..... | 117 |
| Tabela 9 | - Número de amostras (segmentos) coletadas para a classificação por regiões no Monteverdi..... | 117 |
| Tabela 10 | - Número de amostras (<i>pixels</i>) para cada classe na área-teste B..... | 120 |
| Tabela 11 | - Categorização de uma classificação a partir do valor do índice Kappa..... | 122 |
| Tabela 12 | - Resultado dos parâmetros de segmentação de Baatz e Schäpe avaliados pelo SPT em cada área-teste..... | 125 |
| Tabela 13 | - Parâmetros utilizados na segmentação pelo algoritmo de Baatz e Schäpe..... | 129 |
| Tabela 14 | - Parâmetros escolhidos para a segmentação pelo algoritmo de média móvel..... | 132 |
| Tabela 15 | - Número de atributos selecionados e mérito dos subconjuntos selecionados pelo algoritmo CFS no WEKA..... | 132 |

| | | |
|-----------|---|-----|
| Tabela 16 | - Atributos selecionados pelo algoritmo CFS no WEKA. | 133 |
| Tabela 17 | - Resultado dos modelos de classificação por árvore de decisão testados no WEKA..... | 139 |
| Tabela 18 | - Características principais das árvores de decisão geradas em cada <i>software</i> | 141 |
| Tabela 19 | - Área e porcentagem correspondente a cada classe nos mapas de referência das áreas de estudo. | 143 |
| Tabela 20 | - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo InterIMAGE. | 146 |
| Tabela 21 | - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo WEKA. | 146 |
| Tabela 22 | - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo GeoDMA. | 146 |
| Tabela 23 | - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo InterIMAGE.. | 151 |
| Tabela 24 | - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo WEKA. | 152 |
| Tabela 25 | - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo GeoDMA. | 152 |
| Tabela 26 | - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo InterIMAGE.. | 157 |
| Tabela 27 | - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo InterIMAGE, sem a banda MDS. | 157 |
| Tabela 28 | - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo modelo do WEKA. | 157 |
| Tabela 29 | - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo GeoDMA. | 158 |
| Tabela 30 | - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo OTB Monteverdi..... | 164 |

| | | |
|-----------|---|-----|
| Tabela 31 | - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo OTB Monteverdi. | 166 |
| Tabela 32 | - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo OTB Monteverdi. | 167 |
| Tabela 33 | - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo OTB Monteverdi (<i>pixel-a-pixel</i>). | 169 |
| Tabela 34 | - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste A. | 173 |
| Tabela 35 | - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste A. | 174 |
| Tabela 36 | - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste B. | 175 |
| Tabela 37 | - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste B. | 175 |
| Tabela 38 | - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste C. | 176 |
| Tabela 39 | - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste C. | 178 |

LISTA DE QUADROS

| | |
|---|----|
| Quadro 1 - Descrição das classes encontradas nas áreas de estudo..... | 92 |
|---|----|

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------------|--|
| AG | Algoritmos genéticos |
| ARFF | <i>Attribute-Relation File Format</i> (Formato de Arquivo Atributo-Relação) |
| ASCII | <i>American Standard Code for Information Interchange</i> (Código Americano Padrão para Intercâmbio de Informação) |
| BPMA | Batalhão de Polícia Militar Ambiental |
| CART | <i>Classification and Regression Trees</i> (Árvores de Regressão e Classificação) |
| CCD | <i>Charge Coupled Device</i> (Dispositivo de Carga Acomplada) |
| CFS | Correlation-based Feature Selection (Seleção de Atributos baseada em Correlação) |
| CONAMA | Conselho Nacional do Meio Ambiente |
| DAP | Diâmetro a altura do peito |
| FATMA | Fundação do Meio Ambiente do estado de Santa Catarina |
| FBR | Função de Base Radial |
| FOD | Floresta Ombrófila Densa |
| FOM | Floresta Ombrófila Mista |
| FOSS | <i>Free and Open Source Software</i> (Programa Gratuito e de Código Aberto) |
| GeoDMA | <i>Geographical Data Mining Analyst</i> (Analista para Mineração de Dados Geográficos) |
| GIS | <i>Geographic Information System</i> (Sistema de Informações Geográficas) |
| GLCM | <i>Grey Level Co-occurrence Matrix</i> (Matriz de Co-Ocorrência de Nível de Cinza) |
| IBAMA | Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis |
| INPE | Instituto Nacional de Pesquisas Espaciais |
| InterIMAGE | <i>Interpreting Images Freely</i> (Interpretando Imagens Gratuitamente) |

| | |
|--------|---|
| Kcp | Kappa condicional do produtor |
| Kcu | Kappa condicional do usuário |
| MaxVer | Máxima Verossimilhança |
| MDS | Modelo Digital de Superfície |
| MDT | Modelo Digital de Terreno |
| MMA | Ministério do Meio Ambiente |
| MODIS | <i>Moderate Resolution Imaging Spectroradiometer</i> (Espectorradiômetro Imageador de Resolução Moderada) |
| ND | Números Digitais |
| NDVI | <i>Normalized Vegetation Difference Index</i> (Índice de Vegetação por Diferença Normalizada) |
| NIR | <i>Near infrared</i> (Infravermelho próximo) |
| OBIA | <i>Object-Based Image Analysis</i> (Análise de Imagens Baseada em Objeto) |
| OSGeo | <i>Open Source Geospatial Foundation</i> (Fundação Geoespacial para Código Aberto) |
| OTB | <i>Orfeo ToolBox</i> (Caixa de Ferramentas Orfeo) |
| PEC | Padrão de Exatidão Cartográfica |
| PUC | Pontifícia Universidade Católica |
| RBSB | <i>Reference Bounded Segments Booster</i> (Assistente para Segmentos Delimitados como Referência) |
| REM | Radiação Eletromagnética |
| RMA | Rede de ONGs da Mata Atlântica |
| ROIs | Regiões de Interesse |
| SAAPI | Sistema Aerotransportado de Aquisição e Pós-processamento de Imagens Digitais |
| SC | Estado de Santa Catarina |
| SDS | Secretaria de Estado do Desenvolvimento Sustentável |
| SIRGAS | Sistema de Referência Geocêntrico para as Américas |
| SPT | <i>Segmentation Parameters Tuner</i> (Ajustador de Parâmetros de Segmentação) |

| | |
|------|--|
| SVM | <i>Support Vector Machine</i> (Máquina de Vetor de Suporte) |
| VEA | Vegetação em Estádio Avançado |
| VEI | Vegetação em Estádio Inicial |
| VEM | Vegetação em Estádio Médio |
| WEKA | <i>Waikato Environment for Knowledge Analysis</i> (Ambiente Waikato para Análise do Conhecimento) |

SUMÁRIO

| | |
|---|-----------|
| 1. INTRODUÇÃO | 31 |
| 1.1 OBJETIVOS | 35 |
| 1.1.1 Objetivo Geral..... | 35 |
| 1.1.2 Objetivos Específicos | 35 |
| 1.2 ORGANIZAÇÃO DA DISSERTAÇÃO | 36 |
| 2. FUNDAMENTAÇÃO TEÓRICA | 37 |
| 2.1 BIOMA MATA ATLÂNTICA | 37 |
| 2.1.1 Tipologias florestais de Santa Catarina | 39 |
| 2.1.2 Estádios sucessionais no contexto da Legislação Federal | 40 |
| 2.2 SENSORIAMENTO REMOTO | 48 |
| 2.2.1 A radiação eletromagnética e a vegetação | 49 |
| 2.2.2 Imagens aéreas e aplicabilidade de imagens de alta resolução espacial no estudo da vegetação | 54 |
| 2.2.3 Métodos de classificação digital..... | 58 |
| 2.2.4 Segmentação | 63 |
| 2.2.4.1 Algoritmos genéticos..... | 65 |
| 2.2.4.2 Algoritmo de segmentação por crescimento de regiões de Baatz e Schäpe | 70 |
| 2.2.4.3 Algoritmo de segmentação por média-móvel . | 71 |
| 2.2.5 Mineração de dados | 72 |
| 2.2.5.1 Seleção de atributos..... | 74 |
| 2.2.5.2 Árvores de decisão | 75 |
| 2.2.6 Máquina de vetor de suporte (Support Vector Machine- SVM)..... | 80 |
| 3. MATERIAL E MÉTODOS..... | 83 |
| 3.1 ÁREA DE ESTUDO | 83 |
| 3.2 MATERIAL | 86 |
| 3.2.1 Imagens do recobrimento aéreo do estado de Santa Catarina | 86 |
| 3.2.2 Programas utilizados | 89 |

| | |
|---|------------|
| 3.3 CLASSIFICAÇÃO DO USO E COBERTURA DA TERRA | 90 |
| 3.3.1 Definição e caracterização das classes de uso e cobertura da terra | 91 |
| 3.3.2 Classificação por árvore de decisão | 95 |
| 3.3.2.1 Escolha e ajuste dos parâmetros da segmentação..... | 96 |
| 3.3.2.2 Classificação no InterIMAGE | 101 |
| 3.3.2.3 Seleção de atributos e mineração de dados no WEKA | 106 |
| 3.3.2.4 Classificação no QGIS..... | 109 |
| 3.3.2.5 Classificação no GeoDMA | 110 |
| 3.3.3 Classificação por máquinas de vetor de suporte | 115 |
| 3.3.3.1 Classificação por regiões | 116 |
| 3.3.3.2 Classificação pixel-a-pixel..... | 119 |
| 3.4 AVALIAÇÃO DOS RESULTADOS | 121 |
| 4. RESULTADOS E DISCUSSÃO..... | 125 |
| 4.1 SEGMENTAÇÃO DAS IMAGENS | 125 |
| 4.1.1 Segmentação e ajuste dos parâmetros do algoritmo de Baatz e Schäpe | 125 |
| 4.1.2 Segmentação com o algoritmo de média-móvel..... | 129 |
| 4.2 MINERAÇÃO DE DADOS | 132 |
| 4.2.1 Seleção de atributos e descrição das classes de uso e cobertura da terra..... | 132 |
| 4.2.2 Modelos de classificação por árvores de decisão | 138 |
| 4.3 RESULTADOS E AVALIAÇÃO DAS CLASSIFICAÇÕES DO USO E COBERTURA DA TERRA | 143 |
| 4.3.1 Resultado da classificação por árvores de decisão | 144 |

| | | |
|--------------|---|------------|
| 4.3.1.1 | Análise das classificações por árvore de decisão..... | 160 |
| 4.3.2 | Resultado da classificação por máquinas de vetor de suporte..... | 162 |
| 4.3.2.1 | Classificação por regiões..... | 163 |
| 4.3.2.2 | Classificação <i>pixel-a-pixel</i> | 168 |
| 4.3.2.3 | Análise das classificações pelo algoritmo SVM | 169 |
| 4.4 | COMPARAÇÃO ENTRE AS METODOLOGIAS ... | 173 |
| 5. | CONSIDERAÇÕES FINAIS | 185 |
| 5.1 | RECOMENDAÇÕES PARA TRABALHOS FUTUROS | 192 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 195 |
| | APÊNDICES..... | 237 |
| | APÊNDICE A- Atributos gerados no InterIMAGE. | 237 |
| | APÊNDICE B- Atributos gerados no GeoDMA..... | 239 |
| | APÊNDICE C- Descrição dos atributos exportados no InterIMAGE e GeoDMA..... | 242 |
| | APÊNDICE D- Árvores de decisão geradas pelo algoritmo C4.5 no InterIMAGE para cada área-teste. | 245 |
| | APÊNDICE E- Árvores de decisão geradas pelo algoritmo SimpleCart no WEKA para cada área-teste. | 247 |
| | APÊNDICE F- Árvores de decisão geradas pelo algoritmo C4.5 no GeoDMA para cada área-teste. | 248 |

1. INTRODUÇÃO

Com a intensificação das ações antrópicas, as fitofisionomias florestais do estado de Santa Catarina foram submetidas a um intenso processo exploratório. Citam-se como exemplos deste processo a extração de madeira, supressão e queima da vegetação para implantação da agricultura, pecuária e núcleos urbanos (RMA, 2006). Os remanescentes alterados dessas florestas, anteriormente exuberantes, atualmente formam mosaicos contendo diferentes estádios sucessionais com composição florística e estrutura da vegetação diversificada (SANQUETTA, 2005).

Neste cenário, o mapeamento e a classificação dos remanescentes florestais da Mata Atlântica, e seus estádios de sucessão ecológica, estrutura, distribuição e evolução temporal, consistem em uma etapa fundamental para implantação de diversos estudos, ações de fiscalização e manejo ambiental. A proposição de metodologias que contemplem essa temática permite avaliar quantitativamente e qualitativamente as áreas de floresta nativa preservadas, bem como sua distribuição espacial. Essas informações podem subsidiar estudos de fitossociologia, florística, recuperação, educação, proteção e fiscalização ambiental, além do estabelecimento de áreas prioritárias para conservação, entre outros (AMARAL et al., 2009). Podem servir também como base para averbações e adequações de propriedades rurais como, por exemplo, demarcação e monitoramento de áreas de preservação permanente e de reserva legal.

Os mapas de uso e cobertura da terra são importantes para monitorar, entender e prever os efeitos da complexa interação homem-natureza em escala local, regional e global (CLARK et al., 2010). No entanto, os métodos de mapeamento da sucessão vegetal através de trabalhos de campo são dispendiosos em tempo e pessoal envolvido. Os custos resultantes costumam ser elevados e os intervalos temporais

entre os mapeamentos geralmente são grandes, além de irregulares (ANDRESEN et al., 2007). Surge então uma demanda de investigação de novas técnicas e metodologias, que sejam eficientes no mapeamento da cobertura vegetal, vindo aperfeiçoar as técnicas tradicionais de inventário em campo (NAVE, 1999; CINTRA, 2007).

Nesse contexto, as técnicas de Sensoriamento Remoto têm se destacado, principalmente no mapeamento de grandes extensões geográficas e em locais de difícil acesso. Estas técnicas são viáveis e menos custosas que os trabalhos a campo (JANOTH et al., 2007; BARBIER et al., 2011). Os métodos de classificação foram desenvolvidos, visando, especialmente, ao mapeamento da cobertura e ocupação da terra. A classificação pode ser manual, em que o fotointérprete extrai as feições de interesse sobre uma imagem, ou de forma automática ou semiautomática, com algoritmos especializados, maximizando a extração de informações através da simulação da fotointerpretação humana (FLORENZANO, 2002).

Em geral, os trabalhos envolvendo a classificação da cobertura da terra utilizam-se de dados remotamente situados para a extração de classes que possuem características espectrais distintas, como, por exemplo: área urbana, campo, vegetação, corpos d'água, dentre outros. Porém, no mapeamento de estádios de sucessão florestal, as classes espectrais são semelhantes, tornando o processo de classificação das diferentes fases de desenvolvimento florestal uma tarefa desafiadora (VIEIRA et al., 2003). Freitas e Shimabukuro (2007) citam essa dificuldade, descrevendo que, apesar de ser uma tarefa comum diferenciar áreas florestais de outros tipos de cobertura e uso da terra, é difícil discriminar diferentes tipologias florestais. Por isso, a importância da busca de técnicas e metodologias de classificação que contemplem atributos relacionados à forma do alvo, à textura ou relações entre alvos vizinhos no mapeamento das fases de sucessão florestal.

O advento de novos sistemas sensores com melhores qualidades radiométrica, espectral, temporal e espacial exigiu que fossem exploradas novas metodologias para extrair informações desses dados. Segundo Pinho (2005), a utilização de imagens com alta resolução espacial impõe novos desafios para a classificação automática de uso e cobertura da terra. Devido à grande heterogeneidade em detalhes e maior diversidade intraclasse, torna-se inviável a aplicação de técnicas tradicionais de classificação automática que levem em conta apenas a resposta espectral dos alvos, como a convencional classificação *pixel-a-pixel* (RODRIGUES, 2014). Diante disso, a classificação baseada em regiões, em que os *pixels* com características semelhantes são agrupados em segmentos antes da classificação, tornou-se um método mais viável para classificação de imagens de alta resolução espacial. Contudo, apesar de esta técnica ter superado alguns problemas encontrados na classificação dessas imagens, nas últimas décadas tem se destacado outro método, denominado análise baseada em objeto (*Object-Based Image Analysis*- OBIA) que, não só é baseado em regiões, mas também pressupõe a existência de um modelo de conhecimento atrelado ao processo de interpretação da cena (FRANCISCO e ALMEIDA, 2012).

Os sistemas de análise de imagens do tipo OBIA têm sido amplamente utilizados na classificação de uso e cobertura da terra em imagens de alta resolução espacial (RIBEIRO et al., 2011). Esta abordagem foi pesquisada e disseminada a partir da sua incorporação a sistemas comerciais, principalmente o eCognition (BLASCHKE, 2009), que foi o primeiro programa comercial a implementá-la (CARVALHO, 2011). Blaschke (2010) identificou que cerca de 50% a 55% de trabalhos publicados remeteram a este aplicativo.

Seguindo o eCognition, e respondendo às crescentes demandas dos usuários, outros *softwares* comerciais como ENVI e Erdas IMAGINE integraram a extração de feições e componentes de análise em seus pacotes. No entanto, o alto

custo desses pacotes comerciais é um grande obstáculo para muitos pesquisadores e praticantes (VU, 2012). Mais recentemente, sistemas *open source* voltados à OBIA têm se destacado, entre eles *softwares* brasileiros como o *Interpreting Images Freely* (InterIMAGE) (COSTA et al., 2008) e o *Geographic Data Mining Analyst* (GeoDMA) (KÖRTING, 2012), desenvolvidos respectivamente pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) e Instituto Nacional de Pesquisas Espaciais (INPE). Cita-se também o *software* francês *Orfeo ToolBox* (OTB), desenvolvido pela Agência Espacial Francesa (CHRISTOPHE e INGLADA, 2009), que, apesar de não implementar a OBIA, fornece uma interface voltada para classificação de imagens de alta resolução espacial.

Abordagens utilizando a classificação por regiões e a OBIA têm sido utilizadas no estudo da vegetação (YU et al., 2006; AMARAL et al., 2009; ROSÁRIO et al., 2009; SOUSA et al., 2010; FURTADO et al., 2013; PIAZZA, 2014) podendo se tornar importantes aliadas na análise de remanescentes florestais, como a classificação do estágio sucessional da vegetação. No Bioma Mata Atlântica, esta classificação se faz necessária, tanto para nortear o licenciamento de atividades que envolvem supressão da vegetação, quanto para fornecer subsídios suficientes para o correto enquadramento das infrações ambientais, em casos de cometimento de crimes, tendo em vista a legislação tratar cada estágio sucessional de forma diferenciada.

Santa Catarina é um estado que gera riquezas através da pequena propriedade rural. A legislação ambiental surge para subsidiar e regular o uso dos recursos naturais de forma que o pequeno produtor rural possa obter sua renda através da agricultura e pecuária, sem que precise destruir mais áreas cobertas por vegetação natural. Salienta-se, portanto, a relevância deste estudo como subsídio para o estado de Santa Catarina proteger o capital natural das florestas e, ao mesmo

tempo, permitir seu uso múltiplo. Com isso, pautar o desenvolvimento da economia por meio de políticas públicas que garantam o ordenamento territorial e o uso sustentável de seus recursos naturais, além da possibilidade de facilitar a fiscalização e controle dos órgãos ambientais quanto ao uso racional desses recursos naturais.

A partir dessas considerações iniciais, parte-se das seguintes hipóteses:

- a) a vegetação apresenta características espectrais, espaciais e texturais distintas conforme o estágio sucessional de regeneração;
- b) imagens de alta resolução espacial possibilitam a classificação do estágio sucessional da vegetação com precisão, de forma a auxiliar ou substituir trabalhos de campo;
- c) metodologias *open source* de classificação digital de imagens podem ser aplicadas para discriminar diferentes estádios de regeneração da vegetação.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Desenvolver metodologias automatizadas não-paramétricas com base em dados de Sensoriamento Remoto para classificação dos estádios sucessionais da vegetação de Floresta Ombrófila Mista em diferentes áreas de estudo no estado de Santa Catarina.

1.1.2 Objetivos Específicos

- a) Analisar o potencial da utilização de dados de alta resolução espacial obtidos por sensores aerotransportados na definição e diferenciação dos estádios sucessionais da vegetação.

- b) Definir e comparar parâmetros de segmentação das imagens em objetos de forma visual e automática.
- c) Identificar os melhores atributos para a caracterização de cada classe de uso e cobertura da terra por meio de mineração de dados.
- d) Explorar, testar e comparar diferentes sistemas *open source* para a classificação temática das imagens.
- e) Analisar a aplicabilidade das informações extraídas para as áreas de fiscalização, gestão, manejo e recuperação ambiental.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

Este trabalho está organizado em cinco capítulos:

- a) o Capítulo 1 se refere à introdução do trabalho;
- b) o Capítulo 2 apresenta a fundamentação teórica, que aborda os principais temas e conceitos relacionados a esse estudo, como imagens de alta resolução espacial, sensoriamento remoto no estudo da vegetação, análise de imagens baseada em objeto e mineração de dados;
- c) o Capítulo 3 apresenta a caracterização das áreas-teste, o material empregado e as metodologias desenvolvidas;
- d) no Capítulo 4, são apresentados, analisados e discutidos os resultados obtidos na classificação, bem como a avaliação da acurácia das classificações;
- e) o Capítulo 5 reúne as considerações finais e recomendações para trabalhos futuros.

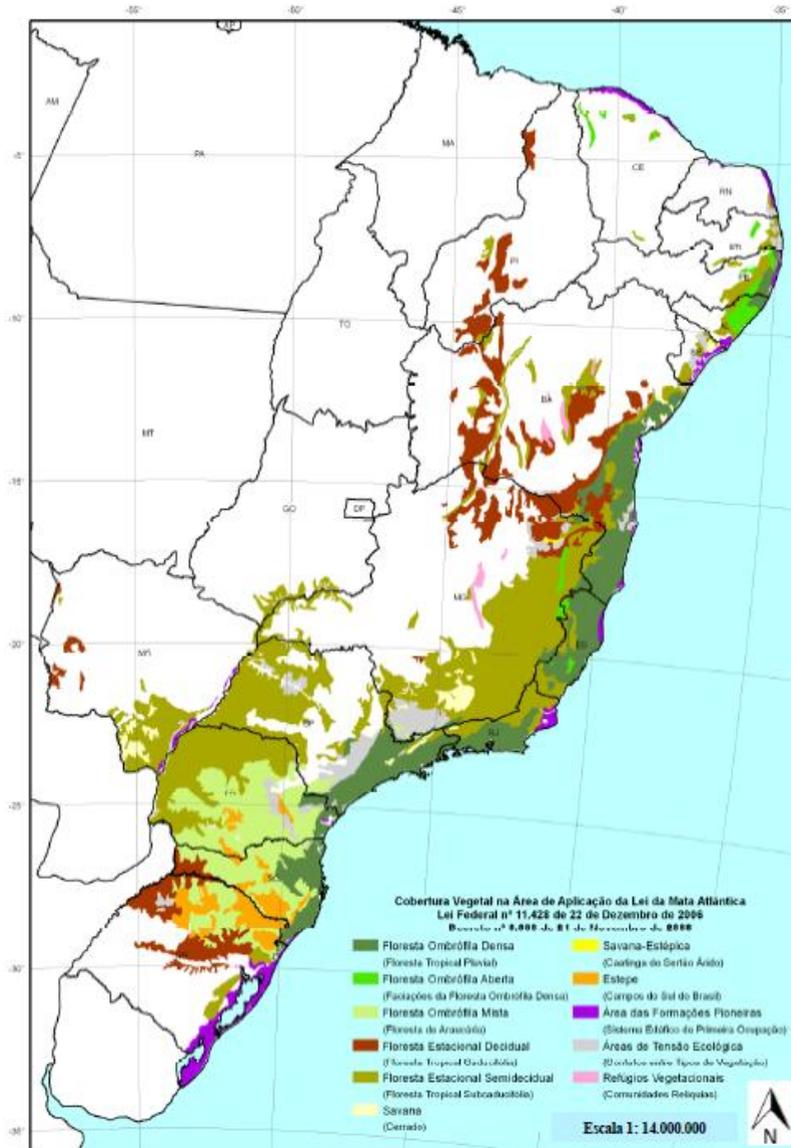
2. FUNDAMENTAÇÃO TEÓRICA

2.1 BIOMA MATA ATLÂNTICA

O Bioma Mata Atlântica e seus ecossistemas associados envolvem uma área de 1,1 milhão de km², abrangendo 13% do território brasileiro, contemplando 17 estados (MMA, 2010). Os remanescentes de vegetação nativa deste bioma estão reduzidos atualmente a aproximadamente 22% de sua cobertura original, sendo que estes se encontram altamente fragmentados e em diferentes estádios de regeneração. Somente 7% ainda estão bem conservados em fragmentos acima de 100 ha (MMA, 2010). Apesar disso, a Mata Atlântica ainda abriga parcela significativa de diversidade biológica do Brasil (SFB, 2010).

Esse bioma é composto por diversidade de formações florestais, como Floresta Ombrófila (densa, mista e aberta), Mata Estacional Semidecidual e Estacional Decidual, manguezais, restingas e campos de altitude associados e brejos interioranos no Nordeste (Figura 1) (SFB, 2010). A Mata Atlântica é também considerada um dos *hotspots* mundiais, ou seja, uma das 25 regiões biologicamente mais ricas e ameaçadas do planeta (IBAMA, 2001).

Figura 1 - Distribuição das tipologias florestais do Bioma Mata Atlântica.



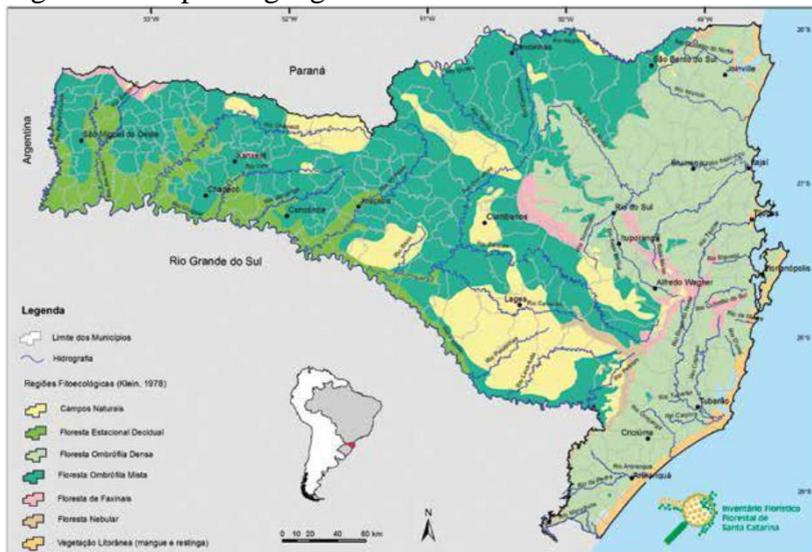
Fonte: Fundação SOS Mata Atlântica e INPE, 2015.

2.1.1 Tipologias florestais de Santa Catarina

Com uma extensão territorial de 95.985 km², dos quais 85%, ou 81.587 km², originalmente cobertos pela Mata Atlântica, Santa Catarina situa-se atualmente como o terceiro estado brasileiro com maior área de remanescentes da Mata Atlântica, resguardando cerca de 1.662.000 ha (16.620 km²), ou 17,46% da área original. Registra-se que a área do estado corresponde somente a 1,12% do território brasileiro (RBMA, 2008).

O mapa fitogeográfico de Klein (1978) é ainda utilizado como base para o estado de Santa Catarina (Figura 2). De acordo com este mapa, a cobertura florestal do estado está dividida em quatro regiões fitoecológicas: Floresta Ombrófila Mista (FOM), que cobria originalmente 43% da sua superfície, Estepe (14%), Floresta Ombrófila Densa (FOD) (30%), Floresta Estacional Decidual (8%) e outras formações, que incluem restingas e manguezais, (2%).

Figura 2 - Mapa fitogeográfico de Santa Catarina.



Fonte: Vibrans et al., 2012.

Shorn et al. (2012), em estudo sobre a estrutura dos remanescentes florestais de Santa Catarina, concluíram que a situação da Floresta Estacional Decidual e da FOM é mais crítica, pois possuem sua estrutura alterada e dominada por espécies secundárias e pioneiras. A FOD ainda mantém remanescentes mais bem conservados, entretanto sua estrutura dominada por espécies secundárias mostra indício de forte pressão antrópica (SCHORN et al., 2012).

Siminski (2004) utiliza o termo “mosaico” ao se referir à cobertura vegetal atual do estado, já que é formada por áreas de formações florestais secundárias em diferentes idades e, em escala reduzida, remanescentes de floresta primária alterada. No Relatório da Fundação SOS Mata Atlântica e INPE (2015) consta que, apesar de a maioria dos estados brasileiros terem diminuído a taxa de desmatamento entre os anos 2012 e 2014, Santa Catarina apresentou um aumento de 3% nessa taxa, com 692 ha de áreas desmatadas, estando em quinto lugar no *ranking* nacional.

De acordo com Siminski (2009), poucos estudos têm sido dedicados a compreender esse processo de redução de florestas remanescentes. Desta forma, fica evidente a necessidade de maior conhecimento sobre a distribuição espacial desses remanescentes, com vistas a proporcionar bases e diretrizes para definir estratégias de uso, restauração e conservação da vegetação.

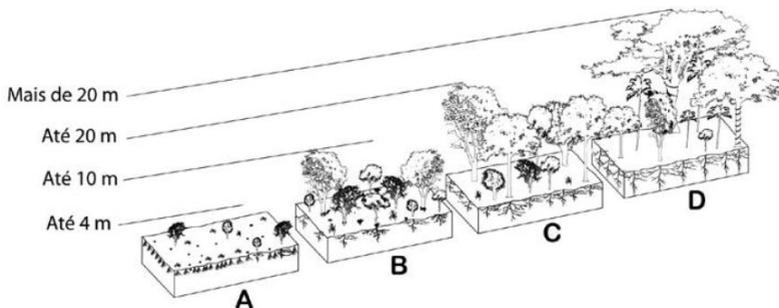
2.1.2 Estádios sucessionais no contexto da Legislação Federal

Antes que a vegetação de um determinado local alcance uma relativa estabilidade em suas características fisiológicas, estruturais e florísticas, ocorre uma série de mudanças nas comunidades. Este processo é denominado sucessão vegetal ou dinâmica (RODRIGUES, 1995). Esta dinâmica estrutural e

florística é causada por uma diversidade de fatores, acarretando heterogeneidade ambiental no espaço e no tempo em uma ampla escala de ação. A simples queda de um indivíduo arbóreo na floresta proporciona um aumento da complexidade estrutural e da diversidade de espécies do local. Outros distúrbios, como a exploração da floresta e sua fragmentação, incêndios, deslizamentos e enchentes, também são conhecidos por ter um forte efeito sobre a dinâmica florestal, e, conseqüentemente, sobre a composição e estrutura das árvores e suas comunidades (JOHNSON e MIYANISHI, 2007 apud NASCIMENTO et al., 2012).

Segundo Odum (1988), quando o processo de formação de uma comunidade se dá sobre um substrato parcialmente desocupado denomina-se sucessão primária, enquanto que aquele que começa em um local anteriormente ocupado por uma comunidade, é denominado sucessão secundária. Assim, as espécies pioneiras são as primeiras a serem instaladas no sistema, dando condições para a introdução das secundárias iniciais e tardias e, na sequência, o estabelecimento das espécies climácicas, as árvores das florestas maduras (Figura 3) (CASTRO, 2012).

Figura 3 - Perfil esquemático de sucessão florestal, com predomínio de pioneiras (A); secundárias iniciais (B); secundárias tardias (C); e climácicas (D).



Fonte: Castro, 2012.

Nesse contexto, os estádios sucessionais são tentativas de categorização das fases de sucessão secundária, e se orientam, essencialmente, pela mudança da composição de espécies e da estrutura do ecossistema que ocorrem ao longo do tempo em uma determinada área (ODUM, 1988). A separação do processo de sucessão secundária em fases ou estádios sucessionais distintos, mesmo arbitrário, é uma estratégia utilizada na busca do entendimento sobre a dinâmica funcional das florestas (KAGEYAMA et al., 1986).

A classificação da vegetação em estádios sucessionais na Mata Atlântica ganhou importância para o licenciamento e fiscalização ambiental quando as modalidades de uso e norma para a supressão da vegetação foram regulamentadas pela legislação de forma diferenciada para cada estágio sucessional (Tabela 1) (SEVEGNANI et al., 2012).

Tabela 1 - Legislações federais que dispõem sobre os estádios sucessionais da vegetação. (Continua)

| Legislação | Descrição |
|---|--|
| Resolução CONAMA 10/1993 (BRASIL, 1993): | Estabelece os parâmetros básicos para análise dos estádios de sucessão da Mata Atlântica. |
| Resolução CONAMA 04/1994 (BRASIL, 1994) | Define vegetação primária e secundária nos estádios inicial, médio e avançado de regeneração da Mata Atlântica, a fim de orientar os procedimentos de licenciamento de atividades florestais em Santa Catarina. |
| Resolução CONAMA 388/2007 (BRASIL, 2007) | Dispõe sobre a convalidação das Resoluções que definem a vegetação primária e secundária nos estádios inicial, médio e avançado de regeneração da Mata Atlântica para fins do disposto no art. 4o § 1o da Lei no 11.428, de 22 de dezembro de 2006. |
| Lei 11.428/2006 (BRASIL, 2006) | Dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica, e dá outras providências. Essa Lei coloca em seu Art. 8º que o corte, a supressão e a exploração da vegetação do Bioma Mata Atlântica far-se-ão de maneira diferenciada, |

Tabela 1 - Legislações federais que dispõem sobre os estádios sucessionais da vegetação. (Conclusão)

| Legislação | Descrição |
|---|---|
| | conforme se trate de vegetação primária ou secundária, nesta última levando-se em conta o estágio de regeneração. |
| Decreto nº 6.660/2008 (BRASIL, 2008) | Regulamenta dispositivos da Lei nº 11.428, de 22 de dezembro de 2006, que dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica. |

Fonte: produção do próprio autor, 2015.

A Resolução CONAMA nº 04/1994 (BRASIL, 1994) determina a observação de um conjunto de critérios e indicadores qualitativos e quantitativos (estruturais) para a caracterização da vegetação primária e secundária nos estádios inicial, médio e avançado de regeneração para o estado de Santa Catarina:

- a) Estádio inicial de regeneração: também chamado de estágio pioneiro, surge logo após o abandono do solo. A vegetação é dominada por arbustos e ervas, possuindo menor diversidade de espécies. Este estágio geralmente dura entre seis e dez anos, dependendo do grau de degradação do solo e do entorno. A altura média da vegetação não ultrapassa 4 m, e o diâmetro médio na altura do peito (DAP) é de no máximo 8 cm, conforme parâmetros definidos na Resolução. Em geral, possui pequeno número de espécies lenhosas, baixa complexidade da estrutura da vegetação, epífitos vasculares ausentes ou raros, trepadeiras herbáceas, serapilheira em camada fina (Figura 4).

Figura 4 - Exemplo de vegetação em estágio inicial de regeneração.



Fonte: produção do próprio autor, 2015.

- b) Estádio médio de regeneração: este estágio pode ocorrer entre seis e quinze anos depois do abandono da área. A diversidade aumenta, existe maior quantidade de arvoretas e arbustos do que herbáceas, a predominância é de espécies de árvores pioneiras. O estágio médio caracteriza-se por possuir, em geral, amplitude diamétrica moderada, podendo variar de 8 a 15 cm de DAP e altura até 12 m. Domínio de espécies lenhosas, epífitos vasculares presentes em maior número que o estágio anterior, trepadeiras lenhosas com pequeno diâmetro e serapilheira possuindo camada variada. Trata-se de estágio em transição entre o inicial e avançado (Figura 5).

Figura 5 - Exemplo de vegetação em estágio médio de regeneração.



Fonte: produção do próprio autor, 2015.

- c) Estádio avançado de regeneração: Nesse estágio da vegetação, a fisionomia arbórea é dominante sobre as demais, formando um dossel fechado e relativamente uniforme, apresentando árvores dominantes. A altura média da vegetação é superior a 12 m, e o DAP médio varia de 15 a 25 cm. Nessa fisionomia, há o predomínio de espécies lenhosas, com complexidade na estrutura da vegetação e grande diversidade de espécies, epífitos vasculares presentes em abundância, trepadeiras lenhosas bem desenvolvidas e serapilheira em camada espessa. Na maioria das vezes, a ocorrência de estágio avançado permite inferir que estas áreas já sofreram algum tipo de interferência antrópica que alterou suas características originais, não sendo caracterizada mais como floresta primária (SETTE, 2009). A Figura 6 ilustra o estágio avançado de regeneração, composto predominantemente por espécies climácias e dossel uniforme.

Figura 6 - Exemplo de vegetação em estágio avançado de regeneração.



Fonte: produção do próprio autor, 2015.

De acordo com a legislação ambiental vigente, qualquer intervenção no ecossistema deve ser precedida de autorização prévia dos órgãos ambientais, estando as possibilidades de intervenção condicionadas pelo estágio sucessional da formação florestal. Nesse cenário, os casos que demandam maior atenção dos órgãos são os de supressão da vegetação, já que em grande parte implicam a mudança no uso da terra (SIMINSKI e FANTINI, 2010).

Em Santa Catarina, as autorizações para supressão de vegetação nativa são concedidas pela Fundação de Meio Ambiente de SC (FATMA), que avalia previamente a área do proponente a fim de analisar pedidos de autorização ambiental. Do ponto de vista técnico, o principal requisito dessa etapa é a determinação do estágio de sucessão da vegetação, que deve ser definida por inventário florestal, realizado por profissional habilitado, com base no que é estabelecido pela Resolução nº 04/1994 do CONAMA (BRASIL, 1994).

Outra situação que se deve atentar é quando ocorre a supressão da vegetação sem autorização prévia dos órgãos ambientais competentes. Nesses casos, incorre-se em crime

ambiental previsto na Lei nº 9.605/98 (BRASIL, 1998), e/ou infração administrativa ambiental prevista no Decreto Federal nº 6.514/08 (BRASIL, 2008). Quando se flagra uma das situações mencionadas, também se faz necessário um inventário florestal da área atingida e seu entorno, a fim de caracterizar o estágio sucessional da vegetação para embasar o Auto de Infração Ambiental, pois o enquadramento será diferenciado conforme o estágio. Na esfera estadual, o Auto de Infração Ambiental pode ser lavrado tanto pela FATMA, quanto pela Polícia Militar Ambiental, conforme a Portaria nº 170/2013/GABP-FATMA/BPMA-SC (SANTA CATARINA, 2013). Esta portaria regula os procedimentos para apuração de infrações ambientais por condutas e atividades lesivas ao meio ambiente no âmbito da FATMA e do Batalhão de Polícia Militar Ambiental (BPMA).

Siminski e Fantini (2004) relatam que a falta de normatização para a amostragem em campo da vegetação, no que diz respeito ao inventário florestal exigido para esta caracterização, compromete os valores limites das variáveis estabelecidos pela referida resolução para diferenciar os estágios de regeneração. Devido a Resolução CONAMA nº 04/94 não citar um diâmetro mínimo de inclusão de plantas no inventário para classificar a vegetação pode acarretar no fato de áreas com características fitofisionômicas semelhantes serem enquadradas em estágios de regeneração diferentes (SIMINSKI, 2009). Jaster (2002) menciona que a classificação dos estágios ocorre principalmente de forma subjetiva, fortemente baseada na experiência do técnico.

Apesar desses problemas, a classificação da Resolução CONAMA nº 04/94 ainda é habitualmente utilizada para fins de fiscalização ambiental no que tange à legislação ambiental vigente, também adotada por pesquisadores que trabalham na área de Sensoriamento Remoto, por ser uma forma prática e objetiva de se definirem os estágios sucessionais.

2.2 SENSORIAMENTO REMOTO

O Sensoriamento Remoto pode ser definido como uma Ciência de obtenção de dados dos objetos da superfície terrestre sem que haja um contato físico direto de qualquer espécie entre o sensor e o objeto (MENESES e ALMEIDA, 2012). Envolve a detecção, aquisição e análise (interpretação e extração de informações) da energia eletromagnética emitida ou refletida pelos objetos terrestres e registradas por sensores terrestres ou a bordo de plataformas aerotransportadas ou orbitais (NOVO, 2008). Dessa forma, o Sensoriamento Remoto dispõe uma visão consistente e repetitiva da Terra, monitorando a longo e curto prazo as mudanças e o impacto das atividades humanas (SCHOWENGERDT, 2007).

No Sensoriamento Remoto, a função de um sensor remoto consiste em captar e medir a quantidade de energia refletida e/ou emitida por alvos e obter informações sobre a natureza e/ou condições desses alvos, de maneira a associá-los com a verdade terrestre (CENTENO, 2004). Essas informações são obtidas através da interação da radiação eletromagnética emitida por fontes naturais (Sol) ou artificiais (Radar; sensores ativos) com esses alvos (FLORENZANO, 2002).

Cabe destacar o Sensoriamento Remoto como um meio eficaz para o monitoramento e mapeamento da superfície terrestre, visto que esta área do conhecimento possibilita a organização e planejamento do espaço físico por tornar possível o tratamento de imagens, fornecendo uma visão sinóptica da superfície terrestre que permite ampla gama de aplicações (INPE, 2001). Além disso, tem sido utilizado amplamente para a verificação do cumprimento da legislação ambiental, tornando-se, dessa forma, uma importante ferramenta para acompanhar e monitorar a preservação do meio ambiente (COSTA, 2012).

2.2.1 A radiação eletromagnética e a vegetação

A vegetação é uma classe de cobertura da terra presente em quase toda a superfície terrestre. Com exceção do Ártico, Antártica e dos desertos, a maior parte da Terra, além das construções humanas, é coberta por vegetação (DI GREGORIO e O'BRIEN, 2012). Di Gregorio e O'Brien (2012) relatam que as plantas podem ser classificadas de acordo com vários critérios diferentes, dependendo de características como: propriedades da própria vegetação (fisionômica, estrutural e florística); propriedades externas da vegetação (sucessão vegetal, *habitat* ou ambiente, localização geográfica das comunidades de espécies); e propriedades que combinam vegetação e ambiente.

No estudo da vegetação, o Sensoriamento Remoto passivo possui a vantagem de gravar a radiação eletromagnética refletida ou emitida pelo objeto de estudo sem interferir com este (JENSEN, 2005). Por ser sistemático no modo de coleta das informações, que ocorre sempre da mesma forma e no mesmo horário, remove as possíveis fontes de erro introduzidas nos levantamentos em campo (SETTE, 2009). Deve-se atentar ao fato de que certas fisionomias vegetais podem apresentar alguma variação fenológica ao longo do ano. Assim, por uma questão sazonal, haverá diferenciação nos padrões de uma mesma fisionomia observada em imagens de diferentes datas nas estações secas e chuvosas, como, por exemplo, as fisionomias do Cerrado (PONZONI et al., 2012). Com o Sensoriamento Remoto, essas variações sazonais podem ser preditas e monitoradas mais facilmente.

Quando se pretende utilizar o Sensoriamento Remoto para obtenção de informações da cobertura vegetal de uma região, deve-se considerar a interação da energia solar com a vegetação. De todos os elementos constituintes da vegetação, a folha constitui o principal deles quando se considera o processo de interação com a radiação eletromagnética (REM). Esta

interação ocorre principalmente através dos pigmentos contidos nas folhas, como, por exemplo, a clorofila, que absorve bem a energia na faixa de comprimento de onda entre 0,45 e 0,67 μm (PONZONI et al., 2012).

O fluxo radiante, ao incidir sobre uma planta, apresenta três formas de interação: absorvância, refletância e transmitância (NOVO, 2008). A radiação absorvida pelos pigmentos da folha corresponde a aproximadamente 50% do total que chega até a planta, auxiliando na fotossíntese, alterando estruturas moleculares e acelerando reações. A refletância é definida pela quantidade de energia refletida em razão da energia que incide na superfície das folhas ou do dossel, e a transmitância é a razão entre o fluxo transmitido através do dossel e das camadas que constituem a folha pelo fluxo incidente sobre ele (JENSEN, 2009; PONZONI e SHIMABUKURO, 2009).

A folha constitui o principal elemento quando se considera o processo de interação acima descrito (PONZONI et al., 2012). Ponzoni et al. (2012) descrevem o comportamento da REM incidente em uma folha da seguinte forma:

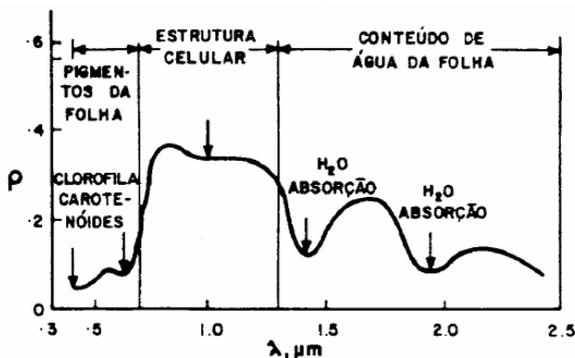
As características de refletância da REM incidente sobre uma folha estão relacionadas à sua composição química, principalmente tipo e quantidade de pigmentos fotossintetizantes, e a sua morfologia interna (distribuição e quantidade de tecidos, espaços intercelulares etc.). Uma folha típica é constituída de três tecidos básicos que são: epiderme, mesófilo fotossintético e tecido vascular. A folha é então coberta por uma camada de células protetoras epidérmicas, na qual, muitas vezes, desenvolve-se uma fina e relativamente impermeável superfície externa. Abaixo da epiderme encontra-se o mesófilo fotossintético, o qual, por sua vez, é frequentemente subdividido em uma camada ou em camadas de células paliçádicas alongadas, arranjas

perpendicularmente à superfície da folha, que formam o parênquima. As células do parênquima são ocupadas por seiva e protoplasma. Os espaços intercelulares cheios de ar localizam-se esparsos através do mesófilo, abrindo-se para fora através dos estômatos. Esta rede de passagens de ar constitui a via de acesso pela qual o CO_2 alcança as células fotossintéticas e o O_2 liberado na fotossíntese retorna à atmosfera externa. Uma terceira característica estrutural da folha é o tecido vascular. A rede de tecidos do sistema vascular não serve somente para suprir a folha com água e nutrientes do solo, mas também constitui a passagem pela qual fluem os produtos da fotossíntese que são produzidos na folha, para as demais partes da planta. As estruturas das células que compõem os três tecidos das folhas são muito variáveis, dependendo da espécie e das condições ambientais.

Os autores ainda citam que o comportamento espectral de uma folha é função de sua composição, morfologia e estrutura interna, e que, portanto, existirão diferenças no comportamento espectral entre grupos geneticamente distintos.

Novo (1989, apud PONZONI et al., 2012) divide em três intervalos a curva do fator de refletância de uma forma verde sadia: visível ($0,4 \mu\text{m}$ - $0,72 \mu\text{m}$), infravermelho próximo ($0,72 \mu\text{m}$ – $1,1 \mu\text{m}$) e infravermelho médio ($1,1 \mu\text{m}$ – $3,2 \mu\text{m}$) (Figura 7).

Figura 7 - Curva de reflectância típica de uma folha verde.



Fonte: Novo, 1989 apud Ponzoni et al., 2012.

Na Figura 7, observa-se que à medida que o espectro eletromagnético se direciona do visível ao infravermelho próximo, em torno de $0,7 \mu\text{m}$, a refletância da vegetação verde e sadia aumenta (LILLESAND e KIEFER, 1999). Segundo Ponzoni et al. (2012), na região do visível, os pigmentos existentes nas folhas dominam a refletância. Os pigmentos predominantes absorvem a REM na região do azul, mas somente a clorofila absorve na região do vermelho. A maioria das plantas é moderadamente transparente na região do verde ($0,540 \mu\text{m}$), região em que se dá um pico na refletância do espectro eletromagnético visível. Na região do infravermelho próximo, ocorre pequena absorção da radiação e considerável espalhamento interno da radiação na folha. Nessa região, fatores externos à folha, como disponibilidade de água, podem alterar a refletância da folha, pois a absorção da água geralmente é baixa nessa faixa do espectro (PONZONI et al., 2012).

No infravermelho médio, a refletância das folhas é afetada pela absorção decorrente da água líquida, que absorve consideravelmente a radiação incidente nessa região espectral (PONZONI et al., 2012). Dessa forma, sensores que trabalham na faixa de comprimento de onda entre $0,7$ e $1,3 \mu\text{m}$

conseguem captar alterações ocasionadas por algum tipo de estresse (CINTRA, 2007).

A aplicação das técnicas de Sensoriamento Remoto no estudo da vegetação inclui a necessidade de compreender o processo de interação da REM, não só com as folhas isoladas, mas também com os demais elementos constituintes de um dossel, como folhas, galhos, frutos, flores, dentre outros. De acordo com Ponzoni et al. (2012), as curvas de refletância de uma folha verde sadia são semelhantes às medições espectrais de dosséis. Os dosséis diferenciam-se espectralmente conforme sua densidade e distribuição espacial, sendo esta última dependente de como foram arranjadas as sementes no plantio (no caso de vegetação cultivada), do tipo de vegetação existente e do estágio de desenvolvimento das plantas. Quanto à densidade da cobertura vegetal, espera-se que em dosséis mais fechados, os valores de refletância espectral referentes à região do visível apresentem uma diminuição quase exponencial, enquanto que para a região do infravermelho próximo estes apresentam aumento também quase exponencial (PONZONI et al., 2012).

O dossel vegetal apresenta valores de refletâncias relativamente baixos na região do visível, devido à ação dos pigmentos fotossintetizantes que absorvem a REM para a realização da fotossíntese, fato evidenciado pela tonalidade escura nas imagens obtidas nesta região. Nas imagens da região do infravermelho próximo, verifica-se que estes valores se apresentam elevados, devido ao espalhamento interno sofrido pela REM em função da disposição da estrutura morfológica da folha, aliado ainda ao espalhamento múltiplo entre as diferentes camadas de folhas. No infravermelho médio, tem-se novamente uma queda destes valores, devido à presença de água no interior da folha (PONZONI et al. 2012). Importante salientar que a refletância no infravermelho próximo aumenta com o número de camadas de folhas em uma copa, podendo

servir de identificação do estágio de desenvolvimento da vegetação (PONZONI et al., 2012).

2.2.2 Imagens aéreas e aplicabilidade de imagens de alta resolução espacial no estudo da vegetação

As técnicas utilizadas no desenvolvimento do Sensoriamento Remoto estão ligadas à evolução da fotografia e das tecnologias espaciais. O filme fotográfico foi o primeiro material construído pelo homem capaz de registrar a radiação eletromagnética e de transformá-la na imagem do objeto fotografado (MENESES e ALMEIDA, 2012).

As fotografias aéreas foram o primeiro produto do Sensoriamento Remoto, dessa forma, métodos e técnicas referentes à Fotogrametria e a Fotointerpretação são antigos, e anteriores ao termo “Sensoriamento Remoto”, que vem a surgir apenas no início dos anos de 1960 por Evelyn L. Pruit e colaboradores (MENESES e ALMEIDA, 2012). Em 1956, foram iniciadas as primeiras aplicações sistemáticas de fotografias aéreas como fonte de informação para o mapeamento de formações vegetais nos Estados Unidos da América. No Brasil, datam de 1958 as primeiras fotografias aéreas na escala 1:25.000 obtidas com o objetivo de levantar as características da Bacia Terciária do Vale do Rio Paraíba como parte do programa de aproveitamento de seus recursos hídricos (INPE, 2001).

Mesmo com o advento dos satélites artificiais de imageamento terrestre, as técnicas e ferramentas utilizadas para levantamentos aerofotogramétricos continuaram evoluindo, sendo ainda utilizados em larga escala para diversas aplicações (INPE, 2001). De acordo com Lillesend e Kiefer (1999) o advento do avião, simultaneamente ao desenvolvimento das câmaras fotográficas, filmes, entre outros, trouxe um grande impulso às aplicações das fotografias para o levantamento de recursos naturais, visto que permitiu a obtenção de dados sob

condições controladas e com a cobertura de áreas relativamente amplas. As tecnologias, tanto de obtenção quanto de processamento e análise de imagens de fotografias aéreas, se desenvolveram fortemente durante anos, inicialmente nos setores militares e, em seguida, na administração pública e em setores privados (LILLESEND e KIEFER, 1999). A partir de imagens aéreas, podem ser extraídas informações para a criação de mapas de diferentes escalas, dependendo basicamente do tipo de câmara utilizada e da altitude de recobrimento aéreo. As imagens aéreas constituem uma importante fonte de dados, devido a sua alta resolução espacial e cobertura flexível (SCHOWENGERDT, 2007).

O levantamento aerofotogramétrico mais recente de Santa Catarina foi realizado pela empresa Engemap nos anos 2010 e 2011, e obteve aproximadamente 57 mil imagens com resolução espacial de 39 cm, em escala nominal de 1:10.000. O último recobrimento dessa natureza realizado no estado de Santa Catarina foi executado no período de 1977 a 1979 pela empresa Aerofoto Cruzeiro SA, em escala nominal de 1:25.000, em que foram geradas ortofotos pancromáticas em diafilme papel fotográfico.

As ortoimagens oriundas do levantamento aerofotogramétrico mais recente do estado constituem uma fonte de informações auxiliares as tarefas de inventário e de avaliação florestal, podendo contribuir para o estudo do uso e ocupação do território catarinense. Estudos têm mostrado grande potencial dessas ortoimagens. Lima (2013) analisou a aplicabilidade das ortoimagens na Perícia Criminal Ambiental. O autor constatou que não houve diferenças significativas entre a aferição de locais de crimes ambientais feita *in loco* com as medições efetuadas somente sob as ortoimagens, concluindo que tais imagens podem ser utilizadas para extração de informações relevantes, e, inclusive, para provar crimes desta natureza. Piovezan (2013) mostrou a aptidão das ortoimagens para auxiliar na identificação de áreas com limitações

ambientais em imóveis rurais no estado. Piazza (2014) realizou a classificação da cobertura da terra através dessas ortoimagens, com enfoque na classificação do estágio sucessional da FOD, e obteve índices Kappa superiores a 0,85.

Imagens de alta resolução espacial aliadas às técnicas de Sensoriamento Remoto constituem importantes fontes de informações para serem utilizadas no estudo da vegetação. Podem ser eficazes para obter dados da estrutura da floresta que não podem ser mensurados em campo, seja devido à dificuldade de medição pelo tamanho da área a ser amostrada, ou por se tratar de locais de difícil acesso (BARBIER et al., 2011). Segundo Asner et al. (2010), o crescente melhoramento e diversificação de sensores embarcados em satélites artificiais sugerem que as combinações de dados de campo e de Sensoriamento Remoto podem resultar em formas mais econômicas para o monitoramento da estrutura da floresta. Ainda segundo os autores, abordagens empregando Sensoriamento Remoto possuem o potencial não só de extrapolar resultados de campo, mas também fornecer dados de difícil obtenção em nível do solo, como altura total e tamanho da copa das árvores do dossel em florestas naturais multiestratificadas.

Barbier et al. (2011) citam que a interpretação visual de imagens aéreas tem sido usada amplamente por décadas para delinear inventários de amostragem em estratos, ou para mapear o mosaico do povoamento florestal através de critérios relacionados a idade, estrutura e dominância de espécies. Ribeiro (2007), ao estudar o potencial de imagens provenientes de sensores remotos no Inventário Florestal Nacional de Portugal, concluiu que é viável seu uso na avaliação e no monitoramento da floresta, principalmente por facilitar a avaliação de grandes áreas e proporcionar uma redução na periodicidade de avaliação das florestas. Yu et al. (2006) mostraram que a classificação semiautomática de imagens de

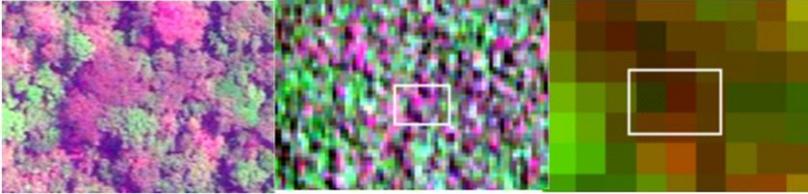
alta resolução espacial tem potencial para auxiliar o inventário florestal em campo.

Boyd e Danson (2005) relatam que o aumento da resolução espacial, espectral e radiométrica dos sensores remotos acrescentaram potencial de informação a ser explorada. Imagens de alta resolução espacial foram utilizadas para a estimação de parâmetros estruturais da vegetação com sucesso em diversos estudos (WULDER et al., 2004; JOHANSEN e PHINN, 2006; WANG et al., 2007; SETTE, 2009; COUTERON et al., 2012), constituindo assim uma ferramenta para caracterização da estrutura florestal e melhoria da acurácia da classificação que utiliza somente parâmetros espectrais.

Esta versatilidade das imagens de alta resolução espacial no estudo da vegetação advém do fato de que as mesmas permitem observar a parte superior das árvores com detalhes, incluindo as copas, os ramos e as partes sombreadas entre eles. De acordo com Disperati e Oliveira-Filho (2005), em imagens com uma resolução espacial de 0,72 m, é possível observar as árvores individualmente. Em imagens com resolução espacial de 1,5 m, começa a se perderem os detalhes, e as árvores deixam de ser distinguíveis em imagens com resolução espacial de 3 a 6 m.

Couteron et al. (2012) fazem um comparativo entre imagens de alta e baixa resolução espacial para discriminar as árvores do dossel da vegetação. Na Figura 8, percebe-se que as grandes árvores de folhas caducifólias emergentes (com “copas rosadas” em falsa-cor) já não são diretamente detectáveis na imagem MODIS, demonstrando assim o potencial de imagens de alta resolução espacial na identificação da vegetação.

Figura 8 - Comparação na diminuição resolução espacial de GeoEye (esquerda) para MODIS (à direita) e SPOT-5 (centro) em uma floresta decídua mista no Sul de Camarões.



Fonte: Couteron et al., 2012.

2.2.3 Métodos de classificação digital

A identificação de alvos em imagens provenientes de sensores remotos mediante a interpretação visual é eficaz quando o interesse é acessar as características geométricas e aparência geral desses alvos. Porém, o fato de a imagem ser composta por *pixel*, limita a visão humana ao extrair informação de *pixels* de forma isolada. Além disso, tem-se o fato de que na interpretação manual, somente três bandas espectrais podem ser visualizadas simultaneamente (PONZONI et al., 2012). Segundo Florenzano (2002), o olho humano é um sensor natural que enxerga somente a luz ou energia visível. Sensores artificiais nos permitem obter dados de regiões de energia invisível ao olho humano, podendo fornecer, dependendo da resolução radiométrica do sensor, 2.048 níveis de cinza ou mais.

A classificação digital é uma técnica que busca reconhecer padrões homogêneos na imagem e associá-los a determinada classe. A primeira etapa deste processo é chamada de treinamento, que pode ser supervisionado, quando o usuário dispõe de informações que permitem a identificação das classes de interesse na imagem, ou não-supervisionado, quando o usuário faz uso dos classificadores para reconhecer classes presentes nas imagens (PONZONI et al., 2012).

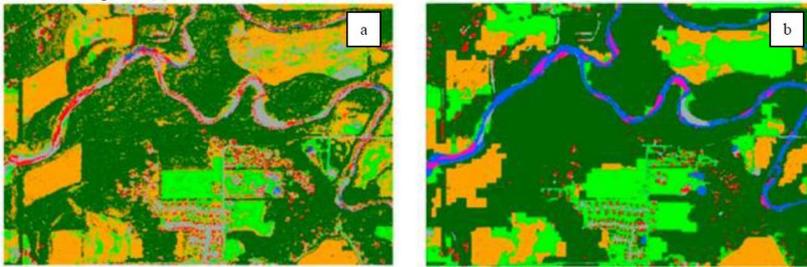
Os algoritmos de classificação podem ser subdivididos em paramétricos e não-paramétricos. Os paramétricos assumem que as classes possuem uma distribuição estatística específica, geralmente a distribuição normal, e requerem estimativas da distribuição dos parâmetros, como o vetor médio e matriz de covariância, para a classificação (HANSEN, 2012). Como exemplo de algoritmo paramétrico mais comum cita-se a Máxima Verossimilhança (MaxVer), que usa um modelo de probabilidade para determinar as fronteiras de decisão.

Os algoritmos não-paramétricos não assumem suposições sobre a distribuição de probabilidade e geralmente são considerados robustos, porque podem trabalhar bem com uma ampla variedade de distribuição de classes, desde que suas assinaturas espectrais sejam razoavelmente distintas. Nas últimas décadas, pesquisas envolvendo a classificação supervisionada da cobertura da terra têm dado ênfase aos algoritmos não-paramétricos como as árvores de decisão, redes neurais artificiais e máquinas de vetor de suporte (HUANG et al., 2002; PAL e MATHER, 2003). Estes algoritmos trabalham na fronteira espectral entre as classes de cobertura da terra, constituindo um avanço frente aos convencionais classificadores paramétricos (HANSEN, 2012).

Quanto à unidade de análise, a classificação pode ser por *pixel* ou por regiões. Na classificação por *pixel*, estes são avaliados isoladamente, e somente a informação espectral é utilizada para definir as regiões homogêneas. Fundamentam-se prioritariamente em métodos estatísticos ou determinísticos (PONZONI et al., 2012). Já a classificação por regiões, que é precedida por uma etapa de segmentação, avalia os segmentos, ou seja, *pixels* agrupados em compartimentos homogêneos da cena. O segmento corresponde a uma região discreta de uma imagem que é internamente coerente e diferente do seu entorno (CASTILLA e HAY, 2008; FURTADO et al., 2013).

As primeiras classificações automáticas de imagens reportadas na literatura consideravam apenas o valor de cada *pixel* (FRANCISCO e ALMEIDA, 2012). Porém, com o surgimento das imagens de alta resolução espacial, os *pixels* deixaram de ser suficientemente grandes para representar classes, como blocos florestais ou manchas urbanas. Sensores de alta resolução espacial fornecem imagens com uma alta variabilidade espectral intraclasse, impondo dificuldades em relação à extração de informação em nível de *pixel* (CAZES e FEITOSA, 2005; CADENA, 2011). Consequentemente, os resultados da classificação tendem a apresentar o efeito “sal-e-pimenta”, ou seja, uma imagem ruidosa, com aparência salpicada de *pixels* isolados fora de contexto, como mostra a Figura 9 (YU et al., 2006; GAO e MAS, 2008). Considera-se, portanto, a classificação por *pixel* limitada quando aplicada em imagens que captam detalhes de cenas muito heterogêneas ou com grande variação interna das classes (NEUBERT et al., 2006; SCHIEWE e TUFTE, 2007).

Figura 9 - Exemplo de imagem de alta resolução espacial classificada por *pixel* (a) e baseada em regiões (b).

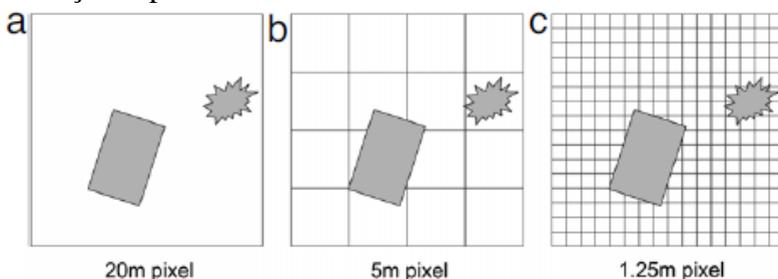


Fonte: Li e Shao, 2014.

Dessa forma, a classificação por regiões tende a ser mais adequada para imagens de alta resolução espacial. A Figura 10 ilustra a relação entre *pixels* e alvos nas imagens de baixa, média e alta resolução espacial. A Figura 10a mostra uma imagem de baixa resolução espacial, em que os *pixels* são

significativamente maiores que os alvos, havendo, portanto, necessidade de se empregar técnicas sub-*pixel* na classificação. A Figura 10b mostra uma imagem de média resolução, na qual se percebe que os tamanhos do *pixel* e dos alvos são similares, técnicas por *pixel* são apropriadas nesse caso. Já a Figura 10c mostra uma imagem de alta resolução geométrica. Nela, os *pixels* são significativamente menores que os alvos, sendo necessário regionalizar os *pixels* dentro de grupos de *pixels*, que correspondem a segmentos ou regiões.

Figura 10 - Relacionamento entre objetos em relação à resolução espacial.



Fonte: Blaschke, 2009.

A segmentação das imagens ganhou um novo impulso quando foi incorporada na OBIA (BLASCHKE e STROBL, 2001; BLASCHKE, 2010). Nela, os *pixels* não só são agrupados em segmentos, mas também são reconhecidos como objetos. Os objetos, diferentemente das regiões ou segmentos, são dotados de significado e identidade, sendo distinguíveis pela sua própria existência e não pelas propriedades que possuem. Deste modo, ainda que dois objetos da mesma classe possuam características idênticas, eles permanecem únicos e independentes (RODRIGUES, 2014). Assim, além das propriedades espectrais, espaciais e texturais, que possuem os segmentos, os objetos apresentam relações contextuais e semânticas, que podem ser utilizados para a análise de imagens, aproximam-se dos processos cognitivos humanos de

interpretação de imagens (NAVULUR, 2006; CAMARGO et al., 2009). A OBIA pressupõe um modelo de conhecimento criado *a priori* pelo intérprete (FRANCISCO e ALMEIDA, 2012).

Muitos estudos têm mostrado a superioridade, tanto da classificação por regiões quanto da OBIA, na classificação da vegetação quando realizada em imagens de alta resolução geométrica. Amaral et al. (2009), ao testarem diferentes classificadores em imagens dos satélites CBERS-2, IRS-P6 e Quickbird, concluíram que métodos de classificação por regiões foram mais eficientes na classificação de estádios de sucessão florestal quando comparados com o método de classificação por *pixel*, apresentando valores mais elevados de índice Kappa. Chubey et al. (2006), ao estudarem métodos para extrair informações de inventário florestal a partir de imagens de alta resolução espacial, identificaram haver relações entre alguns atributos do inventário florestal com as métricas espectrais e espaciais obtidas dos objetos oriundos da segmentação da imagem, como tipos de espécies, fechamento da copa e altura. Hay et al. (2005) mostraram como segmentos da classificação baseada em objeto correspondem à copa das árvores individuais e descrevem como essa ferramenta pode ser utilizada no mapeamento automatizado do inventário florestal. Addink et al. (2007) demonstraram que a precisão da estimativa de parâmetros da vegetação, como biomassa acima do solo e índice de área foliar, foi maior para a OBIA do que para a análise por *pixel*. Li et al. (2014) estudaram a importância de incorporar informações espacial-contextual para melhorar a acurácia da classificação, dentre elas: extração de textura da imagem, campos aleatórios de Markov e a segmentação seguida da OBIA. Sousa et al. (2010) utilizaram imagens Quickbird e classificação baseada em objeto na identificação de espécies florestais em Portugal, obtendo índice Kappa superior a 80%. Li e Shao (2014), ao classificarem a cobertura da terra com imagens aéreas em Midwester, Estados

Unidos da América, concluíram que a OBIA foi mais eficiente que métodos por *pixel* em termos de acurácia da classificação, já que, enquanto abordagens por *pixel* consideraram apenas comportamento espectral, a OBIA permitiu utilizar múltiplos atributos que melhoram a acurácia da classificação.

2.2.4 Segmentação

A segmentação compreende a primeira etapa tanto da classificação por regiões quanto da OBIA. Na segmentação, a imagem é dividida em objetos elementares, *pixels* com características similares em termos tonais e texturais, que são agrupados em grupos homogêneos formando os chamados segmentos, e estes serão as unidades para o processamento da imagem e posterior classificação (DLUGOSZ et al., 2009).

Um *pixel* pode ser interpretado em termos de comportamento espectral, porém não pode ser associado a um objeto da superfície terrestre, pois não possui relação direta ou semântica com alvos de interesse na cena (FRANCISCO e ALMEIDA, 2012). Além disso, o *pixel* não permite a extração de atributos, a exemplo do segmento ou região (BLASCHKE et al., 2008, FRANCISCO e ALMEIDA, 2012). Os segmentos fornecem, além da informação espectral, informações espaciais dos objetos. As informações espaciais da imagem incluem aspectos de textura, contexto, além de atributos geométricos (CHUBEY et al., 2006; BLASCHKE et al., 2008). O nível da segmentação depende do que se pretende expressar no procedimento de classificação. A tarefa de segmentação pode ser considerada “boa”, ou “aceitável”, quando os objetos de interesse do problema estiverem isolados, o que permitirá, em tese, discriminar adequadamente as classes previamente definidas (GONZALES e WOODS, 2001).

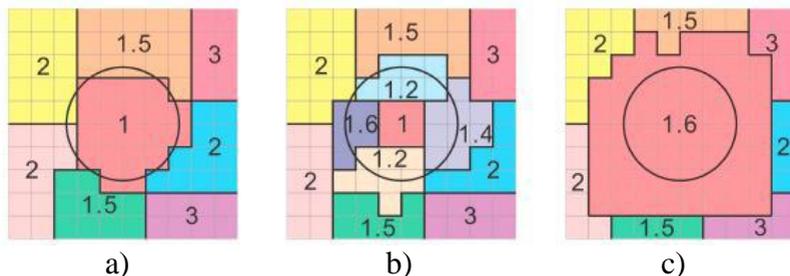
Assim como a classificação, a segmentação pode ser de forma manual ou automática. Na manual o analista considera as cores, textura, formato e conhecimento prévio para delimitar

e categorizar as áreas referentes a cada uso do solo ao delimitar os objetos. Este tipo de segmentação possui alta confiabilidade, mas se torna inviável para a análise de áreas grandes e heterogêneas (SANTOS et al., 2013). Além disso, o resultado não será único, pois diferentes intérpretes podem gerar produtos diferentes a partir de uma mesma imagem. A segmentação automática, por utilizar critérios de reconhecimento de similaridades ou de diferenças mensuráveis entre regiões adjacentes, apresenta resultados sempre iguais quando se repete o processo com os mesmos parâmetros (DLUGOSZ et al., 2009). Também produz resultados satisfatórios em menor tempo do que segmentação manual, sendo usualmente a mais utilizada (MENESES e ALMEIDA, 2012).

Segundo Coutinho (1997), a segmentação é uma etapa que exige maior atenção, devido à inexistência de valores padronizados para as variáveis do algoritmo, o que pode influenciar diretamente na acurácia do seu resultado. Uma segmentação ruim pode resultar em altas taxas de erro na classificação (DURIC et al., 2014).

O processo de segmentação é empírico e ajustável aos diferentes tipos de imagens, com limiares definidos de acordo com a complexidade dos alvos investigados e do problema a ser resolvido (DURIC et al., 2014). Em uma segmentação adequada, os segmentos gerados devem ser representativos dos objetos presentes na cena, como mostra a Figura 11a. Já a supersegmentação (Figura 11b), ou seja, quando são gerados muitos segmentos pequenos pertencentes a um mesmo objeto; e a subsegmentação (Figura 11c), quando um segmento abrange mais de um objeto, devem ser evitadas. Esses problemas podem ser contornados com um adequado ajuste dos parâmetros do algoritmo segmentador empregado.

Figura 11 - Exemplo de resultados satisfatório (a) e não satisfatórios (b,c) da segmentação.



Fonte: Liau, 2014.

A qualidade da segmentação depende, portanto, não apenas do algoritmo selecionado, mas de um apropriado ajuste dos valores de seus parâmetros (CARLEER et al., 2005). Esse ajuste pode ser feito através de uma longa série de tentativas e erros, não havendo garantia de que os valores ótimos serão determinados, ou ainda, de forma automatizada, em que uma função objetivo é dada por uma métrica para avaliação da segmentação (ACHANCCARAY DIAZ, 2014).

2.2.4.1 Algoritmos genéticos

Na análise de imagens de Sensoriamento Remoto, a subjetividade ao se definir parâmetros de segmentação tem impulsionado a pesquisa em abordagens que visam avaliar a qualidade dos segmentos, com base em alguns critérios de qualidade. Pignalberi et al. (2003), Feitosa et al. (2007), Espindola (2009), Saba et al. (2013) e Achancaray Diaz (2014) propõem o uso de algoritmos genéticos (AG) para o ajuste automático dos parâmetros de segmentação, obtendo resultados que confirmam que essa metodologia é uma abordagem promissora para reduzir o tempo investido pelo operador humano no ajuste dos parâmetros de algoritmos de segmentação.

Os AG, propostos por Holland (1975, apud SABA et al., 2013) são baseados no processo de evolução natural. O processamento básico de um AG inicia com a criação de uma população inicial, onde os valores dos genes são determinados aleatoriamente. Na sequência, os indivíduos são avaliados, e os mais aptos são selecionados. Cria-se uma nova população a partir da utilização de operadores genéticos sobre os indivíduos selecionados da geração anterior. Certos percentuais de indivíduos mais aptos são mantidos na nova população (FEITOSA et al., 2007). O procedimento continua até que seja alcançado um determinado número de gerações ou até que um determinado número de indivíduos tenha sido criado (ESPINDOLA, 2009; SABA et al., 2013).

Além de serem aplicados na segmentação de imagens (BHANU et al., 1995; PIGNALBERI et al., 2003; FEITOSA et al., 2006; FREDRICH e FEITOSA, 2008; ESPINDOLA, 2009; FEITOSA et al., 2009; LÜBKER e SCHAAB, 2009; DERIVAUX et al., 2010; FOURIE et al., 2012; LEONARDI et al., 2012), AG são usados em muitas outras aplicações de processamento de imagens, como classificação (TSENG et al., 2008; WEN et al., 2009; STAVRAKLOUDIS et al., 2011), extração de atributos (LI et al., 2011) e seleção de feições ótimas (VAN COILLIE et al., 2007).

A tarefa de ajustar os parâmetros de um algoritmo de segmentação consiste em encontrar os valores que levem o algoritmo a produzir um resultado ótimo, medido a partir de um critério de qualidade apropriado. A qualidade de uma solução (aptidão de um indivíduo) deve refletir o grau de similaridade entre a segmentação produzida a partir destes valores e uma segmentação de referência (FEITOSA et al., 2007). Tal referência pode consistir de um conjunto de segmentos, tipicamente delimitados sobre a imagem analisada por um intérprete, que representam sua percepção do que seja uma segmentação ideal. O que se procura é o conjunto de

valores de parâmetros de segmentação que produza o resultado mais consistente com a referência (FEITOSA et al., 2007).

Segundo Feitosa et al. (2007), dado um conjunto de segmentos S , um conjunto de parâmetros P , e a função objetivo $F(S,P)$, a tarefa do AG consiste em procurar o vetor de parâmetros P para qual o F alcança o seu mínimo, resultando nos parâmetros ótimos (P_{opt}). O problema é formulado matematicamente pela Equação 1.

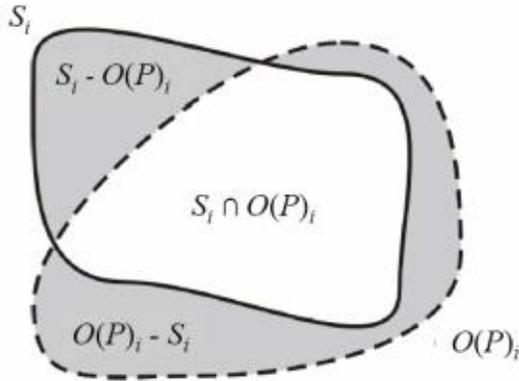
$$P_{opt} = \arg \min_p [F(S,P)] \text{ Equação 1}$$

Seja S_i o conjunto de *pixels* pertencentes ao *i-ésimo* segmento de S , e $O(P)_i$ o conjunto dos *pixels* pertencentes ao segmento produzido a partir de P que possua a maior interseção com S_i , a função objetivo é definida pela Equação 2.

$$F(S,P) = \frac{1}{n} \sum_{i=1}^n \frac{\#(S_i - O(P)_i) + \#(O(P)_i - S_i)}{\#(S_i)} \text{ Equação 2}$$

Em que: “-” representa a operação diferença de conjuntos, “#(.)” representa o número de elementos (cardinalidade) de um conjunto, e n é o número de segmentos em S . A Figura 12 ilustra a função objetivo proposta. A região com contorno sólido representa um segmento de referência S_i , a região com contorno tracejado representa $O(P)_i$. A função objetivo é dada pela média da razão entre a soma das áreas sombreadas e a área de S_i . A correlação perfeita entre os segmentos de referência e a segmentação gerada corresponde a $F = 0$ (FEITOSA et al., 2007).

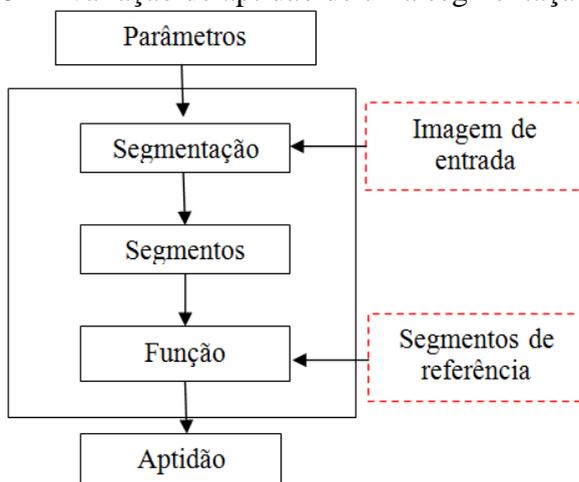
Figura 12 - Representação da função objetivo.



Fonte: Feitosa et al., 2007.

O cálculo da função-objetivo, que deve ser feito para cada indivíduo a cada geração, envolve a execução do procedimento de segmentação. Para avaliar uma solução (indivíduo) compara-se o resultado do procedimento de segmentação obtido a partir dela, com uma segmentação de referência. Ao final do processo o indivíduo mais apto é selecionado como solução aproximada do problema de otimização (AYMA, 2013) (Figura 13).

Figura 13 - Avaliação de aptidão de uma segmentação por AG.



Fonte: Adaptado de Feitosa et al., 2007.

As funções-objetivo são responsáveis por interagir com os algoritmos de segmentação para a determinação de um conjunto de parâmetros ótimos, que podem definir o desempenho do algoritmo de segmentação dada certa função de aptidão (ACHANCCARAY DIAZ, 2014). Como exemplos de funções-objetivos citam-se a Evolução Diferencial (*Differential Evolution*), a Busca Generalizada de Padrões (*Generalized Pattern Search*) e a *Nelder-mead* (NELDER e MEAD, 1965). De acordo com Ayma (2013), a função-objetivo *Nelder-mead* possui um custo computacional mais eficiente para o processo de otimização, já que avalia a função-objetivo só quando necessário.

Uma medida de discrepância, ou função de aptidão, é usada para avaliar a concordância entre os segmentos de referência e os segmentos gerados a partir de cada indivíduo da população (em que um indivíduo é um conjunto de parâmetros e a população um grupo de diferentes conjuntos de parâmetros) (NOVACK et al., 2011).

Alguns aplicativos possuem ferramentas que utilizam os AG na busca e avaliação dos parâmetros de segmentação. Espindola (2009) usou a ferramenta *Genetic Algorithm Tool* do software MATLAB 7.1. Pignalberi et al. (2003) utilizaram o sistema a qual denominaram GASE (*Genetic Algorithm environment*) para otimização de parâmetros de segmentação. Mais recentemente tem se destacado o aplicativo *Segmentation Parameters Tuner* (SPT), desenvolvido pelo Departamento de Engenharia Elétrica da PUC-RJ, utilizado nos trabalhos de Feitosa et al. (2009), Novack (2009), Novack et al. (2011), Achanccaray Diaz (2014). Os trabalhos relatam que este programa possui grande potencial para ajustes de parâmetros de segmentação. Saba et al. (2013), ao estudarem AG para encontrar parâmetros ótimos da segmentação multirresolução de imagens IKONOS, mostraram que a acurácia foi maior quando usado AG do que quando os parâmetros eram selecionados pelo usuário.

2.2.4.2 Algoritmo de segmentação por crescimento de regiões de Baatz e Schäpe

Algoritmos de segmentação por crescimento de regiões, como o algoritmo proposto por Baatz e Schäpe (BAATZ e SCHÄPE, 2000), baseiam-se na similaridade, ou seja, agregam *pixels* em função de sua semelhança com *pixels* vizinhos. Considera, inicialmente, um *pixel* denominado “semente”, como sendo uma região. Na sequência, o algoritmo inicia as comparações com as outras regiões adjacentes, levando em consideração o limiar de similaridade definido pelo analista, executando um teste de médias em que o limiar define a distância entre elas (KAI e MULLER, 1991; BAATZ E SCHÄPE, 2000; MENESES e ALMEIDA, 2012). Se no *pixel* vizinho houver similaridade com relação a algum atributo da região, ele é agregado à região. Este processo continua até que toda a imagem seja segmentada (FONSECA et al., 2000). A

escala é um critério importante no processo da segmentação. Quando o tamanho da região atinge o limite definido no parâmetro de escala definido, o processo de fusão termina (LI e SHAO, 2014).

A medida de heterogeneidade usada no algoritmo Baatz e Schäpe possui um componente espacial e um componente espectral. A heterogeneidade do componente espectral é definida sobre os valores das respostas espectrais dos *pixels* contidos em um segmento. A heterogeneidade do componente espacial baseia-se nos atributos de forma: compactação e suavidade. O grau de compactação é definido como a razão entre o perímetro do segmento e a raiz quadrada de sua área (número de *pixels* que contém). A suavidade é definida como a razão entre o perímetro do objeto e o perímetro do retângulo envolvente mínimo (*bounding box*). O crescimento dos segmentos é condicionado ao ajuste desses critérios. O ajuste pode ser feito pela escolha do parâmetro de escala, dos pesos das bandas espectrais, do fator de cor e do fator de compactação, dependendo da solução computacional empregada (INPE, 2011).

Como exemplos de sistemas utilizando este algoritmo, citam-se o eCognition (LANG e TIEDE, 2007; PIAZZA, 2014), o InterIMAGE (COSTA et al., 2010, MENEGHETTI, 2013) e o GeoDMA (KÖRTING, 2012).

2.2.4.3 Algoritmo de segmentação por média-móvel

A média-móvel (*mean shift*) é um método não-paramétrico e não-supervisionado que tem por objetivo localizar as modas em um conjunto de dados quaisquer. Esse algoritmo foi proposto pela primeira vez por Fukunaga e Hostetler (1975) para o reconhecimento de padrões, e adaptado por Cheng (1995) para a análise de imagens. Comaniciu e Meer (2002) expandiram a média-móvel para as áreas de segmentação de imagens, suavização e detecção de

movimento. Diferentemente dos algoritmos de segmentação por crescimento de regiões, a segmentação por média-móvel adota uma abordagem baseada em *cluster* (agrupamento), em que os dados de uma imagem são vistos como uma nuvem de pontos sobre um eixo unidimensional de escala de cinza, ou em um espaço de cor multidimensional (ACHANCCARAY DIAZ, 2014).

O algoritmo utiliza o conceito de janela-móvel, ou seja, escolhe um ponto arbitrário do conjunto de dados e determina uma janela ao redor desse ponto. Em seguida, ele calcula a média dos pontos que estão dentro dessa janela, assim, o novo centro de massa é determinado e a janela é deslocada de forma que fique centrada no ponto calculado (SCHARDONG, 2011). O procedimento se repete até que a média atinja um critério de parada definido. Esse ponto final é o centro de massa do *cluster*, conhecido como centroide (THIRUMURUGANATHAN, 2010; SCHARDONG, 2011). No procedimento de segmentação pela média-móvel, cada ponto do conjunto de dados é associado ao seu centro de massa. Dessa forma, todos os pontos serão associados a um centroide e cada centroide definirá um *cluster* de dados (SCHARDONG, 2011).

Esse algoritmo possui três parâmetros a serem ajustados quando utilizado na segmentação de imagens: raio espacial, que define a vizinhança de busca; raio ou valor espectral, que define o espaço espectral; e o tamanho mínimo da região, que define os menores objetos a serem formados na segmentação (VU, 2012).

2.2.5 Mineração de dados

Mineração de dados pode ser definida como uma parte de um âmbito maior de pesquisa denominado “Busca pelo Conhecimento em Banco de Dados” (do inglês, *Knowledge Discovery in Database*) que possui uma metodologia própria

para preparação e exploração de dados, interpretação de seus resultados e assimilação dos conhecimentos minerados (CORTÊS et al., 2002). Segundo Körting et al. (2008), as técnicas de mineração de dados podem aumentar o potencial de análises e aplicações de dados de Sensoriamento Remoto, já que estes representam uma grande heterogeneidade de alvos de difícil distinção e, por essa razão, exigem técnicas mais apuradas para a extração de informações. Assim, a mineração de dados permite a classificação de imagens de maneira rápida, em comparação à análise manual (KÖRTING, 2012).

Programas voltados à mineração de dados podem abranger diferentes tarefas, dentre elas: descrição, classificação, regressão, seleção de atributos, agrupamento e associação. O aprendizado na mineração de dados pode ocorrer de duas formas: (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; (2) não-supervisionado, em que não existe a informação de saída desejada, e os algoritmos partem dos dados procurando estabelecer relacionamento entre eles (GOLDSCHMIDT e PASSOS, 2005).

A diferença entre os métodos de aprendizado supervisionados e não-supervisionados está no fato de que os não-supervisionados não precisam de uma pré-categorização para os registros, ou seja, não é necessário um atributo-alvo. Tais métodos geralmente usam alguma medida de similaridade entre os atributos (MCCU, 2007). As tarefas de agrupamento e associação são consideradas como não-supervisionadas. Já no aprendizado supervisionado, os métodos são providos com um conjunto de dados que possuem uma variável-alvo pré-definida, e os registros são categorizados em relação a ela. As tarefas mais comuns de aprendizado supervisionado são a classificação (que também pode ser não-supervisionada) e a regressão (MCCU, 2007). A classificação supervisionada visa identificar a qual classe um determinado registro pertence.

Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (CAMILO e SILVA, 2009).

2.2.5.1 Seleção de atributos

A seleção de atributos é usada para escolher um subgrupo ótimo de atributos a partir do grupo original com base em heurísticas e critérios de avaliação (KOHAVI e JOHN, 1997). Bancos de dados podem conter vários atributos, alguns redundantes ou irrelevantes. Variáveis altamente correlacionadas não agregam informação para a construção de um modelo, e as irrelevantes não contêm informação útil (MERSCHMANN, 2007). Segundo Hall (1998), a existência de informação redundante e irrelevante é um fator que pode degradar significativamente o desempenho de algoritmos de aprendizado. A redução da dimensionalidade dos dados com a seleção de atributos permite que os algoritmos de mineração de dados, como árvores de decisão, trabalhem com mais rapidez e eficiência, podendo, inclusive, proporcionar uma melhoria na qualidade do classificador (KOHAVI e JOHN, 1997; MERSCHMANN, 2007; NOVACK, 2009). Além disso, ganhos podem ser obtidos com o aumento da compreensibilidade do modelo construído (MERSCHMANN, 2007).

Algoritmos de seleção de atributos podem ser divididos em seleção por filtro (*filter*) ou os chamados *wrappers* (YU e LIU, 2003). Os algoritmos de seleção por filtro baseiam-se nas características gerais dos dados de treinamento (amostras) e não envolvem nenhum algoritmo de aprendizado. Já os *wrappers* irão depender do algoritmo de aprendizado para a seleção e usam seu desempenho para avaliar e determinar quais atributos serão selecionados. Devido a essa dependência, os melhores atributos para determinado algoritmo de aprendizado

(redes neurais, por exemplo), podem não ser os melhores para a árvore de decisão. Esse processo tende a alcançar um desempenho superior para encontrar os melhores atributos para certo algoritmo de aprendizado, porém, também é computacionalmente mais caro e demorado (YU e LIU, 2003). Os algoritmos por filtro são computacionalmente mais eficientes quando o número de atributos de um banco de dados é maior. Eles operam independentemente de qualquer algoritmo de mineração de dados, selecionando atributos utilizando apenas as características gerais dos dados (MERSCHMANN, 2007).

Os algoritmos de seleção de atributos por filtro podem ser divididos em dois grupos: aqueles que procuram um subgrupo ótimo de atributos; e aqueles que avaliam individualmente os atributos e os graduam quanto à relevância deles em relação aos conceitos alvos (classes) (KOHAVI e JOHN, 1997). O Filtro Rápido Baseado em Correlação (*Fast Correlation-Based Filter*), RELIEF-F, Razão de Ganho (*Gain Ratio*), Ganho de Informação (*Info Gain*) e Qui-Quadrado (*ChiSquared*), adotam abordagens que consideram cada atributo individualmente, ordenando-os de acordo com as suas capacidades preditivas em um *ranking*. Já a Seleção de Atributos Baseada em Correlação (*Correlation-based Feature Selection*- CFS) se enquadra no primeiro grupo, pois avalia os subconjuntos de atributos com o objetivo de encontrar aquele que maximiza a acurácia preditiva do classificador. São desejados algoritmos de seleção de atributos que eliminam tanto a irrelevância quanto a redundância dos dados, com o menor custo computacional possível (YU e LIU, 2003).

2.2.5.2 Árvores de decisão

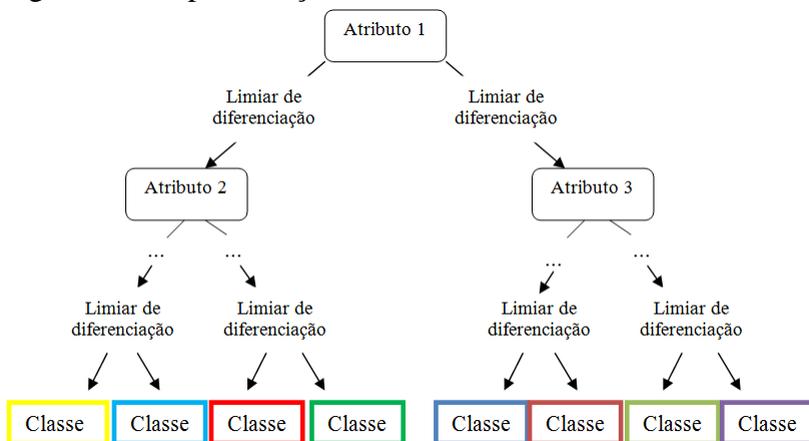
As árvores de decisão pertencem à categoria de algoritmos supervisionados de mineração de dados. A aprendizagem do modelo é baseada em amostras de

treinamentos, cuja informação da classe a qual pertence cada amostra é fornecida pelo usuário. Elas têm sido amplamente utilizadas pelo fato de possuírem uma representação intuitiva, que torna o modelo de classificação fácil de ser interpretado (GONÇALVES, 2007; SATO et al., 2013). De acordo com Quinlan (1993), a abordagem por árvore de decisão tem como vantagem o fato de possuir natureza e propriedades não-paramétricas, podendo classificar imagens com distribuições estatísticas diferentes da gaussiana, heterogêneas e possuidoras de ruídos, como *outliers*.

A árvore de decisão é formada por um conjunto de regras de classificação, em que cada caminho da raiz até uma folha representa uma destas regras, e deve ser definida de forma que, para cada observação da base de dados, haja apenas um caminho da raiz até a folha (GONÇALVES, 2007). Inicialmente, a raiz da árvore contém toda a base de dados das várias classes. Um ponto de separação é escolhido como sendo a condição, ou atributo, que melhor separa ou discrimina as classes. Este atributo separa a base de dados em dois ou mais conjuntos associados a determinado nó-filho. Cada nó-filho, por sua vez, abrange um particionamento que será novamente separado até que os registros dos nós-folha da árvore pertençam inteiramente ou predominantemente a uma mesma classe (GOLDSCHMIDT e PASSOS, 2005; NOVACK, 2009).

Para a construção de uma árvore de decisão é necessário um conjunto de amostras de treinamento, com as quais o usuário define previamente as classes. Assim, a árvore deve ser estruturada de forma que cada nó interno possua como rótulo o nome de um dos atributos previsores, no qual os ramos que saem do nó interno são rotulados com valores do atributo daquele nó, e a folha recebe como rótulo o nome da classe (NOVACK et al., 2011). Essa estrutura é exemplificada na Figura 14.

Figura 14 - Representação de uma árvore de decisão.



Fonte: Adaptado de Piazza, 2014.

Além da fácil interpretação dos seus resultados, já que a classificação é obtida de forma explícita, as árvores de decisão têm a vantagem de serem rapidamente geradas, necessitando menos tempo de processamento em relação a redes neurais, por exemplo (PAL e MATHER, 2003).

Entre os algoritmos de árvore de decisão, cita-se o C4.5, criado por Quinlan (1993) e implementado nos programas InterIMAGE, GeoDMA e WEKA (*Waikato Environment for Knowledge Analysis*), neste último com o nome de J48. De acordo com Ribeiro e Fonseca (2013), esse algoritmo funciona da seguinte forma: ele constrói árvores de decisão em que cada nó da árvore corresponde a um atributo e cada arco corresponde a uma faixa de valor do atributo. O valor de atributo esperado é definido pelo caminho da raiz de cada folha. O atributo mais representativo é associado com cada nó.

A entropia¹ é calculada para avaliar o quanto informativo é um nó. Dessa forma, quanto maior a entropia, mais informações são necessárias para caracterizar os dados, e maior a árvore de decisão gerada. O objetivo do classificador é associar a cada nó o atributo que minimiza a entropia dos dados (SILVA et al., 2008). Após associar o atributo com um nó, o C4.5 define um valor limite para cada arco. O limiar é calculado pelo algoritmo de vizinho mais próximo. Primeiro, o algoritmo calcula a distância euclidiana a partir das amostras de treinamento para uma instância de dados. A instância será atribuída à classe que está mais próxima a ela no espaço de atributos (WITTEN e FRANK, 1999). O algoritmo remove os nós desnecessários através do procedimento de “poda”, produzindo a árvore mais curta possível. O número de instâncias em cada folha também controla o tamanho da árvore, sendo que, quanto menor este número, mais precisa a classificação para o conjunto de treinamento (RIBEIRO e FONSECA, 2013). Segundo Pinho et al. (2008), uma alternativa é testar vários valores de limiares e observar quando a árvore estabiliza, ou seja, quando não ocorre mais significativas mudanças na acurácia da classificação.

Outro algoritmo de árvore de decisão que tem alcançado bons resultados é a Árvore de Regressão e Classificação (*Classification and Regression Trees- CART*), proposta por Breiman et al. (1984), denominado SimpleCart no programa WEKA. O CART consiste de uma técnica não-paramétrica que induz tanto árvores de classificação quanto árvores de regressão, dependendo se o atributo é nominal (classificação) ou contínuo (regressão). Dentre as principais vantagens do CART está a grande capacidade de pesquisa das relações entre

¹Em teoria da informação, a entropia é o valor esperado (média) das informações contidas em cada mensagem recebida. A mensagem significa um evento de amostra, ou de caracteres, desenhados a partir de um fluxo de dados ou distribuição. Entropia caracteriza, assim, a incerteza sobre determinada fonte de informação e aumenta para fontes de maior aleatoriedade. Generalizando, a entropia se refere à desordem ou incerteza (HAN e KOBAYASCHI, 2002).

os dados, mesmo quando elas não são evidentes, bem como a produção de resultados sob a forma de árvores de decisão de grande simplicidade e legibilidade (FONSECA, 1994). Da mesma forma que o algoritmo C4.5, o CART utiliza a técnica de pesquisa exaustiva para definir os limiares que irão dividir os nós correspondentes aos atributos contínuos. Diferente das abordagens adotadas por outros algoritmos, os quais utilizam pré-poda, o CART expande a árvore exaustivamente, realizando pós-poda por meio da redução do fator custo-complexidade (BREIMAN et al., 1984). Por esse motivo, esse algoritmo tende a produzir árvores mais simples, precisas e com boa capacidade de generalização (BREIMAN et al., 1984).

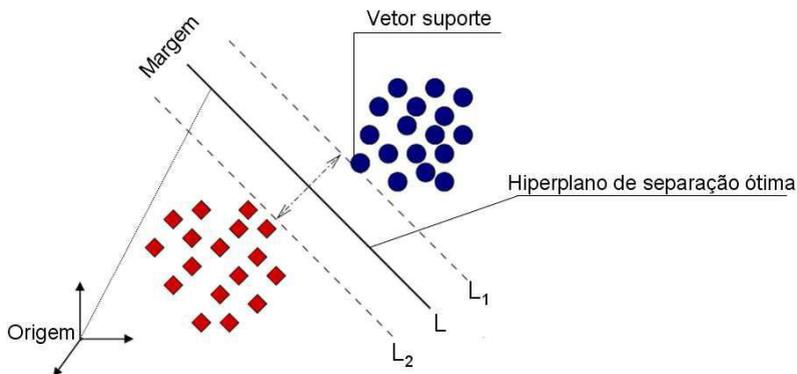
Ambos os algoritmos, C4.5 e CART, são considerados precursores, e diversas variações surgiram deles (CAMILO e SILVA, 2009), como J48Graft, o REPTree e o RandomTree, também disponíveis no WEKA. O algoritmo RandomTree usa uma classificação para a construção de árvore que considera “n” atributos escolhidos aleatoriamente em cada nó e não realiza a poda da árvore de decisão. Já o algoritmo REPTree constrói uma árvore de decisão usando informação de ganho e variância, e poda a árvore usando poda de erro reduzido (GIASSON et al., 2013). O J48Graft gera uma árvore de decisão enxertada do algoritmo J48. A técnica de enxertia é um processo indutivo que adiciona nós às árvores de decisão com a finalidade de reduzir erros de predição (WISAENG, 2013).

Nos últimos anos, vários trabalhos têm se destacado ao utilizar técnicas de mineração de dados, como as árvores de decisão (PAL e MATHER, 2003; PINHO et al., 2008; OTUKEI e BLASCHKE, 2010; CARVALHO, 2011; FRANCISCO e ALMEIDA, 2013; MENEGHETTI, 2013; WISAENG, 2013; RODRIGUES, 2014; PIAZZA, 2014; PASSO et al., 2014) e seleção de atributos (LIU e YU, 2005; MERSCHMANN, 2007; NOVACK, 2009; PIAZZA, 2014).

2.2.6 Máquina de vetor de suporte (*Support Vector Machine*- SVM)

Introduzido por meio da teoria estatística de aprendizagem por Vapnik (1995), o algoritmo *Support Vector Machine* (SVM) é um classificador supervisionado e não-paramétrico, que tem como principal objetivo a determinação de limites de decisão que permitam a separação ótima entre classes a partir da minimização dos erros (NASCIMENTO et al., 2009). Este algoritmo assume como base uma superfície de decisão que separa as classes, maximizando os limites de separação entre elas, sendo este conhecido como hiperplano ideal (Figura 15). Os pontos localizados próximos a essa superfície são definidos como vetores de suporte (*support vectors*), encontrados durante a fase de treinamento (BURGES, 1998). Com base no parâmetro de penalidade (γ), o usuário pode controlar a dualidade entre rigor/permisividade do hiperplano e a aceitação/rejeição de erros de treinamento contidos nas amostras (TSO e MATHER, 2009).

Figura 15 - Esquema de classificação por meio do SVM, em que os pontos e quadrados representam duas classes no espaço de atributos.



Além de realizar a classificação linear, o SVM desempenha eficientemente uma classificação não-linear usando as denominadas funções-núcleos (*kernel*). O SVM possui quatro funções-núcleos principais, sendo elas: (1) linear, (2) quadrática, (3) polinomial e (4) função de base radial (FBR) (NASCIMENTO et al., 2009). A fórmula de cada função e uma análise mais detalhada do SVM pode ser visualizada em Lorena e Carvalho (2007) e Bonesso (2013).

Steinwart e Christmann (2008) identificam três razões para o sucesso do SVM: capacidade de aprender bem com um número pequeno de parâmetros; robustez diante de vários tipos de violações de modelo e da diversidade de modelos e; eficiência computacional em comparação com outros métodos. A robustez no tratamento de dados de alta dimensionalidade, diante de objetos de grandes dimensões, como as imagens de satélites, está relacionada à capacidade de discernimento de dados com maior entropia pelo SVM, ou seja, a quantidade de incerteza (aleatoriedade) na imagem (ANDRADE et al., 2014). Normalmente, outros classificadores encontram problemas com dados de maior entropia, devido à grande ocorrência de *overfitting*, ou seja, um superajustamento dos dados de treinamento (ALIXANDRINI, 2009). Já os SVM, além de extraírem os parâmetros gerais que permitem a generalização, armazenam os ruídos e as peculiaridades, tolerando o reconhecimento de padrões não observados durante a etapa de treinamento (ALIXANDRINI, 2009; ANDRADE et al., 2014). Vários estudos compararam o classificador SVM com outros classificadores, como o MaxVer, redes neurais, e árvores de decisão, e o SVM se equiparou ou superou tais métodos em robustez e acurácia (HUANG et al., 2002; FOODY e MATHUR, 2004; PAL e MATHER, 2005).

Atualmente, estudos têm explorado a integração do SVM na classificação por regiões (TZOTSOS, 2006; TZOTSOS e ARGIALAS, 2008; DEVADAS et al., 2012; RAISSOUNI e

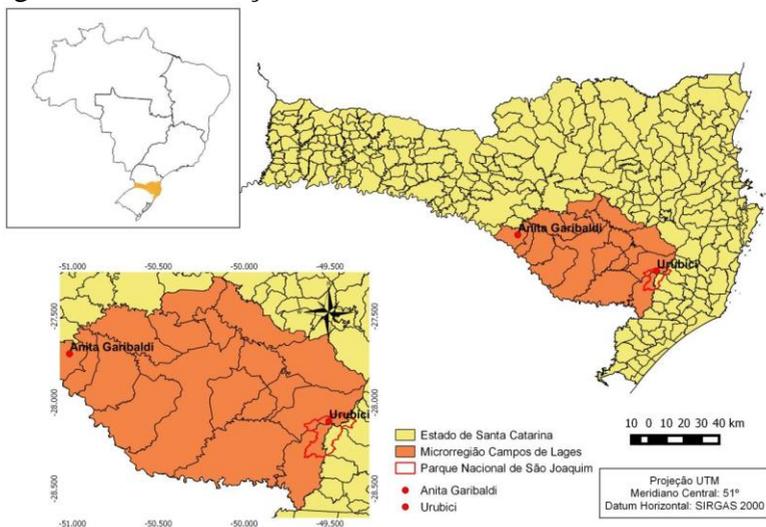
RAISSOUNI, 2013). A principal diferença é que, ao invés de *pixels*, as amostras são objetos oriundos da segmentação, e possuem, além de características espectrais, atributos de forma e textura a serem considerados na classificação. Heumann (2011) utilizou a classificação baseada em regiões SVM para classificar áreas de mangue em imagens Worldview-2 e obteve índice Kappa superior a 0,8. Antunes et al. (2014), ao aplicarem SVM em classificação por regiões em imagens RapidEye para classificar a cobertura da terra, obtiveram porcentagem de acerto de 83%. Devadas et al. (2012), para mapear áreas agrícolas com imagens Landsat, compararam a classificação SVM por objeto com a tradicional classificação por *pixel* pelo algoritmo MaxVer, e obtiveram acurácia de 95% na primeira técnica e 89% na segunda.

3. MATERIAL E MÉTODOS

3.1 ÁREA DE ESTUDO

O estudo foi realizado em três áreas distintas pertencentes à região fitoecológica de Floresta Ombrófila Mista (KLEIN, 1978), situadas na mesorregião Serrana e microrregião Campos de Lages do estado de Santa Catarina (IBGE, 2014). As áreas-teste localizam-se em zonas rurais dos municípios de Anita Garibaldi (áreas-teste A e B) e Urubici (área-teste C), e foram escolhidas pelos seguintes critérios: (a) disponibilidade de dados de levantamento em campo com a classificação dos estádios sucessionais da vegetação (VIBRANS et al., 2012; FAXINA, 2014) e (b) diversidade das classes de uso e cobertura da terra presentes nas imagens (Figura 16).

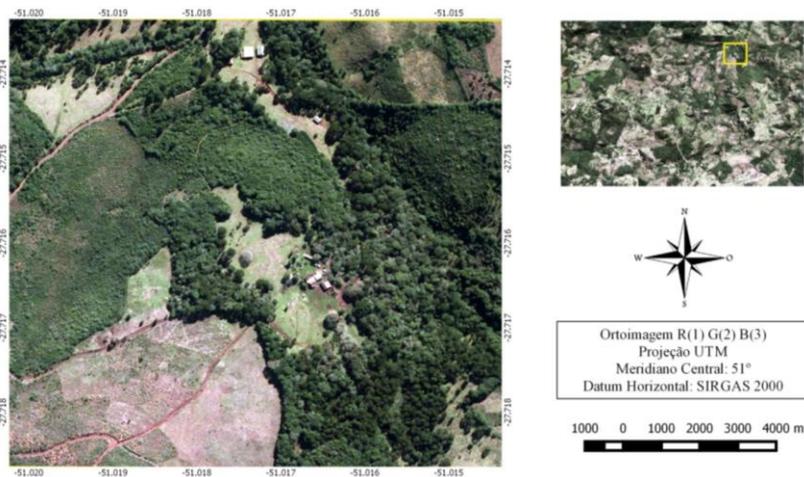
Figura 16 - Localização das áreas de estudo.



Fonte: produção do próprio autor, 2015.

O município de Anita Garibaldi é situado nas coordenadas médias 27°41'21" sul e 51°07'48" oeste, a uma altitude de 885 m, e possui uma área total de 588 km² (IBGE, 2014). A área-teste A caracteriza-se por possuir vegetação em diferentes estádios de sucessão e cultivos agrícolas (Figura 17).

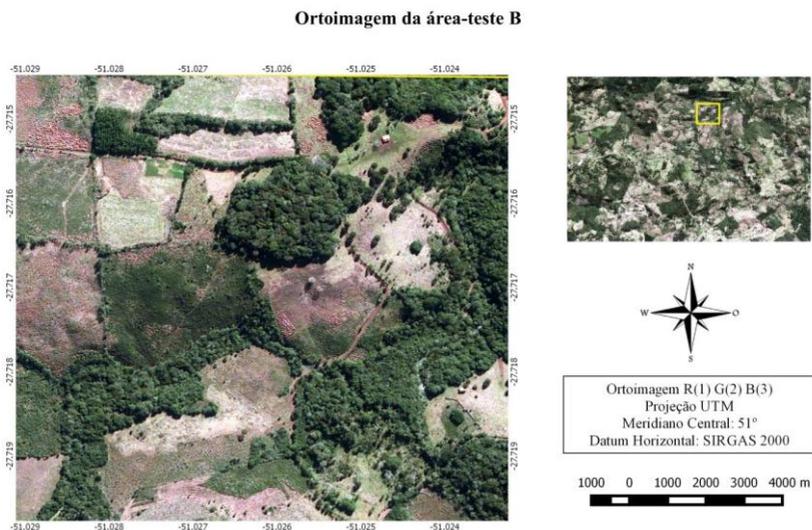
Figura 17 - Localização da área-teste A.
Ortoimagem da área-teste A



Fonte: produção do próprio autor, 2015.

A área-teste B foi escolhida por abranger um dos fragmentos florestais avaliados no Inventário Florístico Florestal de Santa Catarina (VIBRANS et al., 2012), correspondente à Unidade Amostral 261 (Figura 18). As características relatadas sobre esse fragmento florestal por Sevegnani et al. (2012) foram: aproximadamente 8 a 10 anos de regeneração; vegetação com pouca diversidade florística, sinúcias arbóreas empreendendo uma cobertura de menos de 50% constituída por poucos indivíduos de pequena estatura (8 m) e diâmetro, destacando-se as espécies: *Lithrea brasiliensis*, *Styrax leprosus*, *Myrcine coriaceae*, *Machaerium paraguariensi* e *Escallonia bifida* (SEVEGNANI et al., 2012).

Figura 18 - Localização da área-teste B.

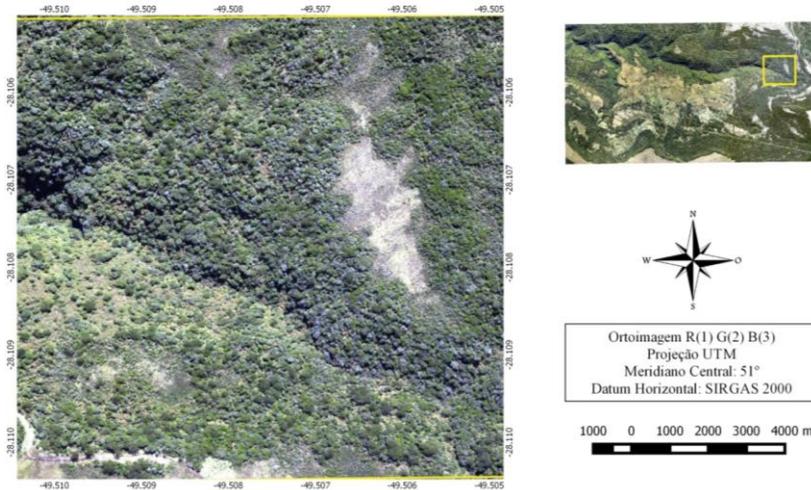


Fonte: produção do próprio autor, 2015.

A área-teste C está localizada no interior do Parque Nacional de São Joaquim, que compreende parte do município de Urubici. O município de Urubici possui área total de 1019 km², e está situado nas coordenadas médias 28°00'54" sul e longitude 49°35'30" oeste, em altitude de 915 m (IBGE, 2014). A criação do Parque Nacional de São Joaquim, em julho de 1961, está ligada à necessidade de proteção dos remanescentes de Matas de Araucárias, que se encontram dentro de seus 49.300 ha (ICMBIO, 2012), abrangendo os municípios de Urubici, Bom Jardim da Serra, Orleans e Grão-Pará. A área-teste C corresponde a uma parcela de vegetação que foi avaliada por Faxina (2014), na qual se constatou que a vegetação pertence, predominantemente, ao estágio avançado de regeneração, com destaque para as seguintes espécies:

Araucaria angustifolia, *Dicksonia sellowiana*, *Clethra uleana*, *Drimys angustifolia*, dentre outras (Figura 19).

Figura 19 - Localização da área-teste C.
Ortoimagem da área-teste C



Fonte: produção do próprio autor, 2015.

Segundo a classificação de Köppen (KÖPPEN e GEIGER, 1928), o clima das três áreas-teste é do tipo "Cfb", mesotérmico, subtropical úmido, com verões frescos, não apresentando estação seca definida.

3.2 MATERIAL

3.2.1 Imagens do recobrimento aéreo do estado de Santa Catarina

O aerolevanteamento fotográfico no Estado de Santa Catarina foi executado com recursos da Secretaria de Estado do Desenvolvimento Econômico Sustentável (SDS) nos anos de 2010 e 2011, e se obtiveram aproximadamente 57 mil ortofotografias. Este levantamento foi realizado com um Sistema

Aerotransportado de Aquisição e Pós-processamento de Imagens Digitais (SAAPI), com sensor CCD (*Charge Coupled Device*, “dispositivo de carga acoplado”), tamanho do *pixel* de 39 μm , e filtro UV-Sky, que filtra a luz ultravioleta e compensa o efeito de névoa atmosférica (PIAZZA, 2014).

De acordo com a Furlanetti (2013), o recobrimento aerofotogramétrico foi realizado do sentido leste para o oeste, acompanhando a estrutura geomorfológica de Santa Catarina. O fato de o estado ter várias regiões com elevação dificultou o planejamento dos voos, e, para evitar distorções na escala, aeronaves acompanharam todo o relevo pelo estado. As ortoimagens foram produzidas na escala 1:10.000, alcançando ampliações até as escalas de 1:2.000 e 1:1.000, mantendo qualidade gráfica/visual satisfatória (PIOVEZAN, 2013).

Os produtos gerados pelo aerolevantamento são a composição em cores verdadeiras (*Red, Green, Blue*- RGB) e a composição colorida no infravermelho próximo (*Near Infrared*- NIR). O levantamento também obteve o modelo digital de terreno (MDT) e o modelo digital de superfície (MDS) correspondente a cada imagem. Esses dados foram recebidos pela SDS com a etapa de pré-processamento, ajuste radiométrico, níveis de contraste, tonalidade, homogeneização das imagens, balanceamento de cores e ortorretificação, já realizadas (ENGEMAP, 2012). A Tabela 2 apresenta as características das ortoimagens.

Tabela 2 - Características das ortoimagens utilizadas. (Continua)

| Característica | Descrição |
|-------------------------------------|------------------|
| Ano de execução do aerolevantamento | 2010 e 2011 |
| Escala | 1: 10.000 |
| Altura média de voo | 3.985 m |
| Distância focal | 50 mm |
| Sensor | CCD |
| <i>Datum</i> | SIRGAS 2000 |

Tabela 2 - Características das ortoimagens utilizadas. (Conclusão)

| Característica | Descrição |
|--------------------------------------|---|
| Tamanho das imagens em <i>pixels</i> | RGB e NIR: 16000 (linha) x 12000 (coluna) MDS e MDT: 12000 (linha) x 7000 (coluna) |
| Área de cada ortoimagem | 25 km ² |
| Resolução espacial | Multiespectral: 39 cm MDS e MDT: 1 m |
| Resolução espectral | Azul: 0,38 – 0,52 µm Verde: 0,52 – 0,63 µm Vermelho: 0,63 - 0,70 µm IR1: 0,76 µm IR2: 0,77 µm IR3: 0,78 µm |
| Resolução radiométrica | Multiespectral: 8 bits MDS e MDT: 32 bits |

Fonte: produção do próprio autor, 2015, com dados de Furlanetti, 2013.

As ortoimagens estão classificadas na Classe A de acordo com o Padrão de Exatidão Cartográfica (PEC) (BRASIL, 1984). Possuem um erro padrão de 0,3 mm x escala, ou seja, 3 m no caso da escala 1:10.000.

Os dados de entrada utilizados no presente estudo compreendem as ortoimagens correspondentes às bandas espectrais do visível e infravermelho próximo, e o MDS, conforme a Tabela 3.

Tabela 3 - Dados de entrada das áreas-teste e respectivas bandas.

| Bandas | Área-teste | |
|---------------|-------------------------------|------------------------------|
| | A e B | C |
| R 1 | SG-22-Y-D-VI-2-SE-F_1.tif | SH-22-X-A-III-2-NE-F_1.tif |
| G 2 | SG-22-Y-D-VI-2-SE-F_2.tif | SH-22-X-A-III-2-NE-F_2.tif |
| B 3 | SG-22-Y-D-VI-2-SE-F_3.tif | SH-22-X-A-III-2-NE-F_3.tif |
| IR1 4 | SG-22-Y-D-VI-2-SE-F_IR1.tif | SH-22-X-A-III-2-NE-F_IR1.tif |
| IR2 5 | SG-22-Y-D-VI-2-SE-F_IR2.tif | SH-22-X-A-III-2-NE-F_IR2.tif |
| IR3 6 | SG-22-Y-D-VI-2-SE-F_IR3.tif | SH-22-X-A-III-2-NE-F_IR3.tif |
| MDS 7 | MDS_SG-22-Y-D-VI-2-SE-F_1.tif | MDS_SH-22-X-A-III-2-NE-F.tif |

Fonte: produção do próprio autor, 2015.

Criou-se uma composição multiespectral contendo todas as bandas espectrais e o MDS, com a ferramenta *concatenate images* (concatenação de imagens) no OTB Monteverdi. As ortoimagens foram recortadas em um tamanho de 1500x1500 *pixels*, visando maior agilidade e redução do custo computacional na aplicação das metodologias propostas. Além disso, alguns dos programas utilizados, como o SPT e InterIMAGE, apresentam limitações ao trabalharem com imagens grandes em computadores convencionais, os quais não dispõem de placa gráfica para processamento, ou de grande número de processadores paralelos e quantidade excessiva de memória volátil (RAM).

3.2.2 Programas utilizados

Para o desenvolvimento da pesquisa priorizou-se o uso de programas *open source* (Tabela 4), que, diferentemente de programas *Freeware*, além de serem totalmente gratuitos e sem nenhum tipo de restrição aos usuários, possuem o código fonte de programação acoplada a eles. Ou seja, os usuários podem modificar as funções do *software* sem a necessidade de nenhum tipo de pagamento aos desenvolvedores.

Tabela 4 - Programas utilizados e finalidade. (Continua)

| Programa | Finalidade |
|-----------------|---|
| SPT 3.0.3 | Avaliação e definição automática de parâmetros de segmentação. |
| InterIMAGE 1.43 | Segmentação, extração de atributos e classificação por árvore de decisão. |
| WEKA 3.7 | Seleção de atributos e geração de modelos de classificação por árvore de decisão. |
| QGIS 2.8.1 | Classificação das imagens com modelo definido no WEKA; geração e avaliação dos mapas temáticos. |
| GeoDMA 0.2.2 | Segmentação, extração de atributos e classificação por árvore de decisão. |

Tabela 4 - Programas utilizados e finalidade. (Conclusão)

| Programa | Finalidade |
|------------------------|--|
| OTB Monteverdi 1.22 | Concatenação das imagens, segmentação, classificação por região e por <i>pixel</i> , por máquinas de vetor de suporte. |

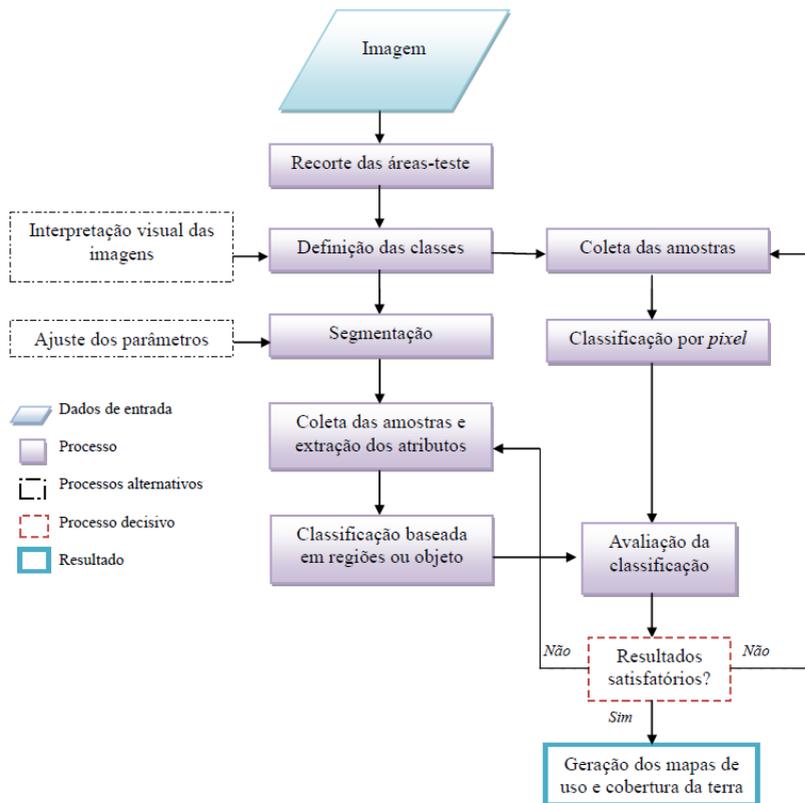
Fonte: produção do próprio autor, 2015.

3.3 CLASSIFICAÇÃO DO USO E COBERTURA DA TERRA

Conforme Giri (2012), existe diferença quando se fala em cobertura da terra ou uso da terra. O primeiro termo refere-se ao conjunto biótico e abiótico que se observa da superfície da Terra e imediato, como florestas, plantações e corpos d'água, sendo que mais recentemente o termo passou a incluir também as estruturas humanas, tais como estradas, áreas construídas, e as características do subsolo imediatas, como as águas subterrâneas. Já o uso da terra é definido como a forma ou maneira em que a terra é utilizada ou ocupada pelos seres humanos. Por exemplo, a área coberta por floresta pode ser utilizada para a produção, recreação, conservação e fins religiosos, dentre outros (GIRI, 2012).

Neste trabalho foram testados cinco diferentes procedimentos envolvendo métodos não-paramétricos para classificação do uso e cobertura da terra. Em três deles utilizou-se técnicas de mineração de dados, com algoritmos de árvores de decisão. Nos outros dois, foram utilizados algoritmos de máquinas de vetor de suporte, compreendendo as tradicionais classificações por *pixel* e por regiões. A sequência e organização de todos os procedimentos metodológicos estão dispostos no fluxograma mostrado na Figura 20. Os passos metodológicos relativos aos processamentos e avaliações específicas de cada metodologia realizada no estudo estão descritos nas seções seguintes.

Figura 20 - Fluxograma metodológico de todas as etapas desenvolvidas.



Fonte: produção do próprio autor, 2015.

3.3.1 Definição e caracterização das classes de uso e cobertura da terra

Nos levantamentos em campo realizados por Vibrans et al. (2012) e Faxina (2014), o estágio da vegetação foi definido conforme os critérios da resolução CONAMA nº 04/94 (BRASIL, 1994), que considera: DAP; altura das árvores; área basal; estratos predominantes; espécies indicadoras; diversidade e dominância de espécies; cobertura do dossel;

presença e características da serapilheira e sub-bosque; existência, diversidade e quantidade de epífitas e trepadeiras.

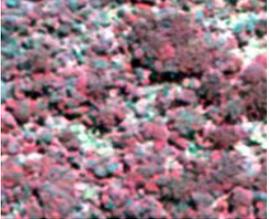
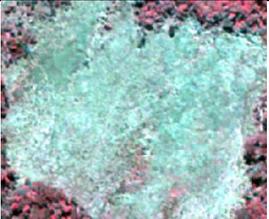
Para as demais classes presentes nas imagens, a definição da verdade terrestre foi feita através da interpretação visual das ortoimagens utilizando o MDS e a ferramenta *3D Analyst* do ArcGIS. Outros estudos também obtiveram a verdade terrestre aliando observações em campo com fotointerpretação visual (ARAÚJO, 2006; DURIEUX et al., 2009; CARVALHO, 2011) ou somente a partir da interpretação visual de imagens de alta resolução espacial (ALMEIDA FILHO et al., 2013; NASCIMENTO et al., 2013).

Entre as principais classes identificadas visualmente nas imagens têm-se: vegetação nos estádios inicial, médio e avançado de regeneração, culturas agrícolas e campo (Quadro 1).

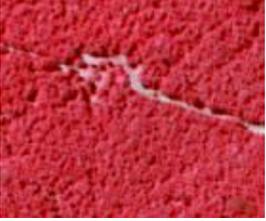
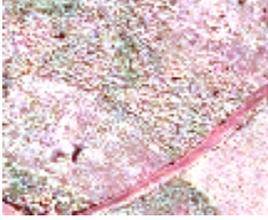
Quadro 1 - Descrição das classes encontradas nas áreas de estudo. (Continua)

| Classe | Composição RGB | Composição NIR |
|--|--|---|
| Vegetação em estádio avançado (VEA) |  |  <p data-bbox="329 1075 932 1286">Fisionomia arbórea com cobertura fechada, formando um dossel relativamente uniforme no porte. A textura se torna mais rugosa e não há sinais de dominância de espécies com distribuição agregada. A textura é heterogênea, com padrão espectral (cor) mais diversificado, o que mostra o aumento na diversidade (CINTRA, 2007).</p> |

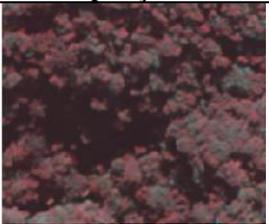
Quadro 1 - Descrição das classes encontradas nas áreas de estudo. (Continua)

| Classe | Composição RGB | Composição NIR |
|---|---|--|
| Vegetação em estágio médio (VEM) |  |  <p data-bbox="393 485 1008 692">Fisionomia arbustivo-arbórea, com cobertura densa e parcialmente fechada. Há início de diferenciação em estratos, observada nas diferenças de altura das árvores em diferentes pontos da imagem. São observadas algumas espécies com distribuição agregada, porém a textura revela um início de heterogeneidade (CINTRA, 2007).</p> |
| Vegetação em estágio inicial (VEI) |  |  <p data-bbox="393 920 1008 1038">Fisionomia herbáceo-arbustiva, com cobertura aberta ou fechada. Geralmente apresenta dominância de uma única espécie observada pela textura e padrão espectral (cor) mais homogêneos (CINTRA, 2007).</p> |
| Campo |  |  <p data-bbox="393 1267 1008 1316">Inclui campos antrópicos ou naturais. Textura lisa e maior reflectância (brilho).</p> |

Quadro 1 - Descrição das classes encontradas nas áreas de estudo. (Continua)

| Classe | Composição RGB | Composição NIR |
|-------------------------------|---|--|
| <p>Campo sujo</p> |  |  <p data-bbox="329 486 930 571">Campos que possuem algum tipo de cobertura que altera seu padrão espectral e/ou textural. Padrão espectral mais heterogêneo que a classe campo.</p> |
| <p>Reflorestamento</p> |  |  <p data-bbox="329 801 930 858">Plantio da espécie <i>Pinus</i> sp. Padrão espectral e espacial homogêneo.</p> |
| <p>Agricultura</p> |  |  <p data-bbox="329 1088 930 1145">Padrão espectral e textural variado conforme a cultura e fase de crescimento. Apresenta padrão espacial linear.</p> |
| <p>Uso antrópico</p> |  |  <p data-bbox="329 1375 930 1423">Edificações e benfeitorias. Tonalidades claras, textura lisa e maior reflectância (brilho) e formas geométricas.</p> |

Quadro 1 - Descrição das classes encontradas nas áreas de estudo. (Conclusão)

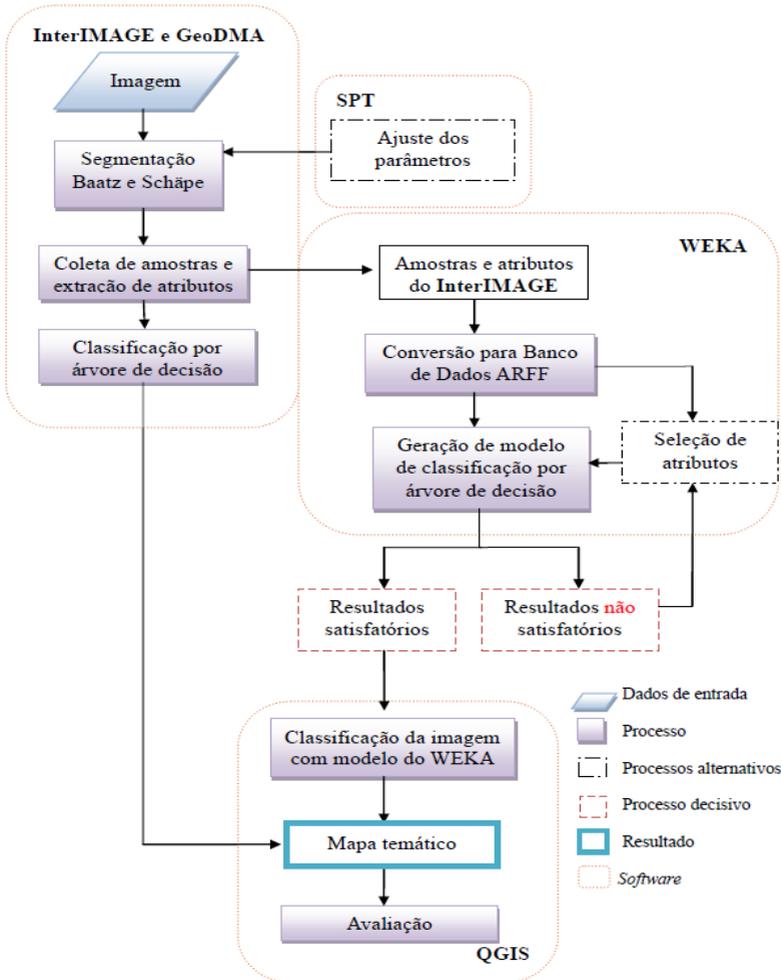
| Classe | Composição RGB | Composição NIR |
|---------------|--|--|
| Sombra |  |  |
| | Áreas escuras, tendendo a preto, textura lisa. Geralmente associada a vegetação de maior porte. | |

Fonte: produção do próprio autor, 2015.

3.3.2 Classificação por árvore de decisão

A classificação baseada em objeto através de árvores de decisão foi executada nos *softwares* InterIMAGE, WEKA aliado ao QGIS e GeoDMA (Figura 21).

Figura 21 - Fluxograma das etapas envolvendo classificação por árvores de decisão.



Fonte: produção do próprio autor, 2015.

3.3.2.1 Escolha e ajuste dos parâmetros da segmentação

Nas metodologias que envolveram a classificação por árvore e decisão, utilizou-se o algoritmo de segmentação proposto por Baatz e Schäpe (2000).

Este segmentador foi escolhido pela velocidade de execução e capacidade de extração de objetos homogêneos em uma mesma escala. Nele, são estabelecidos parâmetros de cor, forma e escala. Os parâmetros cor e forma determinam, respectivamente, o quanto da informação de cor e forma é usado no processo de segmentação. O parâmetro escala determina a média do tamanho do objeto a ser criado (BAATZ e SCHÄPE, 2000), sendo que, quanto mais alto o valor atribuído a esse parâmetro, maiores serão os objetos e, com isto, menor a quantidade de objetos criados. De acordo com Gao et al. (2011), o tamanho médio dos objetos na imagem tem um impacto significativo na acurácia da classificação. O algoritmo também permite atribuir peso a cada banda espectral da imagem de acordo com o seu grau de importância no processo de segmentação. O peso varia entre 0 e 1, e quanto maior o valor, maior o grau de importância conferida à respectiva banda de entrada, sendo que o valor 0 implica a banda é desconsiderada no processo de segmentação (BAATZ e SCHÄPE, 2000).

A escolha dos parâmetros de segmentação pelo algoritmo de Baatz e Schäpe seguiu os seguintes procedimentos: foram definidos alguns cenários de segmentação para as imagens; analogamente, procedeu-se à escolha automática dos parâmetros com o *software* SPT; na última etapa, avaliaram-se os resultados visualmente e através de métricas, selecionando-se os parâmetros mais bem ajustados a cada imagem.

Seguindo a mesma metodologia aplicada por Francisco e Almeida (2013) para a escolha visual dos parâmetros de escala, foram realizados três níveis de segmentação consecutivos, com a diminuição do fator de escala na execução de cada novo procedimento, originando, assim, um novo nível com maior número de objetos de tamanho reduzido. Com cada parâmetro de escala, aumentou-se gradativamente o peso do fator forma e diminuiu-se o do fator cor, e vice-versa. Dessa forma, foram definidos nove cenários de segmentação, conforme Tabela 5.

Em todos os testes, atribuiu-se igual peso para cada banda, tendo em vista que o *software* GeoDMA não permite a manipulação deste parâmetro.

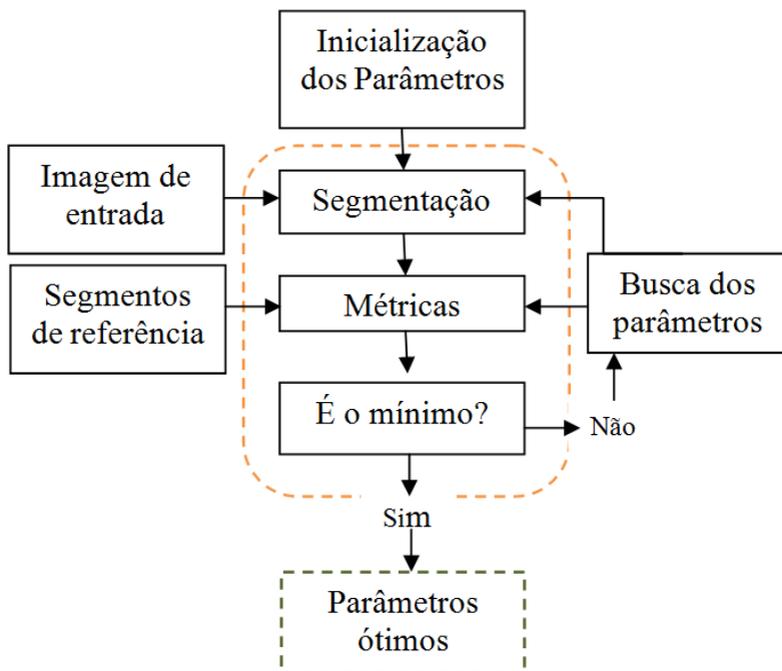
Tabela 5 - Parâmetros selecionados visualmente para a criação dos diferentes cenários de segmentação com algoritmo de Baatz e Schåpe.

| Cor | Forma | Escala |
|------------|--------------|---------------|
| 0.5 | 0.5 | 60, 80, 100 |
| 0.7 | 0.3 | 60, 80, 100 |
| 0.3 | 0.7 | 60, 80, 100 |

Fonte: produção do próprio autor, 2015.

A escolha automática e avaliação dos parâmetros de segmentação foram feitas no SPT, ferramenta gratuita desenvolvida para pesquisa pelo Laboratório de Visão Computacional (LVC) do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ). O SPT fornece uma abordagem para ajustar os valores dos parâmetros da segmentação, no qual um algoritmo de otimização pré-definido procura por valores ótimos. A otimização é estabilizada por uma função objetivo, que mede o grau de concordância entre a segmentação gerada e o conjunto de amostras de referência criados pelo usuário (Figura 22) (SPT 3.0 Guide, 2014).

Figura 22 - Metodologia implementada no SPT.



Fonte: Adaptado do SPT 3.0 Guide, 2014.

Para a escolha automática dos parâmetros, primeiramente criaram-se polígonos (segmentos) de referência que abrangessem classes presentes na imagem. Conforme Feitosa et al. (2007) não é necessário a segmentação de referência cobrir toda a imagem de entrada, o que pode ser muito oneroso computacionalmente, mas apenas delimitar alguns segmentos de referência. Esses segmentos foram utilizados para a escolha automática dos parâmetros de segmentação e para avaliar os parâmetros de segmentação escolhidos visualmente na Tabela 5.

Na etapa de otimização foi feita a busca automática de parâmetros para o algoritmo de Baatz e Schäpe com a função-objetivo *Nelder-mead* (NELDER e MEAD, 1965). No final do processo de otimização, o SPT gera um arquivo *.txt* com os

valores dos parâmetros ajustados para o segmentador, bem como uma imagem *raster* e um ESRI *Shapefile* com a segmentação gerada.

A avaliação da segmentação foi feita através de uma função de aptidão, ou métrica, que indica a qualidade da segmentação. O SPT possui sete métricas para avaliar o resultado da segmentação: *Hoover Index* (índice de Hoover-HI), *Área Fit Index* (índice de ajuste de área- AFI), *Shape Index* (índice de forma- SI), *Rand Index* (índice de Rand- RI), *Precision and Recall* (precisão e revocação- F), *Segmentation Covering* (cobertura da segmentação- S), *Reference Bounded Segments Booster* (Assistente para Segmentos Delimitados como Referência- RBSB). Neste trabalho, optou-se pelo RBSB, métrica proposta por Feitosa et al. (2006), também utilizada por Feitosa et al. (2009), Novack (2009), Kux et al. (2011) e Novack et al. (2011) e Leonardi et al. (2012), que corresponde à razão entre área de dois segmentos fora da interseção com a área de referência. Quanto mais próximo de zero, mais a segmentação gerada se aproxima da segmentação de referência, sendo que zero corresponde ao ajuste perfeito. De acordo com Ferreira (2011), a métrica RBSB mostra boa correlação com a percepção humana de qualidade de segmentação. A métrica é calculada pela Equação 3.

$$\text{RBSB} = \frac{1}{\text{NGT}} \sum_{t=1}^{\text{NGT}} \frac{fnt+fp_t}{fnt+tp_t} \text{Equação 3}$$

em que “t” representa um segmento de referência ou verdade terrestre (*Ground truth- GT*); NGT é o número de regiões na imagem GT; *fn* (*false negative*) *fp* (*false positive*) são falsos negativos e positivos, respectivamente e *tp* (*true positive*) são verdadeiros positivos (ACHANCCARAY DIAZ, 2014).

3.3.2.2 Classificação no InterIMAGE

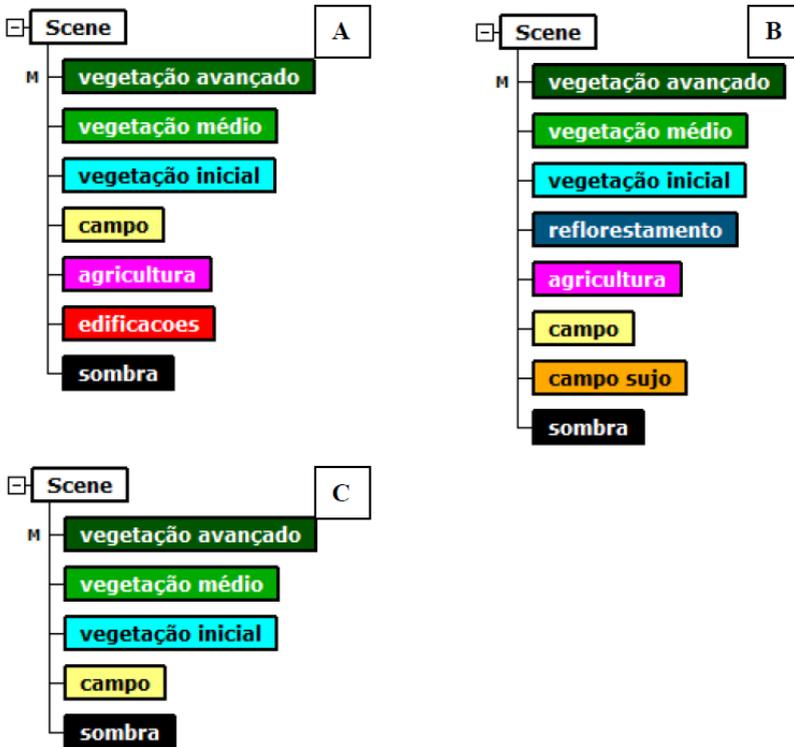
Uma das metodologias desenvolvidas no presente estudo consistiu na classificação automatizada das imagens no InterIMAGE, *software* de domínio público e código aberto desenvolvido pelo Laboratório de Visão Computacional/PUC-RJ em conjunto com a Divisão de Processamento de Imagens/INPE e a Universidade de Hannover (Alemanha) (BÜCKNER et al., 2001). Em termos de estratégia de interpretação, o InterIMAGE possui uma arquitetura flexível, combinando análise comandada por modelo (etapas *top-down*) e dados (etapas *bottom-up*), representando uma melhoria na eficiência computacional (RIBEIRO et al., 2011; MENEGHETTI, 2013).

Os procedimentos realizados na classificação das imagens no InterIMAGE foram: 1) construção das redes semânticas com as classes de uso e cobertura da terra previamente estipuladas; 2) segmentação da imagem com algoritmo de Baatz e Schäpe; 3) coleta de amostras e extração de atributos com a ferramenta *Samples editor* (editor de amostras); 4) classificação da imagem com o algoritmo por árvore de decisão C4.5; 5) avaliação da classificação.

A primeira etapa executada no InterIMAGE foi a elaboração da rede semântica, a qual representa a relação entre os objetos que se espera encontrar na imagem e a descrição hierárquica da resolução do problema de interpretação (NOVACK, 2009). No *software* InterIMAGE, o papel da rede semântica é representar a estratégia sequencial e organizacional da classificação (RODRIGUES, 2014), pois as classes herdam atributos e comportamentos das superclasses que as contêm. Contudo, neste trabalho, as redes semânticas criadas foram apenas operacionais, sem relação conceitual entre as classes, já que o objetivo foi explorar a classificação automatizada com o algoritmo C4.5. Desta forma, as redes semânticas das áreas-teste foram organizadas de forma que cada classe (nós-folha)

estivesse associada ao mesmo nó-pai, ou seja, uma rede sem níveis intermediários (Figura 23).

Figura 23 - Redes semânticas elaboradas para cada área-teste no InterIMAGE.



Fonte: produção do próprio autor, 2015.

Ribeiro et al. (2011) citam que redes com poucos níveis e muitos nós-folha são preferíveis para reduzir o custo computacional. Essa metodologia, no entanto, difere da utilizada por outros autores (NOVACK, 2009; MENEGHETTI, 2013; RODRIGUES, 2014), os quais elaboraram redes semânticas com vários níveis, colocando as classes mais distinguíveis entre si em níveis superiores (nós-pai), e as com maiores confusão nos inferiores (nós-filho). Porém, ressalta-se

que esses trabalhos visavam à classificação de áreas urbanas, e, portanto, com mais classes a serem discriminadas. Além disso, a classificação por árvore de decisão quando usada nesse *software* permite apenas um nível de segmentação na rede semântica.

Após elaboração da rede semântica, efetuou-se a segmentação das imagens com os parâmetros definidos com auxílio do *software* SPT. Com os segmentos gerados, procedeu-se à coleta das amostras na ferramenta “*Samples editor*” do InterIMAGE para as classes estipuladas, através do processo de amostragem aleatória simples (TRIOLA, 2014). As amostras foram coletadas de forma a abranger qualquer variação interna das classes quanto à cor, tonalidade, forma, textura e brilho. A quantidade de amostras também variou conforme a representatividade da área da classe em cada imagem. As classes com as respectivas quantidades de amostras para cada área-teste estão demonstradas na Tabela 6.

Tabela 6 - Número de amostras (objetos) e respectiva classe coletadas no InterIMAGE.

| Classe | Área-teste | | |
|----------------------------|------------|------------|------------|
| | A | B | C |
| Vegetação estágio inicial | 45 | 40 | 70 |
| Vegetação estágio médio | 50 | 60 | 35 |
| Vegetação estágio avançado | 85 | 60 | 70 |
| Reflorestamento | - | 10 | - |
| Agricultura | 60 | 45 | - |
| Campo | 50 | 40 | 30 |
| Campo sujo | - | 90 | - |
| Edificações | 10 | - | - |
| Sombra | 15 | 20 | 15 |
| Total de amostras | 315 | 365 | 220 |

Fonte: produção do próprio autor, 2015.

Em seguida, estipularam-se os atributos a serem extraídos de cada segmento. Foram gerados 47 atributos, sendo 43 deles espectrais e quatro operações entre bandas (Apêndice

A e Apêndice C), estas últimas escolhidas por explorar o contraste que a vegetação apresenta entre as bandas do visível e do infravermelho próximo. Em geral, pode-se considerar que quanto maior for o contraste entre elas, maior vigor terá a cobertura vegetal imageada (PONZONI et al., 2012).

Priorizaram-se atributos estatísticos ao invés dos espaciais, por se tratar predominantemente de áreas naturais, em que os objetos têm formas irregulares. Como citado por Yu et al. (2006), diferentemente da classificação de áreas urbanas, características geométricas têm pouca contribuição para classificação da vegetação em imagens de alta resolução espacial, já que esta não possui um padrão espacial óbvio que poderia ser evidenciado na classificação. Quanto às características texturais, apesar de se ter conhecimento que atributos de textura por Matriz de Co-Ocorrência de Níveis de Cinza (*Grey Level Co-occurrence Matrix- GLCM*) poderiam melhorar a classificação, o InterIMAGE apresentou problemas na extração desses atributos, impossibilitando sua utilização. Meneghetti (2013) e Rodrigues (2014) também relataram problemas relacionados ao tempo de processamento ao tentar extrair tais atributos por esse *software*.

Com as amostras e atributos gerados para os objetos resultantes da segmentação, procedeu-se à classificação das imagens. Essa etapa foi feita através do algoritmo *top down TA_C45_Classifier*, que utiliza o conceito de árvore de decisão proposto por Quinlan (1993). O algoritmo C4.5 foi implementado no *software* InterIMAGE para facilitar o processo de classificação, de forma que fosse realizada no mesmo programa em que se cria a rede semântica do projeto. Diferentemente do *software* WEKA, o InterIMAGE não oferece ao analista a possibilidade de manipular os parâmetros da árvore de decisão, gerando árvores de diversas dimensões (RODRIGUES, 2014). Ao optar por esse algoritmo no InterIMAGE, o usuário define apenas os parâmetros da segmentação, as amostras e atributos gerados, porém, o próprio

algoritmo irá definir, dentre esses atributos, quais serão usados na classificação.

O operador *TA_C45_Classifier* foi associado a apenas um dos nós-filho, o nó correspondente à classe VEA da rede semântica de cada imagem. Nesse nó, habilitou-se a opção *Multiclass*, ficando responsável por repassar as hipóteses aos demais nós da rede. A todos os nós restantes, atribuiu-se o operador *Dummy top-down*. Nesse operador nenhuma hipótese é criada, e as informações são apenas repassadas de nó-pai para nó-filho (RODRIGUES, 2014). No InterIMAGE, também foi executada a classificação da área-teste C sem o componente MDS. Tal procedimento visou verificar se o MDS, mesmo nesse caso sem fornecer a altura dos objetos, mas apenas a variação altimétrica das feições, influenciava positivamente na classificação da cobertura vegetal.

A classificação final resulta em um arquivo *shape* e a árvore de decisão é gerada em um arquivo de texto *.txt* à parte. Ressalta-se que no InterIMAGE, a classificação também pode ser gerada heurísticamente pelo usuário por meio da rede semântica, realizando a análise exploratória dos atributos e descobrindo os seus limiares para a discriminação adequada das classes (RODRIGUES, 2014). De acordo com Almeida (2014, apud RODRIGUES, 2014), há uma diferença entre a rede semântica e a árvore de decisão, são elas:

- a) na rede semântica, uma classe ou nó pode possuir um ou mais atributos, atuando de forma combinada, enquanto que, na árvore de decisão, há apenas um único atributo para diferenciar cada nó, independentemente do seu posicionamento na hierarquia da árvore;
- b) na rede, os limiares dos atributos podem ser *fuzzy* ou *crisp* (rígidos), já na árvore de decisão, os limiares adotados são sempre *crisp*;

- c) em uma rede semântica, uma classe até pode aparecer em diferentes nós eventualmente, porém, isso é raro, já que os nós da rede têm a capacidade de combinar atributos simultaneamente. Por outro lado, numa árvore de decisão, frequentemente uma mesma classe aparece em vários nós;
- d) quanto aos níveis de segmentação, a árvore de decisão opera com um único nível, enquanto a rede semântica pode operar com "n" níveis de segmentação.

Importante mencionar que a árvore de decisão elimina a possibilidade de relações contextuais interníveis, implicando uma subutilização do conceito de segmentação multirresolução, nos casos de plataformas em que esta funcionalidade se encontra disponível (ALMEIDA, 2014 apud RODRIGUES, 2014), como o InterIMAGE. Deve-se considerar, no entanto, que a elaboração de uma rede semântica conceitual consome maior tempo do analista (RIBEIRO et al., 2011), além de incidir sua subjetividade na determinação dos limiares.

3.3.2.3 Seleção de atributos e mineração de dados no WEKA

A ferramenta WEKA incorpora um conjunto de algoritmos de aprendizado de máquina que possibilita a extração do conhecimento (WEKA, 2013). Este aplicativo foi desenvolvido pela Universidade de Waikato da Nova Zelândia, apresentando interface amigável e rápido processamento dos dados. O WEKA possui um formato de arquivo próprio, denominado Formato de Arquivo Atributo-Relação (*Attribute-Relation File Format- ARFF*), arquivo de texto do tipo Código Americano Padrão para Intercâmbio de Informação (*American Standard Code for Information Interchange- ASCII*) que descreve uma lista de instâncias que compartilham um conjunto de atributos (CARVALHO, 2011). No banco de

dados, cada objeto amostral (instância) do conjunto de treinamento possui dois tipos de atributos: o atributo categórico, que indica a classe à qual a instância pertence; e os atributos preditivos (numéricos), que estão associados às respectivas classes.

A metodologia desenvolvida no WEKA compreendeu duas etapas: seleção de atributos e geração de modelo de classificação por árvore de decisão. Primeiramente, as amostras com os respectivos atributos foram exportadas do InterIMAGE em um arquivo *shape*. O banco de dados associado a este arquivo foi convertido para o formato *.csv* (valor separado por vírgula) através do programa Excel, para ser importado no WEKA.

Após o procedimento de conversão, procedeu-se à seleção de atributos, que envolve duas ferramentas a serem escolhidas pelo usuário: o avaliador de atributos e o método de busca. O avaliador determina qual método é usado para atribuir um valor a cada subconjunto de atributos, e o método de busca determina o tipo de busca a ser realizada (WEKA, 2013). Utilizou-se o avaliador CFS associado ao método de busca *BestFirst*. Outros autores, como Chou et al. (2007), Novack (2009) e Karegowda et al. (2010), tiveram bons resultados com este algoritmo. O CFS considera um conjunto de atributos “bom”, quando contém atributos altamente correlacionados com a classe e não-correlacionados entre si. A base deste método é uma heurística de avaliação de subconjuntos que considera não somente a utilidade de atributos individuais, mas também o nível de correlação entre eles (KAREGOWDA et al., 2010). O CFS primeiro calcula uma matriz de correlação de atributo-classe e atributo-atributo (NOVACK, 2009). Um peso (*score*) de um conjunto de atributos é associado usando-se a seguinte fórmula:

$$\text{Mérito (S)} = \frac{k \times r_{ac}}{\sqrt{k + k(k-1)r_{aa}}} \quad \text{Equação 4}$$

em que: $Mérito(S)$ é o mérito de um subconjunto de atributos S contendo k atributos; rac é a média da correlação entre atributo-classe; raa é a média da correlação entre atributo-atributo. O numerador pode ser visto como um indicador do poder preditivo do conjunto de atributos, e o denominador indica o “grau de redundância” que existe entre os atributos. O CFS começa com o conjunto vazio de atributos e usa a heurística *best-first-search* (método da melhor busca inicial) com um critério de parada de cinco subconjuntos consecutivos que não melhoram o mérito. O subconjunto com o maior mérito encontrado pela heurística será selecionado (HALL, 1998).

A etapa seguinte à seleção de atributos foi a classificação supervisionada do banco de dados. Dentre os algoritmos de classificação disponíveis no WEKA, as árvores de decisão se destacam por serem de fácil interpretação e visualização. As árvores de decisão particionam a base de dados em uma hierarquia, até que nas partes finais (nós-folha) restem apenas amostras de uma mesma classe. Esta divisão da base de dados fornece quais atributos e limiares usar na separação de duas ou mais classes (NOVACK, 2009) e pode ser facilmente convertida em regras de classificação (HAN e KAMBER, 2006).

Foram testados os classificadores por árvore de decisão J48, J48Graf, RandomTree, SimpleCart, e REPTree, e observados o índice Kappa e tamanho da árvore gerada. A avaliação dos classificadores foi feita com o teste de conjunto-suporte (*supplied test set*), em que 30% das amostras foram separadas do banco de dados de treinamento para serem usadas na avaliação. Essa etapa, além de permitir a comparação de desempenho entre diferentes classificadores sobre uma mesma base de dados de teste, possibilita uma estimativa da precisão do modelo ao classificar dados que não foram utilizados no processo de construção do modelo (MERSCHMANN, 2007). De maneira geral, buscou-se a árvore que atendia ao critério de

melhor acurácia e menor dimensionalidade, de forma a serem mais facilmente convertidas em regras para a classificação das imagens no QGIS. O classificador que melhor atendeu estes critérios foi o SimpleCart em todas as áreas-teste, sendo o algoritmo selecionado para proceder-se a classificação.

3.3.2.4 Classificação no QGIS

O modelo de classificação obtido no WEKA pode ser implementado em diferentes *softwares*. Autores como Carvalho (2011), Francisco e Almeida (2013) e Piazza (2014) inseriram os limiares do WEKA no *software* eCognition. Fernandes et al. (2013) construíram a árvore de decisão do WEKA no *software* ENVI. Como este trabalho busca soluções *open source*, optou-se por realizar essa etapa no QGIS, da mesma maneira que autores como Nascimento et al. (2013) e Pereira e Richter (2013). Também seria possível implantar a árvore no InterIMAGE, como feito por Novack (2009) e Souza (2012), porém, viu-se no QGIS uma maneira mais rápida de realizar esse processo, além da possibilidade de explorar as potencialidade de um GIS nesse procedimento.

O QGIS é um *software* livre de Sistema de Informações Geográficas (*Geographic Information System- GIS*), licenciado sob a “GNU General Public License”, projeto oficial da Fundação Geoespacial para Código Aberto (*Open Source Geospatial Foundation- OSGeo*). Esse *software* foi utilizado com o intuito de analisar a integração de Sensoriamento Remoto e GIS, tema que tem sido amplamente discutido nos últimos anos, principalmente com o advento de imagens de altíssima resolução espacial. Trabalhos como Ehlers (2007), Schiwe e Tufte (2007) e Leukert (2007) abordam dificuldades e potencialidades na integração de dados de Sensoriamento Remoto e GIS. Kok et al. (2007) citam que a sinergia entre Sensoriamento Remoto e GIS na análise orientada a objeto permite que problemas complexos sejam enfrentados, como a

classificação de dados contendo uma enorme quantidade de feições.

Para proceder com a construção do modelo WEKA no QGIS, primeiramente, converteu-se a árvore de decisão em regras, separando-as conforme a classe de uso e cobertura da terra à qual pertenciam. Utilizou-se no QGIS o arquivo *shape* resultante da segmentação e extração de atributos, ambos gerados no InterIMAGE, correspondente a cada área-teste. Na tabela de atributos desse arquivo, selecionaram-se as feições que satisfaziam as regras correspondentes a cada classe (Figura 24), e o campo “class” era preenchido conforme a classe. Esse processo repetiu-se, mudando-se as regras de decisão conforme a classe. Assim, ao final do processo, todos os segmentos estavam rotulados com as respectivas classes.

Figura 24 - Implementação do modelo WEKA no QGIS.

The screenshot displays the QGIS interface with the following components:

- Table: Tabela de atributo - amostras2 - Total**

| object_id | b4b4 | b4b1 | b4b1 | brilho | class |
|-----------|----------|------------|-------------|-------------|-----------------|
| 153 | 1.362623 | -54.479564 | 0.733079 | 254.962314 | campo_sujo |
| 154 | 1.553003 | 84.429462 | 0.643914 | 278.082120 | agricultura |
| 155 | 0.606900 | 24.060314 | 1.647718 | 155.009464 | pastura |
| 156 | 0.382063 | 102.779930 | 2.617371 | 189.844464 | reflorestamento |
| 157 | 0.110786 | 78.622102 | 3.121020 | 353.201620 | avencado |
| 158 | 0.843341 | 13.512438 | 1.188579 | 177.565223 | campo_sujo |
| 178 | 1.168889 | 0.676117 | 281.961548 | agricultura | agricultura |
| 10.066037 | 0.937017 | 236.862641 | agricultura | agricultura | agricultura |
| 18.221237 | 2.687410 | 176.859168 | avencado | avencado | avencado |
- Expressão Editor:**

Expressão: `"b4-b1" >= 44.4347227 and "maxp1x7" >= 805.471497`

2- Inserção das regras de cada classe
- Map View:**

3- Feições selecionadas de acordo com as regras

Fonte: produção do próprio autor, 2015.

3.3.2.5 Classificação no GeodMA

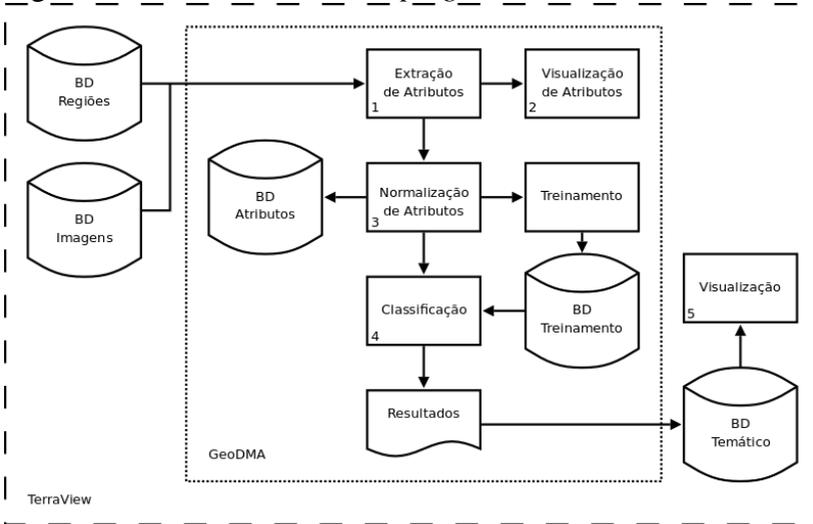
Körting et al. (2013) observaram que grande parte dos trabalhos utiliza mais de um programa computacional para

conduzir todas as etapas de OBIA. Da necessidade de um sistema único capaz de executar as tarefas, como a segmentação, extração de atributos, mineração de dados, reconhecimento de padrão e análise multitemporal, desenvolveu-se o GeoDMA.

O GeoDMA é um *software open source* desenvolvido pela Divisão de Processamento de Imagens no INPE. Assim como outros sistemas de interpretação de imagens, como o eCognition (DEFINIENS, 2009), ENVI-FX (ENVI, 2009), e o InterIMAGE (COSTA et al., 2008), o GeoDMA incorpora ferramentas apropriadas para a classificação orientada a objeto, com mineração de dados. Esse sistema não utiliza o conceito de rede semântica, como o InterIMAGE, pois, de acordo com Körting et al. (2013), essa é uma tarefa que demanda conhecimento e tempo do analista.

Esse aplicativo segue os padrões de código-fonte aberto (*Free and Open Source Software* - FOSS) e foi construído como uma extensão (*plug-in*) do GIS Terra View. Portanto, os mapas temáticos e demais resultados são produzidos rapidamente, aproveitando-se das funcionalidades deste GIS. A Figura 25 apresenta o fluxograma que descreve todas as fases de processamento no sistema GeoDMA. As bases de dados e atributos ficam armazenadas na estrutura de dados do Terra View, bem como a visualização dos resultados, através de mapas temáticos.

Figura 25 - Funcionamento do *plug-in* GeoDMA.



Fonte: Körting et al., 2009.

Na interface do Terra View, primeiramente criou-se um banco de dados e carregou-se a imagem a ser classificada. Para as etapas destinadas à classificação baseada em objeto, usou-se o *plug-in* GeoDMA, onde estão as ferramentas a serem utilizadas para segmentação, extração dos atributos e normalização, treinamento, visualização, classificação e avaliação.

Na segmentação, foi utilizado o algoritmo de Baatz e Schäpe com os mesmos parâmetros utilizados no InterIMAGE. Importante mencionar que, apesar de se utilizar os mesmos parâmetros, a segmentação não ficou idêntica ao InterIMAGE, pois o GeoDMA divide a imagem em blocos como forma de agilizar o procedimento, e segmenta cada bloco separadamente (Figura 26).

Figura 26 - Resultado da segmentação pelo GeoDMA. Visualizam-se os limites entre os blocos.



Fonte: produção do próprio autor, 2015.

Depois da segmentação, foram extraídos, no total, 46 atributos de cada objeto gerado. Nesse *software*, além dos parâmetros espectrais, utilizaram-se dois texturais (entropia GLCM e homogeneidade GLCM) e quatro espaciais (índice de forma- *shape index*, ajuste retangular- *rectangular fit*, densidade e perímetro razão-área). Também foram exportados esses mesmos atributos após o procedimento de normalização pela opção min-max, totalizando, portanto, 92 atributos para cada segmento (Apêndice B e Apêndice C).

Após as etapas da segmentação e extração de atributos, procedeu-se à coleta das amostras para cada classe. Nessa fase, foi criada a tipologia da classe de interesse, definida pelo nome e pela cor. Em seguida, foram selecionadas as amostras na imagem e associadas à respectiva classe criada. A informação da seleção das amostras é armazenada como um rótulo em uma coluna específica do banco de dados dos objetos, e estas são usadas para o treinamento do modelo de classificação e para avaliar os resultados. Ao proceder-se à coleta de amostras, o

software separa automaticamente 40% para serem utilizadas na avaliação da classificação (Tabela 7).

Tabela 7 - Amostras (objetos) e respectiva classe coletadas no GeoDMA, separadas automaticamente entre treinamento e validação.

| Classe/ área-teste | Treinamento | | | Validação | | |
|----------------------------|-------------|------------|------------|------------|-----------|-----------|
| | A | B | C | A | B | C |
| Vegetação estádio inicial | 23 | 20 | 37 | 15 | 14 | 25 |
| Vegetação estádio médio | 27 | 22 | 17 | 17 | 15 | 11 |
| Vegetação estádio avançado | 57 | 20 | 51 | 36 | 13 | 33 |
| Reflorestamento | - | 5 | - | - | 3 | - |
| Agricultura | 27 | 14 | - | 19 | 10 | |
| Campo | 24 | 19 | 15 | 13 | 10 | 12 |
| Campo sujo | - | 49 | - | - | 30 | - |
| Edificações | 4 | - | - | 2 | - | - |
| Sombra | 8 | 7 | 3 | 6 | 4 | 3 |
| Total de amostras | 170 | 156 | 123 | 108 | 99 | 84 |

Fonte: produção do próprio autor, 2015.

A classificação foi feita através do algoritmo supervisionado por árvore de decisão C4.5 (QUINLAN, 1993). Diferentemente do InterIMAGE, o GeoDMA permite escolher quais dos atributos extraídos serão utilizados na classificação por esse algoritmo. Optou-se por selecionar todos, deixando para o algoritmo a definição daqueles que seriam utilizados na construção da árvore de decisão. Além disso, o algoritmo C4.5 do GeoDMA permite intervir nos parâmetros da árvore de decisão, como o número mínimo de objetos por folha. Desta forma, testaram-se algumas variações neste parâmetro, e através da opção de avaliação, que confronta as amostras de treinamento com as amostras de validação, pode-se escolher a de maior acurácia.

3.3.3 Classificação por máquinas de vetor de suporte

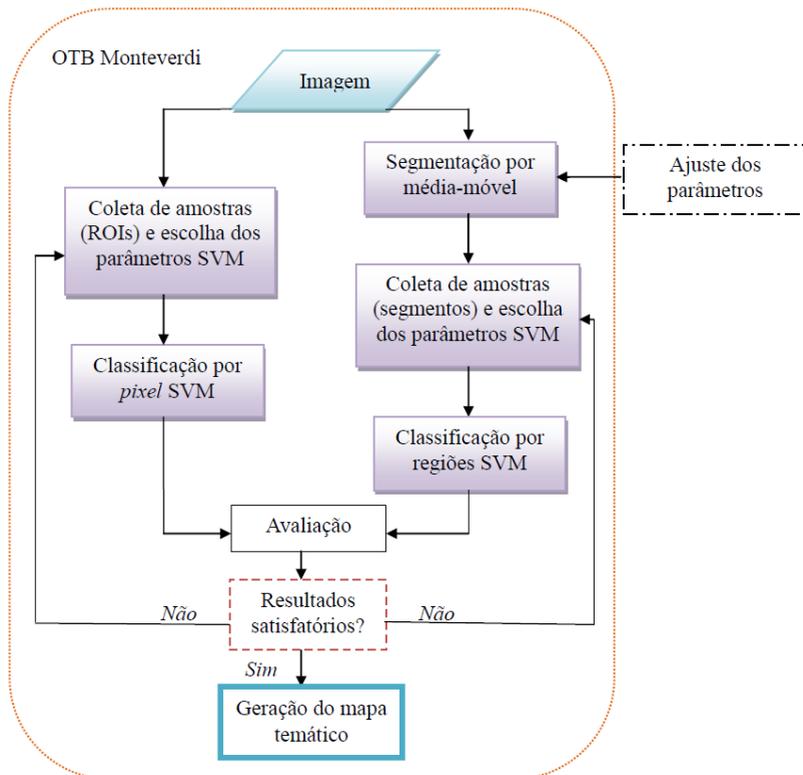
O algoritmo classificador SVM foi utilizado na classificação baseada em regiões e *pixel-a-pixel* no módulo Monteverdi do *software* OTB.

O OTB é um aplicativo de Sensoriamento Remoto e Processamento Digital de Imagens, desenvolvido e mantido pelo Centro Nacional de Estudos Espaciais (*Centre National d'Études Spatiales* - CNES) da França em consórcio com a COSMO-SkyMed da Itália, desde 2006 (CHRISTOPHE e INGLADA, 2009). É distribuído como uma biblioteca de código aberto CeCILL para processamento de imagem e oferece funcionalidades específicas para o processamento de imagens de Sensoriamento Remoto, dispondo de algoritmos direcionados para imagens de alta resolução geométrica (SPOT, Quickbird, WorldView, Ikonos), sensores hiperespectrais (Hyperion) ou radar de abertura sintética (TerraSarX, ERS, PALSAR). O Monteverdi é um módulo do OTB que permite a construção de cadeias de processamento, selecionando rotinas a partir de um conjunto de menus. Suporta dados nos formatos vetoriais e *raster* e dá acesso a funcionalidades OTB em uma arquitetura modular (OTB Guide, 2013).

A classificação orientada a objeto utilizando o conceito de rede hierárquica ou semântica não foi implementada no pacote Monteverdi até o momento. No entanto, ele fornece o algoritmo SVM, que tem sido amplamente utilizado nas classificações de imagens de Sensoriamento Remoto (MOUNTRAKIS et al., 2010), além de permitir a integração do SVM com a classificação por regiões (VU, 2012).

O fluxograma dos procedimentos desenvolvidos no Monteverdi pode ser visualizado na Figura 27.

Figura 27 - Fluxograma das etapas desenvolvidas no OTB Monteverdi.



Fonte: produção do próprio autor, 2015.

3.3.3.1 Classificação por regiões

Para proceder-se à classificação por regiões no Monteverdi, primeiramente foi realizada a escolha dos parâmetros de segmentação da imagem. O algoritmo disponível no Monteverdi para a segmentação é o média-móvel. Este algoritmo permite a escolha de três parâmetros: raio espacial (*spatial radius*), valor espectral (*spectral value*) e o tamanho mínimo de cada segmento (*min region size*).

Como os resultados obtidos com o SPT não foram satisfatórios para esse algoritmo, procedeu-se com a escolha dos parâmetros através de tentativa e erro, partindo de uma supersegmentação até um ajuste, de forma que os segmentos fossem representativos dos alvos presentes na cena (Tabela 8).

Tabela 8 - Parâmetros avaliados visualmente para segmentação pela média-móvel.

| Raio espacial | Valor espectral | Escala |
|---------------|-----------------|--------------|
| 5 | 10 | 80, 100, 120 |
| 5 | 15 | 80, 100, 120 |
| 5 | 20 | 80, 100, 120 |
| 10 | 10 | 80, 100, 120 |
| 10 | 15 | 80, 100, 120 |
| 10 | 20 | 80, 100, 120 |
| 20 | 10 | 80, 100, 120 |
| 20 | 15 | 80, 100, 120 |
| 20 | 20 | 80, 100, 120 |

Fonte: produção do próprio autor, 2015.

A segmentação no Monteverdi gera quatro arquivos de saída: imagem filtrada (*filtered image*), imagem agrupada (*clustered image*), imagem rotulada (*labeled image*) e limites da imagem (*boundaries image*). Ao selecionar o classificador SVM por regiões, selecionou-se o arquivo “imagem rotulada” associado à imagem de interesse, e procedeu-se à coleta das amostras (Tabela 9).

Tabela 9 - Número de amostras (segmentos) coletadas para a classificação por regiões no Monteverdi. (Continua)

| Classe | Área-teste | | |
|----------------------------|------------|----|-----|
| | A | B | C |
| Vegetação estádio inicial | 45 | 60 | 150 |
| Vegetação estádio médio | 50 | 75 | 40 |
| Vegetação estádio avançado | 70 | 85 | 160 |
| Reflorestamento | - | 10 | - |
| Agricultura | 60 | 45 | - |
| Campo | 35 | 45 | 35 |

Tabela 9 - Número de amostras (segmentos) coletadas para a classificação por regiões no Monteverdi. (Conclusão)

| Classe | Área-teste | | |
|--------------------------|------------|------------|------------|
| | A | B | C |
| Campo sujo | - | 135 | - |
| Edificações | 8 | - | - |
| Sombra | 13 | 8 | 15 |
| Total de amostras | 281 | 463 | 400 |

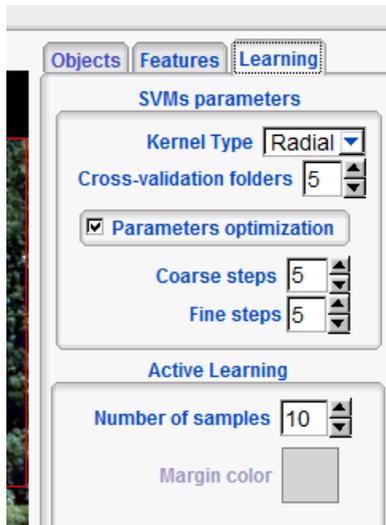
Fonte: produção do próprio autor, 2015.

Para o aprendizado e classificação por regiões, o SVM utiliza atributos espaciais (elongação e forma) e estatísticos (média, variância, assimetria e curtose de cada banda).

Na etapa da classificação, deve-se definir ainda a função-núcleo (*kernel*). Ressalta-se que diferentes definições de núcleos e seus respectivos parâmetros provocam alterações nos resultados fornecidos pelo SVM (SOUSA et al., 2009). Atualmente não existe um método universal para guiar a seleção desses parâmetros (BONESSO e HAERTEL, 2013). No entanto, a FBR é comumente utilizada na literatura para a classificação de imagens de Sensoriamento Remoto (HUANG et al., 2002; NASCIMENTO et al., 2009; ANDREOLA e HAERTEL, 2010; BONESSO e HAERTEL, 2013; ANDRADE et al., 2014). Por esse motivo, optou-se por utilizar esta função-núcleo.

Na classificação SVM por regiões, o aplicativo Monteverdi busca automaticamente os melhores parâmetros da função-núcleo escolhida ao se selecionar a opção *parameters optimization* (otimização de parâmetros). O programa também fornece a acurácia da classificação em função do conjunto de treinamento pelo procedimento “*cross-validation folders*” ou validação cruzada (Figura 28).

Figura 28 - Parâmetros do algoritmo SVM na classificação por regiões.



Fonte: produção do próprio autor, 2015.

A técnica de validação cruzada permite uma avaliação prévia da acurácia da classificação do SVM. Primeiramente, o conjunto de amostras de treinamento é dividido em “n” subconjuntos do mesmo tamanho. Na sequência, um subconjunto é testado usando o classificador treinado nos subconjuntos restantes (n-1). Assim, cada instância de todo conjunto de treinamento é predita uma vez. A acurácia da validação cruzada é a porcentagem de dados que foram corretamente classificados nesse procedimento (HSU et al., 2003). Dessa forma, a validação cruzada pode evitar o problema de *overfitting* ou superajustamento (HSU et al., 2003).

3.3.3.2 Classificação *pixel-a-pixel*

Com o intuito de comparar as classificações anteriores, baseadas em objeto ou regiões, realizou-se uma metodologia

aplicada somente na área-teste B, seguindo os métodos de classificação por *pixel*, com o algoritmo SVM. Nessa metodologia, ao invés dos segmentos, utiliza Regiões de Interesse (ROIs) como amostras. Cabe ao usuário definir o tamanho de cada ROI e o número de *pixels* para cada classe (Tabela 10).

Tabela 10 - Número de amostras (*pixels*) para cada classe na área-teste B.

| Classe | Nº <i>pixels</i> |
|----------------------------|------------------|
| Vegetação estádio inicial | 1678 |
| Vegetação estádio médio | 2698 |
| Vegetação estádio avançado | 2593 |
| Reflorestamento | 567 |
| Agricultura | 2122 |
| Campo | 1988 |
| Campo sujo | 3576 |
| Sombra | 965 |
| Total de amostras | 16187 |

Fonte: produção do próprio autor, 2015.

Optou-se pela função núcleo FBR para a classificação por *pixel*. A metodologia por *pixel* difere da classificação por regiões quanto à escolha dos parâmetros da função núcleo, já que não fornece a opção de otimização. Na função núcleo FBR o usuário define o parâmetro de margem (C), que determina um ponto de equilíbrio razoável entre a maximização da margem e a minimização do erro de classificação, e o parâmetro γ (gama), que controla o raio do núcleo FBR (BONESSO e HAERTEL, 2013). Algumas variações dos parâmetros foram testadas, e escolheu-se a que resultou em uma classificação visual mais satisfatória, que foram os valores de 1,5 para gama e 1 para C.

3.4 AVALIAÇÃO DOS RESULTADOS

Neste estudo, utilizaram-se mapas de referência de cada área-teste, os quais foram elaborados a partir da fotointerpretação, para avaliar os mapas classificados de forma automática. Executou-se o cruzamento dos mapas classificados com os mapas de referência através da ferramenta *intersect* no QGIS. Dessa forma, obteve-se a matriz de confusão, e calculou-se a exatidão global, exatidão do usuário e produtor, Kappa global e Kappa condicional de cada classe. Essas medidas foram calculadas em função das áreas (m²) corretamente e incorretamente classificadas de cada classe, já que a referência foi todo o mapa, e não apenas amostras.

A exatidão global foi calculada, dividindo-se o total de área corretamente classificada de cada classe nos mapas classificados (soma dos elementos ao longo da diagonal principal) pela área correspondente a cada classe no mapa de referência.

Também foi calculada a exatidão do usuário e do produtor. A exatidão do usuário reflete os erros de comissão que indicam a probabilidade de um elemento classificado em uma determinada classe realmente pertencer a essa classe (LILLESAND et al., 2004), ou seja, incluir um objeto na classe que ele não pertence. Essa medida corresponde à razão entre elementos classificados corretamente pelo número de elementos atribuídos a esta classe (Equação 5).

$$Eu = \frac{x_{ii}}{x_{i+}} \text{Equação 5}$$

em que: Eu = exatidão do usuário; x_{ii} = elementos classificados corretamente; x_{i+} = total de elementos classificados para uma classe i .

Já a exatidão do produtor corresponde à razão entre o número de elementos (neste caso, área) de uma classe classificadas corretamente e o número total de elementos

classificados para esta classe (Equação 6) (CONGALTON e GREEN, 1999; LILLESAND et al., 2004), refletindo os erros de omissão, ou seja, a probabilidade de um objeto ser excluído (não classificado) da classe a que ele pertence.

$$E_p = \frac{x_{ii}}{x+i} \text{ Equação 6}$$

sendo: E_p = exatidão do produtor; x_{ii} = elementos classificados corretamente; $x+i$ = total de elementos de referência para uma classe i .

O índice Kappa (CONGALTON e GREEN, 1999) se diferencia da exatidão global por incorporar os elementos fora da diagonal principal. Esse índice foi calculado pela Equação 7.

$$K = \frac{n \sum_{i=1}^c x_{ii} - \sum_{i=1}^c x_i + \sum_{i=1}^c x_i}{n^2 - \sum_{i=1}^c x_i + \sum_{i=1}^c x_i} \text{ Equação 7}$$

em que: K é uma estimativa do coeficiente Kappa; x_{ii} é o valor na linha i e coluna i ; $x_i +$ é a soma da linha i ; $x + i$ é a soma da coluna i da matriz de confusão; n é o número total de elementos de referência; c o número total de classes. Esse índice varia de -1 a 1, e pode ser interpretado de acordo com a Tabela 11.

Tabela 11 - Categorização de uma classificação a partir do valor do índice Kappa.

| Valor Kappa | Qualidade da classificação |
|-------------|----------------------------|
| < 0,00 | Péssima |
| 0,00 – 0,20 | Ruim |
| 0,20 – 0,40 | Razoável |
| 0,40 – 0,60 | Boa |
| 0,60 – 0,80 | Muito Boa |
| 0,80 – 1,0 | Excelente |

Fonte: Landis e Koch, 1977.

O índice Kappa condicional (ROSENFELD e FITZPATRICK-LINS, 1986) permite avaliar a acurácia de uma determinada classe. É calculado com base no mesmo princípio do Kappa utilizado para a avaliação global da classificação, e a interpretação dos resultados obedece aos mesmos princípios (ANDRADE et al., 2014). O índice Kappa condicional do usuário avalia os erros de comissão (Equação 8), e o Kappa condicional do produtor avalia os erros de omissão (Equação 9).

$$K = \frac{(n \times xii) - (xi + \times x + i)}{(n \times xi +) - (xi + \times x + i)} \text{Equação 8}$$

$$K = \frac{(n \times xii) - (xi + \times x + i)}{(n \times x + i) - (xi + \times x + i)} \text{Equação 9}$$

em que: n é o número total de elementos de referência; xii é o número total de elementos classificados corretamente; $xi +$ é o total de elementos classificados para uma categoria i ; $x + i$ é o total de elementos de referência amostrados para uma categoria i .

Depois de obtido o índice Kappa de cada classificação, foi executado o teste z para testar a significância estatística da diferença entre as classificações resultantes de cada metodologia. Esse teste foi executado no aplicativo Excel, descrito conforme Congalton e Green (1999). Atribuiu-se um nível de significância de 5% ($\alpha = 0,05$), com valor crítico de 1,96, ou seja, assumiu-se que se o valor do teste z fosse maior que o valor crítico haveria diferença significativa entre os mapeamentos. A estatística para testar a diferença entre as duas matrizes de erro é obtida pela Equação 10:

$$z = \frac{K2 - K1}{\sqrt{\sigma_{K1}^2 - \sigma_{K2}^2}} \text{Equação 10}$$

em que: K_1 = índice Kappa da imagem 1; K_2 = índice Kappa da imagem 2; e σ^2 = variância do índice Kappa.

4. RESULTADOS E DISCUSSÃO

4.1 SEGMENTAÇÃO DAS IMAGENS

4.1.1 Segmentação e ajuste dos parâmetros do algoritmo de Baatz e Schäpe

A busca automática de parâmetros no SPT para o segmentador de Baatz e Schäpe foi realizada com 100 iterações, escala variando de 30 a 100 e pesos de cor e forma variando de 0 a 1.

Os resultados obtidos podem ser observados na Tabela 12, na qual o RBSB reflete o valor de discrepância da segmentação de referência com a segmentação com parâmetros escolhidos pelo analista, e, em negrito, com os parâmetros escolhidos pela função-objetivo *Nelder Mead*. Os resultados com maior discrepância para cada área-teste estão representados em vermelho.

Tabela 12 - Resultado dos parâmetros de segmentação de Baatz e Schäpe avaliados pelo SPT em cada área-teste. (Continua)

| Área-teste | Escala | wCor | wForma | Discrepância |
|------------|------------|------------|------------|---------------|
| A | 60 | 0,5 | 0,5 | 0,0260 |
| | 60 | 0,7 | 0,3 | 0,1398 |
| | 60 | 0,3 | 0,7 | 0,0488 |
| | 80 | 0,5 | 0,5 | 0,0260 |
| | 80 | 0,7 | 0,3 | 0,0903 |
| | 80 | 0,3 | 0,7 | 0,0912 |
| | 100 | 0,5 | 0,5 | 0,1792 |
| | 100 | 0,7 | 0,3 | 0,0663 |
| | 100 | 0,3 | 0,7 | 0,0486 |
| | 88 | 0,5 | 0,5 | 0,0260 |
| B | 60 | 0,5 | 0,5 | 0,1065 |
| | 60 | 0,7 | 0,3 | 0,0860 |
| | 60 | 0,3 | 0,7 | 0,0439 |
| | 80 | 0,5 | 0,5 | 0,0941 |
| | 80 | 0,7 | 0,3 | 0,0762 |
| | 80 | 0,3 | 0,7 | 0,1367 |

Tabela 12 - Resultado dos parâmetros de segmentação de Baatz e Schäpe avaliados pelo SPT em cada área-teste. (Conclusão)

| Área-teste | Escala | wCor | wForma | Discrepância |
|----------------|-----------|------------|------------|---------------|
| | 100 | 0,5 | 0,5 | 0,0941 |
| | 100 | 0,7 | 0,3 | 0,0704 |
| | 100 | 0,3 | 0,7 | 0,1966 |
| | 81 | 0,5 | 0,5 | 0,0436 |
| C ² | 60 | 0,5 | 0,5 | 0,1517 |
| | 60 | 0,7 | 0,3 | 0,0900 |
| | 60 | 0,3 | 0,7 | 0,0928 |
| | 80 | 0,5 | 0,5 | 0,1114 |
| | 80 | 0,7 | 0,3 | 0,0885 |
| | 80 | 0,3 | 0,7 | 1 |
| | 100 | 0,5 | 0,5 | 0,2014 |
| | 100 | 0,7 | 0,3 | 0,2003 |
| | 100 | 0,3 | 0,7 | 1 |
| | 52 | 0,5 | 0,5 | 0,0842 |

Fonte: produção do próprio autor, 2015.

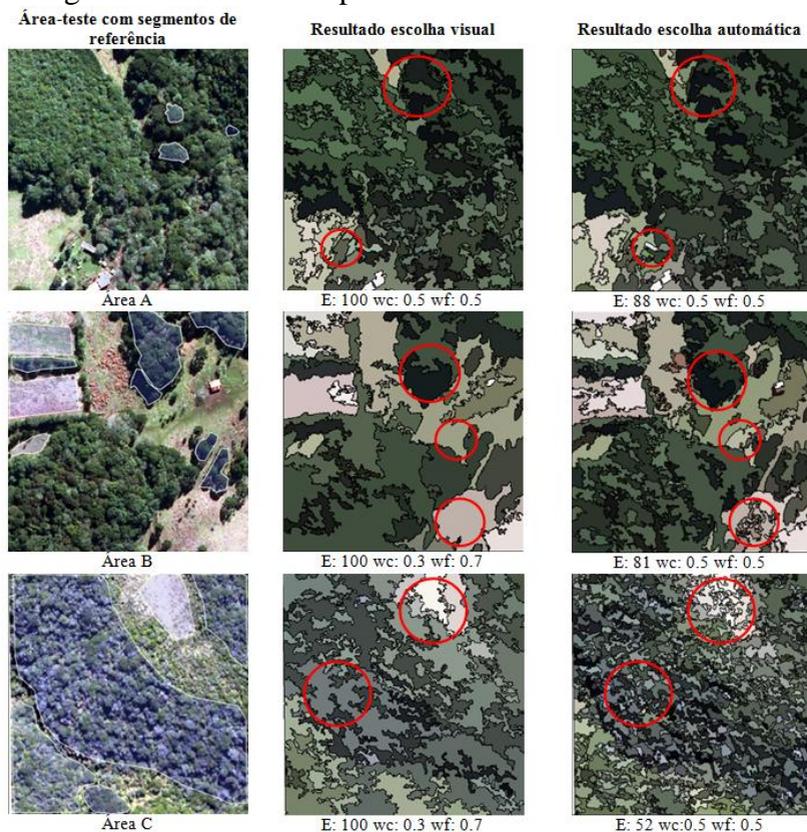
Percebe-se que, em todas as imagens, a segmentação com os parâmetros escolhidos automaticamente com a função-objetivo no SPT obteve menores valores de discrepância. Alguns cenários de segmentação da área-teste C resultaram em discrepância igual a 1, indicando não haver relação entre os segmentos de referência e a segmentação com aqueles parâmetros, o que pode ser parcialmente atribuído à maior complexidade dos segmentos de referência dessa área-teste. Porém, a maioria dos parâmetros ficaram com discrepância próxima a 0, o que indica uma boa delimitação dos segmentos de referência.

A Figura 29 apresenta recortes de imagens nas áreas-teste e respectivos resultados, utilizando-se os parâmetros escolhidos pelo analista que obtiveram maior discrepância, e a escolha automática dos parâmetros. Em vermelho, têm-se os principais detalhes que diferenciam as duas segmentações, e mostram o

² Para esta área-teste escolheu-se os parâmetros E: 80, wc: 0,7 e wf: 0,3.

quanto a escolha dos parâmetros pode vir a influenciar na acurácia da futura classificação.

Figura 29 - Segmentos de referência; resultados da segmentação com escolha visual e automática dos parâmetros do algoritmo Baatz e Schäpe.



Legenda: E= escala; wc= peso da cor; wf= peso da forma.

Fonte: produção do próprio autor, 2015.

Na área-teste A, apesar de a única diferença entre a segmentação com parâmetros definidos visualmente e parâmetros gerados de forma automática ser no fator de escala, de 100 e 80 respectivamente, algumas diferenças podem ser

notadas nos resultados. O círculo vermelho inferior indica local em que uma das edificações não foi delimitada com a escala de 100, o que certamente acarretará erros na classificação dessa classe. No círculo superior da imagem, observa-se que a classe sombra e parte da vegetação estão delimitados em um mesmo segmento, não constituindo, dessa forma, um objeto puro.

Na área-teste B, os círculos inferior e central indicam locais em que o maior fator de escala dos parâmetros visuais (100) e menor peso atribuído à cor (0,3) não delimitaram alguns objetos, como pequenas árvores. O círculo superior indica local em que as classes vegetação e sombra encontram-se no mesmo segmento na segmentação gerada com parâmetros escolhidos pelo analista.

Na área-teste C, a escolha automática dos parâmetros de segmentação não teve um resultado satisfatório com o fator de escala escolhido pelo SPT. Observa-se que ocorreu uma supersegmentação da imagem. Esse fato pode ser atribuído à maior diversidade espectral dentro dos segmentos de referência, já que se tratava de vegetação em estágio avançado de regeneração, e pela maior complexidade dos segmentos de referência. Nessa área optou-se por utilizar os parâmetros visuais que obtiveram a menor discrepância.

Dentre os parâmetros escolhidos visualmente pelo analista, os que obtiveram menor discrepância foram os que priorizaram o peso da cor, com valores de 0,7. Isso pode ser atribuído ao fato de as áreas serem predominantemente naturais, em que as classes têm formas irregulares, e não simétricas, como ocorreriam em áreas urbanas. Souza (2012), ao classificar áreas urbanas com imagens Quickbird, atribuiu valores de 0,9 de forma e 0,4 para a cor, e valores menores de escala. Tedesco et al. (2014), ao realizarem a classificação de voçorocas em imagens IKONOS, atribuíram peso 0,9 à cor e 0,1 à forma.

Dessa forma, observa-se que os parâmetros de segmentação não são genéricos, ou seja, mesmo que sejam

considerados “ideais” para determinada cena, podem não ser adequados para outra, fazendo-se necessária uma avaliação objetiva e quantitativa para a escolha do melhor resultado do segmentador (NASCIMENTO, 1997). Segundo Ribeiro e Kux (2009), os processos de segmentação e de descrição de classe têm de ser adaptados para cada estudo de caso, mesmo que o fluxo de trabalho de classificação permaneça o mesmo.

De acordo com Baatz e Mimler (2007), na maioria dos casos, os objetos de imagem iniciais são resultado de uma segmentação não idêntica com os objetos de interesse, sendo que a sub ou supersegmentação tem sido a regra. Neste caso, preferiu-se a supersegmentação, que resulta em mais segmentos para um mesmo alvo, mas não inclui partes de alvos distintos dentro de um segmento. A Tabela 13 mostra os parâmetros de segmentação utilizados para cada área-teste, tendo sido apenas para a área-teste C escolhido um dos cenários definidos pelo analista, e para as demais, empregaram-se os fornecidos pelo aplicativo SPT.

Tabela 13 - Parâmetros utilizados na segmentação pelo algoritmo de Baatz e Schäpe.

| Área-teste | wCor | wForma | Escala | Objetos gerados |
|------------|------|--------|--------|-----------------|
| A | 0,5 | 0,5 | 88 | 2188 |
| B | 0,5 | 0,5 | 81 | 2683 |
| C | 0,7 | 0,3 | 80 | 1628 |

Fonte: produção do próprio autor, 2015.

4.1.2 Segmentação com o algoritmo de média-móvel

Para escolha dos parâmetros da segmentação pela média-móvel, primeiramente empregou-se o aplicativo SPT, porém, percebeu-se que os parâmetros fornecidos pelo programa não foram viáveis, já que geraram uma supersegmentação das imagens (Figura 30).

Figura 30 - Áreas-teste A e B (esquerda) e resultados da segmentação com parâmetros fornecidos pelo SPT (direita).



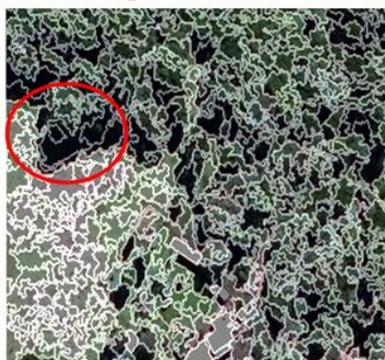
Legenda: RE: raio espacial; VE: valor espectral; E: escala.

Fonte: produção do próprio autor, 2015.

Percebe-se que o SPT não forneceu parâmetros de segmentação adequados para esse algoritmo pela avaliação do analista. Além disso, algumas rotinas executadas com o algoritmo de Baatz e Schäpe, como a avaliação da qualidade da segmentação através de métricas (discrepância), não foram executadas para o algoritmo de média móvel, devido a problemas ocorridos durante o processamento.

Para a escolha dos parâmetros por tentativa e erro, um fator decisivo foi a delimitação da classe sombra. Para as demais classes, percebeu-se que, mesmo alterando-se os parâmetros de forma a reduzir o número de segmentos, a estrutura dos objetos (como edificações) ainda permanecia adequadamente delimitada. No entanto, parte da vegetação começou a ser abrangida no mesmo segmento que a sombra. A Figura 31 mostra essa constatação, com dois resultados obtidos para as áreas-teste A e B. Em vermelho, destaque para alguns segmentos de sombra.

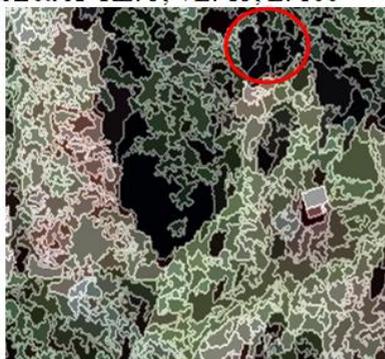
Figura 31 - Resultado de quatro cenários de segmentação avaliados para as áreas-teste A e B.



Área A- RE: 5; VE: 15; E: 100



RE: 15; VE: 20; E: 120.



Área B- RE: 5; VE: 15; E: 80



RE:15; VE:20; E:120

Legenda: RE= Raio espacial; VE= Valor espectral; E= Escala.

Fonte: produção do próprio autor, 2015.

Após a avaliação visual dos resultados, optou-se pelos parâmetros discriminados na Tabela 14.

Tabela 14 - Parâmetros escolhidos para a segmentação pelo algoritmo de média móvel.

| Área-teste | Raio espacial | Valor espectral | Escala | Objetos gerados |
|------------|---------------|-----------------|--------|-----------------|
| A | 5 | 15 | 100 | 5224 |
| B | 5 | 15 | 100 | 6183 |
| C | 10 | 15 | 80 | 2327 |

Fonte: produção do próprio autor, 2015.

4.2 MINERAÇÃO DE DADOS

4.2.1 Seleção de atributos e descrição das classes de uso e cobertura da terra

Entre os 47 atributos gerados no InterIMAGE (Apêndice A), o algoritmo CFS selecionou o subconjunto dos que possuíam maior relevância e que não fossem redundantes para discriminar as classes em cada área-teste (Tabela 15).

Tabela 15 - Número de atributos selecionados e mérito dos subconjuntos selecionados pelo algoritmo CFS no WEKA.

| | Área-teste | | |
|------------------------------|------------|-------|-------|
| | A | B | C |
| Nº de atributos selecionados | 16 | 15 | 18 |
| Mérito do subconjunto | 0,721 | 0,659 | 0,714 |

Fonte: produção do próprio autor, 2015.

Percebe-se que o algoritmo CFS descartou vários dos atributos gerados. A área-teste B teve o subconjunto com o menor mérito. Isto está associado ao fato de que a área possui mais classes espectralmente semelhantes, como quatro tipologias vegetais e duas classes de campo. Dessa forma, atributos do subconjunto selecionado são mais redundantes na discriminação das classes do que nas outras áreas-teste. A

Tabela 16 mostra os atributos selecionados pelo CFS para cada área-teste.

Tabela 16 - Atributos selecionados pelo algoritmo CFS no WEKA.

| Área-teste | | |
|----------------------------|----------------------------|----------------------------|
| A | B | C |
| Divisão da B1 pela B4 | Subtração da B4 pela B1 | Divisão da B1 pela B4 |
| Subtração da B4 pela B1 | Divisão da B4 pela B1 | Subtração da B4 pela B1 |
| Divisão da B4 pela B1 | Média da B4 | Divisão da B4 pela B1 |
| Entropia da B3 | Média da B5 | Entropia da B3 |
| Entropia da B4 | Média do MDS | Entropia da B4 |
| Entropia da B6 | Max. valor de pixel no MDS | Média da B3 |
| Média da B5 | Min. valor de pixel na B6 | Média da B4 |
| Média do MDS | Min. valor de pixel no MDS | Média da B5 |
| Max. valor de pixel na B6 | NDVI ³ | Max. valor de pixel na B3 |
| Max. valor de pixel no MDS | Razão da B4 | Max. valor de pixel no MDS |
| Min. Valor de pixel no MDS | Razão da B5 | Min. valor de pixel no MDS |
| NDVI ³ | Razão da B6 | NDVI ³ |
| Razão da B4 | Desvio padrão da B2 | Razão da B3 |
| Razão da B5 | Desvio padrão da B3 | Razão da B5 |
| Razão da B6 | Desvio padrão da B4 | Razão da B6 |
| Desvio padrão da B4 | | Desvio padrão da B2 |
| | | Desvio padrão da B3 |
| | | Desvio padrão da B7 |

Fonte: produção do próprio autor, 2015.

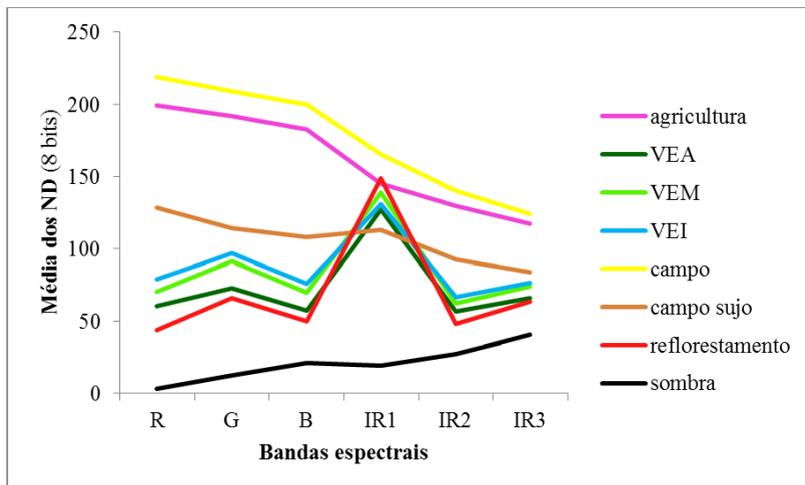
³ Índice de Vegetação por Diferença Normalizada (*Normalized Difference Vegetation Index*- NDVI): Proposto por Rouse et al. (1973), este índice é a normalização da razão simples entre os números digitais (no caso deste estudo) das bandas espectrais correspondentes à região do vermelho (R) e infravermelho próximo (IR1), para o intervalo de -1 a +1 (Apêndice C).

A subtração entre a banda espectral correspondente à região do infravermelho próximo (B4) pela banda correspondente à região do vermelho (B1), além do NDVI, foram atributos selecionados em todas as áreas-teste, mesmo tendo sido calculados em função dos números digitais (ND).

Ponzoni et al. (2012) falam da necessidade da conversão dos ND das imagens para valores de radiância ou reflectância, para possibilitar a caracterização espectral dos alvos, já que um valor de ND de uma imagem em uma banda específica não está na mesma escala de outro ND de outra imagem ou outra banda espectral. Como o aerolevanteamento foi executado por uma empresa privada, tratada diretamente com a SDS, não foi possível adquirir informações detalhadas em relação à calibração do sensor ou do aerolevanteamento, a fim de possibilitar a conversão atmosférica dos ND para reflectância de superfície.

Porém, apesar da impossibilidade de se realizar a conversão neste estudo, percebe-se que o comportamento espectral encontrado em função do ND mostra-se coerente com os reportados em outros trabalhos (VIEIRA et al., 2003; PONZONI e REZENDE, 2004; PIAZZA, 2014) (Figura 32).

Figura 32 - Perfil espectral das classes de uso e cobertura da terra da área-teste B.



Fonte: produção do próprio autor, 2015.

Entre as classes de cobertura vegetal, a vegetação em estágio inicial (VEI) apresentou maiores valores de ND ao longo do espectro, em função da maior regularidade e homogeneidade dos dosséis, com exceção do estágio médio (VEM), que apresentou valores maiores que o inicial na região do IR1. Tal comportamento pode ser atribuído ao fato de que a vegetação em estágio inicial na área-teste B apresenta semelhanças com algumas áreas de campo sujo, classe esta que teve valores menores de ND na banda IR1. Para a vegetação em estágio avançado (VEA), observa-se que a maior heterogeneidade dos dosséis, com maior presença de sombras, resultou em ND menores que os outros estádios.

Ponzoni e Rezende (2004), ao classificarem os estádios sucessionais com imagens Landsat 5 TM, também observaram que o estágio intermediário apresentou valores de reflectância maiores que as demais tipologias vegetais na região do infravermelho próximo. Os autores estranharam esse comportamento, já que seria esperado que, à medida que a

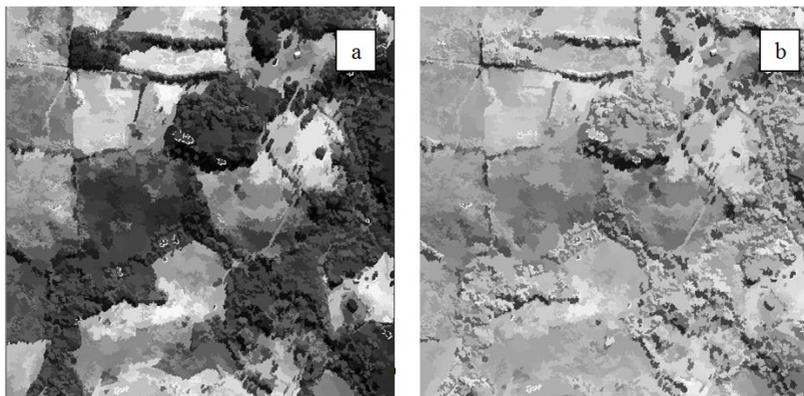
rugosidade do dossel aumentasse, sua reflectância nessa região espectral deveria decrescer devido ao sombreamento mútuo de um estrato ou árvores dominantes que se projetam na parte superior do dossel. Vieira et al. (2003) caracterizaram três estádios de vegetação secundária na Amazônia com imagens Landsat 7 ETM, e observaram que o estágio avançado manteve valores menores em relação aos iniciais, corroborando com o comportamento observado neste estudo.

Piazza (2014), ao classificar os estádios sucessionais da vegetação utilizando ortoimagens advindas do aerolevante fotogramétrico de Santa Catarina, obteve perfil espectral em relação ao ND similares a este estudo, inclusive para a classe de reflorestamento, que teve valores menores de ND que a vegetação natural na região do visível. Segundo Mather e Koch (2011) apud Francisco e Almeida (2012), a classe reflorestamento possui alta correlação entre as bandas do visível, em geral com baixa resposta, e elevada resposta na banda do infravermelho próximo, maior inclusive do que a apresentada por floresta natural. Este comportamento também ficou evidente no perfil espectral desta classe mostrado na Figura 32.

De acordo com Ponzoni et al. (2012), quando o dossel florestal é dividido em dois ou três estratos verticais, espera-se que ele apresente tonalidade mais escura dos que as demais classes da vegetação nas bandas do visível, pela maior atividade fotossintética; tonalidade clara na banda do infravermelho próximo em razão do espalhamento múltiplo da radiação eletromagnética por parte das folhas; e tonalidade novamente escura na imagem da banda do infravermelho médio, devido à maior quantidade de folhas contendo água no seu interior. A Figura 33 ilustra a espacialização das médias da banda espectral vermelho (B1), em que a vegetação aparece em tons escuros, e IR1 (B4), em tons claros, indicando, portanto, valores mais altos de ND. Tal comportamento justifica o fato

de a subtração entre essas duas bandas e o NDVI terem sido atributos selecionados em todas as áreas-teste.

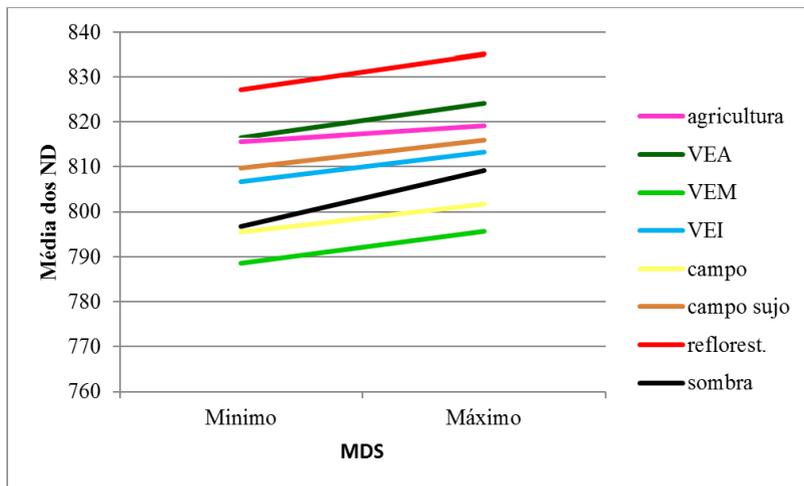
Figura 33 - Espacialização da B1 (a) e B4 (b) na área-teste B.



Fonte: produção do próprio autor, 2015.

Outros atributos selecionados para todas as áreas-teste foram o máximo e mínimo valor de *pixel* do MDS. Como o MDS representa apenas a variação altimétrica das feições em função das altitudes, as alturas dos objetos não são conhecidas, porém, a informação de textura atrelada a essa componente é alta, desempenhando um importante papel ao discriminar as classes, o que pode ser visualizado na Figura 34.

Figura 34 - Comportamento espectral das classes de uso e cobertura da terra em função do MDS na área-teste B.



Fonte: produção do próprio autor, 2015.

Na Figura 34, percebe-se que os valores mínimos e máximos dos ND referentes ao MDS de todas as classes não se sobrepõem, indicando se tratar de um bom atributo a ser utilizado na árvore de decisão. Nas três áreas-teste, também foram selecionados os atributos razão das bandas 5 e 6 (IR2 e IR3) e a média da banda 5, demonstrando a importância da composição infravermelho na determinação das classes.

Piazza (2014), ao executar a seleção de atributos pelo algoritmo *Gain Ratio* (ganho de informação) encontrou entre os mais relevantes o NDVI, média das bandas 3 e 4, menor e maior valor de *pixel* no MDS em ortofotografias do mesmo levantamento aerofotogramétrico que este estudo, corroborando com os resultados encontrados.

4.2.2 Modelos de classificação por árvores de decisão

Após a etapa de seleção de atributos no WEKA, testaram-se os algoritmos de árvores de decisão J48 (C4.5),

J48Graf, RandomTree, REPTree e SimpleCart e obteve-se a acurácia dos modelos através do índice Kappa, bem como o tamanho das árvores geradas, que corresponde ao número de ramificações (Tabela 17).

Tabela 17 - Resultado dos modelos de classificação por árvore de decisão testados no WEKA.

| Modelo | Critério | Área-teste | | |
|------------|----------|------------|------|------|
| | | A | B | C |
| J48 | Tamanho | 45 | 73 | 35 |
| | Kappa | 0,68 | 0,68 | 0,73 |
| J48Graft | Tamanho | 99 | 139 | 36 |
| | Kappa | 0,68 | 0,70 | 0,73 |
| RandomTree | Tamanho | 117 | 145 | 61 |
| | Kappa | 0,66 | 0,62 | 0,70 |
| REPTree | Tamanho | 25 | 29 | 15 |
| | Kappa | 0,68 | 0,68 | 0,72 |
| SimpleCart | Tamanho | 25 | 17 | 15 |
| | Kappa | 0,68 | 0,71 | 0,72 |

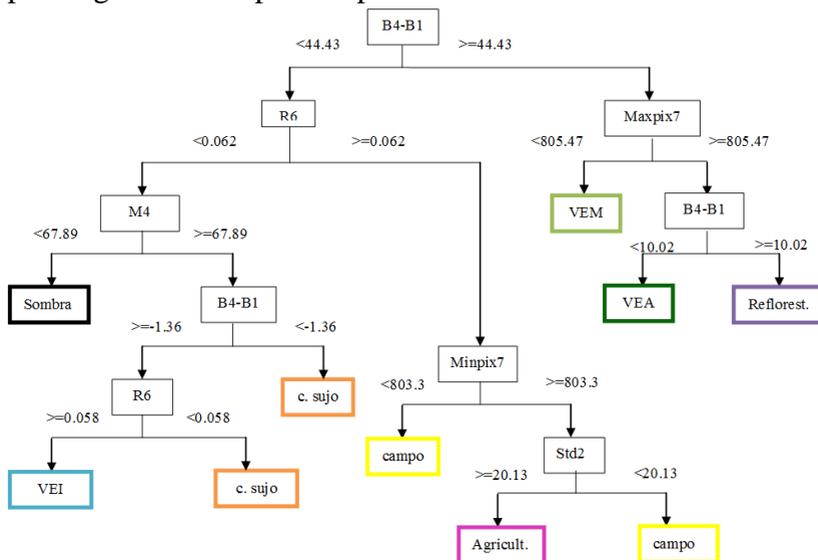
Fonte: produção do próprio autor, 2015.

Na Tabela 12, percebe-se que todos os modelos apresentaram acurácia semelhante em cada área-teste. A principal diferença ocorreu no tamanho da árvore gerada. Os algoritmos J48Graft e RandomTree tenderam a gerar árvores complexas, sendo que o RandomTree apresentou os menores valores de Kappa dentre os modelos. O REPTree e o SimpleCart foram os classificadores que obtiveram melhores resultados, obtendo boa acurácia com árvores menos complexas. Considerando o custo computacional de implantação do modelo no QGIS, optou-se pelo classificador SimpleCart para todas as áreas-teste.

O SimpleCart permite a formação de apenas dois ramos (dois filhos) em cada etapa do processo, ou seja, a árvore de decisão gerada será sempre uma árvore binária (TSO e MATHER, 2009). Sato et al. (2013), ao realizarem um comparativo entre doze algoritmos de árvore de decisão

disponíveis no WEKA, concluíram que, com exceção do DecisionStump, todos obtiveram bons resultados, sendo que o SimpleCart se destacou pela eficiência na determinação das classes com menor número de testes lógicos. A Figura 35 ilustra a árvore de decisão SimpleCart gerada a partir dos atributos selecionados pelo CFS para a área-teste B.

Figura 35 - Demonstração gráfica da árvore de decisão gerada pelo algoritmo SimpleCart para a área-teste B no WEKA.



Fonte: produção do próprio autor, 2015.

Percebe-se que o atributo correspondente à subtração entre as bandas IR1 e vermelho (B4-B1) prevaleceu na árvore de decisão, sendo utilizado como atributo principal e mais duas vezes como atributo secundário, o que corrobora o observado no perfil espectral das classes e espacialização das imagens correspondentes a essas bandas (Figuras 32 e 33). Também se nota que o máximo valor de *pixel* correspondente ao MDS foi o atributo escolhido para discriminar o estágio médio e avançado e a classe reflorestamento, mostrando a contribuição dessa

banda na discriminação das tipologias vegetais. Observa-se que, na área-teste B, o mínimo valor de *pixel* correspondente ao MDS foi utilizado na distinção das classes campo e agricultura, que também são classes espectralmente semelhantes (Figura 32), porém com comportamento espectral distinto no MDS (Figura 34). Na árvore de decisão, vê-se que a vegetação em estágio inicial (VEI) não está no mesmo ramo das demais classes de vegetação, e sim junto com a classe campo sujo (c. sujo). Tal fato deve-se à maior presença de solo na VEI, bem como pelo fato de a classe campo sujo possuir áreas com vegetação mais esparsa.

Por se tratar de três áreas-teste e três metodologias empregando árvore de decisão, foram geradas 10 árvores no total, sendo quatro no InterIMAGE, já que uma das classificações foi sem utilizar o MDS, denominada InterIMAGE (2) (Apêndice D), três no WEKA (Apêndice E) e três no GeoDMA (Apêndice F). As principais características de cada uma delas são discriminadas na Tabela 18.

Tabela 18 - Características principais das árvores de decisão geradas em cada *software*. (Continua)

| Software | Algoritmo | Características da árvore | Área-teste | | |
|----------------|------------|---------------------------|------------|-------|---------|
| | | | A | B | C |
| InterIMAGE | C4.5 | Tamanho | 42 | 52 | 24 |
| | | Nº atributos | 13 | 17 | 10 |
| | | Nº nós | 21 | 27 | 13 |
| | | Atributo principal | B1divB4 | B4-B1 | NDVI |
| InterIMAGE (2) | C4.5 | Tamanho | | | 26 |
| | | Nº atributos | - | - | 11 |
| | | Nº nós | | | 14 |
| | | Atributo principal | | | NDVI |
| WEKA/QGIS | SimpleCart | Tamanho | 25 | 17 | 15 |
| | | Nº atributos | 8 | 6 | 6 |
| | | Nº nós | 13 | 10 | 8 |
| | | Atributo principal | B4divB1 | B4-B1 | B1divB4 |

Tabela 18 - Características principais das árvores de decisão geradas em cada *software*. (Conclusão)

| <i>Software</i> | Algoritmo | Características da árvore | Área-teste | | |
|-----------------|-----------|---------------------------|------------|------------|------------|
| | | | | | |
| GeoDMA | C4.5 | Tamanho | 40 | 32 | 26 |
| | | Nº atributos | 17 | 11 | 11 |
| z | | Nº nós | 21 | 17 | 14 |
| | | Atributo principal | rp_ratio_0 | rp_ratio_4 | rp_ratio_0 |

Fonte: produção do próprio autor, 2015.

A Tabela 18 mostra que as árvores geradas variaram conforme o *software* utilizado e a área-teste. O algoritmo árvore de decisão SimpleCart aliado à seleção de atributos resultou em árvores mais compactas e com menor número de atributos para todas as áreas-teste. Já o InterIMAGE e o GeoDMA resultaram em árvores de tamanhos similares, diferenciando-se apenas de acordo com área-teste e número de atributos utilizados.

Durante a análise da árvore de decisão, os dados mais importantes se encontram nos nós-folhas mais próximos ao nó-raiz (ALMEIDA, 2010). De maneira geral, as operações entre bandas foram os grandes-nós, ou seja, os atributos principais das árvores geradas pelo InterIMAGE e WEKA, demonstrando a relevância de tais atributos ao discriminar a vegetação das demais classes. Nas árvores geradas pelo GeoDMA, por não ter esses atributos, a razão da banda 1 (rp_ratio_0) e razão da banda 5 (rp_ratio_4) foram escolhidas.

Os atributos texturais GLCM extraídos no GeoDMA contribuíram na construção da árvore de decisão, com destaque para a homogeneidade da banda 4 e do MDS, usados para discriminar o estágio médio do avançado nas áreas-teste A e B respectivamente, e entropia da banda 3 para distinção desses estágios na área-teste C (Apêndice F). A aplicação da informação textural é apontada em vários estudos para melhor discriminação de classes de vegetação (YU et al., 2006;

SETTE, 2009; FURTADO et al., 2013), e de classes antrópicas (FRANCISCO e ALMEIDA, 2012; RODRIGUES, 2014).

4.3 RESULTADOS E AVALIAÇÃO DAS CLASSIFICAÇÕES DO USO E COBERTURA DA TERRA

Foram gerados 14 mapas temáticos do uso e cobertura da terra, apresentados na escala 1:4.000. Todos os mapas também mostram a ortoimagem original, composição R1G2B3, e o mapa de referência elaborado pelo fotointérprete, o qual foi utilizado como verdade terrestre.

O mapa de referência fornece a área em m² referente a cada classe de uso e cobertura da terra encontrada nas áreas de estudo (Tabela 19). Esse mapa foi cruzado com os resultados das classificações, possibilitando a obtenção da matriz de confusão. As matrizes de confusão são apresentadas em porcentagem, já que algumas classes são consideravelmente menos representativas em termos de área do que outras. Os dados destas matrizes foram arredondados para uma casa decimal após a vírgula. Desta forma, quando os valores de matrizes mostram 100% de acerto, trata-se de uma aproximação de tais valores.

Tabela 19 - Área e porcentagem correspondente a cada classe nos mapas de referência das áreas de estudo.

| Área-teste A | | | Área-teste B | | | Área-teste C | | |
|--------------|------------------------|------------|-----------------|------------------------|------------|--------------|------------------------|------------|
| Classe | Área (m ²) | Área (%) | Classe | Área (m ²) | Área (%) | Classe | Área (m ²) | Área (%) |
| VEA | 102.541,0 | 36,97 | VEA | 22.266,7 | 9,18 | VEA | 135.644,6 | 47,50 |
| VEM | 65.866,4 | 23,75 | VEM | 62.651,3 | 25,82 | VEM | 21.964,0 | 7,69 |
| VEI | 30.237,2 | 10,90 | VEI | 12.322,2 | 5,08 | VEI | 114.020,2 | 39,93 |
| agricultura | 49.151,4 | 17,72 | agricultura | 17.415,8 | 7,18 | campo | 13.917,0 | 4,87 |
| campo | 29.551,6 | 10,66 | campo | 31.973,5 | 13,18 | sombra | 0,5 | 0,0002 |
| edificação | 0,1 | 0,00003 | campo sujo | 95.974,9 | 39,56 | | | |
| sombra | 1,2 | 0,00041 | reflorestamento | 0,3 | 0,0001 | | | |
| | | | sombra | 0,7 | 0,0003 | | | |
| Total | 277.348,9 | 100 | Total | 242.605,5 | 100 | Total | 285.546,3 | 100 |

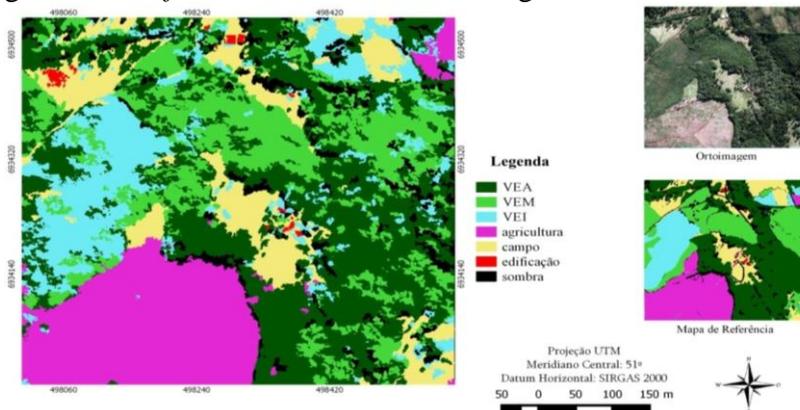
Fonte: produção do próprio autor, 2015.

4.3.1 Resultado da classificação por árvores de decisão

Foram gerados dez mapas e dez matrizes de confusão, como resultado dos três métodos de classificação por árvore de decisão, sendo que, a área-teste C foi classificada duas vezes no InterIMAGE, uma delas sem o MDS.

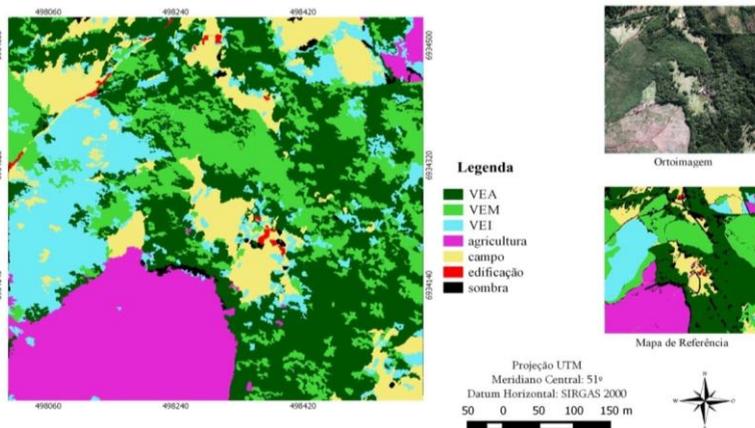
Os mapas temáticos resultantes das classificações no InterIMAGE, modelo gerado no WEKA e implementado no QGIS, e GeoDMA para a área-teste A são apresentados nas Figuras 36, 37 e 38 respectivamente.

Figura 36 - Mapa de uso e cobertura da terra da área-teste A gerado no *software* InterIMAGE com algoritmo C4.5.



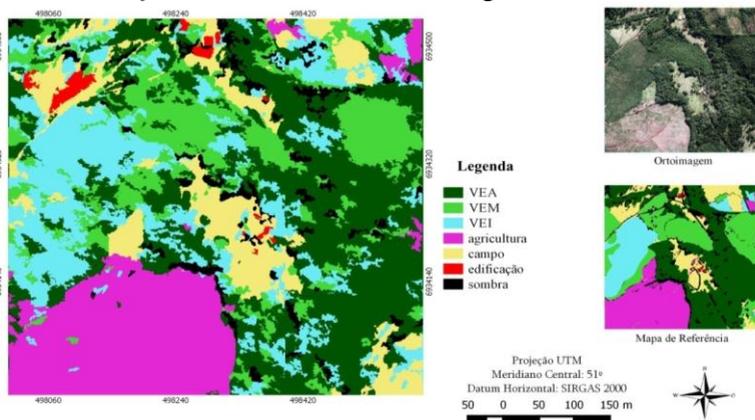
Fonte: produção do próprio autor, 2015.

Figura 37 - Mapa de uso e cobertura da terra da área-teste A gerado a partir do modelo do WEKA com o algoritmo SimpleCart no *software* QGIS.



Fonte: produção do próprio autor, 2015.

Figura 38 - Mapa de uso e cobertura da terra da área-teste A gerado no *software* GeoDMA com o algoritmo C4.5.



Fonte: produção do próprio autor, 2015.

As Tabelas 20, 21 e 22 apresentam as matrizes de confusão geradas a partir do cruzamento dos mapas classificados (linhas) com o mapa de referência (colunas) na área-teste A.

Tabela 20 - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo InterIMAGE.

| Mapa classificado | | Mapa de referência | | | | | | | Total | Eu | Kcu |
|-------------------|--------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|-------------|-------------|
| | | Classe | VEA | VEM | VEI | agricult. | campo | edificação | | | |
| Mapa classificado | VEA | 76,8 | 33,1 | 0,0 | 1,3 | 0,0 | 2,6 | 26,8 | 141 | 80,6 | 0,63 |
| | VEM | 23,2 | 66,9 | 0,0 | 0,2 | 0,0 | 0,0 | 10,4 | 101 | 61,7 | 0,48 |
| | VEI | 0,0 | 0,0 | 100 | 0,7 | 0,0 | 10,1 | 4,6 | 115 | 100 | 1 |
| | agricult. | 0,0 | 0,0 | 0,0 | 93,0 | 0,0 | 0,3 | 4,7 | 98 | 88,1 | 0,88 |
| | campo | 0,0 | 0,0 | 0,0 | 4,4 | 100 | 29,9 | 8,4 | 143 | 100 | 1 |
| | edificação | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 55,7 | 0,0 | 56 | 45,6 | 0,46 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,3 | 0,0 | 1,5 | 45,1 | 47 | 48,2 | 0,48 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | 76,8 | 66,9 | 100 | 93,0 | 100 | 55,7 | 45,1 | | | 77 | |
| Kcp | 0,58 | 0,53 | 1 | 0,93 | 1 | 0,56 | 0,45 | | | | 0,73 |

Fonte: produção do próprio autor, 2015.

Tabela 21 - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo WEKA.

| Mapa classificado | | Mapa de referência | | | | | | | Total | Eu | Kcu |
|-------------------|--------------|--------------------|-------------|------------|------------|-------------|-------------|-------------|-------|-------------|-------------|
| | | Classe | VEA | VEM | VEI | agricult. | campo | edificação | | | |
| Mapa classificado | VEA | 75,0 | 36,9 | 0,0 | 0,0 | 0,0 | 2,2 | 29,0 | 143,1 | 76 | 0,62 |
| | VEM | 25,0 | 46,7 | 0,0 | 0,0 | 0,0 | 0,0 | 9,0 | 80,7 | 54,6 | 0,4 |
| | VEI | 0,0 | 16,4 | 100 | 0,0 | 0,0 | 2,0 | 18,2 | 136,7 | 73,6 | 0,7 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 0,3 | 8,8 | 109,1 | 100 | 1 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 100 | 39,6 | 17,2 | 156,8 | 100 | 1 |
| | edificação | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 55,7 | 0,6 | 56,2 | 66,1 | 0,66 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,1 | 17,2 | 17,4 | 65,2 | 0,65 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | 75,0 | 46,7 | 100 | 100 | 100 | 55,7 | 17,2 | | | 70,7 | |
| Kcp | 0,61 | 0,33 | 1 | 1 | 1 | 0,56 | 0,17 | | | | 0,66 |

Fonte: produção do próprio autor, 2015.

Tabela 22 - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo GeoDMA.

| Mapa classificado | | Mapa de referência | | | | | | | Total | Eu | Kcu |
|-------------------|--------------|--------------------|-------------|------------|------------|-------------|-------------|------------|-------|-------------|-------------|
| | | Classe | VEA | VEM | VEI | agricult. | campo | edificação | | | |
| Mapa classificado | VEA | 65,4 | 36,1 | 0,0 | 0,0 | 0,0 | 1,3 | 32,9 | 136 | 74,9 | 0,53 |
| | VEM | 16,0 | 45,6 | 0,0 | 0,0 | 0,0 | 0,0 | 7,3 | 69 | 63,4 | 0,49 |
| | VEI | 18,7 | 18,2 | 100 | 0,0 | 0,0 | 9,3 | 10,2 | 156 | 46,8 | 0,39 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 100 | 100 | 1 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 100 | 25,0 | 8,4 | 133 | 100 | 1 |
| | edificação | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 63 | 0,2 | 63 | 17 | 0,17 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,3 | 41 | 42 | 54,8 | 0,55 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | 65,4 | 45,6 | 100 | 100 | 100 | 63,0 | 40,9 | | | 73 | |
| Kcp | 0,42 | 0,32 | 1 | 1 | 1 | 0,63 | 0,41 | | | | 0,69 |

Fonte: produção do próprio autor, 2015.

Os índices Kappa obtidos para os resultados das três metodologias por árvore de decisão aplicadas na área-teste A podem ser considerados como muito bom de acordo com Landis e Koch (1977). As maiores confusões ocorreram nas classes VEM, que foi principalmente confundida com o estágio avançado, apresentando erros de omissão e inclusão com esta classe, e entre as classes VEA e sombra. Nos resultados do GeoDMA e WEKA, o estágio médio teve menos de 50% de área classificada corretamente. A classe VEA, além da confusão com o estágio médio, no GeoDMA também foi erroneamente atribuída à VEI.

A metodologia seguida no InterIMAGE, além de ter obtido o maior Kappa dentre as três metodologias, também teve maior Kappa condicional do usuário nas classes de vegetação VEA e VEI, sendo que o GeoDMA teve melhor desempenho na classificação da VEM. As três metodologias tiveram excelentes acertos nas classes VEI, agricultura e campo, por serem classes espectralmente diferentes. No WEKA e GeoDMA, a classe VEI teve o estágio médio erroneamente atribuído a ela, o que reduziu sua exatidão do usuário.

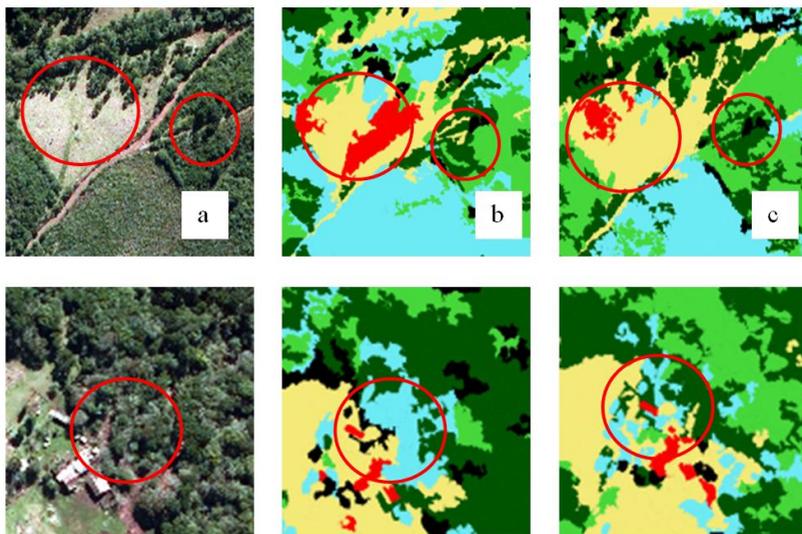
O modelo do WEKA obteve os piores resultados dentre as três metodologias de classificação por árvore de decisão na área-teste A. Para essa classificação, foi utilizada a árvore de decisão mais compacta que outros métodos, após procedimento de seleção de atributos, o que pode ter ocasionado perda de informação útil na discriminação das classes. Segundo Francisco e Almeida (2012), em métodos não-paramétricos, como a árvore de decisão, a redundância pode até ser desejável para a identificação de certas classes na cena, já que há especificidades em cada banda que são imprescindíveis para a discriminação de certos alvos.

Edificação e sombra são classes que tiveram menor índice de acerto. O erro da classe edificação pode advir do fato de a classe ser pouco representativa na imagem, além de ter

tido poucas amostras de treinamento. A classe sombra foi erroneamente atribuída à classe VEA, o que era esperado, visto que a sombra geralmente está associada à vegetação de maior porte. Como observado na Figura 37 e Tabela 21, o mapa gerado pelo WEKA teve o maior erro de omissão na classe sombra que os demais métodos, mostrando que a regra definida para esta classe pelo SimpleCart não foi suficiente para captar todas as áreas desta classe.

A Figura 39 ilustra as confusões comuns encontradas nas classificações da área-teste A. Na parte superior da Figura, o círculo esquerdo mostra uma mancha de campo de alto brilho classificada erroneamente como 'edificação'. Neste caso, a classificação poderia ser melhorada se algum atributo geométrico fosse utilizado na discriminação desta classe. No círculo mais à direita, notam-se áreas de VEM classificadas como estádio avançado, justamente por se tratar de árvores de maior porte, o que torna o comportamento espectral destas áreas descaracterizado. Na parte inferior desta Figura, é mostrado um exemplo em que objetos de vegetação em estádio avançado foram classificados como vegetação em estádio médio ou inicial por apresentarem alto brilho e, conseqüentemente, maiores valores da média dos ND das bandas do visível, característica dos estádios iniciais da vegetação. Além disso, trata-se de locais situados na borda do fragmento florestal, estando em contato com áreas antrópicas e com a classe campo.

Figura 39 - Principais confusões ocorridas na classificação da área-teste A, sendo: (a) imagem original; (b) classificação pelo GeoDMA; (c) classificação pelo InterIMAGE.

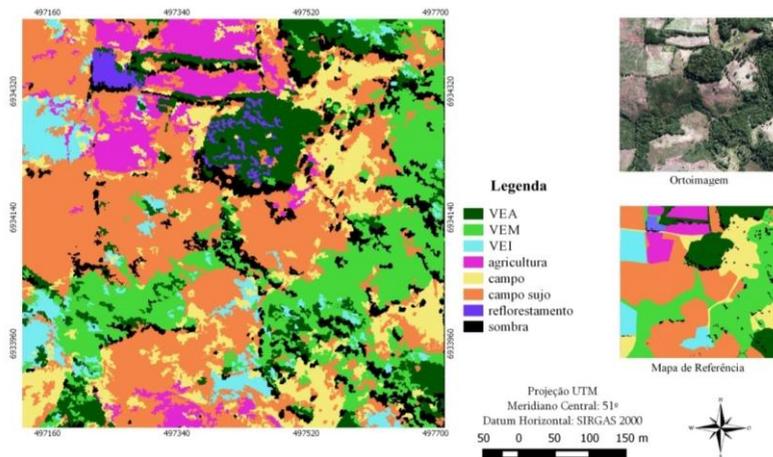


Fonte: produção do próprio autor, 2015.

Apesar de terem sido extraídos atributos geométricos no GeoDMA, estes não foram utilizados na discriminação da classe edificação, o que, teoricamente, poderia agregar acurácia ao resultado desta classe, por apresentar formas regulares, diferentemente das demais. Nesse caso, a interferência do analista na construção da árvore de decisão poderia ser benéfica.

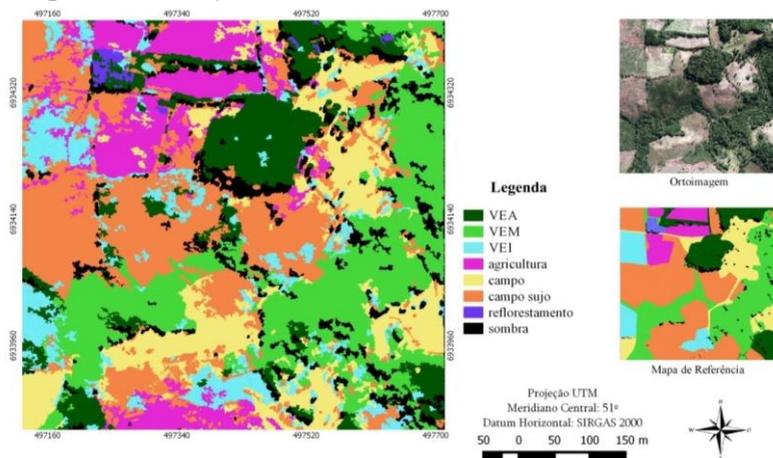
As Figuras 40, 41 e 42 ilustram os mapas gerados a partir da classificação da área-teste B. As Tabelas 23, 24, 25 mostram as matrizes de confusão, exatidão usuário e produtor e Kappa obtidos a partir do cruzamento entre o mapa de referência (coluna) com o mapa resultante da classificação (linha).

Figura 40 - Mapa de uso e cobertura da terra da área-teste B gerado no *software* InterIMAGE com o algoritmo C4.5.



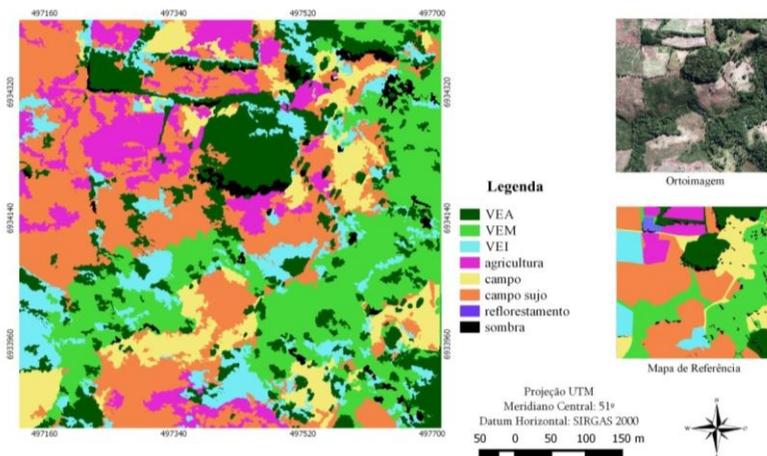
Fonte: produção do próprio autor, 2015.

Figura 41 - Mapa de uso e cobertura da terra da área-teste B gerado a partir do modelo do WEKA com o algoritmo SimpleCart no *software* QGIS.



Fonte: produção do próprio autor, 2015.

Figura 42 - Mapa de uso e cobertura da terra da área-teste B gerado no *software* GeoDMA com o algoritmo C4.5.



Fonte: produção do próprio autor, 2015.

Tabela 23 - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo InterIMAGE.

| Classe | | Mapa de referência | | | | | | | | Total | Eu | Kcu |
|-------------------|------------|--------------------|-------------|------------|-------------|-------------|-------------|-------------|--------|-------------|-------------|-------------|
| | | VEA | VEM | VEI | agricult. | campo | c. sujo | reflorest. | sombra | | | |
| Mapa classificado | VEA | 100 | 16,9 | 0,0 | 0,0 | 0,0 | 0,0 | 16,7 | 3,1 | 136,7 | 67,7 | 0,64 |
| | VEM | 0,0 | 83,1 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 3,2 | 86,2 | 100 | 1 |
| | VEI | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 100,0 | 100 | 1 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 0,7 | 0,0 | 100,7 | 100 | 1 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 65,1 | 15,3 | 2,3 | 0,1 | 82,8 | 58,5 | 0,52 |
| | c. sujo | 0,0 | 0,0 | 0,0 | 0,0 | 34,9 | 84,6 | 5,2 | 2,5 | 127,3 | 87,9 | 0,8 |
| | reflorest. | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 70,3 | 0,5 | 70,8 | 49,2 | 0,49 |
| sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 4,9 | 90,6 | 95,5 | 22,8 | 0,23 | |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | 100 | 83,1 | 100 | 100 | 65,1 | 84,6 | 70,3 | 90,6 | | | 86,7 | |
| Kcp | 1 | 0,78 | 1 | 1 | 0,59 | 0,75 | 0,7 | 0,91 | | | | 0,85 |

Fonte: produção do próprio autor, 2015.

Tabela 24 - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo WEKA.

| Classe | | Mapa de referência | | | | | | | | Total | Eu | Kcu |
|-------------------|--------------|--------------------|------------|------------|------------|------------|-------------|-------------|-------------|-------|-------------|-------------|
| | | VEA | VEM | VEI | agricult. | campo | c. sujo | reflorest. | sombra | | | |
| Mapa classificado | VEA | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 40,4 | 2,4 | 142,8 | 100 | 1 |
| | VEM | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 4,0 | 104,0 | 100 | 1 |
| | VEI | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 1,4 | 2,2 | 0,1 | 103,6 | 92,9 | 0,92 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 16,9 | 0,6 | 0,0 | 117,4 | 60,5 | 0,56 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 100 | 2,5 | 0,1 | 0,0 | 102,6 | 92,4 | 0,91 |
| | c. sujo | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 79,2 | 1,2 | 3,1 | 83,5 | 100 | 1 |
| | reflorest. | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 50,6 | 0,0 | 50,6 | 74,7 | 0,74 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 4,9 | 90,4 | 95,3 | 24,6 | 0,24 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | |
| | Ep | 100 | 100 | 100 | 100 | 100 | 79,2 | 50,6 | 90,4 | | 90 | |
| Kcp | 1 | 1 | 1 | 1 | 1 | 0,7 | 0,5 | 0,9 | | | 0,89 | |

Fonte: produção do próprio autor, 2015.

Tabela 25 - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo GeoDMA.

| Classe | | Mapa de referência | | | | | | | | Total | Eu | Kcu |
|-------------------|--------------|--------------------|-------------|------------|------------|-------------|-------------|-------------|-------------|-------|-------------|-------------|
| | | VEA | VEM | VEI | agricult. | campo | c. sujo | reflorest. | sombra | | | |
| Mapa classificado | VEA | 100 | 18,8 | 0,0 | 0,0 | 0,0 | 0,0 | 91,2 | 26,3 | 236,3 | 63,9 | 0,59 |
| | VEM | 0,0 | 81,2 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 12,5 | 93,7 | 100 | 0,52 |
| | VEI | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 6,1 | 0,7 | 106,9 | 100 | 1 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 16,6 | 0,0 | 0,0 | 116,6 | 46,8 | 0,43 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 100 | 15,2 | 0,0 | 0,0 | 115,2 | 56,1 | 0,52 |
| | c. sujo | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 68,1 | 1,7 | 3,2 | 73,0 | 100 | 1 |
| | reflorest. | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0 | 0 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 57,2 | 58,2 | 53,7 | 0,54 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | |
| | Ep | 100 | 81,2 | 100 | 100 | 100 | 68,1 | 0,0 | 57,2 | | 76 | |
| Kcp | 1 | 0,75 | 1 | 1 | 1 | 0,56 | 0 | 0,57 | | | 0,72 | |

Fonte: produção do próprio autor, 2015.

Na área-teste B os resultados obtidos pelo InterIMAGE e WEKA, com Kappa de 0,85 e 0,89, respectivamente, podem ser considerados excelentes, e 0,72 para o GeoDMA considerado como muito bom, de acordo com Landis e Koch (1977).

Nesta área, a confusão entre as tipologias vegetais foi menor que na área-teste A, alcançando excelente Kappa condicional do produtor e usuário no estágio inicial, e excelente Kappa condicional do produtor na classe VEA. Da mesma forma que a área-teste A, a maior confusão ocorreu na VEM, classificada como VEA nos resultados do InterIMAGE e GeoDMA. Outra confusão recorrente nessa área-teste foi na

classe reflorestamento, classificada predominantemente como VEA. No GeoDMA, o reflorestamento não teve nenhuma área classificada, tendo 100% de erro de omissão, o que causou o menor índice Kappa para essa metodologia. A classe VEA, nos resultados obtidos no InterIMAGE e GeoDMA, apresentou erros de inclusão, devido à classificação errônea da classe reflorestamento e sombra, classes estas que tiveram os maiores erros de omissão. Isto ocorreu pelo fato de a classe reflorestamento, além de poucas amostras de treinamento, ter um padrão espectral escuro parecido com a classe VEA. A confusão da classe sombra ocorreu por ser uma classe estreitamente associada à vegetação de maior porte.

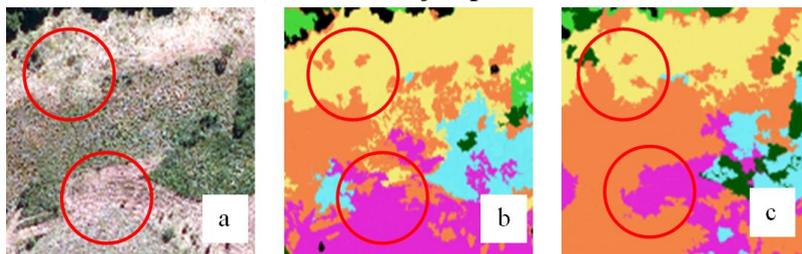
A classe campo sujo teve bom Kappa condicional do usuário, porém, altos erros de omissão, erroneamente atribuída à classe campo nessas três metodologias, e à classe agricultura nos resultados do WEKA e GeoDMA. A mesma apresenta uma variedade espectral intra-classe, já que possui áreas cobertas com uma vegetação esparsa, e outras um pouco mais “limpas”, o que pode ter ocasionado a maior confusão.

Interessante que o WEKA teve apenas uma regra de decisão para as classes de vegetação em estádios inicial, médio e avançado, e da mesma forma, conseguiu excelentes índices de acerto, superando o InterIMAGE e o GeoDMA nesta área-teste. Outro fato observado foi que a classificação derivada da árvore de decisão gerada no WEKA alcançou índice de acerto maior ao ser cruzada com todo o mapa de referência, do que em função das amostras (Tabela 17). Uma possível explicação para isso seria a de que, pelo fato de os *softwares* InterIMAGE e GeoDMA terem gerado árvores de decisão mais complexas, houve o chamado “*overfitting*”, ou seja, um superajuste do modelo às amostras de treinamento, o que impediu de classificar corretamente as classes que não se enquadravam em todos os requisitos elaborados na árvore de decisão. Segundo Körting (2012), uma árvore muito grande pode superajustar os dados, enquanto uma muito pequena pode deixar de capturar

estruturas importantes. Por isso, o autor diz ser preferível uma árvore média, desde que não subestime ou superestime os dados, e também seja facilmente interpretada pelo usuário.

A Figura 43 ilustra os erros recorrentes da classificação da área-teste B, principalmente na classe campo sujo. Percebe-se que as áreas de campo sujo erroneamente classificadas como campo e agricultura (círculo superior e inferior, respectivamente), apresentam semelhanças espectrais com essas classes. Esse erro também pode estar associado ao mapa de referência dessa classe, já que a área classificada como campo sujo no mapa de referência, em certos locais, pode tratar-se de áreas agrícolas abandonadas, assemelhando-se à classe agricultura. Quanto à confusão do campo sujo com a classe campo, percebe-se que ocorreram em regiões em que a classe campo sujo possui maior brilho.

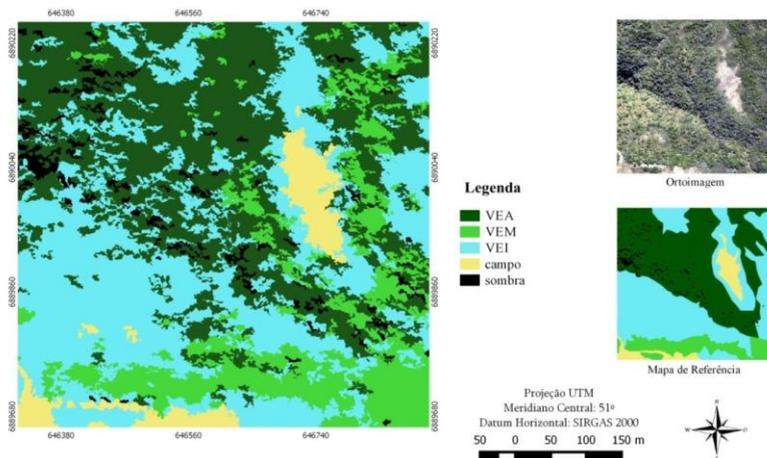
Figura 43 - Principais confusões das classificações da área-teste B, sendo: (a) imagem original; (b) classificação pelo modelo do WEKA; (c) classificação pelo GeoDMA.



Fonte: produção do próprio autor, 2015.

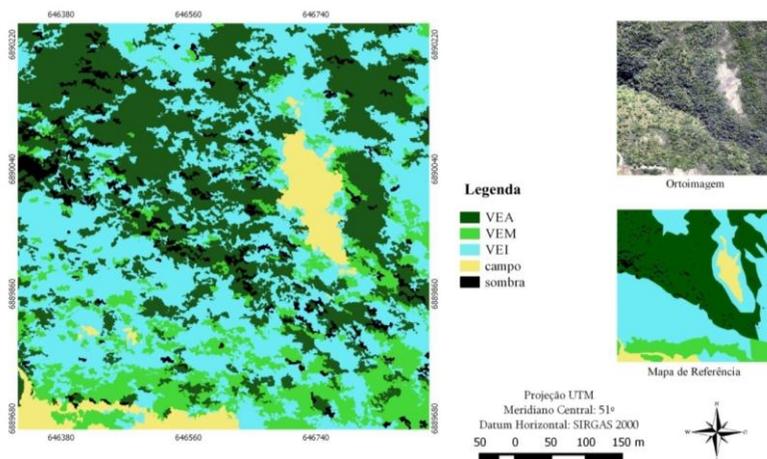
As Figuras 44, 45, 46 e 47 ilustram os resultados da área-teste C obtidas pelas classificações realizadas no InterIMAGE, InterIMAGE sem o componente MDS, modelo do WEKA no QGIS, e GeoDMA, respectivamente.

Figura 44 - Mapa de uso e cobertura da terra da área-teste C gerado no *software* InterIMAGE com o algoritmo C4.5.



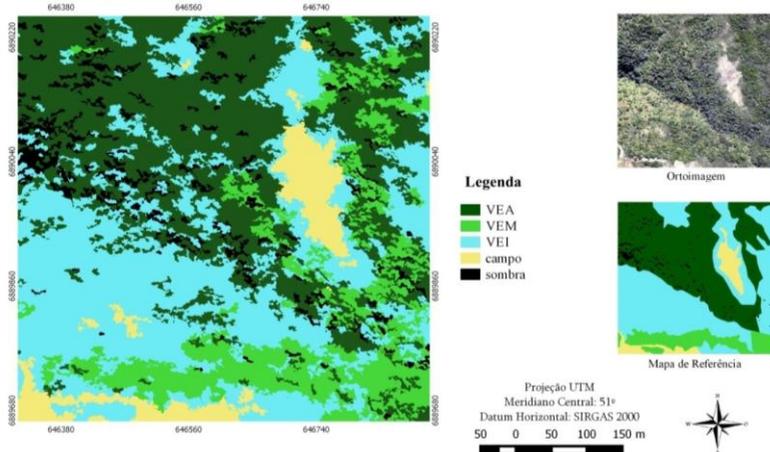
Fonte: produção do próprio autor, 2015.

Figura 45 - Mapa de uso e cobertura da terra da área-teste C gerado no *software* InterIMAGE, sem o MDS com o algoritmo C4.5.



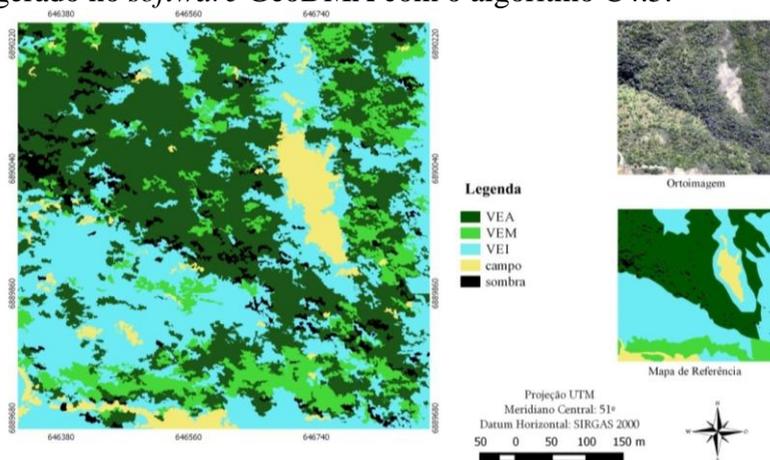
Fonte: produção do próprio autor, 2015.

Figura 46 - Mapa de uso e cobertura da terra da área-teste C gerado a partir do modelo do WEKA com o algoritmo SimpleCart no *software* QGIS.



Fonte: produção do próprio autor, 2015.

Figura 47 - Mapa de uso e cobertura da terra da área-teste C gerado no *software* GeoDMA com o algoritmo C4.5.



Fonte: produção do próprio autor, 2015.

As Tabelas 26, 27, 28 e 29 mostram as matrizes de confusão no cruzamento entre o mapa de referência (colunas) e

mapa classificado (linhas) para a área-teste C, juntamente com o Kappa e exatidão dessas metodologias.

Tabela 26 - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo InterIMAGE.

| | | Mapa de referência | | | | | | Total | Eu | Kcu |
|--------------------|--------|--------------------|------------|-------------|------------|-------------|--------|-------------|-------------|-----|
| | | Classe | VEA | VEM | VEI | campo | sombra | | | |
| Mapa classifica | VEA | 70,1 | 0,0 | 2,7 | 0,0 | 8,8 | 81,7 | 96,8 | 0,94 | |
| | VEM | 11,8 | 100 | 15,5 | 0,0 | 1,4 | 128,7 | 39,4 | 0,34 | |
| | VEI | 18,0 | 0,0 | 81,8 | 0,0 | 15,2 | 115 | 79,2 | 0,65 | |
| | campo | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 100 | 100 | 1 | |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 74,6 | 74,6 | 21,5 | 0,21 | |
| | Total | 100 | 100 | 100 | 100 | 100 | | | | |
| Ep | | 70,1 | 100 | 81,8 | 100 | 74,6 | | 85 | | |
| Kcp | | 0,54 | 1 | 0,69 | 1 | 0,75 | | | 0,82 | |

Fonte: produção do próprio autor, 2015.

Tabela 27 - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo InterIMAGE, sem a banda MDS.

| | | Mapa de referência | | | | | | Total | Eu | Kcu |
|--------------------|--------|--------------------|--------------|-------------|------------|-------------|--------|-------------|-------------|-----|
| | | Classe | VEA | VEM | VEI | campo | sombra | | | |
| Mapa classifica | VEA | 68,8 | 0,0 | 44,3 | 0,0 | 10,7 | 123,8 | 76,8 | 0,45 | |
| | VEM | 0,0 | 100 | 40,8 | 0,0 | 5,4 | 146,2 | 44,1 | 0,39 | |
| | VEI | 31,2 | 0,0 | 15,0 | 0,0 | 10,5 | 56,7 | 18,4 | 0,12 | |
| | campo | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 100 | 100 | 1 | |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 73,4 | 73,4 | 25,3 | 0,25 | |
| | Total | 100 | 100 | 100 | 100 | 100 | | | | |
| Ep | | 68,8 | 100,0 | 15,0 | 100 | 73,4 | | 71 | | |
| Kcp | | 0,36 | 1 | 0,08 | 1 | 0,73 | | | 0,64 | |

Fonte: produção do próprio autor, 2015.

Tabela 28 - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo modelo do WEKA.

| | | Mapa de referência | | | | | | Total | Eu | Kcu |
|--------------------|--------|--------------------|------------|-------------|------------|-------------|--------|-------------|-------------|-----|
| | | Classe | VEA | VEM | VEI | campo | sombra | | | |
| Mapa classifica | VEA | 66 | 0,0 | 21,2 | 0,0 | 9,8 | 97,0 | 76,1 | 0,57 | |
| | VEM | 10,2 | 100 | 13,0 | 0,0 | 0,0 | 123,1 | 38 | 0,34 | |
| | VEI | 14,8 | 0,0 | 65,8 | 0,0 | 3,0 | 83,7 | 81,3 | 0,67 | |
| | campo | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 100 | 100 | 1 | |
| | sombra | 9,0 | 0,0 | 0,0 | 0,0 | 87,2 | 96,2 | 0 | 0 | |
| | Total | 100 | 100 | 100 | 100 | 100 | | | | |
| Ep | | 66 | 100 | 65,8 | 100 | 87,2 | | 84 | | |
| Kcp | | 0,44 | 1 | 0,47 | 1 | 0,87 | | | 0,8 | |

Fonte: produção do próprio autor, 2015.

Tabela 29 - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo GeoDMA.

| | | Mapa de referência | | | | | Total | Eu | Kcu |
|-------------------|--------|--------------------|------------|-------------|------------|-------------|-------|-------------|-------------|
| | | VEA | VEM | VEI | campo | sombra | | | |
| Mapa classificado | Classe | | | | | | | | |
| | VEA | 73,6 | 0,0 | 26,2 | 0,0 | 19,8 | 119,5 | 73,1 | 0,51 |
| | VEM | 13,1 | 100 | 11,4 | 0,0 | 3,6 | 128,1 | 28 | 0,25 |
| | VEI | 13,4 | 0,0 | 62,4 | 0,0 | 1,0 | 76,8 | 82,8 | 0,68 |
| | campo | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 100 | 100 | 1 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 75,6 | 75,6 | 23,6 | 0,24 |
| | Total | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | | 73,6 | 100 | 62,4 | 100 | 75,6 | | 82 | |
| Kcp | | 0,52 | 1 | 0,42 | 1 | 0,76 | | | 0,78 |

Fonte: produção do próprio autor, 2015.

No InterIMAGE e WEKA, obteve-se índice Kappa de 0,82 e 0,8, respectivamente, considerado como excelente. Na classificação feita no InterIMAGE sem o MDS, e no GeoDMA, os índices Kappa de 0,64 e 0,78, respectivamente, podem ser considerados como muito bom, de acordo com Landis e Koch (1977). Diferentemente das outras áreas-teste, que tiveram maiores confusões no estágio médio, na área-teste C a confusão predominou na classe VEI. A explicação seria de que este estágio tem uma variabilidade espectral maior nesta área-teste, pois possui alguns locais em que predomina uma vegetação arbustiva, e em outros, a predominância é de gramíneas. Na classificação sem a componente MDS, a VEI teve 85% de erro de omissão e quase 80% de erro de inclusão. Esse resultado pode ser observado na Figura 45, em que a VEI foi classificada como avançado e médio, e os estádios avançado e médio foram classificados como inicial. Dessa forma, percebe-se que a integração do MDS melhorou a acurácia da classificação das tipologias vegetais, já que, por possuir informação de variabilidade textural dos objetos, foi fundamental para discriminar o estágio avançado das demais classes.

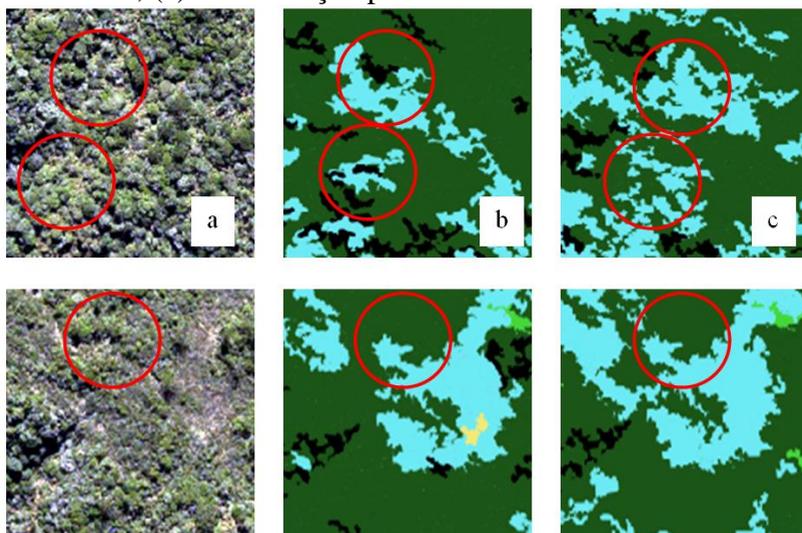
A VEA, por também apresentar maior variabilidade espectral nesta área-teste, foi confundida com os outros dois estádios. A VEM teve maiores erros de comissão, ou seja, as classes estágio inicial e avançado foram erroneamente atribuídas ao médio. Esta classe, por ser menos representativa

na área-teste C, foi a que teve menor número de amostras de treinamento, o que pode ter ocasionado estes erros.

Percebe-se que o InterIMAGE teve o melhor desempenho na classificação do estágio avançado, enquanto nos estádios médios e iniciais as metodologias tiveram desempenhos similares, com exceção da classificação sem a componente MDS, que teve resultados inferiores na classificação do estágio inicial.

A Figura 48 ilustra os principais erros ocorridos nesta área-teste na discriminação dos estádios sucessionais da vegetação. Na parte superior da figura, percebem-se áreas de VEA que possuem maior refletância e foram classificadas erroneamente como VEI. Na parte inferior, são áreas de VEI que, por possuírem padrão espectral mais escuro e maior heterogeneidade, foram classificadas como VEA.

Figura 48 - Principais confusões ocorridas na classificação da área-teste C; (a) imagem original; (b) classificação com modelo do WEKA; (c) classificação pelo InterIMAGE.



Fonte: produção do próprio autor, 2015.

4.3.1.1 Análise das classificações por árvore de decisão

As técnicas que utilizaram a mineração de dados tiveram resultados considerados muito bom e excelente nas áreas testadas. O Kappa variou de 0,64 na classificação realizada sem o MDS na área-teste C, a 0,89, melhor resultado alcançado na classificação com árvore de decisão gerada no WEKA para a área-teste B. Quanto à classificação das tipologias vegetais, os resultados variaram de acordo com a metodologia e com as particularidades de cada área-teste.

Os resultados dessas metodologias ainda podem ser melhorados aplicando-se uma heurística para reduzir erros de comissão resultantes da classificação automática, editando os limiares da árvore, por exemplo. A intervenção do analista poderia ser aplicada principalmente nas classes edificação na área-teste A, e reflorestamento na área-teste B, classes que, por serem menos representativas, tiveram maiores erros de comissão e omissão, respectivamente. Na metodologia seguida no GeoDMA, por exemplo, ocorreram maiores erros na classificação dessas duas classes. Este fato pode ser atribuído à etapa da segmentação neste *software*, que, por dividir a imagem em blocos para então segmentá-la (Figura 26), resulta em fronteiras artificiais que podem ter impactos negativos na classificação, principalmente de alvos menores.

A classificação por árvore de decisão possui a vantagem de que, além de permitir que os limiares sejam alterados, de forma a reduzir os erros encontrados na classificação, também permite que os modelos obtidos sejam ajustados as peculiaridades e replicados em outras áreas com características parecidas de uso e cobertura da terra (CAMARGO et al., 2009). Cintra et al. (2010) mencionam que as árvores de decisão podem ser consideradas “caixas-brancas”, já que permitem a visualização e manipulação das regras de decisão que definem cada classe.

Vários trabalhos obtiveram bons resultados ao utilizar a árvore de decisão na classificação orientada a objeto, tanto quando elaborada de forma heurística através de uma rede semântica, quanto pela mineração de dados. Pinho et al. (2008) e Ribeiro e Fonseca (2013) compararam um experimento em que a árvore de decisão foi elaborada a partir dos limiares definidos pelo usuário, com um baseado na árvore de decisão gerada pelo minerador de dados C4.5. Os resultados encontrados foram similares, com a vantagem que com a mineração de dados, economiza-se tempo. Tedesco et al. (2014) também fizeram esta comparação, porém utilizando o algoritmo de árvore de decisão CART como minerador de dados, para detectar processos erosivos de voçorocas em imagens IKONOS. Esses autores tampouco encontraram diferenças significativas nos dois métodos. Carvalho (2011) testou três metodologias para classificação da cobertura da terra por árvore de decisão, sendo um por modelo gerado via análise exploratória convencional fazendo uso de pertinência *fuzzy*; o segundo com mineração de dados por árvore de decisão, tendo a intervenção do intérprete na especificação do número mínimo de instâncias por folha; e o terceiro por mineração de dados por árvore de decisão sem a intervenção direta do intérprete. O autor teve os melhores resultados para os dois últimos métodos.

Muitos trabalhos também testaram o potencial dos *softwares open source* nesta tarefa. Souza (2012), ao classificar áreas urbanas costeiras no Maranhão com imagens WorldView-2, fez um comparativo entre metodologias que seguiam a análise exploratória dos atributos no InterIMAGE, em que os limiares eram definidos pelo analista, e aliando técnicas de mineração de dados do WEKA, obtendo maior Kappa neste último. Souza e Kux (2013) usaram as ferramentas de mineração de dados do GeoDMA para construir um modelo de conhecimento no InterIMAGE objetivando a classificação de imagens WorldView-2, em que obtiveram Kappa de 0,93.

Kux e Meneghetti (2015) utilizaram o *software* InterIMAGE, imagens de alta resolução espacial do satélite WorldView-2 associadas a técnicas de mineração de dados no GeoDMA nos diversos níveis da rede semântica elaborada no InterIMAGE, visando o mapeamento temático de cobertura da terra, obtendo valores de índices Kappa e exatidão global de 0,92 e 0,94, respectivamente. Körting (2012) fez um comparativo entre a classificação orientada a objeto no InterIMAGE e no GeoDMA pelo algoritmo C4.5 e não obteve diferença entre estes aplicativos. No entanto, o autor utilizou os mesmos atributos em ambos para geração da árvore de decisão. Passo et al. (2014) compararam o desempenho do InterIMAGE na classificação gerada a partir das regras definidas pela análise visual e conhecimento do analista com a classificação resultante do algoritmo C4.5 implementado neste programa, e tiveram melhores resultados nesta última.

Importante destacar que neste estudo houve diferenças na forma pelo que foram conduzidas cada uma das metodologias. No InterIMAGE, a classificação foi com o mesmo algoritmo do GeoDMA, o C4.5, porém os atributos gerados e utilizados na classificação em cada *software* diferiram. No WEKA, no entanto, foram utilizados os mesmos atributos que o InterIMAGE, porém uma árvore de decisão mais simples e compacta gerado pelo SimpleCart foi escolhida pelo analista após o procedimento de seleção de atributos. O resultado desta última metodologia foi excelente na área-teste B, demonstrando que, em certas condições, uma árvore mais simples e generalizável pode ter melhores resultados que as mais complexas.

4.3.2 Resultado da classificação por máquinas de vetor de suporte

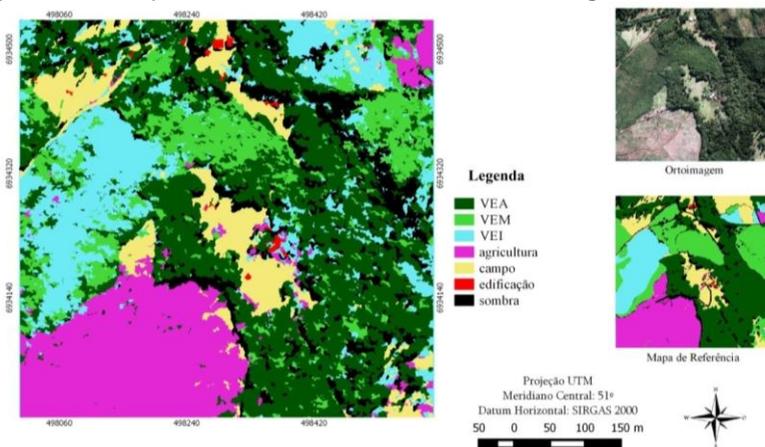
A classificação pelo algoritmo SVM realizada no OTB Monteverdi resultou em quatro mapas, três para classificação

por regiões e um para a classificação por *pixel*, esta última realizada apenas na área-teste B. Os resultados e discussão dessas metodologias são apresentados nas seções seguintes.

4.3.2.1 Classificação por regiões

As Figuras 49, 51 e 52 ilustram o resultado da classificação por regiões realizada no *software* OTB Monteverdi com o algoritmo SVM para as áreas-teste A, B e C, respectivamente. As matrizes de confusão resultantes do cruzamento com o mapa de referência são mostradas nas Tabelas 30, 31 e 32.

Figura 49 - Mapa de uso e cobertura da terra da área-teste A gerado no *software* OTB Monteverdi com o algoritmo SVM.



Fonte: produção do próprio autor, 2015.

Tabela 30 - Matriz de confusão (em porcentagem) da classificação da área-teste A pelo OTB Monteverdi.

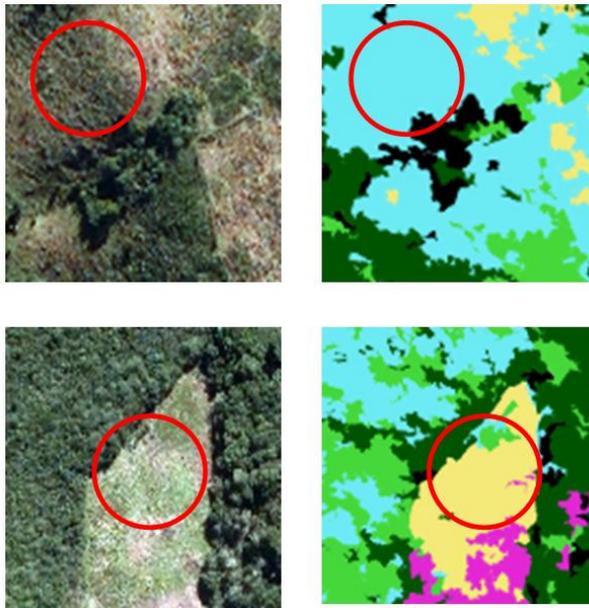
| Classe | | Mapa de referência | | | | | | | Total | Eu | Kcu |
|----------------------|--------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|-------------|-------------|
| | | VEA | VEM | VEI | agricult. | campo | edificação | sombra | | | |
| Mapa classificado | VEA | 83,9 | 12,6 | 1,1 | 0,5 | 0,1 | 0,7 | 26,9 | 125,7 | 90 | 0,84 |
| | VEM | 9,8 | 78,1 | 0,8 | 0,7 | 0,2 | 0,7 | 10,5 | 100,9 | 82,4 | 0,77 |
| | VEI | 6,1 | 9,0 | 97,8 | 1,9 | 55,5 | 16,6 | 4,7 | 191,6 | 54,4 | 0,48 |
| | agricult. | 0,1 | 0,2 | 0,2 | 56,8 | 0,6 | 10,6 | 4,3 | 72,7 | 97,8 | 0,97 |
| | campo | 0,0 | 0,0 | 0,1 | 39,9 | 43,3 | 4,2 | 8,5 | 96,1 | 39,9 | 0,33 |
| | edificação | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 60,1 | 0,0 | 60,1 | 91,9 | 0,92 |
| | sombra | 0,1 | 0,0 | 0,0 | 0,1 | 0,3 | 7,1 | 45,2 | 52,9 | 75 | 0,75 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| Ep | 83,9 | 78,1 | 97,8 | 56,8 | 43,3 | 60,1 | 45,2 | | | 75 | |
| Kcp | 0,75 | 0,72 | 0,97 | 0,52 | 0,36 | 0,6 | 0,45 | | | | 0,61 |

Fonte: produção do próprio autor, 2015.

A classificação por regiões pelo algoritmo SVM na área teste A resultou em um Kappa de 0,61, considerado como muito bom de acordo com Landis e Koch (1977). Este método obteve resultados excelentes na discriminação do estádio médio e avançado da vegetação, com exatidão do usuário de 82% e 90%, respectivamente. As maiores confusões ocorreram entre as classes campo, VEI e agricultura. A classe VEI teve maior erro de comissão, tendo a classe campo atribuída a este estádio. A classe campo teve erros tanto de omissão, em que foi classificada como VEI, quanto de comissão, em que teve 40% das áreas de agricultura foram classificadas como campo.

Na parte superior da Figura 50, o círculo vermelho indica local de campo classificado como VEI. Percebe-se, no entanto, que nesse local o campo apresenta certa rugosidade, o que pode ter ocasionado a confusão. Na parte inferior, tem-se área de agricultura classificada como campo.

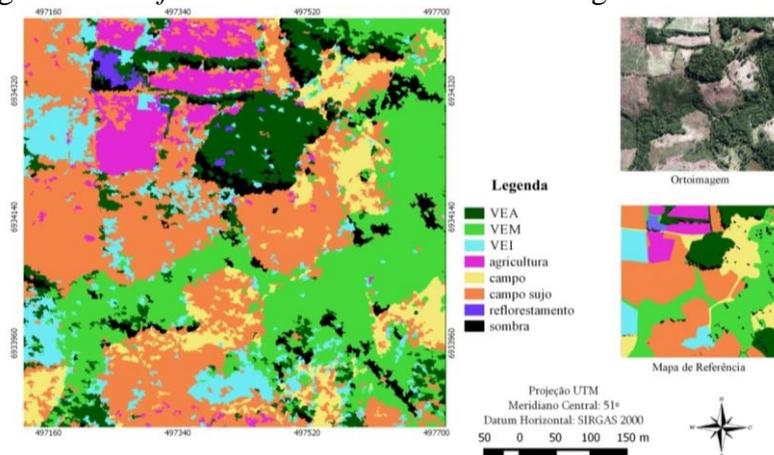
Figura 50 - Principais confusões da classificação por regiões da área-teste A no OTB Monteverdi.



Fonte: produção do próprio autor, 2015.

A Figura 51 e Tabela 31 mostram os resultados da classificação da área-teste B pelo algoritmo SVM no OTB Monteverdi.

Figura 51 - Mapa de uso e cobertura da terra da área-teste B gerado no *software* OTB Monteverdi com o algoritmo SVM.



Fonte: produção do próprio autor, 2015.

Tabela 31 - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo OTB Monteverdi.

| | Mapa de referência | | | | | | | | Total | Eu | Kcu |
|-------------------|--------------------|-----|-----|-----------|-------|---------|------------|--------|-------|------|------|
| | VEA | VEM | VEI | agricult. | campo | c. sujo | reflorest. | sombra | | | |
| Mapa classificado | VEA | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 29,3 | 2,2 | 131,5 | 100 | 1 |
| | VEM | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 11,7 | 111,7 | 100 | 1 |
| | VEI | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 3,5 | 0,0 | 114,3 | 58,7 | 0,56 |
| | agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 0,1 | 0,0 | 100,1 | 100 | 1 |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 94,5 | 0,0 | 0,0 | 94,5 | 100 | 1 |
| | c. sujo | 0,0 | 0,0 | 0,0 | 0,0 | 5,5 | 89,2 | 3,0 | 102,2 | 98,8 | 0,98 |
| | reflorest. | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 48,0 | 48,0 | 72,0 | 0,72 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 16,1 | 81,5 | 97,6 | 27,1 | 0,27 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| | Ep | 100 | 100 | 100 | 100 | 94,5 | 89,2 | 48,0 | 81,5 | 95 | |
| | Kcp | 1 | 1 | 1 | 1 | 0,94 | 0,83 | 0,48 | 0,81 | | 0,88 |

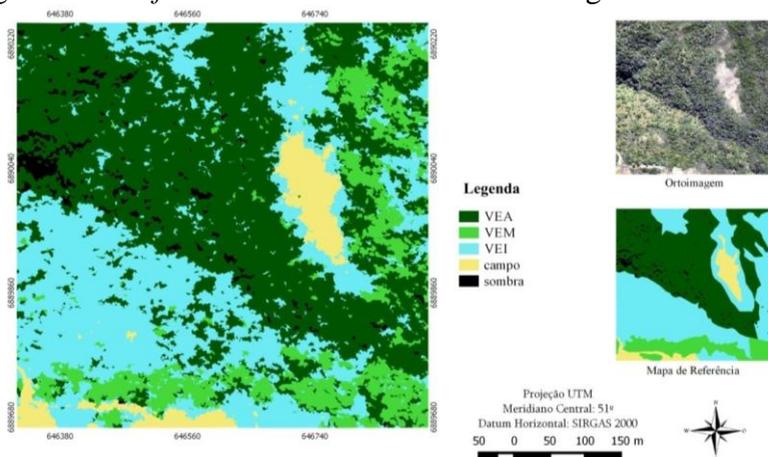
Fonte: produção do próprio autor, 2015.

O resultado da classificação desta área-teste pelo SVM teve Kappa de 0,88, considerado excelente de acordo com Landis e Koch (1977). Percebe-se que os três estádios sucessionais da vegetação tiveram 100% de exatidão do produtor, considerando arredondamento dos dados. As maiores confusões ocorreram na classe reflorestamento, que foi classificada como VEA e sombra, e entre as classes campo e campo sujo. A classe campo sujo teve áreas classificadas como

VEI, o que ocasionou erros de inclusão para esta tipologia vegetal.

A Figura 52 e a Tabela 32 mostram os resultados obtidos na classificação pelo algoritmo SVM no OTB Monteverdi na área-teste C.

Figura 52 - Mapa de uso e cobertura da terra da área-teste C gerado no *software* OTB Monteverdi com o algoritmo SVM.



Fonte: produção do próprio autor, 2015.

Tabela 32 - Matriz de confusão (em porcentagem) da classificação da área-teste C pelo OTB Monteverdi.

| | | Mapa de referência | | | | Total | Eu | Kcu | |
|--------------------------|--------------|--------------------|------------|-------------|------------|-------------|-----------|-------------|-------------|
| | | VEA | VEM | VEI | campo | | | | sombra |
| Mapa classifica do | Classe | VEA | VEM | VEI | campo | sombra | Total | Eu | Kcu |
| | VEA | 89,9 | 0,0 | 29,7 | 0,0 | 8,8 | 128,4 | 72,8 | 0,53 |
| | VEM | 9,1 | 100 | 9,4 | 0,0 | 0,3 | 118,7 | 41,5 | 0,38 |
| | VEI | 0,9 | 0,0 | 60,9 | 0,0 | 81 | 142,8 | 98,7 | 0,97 |
| | campo | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 100 | 100 | 1 |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 10 | 10 | 15 | 0,1 |
| | Total | 100 | 100 | 100 | 100 | 100 | | | |
| | Ep | 89,9 | 100 | 60,9 | 100 | 18 | 77 | | |
| | Kcp | 0,79 | 1 | 0,45 | 1 | 0,17 | | 0,64 | |

Fonte: produção do próprio autor, 2015.

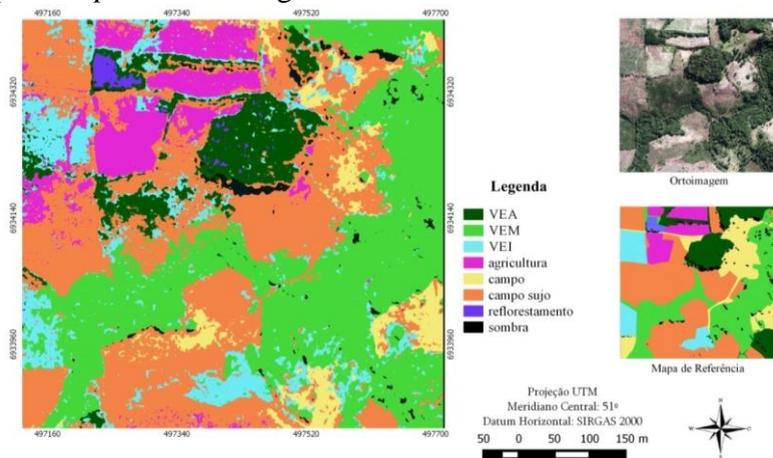
Para a área-teste C, a classificação pelo SVM teve Kappa de 0,64, considerado como muito bom pela literatura. Assim como as metodologias que se basearam em árvore de decisão,

os maiores erros nesta área-teste ocorreram na VEI, que teve aproximadamente 30% de sua área classificada como estádio avançado e 9% como médio. A classe sombra teve 80% de sua área classificada como VEI. Este erro pode advir do fato que o MDS foi obtido por correlação, apresentando problemas em áreas de sombra. Porém, como o processo de classificação pelo SVM não é apresentado de forma explícita, como ocorre com as árvores de decisão, não se vê diretamente atributos e limiares que ele utiliza para discriminar as classes, dificultando sua interpretação.

4.3.2.2 Classificação *pixel-a-pixel*

A área-teste B foi classificada através de metodologia tradicional por *pixel* com o algoritmo SVM no *software* OTB Monteverdi. Os resultados são apresentados na Figura 53, e a Tabela 33 que mostra a matriz de confusão, Kappa e exatidão, resultantes do cruzamento com o mapa de referência.

Figura 53 - Mapa de uso e cobertura da terra da área-teste B gerado no *software* OTB Monteverdi através de metodologia *pixel-a-pixel* com o algoritmo SVM.



Fonte: produção do próprio autor, 2015.

Tabela 33 - Matriz de confusão (em porcentagem) da classificação da área-teste B pelo OTB Monteverdi (*pixel-a-pixel*).

| Classe | | Mapa de referência | | | | | | | | Total | Eu | Kcu | |
|------------------------|--------------|--------------------|------------|------------|------------|-------------|-------------|-------------|-------------|-------|-------------|-------------|--|
| | | VEA | VEM | VEI | Agricult. | campo | c. sujo | Reflorest. | sombra | | | | |
| Mapa de confusão | VEA | 52,9 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 30,0 | 4,6 | 87,5 | 100 | 1 | |
| | VEM | 47,1 | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 23,3 | 170,4 | 83,8 | 0,72 | |
| | VEI | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 62,6 | 6,9 | 0,1 | 169,6 | 51,4 | 0,47 | |
| | Agricult. | 0,0 | 0,0 | 0,0 | 100 | 0,0 | 0,0 | 0,0 | 0,0 | 100,0 | 100 | 1 | |
| | campo | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 50 | 0,44 | |
| | c. sujo | 0,0 | 0,0 | 0,0 | 0,0 | 100 | 37,4 | 2,8 | 0,1 | 140,3 | 28 | 0,19 | |
| | Reflorest. | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 60,3 | 0,0 | 60,3 | 59,9 | 0,6 | |
| | sombra | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 71,9 | 71,9 | 63,4 | 0,63 | |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | |
| | Ep | 52,9 | 100 | 100 | 100 | 0,0 | 37,4 | 60,3 | 71,9 | | | 74 | |
| Kcp | 0,48 | 1 | 1 | 1 | 0 | 0,26 | 0,6 | 0,72 | | | | 0,6 | |

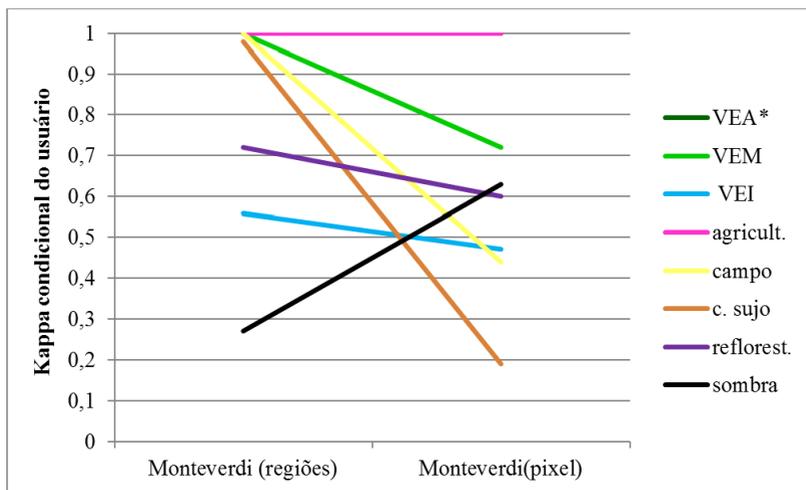
Fonte: produção do próprio autor, 2015.

O Kappa desta classificação foi 0,6, considerado muito bom de acordo com Landis e Koch (1977). A VEA teve 47% de sua área atribuída ao estágio médio. As classes VEM e VEI tiveram excelentes índices de acerto. Os maiores erros dessa metodologia foram na classe reflorestamento, que foi classificada como VEA, e na classe campo, que foi praticamente 100% classificada como campo sujo.

4.3.2.3 Análise das classificações pelo algoritmo SVM

O gráfico da Figura 54 ilustra o Kappa condicional do usuário obtido para a área-teste B na classificação SVM por regiões e por *pixel*.

Figura 54 - Gráfico com o Kappa condicional do usuário da classificação pelo algoritmo SVM por objeto e por *pixel* realizada no OTB Monteverdi na área-teste B.



*A barra correspondente à VEA está abaixo da barra agricultura ($K_{cu}=1$).
 Fonte: produção do próprio autor, 2015.

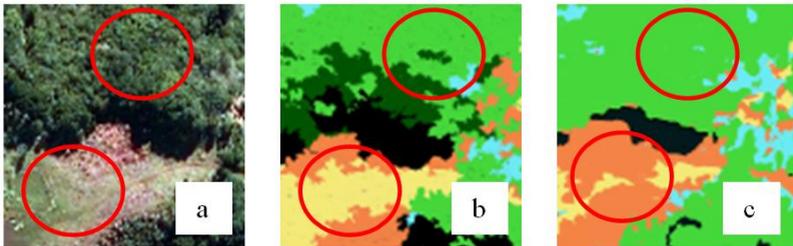
Percebe-se que ambas as metodologias tiveram excelente Kappa condicional do usuário para a classe VEA e agricultura. Com exceção da classe sombra, as demais na classificação por regiões mostraram-se superiores à classificação por *pixel*. A classe sombra teve menor exatidão do usuário na metodologia por regiões porque teve a classe reflorestamento atribuída a ela. A classe reflorestamento teve melhor acerto geral na classificação por *pixel* se comparada à metodologia por regiões (Tabela 26 e 28). Pelo fato de as classes sombra e reflorestamento serem menores, esses erros podem ser atribuídos à etapa da segmentação, que talvez não tenha delimitado os segmentos dessas classes adequadamente.

A metodologia por *pixel* teve desempenho geral péssimo na classificação da classe campo, que foi quase 100% atribuída à classe campo sujo, resultando em um Kappa condicional do

usuário baixo para esta classe. A classe campo, portanto, não foi distinguida bem em nível de *pixel*.

A Figura 55 ilustra erros recorrentes ocorridos na classificação SVM por regiões e por *pixel* na área-teste B.

Figura 55 - Principais erros ocorridos na classificação da área-teste B pelo algoritmo SVM, sendo: (a) imagem original; (b) classificação por regiões; (c) classificação *pixel-a-pixel*.



Fonte: produção do próprio autor, 2015.

O círculo superior mostra local em que a vegetação pertence ao estágio avançado, porém, com dossel espectralmente mais homogêneo que outras áreas. Por este motivo, foi erroneamente classificada como estágio médio. A classificação por regiões captou parte dessa variabilidade. No círculo inferior, percebem-se áreas pertencentes à classe campo, classificadas como campo sujo, principalmente na metodologia por *pixel*. A superioridade da classificação obtida pelo método SVM por regiões pode estar relacionada ao fato de esta metodologia levar em consideração, além de seu valor numérico, o contexto de inserção do *pixel*, o que permitiu discriminar melhor algumas classes que a metodologia *pixel-a-pixel*, como o estágio avançado da vegetação e campo sujo.

Vários estudos têm mostrado que o algoritmo SVM apresenta ótimos resultados na classificação de imagens. Mountrakis et al. (2010) relacionam mais de cem publicações relevantes que usaram o SVM em aplicações no Sensoriamento Remoto. Alguns autores compararam o classificador SVM com a classificação pelo algoritmo paramétrico MaxVer (MATHUR

e FOODY, 2008; DEVADAS et al., 2012; NASCIMENTO et al., 2013; ANDRADE et al., 2014), obtendo resultados superiores com o SVM. Outros autores o compararam com as redes neurais artificiais (DIXON e CANDADE, 2008; SZUSTER et al., 2011; NITZE et al., 2012) e obtiveram resultados similares com os dois métodos, entretanto destacaram a melhor capacidade do SVM em distinguir classes espectralmente semelhantes, além de necessitar menor número de amostras de treinamento para atingir bons resultados.

Porém, diferentemente das árvores de decisão, o SVM é apontado como um classificador “caixa-preta” em alguns trabalhos (CHAVES et al., 2007), ou seja, modelos que não explicam o processo pelo qual um dado resultado é obtido. O mapeamento entrada-saída resultante é composto de uma combinação linear de funções-núcleo, e as decisões que ele toma nem sempre são fáceis de serem explicadas (BARBELLA et al., 2009). Dessa forma, não se veem claramente as razões de certo ponto ou amostra ser atribuída a determinada classe. Já as árvores de decisão podem fornecer modelos facilmente interpretáveis, porém nem sempre alcançam resultados acurados como os do SVM (BARBELLA et al., 2009).

Alguns algoritmos para extrair conhecimento de SVM têm sido propostos na literatura, entre eles o *RulExtSVM* (FU et al., 2004), *SVM+Prototypes* (NUÑEZ et al., 2002), o *FREx_SVM* (CHAVES et al., 2007). Barbella et al. (2009) propuseram duas técnicas, uma para encontrar o vetor de suporte que possui maior contribuição na classificação de determinado ponto, e a segunda com o objetivo de encontrar os limiares de separação das classes.

Outros trabalhos recomendam métodos para guiar a escolha dos parâmetros *kernel*, como o *Grid Search* (MOORE et al., 2011), Algoritmos Genéticos (HUANG e WANG, 2006; CAROLINA e CARVALHO, 2008; ZHANG et al., 2009) ou Recozimento Simulado (BONESSO, 2013).

4.4 COMPARAÇÃO ENTRE AS METODOLOGIAS

Nesta seção, serão apresentadas as comparações feitas através do teste z entre todas as metodologias testadas neste trabalho, bem como a avaliação do desempenho individual das classes nas metodologias testadas de acordo com a área-teste. A Tabela 34 mostra a comparação entre os métodos para a área-teste A.

Tabela 34 - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste A.

| | InterIMAGE | WEKA | GeoDMA | Monteverdi |
|-------------------|--------------------|--------------------|---------------|-------------------|
| Kappa | 0,73 | 0,66 | 0,69 | 0,61 |
| Variância | 0,0003385 | 0,0003804 | 0,0003660 | 0,0004193 |
| InterIMAGE | 0 | | | |
| WEKA | 2,66* | 0 | | |
| GeoDMA | 1,42 ^{ns} | 1,23 ^{ns} | 0 | |
| Monteverdi | 4,38* | 1,74 ^{ns} | 2,96* | 0 |

z estatístico = 1,96: significante a 95% de probabilidade; ns= não-significante.

Fonte: produção do próprio autor, 2015.

Na Tabela 34 percebe-se que algumas metodologias diferiram significativamente, entre elas o InterIMAGE e WEKA, InterIMAGE e Monteverdi, e GeoDMA e Monteverdi. Pelo Kappa, percebe-se que os resultados alcançados no InterIMAGE e no GeoDMA foram superiores aos obtidos no WEKA e Monteverdi.

A comparação entre os acertos específicos para cada classe pode ser visualizada através do Kappa condicional do produtor e usuário obtido para a área-teste A (Tabela 35).

Tabela 35 - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste A.

| Classe | InterIMAGE | | WEKA | | GeoDMA | | Monteverdi | |
|------------|------------|------|------|------|--------|------|------------|------|
| | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu |
| VEA | 0,58 | 0,63 | 0,61 | 0,62 | 0,42 | 0,53 | 0,75 | 0,84 |
| VEM | 0,53 | 0,48 | 0,33 | 0,4 | 0,32 | 0,49 | 0,72 | 0,77 |
| VEI | 1 | 1 | 1 | 0,7 | 1 | 0,39 | 0,97 | 0,48 |
| Agricult. | 0,93 | 0,88 | 1 | 1 | 1 | 1 | 0,52 | 0,97 |
| campo | 1 | 1 | 1 | 1 | 1 | 1 | 0,36 | 0,33 |
| edificação | 0,56 | 0,46 | 0,56 | 0,66 | 0,63 | 0,17 | 0,6 | 0,92 |
| sombra | 0,45 | 0,48 | 0,17 | 0,65 | 0,41 | 0,55 | 0,45 | 0,75 |

Fonte: produção do próprio autor, 2015.

A Tabela 35 mostra que, além das classes sombra e edificação, os menores acertos foram nas classes VEA e VEM. Apesar de o algoritmo SVM ter obtido desempenho geral inferior à árvore C4.5 do InterIMAGE e GeoDMA, ele se destacou por apresentar melhores resultados na classificação do estágio sucessional avançado e médio, superando as árvores de decisão. Isso pode estar relacionado à capacidade do SVM de separar classes espectralmente semelhantes (SZUSTER et al., 2011; HEGDE et al., 2014). Andrade et al. (2014) afirmaram que o método de classificação não-paramétrico SVM apresenta-se como uma alternativa viável ao mapeamento de fisionomias vegetais em imagens de alta resolução. Estes autores obtiveram exatidão do usuário de 89% na discriminação da classe arbórea pelo SVM, enquanto pelo algoritmo MaxVer foi de 60%.

A Tabela 36 mostra a comparação entre as metodologias aplicadas na área-teste B.

Tabela 36 - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste B.

| | InterIMA GE | WEKA | GeoD MA | Monteve rdi | Monteverdi(<i>pixel</i>) |
|--|-------------------|--------------------|------------|----------------|----------------------------|
| Kappa | 0,85 | 0,89 | 0,72 | 0,88 | 0,60 |
| | | 0,00014 | 0,00028 | 0,000156 | |
| Variância | 0,0001871 | 49 | 21 | 2 | 0,00035 |
| InterIMAGE | 0 | | | | |
| WEKA | 2,08* | 0 | | | |
| GeoDMA | 5,75* | 7,86* | 0 | | |
| Monteverdi | 1,5 ^{ns} | 0,58 ^{ns} | 7,28* | 0 | |
| Monteverdi (<i>pixel</i>) | 10,56* | 12,7* | 4,78* | 12,11* | 0 |

z estatístico = 1,96: significativa a 95% de probabilidade; ns= não-significante.

Fonte: produção do próprio autor, 2015.

Percebe-se que todas as metodologias que realizaram a classificação por regiões diferiram significativamente da metodologia por *pixel*, mostrando-se superiores a esta. Os melhores resultados alcançados nesta área-teste foram obtidos pelo algoritmo SVM por regiões no OTB Monteverdi e pela árvore de decisão SimpleCart obtida no WEKA e implementada no QGIS.

O desempenho individual na classificação de cada classe é demonstrado na Tabela 37.

Tabela 37 - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste B.

| Classe | InterIMAGE | | WEKA | | GeoDMA | | Monteverdi | | Monteverdi(<i>pixel</i>) | |
|-----------|------------|------|------|------|--------|------|------------|------|----------------------------|------|
| | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu |
| VEA | 1 | 0,64 | 1 | 1 | 1 | 0,59 | 1 | 1 | 0,48 | 1 |
| VEM | 0,78 | 1 | 1 | 1 | 0,75 | 0,52 | 1 | 1 | 1 | 0,72 |
| VEI | 1 | 1 | 1 | 0,92 | 1 | 1 | 1 | 0,56 | 1 | 0,47 |
| agricult. | 1 | 1 | 1 | 0,56 | 1 | 0,43 | 1 | 1 | 1 | 1 |
| campo | 0,59 | 0,52 | 1 | 0,91 | 1 | 0,52 | 0,94 | 1 | 0 | 0,44 |
| c. sujo | 0,75 | 0,8 | 0,7 | 1 | 0,56 | 1 | 0,83 | 0,98 | 0,26 | 0,19 |
| reforest. | 0,7 | 0,49 | 0,5 | 0,74 | 0 | 0 | 0,48 | 0,72 | 0,6 | 0,6 |
| sombra | 0,91 | 0,23 | 0,9 | 0,24 | 0,57 | 0,54 | 0,81 | 0,27 | 0,72 | 0,63 |

Fonte: produção do próprio autor, 2015.

Os resultados mostram que todas as metodologias tiveram resultados variando de bom a excelente na discriminação dos estádios sucessionais da vegetação. A classificação SVM por *pixel* teve muitas áreas de estágio avançado classificadas como médio, gerando erros de omissão e comissão, respectivamente para estas duas classes. A superioridade da classificação utilizando a árvore de decisão do WEKA e a SVM por regiões está atrelada ao fato de que estas metodologias foram as que conseguiram distinguir melhor as classes campo e campo sujo. Estas duas classes tiveram resultados inferiores na classificação por *pixel*.

A Tabela 38 mostra o comparativo do desempenho individual de cada classe com as metodologias aplicadas na área-teste C.

Tabela 38 - Teste z para significância entre índices Kappa resultantes da classificação de diferentes métodos aplicados na área-teste C.

| | InterIMAGE | InterIMAGE(2) | WEKA | GeoDMA | Monteverdi |
|----------------------|--------------------|--------------------|--------------------|---------|------------|
| Kappa | 0,82 | 0,64 | 0,80 | 0,78 | 0,63 |
| Variância | 0,0003877 | 0,0006069 | 0,0004205 | 0,00045 | 0,0005848 |
| InterIMAGE | 0 | | | | |
| InterIMAGE(2) | 5,5* | 0 | | | |
| WEKA | 0,66 ^{ns} | 4,83* | 0 | | |
| GeoDMA | 1,29 ^{ns} | 4,19* | 0,63 ^{ns} | 0 | |
| Monteverdi | 6,06* | 0,45 ^{ns} | 5,37* | 4,72* | 0 |

z estatístico = 1,96: significante a 95% de probabilidade; ns= não-significante.

Fonte: produção do próprio autor, 2015.

De acordo com a Tabela 38, as metodologias executadas no InterIMAGE, WEKA e GeoDMA mostraram-se significativamente superiores à classificação realizada pelo Monteverdi e InterIMAGE sem a componente MDS, aqui denominada InterIMAGE (2). Isso demonstra que o MDS agrega informações importantes a serem utilizadas na classificação das tipologias vegetais, como já observado por

outros autores. Yu et al. (2006) citam que, para a discriminação de classes como floresta, arbustos, gramíneas, características topográficas tornam-se cada vez mais importantes, enquanto as características espectrais são menos essenciais. Carvalho Júnior et al. (2008) citam que o Modelo Digital de Elevação indica estruturas de vegetação em ambiente com alta variabilidade espacial, porque se correlaciona com fatores ambientais, tais como o microclima, umidade, solo e processos geomorfológicos. Estes autores utilizaram dados de altimetria na classificação por árvore de decisão para discriminar tipologias vegetais da Mata Atlântica, obtendo bons resultados. Piazza e Vibrans (2014), ao classificarem estádios sucessionais da vegetação com as ortoimagens do levantamento aerofotogramétrico de SC utilizando somente as bandas espectrais RGB, obtiveram os maiores erros no estádio inicial e floresta plantada, os quais foram classificados erroneamente como estádio avançado. Posteriormente, Piazza (2014) realizou experimentos com as bandas RGB, NIR, MDS e MDT e obteve aumento de acurácia na classificação das tipologias vegetais.

Tedesco et al. (2014) utilizaram o MDS normalizado na árvore de decisão CART, ou seja, após subtração do MDT, obtendo a altura dos objetos para discriminação entre vegetação arbórea e rasteira, e tiveram acurácia de 100% na discriminação destas duas classes. Francisco e Almeida (2013) demonstraram a importância da inclusão de outras informações, não somente espectrais, para a discriminação das classes de cobertura da terra, sendo que o atributo declividade foi capaz de discriminar duas classes (afloramento rochoso e vegetação herbácea rala) que apresentavam respostas espectrais e texturais semelhantes na árvore de decisão.

Na Tabela 39, apresenta-se o Kappa condicional de cada classe nas metodologias aplicadas na área-teste C.

Tabela 39 - Kappa condicional do produtor (Kcp) e do usuário (Kcu) obtido em cada metodologia para a área-teste C.

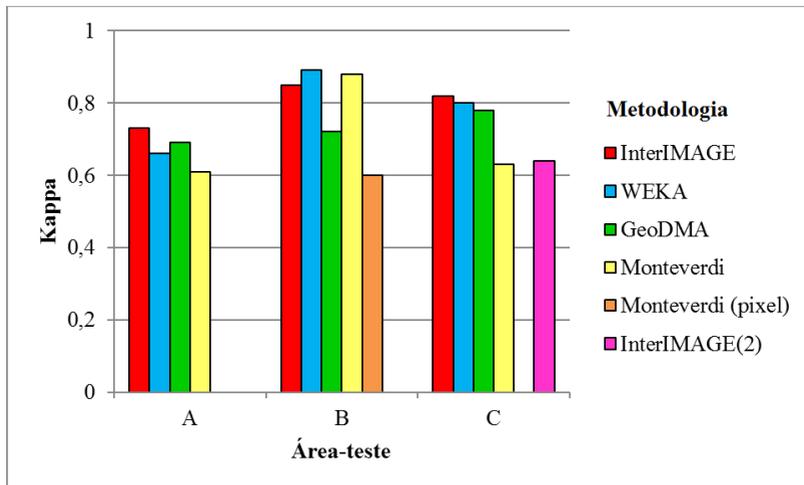
| Classe | InterIMAGE | | InterIMAGE (2) | | WEKA | | GeoDMA | | Monteverdi | |
|--------|------------|------|----------------|------|------|------|--------|------|------------|------|
| | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu | Kcp | Kcu |
| VEA | 0,54 | 0,94 | 0,36 | 0,45 | 0,44 | 0,57 | 0,52 | 0,51 | 0,79 | 0,53 |
| VEM | 1 | 0,34 | 1 | 0,39 | 1 | 0,34 | 1 | 0,25 | 1 | 0,38 |
| VEI | 0,69 | 0,65 | 0,08 | 0,12 | 0,47 | 0,67 | 0,42 | 0,68 | 0,45 | 0,97 |
| campo | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| sombra | 0,75 | 0,21 | 0,73 | 0,25 | 0,87 | 0 | 0,76 | 0,24 | 0,17 | 0,1 |

Fonte: produção do próprio autor, 2015.

Observando-se o Kappa condicional da área-teste C obtido nas diferentes metodologias, percebe-se que esta foi a área em que ocorreram maiores confusões entre os estádios sucessionais da vegetação, ocorrendo, inclusive, no estádio inicial. Como citado anteriormente, a vegetação desta área é mais complexa, apresentando maior heterogeneidade nos diferentes estádios, além de estarem intrinsecamente conectadas e sem influências antrópicas diretas, diferentemente das outras áreas. O estádio médio, por ser menos representativo, apresentou maiores erros de comissão em todas as metodologias, tendo sido os outros dois estádios atribuídos a ele. Novamente, percebe-se que o classificador SVM obteve resultados melhores que as árvores de decisão na distinção dos estádios sucessionais da vegetação.

O acerto geral de todas as metodologias e respectivas áreas-teste podem ser visualizados na Figura 56, a qual ilustra o Kappa global de cada uma delas.

Figura 56 - Kappa global obtido nas metodologias de acordo com a área-teste.



Fonte: produção do próprio autor, 2015.

Todas as metodologias alcançaram valores considerados como muito bom a excelente de acordo com Landis e Koch (1977). O menor Kappa encontrado foi de 0,6 na classificação por *pixel*, e o maior de 0,89 pela árvore construída no WEKA para a área-teste B. Quanto aos classificadores por árvore de decisão, percebe-se que a classificação pelo InterIMAGE com o algoritmo de árvore de decisão C4.5 alcançou os melhores resultados nas áreas-teste A e C, porém, não significativamente superiores ao GeoDMA nestas duas áreas. Na área-teste A, o resultado do WEKA foi inferior ao InterIMAGE e GeoDMA, porém na área-teste B mostrou-se superior a esses aplicativos. Isso demonstra que os classificadores por árvore de decisão podem ter desempenho diferente, conforme as particularidades de cada área a ser classificada. Além disso, os resultados variam conforme os atributos utilizados e a complexidade da árvore gerada. Diante disso, não teria como indicar um deles como sendo o melhor.

Nas metodologias que utilizaram o algoritmo SVM, percebe-se que houve uma variação na acurácia de acordo com a área testada. Nas áreas-teste A e C, o classificador SVM por regiões implementado no OTB Monteverdi foi significativamente inferior aos métodos que utilizaram a árvore de decisão. Porém, na área-teste B, este classificador teve um excelente desempenho, diferindo significativamente do GeoDMA. Isso pode estar atrelado à função-núcleo e respectivos parâmetros escolhidos para o classificador SVM, que podem ser bons para determinadas áreas e inferiores para outras. Além disso, Kalkana et al. (2013) citam que um dos problemas ao se classificar imagens com o SVM é a necessidade de vetorização desses arquivos para comparar com outros classificadores. De acordo com os autores, o processo de vetorização causa perda de informações pela suavização o que é similar à digitalização manual.

Recentemente, alguns trabalhos reportados na literatura realizaram a comparação entre algoritmos de classificação SVM com árvores de decisão. Anjos et al. (2015) compararam o desempenho do algoritmo C4.5, florestas aleatórias e SVM no WEKA e obtiveram resultados superiores com o SVM. No entanto, apenas o C4.5 pode ser reproduzido visualmente de forma a permitir a classificação da imagem em outro aplicativo. Costa et al. (2015) realizaram diferentes experimentos em imagens do sensor MODIS para classificar áreas de formações campestres nativas e florestas cultivadas no Cerrado brasileiro, utilizando os classificadores florestas aleatórias, árvores de decisão e SVM, em programas *open source* como o GeoDMA, WEKA e LibSVM. Em todos os três experimentos, o método SVM apresentou os melhores resultados da classificação, seguido do método de florestas aleatórias e o de árvores de decisão, respectivamente. Os melhores resultados variaram de 0,77 com o SVM, 0,7 com florestas aleatórias e 0,6 com árvores de decisão.

Sabe-se, no entanto, que a metodologia envolvendo o algoritmo SVM não foi aprofundada. Uma análise mais detalhada de algoritmos que automatizam o processo de escolha da função-núcleo e seus parâmetros poderia ser efetuada, de forma a melhorar a acurácia dessas classificações, já que o algoritmo foi o que melhor discriminou os estádios sucessionais da vegetação neste trabalho.

A classe sombra manteve baixos índices de exatidão do usuário e Kappa condicional do usuário na maioria das metodologias testadas. Porém, importante destacar que o fotointérprete não considerou as sombras menores no interior da vegetação na elaboração do mapa de referência, apenas aquelas mais representativas. Por este motivo, na maioria das metodologias, essa classe teve uma superclassificação em relação ao mapa de referência, e, portanto, maiores erros de comissão.

Quanto aos estádios sucessionais da vegetação, as classes tiveram Kappa condicional do usuário superior a 0,5 na maioria das metodologias testadas, com destaque para os estádios avançado e inicial, que chegaram a superar 0,8. A exatidão do usuário indica que estas classes tiveram pouco erro de comissão, ou seja, ao serem classificadas como determinado estágio, existe uma boa chance de se tratar, de fato, desse estágio. Isso é um fator relevante ao se considerar a determinação do estágio sucessional da vegetação como forma de subsidiar inventários em campo.

Em todas as metodologias testadas, as maiores confusões prevaleceram no estágio médio da vegetação, classificado incorretamente como avançado ou inicial, o que pode ser explicado por se tratar da classe intermediária de vegetação, possuindo semelhança espectral tanto com estágio avançado, quanto inicial. Araújo et al. (2006), para melhor discriminar as classes vegetação arbórea e vegetação rasteira, introduziram um atributo de textura que diferenciava ambas as classes quanto ao número de subobjetos contidos no alvo de interesse.

Esta abordagem foi promissora, porque a classe arbórea possui um maior número de subobjetos pela presença de sombra entre as folhas, proporcionando maior textura.

Outros trabalhos reportados na literatura que objetivavam discriminar o estágio da vegetação também encontraram maior confusão no estágio médio. Entre eles, cita-se Andrade et al. (2014), que testaram algoritmos paramétrico (MaxVer) e não-paramétrico (SVM) na classificação da fisionomia vegetal em imagens QuickBird, sendo que apenas a vegetação herbácea-arbustiva, equivalente ao estágio médio no presente estudo, apresentou Kappa condicional do usuário inferior a 0,6 nos experimentos realizados. Amaral et al. (2009), ao utilizarem imagens dos satélites CBERS, IRS e Quickbird para discriminar os estágios sucessionais, concluíram que somente foi eficiente a separação do estágio inicial de sucessão secundária dos estágios avançado e médio. Furtado et al. (2013), ao aplicarem a OBIA na classificação das fisionomias vegetais com imagens de alta resolução espacial, constataram que a classe vegetação arbórea obteve maior valor do índice do usuário, alcançando 0,88 no Kappa condicional, seguida da gramínea rala, com valor de 0,77. Já as classes gramínea e herbácea-arbustiva apresentaram desempenho inferior no estudo dos autores. Sette (2009), ao classificar os estágios sucessionais da vegetação de Floresta Ombrófila Densa em imagens do satélite FORMSAT-2, no Sul da Bahia, obteve acurácia de 60,5% ao utilizar somente as bandas do visível e 91% ao incluir atributos texturais. A autora também relatou que, entre os estágios sucessionais de Mata Atlântica estudados, o estágio médio obteve correlações menores e menos significância com as medidas de textura, o que ela atribuiu à grande amplitude de variação desta classe em relação aos outros estágios. Piazza (2014), ao classificar os estágios sucessionais da Floresta Ombrófila Densa em SC pela OBIA e ortoimagens do levantamento aerofotogramétrico do estado, obteve Kappa entre 0,87 e 0,90, sendo os estágios sucessionais

da vegetação com Kappa condicional superiores a 0,8. Entre os estádios sucessionais, o autor também obteve as maiores confusões entre os estádios médio e avançado, porém todos superaram Kappa de 0,8. Cintra (2007) diferenciou estádios sucessionais da vegetação no Rio de Janeiro com imagens IKONOS, e através da fotointerpretação definiu uma chave de interpretação. Para a autora, a classe ‘estádio inicial’ foi a mais fácil de ser distinguida das outras por sua uniformidade de textura, que geralmente reflete a dominância de poucas espécies.

Os métodos de classificação por regiões e por objeto mostraram ser mais eficientes na classificação de estádios de sucessão florestal em imagens de alta resolução espacial, quando comparados com o método de classificação por *pixel*, apresentando valores mais elevados de índice Kappa. Isso demonstra a importância da inserção de características intrínsecas aos segmentos, como padrões estatísticos e texturais para a classificação dos estádios de sucessão florestal em fragmentos de mata nativa ao se utilizar imagens de alta resolução espacial.

Alguns estudos têm mostrado a superioridade da OBIA e da classificação por regiões na classificação de imagens de alta resolução espacial. Gao e Mas (2008), ao compararem a OBIA com a classificação por *pixel* em imagens com diferentes resoluções espaciais, concluíram que a OBIA teve acurácia superior quando aplicada em imagens de alta resolução espacial. Amaral et al. (2009) compararam o desempenho da classificação por regiões e por *pixel* na classificação dos estádios sucessionais da vegetação, sendo que a classificação por regiões apresentou valores mais elevados de índice Kappa. Os resultados alcançaram Kappa de 0,28 para a classificação por *pixel* pelo algoritmo MaxVer e 0,54 para a classificação por regiões pelo algoritmo baseado na distância de Battacharyya. Cleve et al. (2008), tiveram um aumento de 13,7% na acurácia da classificação da vegetação utilizando a

abordagem OBIA em relação a classificação por *pixel*. Yu et al. (2006), ao classificarem imagens aéreas para distinguir tipologias vegetais, compararam a OBIA com a classificação por *pixel* pelo algoritmo MaxVer, também obtendo resultados mais promissores com a OBIA.

5. CONSIDERAÇÕES FINAIS

Este capítulo destaca as principais considerações em face dos resultados obtidos e dos empecilhos encontrados em cada etapa da realização deste trabalho. Serão também feitas algumas recomendações quanto à continuidade da pesquisa.

Sobre a segmentação:

- a) O aplicativo SPT se mostrou viável para o ajuste automático de parâmetros de segmentação, tanto ao fornecer os melhores parâmetros de segmentação de acordo com a área de estudo, quanto na avaliação dos parâmetros escolhidos pelo usuário, através das medidas de discrepância. Obtiveram-se os menores valores de discrepância para a segmentação com os valores de parâmetros fornecidos automaticamente pelo aplicativo para o algoritmo de segmentação de Baatz e Schäpe, no entanto, tais valores nem sempre resultaram ser os mais adequados na visão do analista, como o caso da área-teste C.
- b) Este aplicativo não teve desempenho satisfatório nas análises conduzidas com o algoritmo de média-móvel. Houve problemas no desempenho do programa, e quando foi possível obter os parâmetros, verificou-se que estes não eram representativos dos objetos na cena, havendo uma supersegmentação. Por este motivo, apenas utilizou-se a tradicional metodologia de tentativa e erro para procedimentos envolvendo este algoritmo.
- c) Considera-se que o ajuste automático dos parâmetros pelo SPT pode ser útil para fornecer um ponto de partida, ou seja, uma ideia preliminar sobre quais parâmetros seriam adequados para a cena. Posteriormente, a realização de testes de parâmetros

através de tentativa e erro ainda constitui uma forma viável na descoberta dos parâmetros mais adequados.

- d) A segmentação foi uma das etapas mais demoradas, mesmo tentando-se automatizar esse processo através do SPT, já que a subjetividade do analista ainda foi necessária nesse procedimento. Além disso, os processos de segmentação e de descrição de classe tiveram de ser adaptados para cada área-teste, o que resulta em maior consumo de tempo.

Sobre a classificação por mineração de dados e SVM:

- a) As ferramentas de mineração de dados permitiram escolher os atributos que melhor discriminam as classes de uso e cobertura do solo, dentre elas os diferentes estádios sucessionais da vegetação. A etapa da seleção de atributos atendeu ao propósito de redução da dimensão e aumento da acurácia do modelo de árvore de decisão gerado. Em geral, o mínimo e máximo valor de *pixel* da componente MDS foram os atributos que mais contribuíram em todas as áreas-teste, além das operações entre as bandas correspondentes à região do NIR e vermelho.
- b) As metodologias empregando a árvore de decisão, tanto o algoritmo C4.5, quanto o SimpleCart, apresentaram grande potencial na análise dos estádios da vegetação. A árvore de decisão possui a vantagem de permitir ao usuário visualizar como o processo de classificação é feito, podendo ser aperfeiçoado de forma a melhorar a acurácia de acordo com as peculiaridades de cada local. A estratégia de uso de uma rede hierárquica e da descrição automática das classes com base em algoritmos supervisionados de mineração de dados mostrou ser uma forma eficaz e de pouca subjetividade de classificação do uso e

cobertura da terra. Além disso, o fato de estas metodologias terem envolvido procedimentos automáticos, faz com que sejam exportáveis para outras áreas e datas, se realizada nova coleta de amostras.

- c) As técnicas de mineração de dados permitiram constatar que a vegetação apresenta características espectrais e texturais distintas conforme o seu estágio sucessional da vegetação, confirmando uma das hipóteses levantadas neste estudo.
- d) O método SVM, apesar de ter apresentado desempenho inferior em duas áreas-teste (A e C), foi o que teve maior índice de acerto na discriminação dos estádios sucessionais da vegetação. Isso corrobora com os estudos anteriormente citados, que mencionam que o algoritmo tem potencial para distinção de classes espectralmente semelhantes. No entanto, diferentemente das árvores de decisão, não se visualiza como o algoritmo chegou àquele resultado, o que dificulta sua interpretação e replicação em outras áreas-teste.

Sobre os aplicativos utilizados:

- a) O *software* SPT, mesmo não fornecendo o resultado mais esperado pelo analista, como para o algoritmo de média-móvel, pode servir como uma ferramenta para construir cenários de segmentação, fornecendo uma ideia preliminar de quais parâmetros seriam mais adequados para certa imagem. O programa, no entanto, é um dos mais limitados em termo de tamanho das imagens a serem utilizadas.
- b) Os *softwares* InterIMAGE e GeoDMA foram promissores e possibilitaram realizar todo o procedimento, desde a segmentação, extração de

atributos, até a classificação das imagens. O classificador C4.5 implementado no InterIMAGE constitui um avanço para o programa e já foi explorado em trabalhos como Meneghetti (2013) e Rodrigues (2014). A única questão é que não tem como escolher os parâmetros da árvore de decisão gerada, diferentemente do que ocorre no GeoDMA.

- c) Uma vantagem na utilização do GeoDMA é a facilidade nos processos de extração, normalização e armazenamento dos atributos dos dados, que podem ser utilizados posteriormente em outras análises (KÖRTING, 2009). Por estar associado a um GIS, não é necessário outro programa para visualização dos atributos inerentes a cada objeto. No entanto, Körting (2012) atenta para o fato de que outros *softwares*, como o InterIMAGE e eCognition, fornecem interfaces amigáveis para a criação de novos atributos a partir dos atributos pré-computados, como o caso das operações entre bandas e NDVI, por exemplo. No GeoDMA, estes procedimentos só podem ser executados manualmente, editando as colunas do banco de dados, tarefa que demanda maior tempo do usuário (KÖRTING, 2012). Quanto ao custo computacional, percebeu-se que o GeoDMA tem uma performance melhor que o InterIMAGE.
- d) O WEKA se mostrou uma interessante ferramenta de mineração de dados. Neste programa a possibilidade de testar diferentes algoritmos de árvore de decisão, e avaliar previamente a acurácia dos modelos gerados em função das amostras, possibilita ao analista escolher o modelo que julgar mais adequado. As ferramentas de seleção de atributos melhoraram a acurácia do modelo gerado e diminuiu a complexidade da árvore de decisão, permitindo que fosse mais facilmente implementada em outro programa.

- e) O OTB Monteverdi também é uma ferramenta adequada que fornece todas as etapas de classificação da imagem, tanto por *pixel* quanto por regiões, e se trata de um aplicativo ainda pouco explorado na comunidade científica. Entre os trabalhos que reportam a esse *software* citam-se Christophe e Inglada (2009) que apresentam as potencialidades gerais do OTB; Grinzonnet e Inglada (2010), que apresentam as funcionalidades do Monteverdi, o aplicativo integrado ao OTB; e Vu (2012) que demonstra a classificação supervisionada por regiões no OTB Monteverdi pelo algoritmo SVM.
- f) De maneira geral, todos os programas *open source* atenderam aos objetivos propostos, mostrando-se excelentes alternativas aos programas comerciais na classificação de imagens que utilizam a classificação baseada em objeto ou por regiões. Porém, ainda se faz necessário um maior amadurecimento desses programas, visto que possuem capacidade limitada quanto ao tamanho da imagem a ser classificada. No SPT, InterIMAGE, GeoDMA e Monteverdi, por exemplo, por várias vezes ocorreram falhas, ou estes pararam de funcionar durante a realização dos processamentos, interrompendo inesperadamente o trabalho, tornando necessária a repetição de etapas realizadas anteriormente. No entanto, percebe-se que em cada nova versão destes aplicativos, melhorias significativas são apresentadas, tanto no desempenho computacional quanto nos recursos implementados.
- g) Confirmou-se a hipótese de que metodologias *open source* de classificação digital de imagens podem ser aplicadas para discriminar diferentes estádios sucessionais da vegetação.

Sobre a acurácia dos resultados:

- a) As metodologias testadas neste trabalho tiveram bons resultados, com Kappa variando de 0,64 a 0,89 na classificação por árvore de decisão, e de 0,6 a 0,88 na classificação pelo algoritmo SVM. Variaram de muito boa a excelente de acordo com a literatura. Quanto à classificação dos estádios sucessionais, o pior resultado encontrado, de acordo com o Kappa condicional, foi de 0,25 na VEM, e os melhores resultados foram próximos a 100% de acerto na VEA e VEI, considerando o arredondamento dos dados.
- b) Visualmente, todas as classificações podem ser avaliadas qualitativamente como muito boas, apresentando coerência espacial entre a distribuição das classes e coerência com o mapa de referência.
- c) O MDS aliado às bandas espectrais, por fornecer informações texturais sobre a estrutura vertical da imagem, resultou em significativa melhora na classificação dos estádios sucessionais da vegetação, com Kappa passando de 0,64 para 0,82 considerando a mesma metodologia empregada.
- d) A classificação automatizada é uma alternativa interessante para o estudo de áreas extensas, economizando assim esforço do operador bem como o tempo de processamento computacional, porém, demanda a análise de acurácia dos resultados.
- e) Dessa forma, pode-se considerar que as imagens de alta resolução espacial e metodologias empregadas neste estudo podem auxiliar os trabalhos a campo, porém, certa cautela é necessária se for considerado um inventário florestal apenas com informações remotas, já que muitas áreas possuem particularidades que devem ser reconhecidas preliminarmente em campo.

Sobre a viabilidade da aplicação operacional das metodologias:

- a) Além do presente estudo, outros trabalhos objetivaram realizar a classificação dos estádios sucessionais do Bioma Mata Atlântica por diferentes imagens e técnicas de Sensoriamento Remoto: Vibrans (2003) utilizou imagens Landsat; Cintra (2007), imagens IKONOS; Sette (2009) utilizou imagens FORMSAT; Amaral et al. (2009) usaram imagens CBERS, IRS e Quickbird; e mais recentemente, Piazza (2014) utilizou imagens do aerolevante fotogramétrico de SC. Ressalta-se, no entanto, que a resolução temporal dos sensores utilizados na maioria dos trabalhos sobre o tema, a qual varia entre média e baixa, não é adequada para ser aplicada de forma efetiva e operacional no trabalho dos órgãos ambientais competentes. Nesse contexto, o desenvolvimento e padronização de metodologias para identificação de estádios de sucessão com o uso de Sensoriamento Remoto, que seja aplicável à fiscalização e licenciamento ambiental, ainda constitui um desafio a ser explorado em estudos futuros.
- b) A Resolução CONAMA nº 04/94 não contempla critérios para identificação dos estádios sucessionais da vegetação por Sensoriamento Remoto. Ela foi elaborada para classificação da vegetação *in loco* com a floresta ainda em pé, considerando-se características observadas e/ou medidas no local, como as variáveis dendrométricas, espécies indicadoras e outros parâmetros fitossociológicos. Critérios para identificação de estádios por Sensoriamento Remoto, possíveis de serem extraídos no texto das Resoluções, seriam de suma importância para utilização pelos órgãos ambientais competentes. Nada impede que se

utilizem essas características atualmente, entretanto tal caracterização pode ser considerada frágil, já que não consta nos textos legais, não estando amparada legalmente (TRÄUCZYNSKI, 2013).

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Seguem abaixo recomendações e propostas para trabalhos futuros:

- a) a fim de melhorar os resultados obtidos pela classificação utilizando a árvore de decisão, pode-se testar a integração da metodologia automatizada com o conhecimento do fotointérprete, gerando uma classificação semiautomatizada;
- b) testar as metodologias utilizando a análise das principais componentes e conversão do sistema de cores aditivo RGB para o IHS (*Intensity* ou Intensidade – I, *Hue* ou Matiz - H, *Saturation* ou Saturação – S), pois outros estudos observaram que tais informações foram mais relevantes na classificação do que as informações das bandas puras (PINHO et al., 2008; FRANCISCO e ALMEIDA, 2012; FURTADO et al., 2013; ANDRADE et al., 2014);
- c) desenvolver melhor as metodologias envolvendo a classificação pelo algoritmo SVM, como variação nas funções-núcleo e seus parâmetros, escolha automática de tais parâmetros, ou abordagens para a extração de informações da SVM, já que foi observado que o algoritmo possui excelente potencial na discriminação dos estádios sucessionais;
- d) aplicar a metodologia em áreas maiores, já que, em virtude da limitação dos *softwares* utilizados, restringiu-se o estudo a áreas menores;

- e) aplicação das metodologias em imagens obtidas por plataformas espaciais, que, por possuírem maior resolução temporal, permitem uma análise da dinâmica da vegetação;
- f) desenvolver metodologias para um estudo multitemporal de fragmentos florestais. Dessa forma, além de permitir o monitoramento de possíveis desmatamentos ilegais, também há a possibilidade de se inferir a idade dos fragmentos florestais examinados, fornecendo subsídios para a identificação de estádios de sucessão da floresta;
- g) realizar a identificação de espécies indicadoras dos diferentes estádios sucessionais através do Sensoriamento Remoto, utilizando laser scanner aerotransportado ou imagens oriundas de sensores hiperespectrais de forma isolada ou integrada, por exemplo, como forma de auxiliar na classificação dessas tipologias.

REFERÊNCIAS BIBLIOGRÁFICAS

ACHANCCARAY DIAZ, P. M. **A Comparasion of Segmentation Algorithms for Remote Sensing**. 84 p. 2014. Dissertação (Mestrado em Engenharia Elétrica)- Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2014.

ADDINK, E.A.; JONG, S.M. de; PEBESMA, E.J. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. **Photogrammetric Engineering & Remote Sensing**, 73 (8), p. 905–912, 2007.

ALIXANDRINI, M. Classificação de imagens Landsat 7 ETM multiespectrais utilizando SVM. In: Simpósio Brasileiro de Sensoriamento Remoto. 14.,2009, Natal, Brasil. **Anais...** São José dos Campos: INPE, p. 7757-7764, 2009.

ALMEIDA FILHO, N. O.; HAYAKAWA, E. H.; PRADO, B. R. Mapeamento da cobertura da terra no município de Alfenas-MG utilizando imagens do satélite RapidEye. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, PR. **Anais...** INPE, p. 1502-1508, 2013.

ALMEIDA, W. G. Processo de Mineração de Dados na Definição de Arquiteturas de Satélites para Missões Espaciais, 2010. Disponível em:
<<http://mtcm19.sid.inpe.br/col/sid.inpe.br/mtcm19/2010/10.05.14.22/doc/publicacao.pdf?languagebutton=pt-BR>>. Acesso em: 15 de dez. 2014.

AMARAL, M. V. F. et al. Avaliação e comparação de métodos de classificação de imagens de satélites para o mapeamento de estádios de sucessão florestal. **Revista Arvore**, Viçosa, v.33, n. 3, Mai./Jun, 2009. ISSN 0100-6762.

ANDRADE, A. C.; FRANCISCO, C. N.; ALMEIDA, C. M. de. Desempenho de classificadores paramétrico e não paramétrico na classificação da fisionomia vegetal. **Revista Brasileira de Cartografia**, Rio de Janeiro, n. 66/2: p. 349-363, 2014. ISSN: 1808-0936.

ANDREOLA, R.; HAERTEL, V. Classificação de Imagens Hiperespectrais Empregando Support Vector Machines. **Boletim de Ciências Geodésicas**, Curitiba, 16, p. 210-231, 2010.

ANDRESEN, T. et al. Análise orientada a objetos de dados de sensoriamento remoto para a obtenção de parâmetros aquáticos/terrestres. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados**. 2. ed.- São Paulo: Oficina de textos. 2007. cap. 22.

ANJOS, C. S. dos.; ALMEIDA, C. M. de. GALVÃO, L. S. Identificação de materiais urbanos por meio de métodos inovadores de classificação de imagens. In: Simpósio Brasileiro de Sensoriamento Remoto, 17., 2015, João Pessoa-PB. **Anais...** São José dos Campos: INPE, p. 4377, 2015.

ANTUNES, D. A.; WIGGERS, K. L.; RIBEIRO, S. R. A. Classificação Orientada a Objetos pelo algoritmo Support Vector Machine de Imagem RapidEye. In: Simpósio Regional de Geoprocessamento e Sensoriamento Remoto – GEONORDESTE. **Anais...** 2014, Aracaju, Brasil, nov., 2014.

ARAÚJO, E. H. G. **Análise multi-temporal de cenas do satélite QuickBird usando um novo paradigma de classificação de imagens e inferências espaciais.** Estudo de caso: Belo Horizonte, MG. 2006. 175 p. (INPE-13956-TDI/1062). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006.

ASNER, G.P. et al. High-resolution Forest carbon stocks and emissions in the Amazon. **Proceedings of the National Academy of Sciences.** 107:16738, 2010.

AYMA, V. A. **Avaliação de Métodos de Otimização aplicados no Modelamento da Estrutura dos Vasos Sanguíneos.** 2013. 77p. Dissertação (Mestrado em Engenharia Elétrica) – Pontífca Universidade Católica do Rio de Janeiro – PUC-Rio. Programa de Pós-Graduação em Engenharia Elétrica, 2013.

BAATZ, M., SCHÄPE, M. **Multiresolution segmentation** — An optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informations- Verarbeitung XII.* Wichmann Verlag, Karlsruhe, p. 12–23, 2000.

BAATZ, M.; MIMLER, M. Objetos de imagens iniciais como suporte para a extração de objetos de interesse. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados.** 2. ed.- São Paulo: Oficina de textos. 2007. cap. 16.

BARBELLA, D. et al. Understanding Support Vector Machine Classifications: A Local Approach. In: *The International Conference on Data Mining (DMIN '09)*, 2009. **Proceedings...** Editors: Robert Stahlbock, Sven F. Crone, and Stefan Lessmann. CSREA Press, p. 305-311, 2009.

BARBIER, N. et al. Linking canopy images to forest structural parameters: potential of a modeling Framework. **Annals of Forest Sciences**, DOI 10.1007/s13595-011-0116-9, 2011.

BHANU B.; LEE S.; MING, J. Adaptive image segmentation using a genetic algorithm. **IEEE Transactions Systems, Man, and Cybernetics**, v. 25, p. 1543-1567, 1995.

BLASCHKE, T., Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, 2009.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, 65, p. 2-16, 2010.

BLASCHKE, T. Object based image analysis: A new paradigm in Remote Sensing? In: Annual Conference Baltimore, Maryland, mar., 2013. **American Society of Photogrammetry and Remote Sensing**. p. 24-28, 2013.

BLASCHKE, T., LANG, S., HAY, G.J. (Eds.). **Object Based Image Analysis**. Heidelberg, Berlin, New York: Springer. 2008. 817 p.

BLASCHKE, T.; STROBL, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. **GIS-Zeitschrift für Geoinformationssysteme**, 14(6), p. 12-17. 2001.

BONESSO, D. **Estimação dos Parâmetros do Kernel em um Classificador SVM na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse**. 108 p. 2013. Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

BONESSO, D.; HAERTEL, V. **Árvore Binária SVM Otimizada na Classificação de Imagem Hiperespectral**. In: Simpósio Brasileiro de Sensoriamento Remoto - SBSR, 16., 2013, Foz do Iguaçu, PR. **Anais...**São José dos Campos: INPE, 2013.

BOYD, D.S.; DANSON, F.M., 2005, Satellite remote sensing of forest resources: three decades of research development. **Progress in Physical Geography**, 29 (1), 1–26. Disponível em: <http://web.ics.purdue.edu/~shao/publications/Forest_RS2005.pdf>. Acesso em: 22 de out. de 2014.

BRASIL. Decreto n. 89.817, de 20 de junho de 1984. Dispõe sobre as instruções reguladoras das normas técnicas da cartografia nacional. **Diário Oficial da União** em 22 de jun. de 1984.

BRASIL. Lei n. 9.605, de 12 de fevereiro de 1998. Dispõe sobre as sanções penais e administrativas derivadas de condutas e atividades lesivas ao meio ambiente, e dá outras providências. Publicada no **Diário Oficial da União** em 12 de fev. de 1998.

BRASIL. Lei n. 11.428, de 22 de dezembro de 2006. Dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica, e dá outras providências. Publicada no **Diário Oficial da União** Seção 1 em 26 de dez. 2006. p. 1.

BRASIL. Decreto n. 6.660, de 21 de novembro de 2008. Regulamenta dispositivos da Lei nº 11.428, de 22 de dezembro de 2006, que dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica. Publicado no **Diário Oficial da União** em 21 de nov. de 2008.

BRASIL. Conselho Nacional do Meio Ambiente. 1993. Resolução CONAMA nº 10/93, de 1 de outubro de 1993. Publicada no **Diário Oficial da União** em 3 de nov. 1993, n. 209.

BRASIL. Conselho Nacional do Meio Ambiente. 1994. Resolução CONAMA nº 04/94, de 4 de maio de 1994. Publicada no **Diário Oficial da União** em 17 jun. 1994, n. 114.

BRASIL. Conselho Nacional do Meio Ambiente. 2007. Resolução CONAMA nº 388/07, de 23 de fevereiro de 2007. Publicada no **Diário Oficial da União** em 26 de fev. 2007, n. 114, p. 8877-8878.

BREIMAN, L. et al. **Classification and Regression Trees**. Belmont, CA: Wadsworth. 1984.

BÜCKNER, J. et al. GeoAIDA – A knowledge based automatic image data analyzer for remote sensing data. In: International Congress on Computational Intelligence: Methods & Applications, 2. 2001, Bangor, Wales, United Kingdom. **Proceedings...** Millet, Alberta: ICSC, p.19-22, 2001.

BURGES, C. J. C. **A Tutorial on Support Vector Machines for Pattern Recognition**. Data Mining and Knowledge Discovery, 2, p. 121-167, 1998.

CADENA, G. T. **Classificação dos tipos de pavimentos das vias urbanas a partir de imagem de alta resolução espacial por meio de análise orientada a objeto**. 2011. 114p.

Dissertação (Mestrado em Ciências Cartográficas)-

Universidade Estadual Paulista, Presidente Prudente, SP, 2011.

CAMARGO, F. F. et al. Geomorphological Mapping Using Object-Based Analysis and ASTER DEM in the Paraíba do Sul Valley, Brazil. **International Journal of Remote Sensing**, (Print), v. 30, p. 6613-6620, 2009.

CAMILO, O. C.; SILVA, J. C. Mineração de dados: Conceitos, Tarefas, Métodos e Ferramentas. **Relatório Técnico**, 001-09, ago., 2009, 28 p.

CARLEER, A. P.; DEBEIR, O.; WOLFF, E. Assesment of very high spatial resolution satellite image segmentations. **Photogrammetric Engineering and Remote Sensing**, v. 71, n. 11, p. 1285–1294, 2005.

CAROLINA, L. A.; CARVALHO, C. A. D. Evolutionary tuning of SVM parameter values in multiclass problems. [S.l.], p.3326–3334, 2008.

CARVALHO JUNIOR, O. A. de et al . Mapeamento da vegetação na floresta atlântica usando classificador de árvore de decisão para integrar dados de sensoriamento remoto e modelo digital de terreno. **Rev. Bras. Geof.**, São Paulo, v. 26, n. 3, p. 331-345, set., 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102261X2008000300007&lng=en&nrm=iso>. Acesso em: 27 jun. 2015.

CARVALHO, M. V. A. **Classificação da cobertura e uso do solo com imagens Worldview-2 utilizando mineração de dados e análise baseada em objetos.** 2011. 201 p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.

CASTRO, D. **Práticas para restauração da mata ciliar.** In: CASTRO, D.; MELLO, R. S. P.; POESTER, G. C. (Org.) Porto Alegre: Catarse– Coletivo de Comunicação, 2012, 60 p.

CAZES, T. B.; FEITOSA, R. Q. **Interpretação Baseada em Conhecimento de Imagens de Sensores Remotos de Alta Resolução.** 2005, 94 p. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

CENTENO, J. A. S. **Sensoriamento Remoto e Processamento de Imagens Digitais.**

Curitiba: Curso de Pós-graduação em Ciências Geodésicas, Universidade Federal do Paraná, 2004. 219 p.

CHAVES, A.; VELLASCO, M.; TANSCHKEIT, R. Extração de regras fuzzy de máquinas vetor suporte para a classificação em múltiplas classes. In: Simpósio Brasileiro de Automação Inteligente, 8., 2007, Florianópolis, SC. **Anais...** Florianópolis: DAS, 2007.

CHENG, Y. Mean shift, mode seeking, and clustering, **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 17, n. 8, p.790–799, 1995.

CHOU, T-S et al. Correlation-Based Feature Selection for Intrusion Detection. In: Design Military Communications Conference, 2007. **MILCOM**, IEEE 29-31, oct. 2007. ISBN: 978-1-4244-1513-7.

CHRISTOPHE, E.; INGLADA, J. Open Source Remote Sensing: Increasing the Usability of Cutting-Edge Algorithms. **IEEE Geoscience and Remote Sensing Society Newsletter**, mar., p. 9-15, 2009.

CHUBEY, M. S.; FRANKLIN, S.E.; WULDER, M.A. Object-based analysis of IKONOS-2 imagery for extraction of forest inventory parameters. **Photogrammetric Engineering & Remote Sensing**, 72 (4), p. 383–394, 2006.

CINTRA, D. P. **Classificação de estágios sucessionais florestais por meio de imagens de alta resolução (Ikonos) no Parque Estadual da Pedra Branca, RJ.** 2007. 87 p. Dissertação (Mestrado em Ciências Ambientais e Florestais)- Instituto de Florestas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2007.

CINTRA, D. P. et al.. PIMAR Project - Monitoring the atlantic rainforest remnants and the urban growth of the Rio de Janeiro city (Brazil) through remote sensing. In: **GEOBIA 2010**, 2010, Ghent, Belgium. **Proceedings** of the **GEOBIA 2010**, 2010.

CLARK, M. L. et al. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. **Remote Sensing of Environment**, v. 114, p. 2816-2832, 2010.

CLEVE, C. et al. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. **Computers, Environment and Urban Systems**, 32, p. 317–326. 2008.

COMANICIU, D.; MEER, P. Mean Shift: a robust approach toward feature space analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.24, n.5, p. 603–619, 2002.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publishers, 1999, 136 p.

CORTÊS, S. C.; PORCARO, R. M.; LIFSHITZ, S. Mineração de Dados: Funcionalidades, técnicas e abordagens. PUC-Rio Inf MCC10-02. mai. 2002. Disponível em: < ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf>. Acesso em: 22 de mar. de 2015.

COSTA, G. A. O. P. et al. INTERIMAGE: uma plataforma cognitiva open source para a interpretação automática de imagens digitais. RBC. **Revista Brasileira de Cartografia**, Rio de Janeiro, v. 60, p. 331-337, 2008.

COSTA, G. A. O. P. et al. Knowledge-based interpretation of remote sensing data with the InterIMAGE system: major characteristics and recent developments. In: **Proceedings of the 3rd GEOBIA**, 2010, Ghent, Belgium, 2010.

COSTA, J. S. **Aplicação de Métodos de Sensoriamento Remoto para mapeamento da Área de Preservação Permanente do lago da Usina Itaipu**. 2012, 40 p. Trabalho de Conclusão de Curso (Graduação em Geografia) - Universidade Estadual de Londrina, Londrina, 2012.

COSTA, W. S.; FONSECA, L. M. G.; KÖRTING, T. S. Classificação de pastagens cultivadas e formações campestres nativas no Cerrado brasileiro a partir da análise de séries temporais extraídas de índices EVI do sensor MODIS. In: Simpósio Brasileiro de Sensoriamento Remoto, 17., 2015, João Pessoa-PB. **Anais...** São José dos Campos: INPE, p.1516-1523.

COUTERON, P. et al. **Linking Remote-Sensing Information to Tropical Forest Structure: The Crucial Role of Modelling.** Apr., 2012. Disponível em <http://earthzine.org/2012/04/23/linking-remote-sensing-information-to-tropical-forest-structure-the-crucial-role-of-modelling/> Acesso em 25 de nov. de 2014.

COUTINHO, A. C. **Segmentação e classificação de imagens Landsat-TM para o mapeamento dos usos da terra na região de Campinas, SP.** 1997. 150 p. Dissertação (Mestrado em Ciências) – Curso de Pós-Graduação em Ciências, Departamento de Ecologia Geral do Instituto de Biociências, Universidade de São Paulo, São Paulo, 1997.

DEFINIENS, 2009. Image analysis with Definiens Cognition Network Technology.

DERIVAUX, S. et al. Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation, **Pattern Recognition Letters**, v. 31, n. 15, p. 2364-2374, 2010.

DEVADAS, R.; DENHAM, R. J.; PRINGLE, M. Support Vector Machine classification of object-based data for crop mapping, using multi-temporal Landsat imagery. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XXXIX-B7, 2012. In: XXII ISPRS Congress, Melbourne, Australia, aug./sep., p. 185-190, 2012.

DI GREGORIO, A. D.; O'BRIEN, D. Overview of Land-Cover Classifications and Their Interoperability. In: GIRI, C. P. **Remote Sensing of Land Use and Land Cover: Principles and Applications** – Taylor & Francis Series in Remote Sensing Applications: Indiana State University Terre Haute, Indiana, U.S.A. 2012, 413 p.

DISPERATI, A. A.; OLIVEIRA FILHO, P.C. Delineamento de copas de árvores em fotografias aéreas de alta resolução, através de suas sombras periféricas: estudo de caso na floresta ombrófila mista. [Versão eletrônica] **Revista Arvore**, Viçosa, 29, (2), p. 195- 202, 2005. Disponível em: <<http://www.scielo.br>>. Acesso em 18 de jan. de 2015.

DIXON, B.; CANDADE, N. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?. **International Journal of Remote Sensing**, v. 29, n. 4, p. 1185–1206, 2008.

DLUGOSZ, F. L. et al. Índice para avaliação de segmentação de imagens. **FLORESTA**, Curitiba, PR, v. 39, n. 1, p. 131-143, jan./mar, 2009.

DURIC, N.; PEHANI, P.; OSTIR, K. Application of In-Segment Multiple Sampling in Object-Based Classification **Remote Sens.** 2014, 6, p. 12138-12165. Disponível em: <www.mdpi.com/journal/remotesensing>. Acesso em: 21 de jan. de 2015.

DURIEUX, A. et al. Object-based land cover classification using a preliminary stratified semantic pixel labeling obtained from a fuzzy spectral prior knowledge classifier. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2009, Natal, PA. **Anais...** São José dos Campos: INPE, p. 137-143, 2009.

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. **Sensoriamento Remoto e SIG Avançados.** 2. ed.- São Paulo: Oficina de textos. 2007. cap. 2.

ENGEMAP Geoinformação. Relatório de produção final - edital de concorrência pública n. 0010/2009. Florianópolis SC, 218 p., 2012.

ENVI, 2009. ENVI Feature Extraction Module. Disponível em: <https://www.exelisvis.com/portals/0/pdfs/envi/ENVI46_FxBrochure.pdf>. Acesso em 12 de mai. de 2015.

ESPINDOLA, G. M. de. Uso de algoritmos genéticos no ajuste de parâmetros da segmentação de imagens. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2009, Natal, PA. **Anais...**São José dos Campos: INPE, p. 6861-6868, 2009.

FAXINA, T. C. **Dilemas da regularização fundiária amigável no Parque Nacional de São Joaquim**: Um estudo de caso – a valorização de áreas silvestres. 2014. Dissertação (Mestrado em Engenharia Florestal)- Universidade do Estado de Santa Catarina, Lages, 2014.

FEITOSA, R. Q. et al. A genetic approach for the automatic adaptation of segmentation parameters. In: International Conference on Object-based Image Analysis – ISPRS **Proceedings...**, v. 36, n. 4/C42, 2006.

FEITOSA, R. Q.; COSTA, G. A. O. P., CAZES, T. B. Ajuste Automático de Parâmetros de Segmentação. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados**. 2. ed.- São Paulo: Oficina de textos. 2007. cap. 12.

FEITOSA, R. Q. et al. Uma avaliação de métodos genéticos para ajuste de parâmetros de segmentação. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2009, Natal, PA. **Anais...**São José dos Campos: INPE, p. 6875-6882, 2009.

FERNANDES, P. J. F.; GIRÃO, R. S.; SILVA, L. C. Técnicas de processamento digital de imagens e mineração de dados aplicadas ao mapeamento de fragmentos florestais a partir de imagens ETM+. **Caderno de Estudos Geoambientais**, v. 4, n.1, Niterói-RJ. p. 23-36. 2013.

FERREIRA, R. **Uma abordagem multiescalar, multicritério para a segmentação de imagens**. 2011. 113p. Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

FLORENZANO, T. G. **Imagens de satélite para estudos ambientais**. São Paulo, Oficina de Textos, 2002.

FONSECA, J.M.M.R. **Indução de Árvores de Decisão**. 151 f. Dissertação (Mestrado em Departamento de Informática) - Universidade Nova de Lisboa, Portugal. 1994.

FONSECA, L. G. M. et al. **Apostila de Processamento Digital de Imagens**. São José dos Campos: INPE. jun., 2000.

FOODY, G. M.; MATHUR, A. A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. **IEEE Transactions on Geoscience and Remote Sensing**, v.42, n.6, 2004.

FOURIE, C.; NIEKERK, A. V.; MUCINA, L. Semi-automated segment generation for geographic novelty detection using edge and area metrics. **South African Journal of Geomatics**, v. 1, n. 2, aug. 2012.

FRANCISCO, C. N.; ALMEIDA, C. M. Mineração de dados e análise de imagens baseada em objeto aplicadas ao mapeamento de cobertura da terra. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, PR. **Anais...** São José dos Campos: INPE, p. 2282-2289, 2013.

FRANCISCO, C. N.; ALMEIDA, C. M. de. Avaliação de Desempenho de Atributos Estatísticos e Texturais em uma Classificação de Cobertura da Terra Baseada em Objeto. **Bol. Ciênc. Geod.**, sec. Artigos, Curitiba, v. 18, no 2, p.302-326, abr/jun, 2012.

FREITAS, S. R.; SHIMABUKURO, Y. E. **New Research on Forest Ecology**. Chapter Diagnosing Tropical Forests through Remote Sensing, p. 101–128. New York, NY: Nova Science Publishers, 2007.

FREDRICH, C. M. B.; FEITOSA R. Q. Automatic adaptation of segmentation parameters applied to inhomogeneous objects detection. In: Conference on geographic object-based image analysis, 2008, Calgary, Canada. **Proceedings...**Calgary, 2008.

FU, X. et al. Extracting the Knowledge Embedded in Support Vector Machines, In: International Joint Conference on Neural Networks, Budapest, Hungary, p. 291-296, 2004.

FUKUNAGA, K.; HOSTETLER, L. D. The Estimation of The Gradient of a Density Function, with Applications in Pattern Recognition. **IEEE Transactions on Information Theory**, [S.l.], v.21, p.32–40, 1975.

FUNDAÇÃO SOS MATA ATLÂNTICA E INPE. **Atlas dos Remanescentes Florestais da Mata Atlântica: Período 2013-2014**. Relatório Técnico, São Paulo: SOS Mata Atlântica e INPE, 2015.

FURLANETTI, Thobias. Apresentação do Aerolevante Fotogramétrico do Estado de Santa Catarina. Diretoria de Recursos Hídricos (DIRH) Secretaria de Estado do Desenvolvimento Econômico e Sustentável. Disponível em: <http://www.concar.ibge.gov.br/plenarias/17a_PI_CONCAR_21set2011_Plano_SC.pdf>. Acesso em 15 de abr. de 2015.

FURTADO, L. F. A.; FRANCISCO, C. N.; ALMEIDA, C. M. Análise de imagem baseada em objeto para classificação das fisionomias da vegetação em imagens de alta resolução espacial. **Geociênc.**, São Paulo, v. 32, n. 3, 2013. Disponível em: <http://papego.igc.usp.br/scielo.php?script=sci_arttext&pid=S010190822013000300004&lng=pt&nrm=iso>. Acesso em: 11 de jun. de 2015.

GAO, Y.; MAS, J. F. A comparison of the performance of pixel based and object based classification over images with various spatial resolutions. **Online Journal of Earth Sciences**, v. 2, p. 27-35, 2008.

GAO, Y., MAS, J.F., KERLE, N., NAVARRETE PACHECO, J.A. Optimal region growing segmentation and its effect on classification accuracy. **Int. J. Remote Sens.** 32 (13), p. 3747–3763, 2011.

GEODMA 0.2.2. São José dos Campos, SP: INPE, 2011.
Disponível em: <<http://www.dpi.inpe.br/geodma>> Acesso em: 30 de abr. de 2015.

GIASSON, E. et al. Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil. **Ciência Rural**, Santa Maria, v.43, n.11, p.1967-1973, nov, 2013.

GIRI, C. P. **Remote Sensing of Land Use and Land Cover: Principles and Applications** – Taylor & Francis Series in Remote Sensing Applications: Indiana State University Terre Haute, Indiana, U.S.A. 2012. 413 p.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining** – um guia prático. Rio de Janeiro: Elsevier, 2005. 257 p. ISBN (85-352-1877-7).

GONÇALVES, E. C. **Extração de Árvores de Decisão com a Ferramenta de Data Mining Weka**. 2007. Disponível em: <<http://www.devmedia.com.br/extracao-de-arvores-de-decisao-com-a-ferramenta-de-data-mining-weka/3388#>>. Acesso em 22 de jan. de 2015.

GONZALEZ, R.C., WOODS, R.E. **Digital Image Processing**. 2. ed. New Jersey: Prentice Hall, 2001. ISBN 0-201-18075-8.

GRINZONNET, M.; INGLADA, J. Monteverdi - Remote Sensing Software from Educational to Operational Context. In: EARSeL Symposium, 2010, p. 749.

HALL, M. A. **Correlation-based feature selection for machine learning**. PhD thesis, (Department of Computer Science) - University of Waikato, Hamilto, New Zealand, 1998.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2006. 743p.

HAN, T. S.; KOBAYASHI, K. **Mathematics of Information and Coding**. Tokyo, Japan: Softcover. 2002. 286 p. ISBN 978-0-8218-4256-0.

HANSEN, M. C. Classification trees and mixed pixels training data. In: GIRI, C. P. **Remote Sensing of Land Use and Land Cover: Principles and Applications** – Taylor & Francis Series in Remote Sensing Applications: Indiana State University Terre Haute, Indiana, U.S.A. 2012, 413 p.

HAY, G. J. et al. An automated object-based approach for the multiscale image segmentation of forest scenes. **International Journal of Applied Earth Observation and Geoinformation**, 7 (4), p. 339–359, 2005.

HAY, G.J., CASTILLA, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: BLASCHKE, T., LANG, S., HAY, G. (Eds.), **Object Based Image Analysis**. Heidelberg, Berlin, New York: Springer, p. 93–112, 2008.

HEGDE, G. et al. Urban Land Cover Classification using Hyperspectral Data. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XL-8, 2014, ISPRS Technical Commission VIII Symposium, Hyderabad, India, dec., 2014.

HEUMANN, B. H. An Object-Based Classification of Mangroves Using a Hybrid Decision Tree—Support Vector Machine Approach. **Remote Sens.**, 3(11), p. 2440-2460, 2011.

HSU, C-W; CHANG, C-C; LIN, C-J. **A Practical Guide to Support Vector Classification**. Taiwan: Department of Computer Science, 2003. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>>. Acesso em: 12 de mai. de 2015.

HUANG, C.; DAVIS, L. S.; TOENSHEND, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, v. 23, n. 4, p. 725-749, 2002.

HUANG, C.-L.; WANG, C.-J. A GA-based feature selection and parameters optimization for support vector machines. **Expert Systems with Applications**, 31(2), p. 231-240, 2006.

INSTITUTO BRASILEIRO DO MEIO AMBIENTE E DOS RECURSOS NATURAIS RENOVÁVEIS - IBAMA. Gestão Biorregional da Costa Catarinense. Documento Básico, versão 2.1, Brasília, 2001.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA- IBGE. **Informações das cidades do Brasil**. 2014. Disponível em: <www.cidades.ibge.gov.br/>. Acesso em: 12 de abr. de 2015.

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE- ICMBIO. 2012. **Parque Nacional de São Joaquim**. Disponível em: <<http://www.icmbio.gov.br/portal/o-que-fazemos/visitacao/ucs-abertas-a-visitacao/196-parque-nacional-de-sao-joaquim>>. Acesso em 22 de julho de 2014.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS- INPE. **Introdução ao Sensoriamento Remoto**. INPE: São José dos Campos- SP. 2001.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS- INPE. **Processamento Digital de Imagens**. São José dos Campos- SP. 2011. Disponível em: <<http://www.dpi.inpe.br/terraview/docs/tutorial/Aula16.pdf>>. Acesso em: 15 de jan. de 2015.

INTERIMAGE 1.43. InterIMAGE - Interpreting Images Freely. Disponível em: <<http://www.lvc.ele.puc.rio.br>> Acesso em: 30 de out. de 2014.

JANOTH, J. et al. Procedimentos Baseados em Segmentação para a Análise de Mudanças e Classificação Florestais com Dados de Satélite de Alta Resolução. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados**. 2. ed.- São Paulo: Oficina de textos. 2007. cap. 9.

JASTER, C. B. **A estrutura como indicadora do nível de desenvolvimento sucessional de comunidades arbóreas da restinga** – Uma proposta metodológica. 2002. Tese (Doutorado em Ciências Florestais)- Universidade Federal do Paraná, Curitiba, 2002.

JENSEN, J. R. **Introductory Digital Image Processing**. 3rd ed. New Jersey: Pearson Prentice Hall, 2005.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres**. Tradução: Epiphanio et al. São José dos Campos: Parêntese, 2009.

JOHANSEN, K.; PHINN S. Mapping structural parameters and species composition of riparian using ikonos and landsat etm+ data in australian tropical savannahs. **Photogrammetric Engineering and Remote Sensing**, 72(1), p. 71–80, 2006.

KAGEYAMA, P. Y.; BRITO, M. A.; BAPTISTON, I. C. Estudo do mecanismo de reprodução de espécies da mata natural. In: KAGEYAMA, P. Y. (Coord.). **Estudo para implantação de matas ciliares de proteção na bacia hidrográfica do Passa Cinco**, Piracicaba, SP. Piracicaba: DAEE/USP/FEALQ, 1986. 236 p.

KAI, L.; MULLER, J.P. Segmentation satellite imagery: a region growing scheme. In: International Geoscience Remote Sensing Symposium. **Proceedings...** Helsink. v.2, p. 1075-1078, 1991.

KALKANA, B. et al. Comparasion of Support Vector Machine and Object Based Classification Methods for Coastline Detection. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, Volume XL-7/W2, 2013 ISPRS2013-SSG, Antalya, Turkey, nov., 2013.

KAREGOWDA, A. G. MANJUNATH, A. S.; JAYARAM, M. A. Comparative of attribute selection using gain ratio and correlation based feature selection. **International Journal of Information Technology and Knowledge Management**, jul./dec., v. 2, n. 2, p. 271-277, 2010.

KLEIN, R. M. **Mapa fitogeográfico do Estado de Santa Catarina**. Itajaí: Herbário Barbosa Rodrigues; Florianópolis: Universidade Federal de Santa Catarina, 1978.

KOHAVI, R.; JOHN, G. Wrappers for feature subset selection. **Artificial intelligence, special issue on relevance**, v. 97, n.2, p. 273–324, 1997.

KOK, R. de et al. Design de projeto modular na Análise Orientada a Objetos. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados**. 2. ed.- São Paulo: Oficina de textos. 2007. cap. 4.

KÖPPEN, W.; GEIGER, R. **Klimate der Erde**. Gotha: Verlag Justus Perthes. 1928. Wall-map 150cmx200cm.

KÖRTING, T. S. et al. GeoDMA - a novel system for spatial Data mining. In: Data Mining Workshops, 2008, Pisa, Italy. ICDMW 08. IEEE International Conference, 2008, Pisa, Italy. **Proceedings...** Pisa: IEEE, 2008.

KÖRTING, T. S. et al. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal, PA. **Anais...**São José dos Campos: INPE, p. 7813-7820, 2009.

KÖRTING, T. S. **GeoDMA**: a toolbox integrating data mining with object-based and multi-temporal analysis of satellite remotely sensed imagery. 2012. 119 p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012.

KÖRTING, T. S., FONSECA, L. M. G.; CÂMARA, G. Interpreting Images with GeoDMA. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XXXVIII-4/C7, 2013.

KUX, H. J. H. et al. Classificação da cobertura do solo urbano usando imagens ópticas de altíssima resolução e o sistema InterIMAGE baseado em conhecimento. In: Simpósio Brasileiro de Sensoriamento Remoto, 15., 2011, Curitiba, PR. **Anais...**São José dos Campos: INPE, p. 0822, 2011.

KUX, H. J. H.; MENEGHETTI, G. T. Imagens de altíssima resolução espacial, GEOBIA e Mineração de Dados para o mapeamento detalhado de cobertura da terra em ambientes costeiros, área-teste: Raposa (MA). In: Simpósio Brasileiro de Sensoriamento Remoto, 17., 2015, João Pessoa, PB. **Anais...**São José dos Campos: INPE, p. 0247-0254, 2015.

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

LANG, S.; TIEDE, D. Definiens developer. **GIS Business** 9, p. 34–37, 2007.

LEONARDI, F. et al. Genetic algorithms and data mining applied to optical orbital and LiDAR data for object-based classification of urban land cover. In: International Conference on Geographic Object-Based Image Analysis, 4. (GEOBIA)., 2012, Rio de Janeiro. **Proceedings...** São José dos Campos: INPE, p. 649-654, 2012.

LEUKERT, K. Utilização de dados de SIG para a obtenção de objetos. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados**. 2. ed.- São Paulo: Oficina de textos. 2007. cap. 13.

LI, M. et al. A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information. **European Journal of Remote Sensing**, 47, p. 389-411, 2014.

LI, S. et al. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. **Knowledge-Based Syst.**, 24(1), p. 40-48, 2011.

LI, X.; SHAO, G. Object-Based Land-Cover Mapping with High Resolution Aerial Photography at a County Scale in Midwestern USA. **Remote Sens.**, 6(11), p. 11372-11390, 2014. Disponível em: <http://www.mdpi.com/journal/remotesensing/special_issues/g_eobia> Acesso em 18 de jan. de 2014.

LIAU, Y-T. Hierarchical Segmentation Framework for Identifying Natural Vegetation: A Case Study of the Tehachapi Mountains, California. **Remote Sens.**, 6, p. 7276-7302, 2014. Disponível em: <www.mdpi.com/journal/remotesensing> Acesso em: 21 de fev. de 2015.

LILLESEND, T.M.; KIEFER, R.W. **Remote Sensing and Image Interpretation**. 4th ed., New York: Wiley, 1999.

LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. W. **Remote sensing and image interpretation**. New York: John Wiley and Sons, 2004. 763 p.

LIMA, C. A. F. **Correções geométricas para a utilização de imagens em perícias criminais ambientais**. 144 p. 2013.

Dissertação (Mestrado profissional em Perícia Criminal) – Universidade Federal de Santa Catarina, Florianópolis, 2013.

LIU, H; YU, L. Toward integrating feature selection algorithms for classification and clustering. **Knowledge and Data Engineering**, IEEE Transactions on, v. 17, p. 491-502, 2005. ISSN 1041-4347.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução às Support Vectors Machines - Tutorial. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43-67, 2007.

LÜBKER T.; SCHAAB G. Optimization of parameter settings for multilevel image segmentation in GEOBIA. In: High resolution earth imaging for geospatial information workshop held, 2009, Hanover, Germany.

MATHUR, A.; FOODY, G. M. Crop classification by support vector machine with intelligently selected training data for an operational application. **International Journal of Remote Sensing**, 29, p. 2227–2240. 2008.

MCCUE, C. Data Mining and Predictive Analysis - Intelligence Gathering and Crime Analysis. **Elsevier**, 2007.

MENEGHETTI, G. T.; RODRIGUES, C. de S. Object oriented classification in urban areas using the GeoDMA plug-in. In: GEOBIA, 4., 2012, Rio de Janeiro. **Proceedings...** p.412, 2012.

MENEGHETTI, G. T. **Classificação da cobertura da terra do município de Raposa (MA) com imagens Worldview-2 utilizando o aplicativo InterIMAGE e mineração de dados.** 2013. 92 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.

MENESES, P. R.; ALMEIDA, T. **Introdução ao Processamento de Imagens de Sensoriamento Remoto.** Brasília: UnB- Embrapa Cerrados. 2012. 266 p.

MERSCHMANN, L. H. C. **Classificação probabilística baseada em análise de padrões.** 2007. 117 p. Tese (Doutorado em Otimização Combinatória) – Universidade Federal Fluminense, Niterói, 2007.

MINISTÉRIO DO MEIO AMBIENTE- MMA. Bioma Mata Atlântica. 2010. Disponível em:
<<http://www.mma.gov.br/biomas/mata-atlantica>>. Acesso em 14 de fev. de 2014.

MOUNTRAKIS, G.; JUNGHO IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, p. 247-259, 2010.

MOORE, G.; BERGERON, C.; BENNET, K. Model selection for primal SVM. **Machine Learning**, 85, p. 175-208, 2011.

NASCIMENTO, A. F. et al. Classificação da Cobertura da Terra, Utilizando os Programas Livres: InterIMAGE, WEKA e QuantumGIS. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, PR. **Anais...**São José dos Campos: INPE, p. 2258-2265, 2013.

NASCIMENTO, P. S. R. **Avaliação de técnicas de segmentação e classificação por regiões em imagens Landsat-TM visando o mapeamento de unidades de paisagem na Amazônia.** 1997. 120 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.

NASCIMENTO, R. F. F. et al. O algoritmo Support Vector Machine (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal, PA. **Anais...**São José dos Campos: INPE, p. 2079-2086, 2009.

NASCIMENTO, R. G. M. et al. Modelo de projeção por classe diamétrica para florestas nativas: enfoque na função probabilística de Weibull. **Pesquisa Florestal Brasileira.** Brazilian Journal of Forestry Research. doi: 10.4336/2012.pfb.32.70.93, 2012.

NAVE, A. G. **Determinação de unidades ecológicas num fragmento de floresta nativa, com auxílio de sensoriamento remoto.** 1999. 167p. Dissertação (Mestrado em Ciências Florestais) - Escola Superior de agricultura Luis de Queiroz, Piracicaba, 1999.

NELDER, J. A.; MEAD, R. Simplex Method for Function Minimization. **The Computer Journal**, v. 7, n. 4, p. 308-313, 1965.

NITZE, I.; SCHULTHESS, U.; ASCHE, H. Comparison of machine learning algorithms Random Forest, Artificial Neural Network e Support Vector Machine to Maximum Likelihood for supervised crop type classification. In: GEOBIA, 4., 2012, Rio de Janeiro. **Proceedings...**, p. 035, 2012.

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

NOVACK, T.; RIBEIRO, B. M. G.; KUX, H. J. H. Análise dos dados do satélite WorldView-2 para a discriminação de alvos urbanos semelhantes com base em algoritmos de seleção de atributos. In: Simpósio Brasileiro de Sensoriamento, 15., 2011, Curitiba. **Anais...** São José dos Campos: INPE, p. 7815-7821, 2011.

NOVO, E. M. L. M. **Sensoriamento Remoto: princípios e aplicações.** 3 ed. São Paulo: Edgard Blücher, 2008.

NEUBERT, M.; HEROLD, H.; MEINEL, G. Evaluation of remote sensing image segmentation quality – further results and concepts. 2006. Disponível em: <http://www2.ioer.de/recherche/pdf/2006_neubert_obia.pdf>. Acesso em: 20 de jul. de 2013.

NUÑEZ, H.; ANGULO, C.; CARALÀ, A. Rule Extraction From support vectors machines. In: European Symposium on Artificial Neural Networks, 10., Bruges, Belgium, p. 107-112, 2002.

ODUM, E. P. **Ecologia.** Rio de Janeiro: Guanabara, 1988. 434 p.

ORFEO ToolBox. **Open Source processing of remote sensing images.** Disponível em: <<https://www.orfeo-toolbox.org/>>. Acesso em: 13 de mar. de 2014.

OTB Software Guide. Disponível em: <<https://www.orfeo-toolbox.org/SoftwareGuide/index.html>>. Acesso em: 22 de ago. de 2014.

OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation**, 12, Supplement 1, p. 27-31, 2010.

PAL, M ; MATHER, P. M. An assessment of the effectiveness of decision tree methods for land cover classification. **Remote Sens. Environ**, 86, p. 554–565, 2003.

PAL, M.; MATHER, P.M. Support vector machines for classification in remote sensing. **International Journal of Remote Sensing**, v. 26, n. 5, p. 1007-1011, 2005.

PASSO, D. et al. Comparison of the accuracy of classifications generated by interimage and by interimage integrated with data mining. **South-Eastern European Journal of Earth Observation on and Geomatics Issue** v. 3, n. 2S, 2014 93
©Aristotle University of Thessaloniki, Greece Published online May, 2014.

PEREIRA, E. G. S.; RICHTER, M. Classificação da Cobertura da Terra em Angra dos Reis - RJ, a partir de Análise de Imagem Baseada em Objeto e Mineração de Dados. In: Congresso Brasileiro de Geoprocessamento, 4., 2013, Rio de Janeiro. **Anais...** Disponível em: <http://www.cartografia.org.br/cbg/trabalhos/90/15/obia_pereiraegs_richterm_1376400711.pdf>. Acesso em: 20 de ago. de 2014.

PIAZZA, G. A. **Processamento digital de imagens de alta resolução espacial com enfoque na classificação dos estágios sucessionais iniciais da Floresta Ombrófila Densa em Santa Catarina**. 2014. 128 p. Dissertação (Mestrado em Engenharia Ambiental)- Centro de Ciências Tecnológicas, FURB, Blumenau, SC, 2014.

PIAZZA, G. A.; VIBRANS, A. C. Classificação orientada a objetos para mapear florestas secundárias em Santa Catarina, Sul do Brasil. **Revista Brasileira de Cartografia**, n. 66/5, p. 993-1005, 2014. ISSN: 1808-0936.

PIGNALBERI, G.; CUCCHIARA, R.; CINQUE, L.; Levialdi, S. 2003. Tuning range image segmentation by genetic algorithm. *EURASIP Journal on Applied Signal Processing*, n. 8, p. 780–790, 2003.

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 intra-urbano: o caso de São José dos Campos**. 2005. 180 p. (INPE- 14183-TDI/1095). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2005.

PINHO, M. D. et al. Intra-urban Land Cover Classification From High-resolution Images Using the C4.5 Algorithm. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**. v. 37. Part B7. Beijing, 2008.

PIOVEZAN, M. **Obtenção de Variáveis em Ambientes SIG como contribuição para Avaliação de Imóveis Rurais**. 103 p. 2013. Dissertação (Mestrado em Engenharia Civil)- Universidade Federal de Santa Catarina, Florianópolis, 2013.

PONZONI, F. J.; REZENDE, A. C. P. Caracterização espectral de estágios sucessionais de vegetação arbórea secundária arbórea em Altamira (PA), através de dados orbitais. **Revista Árvore**, Viçosa, v. 28, n.4, p. 535-545, 2004.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento Remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2009.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento Remoto da Vegetação**. 2. ed. São Paulo: Oficina de Textos, 2012.

QGIS. A Free and Open Source Geographic Information System. Disponível em <<http://www.qgis.org/en/site/>> Acesso em 10 de jan. de 2014.

QUINLAN, R. **C4.5: programs for machine learning**. San Francisco: Morgan Kaufmann. 1993.

RAISSOUNI, O. B.; RAISSOUNI, N. Support Vector Machines for Object Based Building Extraction in Suburban Area using Very High Resolution Satellite Images, a Case Study: Tetuan, Morocco. **IAES International Journal of Artificial Intelligence (IJ-AI)**, v. 2, n. 1, p. 43-50, mar., 2013.

RESERVA DA BIOSFERA DA MATA ATLÂNTICA (RBMA). Revisão e atualização dos limites e zoneamento da reserva da biosfera da Mata Atlântica em base cartográfica digitalizada. LINO, C., F. (Coord.). Fase VI. São Paulo, 2008.

REDE MATA ATLÂNTICA (RMA) – uma rede pela floresta. CAMPANILI, M.; PROCHNOW, M. (Org.). Brasília: RMA, 2006. 332p. ISBN: 85-99824-01-5.

RIBEIRO, J. R. P. **A detecção remota no inventário florestal: análise das potencialidades da utilização de imagens de satélite**. 2007. 138p. Dissertação (Mestrado em Estatística e Gestão da Informação)- Universidade Nova de Lisboa, Lisboa, 2007. Disponível em:
<http://www.igeo.pt/gdr/pdf/Tese_Jo%E3o_Ribeiro.pdf>
Acesso em: 11 de nov. de 2014.

RIBEIRO, B. M. G.; KUX, H. J. H. Classificação Orientada a Objeto para Mapeamento do Uso do Solo – Métodos de Análise de Expansão Urbana. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2009, Natal, PA. **Anais...** São José dos Campos: INPE, p. 7893- 7900, 2009.

RIBEIRO, B. M. G.; FONSECA, L. M. G.; KUX, H. J. H. Avaliação das imagens WorldView-II para o mapeamento da cobertura do solo urbano utilizando o sistema InterIMAGE. In: Simpósio Brasileiro de Sensoriamento Remoto, 15., 2011, Curitiba, PR, Brasil. **Anais...** São José dos Campos: INPE, p.0722, 2011.

RIBEIRO, B. M. G.; FONSECA, L. M. G. Urban Land Cover Classification using WorldView-2 Images and C4.5 Algorithm. In: JURSE, 2013, São Paulo. **Proceedings...** São Paulo, Brasil.

RODRIGUES, R. R. A sucessão florestal. In: MORELLATO, P. C., LEITÃO FILHO, H. F. (Orgs.). **Ecologia e preservação de uma floresta tropical urbana: Reserva de Santa Genebra**. Campinas: UNICAMP, p. 30-36, 1995.

RODRIGUES, T. C. S. **Classificação da cobertura e do uso da terra com imagens WorldView-2 de setores norte da Ilha do Maranhão por meio do aplicativo InterIMAGE e de mineração de dados.** 2014. 87 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

ROSÁRIO, C. B.; TEMPONI, L. M.; SANTOS, P. R. A. dos. Análise Orientada a Objeto no Mapeamento dos Estágios Sucessionais da Vegetação na Escala 1:25.000 – um Estudo de Caso da Reserva União, RJ. In: Simpósio Brasileiro de Sensoriamento Remoto, 14., 2009, Natal, PA. **Anais...** São José dos Campos: INPE, p. 7119-7126, 2009.

ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric Engineering & Remote Sensing**, v. 52, n. 2, p. 223-227, 1986.

ROUSE, J. W. et al. **Monitoring vegetation systems in the Great Plains with ERTS.** In: Third ERTS Symposium, NASA SP-351 I, 309- 317, 1973.

SABA, F.; VALADANZOUJ, M. J.; MOKHTARZA, M. de. The optimization of multi resolution segmentation of remotely sensed data using genetic algorithm. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 40-1/W3, SMPR 2013, Tehran, Iran, oct., 2013.

SANQUETTA, C. R. Perspectivas da recuperação e do manejo sustentável das florestas de araucária. Disponível em: <<http://www.comciencia.br/reportagens/2005/08/09.shtml>>. Acesso em: 28 de fev. de 2014.

SANTA CATARINA, 2013. Portaria N° 170/2013/GABP-FATMA/BPMA-SC. Regula os procedimentos para apuração de infrações ambientais por condutas e atividades lesivas ao meio ambiente no âmbito da Fundação do Meio Ambiente - FATMA e do Batalhão de Polícia Militar Ambiental – BPMA. Publicada no **Diário Oficial do Estado** em 04 de out. de 2013.

SANTOS, A. A. et al. Análise comparativa entre classificação orientada ao objeto e a fotointerpretação do uso do solo no Parque Estadual da Serra do Rola Moça, MG. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, PR. **Anais...** São José dos Campos: INPE. p. 1190, 2013.

SATO, L. Y.; SHIMABUKURO, Y. E.; GOMES, V. C. Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, PR. **Anais...** São José dos Campos: INPE. p. 2352-2360, 2013.

SCHARDONG, G. G. **Acelerando o Mean-shift com Opencl.** 50 p. 2011. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade Federal de Santa Maria, Santa Maria, 2011.

SCHIEWE, J.; TUFTE, L. O potencial de procedimentos baseados em regiões para a avaliação integrada de dados de SIG e Sensoriamento Remoto. In: BLASCHKE, T.; KUX, H. **Sensoriamento Remoto e SIG Avançados.** 2. ed.- São Paulo: Oficina de textos. 2007. cap. 5.

SCHORN, L. A. et al. Síntese da estrutura dos remanescentes florestais em Santa Catarina. In: VIBRANS, A. C. et al. **Inventário Florístico Florestal de Santa Catarina.** Blumenau: Edifurb, 2012. Cap. 7, p. 125-140.

SCHOWENGERDT, R.A. **Remote sensing, models and methods for image processing**, 3.ed. San Diego: Academic Press, 2007. 558p.

SETTE, P. G. C. **Análise da textura de imagem de alta resolução para classificação de estágios sucessionais de Mata Atlântica do sul da Bahia**. 2009. 73p. Dissertação (Mestrado em Análise e Modelagem de Sistemas Ambientais)- Instituto de Geociências, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

SEVEGNANI, L. et al. Estádios sucessionais da Floresta Ombrófila Mista em Santa Catarina. In: VIBRANS, A. C. et al. **Inventário Florístico Florestal de Santa Catarina**. Blumenau: Edifurb, 2012. v. 3, cap. 9.

SHACKELFORD, A.K., DAVIS, C.H. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. **IEEE Transactions on Geoscience and Remote Sensing**. n. 41, v. 9, p.1920–1932, 2003.

SILVA, M. P. S. et al. Mining patterns of change in remote sensing image databases. In: IEEE Internacional Conference on data mining, 15., 2005.

SILVA, M. P. S. et al. Remote Sensing Image Mining: Detecting Agents of Land Use Change in Tropical Forest Areas. **International Journal of Remote Sensing**, v. 29, n. 16, p. 4803-4822, 2008. Disponível em: <http://www.dpi.inpe.br/gilberto/papers/rs_mining.pdf>. Acesso em: 05 de abr. de 2015.

SIMINSKI, A. **Formações florestais secundárias como recurso para o desenvolvimento rural e a conservação ambiental no litoral de Santa Catarina.** 102 p. 2004. Dissertação (Mestrado)- Universidade Federal de Santa Catarina, Florianópolis, 2004.

SIMINSKI, A.; FANTINI, A. C. Classificação da Mata Atlântica do litoral catarinense em estádios sucessionais: ajustando a lei ao ecossistema. **Floresta e Ambiente**, Seropédica, v.11, n.2, p. 20- 25, 2004.

SIMINSKI, A. **A floresta do futuro: conhecimento, valorização e perspectivas de uso das formações florestais secundárias no estado de Santa Catarina.** 2009, 153p. Tese (Doutorado em Ciências)- Universidade Federal de Santa Catarina, Florianópolis, 2009.

SIMINSKI, A.; FANTINI, A. C. A Mata Atlântica cede lugar a outros usos da terra em Santa Catarina, Brasil. **Biotemas**, 23 (2): 51-59, jun., 2010.

SPT. Segmentation Parameters Tuner. Disponível em: <<http://www.lvc.ele.puc-rio.br/wp/?p=1403>> Acesso em: 12 de out. de 2014.

SPT. User Guide 3.1.0. Disponível em: <http://www.lvc.ele.puc-rio.br/downloads/arquivos_download/User_Guide_3.1.0.pdf>. Acesso em: 10 de jan. de 2015.

SOUSA, B. F. S.; TEIXEIRA, A. S.; SILVA, F. A. T. F. da. Classificação de bioma caatinga usando Support Vector Machines (SVM). In: Simpósio Brasileiro de Sensoriamento Remoto. 14.,2009, Natal, Brasil. **Anais...** São José dos Campos: INPE, p. 7917-7924, 2009.

SOUSA, A. M. O. et al. Segmentação e classificação de tipologias florestais a partir de imagens QUICKBIRD. **Ambiência Guarapuava**, (PR) v.6 Ed. Especial, p. 57- 66, 2010. ISSN 1808 – 0251.

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**. 2012. 112p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.

SOUZA, U. D. V.; KUX, H. J. H. 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. In: Simpósio Brasileiro de Sensoriamento Remoto, 16., 2013, Foz do Iguaçu, SC. **Anais...**São José dos Campos: INPE, 2013.

STAVRAKOUDIS, D. G., THEOCHARIS, J. B., ZALIDIS, G. C. A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery. **ISPRS J. Photogramm. Remote Sens.**, 66(4), p. 529-544, 2011.

STEINWART, I., CHRISTMANN, A. **Support Vector Machines**. New York: Springer, 2008.

SZUSTER, W. B.; CHEN, Q.; BORGER, M. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. **Applied Geography**, 31, p. 525-532, 2011.

TAUBENBÖCK, H. et al. Object-based feature extraction using high spatial resolution satellite data of urban areas. **Journal of Spatial Science**, v. 55, n. 01, p. 117-132, 2010.

TEDESCO, A. et al. Detecção de formação erosiva (voçoroca) por meio de classificação hierárquica e por árvore de decisão. **Bol. Ciênc. Geod.**, Curitiba, v. 20, n. 4, out./dez, 2014.

THIRUMURUGANATHAN, S. **Introduction to Mean Shift Algorithm**. Disponível em: <
<http://saravananthirumuruganathan.wordpress.com/2010/04/01/introduction-to-meanshift-algorithm/>>. Acesso em: 19 de jan. de 2015.

TRAUCZYNSKI, R. A. **Perícias criminais em delitos contra a flora no Estado de Santa Catarina: Diagnóstico, metodologia e perspectivas**. 2013. 88 p. Dissertação (Mestrado profissional em Perícia Criminal)- Universidade Federal de Santa Catarina, Florianópolis, 2013.

TRIOLA, Mario F. **Introdução à estatística: atualização da tecnologia**. 11. ed. Rio de Janeiro: LTC, 2014. 707 p.

TSENG, M.H., CHEN, S-J.; HWANG, G-H; SHEN, M-Y. A genetic algorithm rule-based approach for land-cover classification. **ISPRS J. Photogramm. Remote Sens.**, 63(2), p. 202-212, 2008.

TSO, B.; MATHER, P. M. **Classification Methods for Remotely Sensed Data**. Second Edition. CRC Press: Florida, 2009.

TZOTSOS, A.; ARGIALAS, D. Support vector machine classification for object-based image analysis. **Object-Based Image Analysis**. Section 7, p. 663-677, 2008.

TZOTSOS, A. A support vector machine approach for object based image analysis. In: OBIA, 2006, Salzburg, Austria, 2006. **Proceedings...** Disponível em: <<http://www.datascienceassn.org/sites/default/files/A%20Support%20Vector%20Machine%20Approach%20for%20Object%20Based%20Image%20Analysis.pdf>> Acesso em: 02 de fev. de 2015.

VAN COILLIE, F. M. B.; VERBEKE, L.P.C.; WULF, de R. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. **Remote Sens. Environ.**, 110(4), p. 476–487, 2007.

VAPNIK, V. N. **The nature of Statistical learning theory.** Springer-Verlag, New York, 1995.

VIBRANS, A. C. et al. **Inventário Florístico Florestal de Santa Catarina.** v. 1. Diversidade e conservação dos remanescentes florestais. Blumenau: Edifurb, 2012. 344 p.

VIBRANS, A. C. **A cobertura florestal da bacia do rio Itajaí - elementos para uma análise histórica.** 2003. 238 p. Tese (Doutorado em Geografia) - Departamento de Geociências, Universidade Federal de Santa Catarina, Florianópolis, 2003.

VIEIRA, I. C. G. et al. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. **Remote Sensing of Environment**, v. 87, p. 470-481, 2003.

VU, T. T. Object-Based Remote Sensing Image Analysis with OSGeo Tools. In: Applied Geoinformatics for Society and Environment, 2012. Karlsruhe, Germany. 2 nd, extended edition, p. 80-84, 2012.

WANG, X., LI, Z., LIU, X., DENG, G., JIANG, Z. Estimating stem volume using quickbird imagery and allometric relationships for open populus xiaohai plantations. **Journal of Integrative Plant Biology**, 49(9), p. 1304–1312, 2007.

WEN, X., et al. Multi scale modeling for classification of SAR imagery using hybrid EM algorithm and genetic algorithm. **Prog. Nat. Sci.**, 19(8), p. 1033–1036, 2009.

WEKA. **Data Mining Software in Java**. Desenvolvido pela Universidade de Waikato. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>> Acesso em: 20 de set. 2014.

WEKA. **Manual for version 3-7-8**. 2013. Disponível em: <http://statweb.stanford.edu/~lpekelis/13_datafest_cart/WekaManual-3-7-8.pdf> Acesso em: 22 de jan. de 2015.

WISAENG, K. A comparison of different classification techniques for bank direct marketing. **International Journal of Soft Computing and Engineering (IJSCE)**, v. 3, sept., p. 116-119, 2013.

WITTEN, I. H.; FRANK, H. **Data mining: practical machine learning tools and techniques with java implementations**. San Francisco: Morgan Kaufman, 1999.

WULDER, M. A., HALL, R. J., COOPS, N., FRANKLIN S. High spatial resolution remotely sensed data for ecosystem characterization. **Bioscience**, 54, p. 511–521, 2004.

YU, L.; LIU, H. Feature selection for high-dimensional data: a fast correlation based filter solution. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML-2003), 20., 2003, Washington. **Proceedings...** Washington DC: [s.n.], 2003. Disponível em:
<<http://www.public.asu.edu/~huanliu/papers/icml03.pdf>>.
Acesso em: 13 de abr. de 2015.

YU, Q. et al. Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. **Photogrammetric Engineering & Remote Sensing**, v. 72, n. 7, jul., p. 799–811, 2006.

ZHANG, C. et al. Application and Evaluation of Object-oriented Technology in High-resolution Remote Sensing Image Classification. **National Key Laboratory of Remote Sensing Information and Imagery Analyzing Technology**, Beijing Research Institute of Uranium Geology. Beijing, China, 2011.

ZHANG, Q. et al. Parameters optimization of Support Vector Machine based on Simulated Annealing and Genetic Algorithm., p. 1302-1306, dec., 2009.

APÊNDICES

APÊNDICE A- Atributos gerados no InterIMAGE. (Continua)

| Atributo | Layers | Denominação |
|-----------------------|------------------------------------|--------------------|
| Média | mean('imagem_layer1') | M1 |
| | mean('imagem_layer2') | M2 |
| | mean('imagem_layer3') | M3 |
| | mean('imagem_layer4') | M4 |
| | mean('imagem_layer5') | M5 |
| | mean('imagem_layer6') | M6 |
| | mean('imagem_layer7') | M7 |
| Razão | Ratio('imagem_layer1') | R1 |
| | Ratio('imagem_layer2') | R2 |
| | Ratio('imagem_layer3') | R3 |
| | Ratio('imagem_layer4') | R4 |
| | Ratio('imagem_layer5') | R5 |
| | Ratio('imagem_layer6') | R6 |
| | Ratio('imagem_layer7') | R7 |
| Entropia | Entropy('imagem_layer1') | E1 |
| | Entropy('imagem_layer2') | E2 |
| | Entropy('imagem_layer3') | E3 |
| | Entropy('imagem_layer4') | E4 |
| | Entropy('imagem_layer5') | E5 |
| | Entropy('imagem_layer6') | E6 |
| | Entropy('imagem_layer7') | E7 |
| Desvio Padrão | standardDeviation('imagem_layer1') | Std1 |
| | standardDeviation('imagem_layer2') | Std2 |
| | standardDeviation('imagem_layer3') | Std3 |
| | standardDeviation('imagem_layer4') | Std4 |
| | standardDeviation('imagem_layer5') | Std5 |
| | standardDeviation('imagem_layer6') | Std6 |
| | standardDeviation('imagem_layer7') | Std7 |
| Brilho | brightness('imagem') | brilho |
| Máximo valor de pixel | maxPixelValue('imagem_layer1') | Maxpix1 |
| | maxPixelValue('imagem_layer2') | Maxpix2 |
| | maxPixelValue('imagem_layer3') | Maxpix3 |
| | maxPixelValue('imagem_layer4') | Maxpix4 |
| | maxPixelValue('imagem_layer5') | Maxpix5 |
| | maxPixelValue('imagem_layer6') | Maxpix6 |
| | maxPixelValue('imagem_layer7') | Maxpix7 |
| Mínimo valor de pixel | minPixelValue('imagem_layer1') | Minpix1 |
| | minPixelValue('imagem_layer2') | Minpix2 |
| | minPixelValue('imagem_layer3') | Minpix3 |
| | minPixelValue('imagem_layer4') | Minpix4 |

APÊNDICE A- Atributos gerados no InterIMAGE.
(Conclusão)

| | | |
|------|--|---------|
| | minPixelValue('imagem_layer5') | Minpix5 |
| | minPixelValue('imagem_layer6') | Minpix6 |
| | minPixelValue('imagem_layer7') | Minpix7 |
| | bandMeanDiv('imagem_layer1','imagem_layer4') | B1divb4 |
| | BandMeanSub('imagem_layer4','imagem_layer1') | B4-b1 |
| NDVI | $(m4-m1)/(m4+m1)$ | NDVI |
| | BandMeanDiv('imagem_layer4','imagem_layer1') | B4divb1 |

Fonte: produção do próprio autor, 2015.

APÊNDICE B- Atributos gerados no GeoDMA. (Continua)

| Atributo | Layer |
|---------------------------|----------------|
| Média | rp_mean_0 |
| | rp_mean_1 |
| | rp_mean_2 |
| | rp_mean_3 |
| | rp_mean_4 |
| | rp_mean_5 |
| | rp_mean_6 |
| Média normalizada | Nm_rp_mean_0 |
| | Nm_rp_mean_1 |
| | Nm_rp_mean_2 |
| | Nm_rp_mean_3 |
| | Nm_rp_mean_4 |
| | Nm_rp_mean_5 |
| | Nm_rp_mean_6 |
| Desvio padrão | rp_stddev_0 |
| | rp_stddev_1 |
| | rp_stddev_2 |
| | rp_stddev_3 |
| | rp_stddev_4 |
| | rp_stddev_5 |
| | rp_stddev_6 |
| Desvio padrão normalizado | Nm_rp_stddev_0 |
| | Nm_rp_stddev_1 |
| | Nm_rp_stddev_2 |
| | Nm_rp_stddev_3 |
| | Nm_rp_stddev_4 |
| | Nm_rp_stddev_5 |
| | Nm_rp_stddev_6 |
| Razão | rp_ratio_0 |
| | rp_ratio_1 |
| | rp_ratio_2 |
| | rp_ratio_3 |
| | rp_ratio_4 |
| | rp_ratio_5 |
| | rp_ratio_6 |
| Razão normalizada | Nm_rp_ratio_0 |
| | Nm_rp_ratio_1 |
| | Nm_rp_ratio_2 |
| | Nm_rp_ratio_3 |

APÊNDICE B- Atributos gerados no GeoDMA. (Continua)

| | |
|----------------------------|-------------------|
| | Nm_rp_ratio_4 |
| | Nm_rp_ratio_5 |
| | Nm_rp_ratio_6 |
| Amplitude | rp_amplitude_0 |
| | rp_amplitude_1 |
| | rp_amplitude_2 |
| | rp_amplitude_3 |
| | rp_amplitude_4 |
| | rp_amplitude_5 |
| | rp_amplitude_6 |
| Amplitude normalizada | Nm_rp_amplitude_0 |
| | Nm_rp_amplitude_1 |
| | Nm_rp_amplitude_2 |
| | Nm_rp_amplitude_3 |
| | Nm_rp_amplitude_4 |
| | Nm_rp_amplitude_5 |
| | Nm_rp_amplitude_6 |
| Entropia GLCM ⁴ | rp_entropy_0 |
| | rp_entropy_1 |
| | rp_entropy_2 |
| | rp_entropy_3 |
| | rp_entropy_4 |
| | rp_entropy_5 |
| | rp_entropy_6 |
| Entropia normalizada GLCM | Nm_rp_entropy_0 |
| | Nm_rp_entropy_1 |
| | Nm_rp_entropy_2 |
| | Nm_rp_entropy_3 |
| | Nm_rp_entropy_4 |
| | Nm_rp_entropy_5 |
| | Nm_rp_entropy_6 |
| Homogeneidade GLCM | rp_homogeneity_0 |
| | rp_homogeneity_1 |
| | rp_homogeneity_2 |
| | rp_homogeneity_3 |

⁴ Os atributos texturais baseados na GLCM foram calculados a partir de uma janela de tamanho 3x3, na direção SE, tendo em vista que é a única disponível na plataforma GeoDMA.

APÊNDICE B- Atributos gerados no GeoDMA. (Conclusão)

| | |
|-----------------------------------|---------------------------|
| | rp_homogeneity_4 |
| | rp_homogeneity_5 |
| | rp_homogeneity_6 |
| Homogeneidade normalizada GLCM | Nm_rp_homogeneity_0 |
| | Nm_rp_homogeneity_1 |
| | Nm_rp_homogeneity_2 |
| | Nm_rp_homogeneity_3 |
| | Nm_rp_homogeneity_4 |
| | Nm_rp_homogeneity_5 |
| | Nm_rp_homogeneity_6 |
| Rectangular fit | p_rectangular_fit |
| Rectangular fit normalizado | Nm_p_rectangular_fit |
| Perímetro area razão | p_perimeter_area_ratio |
| Perímetro area razão normalizado | Nm_p_perimeter_area_ratio |
| Índice de forma | p_shape_index |
| Índice de forma normalizado | Nm_p_shape_index |
| Densidade | p_density |
| Densidade normalizada | Nm_p_density |

Legenda: rp= atributo espectral; p= espacial e nm= normalizado.

Fonte: produção do próprio autor, 2015.

APÊNDICE C- Descrição dos atributos exportados no InterIMAGE⁵ e GeoDMA⁶. (Continua)

| Atributo | Descrição | Fórmula |
|---|--|---|
| Média ^{5,6} | Valor médio dos ND de uma banda espectral para cada objeto. | $\frac{\sum_{i=1}^N ND_{bi}}{n}$ <p>n = número de pixels de um segmento; e ND_{bi} = valor do número digital do <i>pixel</i> i na banda b.</p> |
| Desvio padrão ^{5,6} | Variação dos ND de cada segmento de uma banda espectral. | $\sqrt{\frac{1}{n-1} \sum_{i=1}^N (ND_{bi} - \mu)^2}$ <p>n = número de pixels de um segmento; ND_{bi} = valor do número digital do <i>pixel</i> i na banda b; e μ = média espectral do segmento na banda b.</p> |
| Razão ^{5,6} | Razão entre a média do segmento de uma banda pelo somatório das médias de todos os segmentos na mesma banda. | $\frac{Mb}{\sum_{j=1}^t Mb_j}$ <p>Mb = média do segmento na banda b; j = identificação do segmento; e t = total de segmentos na banda b.</p> |
| Amplitude ⁶ | Definido pela subtração do maior valor de <i>pixel</i> pelo menor valor. | $ND_{bimax} - ND_{bimin}$ <p>ND_{bi} = valor do número digital do <i>pixel</i> i na banda b.</p> |
| Máximo valor de <i>pixel</i> ⁵ | Máximo valor de <i>pixel</i> de todos os n <i>pixels</i> dentro do segmento. | |
| Mínimo valor de <i>pixel</i> ⁵ | Mínimo valor de <i>pixel</i> de todos os n <i>pixels</i> dentro do segmento. | |

⁵ Atributos exportados no InterIMAGE.

⁶ Atributos exportados no GeoDMA.

APÊNDICE C- Descrição dos atributos exportados no InterIMAGE e GeoDMA. (Continua)

| | | |
|------------------------------|---|---|
| NDVI ⁵ | É expresso como a diferença entre a banda do infravermelho próximo e vermelho, normalizada pela soma das bandas. | $\frac{(NIR - RED)}{(NIR + RED)}$ <p>NIR= ND da banda correspondente à região do infravermelho próximo; RED= ND da banda correspondente à região do vermelho</p> |
| B4-B1 ⁵ | Subtração da banda correspondente a região do Infravermelho próximo pela banda região do vermelho. | $NIR - RED$ <p>NIR= ND da banda correspondente à região do infravermelho próximo; RED= ND da banda correspondente à região do vermelho</p> |
| B4divB1 ⁵ | Produzido por uma simples divisão de valores contidos em bandas do infravermelho próximo por aqueles contidos na banda do vermelho. | $\frac{NIR}{RED}$ <p>NIR= ND da banda correspondente à região do infravermelho próximo; RED= ND da banda correspondente à região do vermelho</p> |
| B1divB4 ⁵ | Produzido por uma simples divisão de valores contidos em bandas do vermelho por aqueles contidos na banda do infravermelho. | $\frac{RED}{NIR}$ <p>NIR= ND da banda correspondente à região do infravermelho próximo RED= ND da banda correspondente à região do vermelho</p> |
| Entropia (GLCM) ⁶ | Mede a desordem de uma imagem. Quando a imagem não é uniforme, muitos elementos GLCM têm valores menores, resultando em maior entropia. | $-\sum_{i=1}^{D-1} \sum_{j=1}^{D-1} pij \times \log pij$ |

APÊNDICE C- Descrição dos atributos exportados no InterIMAGE e GeoDMA. (Conclusão)

| | | |
|---|--|--|
| Homogeneidade (GLCM) ⁶ | Assume valores maiores para as menores diferenças no GLCM. | $\sum_{i=1}^{D-1} \sum_{j=1}^{D-1} \frac{p_{ij}}{1 + (i - j)^2}$ |
| Ajuste rectangular (<i>Rectangular fit</i>) ⁶ | Razão entre os <i>pixels</i> dentro da região e o retângulo mínimo fora da região. Valores maiores representam regiões similares a um retângulo. | |
| Perímetro área razão (<i>Perimeter area-ratio</i>) ⁶ | Calcula a razão entre o perímetro e a área de uma região. É a medida da complexidade de forma da região. | $\frac{\textit{perimeter}}{N}$ <i>Perimeter</i> = perímetro do segmento; N= área do segmento. |
| Índice de forma (<i>Shape_index</i>) ⁶ | É o índice de forma da região, sendo igual a 1 quando é a forma mais compacta, aumentando seu valor conforme a irregularidade da forma. | |
| Densidade ⁶ | Descreve a compactação dos segmentos. É descrito pela razão da área que cobre um segmento pelo seu raio. | $D = \frac{\sqrt{n}}{1 + \sqrt{\textit{var}X} + \sqrt{\textit{var}Y}}$ <i>n</i> = número de <i>pixels</i> de um segmento; <i>var(X)</i> = componente na direção de X; e <i>var(Y)</i> = componente na direção de Y. |

Fonte: produção do próprio autor, com dados de Silva et al. (2005); Körting (2012); Piazza (2014).

APÊNDICE D- Árvores de decisão geradas pelo algoritmo C4.5 no InterIMAGE para cada área-teste.

| | |
|--|---|
| <pre> B1DIVB4 <= 0.9 R4 <= 0.1 -> sombra R4 > 0.1 B1DIVB4 <= 0.5 MAXPIX1 <= 91.0 -> avancado MAXPIX1 > 91.0 M7 > 811.8 -> avancado M7 <= 811.8 MINPIX4 > 66.0 -> medio MINPIX4 <= 66.0 MINPIX5 > 32.0 -> avancado MINPIX5 <= 32.0 STD4 <= 22.3 -> medio STD4 > 22.3 MINPIX1 <= 0.0 -> medio MINPIX1 > 0.0 E1 <= 5.0 -> avancado E1 > 5.0 B1DIVB4 <= 0.5 -> medio B1DIVB4 > 0.5 -> avancado B1DIVB4 > 0.5 M4 <= 154.1 STD4 > 24.7 -> avancado STD4 <= 24.7 M4 <= 135.7 -> inicial M4 > 135.7 STD7 <= 1.8 -> avancado STD7 > 1.8 M4 <= 140.3 -> medio M4 > 140.3 -> inicial M4 > 154.1 MAXPIX7 <= 795.4 -> medio MAXPIX7 > 795.4 -> avancado B1DIVB4 > 0.9 R6 > 0.1 -> edificacoes R6 <= 0.1 MAXPIX7 <= 807.5 R5 <= 0.1 -> inicial R5 > 0.1 -> campo MAXPIX7 > 807.5 M7 <= 806.0 -> campo M7 > 806.0 -> agricultura </pre> | <pre> B4-B1 <= 44.4 M4 <= 70.3 -> sombra M4 > 70.3 R6 <= 0.1 NDVI <= -0.1 -> campo_sujo NDVI > -0.1 R6 <= 0.1 -> campo_sujo R6 > 0.1 E7 <= 6.8 -> campo E7 > 6.8 MINPIX2 > 23.0 -> inicial MINPIX2 <= 23.0 E1 <= 5.4 -> inicial E1 > 5.4 -> campo_sujo R6 > 0.1 MINPIX7 <= 803.3 E1 > 5.3 -> campo_sujo E1 <= 5.3 MINPIX5 <= 104.0 -> campo MINPIX5 > 104.0 B1DIVB4 <= 1.3 -> campo B1DIVB4 > 1.3 -> campo_sujo MINPIX7 > 803.3 STD2 <= 19.3 -> campo STD2 > 19.3 MINPIX6 <= 68.0 -> campo_sujo MINPIX6 > 68.0 -> agricultura B4-B1 > 44.4 MAXPIX7 <= 805.4 M7 <= 787.0 -> medio M7 > 787.0 STD4 <= 20.2 MINPIX6 <= 50.0 -> inicial MINPIX6 > 50.0 -> medio STD4 > 20.2 STD7 <= 1.7 -> medio STD7 > 1.7 STD4 <= 22.4 -> medio STD4 > 22.4 B1DIVB4 <= 0.5 -> avancado B1DIVB4 > 0.5 -> medio MAXPIX7 > 805.4 M7 <= 826.3 E7 > 7.2 -> inicial E7 <= 7.2 MAXPIX3 <= 145.0 -> avancado MAXPIX3 > 145.0 M4 <= 121.1 -> medio M4 > 121.1 -> avancado M7 > 826.3 B1DIVB4 <= 0.4 -> </pre> |
|--|---|

```

NDVI <= -0.0 -> campo
NDVI > -0.0
| R6 <= 0.0 -> sombra
| R6 > 0.0
| | NDVI <= 0.1
| | | R3 <= 0.0 -> inicial
| | | R3 > 0.0
| | | | R5 <= 0.0 -> avancado
| | | | R5 > 0.0 -> inicial
| | NDVI > 0.1
| | | STD3 <= 6.8
| | | | M7 <= 1531.2
| | | | | R3 <= 0.0 -> inicial
| | | | | R3 > 0.0 -> avancado
| | | | | M7 > 1531.2
| | | | | R4 > 0.0 -> medio
| | | | | R4 <= 0.0
| | | | | E3 <= 4.6 -> medio
| | | | | E3 > 4.6 -> avancado
| | | STD3 > 6.8
| | | | B4DIVB1 > 1.3 -> avancado
| | | | B4DIVB1 <= 1.3
| | | | | MAXPIX4 <= 163.0 -> inicial
| | | | | MAXPIX4 > 163.0 -> avancado

```

| |
|---|
| C |
|---|

reflorestamento

| | | B1DIVB4 > 0.4 -> avancado

Fonte: produção do próprio autor, 2015.

APÊNDICE E- Árvores de decisão geradas pelo algoritmo SimpleCart no WEKA para cada área-teste.

| | | | |
|---|---|---|---|
| <pre> b4divb1 < 1.160111 m7 < 806.81787958 b1divb4 < 1.3589075: campo b1divb4 >= 1.3589075: edificacoes m7 >= 806.81787958: agricultura b4divb1 >= 1.160111 b4divb1 < 1.8785525 r4 < 0.106675 e4 < 6.646203 e3 < 4.170029: sombra e3 >= 4.170029: inicial e4 >= 6.646203: avançado r4 >= 0.106675 e6 < 5.3283305: medio e6 >= 5.3283305: avançado b4divb1 >= 1.8785525 e3 < 3.495915: sombra e3 >= 3.495915 maxpix7 < 800.4531258 r4 < 0.114505 maxpix7 < 793.5378428: avançado maxpix7 >= 793.5378428: medio r4 >= 0.114505: medio maxpix7 >= 800.4531258: avançado </pre> | A | <pre> b4-b1 < 44.4347227 r6 < 0.0619405 m4 < 67.8951227: sombra m4 >= 67.8951227 b4-b1 < -1.356860957: campo_sujo b4-b1 >= -1.356860957 r6 < 0.057853: campo_sujo r6 >= 0.057853: inicial r6 >= 0.0619405 minpix7 < 803.30178858: campo minpix7 >= 803.30178858 std2 < 20.1299617: campo std2 >= 20.1299617: agricultura b4-b1 >= 44.4347227 maxpix7 < 805.4714978: medio maxpix7 >= 805.4714978 b4-b1 < 100.1859968: avançado b4-b1 >= 100.1859968: reflorestamento </pre> | B |
| <pre> b1divb4 < 0.9863025 b4divb1 < 1.294170 std3 < 8.9789465: inicial std3 >= 8.9789465: avançado b4divb1 >= 1.294170 std3 < 6.8597805 m3 < 59.4770177: sombra m3 >= 59.4770177 maxpix7 < 1537.1376349: inicial maxpix7 >= 1537.1376349 r5 < 0.0329655: avançado r5 >= 0.0329655: medio std3 >= 6.8597805: avançado b1divb4 >= 0.9863025: campo </pre> | C | | |

Fonte: produção do próprio autor, 2015.

APÊNDICE F- Árvores de decisão geradas pelo algoritmo C4.5 no GeoDMA para cada área-teste.

```

rp_ratio_0 <= 0.0808352903
| nm_rp_mean_3 <= 0.3194212914
| | nm_p_shape_index <= 0.2063975930 -> sombra
| | nm_p_shape_index > 0.2063975930
| | | nm_rp_stddev_4 <= 0.4864028990 -> sombra
| | | nm_rp_stddev_4 > 0.4864028990 -> avançado
| nm_rp_mean_3 > 0.3194212914
| | nm_rp_stddev_4 <= 0.5981923938
| | | nm_rp_stddev_1 <= 0.4185225964 -> inicial
| | | nm_rp_stddev_1 > 0.4185225964
| | | | rp_ratio_3 <= 0.0696008280 -> inicial
| | | | rp_ratio_3 > 0.0696008280
| | | | | nm_rp_amplitude_2 <= 0.8666666746
| | | | | nm_rp_stddev_0 <= 0.3116489053 -> medio
| | | | | nm_rp_stddev_0 > 0.3116489053
| | | | | | nm_rp_stddev_2 <= 0.5592207909
| | | | | | | nm_p_perimeter_area_ratio > 0.1036953032 -> medio
| | | | | | | nm_p_perimeter_area_ratio <= 0.1036953032
| | | | | | | | nm_rp_amplitude_0 <= 0.5529412031 -> avançado
| | | | | | | | nm_rp_amplitude_0 > 0.5529412031
| | | | | | | | | nm_rp_amplitude_3 <= 0.7361111045 -> medio
| | | | | | | | | nm_rp_amplitude_3 > 0.7361111045 -> avançado
| | | | | | | nm_rp_stddev_2 > 0.5592207909
| | | | | | | | nm_rp_stddev_0 <= 0.3668682873 -> avançado
| | | | | | | | nm_rp_stddev_0 > 0.3668682873
| | | | | | | | | nm_rp_mean_0 <= 0.2603096068 -> medio
| | | | | | | | | nm_rp_mean_0 > 0.2603096068 -> inicial
| | | | | | | nm_rp_amplitude_2 > 0.8666666746
| | | | | | | | nm_rp_stddev_3 <= 0.5763316751 -> inicial
| | | | | | | | nm_rp_stddev_3 > 0.5763316751
| | | | | | | | | nm_rp_homogeneity_3 <= 0.0559089817 -> medio
| | | | | | | | | nm_rp_homogeneity_3 > 0.0559089817 -> avançado
| | | nm_rp_stddev_4 > 0.5981923938
| | | | nm_rp_amplitude_3 <= 0.7638888955 -> inicial
| | | | nm_rp_amplitude_3 > 0.7638888955 -> agricultura
rp_ratio_0 > 0.0808352903
| rp_ratio_5 > 0.0686780363 -> edificacao
| rp_ratio_5 <= 0.0686780363
| | nm_rp_mean_6 <= 0.6141895056 -> campo
| | nm_rp_mean_6 > 0.6141895056 -> agricultura

rp_ratio_4 <= 0.0606101304
| nm_rp_mean_3 <= 0.2258504927 -> sombra
| nm_rp_mean_3 > 0.2258504927
| | nm_rp_mean_6 <= 0.4305042028
| | | nm_rp_ratio_3 > 0.7709432244 -> medio
| | | nm_rp_ratio_3 <= 0.7709432244
| | | | nm_rp_mean_0 > 0.2095167041 -> inicial
| | | | nm_rp_mean_0 <= 0.2095167041
| | | | | nm_p_rectangular_fit <= 0.2891766131 -> medio
| | | | | nm_p_rectangular_fit > 0.2891766131 -> avançado
| | nm_rp_mean_6 > 0.4305042028

```

A

B

```

| | | nm_rp_ratio_3 > 0.8217931986 -> avancado
| | | nm_rp_ratio_3 <= 0.8217931986
| | | nm_rp_ratio_5 <= 0.5994511247 -> avancado
| | | nm_rp_ratio_5 > 0.5994511247
| | | | nm_rp_ratio_3 > 0.7453479767 -> inicial
| | | | nm_rp_ratio_3 <= 0.7453479767
| | | | nm_rp_mean_6 <= 0.7618234158 -> campo_sujo
| | | | nm_rp_mean_6 > 0.7618234158 -> inicial
rp_ratio_4 > 0.0606101304
| nm_rp_homogeneity_2 <= 0.0456966497 -> campo_sujo
| nm_rp_homogeneity_2 > 0.0456966497
| rp_ratio_3 > 0.0878239572 -> campo
| rp_ratio_3 <= 0.0878239572
| nm_rp_amplitude_5 <= 0.4078947008 -> campo
| nm_rp_amplitude_5 > 0.4078947008
| | nm_rp_amplitude_1 > 0.9098039269 -> campo_sujo
| | nm_rp_amplitude_1 <= 0.9098039269
| | | nm_rp_mean_6 <= 0.4129633904 -> campo_sujo
| | | nm_rp_mean_6 > 0.4129633904
| | | nm_rp_ratio_3 <= 0.6138870120 -> campo_sujo
| | | nm_rp_ratio_3 > 0.6138870120 -> agricultura

rp_ratio_0 <= 0.0485177711
| nm_rp_amplitude_5 <= 0.3020833135 -> sombra
| nm_rp_amplitude_5 > 0.3020833135
| | nm_rp_amplitude_2 <= 0.4432989955
| | | rp_ratio_1 <= 0.0333494097 -> avancado
| | | rp_ratio_1 > 0.0333494097
| | | nm_rp_mean_6 <= 0.6300721169 -> inicial
| | | nm_rp_mean_6 > 0.6300721169
| | | | nm_rp_amplitude_3 <= 0.5874125957 -> avancado
| | | | nm_rp_amplitude_3 > 0.5874125957
| | | | nm_rp_homogeneity_6 <= 0.1218764037 -> medio
| | | | nm_rp_homogeneity_6 > 0.1218764037 -> avancado
| | nm_rp_amplitude_2 > 0.4432989955
| | | nm_rp_entropy_2 <= 0.1428025961
| | | | nm_rp_amplitude_6 > 0.1754952073 -> avancado
| | | | nm_rp_amplitude_6 <= 0.1754952073
| | | | nm_rp_homogeneity_3 > 0.0620892718 -> inicial
| | | | nm_rp_homogeneity_3 <= 0.0620892718
| | | | | nm_rp_amplitude_6 <= 0.0643964410 -> inicial
| | | | | nm_rp_amplitude_6 > 0.0643964410 -> avancado
| | | nm_rp_entropy_2 > 0.1428025961
| | | nm_rp_mean_3 <= 0.6599115133 -> inicial
| | | nm_rp_mean_3 > 0.6599115133 -> medio
rp_ratio_0 > 0.0485177711
| nm_rp_mean_3 <= 0.7863097787 -> campo
| nm_rp_mean_3 > 0.7863097787 -> inicial

```

Fonte: produção do próprio autor, 2015.