

Exatidão na classificação de fragmentos de matas em imagem do satélite Cbers-CCD, no município de Lavras, MG.

Elizabeth Ferreira¹
Antonio Augusto Aguilar Dantas¹
Augusto Ramalho de Moraes²

¹Departamento de Engenharia
²Departamento de Ciências Exatas
Universidade Federal de Lavras
Caixa Postal 3037, CEP 37200-000 – Lavras - MG
{bethf, auau, amorais}@ufla.br

Abstract: In this work an image from Cbers 2 - CCD was analyzed with the subject to identify small areas of native forest and to evaluate the classification accuracy. The study was made in Lavras region, MG, using the SPRING GIS with the appropriate functions to make the digital classification. The image's classification and accuracy assessment employed the overall accuracy, the user's accuracy, the producer's accuracy and the kappa coefficient. The results showed that the overall accuracy was 91%, the kappa coefficient was 82%, the user's accuracy was 93% and the producer's accuracy was 85%. These results allowed to conclude that the estimated coefficients to native forest maps were very good, but some forest areas were underestimated.

Keywords: remote sensing, accuracy, forest, thematic maps.

Palavras-chave: sensoriamento remoto, exatidão, mapas temáticos.

1. Introdução

As imagens de satélites de média resolução são comumente utilizadas no mapeamento de cobertura vegetal. As imagens do satélite Landsat 7 foram muito utilizadas, mas atualmente o mesmo encontra-se inoperante. O satélite sino brasileiro Cbers possui características que o colocaram na condição de prosseguir a geração de imagens com a vantagem das mesmas serem disponibilizadas sem custos para os usuários. A avaliação dos produtos do satélite Cbers, utilizando para isto medidas estatísticas é uma das formas de se estabelecer a validação da continuidade da utilização de imagens de média resolução e verificar as necessárias correções.

Dentre os métodos mais simples para geração de imagens classificadas, a partir de imagens de satélite, o algoritmo de máxima verossimilhança tem sido o mais utilizado. O mapeamento temático produzido a partir deste algoritmo apresenta bons resultados quando são comparados os tipos de cobertura de solo e os usos existentes no mapa e no campo. A avaliação da verdade de campo pode ser obtida construindo-se uma matriz de erro, como ponto de partida, para a aplicação de uma série de técnicas estatísticas descritivas e analíticas dos dados.

As diferentes técnicas de interpretação e a escolha dos produtos (sensor, resolução, faixa espectral) influenciam nas características dos mapas produzidos. Além destes fatores, os critérios de interpretação, utilizados pelos intérpretes podem originar diferentes classificações.

Costa (1999) comparou diferentes classificadores automáticos, em uma imagem Landsat-TM, do município de Viçosa, MG, avaliando a percentagem de área não classificada e o tempo de processamento, concluindo que o algoritmo de máxima verossimilhança foi o melhor, com exatidão global de 77%.

Carvalho (2001) utilizou uma série de imagens Landsat-TM, da região do Alto do Rio Grande, MG, e diferentes técnicas de classificação e integração de dados, com o objetivo de

discriminar remanescentes florestais de cultivos de café e eucalipto, obtendo exatidões que variaram de 34 a 51%.

O uso de critérios para avaliar e comparar a exatidão de mapeamentos temáticos é muito bem descrito por Congalton (1991). Segundo o autor, uma matriz de erro pode ser usada como o ponto de partida para a aplicação de uma série de técnicas estatísticas descritivas e analíticas. Através da estatística descritiva obtém-se o índice de exatidão global, em que se divide o total de pixels classificados corretamente (diagonal da matriz) pelo número total de pixels da matriz. A exatidão das categorias estudadas é feita de forma similar. Como técnica analítica, o índice kappa é bastante utilizado para inferir a exatidão.

Fidalgo (1995) utilizou vários métodos estatísticos para estimar a exatidão de mapeamentos de remanescentes de Mata Atlântica, através de imagens Landsat-TM, produzidos por diferentes intérpretes e processos de digitalização, concluindo que a estimativa com o coeficiente kappa para todos os produtos foram inferiores às respectivas estimativas de exatidão global.

Segundo Lillesand et al. (2004) a matriz de erro, também denominada matriz de confusão ou tabela de contingência, compara basicamente, categoria por categoria, a relação entre os dados de referência conhecidos e os resultados correspondentes de uma classificação automatizada. Chuvieco (1990) explica que os dados de referência podem ser obtidos a partir de trabalhos de campo ou de documentos auxiliares, como fotografias aéreas e cartas.

A partir da matriz de confusão pode desenvolver-se uma série de medidas estatísticas que concluem o processo de validação, sendo descritas por Chuvieco (1990), Congalton (1991), Fidalgo (1995) e Lillesand et al. (2004).

Entre estas, a *exatidão global* é a medida mais simples e relaciona os elementos da diagonal com o total de pontos amostrados, sendo dada pela fórmula:

$$F_m = \frac{\sum X_i}{N} \times 100$$

F_m = exatidão global

X_i = elementos na diagonal

N = total de elementos amostrados

Em uma análise mais rigorosa deve-se levar em consideração os valores nas células marginais da matriz. No caso das linhas, os valores indicam o número de amostras que, pertencendo a uma determinada categoria, não foram incluídos nela. Os *Erros de Omissão* são calculados por:

$$E_o = \frac{X_{i+} - X_{ii}}{X_{i+}}$$

E_o = erros de omissão

X_{i+} = marginal da linha

X_{ii} = diagonal daquela linha

De igual forma as células nas diagonais das colunas expressam o *erros de comissão*, isto é, pixels que não se incluem em uma determinada categoria, pertencente realmente a outra:

$$E_{co} = \frac{X_{+i} - X_{ii}}{X_{+i}}$$

E_{co} = erros de comissão

X_{+i} = marginal da coluna

X_{ii} = diagonal daquela coluna

Os *erros de omissão e comissão* enfocam o mesmo problema. A omissão se refere a uma definição imperfeita da categoria e a comissão se refere a uma delimitação excessiva da categoria. Alguns autores enfocam-na como *exatidão do usuário (consumidor) e exatidão do produtor*:

$$F_u = \frac{X_{ii}}{X_{i+}} \times 100$$

F_u = exatidão do usuário

$$F_p = \frac{X_{ii}}{X_{+i}} \times 100$$

F_p = exatidão do produtor

A *estimativa de Kappa* expressa a medida da diferença entre a concordância dos dados de referência e a classificação automática, e a probabilidade de concordância entre os dados de referência e a classificação aleatória. Conceitualmente, o índice kappa pode ser definido como:

$$K = \frac{N \sum X_{ii} - \sum X_{i+} X_{+i}}{N^2 - \sum X_{i+} X_{+i}}$$

X_{ii} = concordância observada

X_{+i} e X_{i+} (produto das marginais), sendo a concordância esperada

N = total de elementos observados

Assim, neste trabalho o objetivo foi elaborar um mapeamento de fragmentos de mata nativa utilizando a classificação digital de imagem do satélite Cbers e avaliar por meio de índices de exatidão a qualidade dos resultados obtidos.

2. Material e Métodos

O trabalho foi realizado no município de Lavras, MG, com área total de 548,4024 km², localizado entre as coordenadas UTM 7.633 km, 7.664 km N e 481 km, 514 km E, meridiano central 45° WGr, *datum* Córrego Alegre.

O trabalho foi realizado em três etapas. Na primeira etapa foi criada uma base de dados para desenvolvimento do trabalho no aplicativo SPRING (CÂMARA et al.1996).

Nesta base, o projeto “Lavras” foi criado, contendo categorias no modelo imagem (“Imagem”), MNT (“Altimetria”) e temáticas (“Limite”, “Drenagem” e “Vegetação”).

A base planialtimétrica (curvas de nível, estradas e cursos d’água), referentes às folhas Lavras, Itumirim, Carmo da Cachoeira e Nepomuceno, foram importadas dos arquivos digitais de cartas do IBGE.

A imagem do satélite Cbers2-CCD, órbita-ponto 153/124, bandas 423/RGB de 04/07/04 foi utilizada para o trabalho.

O georreferenciamento da imagem Cbers foi feito pela captura de pontos de controle utilizando-se como plano vetorial, principalmente os cursos d’água, das cartas do IBGE.

Para a classificação supervisionada foi estabelecida uma chave de interpretação utilizando fotos aéreas e inspeções “in situ”. A principal fisionomia de vegetação identificada foi a floresta semidecídua ou mata nativa, que, segundo Oliveira-Filho & Fluminham-Filho (1999), é encontrada no fundo dos vales, nas adjacências aos cursos d’água (mata ciliar) e encostas cujos solos são mais profundos (mata de encosta). Os fragmentos se apresentaram

em forma irregular, tamanho médio a pequeno, tonalidade vermelha escura, textura rugosa, localizados próximo às drenagens ou nas encostas.

Na classificação digital das imagens utilizou-se o método supervisionado adquirindo-se 20 amostras de treinamento, com limiar de aceitação de 99%. O tamanho da amostra de treinamento foi o menor polígono possível, com a imagem na escala 1:50.000. Não foi feita nenhuma transformação sobre as bandas espectrais e o critério utilizado para classificação foi o de máxima verossimilhança.

Após a classificação, partiu-se para a elaboração dos mapeamentos de fragmentos de vegetação. Nestes mapas, os polígonos com áreas menores que 5000 m² foram eliminados, e transformados para o formato matricial com resolução espacial de 20 x 20 m. Com os mapas temáticos passou-se a definir o tipo e o tamanho e de amostragem.

A amostragem utilizada foi a aleatória simples, atribuindo-se a cada polígono um número único para, depois, selecionar alguns desses elementos de maneira casual. Foram gerados 1724 polígonos para a imagem Cbers.

$$n = \frac{1,96^2 \times 85 \times 15}{5^2} = 195,92$$

Para verificação em campo foram selecionados, aleatoriamente 81 pontos das áreas com floresta e 115, referentes a outros usos.

Uma vez estabelecido o tipo e o tamanho da amostra, os dados para avaliação da exatidão foram coletados, com o objetivo de expressar em uma tabela, a classes reais e as deduzidas da classificação, denominada matriz de erro.

Nas visitas a campo foi utilizado um GPS de navegação com 15 m de precisão e observada a presença ou ausência da categoria (sim/não). Para avaliação da categoria não floresta (outros) foram utilizadas ortofotos e fotografias aéreas de 1986.

A comparação entre os mapeamentos gerados e a verdade de campo foi feita com uso dos índices de exatidão (Chuvieco, 1990; Congalton, 1991; Fidalgo, 1995).

Os produtos foram então finalizados utilizando os módulos Scarta e Iplot do SPRING.

3. Resultados e Discussão

Na **Figura 1** estão destacados (por um círculo) os pontos que foram classificados como floresta e no campo eram referentes a outros usos. Os principais usos confundidos na classificação foram café (6 pontos), reflorestamento (4 pontos) e capoeira (2 pontos).

Na **Tabela 1** estão dispostos os valores da matriz de confusão, gerada pela verificação em campo e pela imagem classificada, a partir dos quais foram obtidos os índices de exatidão global, exatidão do consumidor, exatidão do produtor e índice kappa.

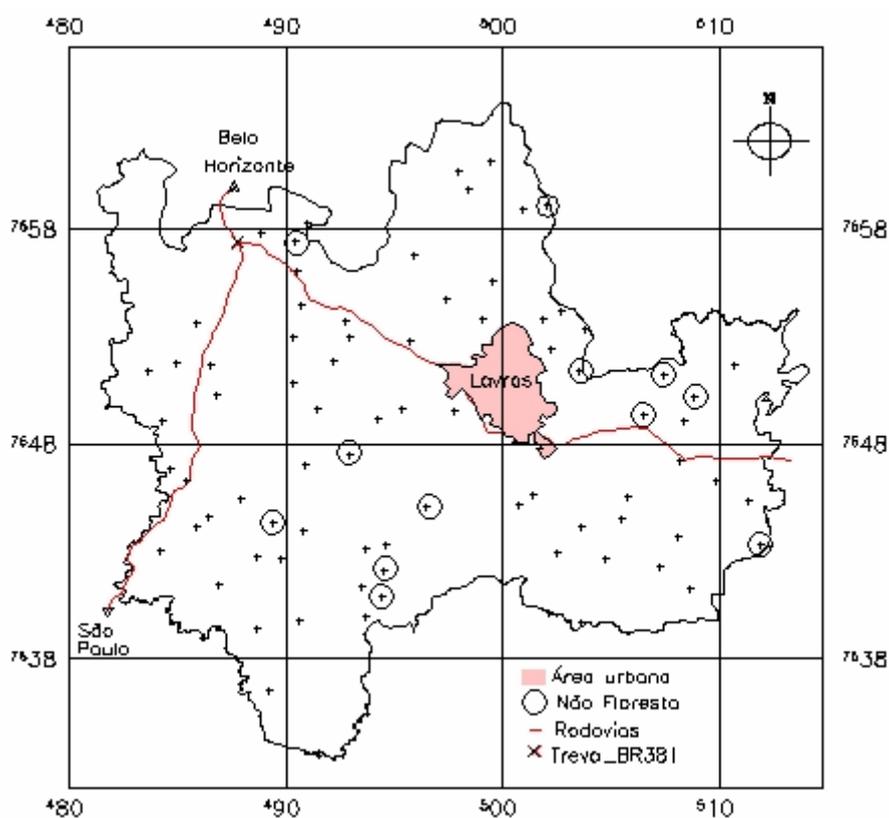


Figura 1. Pontos selecionados para verificação dos fragmentos de vegetação no campo.

Tabela 1. Erros de omissão, de comissão, exatidão do consumidor, do produtor, índice de exatidão global e kappa, em porcentagem, obtidos pela verificação em campo e pela imagem classificada.

Classificação	Verdade de campo			Erro de omissão	Exatidão do consumidor
	Floresta	Outros	Total		
Floresta	67	5	72	7%	93%
Outros	12	112	124	10%	90%
Total	79	117	196		
Erro de comissão	15%	4%			
Exatidão do Produtor	85%	96%			
Exatidão Global = 91%				Índice kappa = 82%	

Os cálculos dos erros de omissão (E_o), de comissão (E_{co}), exatidão do consumidor (F_u), exatidão do produtor (F_p), calculados, para as categorias floresta e não floresta (outros) estão expressos na **Tabela 2**.

A exatidão global (F_m) e o índice kappa (k), foram calculados, para as categorias, obtendo-se resultados de 91 e 82%.

Tabela 2. Cálculos dos erros de omissão (E_o), de comissão (E_{co}), exatidão do consumidor (F_u), do produtor para as categorias floresta e não floresta (outros).

Floresta	Outros
$E_o = \frac{72-67}{72} \times 100 = 7\%$	$E_o = \frac{124-112}{124} \times 100 = 10\%$
$E_{co} = \frac{79-67}{79} \times 100 = 15\%$	$E_{co} = \frac{117-112}{117} \times 100 = 4\%$
$F_u = \frac{67}{72} \times 100 = 93\%$	$F_u = \frac{112}{124} \times 100 = 90\%$
$F_p = \frac{67}{79} \times 100 = 85\%$	$F_p = \frac{112}{117} \times 100 = 96\%$

$$F_m = \frac{179}{196} \times 100 = 91\%$$

$$\kappa = \frac{196(67+112) - [(79 \times 72) + (117 \times 124)]}{(196)^2 - [(79 \times 72) + (117 \times 124)]} \times 100$$

$$\kappa = \frac{35.084 - 20.196}{38.416 - 20.196} \times 100 = \frac{14.888}{18.220} = 82\%$$

Os índices de exatidão indicam a probabilidade das categorias estudadas (floresta e outros) classificadas na imagem Cbers terem sido também verificadas em campo.

O índice de exatidão global foi de 91%. Baseado apenas neste resultado, poderia se concluir que este índice foi excelente. No entanto, Congalton (1991) alerta que, com base apenas no índice de exatidão global, as inferências feitas podem gerar equívocos e conclui que o cálculo e a análise da “user’s accuracy” confere maior significado aos dados, revelando se há confusão entre as categorias.

Na **Tabela 2** observa-se que a exatidão do produtor foi de 85% (floresta) e 96% (outros). Apesar da proposta da classificação do mapa ter sido para a categoria floresta, o índice de exatidão obtido para a categoria “outros” foi melhor. Este resultado não pode ser extrapolado para as classes que obviamente existem dentro da categoria “outros”, como água, urbano, cultivos agrícolas entre outras. De qualquer forma, o resultado obtido para a categoria floresta foi muito bom, embora a exatidão global tenha sido melhor (91%).

A exatidão do consumidor foi de 93% (floresta) e 90% (outros). No caso da floresta, resume-se que, embora 85% das áreas tenham sido identificadas como florestas, a exatidão do consumidor mostra que a verificação em campo seria floresta em 93% das visitas. Para categoria “outros”, a chance de encontrá-la seria em 90% das visitas a campo. Para ambas categorias os resultados obtidos para a exatidão foram excelentes.

Uma outra forma de analisar a matriz de confusão seria através dos erros. Os erros de omissão foram de 15% (floresta) e 4% (outros), significando que 15%, ou seja, 12 pontos visitados que deveriam ser classificados como floresta, não foram. Para a categoria “outros”, apenas 4%, 5 pontos, não corresponderam à categoria na realidade, sendo omitidos desta.

Os erros de comissão mostram os pontos que foram indevidamente incluídos em uma categoria. Pela **Tabela 2** pode ser observado que 7% (5 pontos) da categoria floresta foram indevidamente incluídos na categoria outros. De modo igual, 10% (12 pontos) da categoria

outros foram incluídos na categoria floresta. A categoria floresta apresentou melhor resultado do que a categoria outros.

O índice kappa resultou em valor de 82%, sendo uma avaliação muito boa de acordo com a tabela de referência de Lands & Koch, (1977). Os índices de exatidão global foram maiores que os índices kappa sendo que resultados semelhantes foram encontrados por vários autores (Congalton, 1991; Fidalgo, 1995 e Costa, 1999).

Segundo Lillesand (2004) o índice kappa serve como um indicador do grau que a percentagem dos valores corretos da matriz de erro é devido à “verdade” versus a “probabilidade” de concordância. Neste caso, o índice kappa obtido de 82% deve ser entendido como uma indicação de que a classificação obtida foi 82% melhor do que seria um resultado originado através de uma assinatura aleatória dos pixels.

Na **Figura 2** pode ser observada a distribuição espacial das áreas de fragmentos de mata nativa. A área total obtida foi de 54,15 km², correspondendo a 10% da área total do município. Pode se observar também a baixa sensibilidade do sensor CCD para detecção de fragmentos dentro da área urbanizada. Uma análise estatística das áreas dos polígonos resultou em valores com média de 8,38 ha, sendo o mínimo de 1,23 ha e o máximo de 73,29 ha. Alguns autores (Chuvienco, 1990 e Lillesand, 2004) afirmam que a amostragem aleatória simples pode levar à subestimação de pequenas áreas. Os erros que podem ser inseridos nos índices de exatidão são bem detalhados em trabalhos de Verbyla & Haymond (1995).

Apesar dos altos índices de exatidão alcançados pode ser observado na **Figura 2** que a região noroeste da imagem não teve nem um polígono classificado e em consequência, nenhum ponto foi visitado no campo. Isto pode ter levado a uma subestimativa dos valores de área, mas este erro não afetou os índices de exatidão.

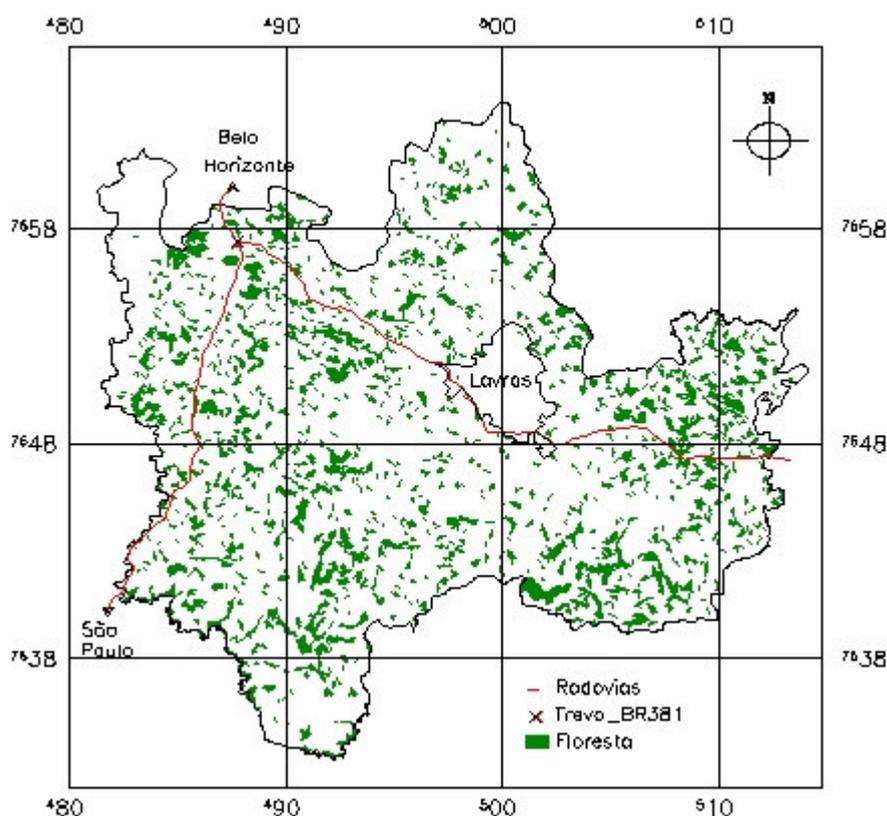


Figura 2. Áreas de fragmentos florestais obtidas a partir das imagens Cbers, com índice kappa igual a 82%. Município de Lavras, MG.

4. Conclusões

Com base nos resultados obtidos pode-se concluir que o mapa de fragmentos de mata nativa produzido a partir da classificação digital da imagem Cbers apresentou índices de exatidão considerados muito bons e excelentes.

Apesar dos altos índices de concordância houve uma subestimativa de áreas com mata nativa em algumas regiões do município.

Os produtos Cbers podem ser utilizados com segurança para substituir as imagens Landsat 7 que foram descontinuadas.

5. Referências Bibliográficas

Câmara G., Souza, R.C.M, Freitas, U.M, Garrido, J. Spring: Integrating remote sensing and GIS by object-oriented data modelling. **Computers & Graphics**. V.20, n. 3, p.395-403, 1996.

Carvalho, L. **Mapping and monitoring forest remnants: a multi-scale analysis of spatio-temporal data**. 2001. 140p. (ISBN 90-5808-545-7). Thesis – Wagenigen Universiteit, Wagenigen , 2001.

Congalton, R.G. A review of assessing the accuracy classificatons of remotely sensed data. **Remote Sensing Environment**, v. 37, p.35-46, 1991.

Costa, T. C. e C da. **Avaliação de estratégias para classificação de imagens orbitais em mapeamento de uso e cobertura da terra**. 1999. 109p. (UFV-621.3678). Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 1999.

Chuvieco, E. **Fundamentos de teledetection espacial**. Madrid: Unigraf, 1990. 453p.

Fidalgo, E. C.C. **Exatidão no processo de mapeamento temático da vegetação de uma área de mata atlântica no estado de São Paulo, a partir de imagens TM-Landsat**. 1995. 186p. (INPE-5944-TDI/570). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1995.

Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-154, 1977.

Lillesand, T.M., Kiefer, R.W., Chipan, J.W. **Remote sensing and interpretation**. 5ed. Madison: Wiley, 2004. 763p.

Oliveira-Filho, A. T.; Fluminhan-Filho, M. Ecologia da vegetação do parque florestal do Rio Bonito. **CERNE**, v. 5, n. 2,p.51-64, 1999.

Verbyla, D.L., Hammond, T. O. Conservative bias in classification accuracy assessment due to pixel by pixel comparison of classified images with reference grids. **International Journal of Remote Sensing**, v. 16, p. 581-587, 1995.