

## **Análise comparativa de técnicas de classificação de imagens do sensor Landsat/TM para caracterização de áreas desflorestadas**

DIÓGENES S. ALVES, JORGE L.G. PEREIRA, CLAUDIA L. DE SOUSA, JOÃO V. SOARES,  
JOSÉ C. MOREIRA, JUSSARA O. ORTIZ, YOSIO E. SHIMABUKURO, FERNANDO YAMAGUCHI

Instituto Nacional de Pesquisas Espaciais (INPE)  
Av. dos Astronautas 1758, C.P. 515, CEP 12201-970  
São José dos Campos, SP, Brazil

**Abstract:** Five classification algorithms - image thresholding, unsupervised and supervised region classifiers, maximum likelihood classification of vegetation, soil and shade bands, and of TM bands 3, 4 and 5 - were tested for classification of TM imagery for mapping of deforested areas. Deforestation included pasture, annual and perennial plantations and secondary vegetation. All five algorithms resulted in excellent levels of agreement with a reference map of deforestation ( $\kappa > 0.8$ ), showing that digital classification can be feasible for deforestation surveys.

**Keywords:** image classification, image segmentation, deforestation, Amazônia

### **1 Introdução**

Levantamentos baseados em imagens dos sensores Landsat *Multispectral Scanner Subsystem* (MSS) e *Thematic Mapper* (TM) tem permitido acompanhar o aumento das áreas desflorestadas na Amazônia Legal nas últimas décadas (Tardin *et al.* 1980, INPE 1998). O uso de imagens de satélite também tem possibilitado o estudo do desflorestamento em outras áreas do planeta, particularmente em regiões tropicais (*e.g.* Chomentowski *et al.* 1994), onde extensões significativas de florestas não poderiam ser monitoradas facilmente sem o recurso aos satélites de sensoriamento remoto.

A metodologia utilizada nos levantamentos de áreas desflorestadas realizados pelo INPE inclui duas etapas principais: a interpretação visual de reproduções fotográficas de imagens do sensor TM, seguida da elaboração de mapas digitais a partir dos originais interpretados (Alves *et al.* 1992). A área de estudo corresponde a aproximadamente 4.000.000 km<sup>2</sup>, sendo que um levantamento exige a análise de até duas centenas de cenas. A elaboração dos mapas digitais passa por uma série de etapas que incluem cópia em nanquim dos originais de interpretação, a digitalização dos originais copiados em nanquim e a produção de mapas em *plotters* para verificação e controle de qualidade. Essas

etapas tornam o trabalho complexo e caro, sendo necessária a busca de novas metodologias para sua elaboração.

Técnicas de classificação digital de imagens do sensor TM estão sendo utilizadas com sucesso no levantamento das áreas desflorestadas na região tropical do mundo inteiro conforme descrito por Chomentowski *et al.* (1994) e num sistema para o monitoramento de áreas críticas desenvolvido no Instituto Brasileiro de Recursos Naturais e Renováveis (IBAMA) (C. Paiva comunicação pessoal 1997). Em particular, essas técnicas podem contribuir para:

- (1) Reduzir o custo dos levantamentos da área desflorestada.
- (2) Utilizar imagens de diversos sensores e satélites de forma combinada.
- (3) Criar um sistema de informações reunindo imagens de satélite, mapas de áreas desflorestadas e dados relacionados.

Este trabalho pretende contribuir para um melhor entendimento do uso de técnicas de classificação digital em levantamentos de áreas desflorestadas. Ele analisa o desempenho de cinco técnicas diferentes de classificação de imagens do sensor TM para a identificação de áreas desflorestadas. Os classificadores utilizados foram: (1) fatiamento ou limiarização (Research Systems 1997) da banda TM5, (2) classificação não supervisionada de regiões utilizando as bandas TM3, TM4 e TM5, (3) classificação supervisionada de regiões utilizando as bandas TM3, TM4 e TM5 (<http://www.dpi.inpe.br/spring/usuario>, outubro de 1997), classificação pelo algoritmo de máxima verossimilhança (4) das bandas de proporções dos componentes vegetação, solo e sombra produzidas por combinações lineares das bandas TM3, TM4 e TM5 (Shimabukuro *et al.* 1997) e (5) das bandas TM3, TM4 e TM5. As técnicas foram avaliadas em uma região de Rondônia em que foram implantados projetos de agricultura e pecuária de tamanhos variando entre 25 e 1000 ha (INCRA 1996), compondo um mosaico relativamente complexo de padrões de uso e cobertura da terra.

## **2 Área de Trabalho**

A área de trabalho pertence à cena do sensor TM de índice WRS/TM 231/67 e inclui parte dos municípios de Ariquemes, Machadinho d'Oeste, Rio Crespo e Theobroma, na região central do

estado de Rondônia. Suas dimensões correspondem a uma área de 333.000 hectares ou uma imagem de 1850x2000 pixels na resolução de 30 metros do sensor TM (**Figura 1**).

A área é ocupada por propriedades de porte médio (1.000 hectares) e pequeno (25-100 hectares), pertencentes a projetos implantados na região a partir da década de setenta (INCRA 1996). As propriedades maiores são dedicadas em sua maior parte à pecuária, enquanto as menores pertencem a projetos de assentamento, onde podem ser encontradas culturas anuais e perenes e, também, pecuária. A disposição das propriedades em meio à floresta e os diferentes tipos de culturas, de gramíneas e de manejo dos pastos compõem um mosaico relativamente complexo de padrões de uso e cobertura da terra (**Figura 1**), que poderia ser considerado um interessante desafio para o uso de técnicas de classificação automática. A cobertura vegetal nativa da região é na sua maioria de florestas tropicais densas ou abertas, com relevo levemente ondulado (RADAMBRASIL, 1978). Algumas áreas apresentam características de difícil tratamento por classificação automática, como os afloramentos rochosos em que ocorre uma vegetação pouco desenvolvida, e o relevo mais movimentado da porção sudoeste da imagem (**Figura 1**).

### **3 Materiais e métodos**

O trabalho utilizou as bandas TM3, TM4 e TM5 da imagem de índice WRS/TM 231/67 de 3 de agosto de 1995. As bandas foram lidas de um CD e georeferenciadas utilizando-se pontos de controle adquiridos e processados com equipamento GPS diferencial. Em seguida a imagem foi classificada conforme descrito abaixo.

O objetivo da classificação foi o de discriminar duas classes – floresta e área desflorestada – na imagem da **figura 1**. Como áreas desflorestadas consideraram-se as áreas em que tenha sido possível detectar a retirada da cobertura primária de vegetação em uma série de imagens dos anos de 1985, 1986, 1988, 1990, 1992, 1994 e 1995, recobrando a área de estudo. Tais áreas incluíram aquelas que em 1995 eram ocupadas por pastagens, culturas perenes e anuais, assim como áreas abandonadas em que houvesse ocorrido crescimento de vegetação secundária (“capoeiras” ou “juquiras”). Por oposição, áreas de floresta foram consideradas aquelas em que a remoção da cobertura de vegetação primária não pudesse ser inferida a partir da análise das mesmas imagens.

As classificações da imagem foram executadas segundo cinco algoritmos diferentes:

1. Fatiamento. O valor de corte foi definido após estimativa das estatísticas básicas da banda TM5 (mínimo, máximo, média e desvio padrão) para as áreas de pasto, rebrota, culturas e floresta.

2. Classificação não supervisionada de regiões. Foi utilizado o algoritmo de segmentação do SPRING com limiar 8 e área mínima de 10 pixels, seguido do algoritmo de classificação ISOSEG (<http://www.dpi.inpe.br/spring/usuario>, outubro de 1997). Os resultados do classificador foram então analisados, com a associação (mapeamento) das classes seguindo um processo de interpretação visual.

3. Classificação supervisionada de regiões. Também utilizou-se o algoritmo de segmentação do SPRING com o limiar 8 e área mínima de 10 pixels, seguido do algoritmo de classificação Bhattacharya (<http://www.dpi.inpe.br/spring/usuario>, outubro de 1997), identificando-se amostras de regiões de pasto, culturas e floresta.

4. Classificação de uma imagem composta de bandas de vegetação, solo e sombra. A imagem foi produzida a partir de transformações lineares das bandas TM3, TM4 e TM5 (Shimabukuro 1997), e em seguida classificada pelo algoritmo de máxima verossimilhança. O classificador utilizou amostras de pasto, culturas, vegetação secundária nos estágios iniciais, intermediário e avançado e também, efeitos de relevo sobre as áreas de floresta.

5. Classificação das bandas TM3, TM4 e TM5 pelo algoritmo de máxima verossimilhança. Como no caso 4, o classificador utilizou um conjunto de amostras procurando representar o maior número possível de classes de pasto, culturas, vegetação secundária nos estágios iniciais, intermediário e avançado e também, efeitos de relevo sobre as áreas de floresta.

Após classificação, (1), (4) e (5) foram tratados por um filtro não linear para eliminação das áreas menores que 10 pixels. Os resultados dos classificadores foram comparados com um mapa de referência das áreas desflorestadas da região (Alves *et al.* 1998b), calculando-se o coeficiente kappa ( $\kappa$ ) (Fleiss 1981, Hudson e Ramm 1987). Esse mapa foi elaborado a partir do fatiamento da banda

TM5, seguido de um processo iterativo de inspeção visual e correção, que utilizou dados de campo e as imagens Landsat do período 1985-1995 disponíveis para eliminação dos erros de classificação.

#### 4 Resultados e discussões

Os resultados dos classificadores são apresentados na **Figura 2** e na **Tabela I**. A figura permite apreender o desempenho dos classificadores em grandes linhas, sendo possível observar as regiões maiores delineadas por cada um deles, assim como omissões e ruídos. Os dados da **Tabela I** mostram que a fração de área desflorestada classificada com sucesso conservou-se acima dos 89%, variando entre 89% (algoritmo de máxima verossimilhança) e 95% (bandas de vegetação/solo/sombra classificadas por esse mesmo algoritmo).

Uma melhor avaliação do grau de concordância entre os classificadores e o mapa de referência é dada pelo coeficiente  $\kappa$  da **Tabela I**. Observa-se  $\kappa > 0,8$  em todos os casos, com os classificadores apresentando um excelente grau de concordância (Fleiss 1981) com o mapa de referência da **Figura 2.f**.

**Tabela I – Desempenho dos cinco algoritmos de classificação na área de estudo**

Desflorestado corresponde à fração de acerto ou erro que cada algoritmo apresentou para as áreas desflorestadas da imagem de referência

Floresta corresponde à fração de acerto ou erro que cada algoritmo apresentou para as áreas de floresta da imagem de referência

Algoritmo de Classificação	$\kappa$	Desflorestado		Floresta	
		Acerto (%)	Erro (%)	Acerto (%)	Erro (%)
(1) Fatiamento	0.885	92.4	7.6	96.1	3.9
(2) Reg. não superv.	0.833	90.6	9.4	93.4	6.6
(3) Reg. superv.	0.862	90.8	9.2	95.3	4.7
(4) Veg/Sol/Sombr	0.876	94.8	5.2	94.2	5.8
(5) Máx. Verossim.	0.851	89.2	10.8	95.5	4.5

Apesar dos coeficientes  $\kappa$  terem apresentado variações, deve-se considerar que as classificações produzidas não representaram necessariamente os melhores resultados que poderiam ser produzidos por cada algoritmo. Nessa ótica, deve-se salientar que nenhum dos classificadores apresentou

desempenho que não fosse classificado como excelente segundo os critérios apresentados por Fleiss (1981), mostrando que a classificação digital pode permitir elaborar mapas de áreas desflorestadas com resultados satisfatórios.

Os erros de classificação apresentados na **Tabela I** variaram entre 3,9% e 10,8%. As diferenças entre os resultados das classificações e o mapa de referência tiveram várias fontes, entre as quais identificaram-se quatro que apresentaram maior frequência na área de estudo: confusão entre áreas desflorestadas e vegetação nativa arbustiva, confusão entre floresta e vegetação secundária em estágio avançado de regeneração, confusão entre vegetação secundária e floresta iluminada (em áreas de relevo movimentado) e diferenças de borda. A **Figura 3** apresenta alguns exemplos desses problemas, que podem ser considerados típicos para o problema de classificação de áreas desflorestadas.

Além dos problemas acima, poderiam citar-se outros que ocorreram com menor frequência na área de estudo, mas que podem ser significativos em imagens de outras datas ou outras regiões, como a ocorrência de áreas queimadas, baixa visibilidade da imagem devido a condições atmosféricas, ocorrência de áreas inundáveis, entre outras.

Assim, analisando-se os problemas encontrados, pode-se considerar que nenhum dos algoritmos apresentou vantagens ou desvantagens evidentes sobre os outros, do ponto de vista de identificação das áreas desflorestadas. Embora uma avaliação mais detalhada do desempenho dos algoritmos possa ser realizada, o problema de classificação de áreas desflorestadas deve também considerar outras características, de natureza operacional, que devem ser cogitadas. Essas características são enumeradas na **Tabela II**.

Considerando-se estritamente as características apresentadas na **Tabela II**, pode-se observar que o método de implementação mais simples seria o fatiamento, já que ele requer somente a geração da imagem fatiada, pode ser executado rapidamente, e exige uma análise relativamente simples das estatísticas básicas da banda utilizada (TM5). Os classificadores baseados na segmentação apresentam como principal desvantagem o tempo de processamento. De uma forma geral, o tempo de processamento pode ser reduzido através da otimização dos programas, do uso de pré-processamento dos dados (Shimabukuro *et al.* 1997), ou pela aquisição de equipamentos de alto

desempenho. No entanto, essa redução pode ser difícil quando o problema necessita do tratamento de grandes volumes de dados. Os classificadores que exigem intervenção do operador, como o caso dos algoritmos que requerem aquisição de amostras, podem apresentar vantagens para a classificação, graças à possibilidade de uma melhor separabilidade entre classes. No entanto, deve-se observar que intervenções do operador podem ser trabalhosas e exigir conhecimentos mais profundos do problema tratado. Para o tratamento de grandes volumes de dados, essa situação pode levar à necessidade de um número excessivo de pessoal de alto nível de especialização.

**Tabela II – Algumas características operacionais dos algoritmos (1)-(5)**

<b>Algoritmo de Classificação</b>	<b>Dados gerados</b>	<b>Tempo de execução</b>	<b>Etapas difíceis na operação</b>
(1) Fatiamento	Imag. Class.	Muito rápido	Aquisição de amostras
(2) Reg. não sup.	Imag. Segm. Imag. Class.	Lento	Mapeamento dos resultados da classificação
(3) Reg. superv.	Imag. Segm. Imag. Class.	Lento	Aquisição de amostras
(4) Veg/Sol/Sombr	Bandas proporção Imag. Class.	Rápido	Criação veg/sol/sombr. Aquisição de amostras
(5) Máx. Veros.	Imag. Class.	Rápido	Aquisição de amostras

Imag. Class. = imagem classificada; Imag. Segm. - Imagem segmentada.

## 5 Considerações finais

Este trabalho mostrou que técnicas de classificação digital podem apresentar bons resultados em tarefas de levantamento de áreas desflorestadas. Os resultados devem ser vistos à luz de um problema muito específico: o levantamento de áreas desflorestadas para uma área de trabalho de grandes dimensões, como a Amazônia. Neste problema, o volume de trabalho e a diversidade dos padrões de uso e cobertura na Amazônia contrapõem-se à relativa simplicidade da classificação em duas categorias (desflorestado/não desflorestado).

Considerando-se as restrições de capacidade de armazenamento e processamento de dados e de pessoal qualificado, assim como a necessidade de correções para elaboração de um mapa, um sistema deve apresentar bom desempenho na classificação, mas também, em outros aspectos, como,

por exemplo, a facilidade de uso, a rapidez de processamento e a economia de utilização de meios computacionais. Seu funcionamento exige, também, a participação de especialistas na região e em processamento de imagens. Por outro lado, o sistema deve oferecer uma variedade de ferramentas já que os problemas de classificação para uma sistema de levantamento de áreas desflorestadas transcendem a simples dicotomia desflorestado/não desflorestado.

### **Agradecimentos**

O autores agradecem o apoio da FAPESP (processo 95/2432-5), do CNPq (301474/91-5) e FINEP (0816/95, contrato 6.6.96.0474.00). Agradecem também o constante apoio da Secretária de Desenvolvimento Ambiental de Rondônia (SEDAM) e, em particular, de Eraldo Matricardi, dessa Secretaria.

### **Referências**

- Alves, D.S., Meira Filho, L.G., d'Alge, J.C.L., Mello, E.K., Moreira, J.C., de Medeiros, J.S. The Amazonia Information System. **ISPRS Archives**, Commission VI, **28**, 1992, 259-266.
- Alves D.S., da Costa W.M., Escada M.I.S., Lopes E.S.S., de Souza R.C.M., Ortiz J.D. **Uma análise das regiões que apresentaram as maiores taxas de desflorestamento na Amazônia Legal nos períodos 1991-1992 e 1992-1994**. São José dos Campos: Instituto de Pesquisas Espaciais, 1998a (Relatório técnico INPE-AMZ-R03/97).
- Alves, D.S., Pereira, J.L.G., Souza, C.L., Soares, J.V., Yamaguchi, F. Classification of the deforested area in central Rondônia using TM imagery. Submetido ao IX Simpósio Brasileiro de Sensoriamento Remoto, 1998b.
- Chomentowski W., Salas, W., e Skole, D. Landsat Pathfinder Project Advances Deforestation Mapping. **GIS World**, **7**, 1994, 34-38.
- Fleiss, J.L. **Statistical Methods for Rates and Proportions** (New York: John Wiley & Sons), 1981.
- Hudson, W.D. e Ramm C.W. Correct formulation of the kappa coefficient of agreement. **Photogrammetric Engineering and Remote Sensing**, **V53(4)**, 1987, 421-422.

- Instituto Nacional de Colonização e Reforma Agrária (INCRA). **Dossiê (Divisão de Assentamento)** (Porto Velho: INCRA), Dezembro de 1996 (Relatório Interno).
- Instituto Nacional de Pesquisas Espaciais (INPE), INPE atualiza os dados do desflorestamento na Amazônia, de 95 a 97. **INPE Notícias**. No 13, 1-2, jan.-fev. de 1998, 1-2.
- RADAMBRASIL, Projeto. *Levantamento de Recursos Naturais, Folha SC.20 - Porto Velho, Vol. 16*. DNPM/Projeto RADAMBRASIL, Rio de Janeiro, 1978.
- Research Systems **ENVi user's guide**. Boulder: Research Systems Inc., 1997.
- Shimabukuro, Y.E., Mello, E.M.K., Moreira, J.C., Duarte V.D. **Segmentação e classificação da imagem sombra do modelo de mistura para mapear desflorestamentos na Amazônia**, São José dos Campos: INPE, 1997 (Relatório INPE-6147-PUD/029).
- Tardin, A.T., Lee, D.C.L., Santos, R.J.R., Assis, O.R., Barbosa, M.P.S., Moreira, M.L., Pereira, M.T., Silva, D., Santos Filho, C.P, 1980, **Subprojeto desmatamento: convênio IBDF/CNPq -- INPE**. São José dos Campos: Instituto de Pesquisas Espaciais, 1980. (INPE-1649-RPE/103).

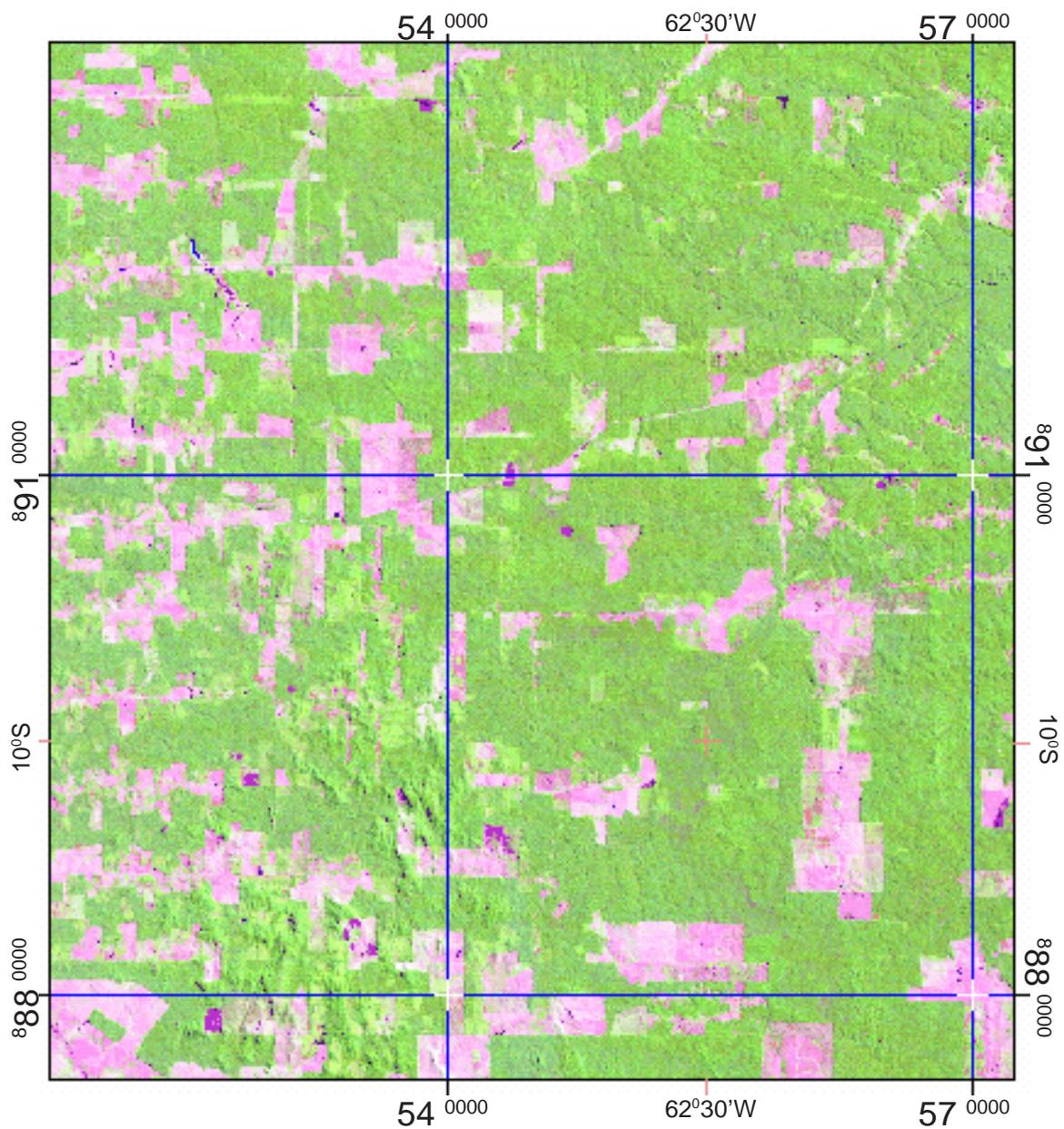


Figura 1 - Área de estudo numa composição colorida das bandas TM5 (vermelho), TM4 (verde) e TM3 (azul). As áreas desflorestadas aparecem principalmente em tons avermelhados e a floresta em tons de verde. A vegetação secundária corresponde a algumas áreas verde-claro. Efeitos do relevo podem ser vistos no quadrante inferior esquerdo. Projeção UTM. Data da imagem: 03/08/1995.

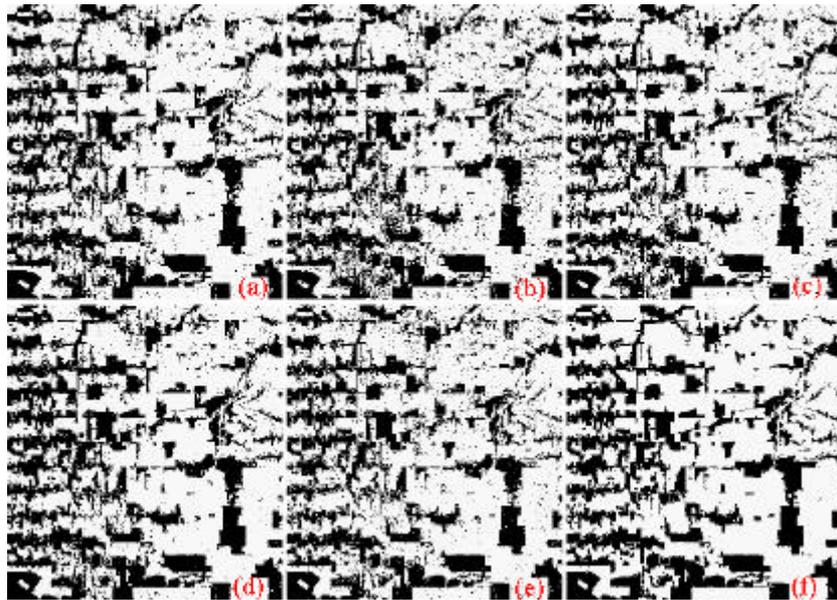


Figura 2 - Resultados gerados pelo fatiamento (a), pela classificação de regiões não supervisionada (b) e supervisionada (c), pela classificação supervisionada das bandas de vegetação, solo e sombra (d) e pelo algoritmo de máxima verossimilhança das bandas TM3, TM4 e TM5 (e). O mapa de referência aparece em (f).

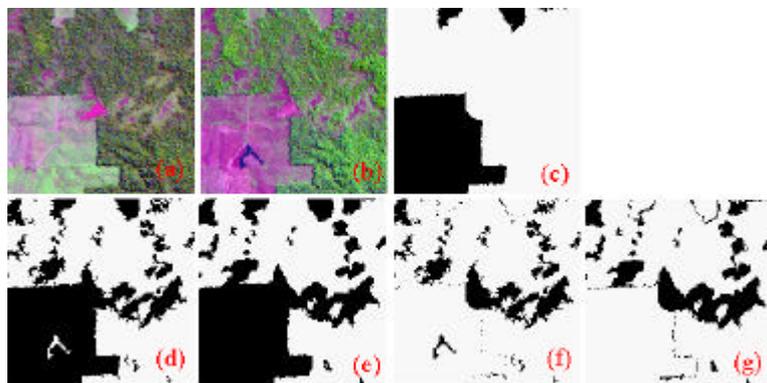


Figura 3 - Áreas desflorestadas (em preto): mapa de referência (c), fatiamento da banda TM5 (d) e classificação das bandas de vegetação, solo e sombra (e). Os erros de classificação de (d) e (e) correspondem, respectivamente, às figuras (f) e (g) (em preto). Pode-se observar que as áreas de erro correspondem à vegetação natural arbustiva observada nas imagens de 1992 (a) e 1995 (b) e por diferenças de bordas.