

EXATIDÃO DE CLASSIFICAÇÕES AUTOMÁTICAS DE UMA IMAGEM LANDSAT 5 TM PARA A REGIÃO CAFEIEIRA DE MACHADO, MG

Walbert J. R. Santos¹; Tatiana G. C. Vieira²; Tiago Bernardes³; Helena M. R. Alves⁴; Sandra P. S. Silveira³

¹Bolsista do CBP&D/Café, E-mail: walbert@epamig.ufla.br; ²Empresa de Pesquisa Agropecuária de Minas Gerais/Instituto Mineiro de Agropecuária, Lavras, MG; ³Empresa de Pesquisa Agropecuária de Minas Gerais, Lavras, MG; ⁴Embrapa Café, Lavras, MG.

Resumo:

O presente trabalho avalia a acurácia de dois classificadores automáticos disponíveis no Sistema de Informação Geográfica SPRING para o mapeamento de uso e ocupação da terra da região cafeeira de Machado, MG. Foram utilizados os classificadores: Battacharya, “por região”; e o Máxima Verossimilhança (Maxver), “pixel a pixel” para a imagem Landsat 5 TM e como referência foi utilizada a imagem do satélite Spot 4 classificada visualmente e validada em campo. Ambos algoritmos demonstraram baixo rendimento para a região. Posteriormente às classificações, foram criadas matrizes de erro, comparando as classificações automáticas com a visual. O desempenho das classificações foi avaliado pelos índices de exatidão Global, Kappa e Tau. O baixo desempenho das classificações pode ser explicado pela região cafeeira de Machado possuir um relevo muito movimentado, proporcionando regiões sombreadas e lavouras cafeeiras contíguas a fragmentos de vegetação nativa, os quais apresentam resposta espectral semelhante.

Palavras-chave: geotecnologia, café, classificação automática, SIG.

ACCURACY OF AUTOMATIC CLASSIFICATION OF A LANDSAT 5 TM IMAGE FOR THE COFFEE REGION OF MACHADO IN THE STATE OF MINAS GERAIS.

Abstract:

This work evaluates the accuracy of two automatic classifiers available on the Geographic Information System SPRING to map land use in the coffee region of Machado, in the state of Minas Gerais, Brazil. The classifiers used were Battacharya, a “per region” classifier, and Maxver (maximum likelihood) a “per pixel” classifier. The algorithms were used to classify a Landsat 5 TM image. For the comparison, a Spot 4 satellite image, which was classified by visual interpretation and verified on the field, was used as ground truth. Both algorithms presented low performance for the region. After the classifications, error parameters were done comparing the automatic classifications with the visual interpretation. The accuracy of the classifications were evaluated through the Global, Kappa and Tau indices. The classifications’ low performance may be due to the region’s relief, which provides shadowed areas, and the characteristics of the coffee fields, usually contiguous to fragments of native vegetation, both of which present similar spectral signatures.

Key words: geotechnology, coffee, automatic image classification, GIS.

Introdução

O levantamento do uso da terra é de grande importância para a cafeicultura, pois existe a necessidade de atualização constante dos registros de uso e ocupação do solo para que suas tendências possam ser analisadas com o objetivo de fornecer subsídios às ações do planejamento regional (Vieira *et al.*, 2003). As primeiras classificações de uso da terra baseavam-se em trabalhos de campo. Posteriormente, a partir da década de 50, um grande número de pesquisadores em várias partes do mundo dedicou-se à identificação detalhada de culturas agrícolas com uso de fotografias aéreas e posteriormente imagens orbitais (Steiner, 1970).

A classificação visual de fotografias aéreas e imagens de satélite, embora de grande utilidade para o mapeamento do uso e ocupação da terra, pode tornar-se inviável, quando aplicada a áreas muito extensas, pela grande demanda de tempo (Vasconcelos & Novo, 2001). Com isso os procedimentos de classificação automática têm sido largamente utilizados (Gamba & Luchiari, 2005), pois podem agilizar todo o processamento das imagens.

Dois tipos de classificação automática supervisionada serão avaliados neste trabalho. A classificação por máxima verossimilhança (Maxver) exige a definição de um conjunto de amostras de treinamento para cada classe a ser diferenciada na imagem classificada. Essa técnica assume que as classes têm distribuições de probabilidade normal. Uma amostra de treinamento é constituída por um polígono que delimita uma região da imagem pertencente a uma determinada classe. A partir das amostras de treinamento, uma função de probabilidade normal é estimada para cada classe. Essa função descreve a probabilidade de se encontrar um pixel de uma determinada classe em uma determinada posição. A partir daí, uma regra de decisão é definida e cada pixel avaliado. Um pixel pertencerá à classe com maior função de distribuição de probabilidade (INPE, 2005). A classificação Battacharya utiliza, além da informação espectral de cada “pixel”, a informação espacial que envolve a relação entre os “pixels” e seus vizinhos.

Índices relativos à confiabilidade e exatidão destes métodos de mapeamento devem ser adotados. Usualmente os índices adotados são: o índice Global (Gong & Howart, 1990), Kappa (Landis & Kock 1977) e Tau (Ma & Redmond,

1995), calculados por meio da montagem de matrizes de erro (também denominadas de confusão ou de contingência). Os índices são calculados a partir da comparação dos dados de referência (verdade terrestre) com os dados resultantes da classificação. Uma matriz de erro é um conjunto de números colocados na forma de matriz, em que as linhas normalmente correspondem aos dados de referência e as colunas aos resultados da identificação ou classificação. A diagonal principal indica os pontos concordantes entre verdade terrestre e os resultados da classificação (Story & Congalton, 1986).

O Objetivo deste trabalho é avaliar a acurácia de duas classificações automáticas (Maxver e Battacharya), por meio de índices de exatidão para gerar o mapa de uso e ocupação da terra da região cafeeira de Machado MG.

Caracterização da área

A área de estudo delimita-se geograficamente pelas coordenadas 21°31'09" e 21°42'04" de latitude sul; 46°42'04" e 45°42'04" de longitude oeste, situada na região Sudeste do Brasil, no sul do estado de Minas Gerais. A região apresenta-se dividida em dois grandes ambientes principais: o ambiente geomorfo-pedológico N-NE-E, com domínio de Latossolos, em relevo predominantemente plano a ondulado, ocorrendo nas regiões norte, nordeste e leste em relação à área urbana de Machado; e o ambiente geomorfo-pedológico W-NW, com domínio de solos com horizonte B textural, além de ocorrências de Cambissolos e Solos Litólicos, em relevo predominantemente ondulado a montanhoso, ocorrendo nas regiões oeste e noroeste em relação à área urbana de Machado (Lacerda *et al.*, 2000).

Material e Métodos

Por meio do software SPRING (INPE, 2005), foi gerada uma categoria temática "uso da terra" para abrigar os mapas de classificação visual e os da classificação automática. As classes temáticas identificadas foram: *Café em produção*: lavouras com idade superior a 3 anos; *Café em formação*: lavouras com até 3 anos; *Mata*: formação florestal densa; *Solo exposto*: áreas em preparo para plantio ou com culturas em fase de germinação; *Corpos D'água*: laminais d'água; *Outros usos*: áreas com culturas anuais em diversos estágios de desenvolvimento, pastagens e vegetação de brejo; *Reflorestamento*: plantação de eucalipto; e *Área urbana*.

Na geração desses mapas foi utilizada imagem Landsat 5 TM bandas 3, 4 e 5, órbita/ponto 219/75, com resolução espacial de 30m, e a fusão das bandas 1, 2 e 3 (20m de resolução espacial) da SPOT 4 com a pancromática (10m de resolução espacial), para obter uma melhor resolução espacial (10m).

O processo de interpretação visual da SPOT 4 foi realizado pela observação simultânea dos elementos de reconhecimento, como tonalidade, cor, textura, forma, tamanho, padrão, sombra e associação de evidências, descritos por Marchetti & Garcia (1996) e posteriores campanhas de campo para checagem de áreas com dúvidas, resultando o mapa de uso e ocupação da terra, mostrado na Figura 1.

A classificação automática por região foi a Battacharya, disponível no SPRING (INPE, 2005). Para executá-la, assim como qualquer classificação automática por região, é necessária a segmentação da imagem. Este processo divide a imagem em regiões que devem corresponder às áreas de interesse da aplicação. Entende-se por regiões, um conjunto de "pixels" contíguos, que se espalham bidirecionalmente e que apresentam uniformidade (INPE, 2005).

Na classificação pelo método Battacharya a imagem Landsat 5 TM foi segmentada e amostras foram coletadas posteriormente para o treinamento do classificador, fazendo a subdivisão das classes que mostraram uma maior diversidade nos elementos analisados na classificação visual, procurando obter amostras uniformes. As classes que exigiram uma maior diversidade dos temas foram: *mata* (3 temas), *café em formação* (2 temas), *café em produção* (3 temas). Segundo Bernardes (2006), este procedimento permite a distinção do maior número possível de classes dentro de uma imagem, ainda que sem nenhum conteúdo semântico e evitando que regiões não sejam classificadas por falta de amostras para treinamento. Após classificada, na fase do mapeamento, as classes temáticas *mata1*, *mata2* e *mata3* uniram-se numa nova classe *mata*, e o mesmo foi feito com os demais temas, obtendo o mapa de uso pela classificação automática pelo algoritmo Battacharya (Figura 2).

Também foi executada a classificação automática "pixel a pixel" pelo algoritmo máxima verossimilhança (Maxver). Este algoritmo considera a ponderação das distâncias entre médias dos níveis digitais das classes, utilizando parâmetros estatísticos. A distribuição espectral das classes de uso do solo é considerada como uma distribuição normal, ou seja, gaussianas (Moreira, 2005). Para o treinamento foram importados os temas com as respectivas amostras do método Battacharya, por apresentar números significativos de amostras e as subdivisões de temas de acordo com a variação espectral observada. Após classificada, na fase do mapeamento, as classes temáticas *mata1*, *mata2* e *mata3* também foram unidas numa nova classe *mata*, e o mesmo foi feito com os demais temas, obtendo-se o mapa de uso pela classificação automática pelo algoritmo Maxver (Figura 3).

Posteriormente às classificações, foram criadas matrizes de erro (Tabelas 1 e 3), comparando as classificações automáticas com a visual, para verificar a exatidão das mesmas com os índices: Global, Kappa e Tau.

Resultados e Discussão

As Figuras 1, 2 e 3 ilustram os mapeamentos realizados por meio de interpretação visual e automática pelas abordagens Battacharya e Maxver, respectivamente. As matrizes resultantes dos cruzamentos entre os referidos mapeamentos são apresentadas nas Tabelas 1 e 3, com os erros de omissão, e nas Tabelas 2 e 4 erros de comissão.

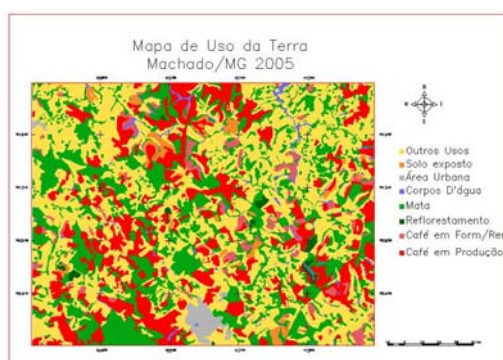


Figura 1 – Mapa de Uso da Terra - Classificação visual.

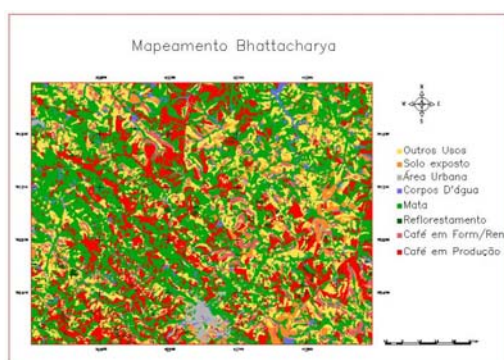


Figura 2 – Mapa de Uso da Terra - Classificação Battacharya

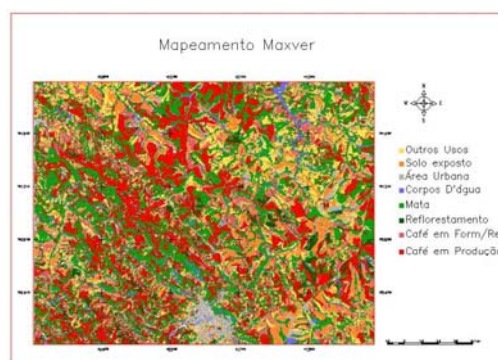


Figura 3 – Mapa de Uso da Terra - Classificação Maxver

Classificação automática pelo método Battacharya

A classe que apresentou melhor resultado na classificação Battacharya foi a *área urbana*, com um acerto de 63% dos pixels, apresentando baixo erro de omissão (pixels classificados em outras classes) de 37%. A que apresentou resultado menos preciso foi a classe *café em formação* com apenas 27% de acerto e elevado erro de omissão 73% (Tabela 2), sendo que 31% dos pixels foram classificados erroneamente como outros usos.

Os erros de comissão caracterizam os pixels classificados como uma determinada classe que não correspondem espacialmente ao dado considerado como referência. Neste caso, as classes que obtiveram melhor resultado foram *café em produção* com 46% e *Outros usos* com 41% (Tabela 2).

Os índices alcançados na classificação Battacharya, foram: Global com 38,71%, Kappa 20,2% e o Tau 29,95%. Segundo Brites (1996), o Índice Global sempre resultará em valores mais altos que Kappa e Tau, superestimando a avaliação da exatidão, uma vez que não considera os erros de omissão e comissão implícitos no restante da matriz.

De acordo com a Tabela 5, proposta por Lands & Koch (1977), os valores alcançados com o índice Kappa podem ser considerados como razoáveis.

Tabela 1 – Matriz de erro para a comparação da classificação visual (linhas) com a classificação automática Battacharya (colunas); diagonal principal apresenta a percentagem de acerto.

	1	2	3	4	5	6	7	8	TOTAL	OMISSÃO
	Percentagem									
1	35%	3%	40%	5%	0%	14%	1%	2%	100%	65%
2	6%	27%	17%	9%	1%	31%	1%	10%	100%	73%
3	14%	5%	53%	8%	0%	16%	1%	2%	100%	47%
4	6%	13%	23%	38%	0%	17%	1%	1%	100%	62%
5	5%	7%	29%	11%	28%	12%	4%	2%	100%	72%
6	5%	8%	31%	15%	1%	31%	3%	5%	100%	69%
7	1%	0%	14%	7%	1%	8%	63%	6%	100%	37%
8	5%	12%	12%	7%	1%	11%	9%	44%	100%	56%

1 - Café em produção; 2 - Café em formação; 3 - Mata; 4 - Reflorestamento; 5 - Corpos D'água; 6 - Outros usos; 7 - Área urbana; 8 - Solo Exposto

Tabela 2 – Matriz de erro em percentagem, erros de comissão (Batacharya)

	1	2	3	4	5	6	7	8
	Percentagem							
1	54%	10%	23%	10%	9%	13%	9%	9%
2	2%	45%	2%	3%	2%	5%	1%	9%
3	27%	19%	39%	22%	10%	20%	13%	11%
4	0%	1%	0%	2%	0%	0%	0%	0%
5	0%	1%	1%	1%	38%	1%	2%	1%
6	16%	50%	34%	59%	36%	59%	44%	50%
7	0%	0%	0%	1%	2%	0%	26%	2%
8	1%	3%	1%	1%	3%	1%	5%	48%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%
COMISSÃO	46%	85%	61%	98%	62%	41%	74%	82%

1 - Café em produção; 2 - Café em formação; 3 - Mata; 4 - Reflorestamento; 5 - Corpos D'água; 6 - Outros usos; 7 - Área urbana; 8 - Solo Exposto

Classificação automática pelo método de Máxima Verossimilhança MAXVER

Na classificação automática Maxver, a classe que apresentou melhor resultado foi *Mata*, com uma precisão de 53% dos pixels e erro de omissão de 47% (Tabela 3)., A classe que apresentou pior resultado foi *Café em formação* com apenas 8% de acerto e elevado erro de omissão, 92%, sendo que 47% dos pixels foram classificados erroneamente como *Outros Usos* conforme explicitado na Tabela 3. Esta classificação também apresentou elevados erros de comissão, sendo os mais altos relativos às classes *Reflorestamento* com 88%, *Corpos D'água* e *Área Urbana*, ambas com 87% (Tabela 4).

Os índices obtidos para a classificação Maxver foram: Global, com 41,25%, Kappa 19,24% e o Tau 32,85%. De acordo com a Tabela 5, o índice Kappa obteve uma qualidade ruim.

Tabela 3 - Matriz de erro para a comparação da classificação visual (linhas) com a classificação automática Maxver (colunas); diagonal principal apresenta a percentagem de acerto.

	1	2	3	4	5	6	7	8	TOTAL	OMISSÃO
	Percentagem									
1	39%	1%	41%	2%	2%	14%	1%	1%	100%	61%
2	11%	8%	21%	6%	2%	47%	1%	3%	100%	92%
3	19%	1%	53%	2%	2%	20%	2%	1%	100%	47%
4	22%	1%	19%	45%	1%	11%	0%	1%	100%	55%
5	9%	2%	28%	3%	31%	22%	4%	1%	100%	69%
6	10%	3%	38%	2%	4%	38%	3%	2%	100%	62%
7	4%	1%	16%	1%	5%	38%	25%	10%	100%	75%
8	7%	14%	13%	2%	3%	28%	2%	31%	100%	69%

1 - Café em produção; 2 - Café em formação; 3 - Mata; 4 - Reflorestamento; 5 - Corpos D'água; 6 - Outros usos; 7 - Área urbana; 8 - Solo Exposto

Tabela 4 – Matriz de erro em percentagem, erro de comissão (Maxver).

	1	2	3	4	5	6	7	8
	Percentagem							
1	46%	5%	21%	19%	11%	11%	11%	9%
2	2%	45%	2%	10%	3%	7%	2%	5%
3	28%	14%	36%	23%	20%	20%	18%	11%
4	1%	0%	0%	42%	0%	0%	0%	0%
5	1%	1%	1%	2%	43%	1%	2%	1%
6	22%	53%	39%	32%	49%	58%	53%	44%
7	0%	1%	0%	0%	2%	2%	43%	6%
8	1%	11%	1%	2%	2%	2%	1%	24%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%
COMISSÃO	54%	85%	64%	88%	87%	42%	87%	76%

1 - Café em produção; 2 - Café em formação; 3 - Mata; 4 - Reflorestamento; 5 - Corpos D'água; 6 - Outros usos; 7 - Área urbana; 8 - Solo Exposto

Tabela 5 – Qualidade da classificação associada aos valores da estatística Kappa.

Valor Kappa	Qualidade do mapa temático
< 0,00	Péssima
0,00 - 0,20	Ruim
0,20 - 0,40	Razoável
0,40 - 0,60	Boa
0,60 - 0,80	Muito boa
0,80 - 1,00	Excelente

O baixo desempenho das classificações pode ser explicado pela região cafeeira de Machado possuir um relevo muito movimentado, proporcionando regiões sombreadas, e lavouras cafeeiras contíguas a fragmentos de vegetação nativa, os quais apresentam resposta espectral semelhante. Também podem influenciar no comportamento espectral as características da lavoura de café como espaçamento, idade, cultivar, face do relevo e tratos culturais (Moreira, 2005). Outra dificuldade encontrada para obter uma classificação automática mais próxima da realidade, é que durante o processo de imageamento da superfície terrestre, pelos sistemas de sensores, a radiação captada é representada pela integração das radianças dos diferentes alvos, ou seja, os níveis de cinza dos pixels estão associados informações de mais de um tipo de

cobertura do solo e como a imagem Landsat 5 TM possui uma resolução espacial baixa (900m²), existe uma grande confusão espectral no imageamento.

Conclusões

As classificações pelos dois métodos, Maxver e Batthacharya, apresentaram baixa correlação com os dados considerados como referência, oriundos da interpretação visual da imagem Spot 4, refletindo em baixos índices, seja pelo coeficiente Global, Kappa ou Tau. Os erros de comissão e omissão também ratificam o baixo desempenho destes algoritmos para classes individuais para a região em questão.

Referência Bibliográfica

Bernardes, T. (2006) *Caracterização do ambiente agrícola do complexo serra negra por meio de sensoriamento remoto e sistemas de informação geográfica*, Dissertação (Mestrado em solos e nutrição de plantas) – Universidade Federal de Lavras, Lavras, MG

Brites, R. S. (1996) *Verificação da exatidão em classificação de imagens digitais orbitais: efeitos de diferentes estratégias de amostragens e avaliação de índices de exatidão*, 101p. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, MG.

Gamba, C. T. de; Luchiari, A. (2005) *Avaliação da exatidão de mapeamentos do uso e da cobertura da terra obtidos a partir de classificadores baseados na concepção Fuzzy*, Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, 16-21 abril 2005, INPE, p. 4073-4080.

Gong, P.; Howart, P. J. (1990) An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 56, n. 5, p. 597-603.

INPE, (2005) *Sistema de processamento de informações georreferenciadas (SPRING)*.

Lacerda, M. P. C.; Vieira, T. G. C.; Alves, H. M. (2000) *Estimativas de áreas cafeeiras ocupadas pela cafeicultura em regiões produtoras de Minas Gerais por meio de geoprocessamento*, In: Simpósio de Pesquisa dos Cafés do Brasil, 1, Poços de Caldas. Resumos expandidos...: Brasília: Embrapa Café, 2000 v. 2, p.1356-1359

Landis, J. R.; Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, St. Louis, v. 33, n. 1, p. 159-174.

Ma, Z.; Redmond, R. L. Tau. (1955) Coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 61, n. 4, p. 453-459.

Marchetti, D. A. B.; Garcia G. J. (1996) *Princípios de fotogrametria e fotointerpretação*. São Paulo – SP: Nobel, 1996. 264p

Moreira, M. A. (2005) *Fundamentos do sensoriamento remoto e metodologias de aplicação*, Ed. UFV, 3ed., 320p.

Steiner, D. (1970) *Time dimension for crop surveys from space*. Photogrammetric Engineering. Falls Church, v.36, n.2, p.187-194.

Story, M.; Congalton, R. (1986) Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, Bethesda, v. 52, n. 3, p. 397-399.

Vasconcelos, C. H.; Novo, E. M. L. M. (2001) *Mapeamento da Cobertura da Terra Utilizando Técnicas de Segmentação e Classificação por Regiões em Imagens – Fração, Obtidas do Sensor TM/LANDSAT, na Região do Reservatório de Tucuruí – PA*, Anais X SBSR, Foz do Iguaçu, 21-26 abril 2001, INPE, p. 1035-1038, Sessão Pôster.

Vieira, T. G. C.; Alves, H. M. R.; Bertoldo, M. A. (2003) Mapeamento de Áreas Cafeeiras de Minas Gerais por Imagem de Satélite - Parte II : Machado. In: III Simpósio de Pesquisa dos Cafés do Brasil Workshop Internacional de Café & Saúde, III Simpósio de Pesquisa dos Cafés do Brasil Workshop Internacional de Café & Saúde.