



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

MINERAÇÃO DE DADOS ESPACIAIS UTILIZANDO MÉTRICAS DE PAISAGEM

Marcio Azeredo

Relatório da Disciplina de Mineração de Dados

INPE
São José dos Campos
2008

1 INTRODUÇÃO

Nas últimas décadas, tem sido constatada uma intensificação no processo de desmatamento da região amazônica, muitas vezes como consequência de um processo desordenado de ocupação humana. A possibilidade de associação desse tipo de ocupação com determinados padrões de desmatamento é de grande valia para o planejamento e emprego de políticas públicas sócio-ambientais.

Neste contexto, o presente trabalho tem por objetivo o emprego de técnicas de mineração de dados, na tentativa de viabilizar tal associação.

Segundo Ester et al., 1999, Mineração de Dados consiste da busca, automática ou semi-automática, em grandes quantidades de dados com o objetivo de descobrir padrões importantes, utilizando algoritmos. Sendo assim, foram implementadas métricas de paisagem, na linguagem C++, responsáveis por quantificar determinadas características espaciais dos polígonos de desmatamento. Uma vez valoradas tais características, uma árvore de decisão binária é utilizada para classificar todas as subdivisões da área de estudo, tendo como base os padrões pré-estabelecidos.

Vale o registro, que o estudo foi desenvolvido dentro do contexto do projeto GeoDMA (*Geographical Data Mining Analyst*), projeto este que funciona na DPI (Divisão de Processamento de Imagens) e tem por objetivo o desenvolvimento de um produto para a identificação e análise de padrões espaciais.

2 BASE DE DADOS

Para a realização do trabalho foi utilizada a base de dados de desmatamento do PRODES, mais especificamente da região denominada Terra do Meio, localizada no estado do Pará, e que estende-se sobre os territórios de Trairão, Altamira e São Félix do Xingu.

Como o PRODES estima taxa anual de desmatamento em formações florestais na Amazônia Legal, foram utilizadas as bases de desmatamento referentes aos

anos 1997, 2000, 2003 e 2006, de modo a atender necessidades do projeto GeoDMA, .

A evolução do desmatamento durante o período de estudo pode ser observado nas figuras 1 e 2.

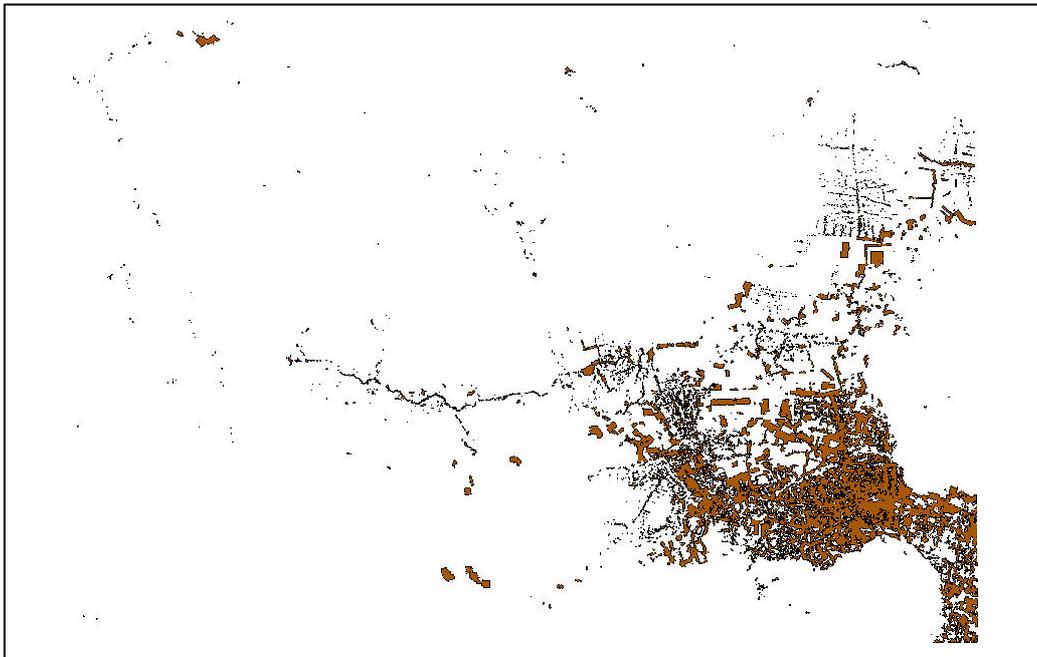


Figura 1 - Polígonos de desmatamento -1997.

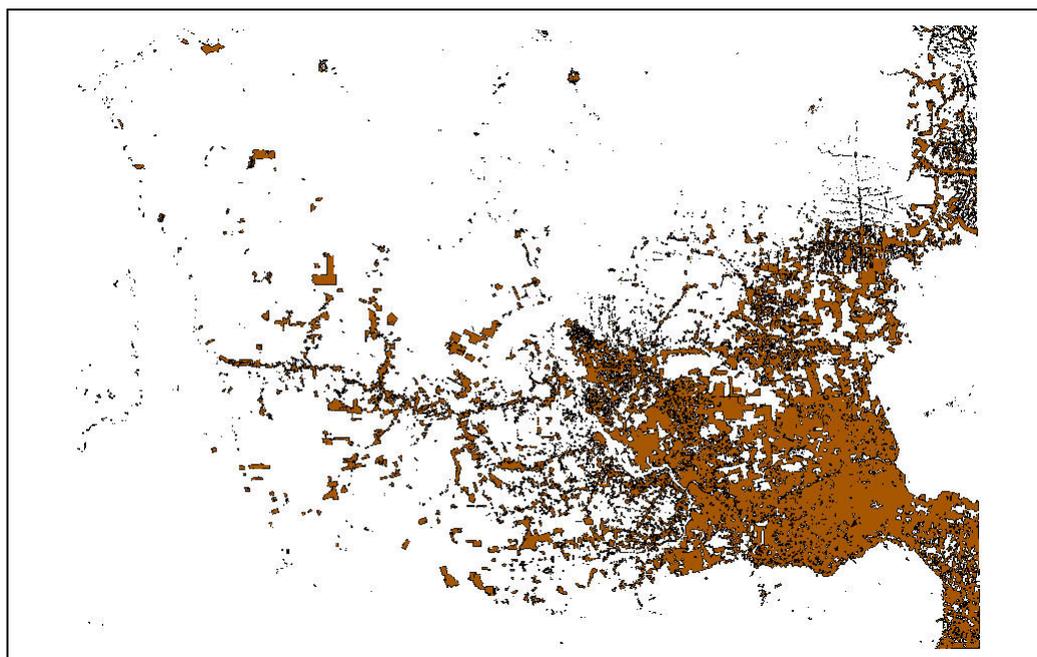


Figura 2 – Polígonos de desmatamento - 2006.

Na tabela 1 é apresentado um comparativo entre a quantidade de polígonos de desmatamento registrados nas bases anuais estudadas.

Tabela 1 – Comparativo entre a quantidade de polígonos (1997-2006).

Ano de Estudo	Quantidade de Polígonos de Desmatamento
1997	3.476
2000	4.066
2003	4.493
2006	5.165

3 METODOLOGIA

Para o desenvolvimento do estudo foi utilizada como base a metodologia proposta por SILVA, 2006.

3.1. DEFINIÇÃO DO TAMANHO DA CÉLULA

Uma preocupação inicial foi a definição da dimensão do tamanho da célula da paisagem a ser estudada. Foram então geradas paisagens de 5km, 10km, 20km e 50km de lado, onde observou-se que, nas de 50km, 20km e 5km, havia uma grande confusão de padrões. Na de 10km, os padrões puderam ser melhor identificados, sendo esta a subdivisão utilizada no trabalho (figura 3).

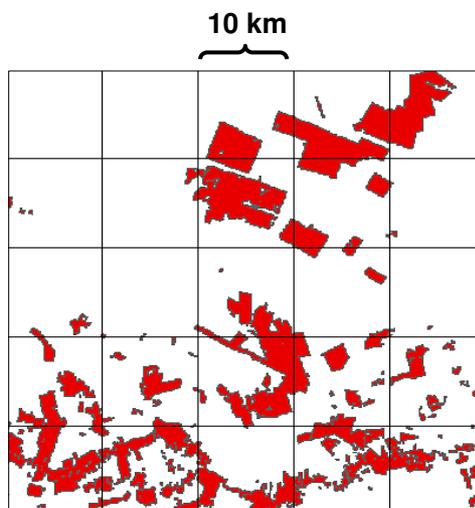
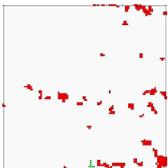
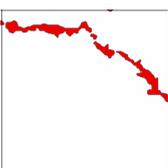
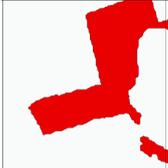
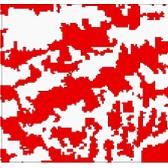
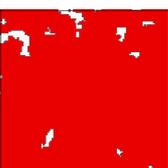


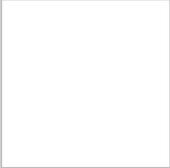
Figura 3 – Célula de paisagem de 100km².

3.2. DEFINIÇÃO DA TIPOLOGIA DE PADRÕES DE OCUPAÇÃO

Foram definidos os tipos de padrão de desmatamento observados nas bases de dados e seus respectivos padrões de ocupação humana associados (tabela 2).

Tabela 2 – Padrões de desmatamento e ocupação humana (ESCADA, 2008).

Métrica	Tipo	Descrição	Semântica
	Difuso	-Manchas pequenas -Manchas Isoladas -Baixa a média densidade -Distribuição uniforme	-Início da ocupação -Ocupação espontânea (não Planejada) -Pequenos produtores rurais. -Ocupação Ribeirinha (+ Rios)
	Linear	Manchas alongadas e contínuas Baixa densidade Unidirecional	Início da ocupação ao longo da estrada Ocupação espontânea (raramente planejada) Pequenos produtores rurais
	Geométrico	Manchas médias a grandes e isoladas Forma geométrica regular Baixa a média densidade	Estágios iniciais de ocupação Médias e Grandes Fazendas
	Multirecional Desordenado	Manchas pequenas a médias que se uniram Formas variadas (irregular, geométrica, linear) Média a alta densidade Multidirecional	Ocupação em expansão, inicialmente espontânea Pode haver concentração fundiária Pequenos e médios produtores rurais
	Consolidado	Manchas grandes e contínuas Densidade baixa e áreas pequenas de remanescentes florestais Manchas compactas e contínuas	Estágios avançados de ocupação Concentração fundiária Pequenos, médios e grandes produtores rurais Esgotamento da floresta Ocupação consolidada

	<p>Floresta</p>	<p>Não há a presença de polígonos desmatamento</p>	<p>Não há registros de ocupação humana</p>
---	-----------------	--	--

3.3. SELEÇÃO DAS MÉTRICAS DE PAISAGEM

Foram selecionadas quais as métricas que seriam aplicadas aos polígonos de desmatamento, de forma a valorar suas características geométricas.

Tabela 3 – Métricas de paisagem utilizadas.

Métrica	Finalidade
Class Area (CA)	Calcula a área total das “manchas” de desmatamento de uma paisagem.
Percent LAND (%LAND)	Calcula a porcentagem de área da paisagem atualmente desmatada.
Patch Density (PD)	Número de “manchas” de desmatamento por km ² da paisagem.
Mean Patch Size (MPS)	Calcula o tamanho médio das “manchas” de desmatamento da paisagem.
Landscape Shape Index (LSI)	Calcula a complexidade da forma das “manchas” de desmatamento com base no perímetro das mesmas, e na área da paisagem.
Mean Shape Index (MSI)	Calcula a média da complexidade da forma das “manchas” de desmatamento com base no perímetro das mesmas, e na área da paisagem.
Area Weighted Mean Shape Index (AWMSI)	Calcula a média da complexidade da forma das “manchas” de desmatamento ponderada pelo tamanho da área das mesmas.
Mean Patch Fractal Dimension (MPFD)	Calcula a média da complexidade da forma das “manchas” de desmatamento utilizando logaritmos.

O resultado da aplicação das métricas é um arquivo no formato de leitura do *software* Weka com os respectivos valores correspondentes à todos os

polígonos. Ao todo, foram gerados quatro arquivos: um para cada ano da base de dados.

3.4. EMPREGO DAS TÉCNICAS DE DATA MINING

A técnica de mineração de dados utilizada no trabalho foi a árvore de decisão binária implementada dentro do *software* Weka.

Foram selecionadas quarenta amostras de treinamento para treinamento da árvore de decisão. Algumas dessas amostras são observadas na figura 4.

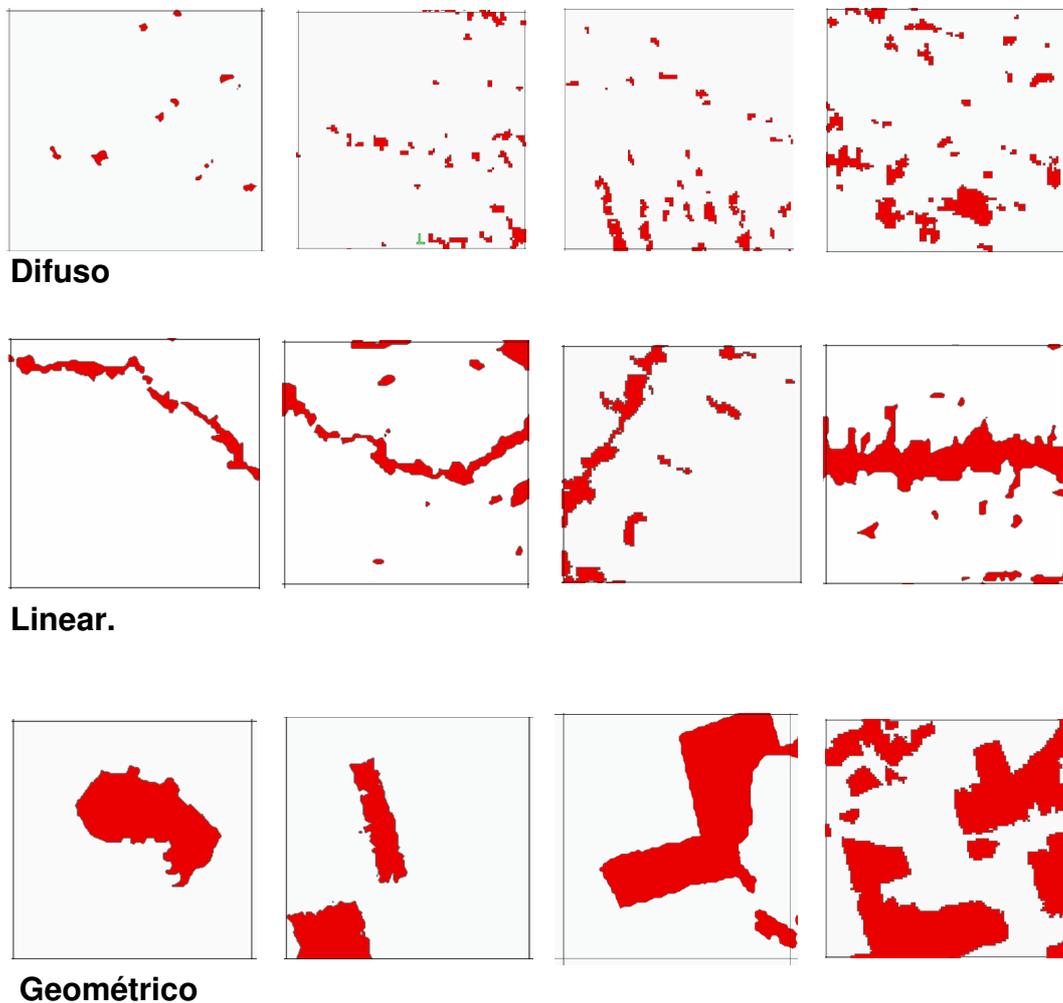


Figura 4 – Exemplos de amostras de treinamento da árvore.

Com o arquivo de amostras já classificadas visualmente, foi aplicado o algoritmo de árvore de decisão J48. A árvore de treinamento gerada é apresentada na figura 5.

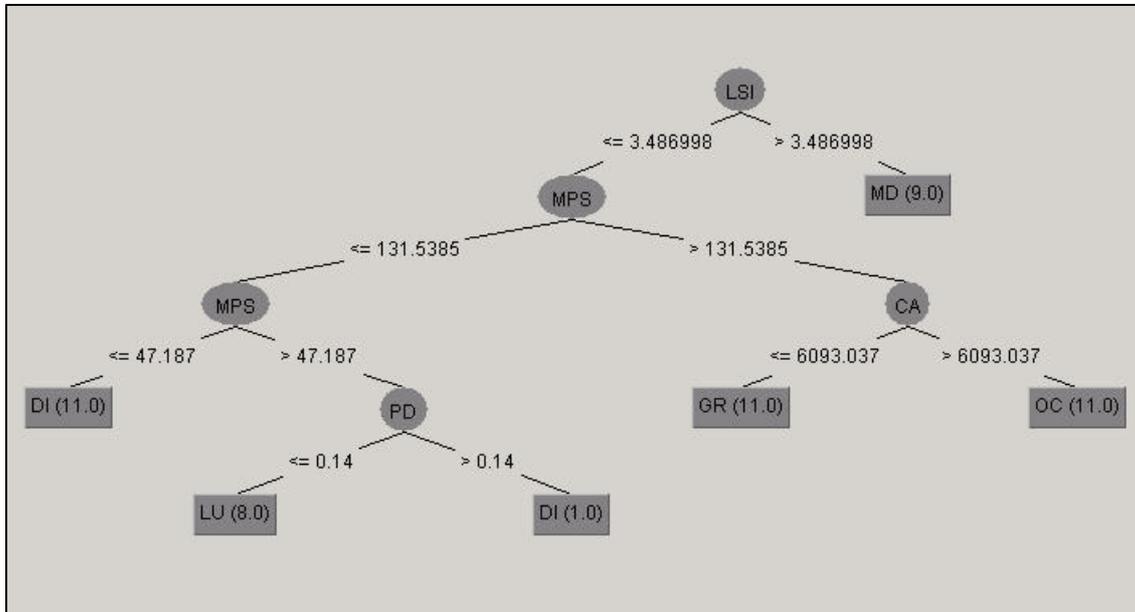


Figura 5 – Árvore de decisão binária gerada pelo *software* Weka.

Uma vez gerada a árvore, a próxima etapa foi sua implementação no algoritmo em C++ desenvolvido no trabalho para a classificação das demais células.

4 RESULTADOS E CONCLUSÕES

Após todas as bases de dados terem sido classificadas, o seguinte resultado foi obtido para o período compreendido entre os anos de 1997 e 2006 (figura 6).

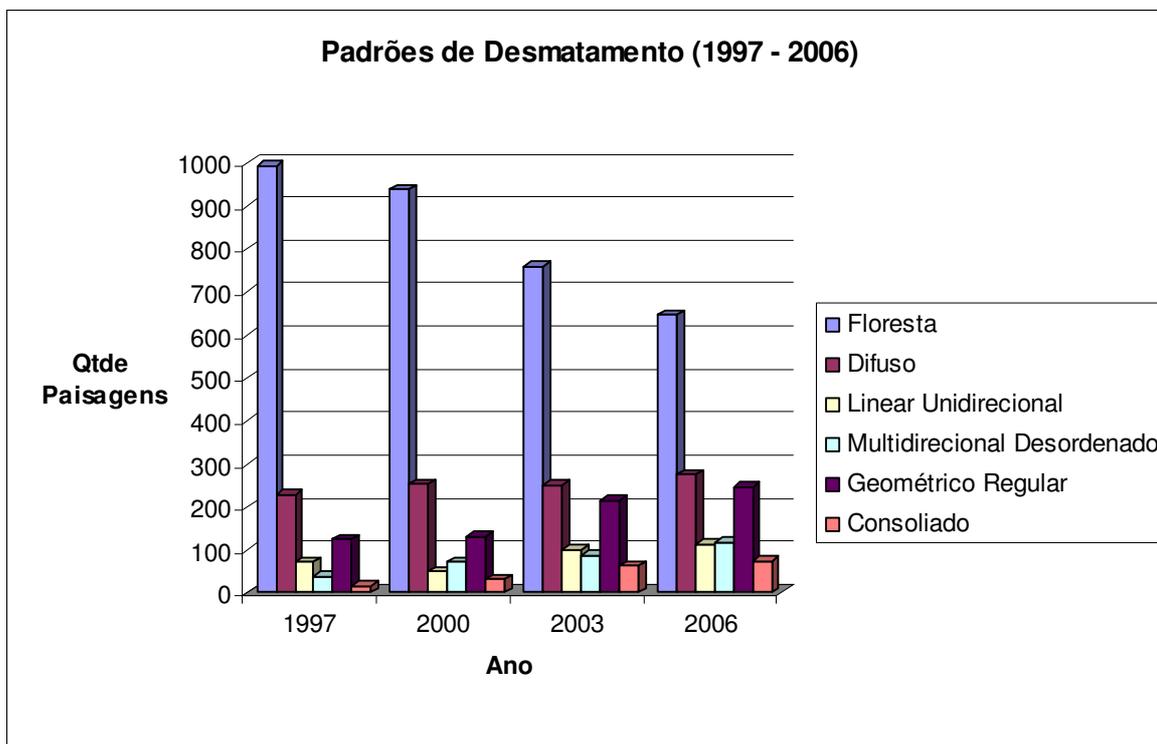


Figura 6 – Classificação das células de paisagem 1997-2006.

O gráfico mostra um aumento significativo de áreas com estágio avançado de desflorestamento (consolidado) em paralelo a uma diminuição de área de floresta sem indícios de ocupação. Uma vez realizada a classificação, foram gerados mapas com os resultados para cada ano de estudo. Na figura 7 é possível observar o resultado obtido para o ano de 1997.

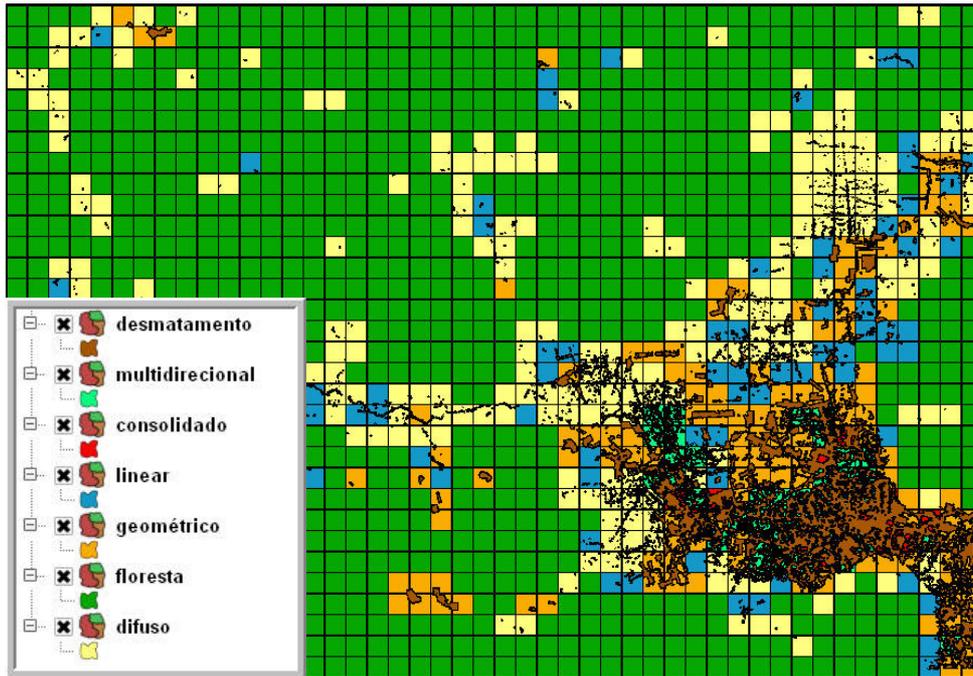


Figura 7 – Vista geral da classificação das células -1997.

De uma maneira geral, todas as tipologias definidas foram classificadas. Uma visão ampliada é apresentada na figura 8.

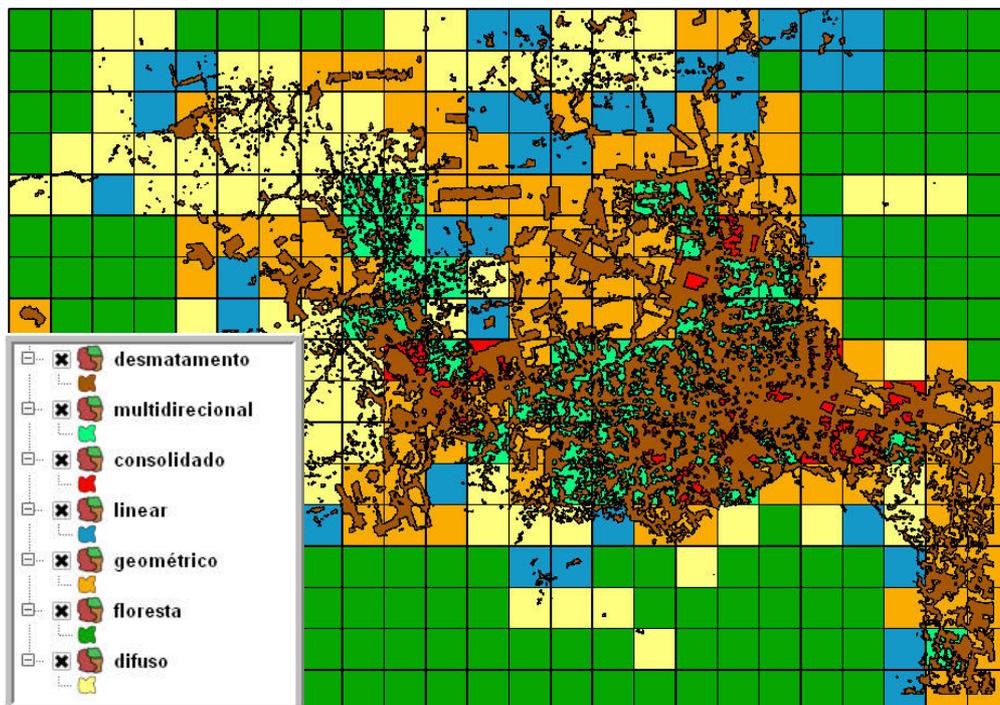


Figura 8 – Vista ampliada da classificação das células (1997).

Analisando o resultado da classificação foi possível identificar que a tipologia linear apresentou o pior resultado entre as tipologias definidas. Mesmo assim o resultado foi considerado satisfatório. Tal confusão é aceitável em alguns casos, pois se trata de configurações de paisagens limítrofes.

Como sugestão para trabalhos futuros, é interessante a implementação de novas métricas de paisagem que possam obter resultados mais refinados; a criação de uma tipologia para associação com outros tipos de padrões (ex: padrões de exploração e uso do solo); e ainda um estudo evolutivo dos padrões de desmatamento utilizados no trabalho ao longo de todo o período estudado.

O objetivo do trabalho foi considerado alcançado, pois com as métricas implementadas em conjunto com a utilização técnicas de mineração de dados, foi possível classificar padrões de desmatamento e associá-los à padrões de ocupação humana.

5 REFERÊNCIAS BIBLIOGRÁFICAS

Silva, M. P. S.; Câmara, G.; Escada, M. I.; Souza, R. C. M. 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.

Silva, M. P. S. Mineração de Padrões de Mudança em Imagens de Sensoriamento Remoto. Tese - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006.

de Oliveira Filho, F. J. B.; Metzger, J. P. Thresholds in Landscape Structure for Three Common Deforestation Patterns in the Brazilian Amazon. *Landscape Ecology*, n. 21, p. 1061–1073, 2006.

Frohn, R. C.; Hao, Yongping. Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil. *Remote Sensing of Environment*, v. 100, n. 2, p. 237-251, 2006.

da Silva, F. C.; Korting, T. S.; Fonseca, L. M. G.; Escada, M. I. S. Deforestation Pattern Characterization in the Brazilian Amazonia. XIII Simpósio Brasileiro de Sensoriamento Remoto. p. 6207 - 6214, 2007.