



Ministério da
Ciência e Tecnologia



ANÁLISE DO ESTADO DE CONSERVAÇÃO DA VEGETAÇÃO INTRA-URBANA E SUAS RELAÇÕES COM DADOS SOCIOECONÔMICOS

Giovanni de Araujo Boggione

Relatório da disciplina Princípios e Aplicações de Mineração de Dados
(CAP-359) ministrada pelo Dr. Rafael Santos

INPE
São José dos Campos
2009



Ministério da
Ciência e Tecnologia



ANÁLISE DO ESTADO DE CONSERVAÇÃO DA VEGETAÇÃO INTRA-URBANA E SUAS RELAÇÕES COM DADOS SOCIOECONÔMICOS

Giovanni de Araujo Boggione

Relatório da disciplina Princípios e Aplicações de Mineração de Dados
(CAP-359) ministrada pelo Dr. Rafael Santos

INPE
São José dos Campos
2009

Boggione, Giovanni de Araujo.

Análise do estado de conservação da vegetação intra-urbana e suas relações com dados socioeconômicos / Giovanni de Araujo Boggione - São José dos Campos: INPE, 2009.

65p.

Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2009. Orientadores: Rafael Santos, Leila Maria Garcia Fonseca.

1. Vegetação. 2. Sensoriamento Remoto Urbano. 3. Mineração de Dados. 4. SIG. 5. Fusão. I. Título.

RESUMO

No Brasil a falta de planejamento para o crescimento das cidades tem contribuído para a degradação da qualidade ambiental e de vida da população, principalmente naquelas onde a aglomeração humana e de atividades já atingiu um nível considerado.

Embora se acredite na importância da existência e manutenção da vegetação intra-urbana, o que se constata com grande frequência é que essas áreas não têm apresentado crescimento proporcional e com a mesma intensidade em que se dá o crescimento populacional nas cidades. A falta de organização no desenvolvimento das cidades gera ambientes urbanos com elevados níveis de degradação, não somente porque o planejamento urbano não consegue alcançar o rápido processo de urbanização, mas também porque se observa um desinteresse político para a criação e implementação de mecanismos de combate ao declínio da qualidade de vida no ambiente urbano.

Neste contexto, o sensoriamento remoto se apresenta como uma interessante alternativa para o monitoramento e controle da cobertura vegetal. As características atuais dos sensores e a disponibilidade dos dados favorecem este cenário. Este trabalho apresenta uma adaptação da metodologia de Domingos (2005) e propõe a análise de dados oriundos de imagens de satélite e dados socioeconômicos utilizando, para tanto, técnicas de mineração de dados.

ABSTRACT

In Brazil the lack of city growth planning has been contributing to the degradation of the environmental quality and population life, mainly in those where human activity and gathering already reached a considerable level.

Although it is believed in the importance of the existence and maintenance of the intra-urban vegetation, what is verified with great frequency is that those areas have not been presenting proportional growth and with the same intensity of the population growth in the cities.

The lack of organization in the cities development generate urban atmospheres with high degradation levels, not only because the urban planning doesn't get to reach the fast urbanization process, but also because a political indifference is observed on the creation and implementation of combat mechanisms that decline the life quality in the urban atmosphere.

In this context, the remote sensing becomes an interesting alternative for vegetable covering control. The current characteristics of the sensor and the availability of data contribute to this scene. This work presents an adaptation of Domingos methodology (2005) and it proposes the analysis of satellite images originated data and socioeconomic data using, as well as data mining technique.

SUMÁRIO

Pág

LISTA DE FIGURAS

CAPÍTULO 1 INTRODUÇÃO	11
1.1 Proposta do trabalho	12
1.2 Organização da monografia	13
CAPÍTULO 2 REVISÃO BIBLIOGRÁFICA	15
CAPÍTULO 3 PROCESSAMENTO DIGITAL DE IMAGENS DE S. R.	17
3.1 Mosaico de Imagens	18
3.2 Restauração de Imagens	19
3.3 Georreferenciamento e Correção Geométrica.....	19
3.4 Reamostragem por Interpolação	20
3.5 Fusão de Imagens.....	21
3.6 Índice de Vegetação por Diferença Normalizada (NDVI) ...	24
3.7 Subtração de Imagens	25
3.8 Classificação de Imagens.....	25
CAPÍTULO 4 ESTUDO DE CASO: GOIÂNIA/GO	29
CAPÍTULO 5 MINERAÇÃO DE DADOS	31
CAPÍTULO 6 MATERIAIS E METODOLOGIA	35
6.1 Materiais	35
6.2 Metodologia	36
6.3 Pré-Processamento das Imagens	38
6.4 Processamento das Imagens	40
CAPÍTULO 7 RESULTADOS	49
CAPÍTULO 8 CONCLUSÃO	61
REFERÊNCIAS BIBLIOGRÁFICAS	63

LISTA DE FIGURAS

	<u>Pág</u>
1.1 – Vegetação intra-urbana apresentada como praça.	11
4.1 – Município de Goiânia. FONTE: Ferreira et all. (2007).	29
4.2 – Crescimento da população residente no município de Goiânia.....	30
6.1 – Diagrama de Transformações.	36
6.2 Fluxograma metodológico do processamento de imagens.	37
6.3 – Mosaico das cenas (a) HRC e (b) TM.	38
6.4 – Geração da máscara (em vermelho).....	40
6.5 – Aplicação da Máscara.	40
6.6 – Estádio Serra Dourada. Composição 3R4G2B. CCD HIBRIDA.	41
6.7 – Aeroporto de Goiânia. Composição 3R4G2B. TM HIBRIDA.	42
6.8 – Imagem Matiz identificando a vegetação.	42
6.9 – Subtração de NDVI. Época chuvosa e época seca.....	44
6.10 – Resultado da classificação pelo método Distância Mínima.	45
6.11 – Arquivos vetoriais gerados.	45
6.12 – Fluxograma metodológico do SIG.	47
6.13 – Área por Distrito Censitário.	47
6.14 – PMporD, AMporD e VTporD.....	48
7.1 – Distribuição Espacial entre áreas de pouca mudança (verde) e alta mudança (branco).	49
7.2 – Área apontada como “Pouca mudança”.	50

7.3 – Distribuição Espacial entre áreas de pouca mudança (verde) e alta mudança (branco) e renda.	50
7.4 – Consulta por atributos.	52
7.5 – Áreas de alta mudança. Vale do Rio Meia Ponte.	52
7.6 – Consulta por atributos.	53
7.7 – Áreas de pouca mudança. Centro e Campinas.	53
7.8 – Consulta por atributos.	54
7.9 – Áreas de pouca mudança. Centro e Campinas.	54
7.10 – Processamento no WEKA.	57
7.11 – Processamento no WEKA.	58

CAPÍTULO 1

INTRODUÇÃO

A partir dos anos 70 os municípios brasileiros sofreram intensas transformações. A carência de políticas eficazes de ordenação do desenvolvimento das cidades e o rápido crescimento das mesmas estão relacionados, na maioria das vezes, às inúmeras conseqüências da urbanização. A busca pela compreensão da diversidade dos aspectos do espaço urbano, relacionados às suas dimensões físico-territoriais e seus habitantes tornou-se uma preocupação para o planejamento e gestão urbana. Neste sentido, as discussões relacionadas à qualidade ambiental urbana têm sido cada vez mais freqüentes. Na literatura é fácil encontrar citações que consideram a vegetação como importante indicador de qualidade ambiental urbana. No entanto, há divergências conceituais, pois nota-se que termos como áreas verdes, espaços livre, áreas de lazer, por exemplo, são utilizados indistintamente como sinônimos para referência à presença de áreas de cobertura vegetal. Embora não haja uma definição consensual, o termo mais utilizado para designar a vegetação intra-urbana é “áreas verdes”. A Figura 1.1 apresenta uma área verde sob a forma de uma praça bem cuidada no município de Goiânia, GO.

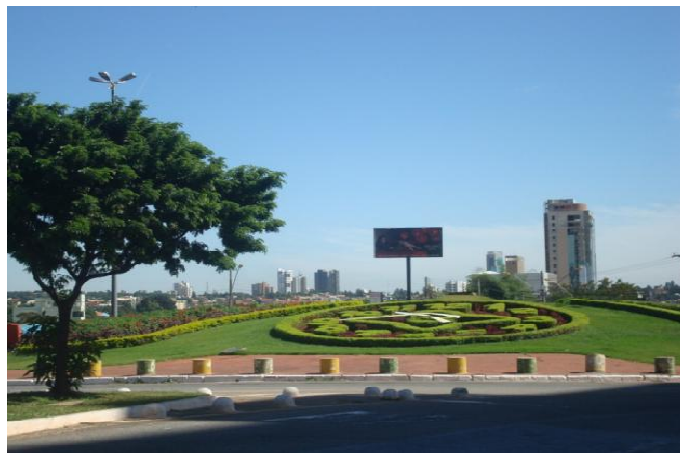


Figura 1.1 – Vegetação intra-urbana apresentada como praça.

O conhecimento da dinâmica da cobertura vegetal nas áreas urbanas em seus aspectos quantitativos e qualitativos é fundamental nos dias atuais ao bem estar da população e ao desenvolvimento econômico, social e ambiental de uma determinada região. Para tanto, é indispensável o desenvolvimento de ferramentas que permitam o monitoramento e a análise de informações destas alterações na paisagem de forma sistematizada. No intuito de automatizar e espacializar tais análises, as geotecnologias vêm sendo cada vez mais utilizadas, já que oferecem elementos que possibilitam subsidiar a dinâmica da análise da cobertura vegetal, viabilizando, assim, o planejamento e a detecção de mudanças ocorridas. Com isso, estudos de métodos referentes a técnicas que possibilitam a utilização de imagens oriundas de diferentes sensores e sua posterior integração com Sistemas de Informações Geográficas (SIG) vêm sendo de suma importância, visto que tais métodos permitem uma investigação da dinâmica da paisagem e o suposta correlação com variáveis sociais e econômicas.

1.1 Proposta do trabalho

É intenção deste trabalho responder às seguintes questões:

- É possível extrair a vegetação intra-urbana utilizando técnicas de PDI em imagens que não apresentem uma resolução fina?
- É possível determinar o estado da vegetação intra-urbana? A metodologia proposta é aplicável?
- É possível determinar uma relação entre o estado da vegetação intra-urbana e dados socioeconômicos utilizando técnicas de mineração de dados?

Para responder a estas questões parte-se das seguintes premissas:

- A cobertura vegetal intra-urbana pode ser observada e quantificada a partir de imagens de sensoriamento remoto orbital processadas para que sejam ajustadas às condições adequadas à investigação em função das características espaciais e espectrais das mesmas.

- A utilização de um conjunto de imagens de datas diferentes de sensores orbitais obtidas em períodos de seca e de chuva para determinar se são perceptíveis os efeitos sazonais do clima. Este efeito pode ser notado em imagens de satélite registradas nas estações seca e chuvosa. Caso a cobertura vegetal em questão esteja recebendo cuidados periódicos, essas variações sazonais devem ser mais suaves do que as da cobertura vegetal que não recebem cuidados artificiais.

Segundo Domingos (2005), o Estado, no papel das prefeituras municipais, responde pelas áreas públicas e se compromete a oferecer ambiente digno a população. Neste quadro, a Prefeitura Municipal tem o compromisso de zelar pela conservação e manutenção da cobertura vegetal urbana pública. Já as áreas particulares devem ser mantidas por seus respectivos proprietários.

Um estudo de caso na cidade de Goiânia, GO para o intervalo dos anos de 2008 e 2009 foi elaborado. Este estudo teve como propósito a verificação dos métodos e técnicas propostos, bem como a identificação das condições atuais em relação à vegetação existente.

Desta forma, torna-se possível diagnosticar melhor a situação da cidade em relação a sua vegetação intra-urbana. Com este diagnóstico, podem-se definir indicadores espaciais que tratam da condição de manejo da cobertura vegetal.

Este trabalho, portanto, apresenta como o uso das técnicas de processamento de imagens e mineração de dados torna possível levantar questionamentos a respeito das políticas públicas de manutenção de áreas verdes comuns aos cidadãos. Esses questionamentos permitem corroborar hipóteses sobre as relações entre o planejamento do crescimento das cidades e a degradação da qualidade ambiental e de vida da população no Brasil.

1.2 Organização da monografia

O Capítulo 2 apresenta a revisão bibliográfica dos temas abordados no trabalho, a pesquisa chave do projeto.

No Capítulo 3, as técnicas de processamento de imagens utilizadas neste trabalho são apresentadas como fundamentação teórica.

Uma descrição da área de estudo é apresentada no Capítulo 4. Detalhes sobre o município de Goiânia são descritos neste capítulo.

O Capítulo 5 apresenta o cenário da mineração de dados e as técnicas utilizadas nesta pesquisa.

No Capítulo 6 é apresentada a metodologia do trabalho e os materiais utilizados.

O Capítulo 7 mostra resultados do cruzamento das informações geradas pela metodologia e os dados socioeconômicos.

Finalmente, conclusões e uma visão geral do trabalho são apresentadas no Capítulo 7.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

Para a elaboração desta pesquisa, alguns trabalhos foram utilizados como referência principal.

A dissertação de mestrado pelo Instituto Nacional de Pesquisas Espaciais de Paulina Lopes Hoffmann Domingos, **APLICAÇÃO DE DADOS DE SENSORIAMENTO REMOTO ORBITAL DE ALTA RESOLUÇÃO À ANÁLISE DA COBERTURA VEGETAL INTRA-URBANA E SEU ESTADO DE CONSERVAÇÃO** (2005) foi o trabalho utilizado como referência principal porque trata da metodologia de extração da vegetação e também da análise do estado de conservação de áreas verdes urbanas utilizando imagens de sensoriamento remoto. Domingos apresenta detalhadamente os métodos adotados o que facilitou a adaptação do método proposto neste trabalho.

Outra pesquisa utilizada foi a **Análise de vetores de crescimento para a quantificação das transformações urbanas no município de Goiânia de Silva, Ferreira e Boggione do CEFET de Goiás**, onde os autores apresentam um cenário da tendência de crescimento do município de Goiânia levando em consideração, diversas variáveis e integrando imagens de sensoriamento remoto, cartas topográficas e sistemas de informações geográficas.

Para o georreferenciamento das imagens foi utilizada a ortofoto de Goiânia e arquivos complementares do MUBDG (Mapa Urbano Básico Digital de Goiânia). Estes dados são amplamente discutidos no trabalho de Ferreira, Nazareno e Macedo do CEFET de Goiás. **AVALIAÇÃO DA EXATIDÃO CARTOGRÁFICA DA ORTOFOTO DIGITAL DO MUNICÍPIO DE GOIÂNIA – GO.**

A conceituação que envolve o assunto de vegetação intra-urbana é amplamente discutida por Caporusso e Matias da UNICAMP no trabalho **ÁREAS VERDES URBANAS: AVALIAÇÃO E PROPOSTA CONCEITUAL.**

A utilização de mineração de dados em dados geográficos é abordado nos trabalhos de Vania Bogorny: **Uma Linguagem de Consulta para Mineração de**

Dados em Banco de Dados Geográficos Orientado a Objetos e GEOARM: an Interoperable Framework to Improve Geographic data Preprocessing and Spatial Association Rule Mining

CAPÍTULO 3

PROCESSAMENTO DIGITAL DE IMAGENS DE SENSORIAMENTO REMOTO

O Processamento Digital de Imagens (PDI) visa fazer a identificação e extração dos dados do sensor, servindo de fonte atualizada de informações para atender às grandes inovações que vêm surgindo com o passar dos anos, nas mais diversas áreas.

As técnicas de PDI podem ser divididas em atividades de pré-processamento e de processamento. Os métodos de pré-processamento preparam a imagem, adequando-a do ponto de vista geométrico, radiométrico e minorando os efeitos atmosféricos, enquanto os métodos de processamento realçarão o comportamento espectral ou textural de um objeto ou fenômeno.

O PDI tem como objetivo manipular e interpretar os dados, de maneira que possa melhorar o aspecto visual da imagem. E com isso realçar todas as informações de interesse para o uso do analista humano e conseqüentemente gerar uma gama de aplicações. A necessidade de processamento advém do fato de a imagem conter uma quantidade maior de informações do que o olho humano conseguiria perceber. (Crosta, 1992).

Várias são as formas que podem ser utilizadas para a manipulação de imagens. De um modo geral, podem ser categorizadas em um ou mais procedimentos que incluem quatro tipos abrangentes de operações computacionais que são: a retificação e restauração de imagens estão relacionadas com as operações realizadas para corrigir as distorções e degradações dos dados de uma imagem, com finalidade de criar uma representação mais fiel da cena.

O realce de imagens inclui os procedimentos aplicados aos dados de uma imagem com o objetivo de melhorar efetivamente a visualização da cena, para subsequente interpretação visual ou digital. Já a classificação de imagens são operações cuja finalidade é substituir a análise visual dos dados por técnicas quantitativas de análise automática, visando à identificação das feições presentes na cena. Por fim, as combinações de dados (data merging) são

procedimentos utilizados para combinar os dados de uma imagem, referente à certa área geográfica, com outros conjuntos de dados referenciados geograficamente para a mesma área.

3.1 Mosaico de Imagens

O mosaico de imagens é um produto gerado a partir de técnicas de registro de imagens. O processo de geração é considerado complexo, pois envolve um conjunto de procedimentos. Inicialmente é feito o registro da imagem de referência com a imagem de ajuste; é obtida em seqüência uma imagem resultante da combinação das duas. Assim, podem-se combinar sucessivamente outras imagens até que seja obtida uma imagem que cubra a área de interesse.

Alguns autores atentam que para a elaboração de um bom mosaico é imprescindível que se tenha um bom registro das cenas e sempre observar as áreas de transição entre uma cena e outra, devido às variações nas tonalidades dos níveis de cinza na imagem. Um mosaico de imagens consiste em uma composição de imagens adquiridas de diferentes pontos de vista para construir uma imagem maior, dando uma visão global da cena.

Define-se mosaico como sendo o conjunto de fotos de escala aproximada, de uma determinada região, recortada e montada, técnica e artisticamente de tal forma a dar a impressão de que todo o conjunto é uma única fotografia.

Quando se deseja combinar imagens tomadas em épocas diferentes (multi-temporais), o mosaico de imagens torna-se mais complexo de se elaborar. Pois devido às mudanças sazonais, o mesmo alvo aparece com forma (cor, formato) diferente nas imagens a serem registradas.

Um exemplo simples que o analista pode fazer é demarcar a região próxima às bordas das duas imagens e escolher a imagem sobre a qual esta região vai ser sobreposta, podendo utilizar um rio como a divisão entre as imagens. Outro problema a ser observado é a diferença de iluminação, isto ocorre mesmo quando as imagens são de um mesmo sensor, pode ser solucionado através

da equalização das imagens. Isto é feito através da identificação das médias e variâncias na região de sobreposição e normalizar uma das imagens considerando a outra como a referência.

3.2 Restauração de Imagens

O objetivo da restauração é realçar os detalhes que na imagem aparecem suavizados. Essas degradações inseridas no processo de formação de imagens atuam como filtros passa-baixas, eliminando ou atenuando o conteúdo em altas frequências da cena, responsáveis pelos detalhes da imagem (Fonseca, 1988). O que temos, portanto, é uma cópia borrada da cena. Aplicando o processo inverso do fenômeno de degradação na restauração, obtém-se uma imagem com aspecto realçado.

Os métodos de restauração baseiam-se em um modelo de degradação, enquanto que as técnicas de realce são empíricas e baseiam-se nas limitações do Sistema Visual Humano (SVH).

3.3 Georreferenciamento e Correção Geométrica

Ao fazer a aquisição de imagens brutas, geralmente estas contêm distorções geométricas, que inviabilizam sua utilização como mapa. Dentre as fontes de distorções geométricas estão a rotação da terra, distorções panorâmicas, curvatura da terra, arrastamento da imagem durante uma varredura, variação de altitude, altitude e velocidade do satélite e outros.

Partindo desse pressuposto, é necessário fazer o uso de técnicas que vão corrigir geometricamente as imagens escolhidas para o estudo. O georreferenciamento como correção geométrica é realizado com o objetivo de eliminar distorções sistemáticas e conferir maior precisão cartográfica às imagens, tratando-se de uma operação necessária para que se possa fazer a manipulação destas. Esta operação é feita pelo ajuste de uma imagem ou de um arquivo vetorial a um espaço definido por um sistema de coordenadas de referência. Com relação às imagens, uma forma comum de correção

geométrica se faz pela indicação do sistema referência (p.ex., latitude/longitude), dos sistemas de projeção (p.ex., Gauss, UTM) e das coordenadas dos limites: esquerdo, direito, superior e inferior da imagem. O mesmo é válido para dados vetoriais.

De acordo com Crósta (1992), o processo de correção geométrica envolve três etapas: a) Determinação da relação entre o sistema de coordenadas do mapa e da imagem; b) estabelecimento de um conjunto de pontos definindo os centros dos pixels na imagem corrigida, conjunto esse que define uma grade com as propriedades cartográficas do mapa de referência; c) o cálculo dos números digitais (DN's) dos pixels na imagem corrigida, por interpolação dos DN's dos pixels da imagem original.

Para obter um bom resultado na correção geométrica de imagens é necessário algum conhecimento prévio dos erros que interferem no processo de formação das imagens. Saber escolher o modelo matemático mais adequado ao tratamento de cada caso é importante. Outro fator a ser observado é que o programa utilizado na manipulação da imagem, que deve sempre propiciar ferramentas para que o resultado de uma correção geométrica possa ser avaliado e, conseqüentemente, validado.

O processo de correção geométrica de imagens inicia-se com uma transformação geométrica, também denominada mapeamento direto, que estabelece uma relação entre coordenadas de imagem (linha e coluna) e coordenadas geográficas (latitude e longitude). Trata-se de uma etapa em que se eliminam as distorções existentes e se define o espaço geográfico a ser ocupado pela imagem corrigida. Em seguida, faz-se o mapeamento inverso, que inverte a transformação geométrica usada no mapeamento direto, permitindo que se retorne à imagem original para que se definam os níveis de cinza que comporão a imagem corrigida. Esta definição de níveis de cinza ocorre na última etapa, chamada de reamostragem.

3.4 Reamostragem por Interpolação

Em processamento digital de imagens, a técnica que permite estimar valores de amostras de uma imagem em pontos de uma grade desejada, a partir de amostras da grade original, é conhecida por reamostragem. Quando existe a necessidade de estimar valores radiométricos em pontos intermediários situados entre os pontos em que os valores de radiometria são conhecidos, utiliza-se o processo de interpolação. Essa técnica é comumente utilizada nos processos de correção geométrica, registro e ampliação de escala.

Algumas técnicas de interpolação se destacam ou pela sua eficiência computacional ou pelos bons resultados obtidos. Entre estas técnicas podemos citar os métodos de Vizinho Mais Próximo, Bilinear e Convolução Cúbica (Richards, 1993). Como a teoria sobre estes interpoladores está bem estabelecida na literatura, neste trabalho não será dada uma atenção especial a este tópico. Para maiores detalhes ver Mather (1987). Vale a pena ressaltar que o uso de cada um destes interpoladores depende da aplicação a que se propõe. O interpolador Vizinho Mais Próximo tem sido utilizado em aplicações em que não se deseja alterar as estatísticas dos dados, enquanto o interpolador de Convolução Cúbica (ICC) tem sido usado em aplicações em que se deseja uma aparência mais natural da imagem. E o interpolador Bilinear é mais rápido que o ICC, porém degrada as feições na imagem.

Embora alterem a imagem significativamente, estes interpoladores são os mais utilizados já que estão implementados na maioria dos sistemas de processamento de imagens.

3.5 Fusão de Imagens

Há várias definições de fusão de imagem. De forma global, podemos descrever que a fusão de imagem consiste em combinar duas ou mais imagens diferentes para formar uma nova imagem, utilizando um algoritmo apropriado, com o objetivo de refinar e/ou melhorar a informação.

A fusão digital de diferentes conjuntos de dados espectrais e não espectrais permitem a análise simultânea de diferentes tipos de informação para uma

mesma área do terreno. A técnica de fusão vai realçar as imagens intensificando as sensações de brilho, cor e saturação.

Mesmo sendo um processo bastante difundido entre a comunidade de sensoriamento remoto, ainda há muita controvérsia quanto à utilização de um termo específico para designá-lo, sua definição varia de acordo com sua finalidade. Em outras palavras, qualquer integração de dados que possibilite um ganho de análise é considerada como fusão.

Para a maioria dos satélites comerciais de alta resolução são previstas duas categorias de imagens. As pancromáticas e as multiespectrais, sendo a primeira uma imagem de uma única banda, com resolução espacial melhor; enquanto que a segunda possui resolução espectral melhor, porém com pior resolução espacial. Quando se gera uma imagem híbrida a partir da técnica de fusão de dados de diferentes características (espectrais, espaciais e temporais), poder-se-á obter uma visão mais abrangente da área a ser observada. Desta forma, espera-se uma melhor capacidade de interpretação; ter-se-á uma imagem com um melhor aspecto visual, permitindo melhor diferenciação de áreas urbanas, tipos de vegetação, água, uso do solo etc.

Para que a fusão de imagens possa ser realizada é necessário que elas estejam sob um mesmo sistema de coordenadas, ou que tenham sido pré-processadas, de tal forma que estejam representando a mesma região espacial. Este procedimento é fundamental para o êxito da fusão. Para tal procedimento devem-se considerar alguns fatores prévios como: o pré-processamento envolvido; qual a melhor combinação dos dados; qual é o objetivo e também saber escolher a melhor técnica de fusão a ser utilizada para cada caso particular.

Alguns métodos têm sido utilizados para efetuar a fusão, como a conversão entre espaço de cores, a aplicação de filtro de detecção de bordas e a soma do resultado a uma das bandas, a utilização das componentes principais e de wavelets.

Estes métodos têm como finalidade a geração de produtos híbridos que melhoram a interpretação visual, embora alterem as informações espectrais das bandas originais. Cada método tem apresentado vantagens e

desvantagens, e pressupostos, os quais seguidos, produzem bons resultados finais.

Um dos métodos de fusão mais utilizados é o sistema IHS de representação de cores, que se constitui em uma alternativa ao sistema RGB. Dentro do espaço IHS, as cores são representadas através de três parâmetros independentes do ponto de vista da percepção visual. A intensidade (*Intensity*) que representa o brilho, o matiz (*Hue*) que está relacionado à cor (cromaticidade) e a saturação (*Saturation*) que é a pureza.

Sabins (1987), citado por (Napoleão, 2001), definiu os atributos do espaço IHS como sendo: a intensidade que é a medida da energia total refletida em todos os comprimentos de onda, e responsável, portanto, pela sensação de brilho; o matiz, que é determinado pelo comprimento de onda médio da energia emitida ou refletida pelo objeto, definindo sua cor; e a saturação, definida como o raio do comprimento de onda médio no qual a energia é refletida, expressando a pureza tonal do objeto.

Vários autores afirmam que a transformação IHS apresenta um melhor desempenho e flexibilidade devido aos seguintes fatores: produz um amplo intervalo de cores; preserva as informações espectrais e espaciais dos dados de entrada, razão pela qual as cores obtidas nas imagens de saída refletem as informações contidas nos dados podendo controlar cada parâmetro de cor.

Um exemplo de aplicações da transformação IHS é a fusão de imagens multiespectrais com uma imagem pancromática (composta apenas por uma banda), como por exemplo, bandas espectrais do sistema CBERS/CCD ou LANDSAT/TM com a banda pancromática CBERS/HRC ou a integração de imagens de satélite com informação geofísica. No caso de fusão CCD/HRC, o procedimento envolve calcular os componentes H, S e I a partir das três bandas selecionadas do CCD; substituir o componente I pela imagem HRC e após aplica-se a transformação inversa IHS—RGB. Desse modo, é possível a obtenção de uma composição colorida com a resolução espectral correspondente às três bandas CCD e a resolução espacial da imagem HRC.

3.6 Índice de Vegetação por Diferença Normalizada (NDVI)

Segundo Moreira e Shimabukuro (2004), na literatura são encontrados mais de cinquenta índices de vegetação sendo quase todos obtidos de medidas da reflectância nas faixas espectrais do vermelho e infravermelho próximo do espectro eletromagnético. Os dois tipos mais comumente utilizados, no entanto, são: Razão Simples (RVI) e o Índice de Vegetação por Diferença Normalizada (NDVI). Utilizou-se neste trabalho o NDVI que é definido por:

$$NDVI = (\rho_{ivp} - \rho_v) / (\rho_{ivp} + \rho_v) \quad (1)$$

Onde: ρ_{ivp} é a reflectância no infravermelho próximo;

ρ_v é a reflectância no vermelho.

O NDVI é o índice mais utilizado devido às vantagens que apresenta, tais como: menor influência de variações atmosféricas e a variação do resultado final é controlada, ou seja, varia de -1 a +1 enquanto outros mais simples variam de 0 a infinito.

No caso das imagens orbitais utilizadas neste trabalho (IKONOS 2), as bandas IVP e V correspondem às bandas 4 e 3 respectivamente. Quanto maior a diferença entre as respostas espectrais nessas duas faixas, maior o valor da razão entre as bandas. Segundo Jensen (2000), a resposta da banda da faixa do comprimento do vermelho (600 a 700nm) para a vegetação é inversamente proporcional a da banda do verde (500 a 600nm), sendo que a faixa espectral do verde reflete a quantidade de clorofila presente na planta. A resposta espectral da vegetação na faixa do infravermelho próximo (700 a 900nm) é relacionada, entre outras coisas, a sua concentração hídrica.

Para analisarmos a alteração da cobertura vegetal basta acompanhar os efeitos da variação do período na vegetação. Na época chuvosa a vegetação se apresenta mais aparente e dessa forma o valor de NDVI será mais alto e no período de seca, devido à menor concentração hídrica, o NDVI será menor.

O NDVI foi usado neste trabalho para ressaltar o estado da vegetação intra-urbana segundo a influência do período.

3.7 Subtração de Imagens

A técnica de subtração de imagens tem o objetivo de realçar possíveis diferenças entre o conjunto de imagens de entrada, desde que ambas estejam devidamente co-registradas. Neste trabalho, a operação de subtração foi utilizada para detectar as mudanças ocorridas na cobertura vegetal entre duas datas.

Ao subtrair a imagem NDVI da época chuvosa da imagem NDVI da época seca, tem-se uma nova imagem, cujos níveis de cinza correspondem às alterações detectadas entre as imagens em questão.

3.8 Classificação de Imagens

As técnicas de classificação de imagem podem ser definidas como um processo utilizado para extrair o máximo de informações possível numa imagem a fim de discernir e determinar critérios para os diferentes alvos e objetos, como vegetação, solo, área urbana, água etc. Essas técnicas consistem, basicamente, em separá-los, colocando-os em grupos previamente definidos, onde vai ser estabelecida uma relação entre os pixels da imagem a uma determinada categoria ou classe, que deve ser escolhida anteriormente pelo usuário, tendo como resultado final um mapa temático. No processamento de imagem, atributos relacionados à intensidade dos pixels são naturalmente escolhidos para representarem objetos ou regiões das imagens num classificador, ressaltam.

Segundo (Novo, 1989), as técnicas de classificação digital implicam na implementação de um processo de decisão para que o computador possa atribuir certo conjunto de pontos (pixels) a uma determinada classe.

A classificação de objetos ou fenômenos é feita pela escolha das características que os descrevem para diferenciá-los entre si. Na definição matemática, usa-se o espaço de atributos ("feature space"). O espaço de

atributos é definido como um processo onde os elementos da imagem pertencentes a uma classe vão ser aglomerados como uma nuvem de pontos, onde o algoritmo classificador vai detectar o alvo ou objeto. Este espaço vai ser dividido em regiões onde cada pixel vai corresponder a uma classe à medida que vão se agregando, vão formando as amostras.

Crosta (1992) afirma que no processamento digital, a cada eixo desse espaço são atribuídos os níveis de cinza de uma determinada banda espectral. E ainda adverte que pode haver situações em que o uso de uma ou duas bandas não gera uma boa caracterização espectral da superfície de escolha. Pode-se definir como regra geral, que quanto maior o número de bandas espectrais, maior será a precisão da classificação.

Entretanto, existem ainda vários tipos de classificadores que podem ser utilizados conforme o processo de classificação. São eles: pixel a pixel em que se utiliza a informação espectral isolada de cada pixel, para fazer a identificação da região homogênea. E a classificação por região, em que a imagem é dividida em região (segmentação) e depois cada região é unida a uma classe. A principal vantagem de trabalhar com classificação por regiões é que essa técnica fornece informações de natureza espacial, que não podem ser extraídas individualmente de pixels e que são normalmente consideradas durante o processo de análise visual de imagens.

Dentre os métodos de classificação mais utilizados, destacam-se a classificação supervisionada e não supervisionada.

Na classificação não supervisionada não há qualquer conhecimento prévio do classificador sobre os atributos das classes pertinentes à cena, enquanto que na classificação supervisionada, o classificador orienta sua busca de classes a partir de amostras de treinamento feitas anteriormente com as classes de interesse da cena.

A classificação supervisionada constitui em uma ferramenta analítica muito usada para extração de informação quantitativa a partir dos dados multiespectrais. Essa técnica é executada usando algoritmos adequados para a determinação dos pixels a fim de se fazer uma representação da classe.

O processo é dito supervisionado devido à constante interação entre o intérprete e o sistema de análise. O intérprete humano deverá ter um conhecimento prévio da área, para que possa estar apto a fornecer informações específicas, possibilitando assim fazer o treinamento do algoritmo de classificação. Neste tipo de classificador torna-se essencial que se tenha uma seleção de amostras homogêneas da classe e que, ao mesmo tempo, deva conter todas as variações dos níveis de cinza da imagem. Esta etapa é denominada como treinamento.

De acordo com (Crosta, 1992), uma área de treinamento é definida pelo usuário traçando-se seus limites diretamente sobre a imagem, no monitor e várias áreas de treinamento podem ser definidas para uma mesma classe, para poder assegurar que os pixels a ela pertencentes são realmente representativos a dela.

CAPÍTULO 4

ESTUDO DE CASO: GOIÂNIA/GO

O município de Goiânia está localizado a aproximadamente 190 km do Distrito Federal, entre as coordenadas 49°27' W, 16°50' S e 49°04' W, 16°27' S, conforme pode ser observado na Figura 4.1. Ocupa uma área total de 740,53 km², sendo que cerca de 40% da área do município já está urbanizada.

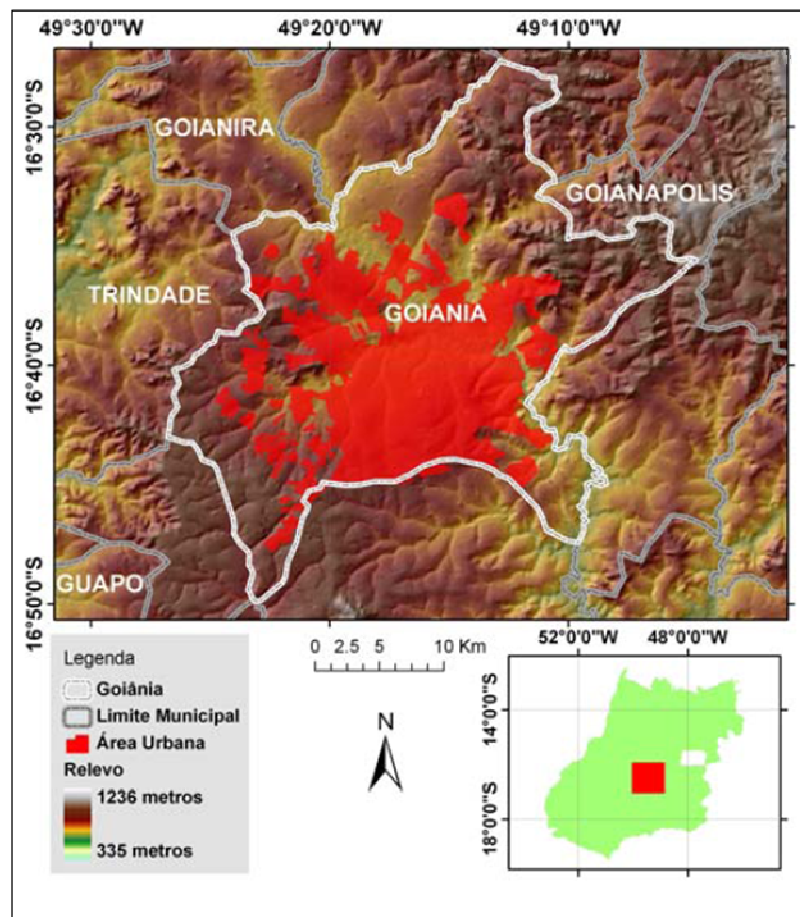


Figura 4.1 – Município de Goiânia. FONTE: Ferreira et al. (2007).

Com uma população residente total de 1.093.007 habitantes (IBGE, 2000), no município de Goiânia ocorreu uma significativa explosão populacional entre a década de 1950 até 1980. Neste período, a população do município

praticamente duplicava a cada dez anos, provavelmente devido à mudança da capital federal para Brasília e também devido aos projetos do governo federal de infra-estrutura e de incentivos para a ocupação e uso do bioma cerrado para prática de agropecuária. Após a década de 1980, o crescimento populacional de Goiânia ainda continuou alto, com cerca de 20% a cada 10 anos, conforme se pode observar na Figura 4.2.

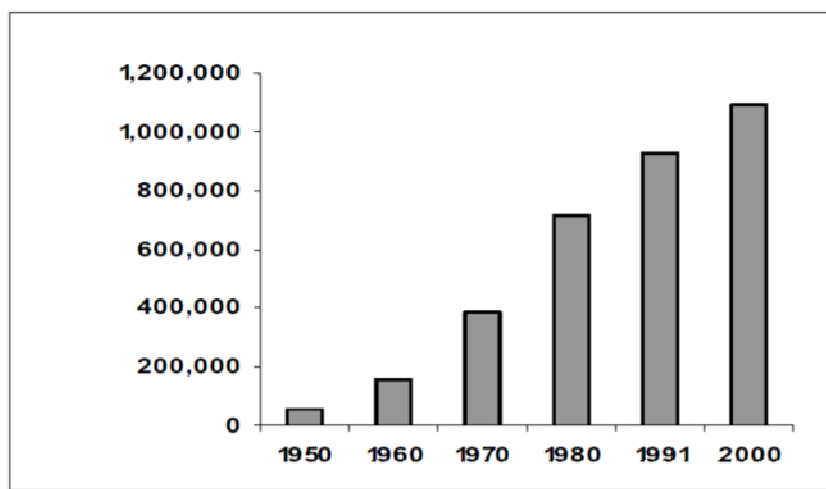


Figura 4.2 – Crescimento da população residente no município de Goiânia.

FONTE: IBGE (2000).

O rápido crescimento populacional ocorrido em Goiânia induziu também o rápido crescimento da área urbana do município. O mais interessante é que apesar da maior parte do município ainda pertencer à zona rural (cerca de 60%), praticamente toda a população de Goiânia (99,34%) da população total, reside em área urbana, e o restante da população, apenas (0,66%) reside em áreas rurais.

CAPÍTULO 5

MINERAÇÃO DE DADOS

Descoberta de conhecimento em bancos de dados, é o processo não trivial de identificar em dados padrões que sejam válidos, novos (previamente desconhecidos), potencialmente úteis e compreensíveis, visando melhorar o entendimento de um problema ou um procedimento de tomada de decisão [Fayyad et al. 1996].

Um dos objetivos da descoberta de conhecimento em bancos de dados é tornar padrões compreensíveis para humanos, visando promover uma melhor compreensão dos próprios dados. Embora seja um tanto subjetivo medir compreensibilidade, um dos fatores freqüentes é a medida de simplicidade. O fator de compreensão dos dados está relacionado à intuitividade da representação destes, bem como da granularidade alta o suficiente para que estes sejam compreendidos.

5.1. Etapas do Processo de Descoberta de Conhecimento em Bancos de Dados

O processo de Descoberta de conhecimento em bancos de dados é interativo, iterativo, cognitivo e exploratório, com muitas decisões sendo feitas pelo analista (que é um especialista do domínio dos dados, ou um especialista de análise dos dados).

1. Definição do tipo de conhecimento a descobrir, o que pressupõe uma compreensão do domínio da aplicação bem como do tipo de decisão que tal conhecimento pode contribuir para melhorar.
2. Criação de um conjunto de dados alvo: selecionar um conjunto de dados, ou focar num subconjunto, onde a descoberta deve ser realizada.
3. Limpeza de dados e pré-processamento: operações básicas tais como remoção de ruídos quando necessário, coleta da informação necessária para modelar ou estimar ruído, escolha de estratégias para manipular campos de dados ausentes, formatação de dados de forma a adequá-los à ferramenta de mineração.

4. Redução de dados e projeção: localização de características úteis para representar os dados dependendo do objetivo da tarefa, visando a redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento semântico das informações.

5. Mineração de dados (Data Mining): selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação ou conjunto de representações; busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão.

6. Interpretação dos padrões minerados, com um possível retorno aos passos 1-6 para posterior iteração.

7. Implantação do conhecimento descoberto: incorporar este conhecimento ao desempenho do sistema, ou documentá-lo e reportá-lo às partes interessadas.

Etapas de mineração de dados utilizam técnicas e algoritmos de diferentes áreas do conhecimento, principalmente inteligência artificial (especialmente aprendizagem de máquina), banco de dados (recursos para manipular grandes bases de dados) e estatística (comumente na avaliação e validação de resultados). Uma questão que freqüentemente surge é a seguinte: porque não utilizar tão somente os conhecidos procedimentos estatísticos para obter informações relevantes a partir de um conjunto de dados?

Conforme mencionado, procedimentos estatísticos são utilizados nas etapas de descoberta de conhecimento em bancos de dados e mais especificamente na mineração de dados. Entretanto, o volume, complexidade e peculiaridades dos eventos e dos dados por eles originados impõem severas limitações a metodologias puramente estatísticas, dentre elas:

- Dados nem sempre possuem independência estatística entre eles, ou seja, muitos domínios possuem inter-relação entre seus objetos e respectivos atributos, comprometendo a aplicação de métodos estatísticos;
- A análise estatística demanda um grau de conhecimento e domínio desta área que apenas estatísticos e profissionais de áreas correlatas possuem, restringindo assim a atuação da grande maioria dos potenciais usuários de procedimentos analíticos;

- Métodos estatísticos manipulam muito bem dados numéricos, mas não manipulam bem valores simbólicos, incompletos ou inconclusivos;
- Estes métodos são computacionalmente caros quando se trata de grandes bases de dados. Desta forma, percebe-se claramente que a mineração de dados possui grande relevância, contribuição e abrangência no que diz respeito a aplicações.

5.2 APRIORI

APRIORI é um algoritmo seminal proposto em (SIKRANT; AGRAWAL, 1994) para a mineração de conjuntos de itens freqüentes. Usando este algoritmo, podemos descobrir todos os conjuntos de itens que satisfaçam um limite mínimo de freqüência, independente do seu tamanho, i.e., do número de itens que ele contém.

Esse algoritmo se baseia na seguinte propriedade: *“Todo subconjunto não vazio de um conjunto de itens freqüentes também dever ser freqüente.”* Isto significa que dentro de um conjunto de itens freqüente de tamanho k , todos os seus subconjuntos de tamanho m ($0 < m < k$) também são freqüentes. O que é lógico; pois um conjunto não tem como ser mais freqüente que um de seus subconjuntos.

A partir dessa premissa, o algoritmo APRIORI faz o caminho inverso. A partir de todos os conjuntos de itens freqüentes de tamanho k , ele forma todos os possíveis conjuntos de itens freqüentes de tamanho $k+1$, porque *“apriori”*, um conjunto freqüente só pode ser formado por subconjuntos freqüentes. Assim, iterativamente, o algoritmo vai procurando por conjuntos cada vez maiores, baseando-se na combinação de conjuntos já encontrados.

O algoritmo funciona da seguinte maneira: começamos descobrindo todos os conjuntos de itens freqüentes de tamanho 1, formando a lista $L1$. A partir dessa lista, podemos gerar os conjuntos candidatos à lista $L2$, i.e., todos os conjuntos de tamanho passíveis de serem freqüentes. Segundo a propriedade *APRIORI*, todos os elementos de $L2$ são formados por subconjuntos oriundos de $L1$; Assim, combinando os elementos de $L1$ entre si, de forma a obtermos conjuntos de tamanho 2, e levando em consideração que um item só pode

aparecer uma vez num conjunto, obtemos os candidatos a serem elementos da lista L_2 . O próximo passo é verificar quais destes conjuntos candidatos atendem à frequência mínima; eliminando-se aqueles que não atendem, os que restarem serão os elementos de L_2 . Prosseguindo dessa forma, iterativamente, podemos combinar os elementos de L_i para obter os elementos de L_{i+1} , até que não haja mais conjuntos maiores frequentes.

CAPÍTULO 6

MATERIAIS E METODOLOGIA

6.1 Materiais

As imagens de sensoriamento remoto utilizadas para o desenvolvimento deste trabalho foram adquiridas através do site do (INPE) Instituto Nacional de Pesquisas Espaciais (<http://www.dgi.inpe.br>). Segue a descrição:

- Imagem do satélite LANDSAT-5, sensor TM, bandas 2 (verde), 3 (vermelho) e 4 (infravermelho), Órbitas Ponto 222/071 e 222/072, correspondente à região de Goiânia, com data de aquisição em 03/03/2009, período chuvoso, resolução espacial de 30m.
- Imagem do satélite CBERS 2B, sensor CCD, bandas 2 (verde), 3 (vermelho) e 4 (infravermelho), Órbita Ponto 158/119, data da aquisição 19/08/2008, período seco, resolução espacial de 20 metros.
- Imagem do satélite CBERS 2B, sensor HRC, banda pancromática, Órbitas Ponto 158/A/119/1 e 158/A/119/2, data da aquisição 10/10/2008, período seco, resolução espacial de 2,5 metros.

Para o georreferenciamento das imagens, foi utilizada uma ortofoto de 2006, e para a validação da extração da vegetação foi utilizada uma Imagem QuickBird de 2002. Ambas as imagens possuem resolução espacial nominal de 60 cm, composição natural (sem infravermelho) e foram adquiridas junto à prefeitura de Goiânia.

Dados Censitários foram adquiridos junto ao IBGE e correspondem ao censo de 2000. Estes dados foram disponibilizados agregados em Distritos Censitários.

Dados vetoriais complementares foram adquiridos junto à prefeitura de Goiânia sob a forma do MUBDG (Mapa Urbano Básico Digital de Goiânia) e foram atualizados em 2008.

Os softwares utilizados neste trabalho foram o SPRING para o processamento de imagens, o TerraView para o SIG e o WEKA para a mineração de dados.

6.2 Metodologia

A seguir é apresentado o modelo OMT-G (Figura 6.1) em forma de Diagrama de Transformação.

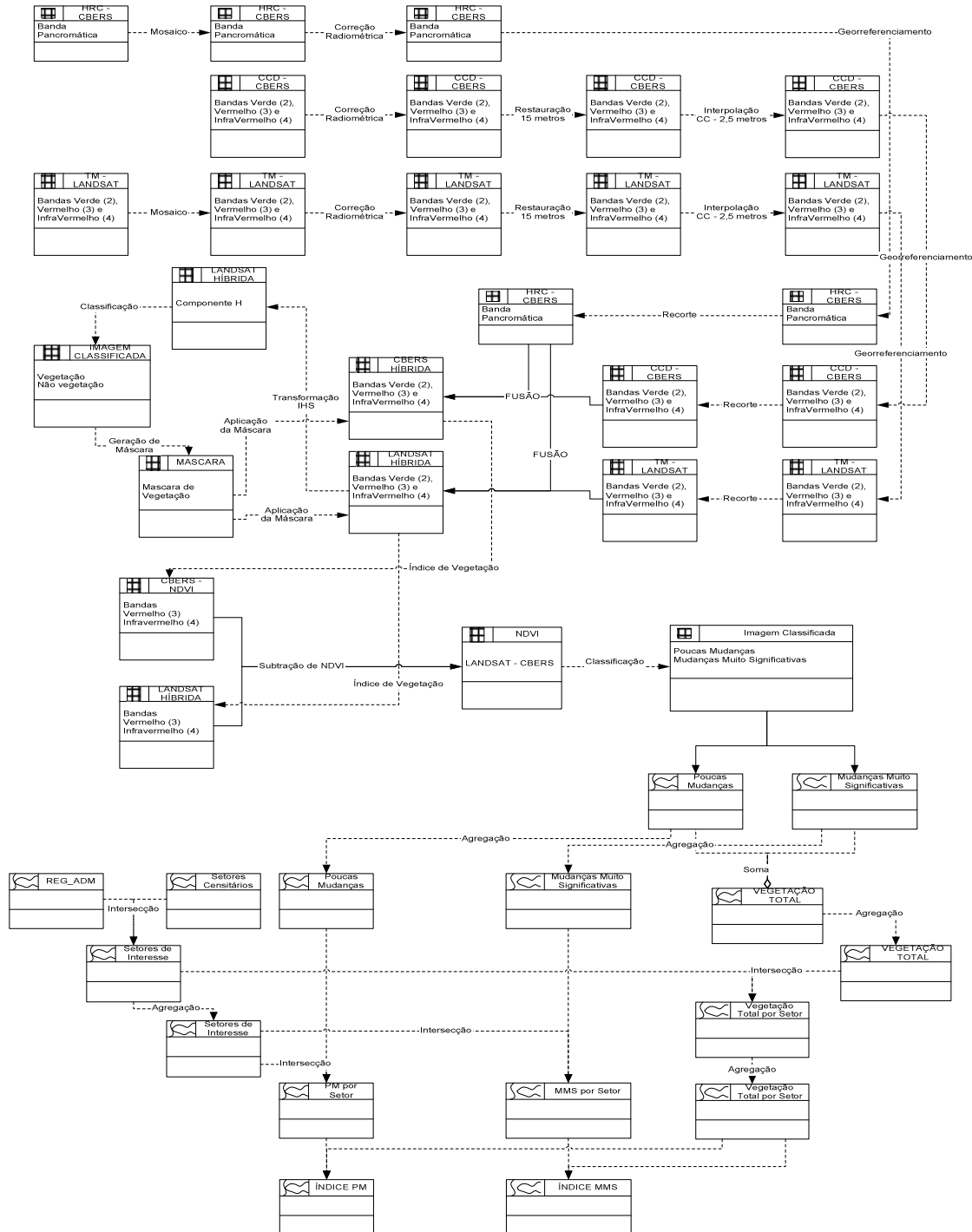


Figura 6.1 – Diagrama de Transformações.

O fluxograma apresentado na Figura 6.2 representa uma simplificação da parte de PDI do modelo OMT-G e tem a funcionalidade de mostrar de forma esquemática a metodologia adotada para a geração do dado vetorial através de imagens de SR que será utilizado no SIG.

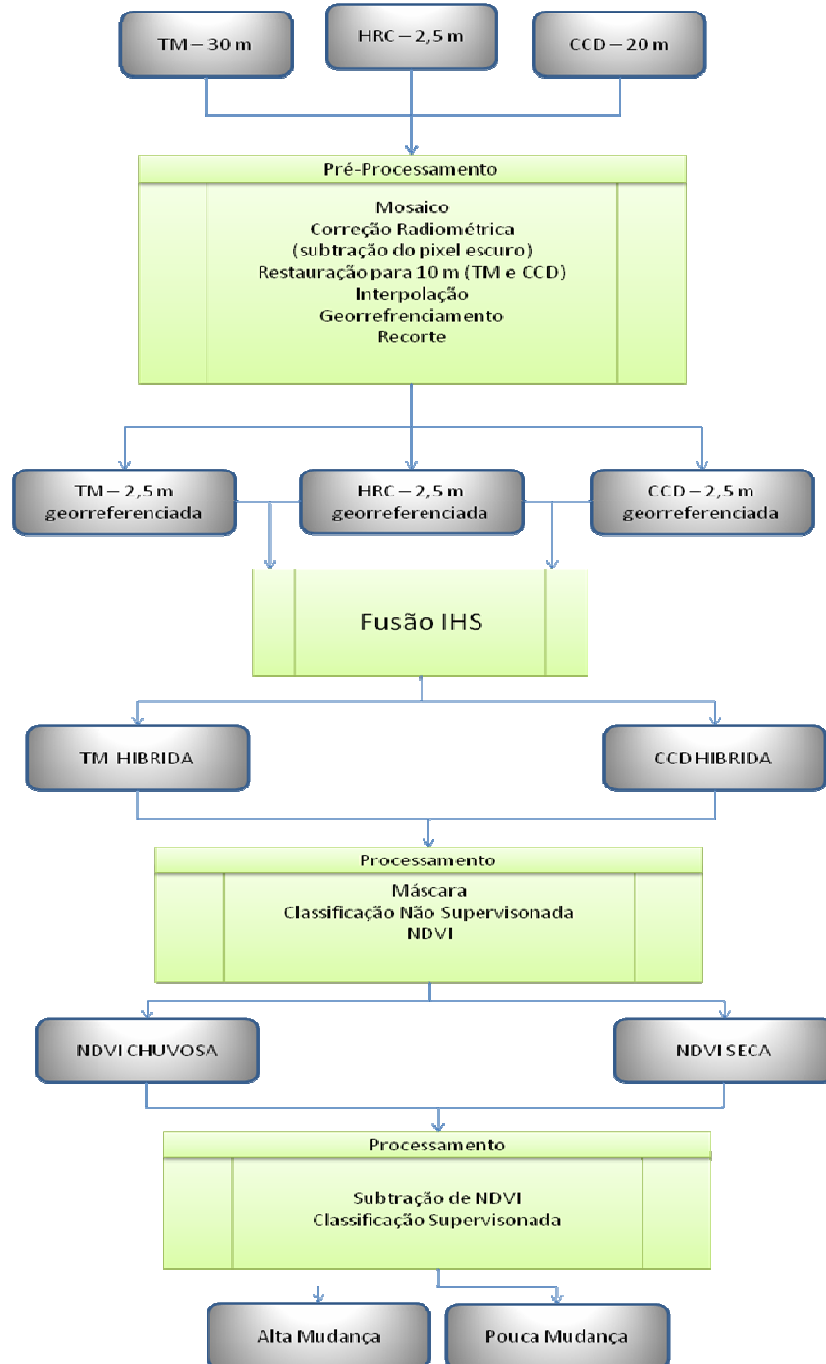


Figura 6.2 Fluxograma metodológico do processamento de imagens.

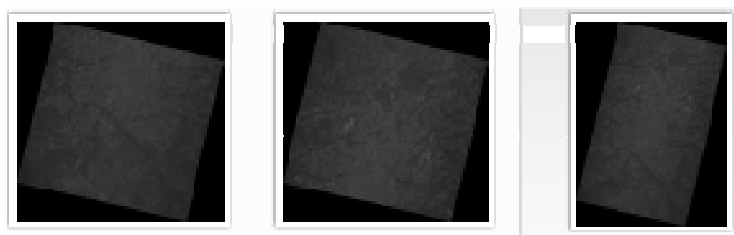
Como demonstra a figura, a execução deste fluxograma foi realizada utilizando as técnicas de PDI apresentadas anteriormente e foram dispostas em blocos para uma melhor compreensão do processo proposto.

O pré-processamento e o processamento de imagens foram executados no SPRING.

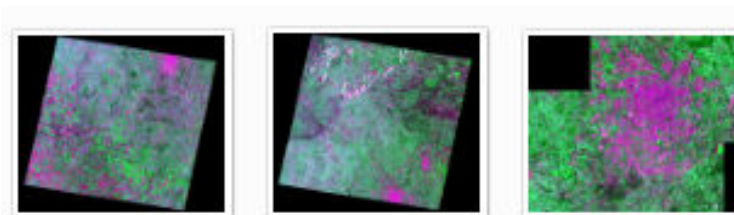
6.3 Pré-Processamento das Imagens

Em função das diferenças existentes nas imagens geradas pelos sensores é natural que haja distorções entre as imagens, sendo necessária a realização de algumas etapas de pré-processamento para que o conjunto de dados se torne coerente para o procedimento proposto.

O ajuste em questão iniciou-se com mosaico das cenas das imagens TM e HRC conforme mostra a Figura 6.3.



(a) Órbita/Ponto: 158_A_119_1 + 158_A_119_2



(b) Órbita/Ponto: 222_071 + 222_072

Figura 6.3 – Mosaico das cenas (a) HRC e (b) TM.

A etapa seguinte consistiu na correção radiométrica pela técnica da subtração do pixel escuro. Este procedimento se fez necessário para equilibrar a intensidade do conjunto de dados.

Para estabelecer o mesmo tamanho de pixel para as imagens e garantir às mesmas uma melhor qualidade visual foi realizada a restauração das imagens CCD e TM, alterando o tamanho do pixel de 20 e 30 metros respectivamente, para 10 metros.

A próxima etapa é a reamostragem das imagens usando o interpolador de convolução cúbica que apresenta um aspecto mais natural. Os dados foram interpolados para 2,5 metros para que apresentassem o mesmo tamanho de pixel da imagem HRC. Condição fundamental para o bom resultado da fusão.

No georreferenciamento das imagens HRC, CCD E TM foi utilizada como referência a ortofoto de 2006. A ortofoto apresentava georeferenciamento no sistema de Projeção UTM com Datum Sad 69 e com isso as imagens a serem ajustadas ficaram com o mesmo sistema de coordenadas. Para isso, coletou-se um conjunto de 47 pontos de controle identificáveis e bem distribuídos por toda área de estudo, ou seja, sobre as imagens a serem ajustadas, obtendo um erro menor que 0,38 pixel. Para minimizar ainda mais o erro, foi utilizado o aplicativo REGEMY, versão 0.2.43 que permitiu um melhor refinamento dos pontos de controle e erro inferior a 0,17 pixel.

O passo seguinte foi o recorte da área a ser trabalhada. Como as cenas ultrapassavam a área de interesse após serem mosaicadas, foi preciso sobrepor na imagem um arquivo vetorial com os limites da região, para que pudesse eliminar a área que não seria utilizada. Em virtude dessa situação foi utilizado um vetor do limite da área piloto escolhida em Goiânia (Figura 6.4) que contém 5 Macrozonas (Campinas, Sul, Central, Macambira e Vale do Maia Ponte) que foi obtido junto à prefeitura no MUBDG, para posterior confecção de uma máscara e a partir dela, um recorte contendo apenas a área de interesse, ao invés da cena inteira, como mostra a Figura 6.5 a seguir. Estas áreas foram escolhidas por apresentarem maior quantidade de população e grande densidade de construções.

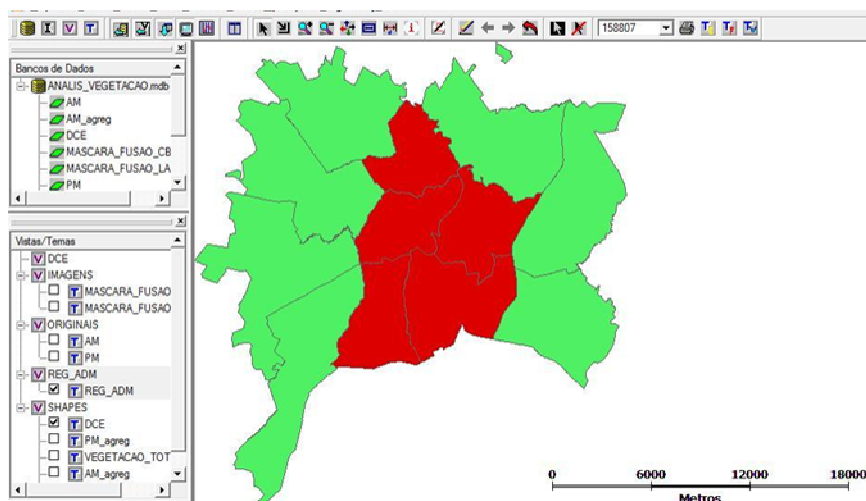


Figura 6.4 – Geração da máscara (em vermelho).



Figura 6.5 – Aplicação da Máscara.

6.4 Processamento das Imagens

A combinação da banda HRC com os dados multiespectrais das imagens TM e CCD foram realizadas através da técnica de fusão por transformação inversa no espaço IHS de cores.

O procedimento de transformação RGB para IHS envolveu três etapas de trabalho. A primeira compreendeu a transformação dos componentes R (*red*, vermelho), G (*green*, verde) e B (*blue*, azul) na composição 3R4G2B do TM para os componentes I (*intensity* – intensidade, representa o brilho da imagem),

H (*hue* – matiz, representa a cor propriamente dita) e S (*saturation* – saturação, representa a pureza da cor) a partir de três bandas do sistema óptico. Na segunda etapa realizou-se a substituição do componente I pela banda HRC da imagem TM. Por fim, a última etapa do processo de fusão foi a transformação inversa IHS→RGB. O mesmo procedimento foi realizado para a fusão CCD-HRC.

A substituição da banda HRC pela componente I é possível uma vez que esta componente não está correlacionada com as demais, o que não acontece com H e S, que estão estatisticamente correlacionadas.

A operação de inversão das componentes do espaço IHS para o RGB é necessária; com a aplicação de uma transformação inversa, volta-se às coordenadas do espaço RGB, obtendo-se assim as bandas híbridas, as quais possuem a resolução espacial de 2,5m, e características espectrais semelhantes às bandas 2, 3 e 4 originais dos sensores TM e CCD.

Com isso, obtém-se uma composição colorida que tenta manter a informação espectral das imagens multiespectrais e a informação espacial da banda pancromática, gerando, como produto final, uma imagem híbrida. As Figuras 6.6 e 6.7 mostram o resultado da fusão.

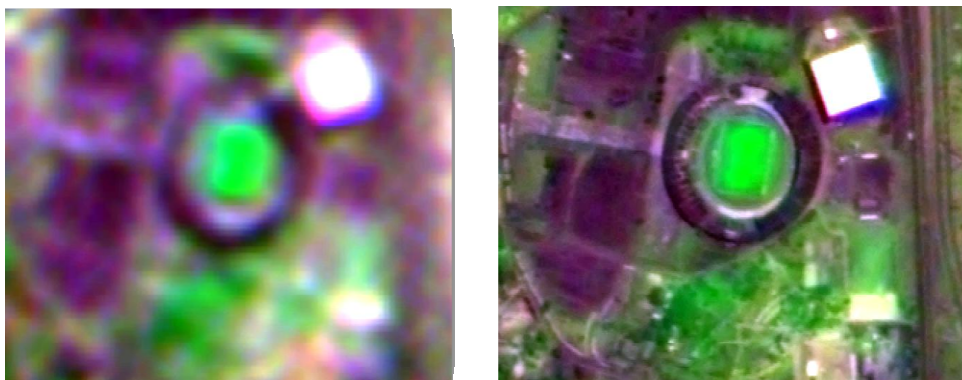


Figura 6.6 – Estádio Serra Dourada. Composição 3R4G2B. CCD HIBRIDA.



Figura 6.7 – Aeroporto de Goiânia. Composição 3R4G2B. TM HIBRIDA.

A próxima etapa na metodologia é gerar uma máscara que identifique apenas as áreas verdes nas imagens híbridas. Novamente foi realizada a transformação de cores do espaço IHS, só que agora com a intenção de utilizar a componente Matiz (H) resultante da transformação da composição 3R4G2B da imagem HIBRIDA TM para identificar a vegetação na imagem. Nesta etapa foi utilizada apenas a imagem HIBRIDA TM porque pela hipótese inicialmente levantada esta imagem deveria conter mais áreas e maior resposta da vegetação por ser da época chuvosa. A Figura 6.8 apresenta a imagem Matiz onde se observa em tons claros a predominância da vegetação.

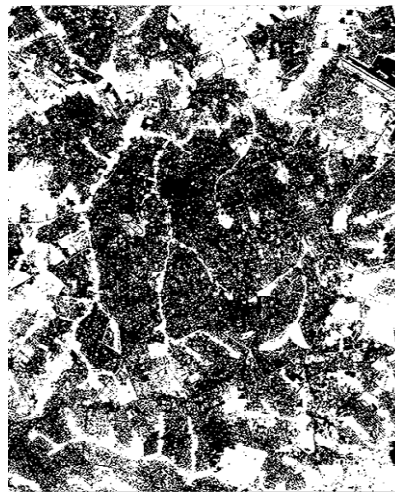


Figura 6.8 – Imagem Matiz identificando a vegetação.

Para identificar e separar os alvos de vegetação dos alvos não-vegetação, a imagem Matiz foi classificada em duas classes: “vegetação” e “não-vegetação”. Um método simples de classificação, quando se tem o limite entre classes bem definidos é o classificador Isepeg que é um algoritmo de agrupamento de dados não-supervisionado, aplicado sobre o conjunto de regiões, que por sua vez são caracterizadas por seus atributos estatísticos de média, matriz de covariância, e também pela área.

Um algoritmo de "clustering" não supõe nenhum conhecimento prévio da distribuição de densidade de probabilidade dos temas. É uma técnica para classificação que procura agrupar regiões, a partir de uma medida de similaridade entre elas.

Para a verificação da qualidade do dado gerado foi feito o cruzamento da classificação com uma imagem QuickBird para validar o método. Desta maneira foi gerada a máscara de acordo com o resultado da classificação.

O procedimento seguinte foi produzir as imagens NDVI de acordo com as máscaras obtidas pela imagem Matiz. Foi gerada uma imagem para cada época, NDVI CHUVOSA e NDVI SECA. De acordo com a premissa do trabalho, as imagens NDVI de épocas que se diferenciam principalmente pela precipitação devem ter aspectos diferentes, portanto, uma subtração entre estas imagens deverá apresentar as diferenças que eventualmente possam existir. O resultado desta operação é apresentado na Figura 6.9.

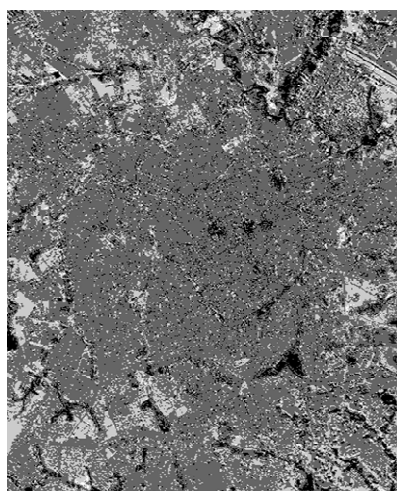


Figura 6.9 – Subtração de NDVI. Época chuvosa e época seca.

A imagem resultante da subtração das imagens NDVI foi classificada para que se obtivesse um reconhecimento melhor dos padrões e objetos homogêneos.

O método adotado nesta etapa do trabalho foi a classificação supervisionada de Distância Mínima que permite o treinamento e assim a separabilidade dos elementos de cena se apresenta mais coerente. Para tal procedimento foram tomadas 2 classes: Alta Mudança (valores mais altos na imagem subtração) e Pouca Mudança (valores mais baixos na imagem subtração).

O Método de Distância Mínima é considerado o classificador mais simples e é indicado quando o tamanho dos conjuntos de treinamento for pequeno. Este algoritmo vai associar cada pixel desconhecido à classe cuja média está mais próxima a ele. As classes podem ser definidas com base em dados estatísticos, calculando-se a média de cada uma, em cada banda espectral.

Neste processo, foram identificadas na imagem as classes de informações de interesse, as quais foram separadas em regiões para que melhor fossem representadas. Dentro destas regiões separaram-se as áreas de treinamento onde se encontram os pixels que foram usados como representantes das mesmas. A partir destes pixels são extraídas determinadas características, que serão usadas para a classificação dos demais pixels. O resultado obtido é demonstrado na figura 6.10.

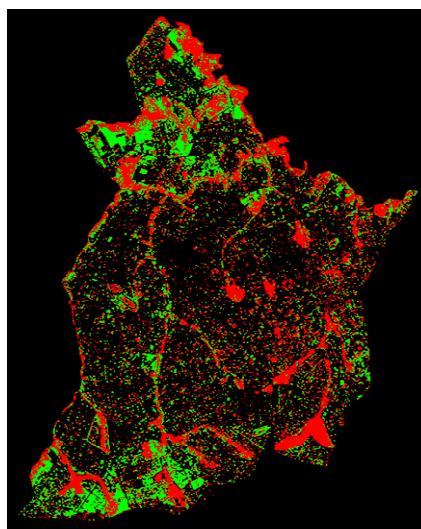


Figura 6.10 – Resultado da classificação pelo método Distância Mínima.

O último procedimento que compreendeu a etapa de processamento de imagens foi a geração dos vetores em função das classes separadas na classificação. Este procedimento é conhecido como vetorização e gera polígonos a partir de cada célula classificada na imagem. Cada classe gerou um conjunto de polígonos com uma única identificação. Estes vetores foram exportados em formato shape (formato padrão de dados vetoriais) como 2 arquivos: Alta Mudança (AM) e Pouca Mudança (PM). (Figura 6.11).

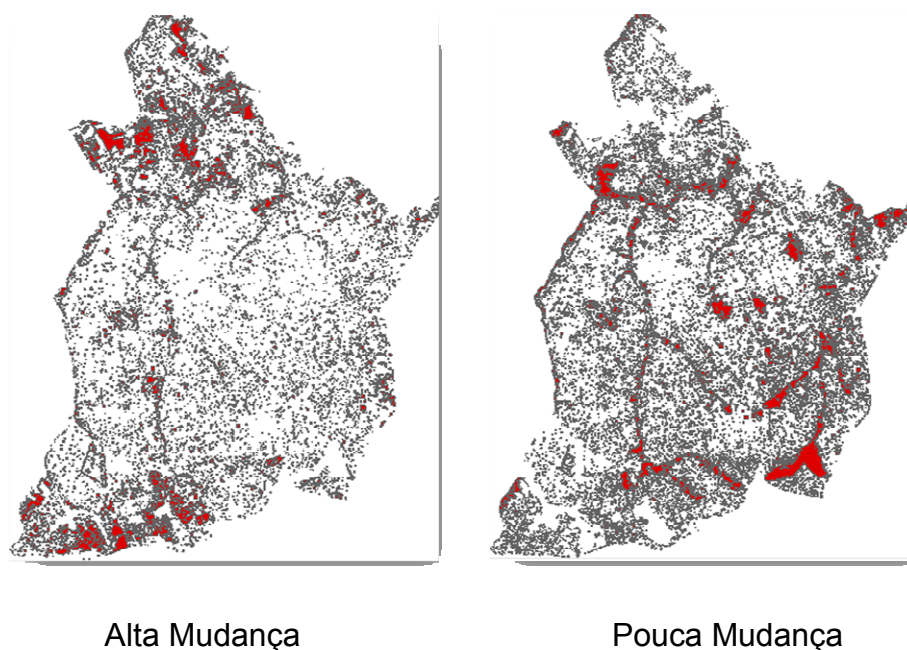


Figura 6.11 – Arquivos vetoriais gerados.

A metodologia seguinte é apresentada no fluxograma (Figura 6.12) e mostra o cruzamento de dados no SIG.

A base de dados está na mesma projeção cartográfica dos arquivos gerados pelo processamento de imagens: UTM, SAD-69. O vetor de MACROZONAS é primeiramente editado para conter apenas as Áreas de Interesse. Este vetor contém a área das regiões. O primeiro cruzamento é a intersecção entre as

Áreas de Interesse e os Distritos Censitários. O resultado é a área por Distrito Censitário. (Figura 6.13).

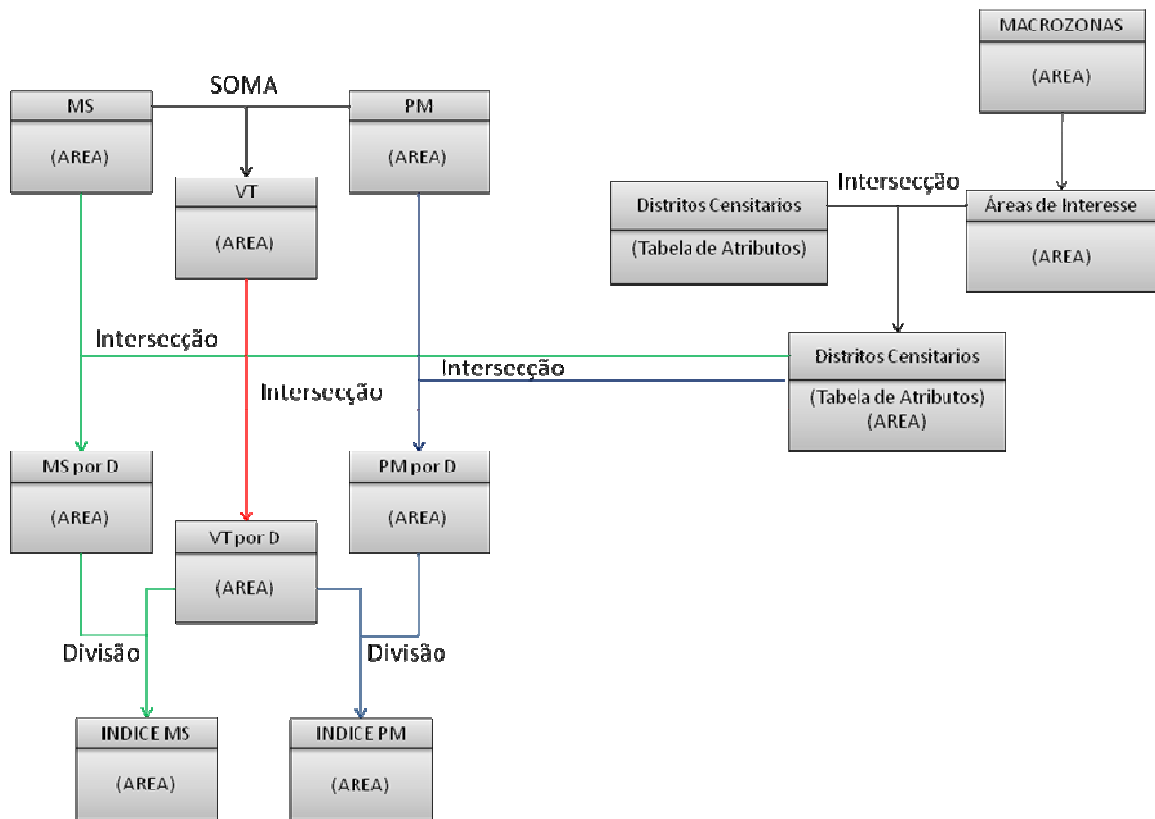


Figura 6.12 – Fluxograma metodológico do SIG.

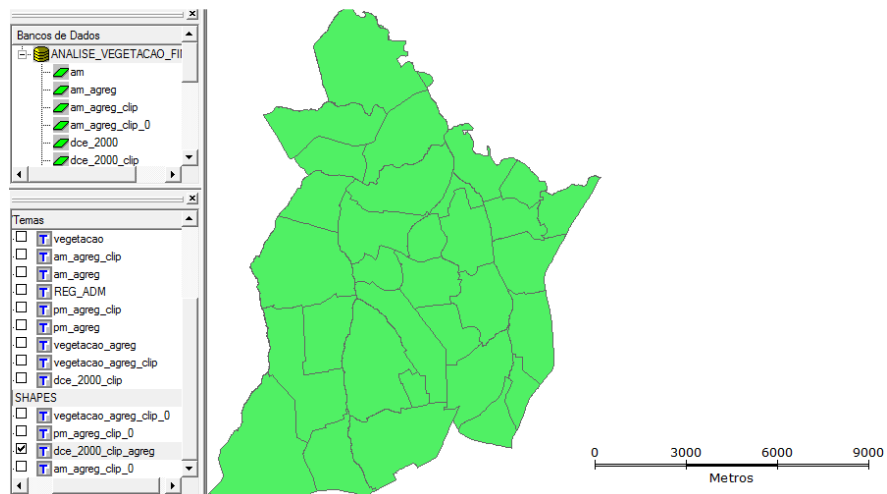


Figura 6.13 – Área por Distrito Censitário.

Na outra ponta do fluxograma é feita a soma dos dados de PM e MS para que obtenha a Vegetação Total (VT). É feita mais uma intersecção entre os vetores (PM, MS, VT) e os Distritos Censitários e deste cruzamento são gerados os vetores de PM, MS e VT por Distrito (Figura 6.14).

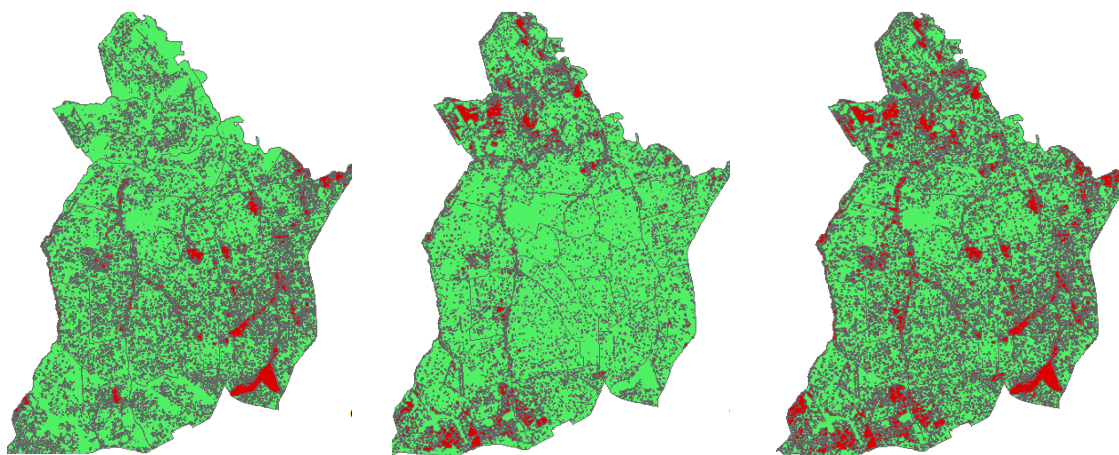


Figura 6.14 – PMporD, AMporD e VTporD.

O último procedimento é realizar a divisão entre os vetores PMporD e MSporD por VporD para gerar os ÍNDICE PM e ÍNDICE AM, respectivamente. Tal procedimento é feito para garantir que nenhum distrito que tenha uma área pequena e muita vegetação seja colocado em igualdade com outro que tenha a mesma quantidade de vegetação, mas em uma área maior.

CAPÍTULO 7 RESULTADOS

Neste capítulo são apresentados os resultados obtidos pela relação dos dados gerados pela metodologia apresentada no capítulo 6 e os dados censitários adquiridos junto ao IBGE. Outra abordagem é avaliar os resultados da classificação quanto a sua eficiência na delimitação da vegetação intra-urbana, assim como em fazer uma análise do estado de conservação da vegetação.

A investigação dos dados procurou apresentar um exemplo de cada tipo de resultado obtido e seus possíveis motivos considerando o contexto em que aparecem. A distribuição de áreas dessas classes pode ser vista na Figura 7.1. A classe “Poucas mudanças” é a que mais aparece no mapa, com 52% da cobertura vegetal. A classe “Alta Mudança” corresponde a 48% do total da cobertura vegetal na área teste.

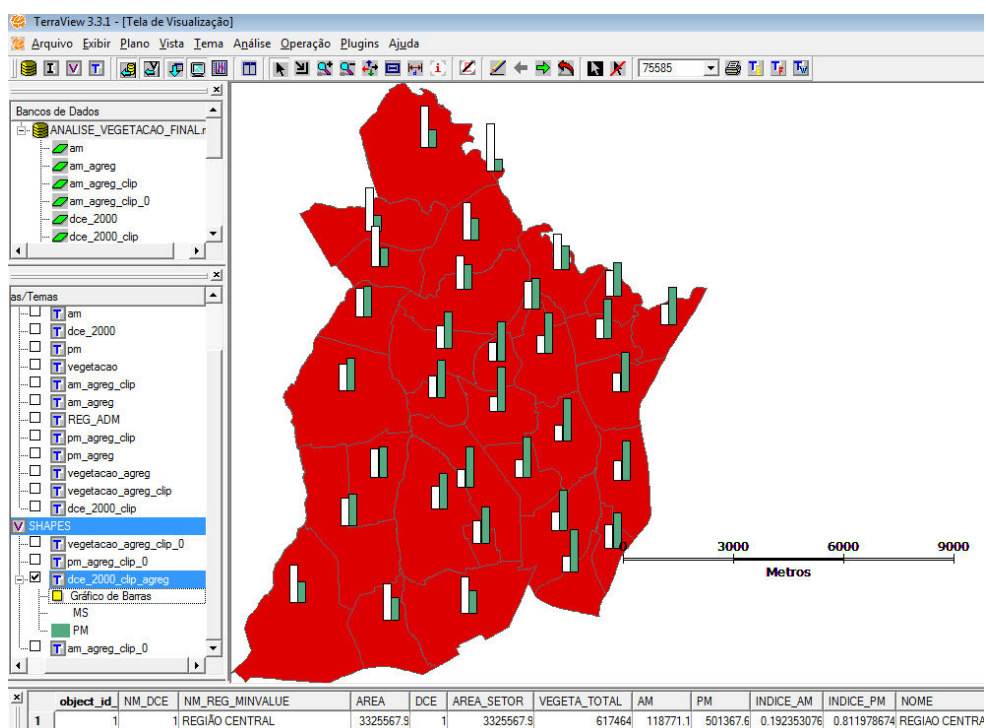


Figura 7.1 – Distribuição Espacial entre áreas de pouca mudança (verde) e alta mudança (branco).

A Figura 7.2 mostra um caso, em que há predominância de áreas com vegetação conservada. Nela, fica evidente o cuidado com a praça e que o sistema retornou como pertencente à classe “Poucas mudanças”.



Figura 7.2 – Área apontada como “Pouca mudança”.

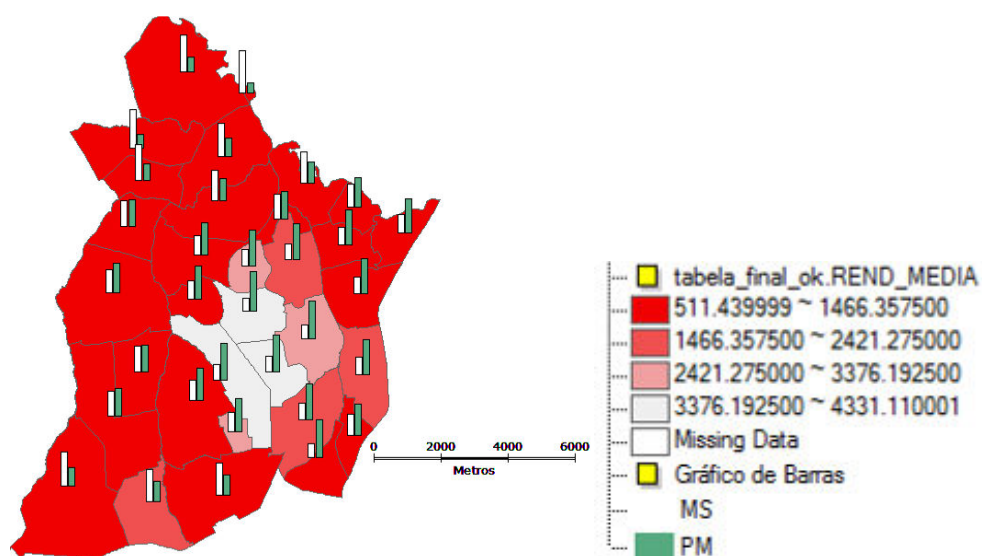


Figura 7.3 – Distribuição Espacial entre áreas de pouca mudança (verde) e alta mudança (branco) e renda.

Neste problema especificamente, a facilidade de manipulação e visualização dos dados em um Sistema de Informações Geográficas, veio corroborar com as premissas definidas por este trabalho. A Figura 7.4 apresenta uma consulta espacial envolvendo a variável AM e renda. Foi perguntado ao sistema quais eram os distritos que apresentavam AM maior que 0,8 e renda menor do que aproximadamente 2 salários mínimos. O sistema retornou como resposta os distritos do Vale do Meia Ponte, o que pelo trabalho de campo acaba por confirmar o processamento do sistema como é apresentado na Figura 7.3. Esta área apresenta canteiros cobertos por vegetação sem cuidado e praças abandonadas.

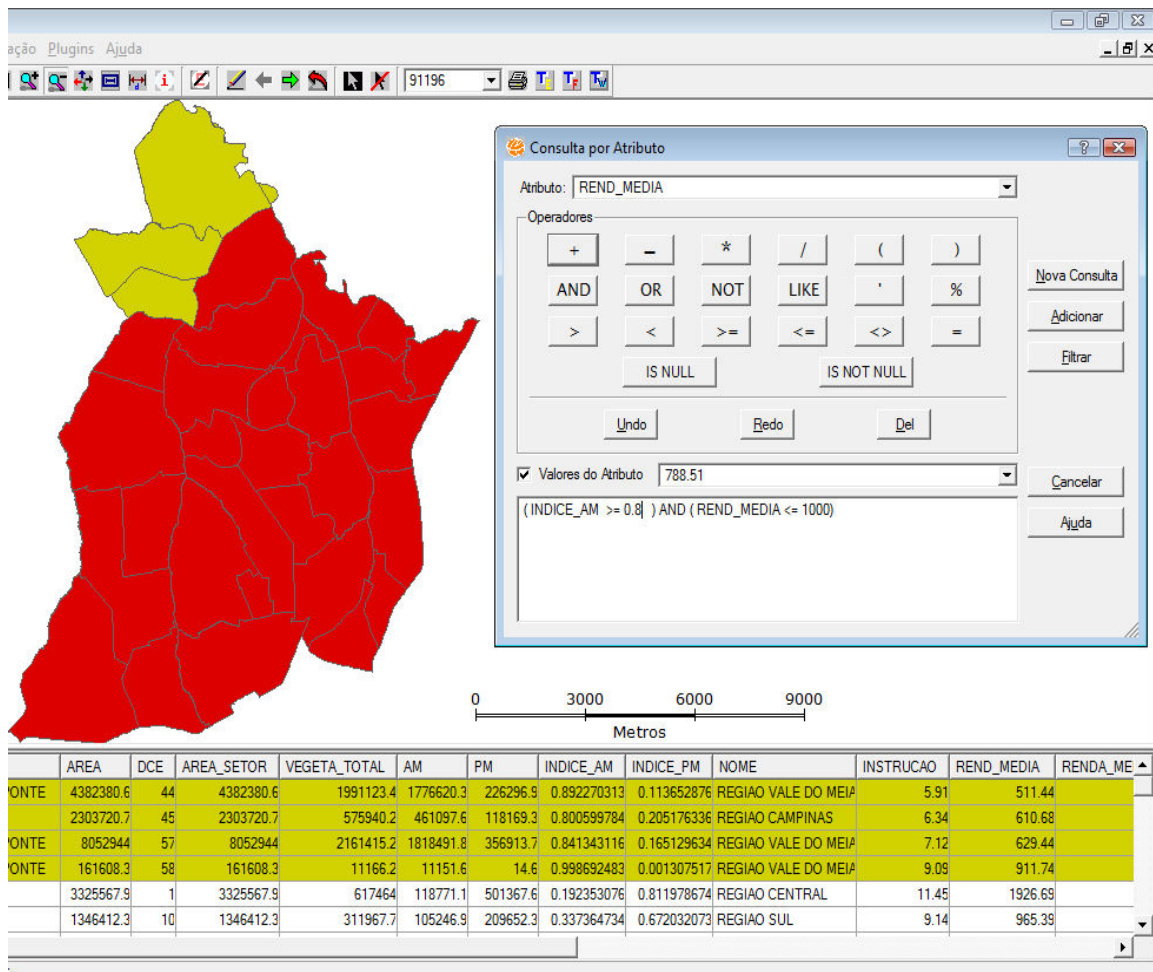


Figura 7.4 – Consulta por atributos.



Figura 7.5 – Áreas de alta mudança. Vale do Rio Meia Ponte.

Outra consulta tinha por finalidade saber os distritos que apresentavam AM

menor que 0,2 e renda maior do que 3500,00 reais como mostra a Figura 7.6. O Setor Central e o Setor Campinas aparecem como distritos de melhor renda e baixa mudança. Este “conforto” da vegetação pode ser confirmado pela Figura 7.7.

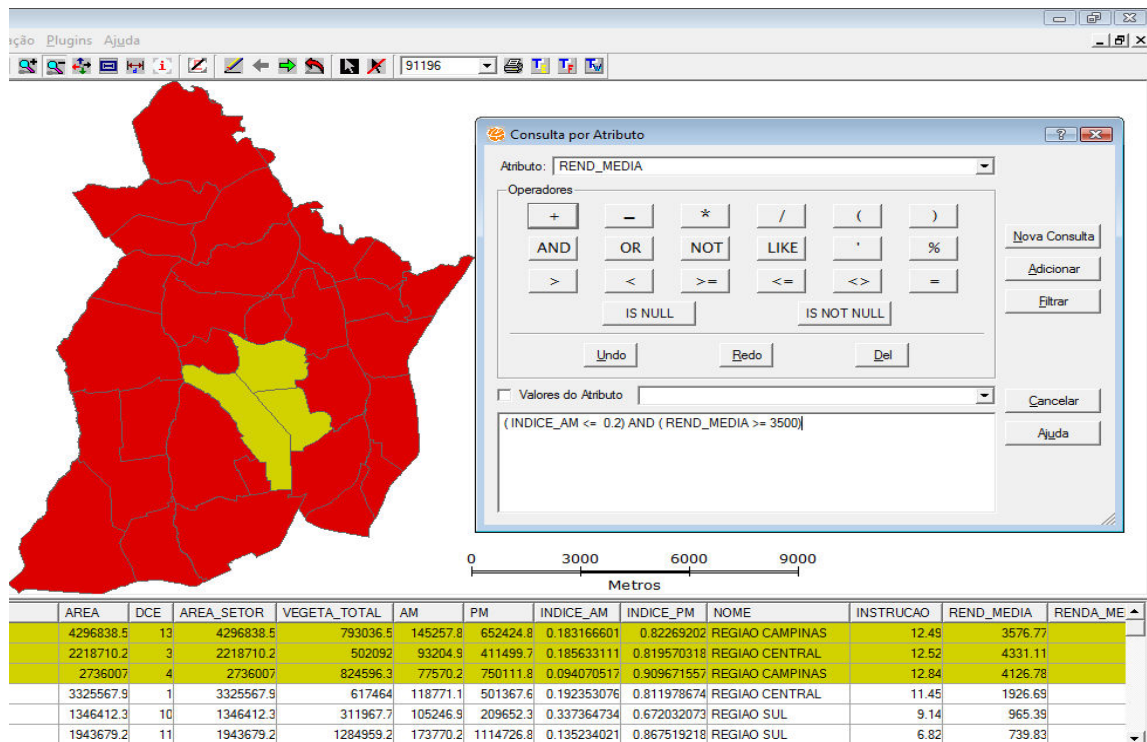


Figura 7.6 – Consulta por atributos.



Figura 7.7 – Áreas de pouca mudança. Centro e Campinas.

Embora inúmeras consultas e análises possam ser feitas, não é objetivo deste trabalho esgotar a discussão. Um último cruzamento será apresentado. Nesta

consulta, os dados de instrução foram agregados em função do número de anos de estudo e depois correlacionado com a medida de AM. É interessante observar que os distritos que apresentam maior índice de instrução, também possuem um índice de AM baixo conforme é apresentado na Figura 7.8.

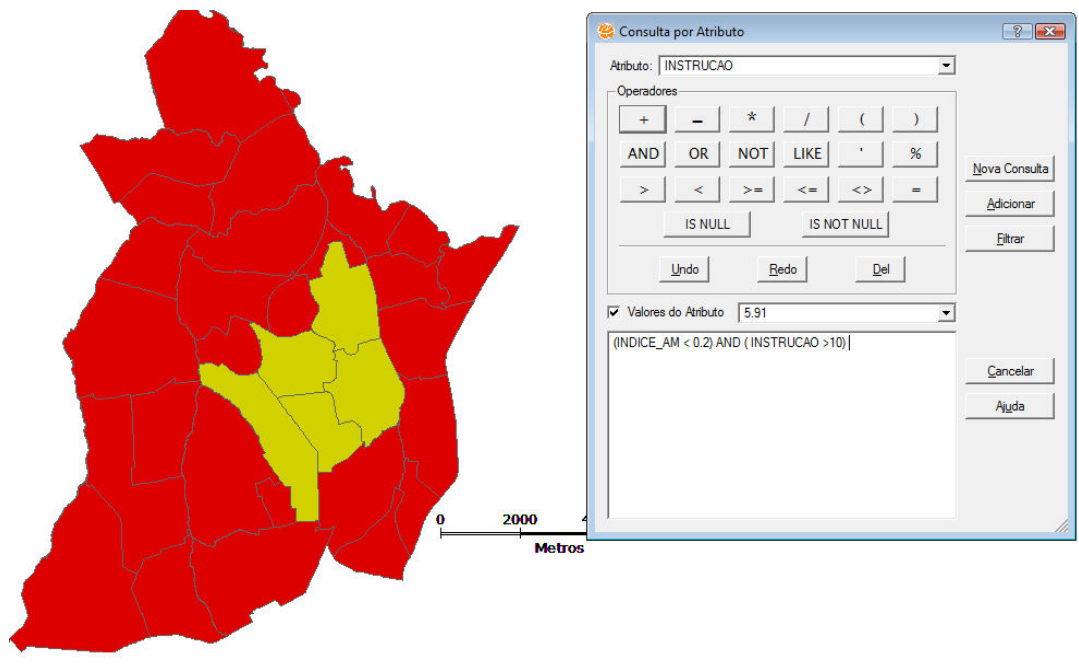


Figura 7.8 – Consulta por atributos.



Figura 7.9 – Áreas de pouca mudança. Centro e Campinas.

Na verdade, este perfil era esperado já que renda média e instrução são altamente correlacionadas. A Figura 7.9 mostra praças e canteiros com índice de AM baixo. Esta foto apresenta um componente adicional já que aponta para o fato de que os moradores também assumem a responsabilidade de

manutenção da vegetação intra-urbana através de associações de bairros que se organizam e prezam o bem comum.

Esta abordagem do problema permite estabelecer relações que descrevem o cenário de estudo de maneira pontual, ou seja, para cada consulta tem-se a resposta para o questionamento levantado e não para todo o conjunto de dados. Desta forma, uma alternativa interessante é transformar os dados em tabelas para que possam ser processadas e utilizar técnicas de mineração de dados para extrair possíveis relações existentes entre as variáveis em discussão.

Para resolver a questão, foram utilizados dois métodos:

1. A tabela gerada é em função dos distritos, ou seja, tem-se 40 distritos e, portanto, 40 registros com atributos socioeconômicos e de mudança da vegetação.
2. A tabela gerada é em função de um espaço celular, com resolução de 20 metros, gerado apenas nas áreas extremas de menor e maior alteração na vegetação.

Em ambos os métodos, a idéia era estabelecer possíveis associações entre as variáveis em discussão. Para tanto, um primeiro passo foi discretizar os valores de entrada para que o algoritmo de associação pudesse gerar as regras.

O software utilizado foi o WEKA que permite utilizar diversas técnicas de mineração de dados. Entre elas, o algoritmo de associação APRIORI foi o escolhido.

Pelo método 1, os dados socioeconômicos disponibilizados sob forma de valores numéricos foram discretizados conforme o Quadro 7.1 abaixo:

Quadro 7.1 – Discretização dos atributos pelo método 1.

<i>ATRIBUTO</i>	<i>VALORES</i>		
<i>INSTRUCAO</i>	ATE_1_GRAU	2_GRAU	SUPERIOR

<i>REND_MEDIA</i>	ATE_2_SAL	ENTRE_3_6_SAL	ENTRE_3_6_SAL
<i>INDICE_LIXO</i>	ATE_90	ENTRE_91_95	ENTRE_96_99 TOTAL
<i>IDADE</i>	ATE_20_ANOS	ENTRE_21_25	ENTRE_26_29 MAIOR_QUE_30
<i>DENSIDADE_DOM</i>	BAIXA	MEDIA	ALTA
<i>DENSIDADE_POP_DOM</i>	BAIXA	MEDIA	ALTA
<i>DENSIDADE_POP</i>	BAIXA	MEDIA	ALTA
<i>INDICE_PON</i>	BAIXO	MEDIO	ALTO
<i>DRENAGEM</i>	TRUE		FALSE
<i>UC_INSIDE</i>	TRUE		FALSE
<i>APP_INSIDE</i>	TRUE		FALSE
<i>VEGETA</i>	BAIXA_MUDANCA		ALTA_MUDANCA

Esta separação foi realizada em função da natureza dos dados. Os valores numéricos foram fatiados em intervalos que pudessem conter significativamente o conjunto de dados a ser representado.

De maneira análoga, o método 2 utilizou uma tabela de valores numéricos discretizados conforme o Quadro 7.2:

Quadro 7.2 – Discretização dos atributos pelo método 2.

ATRIBUTO	VALORES			
<i>INSTRUCAO</i>	ATE_1_GRAU	2_GRAU	SUPERIOR	
<i>REND_MEDIA</i>	1_SAL	2_SAL	ENTRE_8_10_SAL	MAIS_QUE_10_SAL
<i>INDICE_LIXO</i>	PARCIAL		TOTAL	

<i>IDADE</i>	ATE_25_AN OS	ENTRE_26_3 0	ENTRE_31_34	MAIS_QUE_35
<i>DENSIDADE_POP_DOM</i>	BAIXA		MEDIA	ALTA
<i>DENSIDADE_POP</i>	BAIXA			ALTA
<i>INDICE_PON</i>	BAIXA	MEDIA	ALTA	
<i>VEGETA</i>	BAIXA_MUDANCA			ALTA_MUDANCA

A diferença na discretização dos valores entre os dois métodos foi proposta para permitir uma melhor avaliação dos fatiamentos, obviamente, baseados em critérios subjetivos. Como não é abordagem do trabalho comparar qual o melhor método aplicado, e sim, como melhor ajustar os dados para que sejam processados, esta diferenciação torna-se uma alternativa interessante.

O processamento no WEKA é apresentado nas Figuras 6.10 e 6.11:

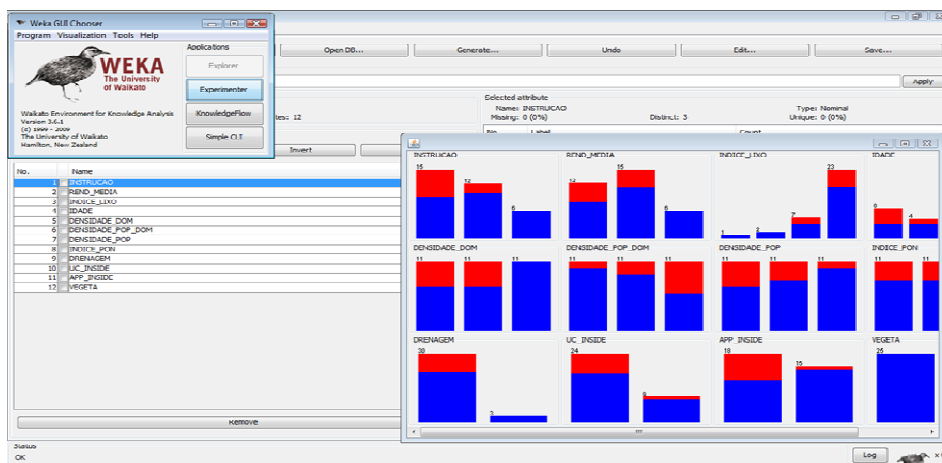


Figura 7.10 – Processamento no WEKA.

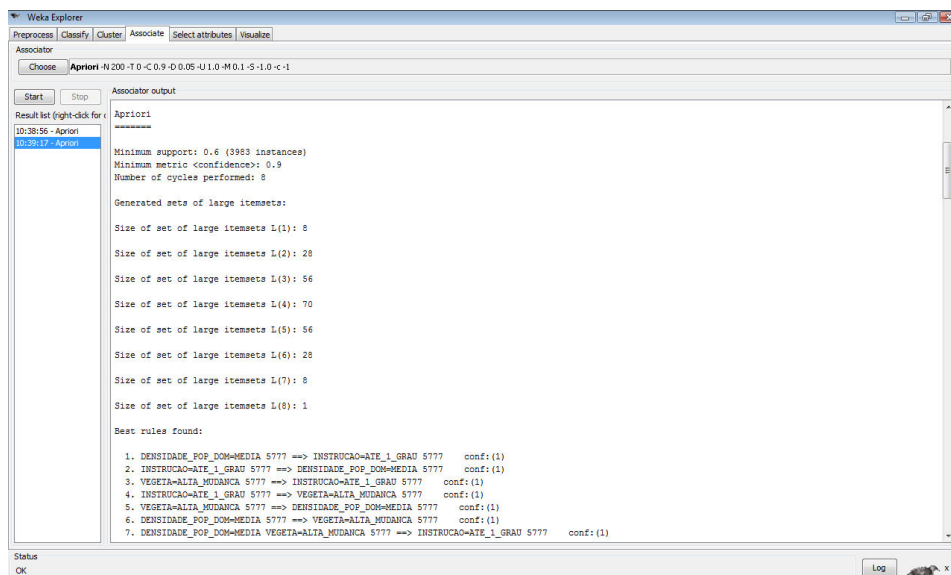


Figura 7.11 – Processamento no WEKA.

Os resultados mais relevantes obtidos com o algoritmo APRIORI são apresentados a seguir. Para ambos os métodos foram geradas 300 regras de associação.

Método 1:

- VEGETA=ALTA_MUDANCA ==> DRENAGEM=TRUE e APP_INSIDE=TRUE
- UC_INSIDE=TRUE e VEGETA=ALTA_MUDANCA==> INSTRUCAO=ATE_1_GRAU e DRENAGEM=TRUE
- INSTRUCAO=SUPERIOR e REND_MEDIA=MAIOR_6_SAL ==> INDICE_LIXO=TOTAL e IDADE=MAIOR_30 e VEGETA=BAIXA_MUDANCA
- INSTRUCAO=2_GRAU e DENSIDADE_POP_DOM=BAIXA e DRENAGEM=TRUE ==> REND_MEDIA=ENTRE_3_6_SAL e INDICE_LIXO=TOTAL e UC_INSIDE=TRUE VEGETA=BAIXA_MUDANCA
- INSTRUCAO=2_GRAU e REND_MEDIA=ENTRE_3_6_SAL e DENSIDADE_POP_DOM=BAIXA==> INDICE_LIXO=TOTAL e DRENAGEM=TRUE e UC_INSIDE=TRUE e VEGETA=BAIXA_MUDANCA

Pelo Método 2, os resultados mais relevantes são apresentados a seguir.

- INSTRUCAO=ATE_1_GRAU==> VEGETA=ALTA_MUDANCA
- DENSIDADE_POP_DOM=MEDIA ==> VEGETA=ALTA_MUDANCA
- REND_MEDIA=1_SALARIO_MINIMO ==> VEGETA=ALTA_MUDANCA
- REND_MEDIA=1_SALARIO_MINIMO e INDICE_LIXO=PARCIAL==> VEGETA=ALTA_MUDANCA
- REND_MEDIA=1_SALARIO_MINIMO e DENSIDADE_POP_DOM=MEDIA ==> VEGETA=ALTA_MUDANCA

Percebe-se que em ambos os métodos a manutenção da vegetação tem uma relação com a qualidade socioeconômica da região. Nas áreas em que a renda é menor e o grau de instrução também é, as alterações na vegetação são mais freqüentes. O cenário oposto também é verdadeiro. Áreas com maior qualidade socioeconômica apresentam índices de alterações da vegetação intra-urbana menores.

Outras associações aparecem e não são tão intuitivas assim, como por exemplo, a relação entre densidade populacional por domicílio e baixo salário levar a uma condição de alta mudança de vegetação.

CAPÍTULO 8

CONCLUSÃO

Promover o uso e controle adequado dos recursos tecnológicos hoje disponíveis, sem dúvida é um dos maiores desafios neste início de milênio. Entre tantas necessidades imediatas para a sistematização de dados, direta ou indiretamente associadas ao uso de geotecnologias, a busca pela integração de diferentes informações advindas de fontes múltiplas tem tido papel relevante.

O presente trabalho apresentou uma metodologia de integração de SR e SIG, que pudesse ser aproveitada para relacionar variáveis tão díspares como vegetação intra-urbana e dados socioeconômicos.

Foi avaliado que a metodologia é aplicável utilizando para tal, dados e softwares sem nenhum custo. Tanto para a identificação quanto para a classificação do estado de conservação da vegetação os métodos se comportaram da maneira esperada.

Outro ponto a ser observado é a possibilidade de análise exploratória utilizando dados socioeconômicos. A visualização das possíveis relações de variáveis tão distintas como

Alta Mudança da vegetação e Renda Média permite levantar questionamentos a respeito das políticas públicas de manutenção de áreas verdes comuns aos cidadãos. As técnicas de mineração de dados ajudam a corroborar estas hipóteses ao apresentarem o cenário das associações entre as variáveis em discussão.

REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, R., SRIKANT, R. : **Fast Algorithms for Mining Association Rules.**

Proc. 20th Int.

Conf. Very Large Data Bases, VLDB, 1994.

BOGORNY, V. ; ENGEL, PAULO MARTINS ; ALVARES, LUIS OTAVIO .

GEOARM: an Interoperable Framework to Improve Geographic data

Preprocessing and Spatial Association Rule Mining. In: Software

Engineering and Knowledge Engineering (SEKE), 2006, San Francisco.

Proceedings of the Eighteenth International Conference on Software

Engineering and Knowledge Engineering. San Francisco Bay : Knowledge

Systems Institute, 2006. v. 18. p. 79-84.

BIGOLIN, NARA MARTINI ; BOGORNY, V. ; ÁLVARES, LUIS OTÁVIO CAMPOS

. **Uma Linguagem de Consulta para Mineração de Dados em Banco de**

Dados Geográficos Orientado a Objetos. In: XXIV Conferência

LatinoAmericana de Informática, 2003, La Paz. Clei. La Paz : Universidad

Mayor de San Andrés, 2003.

CAPORUSSO, D; MATIAS, L. F. UNICAMP. **Áreas verdes Urbanas: Avaliação**

e Proposta Conceitual. Rio Claro, SP, 2008

CROSTA, A. P. **Processamento Digital de Imagens de Sensoriamento**

Remoto. UNICAMP/IG, Campinas, São Paulo, 1992.

DOMINGOS, P. L. H. **Aplicação de dados de sensoriamento remoto orbital de**

alta resolução à análise da cobertura vegetal intra-urbana e de seu estado

de conservação. 2005. (INPE--TDI/). Dissertação (Mestrado em

Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2005.

FERREIRA, N. C.; NAZARENO, N. X.; MACEDO, F. C. **Avaliação da exatidão cartográfica da ortofoto digital do município de Goiânia-GO**. II Simpósio Brasileiro de Geomática Presidente Prudente - SP, 24-27 de julho de 2007.

FONSECA, L.M.G. Restauração e interpolação de imagens do satélite **Landsat por meio de técnicas de projeto de filtros FIR**. São José dos Campos. 148 p. (INPE-6628-TAE/30). Dissertação (Mestrado em Engenharia Elétrica) - Instituto Tecnológico da Aeronáutica, 1988.

GONZALEZ, R. C. e WOODS, R. E., **Processamento de Imagens Digitais**, Editora Edgard Blücher Ltda, 2000.

IBGE. Dados do censo populacional dos municípios do Brasil. Disponível em <<http://www.ibge.gov.br>>.

MATHER, P.M. **Computer processing of remotely-sensed images**. New York: John Wiley & Sons, 1987. 293p.

MOREIRA, M. A.; SHIMABUKURO, Y. E. Cálculo do índice de vegetação a partir do sensor AVHRR. In: FERREIRA, Nelson Jesus (Ed.). **Aplicações ambientais brasileiras dos satélites NOAA e Tiros-N**. São José dos Campos: Oficina de Textos, 2004. p. 79-101. ISBN 85-86238-35-X. (INPE-11458-PRE/6879).

NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. São Paulo: Edgard Blücher, 1989. 308 p.

SILVA, W. V.; FERREIRA, N. C.; BOGGIONE, G. A. Análise de vetores de crescimento para a quantificação das transformações urbanas no município de Goiânia. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12.

(SBSR), 2005, Goiânia. **Anais...** São José dos Campos: INPE, 2005. p. 681-688. CD-ROM, On-line. ISBN 85-17-00018-8.