

Avaliação de classificadores para o mapeamento de uso da terra

Ariadiny Monteiro da Silva¹
Jefferson Rodrigo Batista de Mello¹
Raquel Stucchi Boschi¹
Jansle Vieira Rocha¹
Rubens Augusto Camargo Lamparelli^{1,2}

¹Faculdade de Engenharia Agrícola - FEAGRI/UNICAMP
Av. Candido Rondon, 501 - Caixa Postal 6011
13083-875 - Campinas - SP, Brasil
{ariadiny.silva; raquel.boschi; jansle.rocha}@feagri.unicamp.br
jerso-mello@hotmail.com

²Centro de Pesquisas Meteorológicas e Agrícolas aplicadas a Agricultura-Cepagri/Unicamp
Cidade Universitária “Zeferino Vaz” - 13083-970 - Campinas - SP
rubens@cpa.unicamp.br

Abstract: In the classification it seeks to assign a class to each pixel of the image, according to land use, being that the main objective of the process is to distinguish and identify different surface materials, whether vegetation, rocks and others. This recognition of the targets and the distinction between them is only possible due to the fact that surface materials have specific behavior along the electromagnetic spectrum. The aim of this study was to identify the best algorithm for classifying different types of land use in an image Landsat-5/TM using supervised classifiers and unsupervised. The study area is located in the extreme west of the state of Bahia. It was tested the methods of the Minimum Distance, of the Malahanobis, of the Maximum Likelihood and the ISODATA method. The best results were obtained by merging the results of the Minimum Distance method and the ISODATA method, with a kappa index of 0.8615 and a value of 85.88% of Overall Accuracy. Without the merger, the method that produced the best results was the ISODATA method, with a value of overall accuracy of 85% and a Kappa index of 0.8465. Following, there was the method of maximum likelihood value with accuracy of 84.23% and Kappa index of 0.83, followed by the Minimum Distance method with 80% of overall accuracy and a Kappa index of 0.79. Finally, there was the Mahalanobis method, with the same value of overall accuracy of the previous method and with a Kappa of 0, 73.

Palavras-chave: processamento de imagem, classificação supervisionada e não-supervisionada, índice Kappa, image processing, supervised and unsupervised classification, Kappa index.

1. Introdução

No processo de classificação procura-se atribuir uma classe a cada pixel da imagem, de acordo com a ocupação do solo, sendo que o objetivo principal do processo é a distinção e identificação de diferentes materiais superficiais, sejam eles vegetação, rochas e outros. Este procedimento é realizado por algoritmos específicos de reconhecimento de padrão, também denominados classificadores (Moreira, 2003). Esse reconhecimento dos alvos e a distinção entre eles só é possível devido ao fato dos materiais superficiais terem comportamento específico ao longo do espectro eletromagnético (Crosta, 2002).

O algoritmo utilizado define o tipo de classificação como não-supervisionada ou supervisionada.

A classificação não-supervisionada é um método estatístico cujo procedimento constitui em um processo de análise de pixels de forma isolada, que agrupa pixels com intervalos de níveis de cinza aproximados.

O algoritmo ISODATA (Self-Organizing Data Analysis Techniques) é semelhante ao algoritmo k-means, porém é mais sofisticado pois permite o ajuste do número de clusters automaticamente durante a interação por fusão de agrupamentos semelhantes e divisão de clusters com grande desvio padrão (Jensen, 1996).

Já na classificação supervisionada, o reconhecimento de padrões espectrais se faz com base em amostras fornecidas pelo analista, em que toda a imagem é analisada estatisticamente, a partir das assinaturas espectrais geradas, de acordo com o algoritmo escolhido (Richards e Jia, 2006).

Dentre os classificadores supervisionados, Richards e Jia (2006) citam que o método Mahalanobis é um classificador direcional sensível à distância e que utiliza dados estatísticos para cada classe.

Já Crosta (2002) menciona que o método da máxima verossimilhança considera a ponderação das distâncias das médias e utiliza parâmetros estatísticos. Já o método da distância mínima calcula a média para cada tema em cada banda espectral, sendo que, para cada pixel é atribuída a classe com a média mais próxima dele, ou seja, todos os pixels da imagem são classificados.

Após o processo de classificação é necessário validar os resultados, a fim de analisar a acurácia do processo. A forma mais comum de se avaliar é por meio da matriz de erro ou matriz de confusão, que compara as amostras da imagem de referência com os resultados do mapa classificado (Story e Congalton, 1986; Cohen, 1960).

Com base na matriz de erro podem ser calculados os erros de inclusão e de omissão, assim como a Exatidão Global e o Índice Kappa, quanto maior o valor do índice, melhor será considerado o resultado da classificação.

Deste modo, o objetivo deste trabalho foi identificar o melhor algoritmo de classificação na identificação de diferentes tipos de uso da terra da imagem Landsat-5/TM, sobre uma região do município de Formosa do Rio Preto, na região do Extremo Oeste Baiano, estado da Bahia.

2. Metodologia de Trabalho

A área de estudo está localizada na região do Extremo Oeste do estado da Bahia entre os paralelos $10^{\circ} 40'$ a $12^{\circ} 29'$ de latitude Sul e entre os meridianos $44^{\circ} 24'$ a $46^{\circ} 30'$ de longitude Oeste, conforme demonstra a Figura 1.

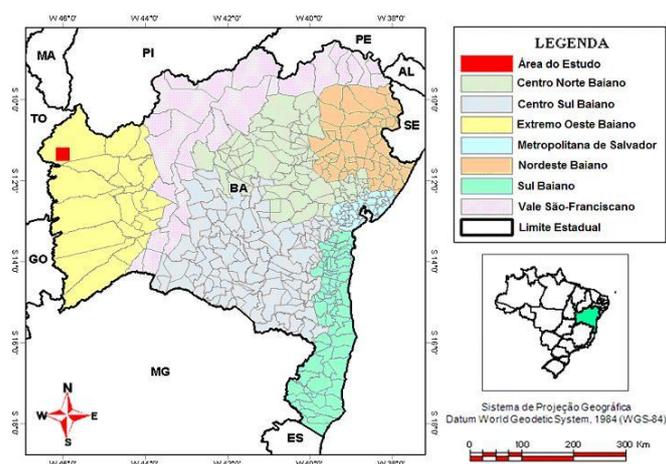


Figura 1. Localização da área de estudo na região do Extremo Oeste Baiano.

Os processamentos aplicados na imagem, como realces, contrastes, classificações e fusão foram realizados no software SIG TNTmips Pro 2008:74.

A imagem utilizada do satélite Landsat-5/TM, imageada em fevereiro de 2010, composição RGB543 foi disponibilizada pelo INPE e georreferenciada com base na imagem Landsat 7/TM+, ano 2000, corrigida geometricamente pela NASA.

Foi utilizado o sistema de projeção UTM (Universal Transversa de Mercator), *datum* WGS-84 (World Geodetic System, 1984), Meridiano Central 45° Oeste (Fuso 23° Sul).

Para classificação não-supervisionada foi utilizado o “ISODATA Automatic”.

Nas classificações supervisionadas foram testados os métodos da “Distância Mínima”, da “Máxima Verossimilhança” e da “Distância de Mahalanobis”.

Para avaliar a acurácia dos resultados foi construída a matriz de confusão a partir de 260 pontos de amostragem pelo método da amostragem estratificada aleatória não-linear.

A partir da matriz de confusão pode-se analisar e comparar a imagem de referência com o mapeamento resultante de cada classificador.

Com base nesta matriz foram calculados os erros de inclusão e de omissão, assim como a Exatidão Global (Equação 1) e o Índice Kappa (Equação 2).

$$Exatidão\ Global = \frac{A}{n} * 100 \quad (1)$$

A= total de acertos; n= número de pontos amostrais.

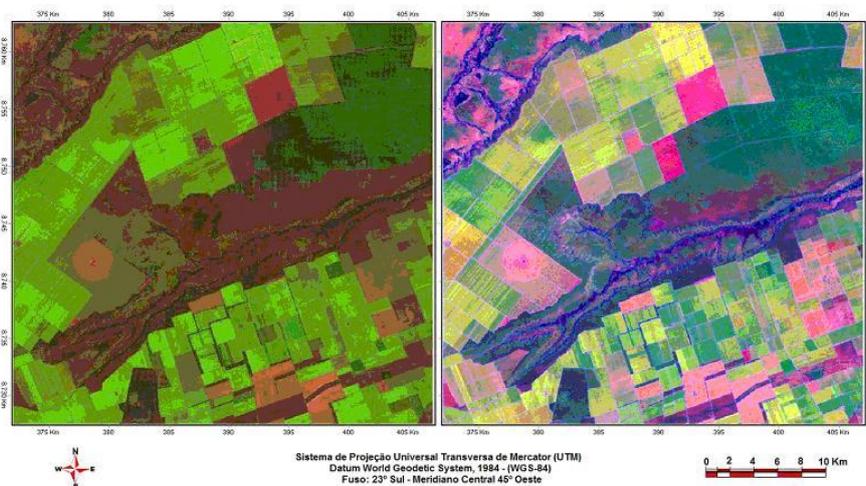
$$Índice\ Kappa = \frac{[n * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_i * x_j)]}{[n^2 - \sum_{i=1}^r (x_i * x_j)]} \quad (2)$$

n = número de observações (pontos amostrais); r = número de linhas da matriz de erro; x_{ij} = observação na linha i e coluna j; x_i = total marginal da linha i; x_j = total marginal da coluna j.

3. Resultados e Discussão

Na classificação pelo método ISODATA a área foi inicialmente dividida em 12 classes temáticas. No entanto, no decorrer do processo estas classes foram ajustadas de forma interativa, resultando em 5 classes temáticas relevantes no mapeamento final: cultura anual de cereais e algodão, cerrado, juquira (vegetação secundária), solo exposto e área degradada.

A Figura 2(a) apresenta a imagem de referência Landsat-5/TM na composição RGB 543 sem atividades de realce e contraste. Já a Figura 2(b) apresenta a imagem na composição RGB 543 após atividades de realce e contrastes, com aplicação de contraste equalizado na Banda 5, contraste normalizado nas bandas 4 e 3, bem como filtro de passa alta Laplaciano na banda 3 para detecção de bordas.



(a)

(b)

Figura 2. (a) Imagem original RGB 543; (b) Imagem realçada e contrastada RGB 543.

A Figura 3 compara visualmente a imagem de referência com o mapa classificado pelo método ISODATA.

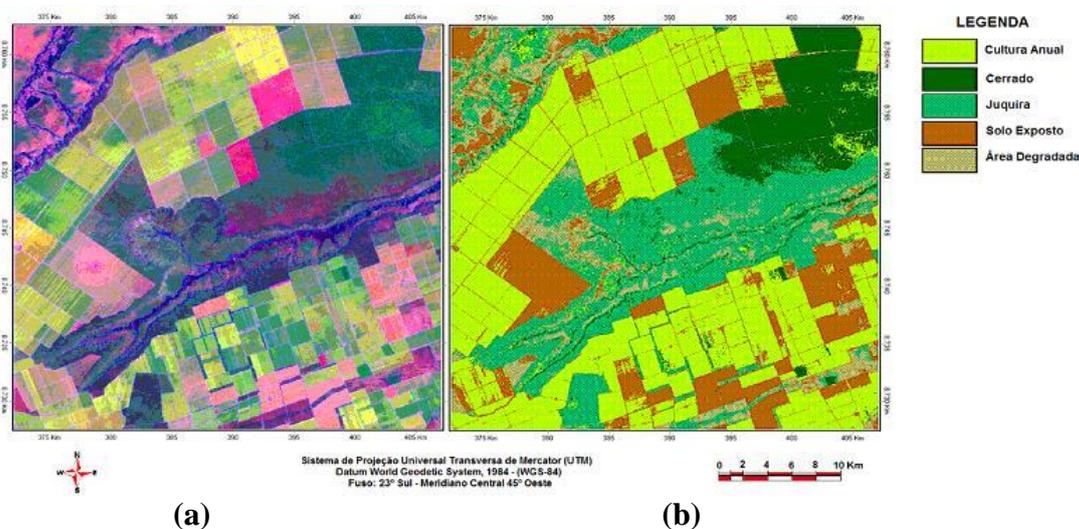


Figura 3. (a) Imagem de referência; (b) Mapa classificado pelo método ISODATA.

A Tabela 1 apresenta a matriz de erro, na qual se observa que das 260 amostras distribuídas na imagem e no mapa classificado, observou-se que 221 foram classificadas corretamente, resultando numa Exatidão Global de 85% e Índice Kappa de 0,8465, classificado por Landis e Kock (1977) como “Excelente”.

Tabela 1. Matriz de erro entre a imagem de referência e o mapa classificado pelo método não-supervisionado - ISODATA.

		Matriz de Erro/Confusão						Acerto do Usuário	Erro de Inclusão
		Referência Terrestre (imagem RGB453)							
		Área Degrad.	Cerrado	Cult Anual	Juquira	Solo Exp.			
Mapa Classificado	Área Degrad.	14	3	0	4	2	60,9%	39,1%	
	Cerrado	0	24	2	2	0	85,7%	14,3%	
	Cult. Anual	0	3	95	3	4	90,5%	9,5%	
	Juquira	2	6	1	48	0	84,2%	15,8%	
	Solo Exp.	1	0	5	1	40	85,1%	14,9%	
	Acerto	82,4%	66,7%	92,2%	82,8%	87,0%			
	Erro de Omissão	17,6%	33,3%	7,8%	17,2%	13,0%			

Os três métodos da classificação supervisionada (Método da Distância Mínima, Método de Mahalanobis e o Método da Máxima Verossimilhança) geraram os seguintes resultados:

O **Método da Distância Mínima** classificou corretamente 209 das 260 amostras, conforme demonstra a matriz de erro na Tabela 2, resultando numa Exatidão Global de 80% e Índice Kappa de 0,7959, classificado como “Muito Bom”.

A classificação pelo método da Distância Mínima apresentou uma acurácia menor em relação à classificação pelo ISODATA. Segundo Crosta (2002), a desvantagem desse método reside no fato de que todos os pixels da imagem serão classificados, o que pode acarretar em mistura da resposta espectral dos temas.

A Figura 4 apresenta a imagem de referência comparada ao mapa classificado pelo Método supervisionado da Distância Mínima.

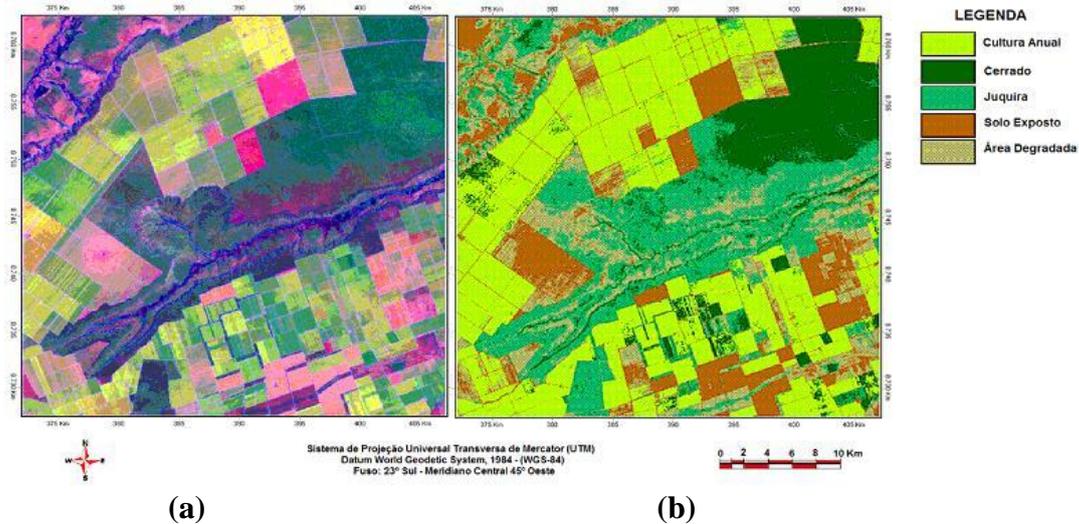


Figura 4. (a) Imagem de referência; (b) Mapa classificado pelo método da Distância Mínima.

Tabela 2. Matriz de erro entre a imagem de referência e o mapa classificado pelo método supervisionado da Distância Mínima.

		Matriz de Erro/Confusão						
		Referência Terrestre (imagem RGB453)					Acerto do Usuário	Erro de Inclusão
		Área Degrad.	Cerrado	Cult. Anual	Juquira	Solo Exp.		
Mapa Classificado	Área Degrad.	15	3	5	11	11	33,3%	66,7%
	Cerrado	0	28	3	3	0	82,4%	17,6%
	Cult. Anual	0	0	89	0	2	97,8%	2,2%
	Juquira	1	5	4	44	0	81,5%	18,5%
	Solo Exp.	1	0	2	0	33	91,7%	8,3%
	Acerto	88,2%	77,8%	86,4%	75,9%	71,7%		
Erro de Omissão		11,8%	22,2%	13,6%	24,1%	28,3%		

O Método da Máxima Verossimilhança classificou corretamente 219 amostras, de acordo com a matriz de confusão mostrada na Tabela 3, resultando numa Exatidão Global de 84,23% e Índice Kappa geral foi de 0,8383, classificado como “Excelente”.

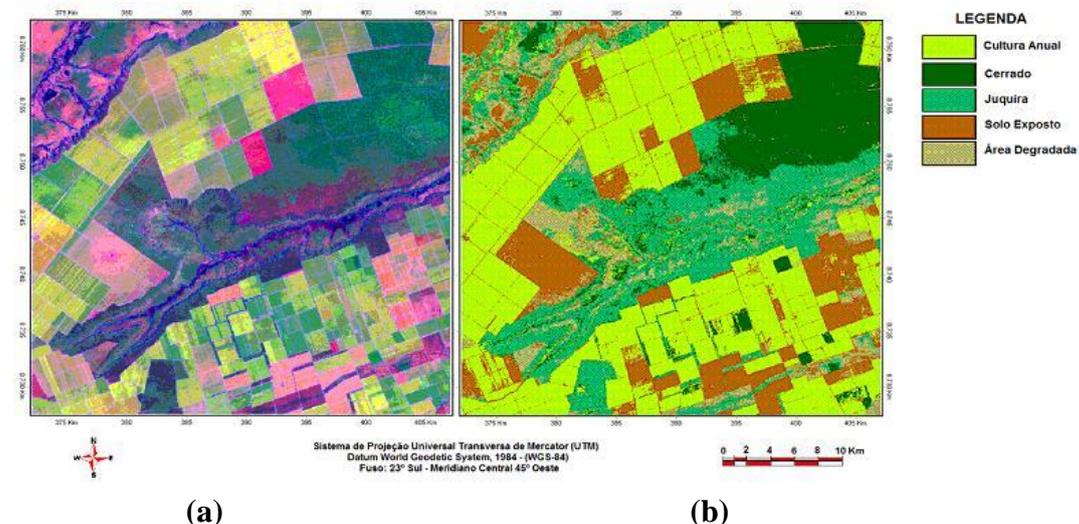


Figura 5. (a) Imagem de referência; (b) Mapa classificado pelo método da Máxima Verossimilhança.

A classificação pelo método MaxVer apresentou uma acurácia menor em relação à classificação ISODATA, porém maior em relação ao método da Distância Mínima.

Tabela 3. Matriz de erro entre a imagem de referência e o mapa obtido a partir do método da MaxVer.

		Matriz de Erro/Confusão						
		Referência Terrestre (imagem RGB453)						
		Área Degrad.	Cerrado	Cult. Anual	Juquira	Solo Exp.	Acerto do Usuário	Erro de Inclusão
Mapa Classificado	Área Degrad.	13	2	0	4	3	59,1%	40,9%
	Cerrado	0	27	2	7	0	75,0%	25,0%
	Cult. Anual	0	2	95	2	3	93,1%	6,9%
	Juquira	3	5	1	44	0	83,0%	17,0%
	Solo Exp.	1	0	5	1	40	85,1%	14,9%
	Acerto	76,5%	75,0%	92,2%	75,9%	87,0%		
Erro de Omissão	23,5%	25,0%	7,8%	24,1%	13,0%			

O Método Mahalanobis classificou corretamente 208 amostras, como pode ser verificado na Tabela 4, resultando numa Exatidão Global de 80% e Índice Kappa geral foi de 0,7916, classificado como “muito bom”.

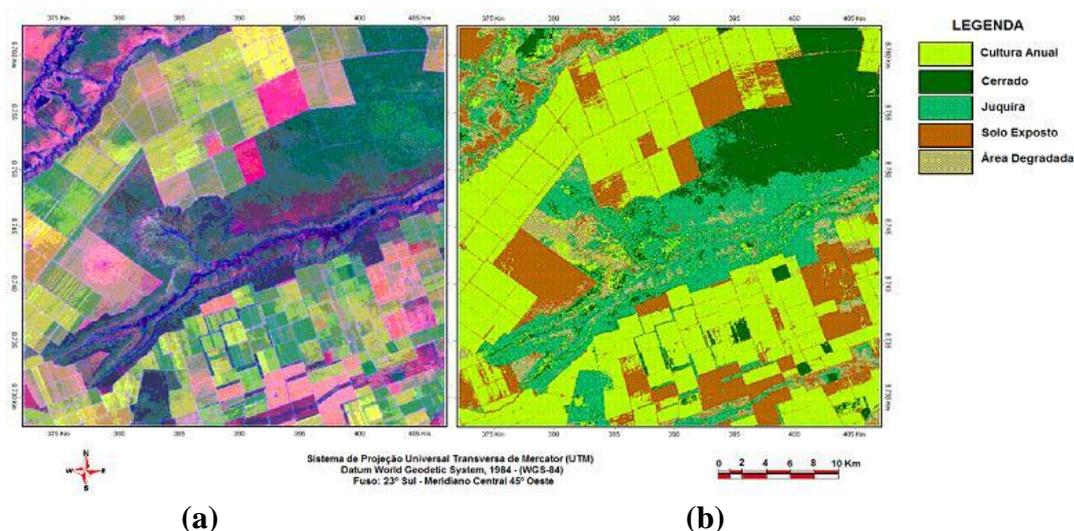


Figura 6. (a) Imagem de referência; (b) Mapa classificado pelo método de Mahalanobis.

Tabela 4. Matriz de erro entre a imagem de referência e o mapa obtido pelo método supervisionado da Distância de Mahalanobis.

		Matriz de Erro/Confusão						
		Referência Terrestre (imagem RGB453)						
		Superfície					Acerto do Usuário	Erro de Inclusão
		Área Degrad.	Cerrado	Cult. Anual	Juquira	Solo Exp.		
Mapa Classificado	Área Degrad.	9	0	0	4	3	56,3%	43,8%
	Cerrado	0	27	2	8	0	73,0%	27,0%
	Cult. Anual	0	2	97	9	4	86,6%	13,4%
	Juquira	6	6	0	36	0	75,0%	25,0%
	Solo Exp.	2	1	4	1	39	83,0%	17,0%
	Acerto	52,9%	75,0%	94,2%	62,1%	84,8%		
Erro de Omissão	47,1%	25,0%	5,8%	37,9%	15,2%			

A classificação pelo método de Mahalanobis apresentou uma acurácia menor em relação aos demais classificadores.

O método da distância mínima classificou melhor as classes de Cultura Anual e Solo Exposto. Entretanto, o método não-supervisionado ISODATA classificou melhor as classes de Cerrado, Juquira e Área Degradada.

Diante disto, optou-se pela geração de um mapa de uso da terra a partir da fusão das classes que melhor foram classificadas, ou seja, foi feita a fusão entre os resultados dos classificadores ISODATA e da distância mínima, conforme mostra a Figura 7.

Após a criação do mosaico, a imagem foi submetida ao processo de filtragem de moda, visando a homogeneização dos grupos classificados e eliminação de ruídos.

A Tabela 5 apresenta a matriz de confusão gerada a partir da análise dos 260 pontos de amostragem distribuídos na imagem de referência e no mapa resultante da fusão.

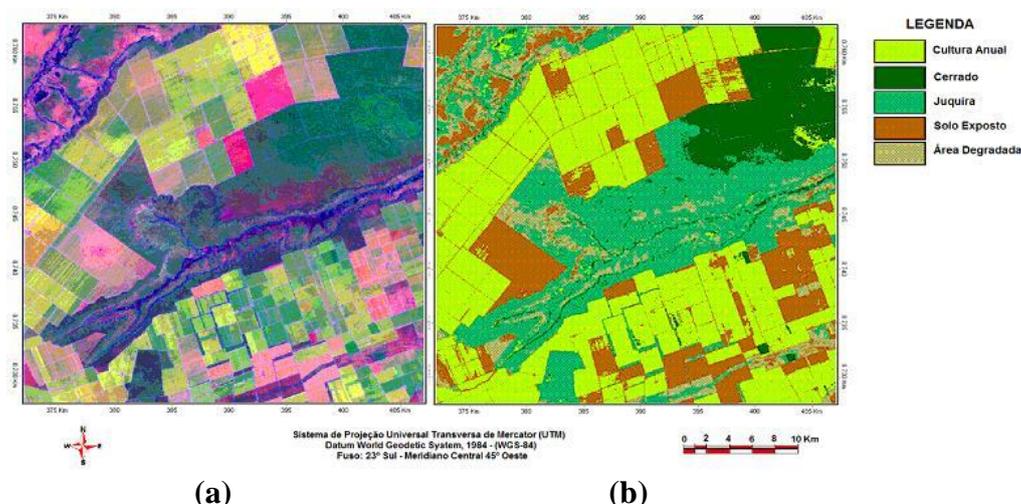


Figura 7. (a) Imagem de referência; (b) Mosaico com as classes extraídas pelos métodos ISODATA e da Distância Mínima.

Tabela 5. Matriz de erro entre a imagem de referência e o mapa obtido pela fusão.

Classes		Referência Terrestre (imagem RGB453)						Acerto do Usuário	Erro de Inclusão
		Área Degrad.	Cerrado	Cult. Anual	Juquira	Solo Exp.			
Mapa Classificado	Área Degrad.	12	3	0	5	2	54,5%	45,5%	
	Cerrado	0	26	1	1	0	92,9%	7,1%	
	Cult. Anual	0	1	96	2	3	94,1%	5,9%	
	Juquira	3	6	1	49	0	83,1%	16,9%	
	Solo Exp.	2	0	5	1	41	83,7%	16,3%	
	Acerto	70,6%	72,2%	93,2%	84,5%	89,1%			
Erro de Omissão		29,4%	27,8%	6,8%	15,5%	10,9%			

A Fusão dos classificadores dos métodos ISODATA e Distância Mínima demonstrou melhores resultados, comparados aos resultados obtidos por apenas um único método de classificação.

Os resultados da Exatidão Global e Índice Kappa da fusão foram de 0,8615 e 0,8588 respectivamente, sendo superior aos valores do método ISODATA que foram, os melhores resultados totais obtidos das classificações por um único método, sendo 0,85 e 0,8465, respectivamente para a Exatidão Global e o Índice Kappa. Ou seja, consoante Landis e Koch (1977) o mapa da fusão ficou classificado como “Excelente”.

Em suma, verifica-se na Tabela 6 a comparação entre os valores de exatidão global e Índice Kappa de cada classificador, bem como da fusão.

Tabela 6. Valores de Exatidão Global e Índice Kappa de cada classificador.

Classificadores	Exatidão Global	Índice Kappa
Fusão	86,15%	0,8588
ISODATA	85%	0,8465
MaxVer	84,23%	0,8383
Distância Mínima	80%	0,7959
Mahalanobis	80%	0,7316

De modo geral, os classificadores que apresentaram melhor acurácia seguem em ordem decrescente: ISODATA, MaxVer, Distância Mínima e Mahalanobis.

Conclusões

Neste trabalho, o classificador que apresentou melhor acurácia foi obtido pela fusão entre os resultados do método da Distância Mínima e do método ISODATA, em que aproximadamente 14% dos pontos de amostragem foram classificados erroneamente, correspondendo a um Índice Kappa de 0,86.

As técnicas de pré-processamento mostraram-se eficientes para análise visual dos alvos, com destaque para as técnicas de realce (contraste e filtragem).

Referências Bibliográficas

- COHEN, J. A. **Coefficient of agreement for nominal scales.** *Educational and Psychological Measurement*, Durham, v.20, n.1, p.37-46, 1960.
- CRÓSTRA, A. P. **Processamento digital de imagens de sensoriamento remoto.** 4 ed. rev. Campinas: Instituto de Geografia – UNICAMP, 2002. p164.
- JENSEN, J. R. **Introductory digital image processing: A remote sensing perspective.** In Prentice Hall, 1996.
- LANDIS, J.R., KOCH, G.G. **The measurement of observer agreement for categorical data.** *Biometrics*, v.33, p.159-174, 1977.
- MOREIRA, M.A. **Fundamentos do sensoriamento remoto e metodologias de aplicação.** Viçosa, SP: Ed. UFV, 2005. 320p.
- RICHARDS, J. A.; JIA, X. **Remote sensing digital image analysis: an introduction.** Fourth Edition. New York: Springer Verlag, 431p., 2006.
- STORY, M.; CONGALTON, R. G. **Accuracy assessment: a user's perspective.** *Photogrammetric Engineering and Remote Sensing*, Bethesda, v.52, n.3, p.397-399, 1986.