

Classificador fuzzy não supervisionado aplicado à detecção de biomas: uma aplicação do algoritmo FCM e dos índices de validação de clusters.

Carlos Eduardo Guerra¹
Nayara Safira da Silva Caldas²

¹Universidade Federal do Oeste Pará - UFOPA
Santarém, PA, Brasil
guerra@ufpa.br

²Universidade Federal do Oeste Pará - UFOPA
Santarém, PA, Brasil
safira_newton@hotmail.com

Abstract - This paper aims to show a system, for unsupervised classification, using as selection tool classes, a Fuzzy Clustering algorithm (FCM), combined with a Fuzzy system classification. Among the most common classification techniques, like the binary, which classifies the pixel as the relevance or not in the range of gray levels, most of techniques have some limitations for the separation of two or more classes. According to Meirelles and House (2007), the separation between different biomes does not occur abruptly and there is a transition zone that can not be fully characterized as belonging to a particular class. The algorithm FCM (Fuzzy C-means clustering) groups the pixel`s in clusters, in which we can obtain intervals classes that fed into the Fuzzy system. Finally, the Fuzzy classifier is applied to LANDSAT TM, with the aim of discriminating features of type savanna. The system can be used to discriminate against other objects, such as lakes, forests and urban areas, may also be useful for studying changes in land cover due to human occupation, deforestation, among others.

Palavras-chave: Processamento de imagens, Classificador Fuzzy e Algoritmo de Clusterização.

1. Introdução

Os estudos de imagens de satélite são de grande importância para o monitoramento de regiões, caracterizando mudanças no solo e na cobertura vegetal. No entanto, para que essas imagens forneçam dados significativos muitas vezes faz-se necessário a utilização de ferramentas do processamento digital de imagens, dentre as quais podemos destacar os classificadores.

Os procedimentos de classificação são frequentemente utilizados nos estudos ambientais, com o objetivo de discriminar objetos de interesses (biomas, zonas urbanas, corpos d'água, etc.). Segundo Fonseca (2008), a classificação consiste em dividir o espaço de atributos da imagem em regiões correspondentes as classes temáticas. Dentre os tipos de classificações, temos a classificação supervisionada e a não supervisionada. Genericamente, na classificação não-supervisionada, cada pixel da imagem é associado a uma classe sem que o usuário tenha conhecimento prévio do número ou da identificação das diferentes classes presentes na área imageada. A classificação é feita através de agrupamentos baseados em similaridades (FONSECA, 2008). Na classificação supervisionada, é necessário um conhecimento prévio da área estudada para se definir as fronteiras de classe. Nos processos de classificação de imagens, utiliza-se normalmente ferramentas de limiarização, que tem como objetivo, o agrupamento de pixel`s em regiões com características específicas, pertencentes a um intervalo comum de níveis de cinza ou limiar (Caldas & Guerra, 2009). Conforme o algoritmo de limiarização e do classificador a ser empregado, os sistemas de classificações podem ser divididos em duas abordagens: **Classificadores "pixel a pixel"**, que utilizam apenas a informação espectral, isoladamente, de cada pixel para achar regiões homogêneas; e os **classificadores por regiões**, que utilizam, além de informação espectral de cada "pixel", a informação espacial que envolve a relação entre os "pixels" e seus vizinhos (Florenzano, 2002).

Neste trabalho, aplicaremos uma rotina de clusterização (FCM) para seleção automática (não supervisionada) de classes. As classes obtidas, em forma de intervalos, serão utilizadas em um sistema Fuzzy, que realizará uma classificação pixel a pixel. Finalmente, aplicaremos o sistema classificador completo em imagens TM, para a detecção de savanas na região de Santarém. A validação do número de classes será feita através do índice Dunn, que apontará o número máximo de clusters para uma boa classificação.

Savanas são ecossistemas com vegetação esclerofila mista; de espécies arbóreo-arbustivas esparsas, cujo estrato inferior é dominado por espécies herbáceas. Segundo Holanda et al. (2009), as savanas na região de Santarém estão inseridas numa região de tensão ecológica entre florestas e apresentam intensas mudanças na sua cobertura vegetal, associadas ao regime pluviométrico, dinâmica natural da sua vegetação e a ação antrópica (agricultura, ocupação, etc.).

2. Objetivos

O objetivo deste trabalho é discriminar objetos, do tipo savanas (na Região de Santarém), valendo-se de uma metodologia de classificação não supervisionada, que utiliza o algoritmo Fuzzy C- means Clustering (FCM), para seleção automática de intervalos de classes, aliado a um sistema Fuzzy de classificação.

3. Materiais e métodos

3.1 Fundamentos teóricos

Definimos uma imagem de satélite por um mapeamento ψ (Equação 1) de \mathbf{A} em \mathbf{Z} , onde \mathbf{A} é uma matriz com elementos \mathbf{a}_{ij} , que indicam a posição de cada pixel. Os valores de $\psi(\mathbf{a}_{ij})$ assumidos pelo mapeamento são denominados por níveis de cinza.

$$\psi: \mathbf{a}_{ij} \rightarrow \psi(\mathbf{a}_{ij}), \mathbf{a}_{ij} \in \mathfrak{R} \text{ e } \psi(\mathbf{a}_{ij}) \in \mathbf{Z}. \quad (1)$$

Um sistema de classificação binária é uma função do tipo:

$$\tau: \psi_{ij} \rightarrow \{0,1\}, \quad (2)$$

que levam os pixel's da imagem em níveis cinza a valores 0 ou 1. Nesse caso, os pixel's são mapeados de acordo com limiares definidos segundo as características espectrais dos alvos que se deseja separar (Gonzales, 2003). Um sistema de classificação binária tem como base a lógica clássica, i.é, a atribuição de valores fixos para relacionar elementos a um conjunto, baseia-se no princípio do terceiro excluído. Atribui-se valor 1 e 0 para a pertinência e não pertinência, respectivamente. O procedimento binário normalmente começa por uma análise prévia do histograma de uma das componentes espectrais da imagem; e define-se um limiar, ou seja, um intervalo de níveis de cinza, que corresponda à feição que se deseja extrair. Aos pixel's pertencentes ao intervalo alvo atribui-se valor 1 e aos demais, 0. Seja A um conjunto no universo X , a pertinência ou não de $x \in X$ em A pode ser caracterizada, pela lógica clássica por:

$$\mu_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad (3)$$

3.1.1 Lógica Fuzzy

Um conjunto de abordagens que permitem tratar incertezas, como um especialista da área de sensoriamento remoto o faria, são denominados por sistemas especialistas, dentre os quais destacam-se os algoritmos genéticos, computação evolutiva, as redes neurais e a lógica fuzzy (Barron, 1993; Aguiar, 1999 e Gomide et. al., 1994).

O conceito de conjunto Fuzzy foi introduzido por Lotfi A.Zadeh, dando importantes contribuições à lógica e a inteligência artificial. Na lógica Fuzzy, os elementos podem assumir graus de pertinência entre os valores 0 e 1. A função de pertinência é assim definida:

$$\mu_A(x) \rightarrow [0,1] \quad , \quad (4)$$

ou seja, um mapeamento dos elementos x , do universo de discurso X , no intervalo $[0, 1]$. Determinamos o conjunto fuzzy A , pela união dos pares ordenados $(x, \mu_A(x))$. No intervalo $[0,1]$, 0 significa a não pertinência de x em A , 1 significa a pertinência total em A e os valores entre 0 e 1 representam grau de pertinência em A . Em sistemas fuzzy, a classificação é feita através de regras de inferências, construídas a partir de operadores fuzzy definidos abaixo:

Definição 3.1.1 Os operadores usados para elaboração das regras fuzzy são:

- \wedge operador de conjunção (and).
- \vee operador de disjunção (or).
- \Rightarrow operador de implicação (if-then).
- \Leftrightarrow operador de dupla implicação (if and only if).

Definição 3.1.2 Dados os conjuntos fuzzy W e Q e as funções W_μ e Q_μ , definimos o operador Norma-Tmínimo por:

$$T\text{-min}(W,Q)=\min(W_\mu, Q_\mu)=W_\mu \wedge Q_\mu. \quad (5)$$

Definição 3.1.3 O operador Conorma-T máximo :

$$T\text{-max}(W,Q)=\max(W_\mu, Q_\mu)=W_\mu \vee Q_\mu. \quad (6)$$

Como exemplo, a regra $R_1: V_{1ij} \wedge V_{2ij} \Rightarrow V_{rij} = \otimes \min(V_{1ij} \wedge V_{2ij}, V_{rij})$ significa:

R_1 : Se "entrada 1 é vegetação" e "entrada 2 é vegetação" então " saída é vegetação"

3.1.2 Fuzzy Clustering

A metodologia FCM foi desenvolvido por Dunn(1973) e melhorado por Bezdek em 1981. O "Fuzzy C-means Clustering" é um algoritmo que organiza um determinado conjunto de objetos em grupos, clusters. Um cluster é um conjunto de objetos agrupados segundo características em comum e que os diferem de outros clusters. Esse algoritmo baseia-se na minimização da função:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, 1 \leq m < \infty, \quad (7)$$

onde m é qualquer número maior que 1, u_{ij} é o grau de pertinência de x_i no cluster j , x_i é o i -ésimo dado, c_j é o centro do cluster e $\|\cdot\|$ é qualquer norma que expresse similaridades entre o centro e os dados processados.

3.2 Metodologia

A metodologia que implementa o classificador consiste das seguintes etapas:

3.2.1 Seleção da área estudada e imagens

O alvo a ser estudado são as savanas da região de Santarém-PA. Primeiramente, escolhemos as imagens a serem processadas, no que se refere às bandas espectrais e sensores típicos para estudo de vegetação. Na **Figura 1** temos uma composição falsa cor e um recorte das bandas do vermelho, infravermelho próximo e infravermelho médio, referentes à zonas de savanas em Alter do Chão.

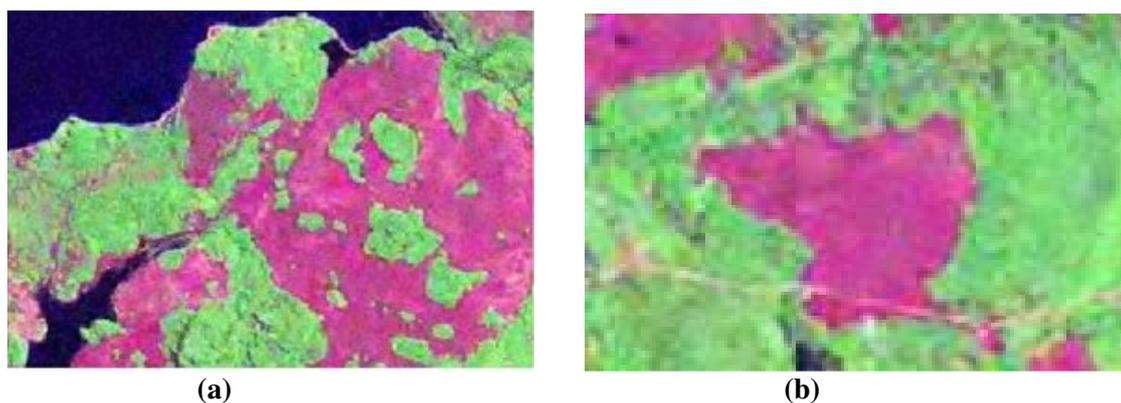


Figura 1: (a) Composição colorida TM (banda 3,4 e 5) de uma região de Savanas (Região do Eixo-Forte em Santarém) com baixa cobertura Vegetal; (b) Recorte da savana em um trecho ao longo de uma Rodovia.

As bandas espectrais escolhidas como entrada para o classificador foram o vermelho e infravermelho próximo, pois segundo Ponzoni & Shimabukuro (2007), estas banda caracterizam melhor as estruturas com baixa e alta cobertura vegetal. Na **Figura 2.(a)** e **2.(c)** as zonas mais claras correspondem as feições com baixa cobertura vegetal, i.é, savanas, estradas, desmatamentos, solos, dentre outras; enquanto que na **Figura 2.(b)** e **2.(d)** estas mesmas estruturas são representadas pelas zonas mais escuras.

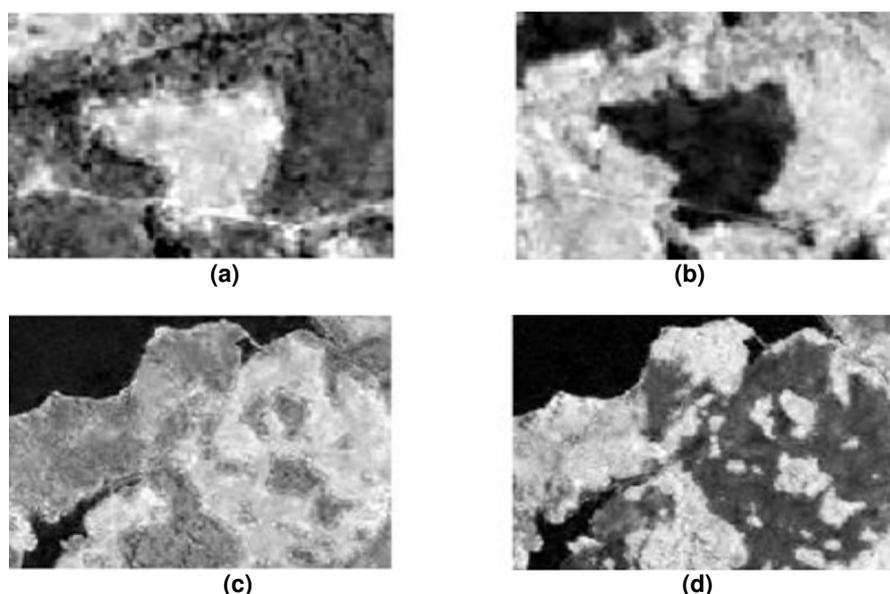


Figura 2: Na figura (a) e (c), Imagem TM na banda 3(vermelho) e banda 4 (b) e (d), infravermelho próximo, referentes as imagens das **Figuras 1) e 2).**

3.2.2 Aplicação do FCM e determinação de intervalos de classe.

Após a escolha das componentes a serem trabalhadas, as imagens são transformadas em vetores que armazenam os valores de níveis de cinza. Para a imagem da **Figura 1.(a)** o número de clusters a serem determinados pelo FCM é 3, pois é visível a presença de 3 estruturas bem definidas na imagem, que correspondem a savana (magenta), cobertura vegetal(verde) e a água(azul escuro). Para a imagem da **Figura 1.(b)** o número de Clusters que pode ser inferido é apenas 2, visto que há predominância de savanas e cobertura vegetal. Os clusters podem ser observados na **Figura 3.(a)** e **Figura 3.(b)**. Os pontos pretos em destaques correspondem aos centros de cada cluster. O eixo horizontal e vertical são os valores dos pixel's, em nível de cinza, para as componentes vermelha e infravermelho próximo, respectivamente.

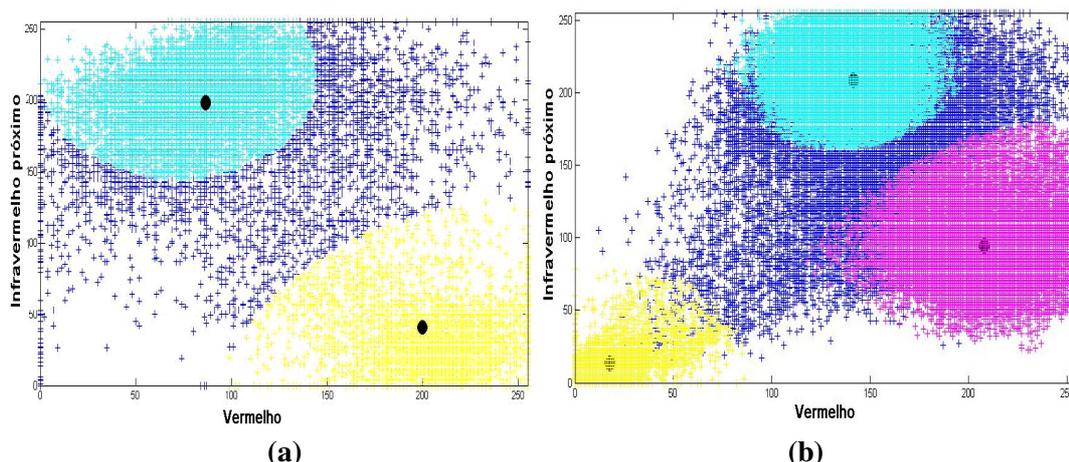


Figura 3: Em (a), o agrupamento dos pixel's em clusters. O cluster azul claro corresponde às zonas com cobertura vegetal e o amarelo às zonas de savana. Em (b), O cluster amarelo corresponde à água, o azul claro, as zonas com cobertura vegetal densa e o magenta, savanas.

3.2.4 Validação do número de clusters

Os índices de validação de clusters são utilizados para se obter uma estimativa do número de clusters que proporcione uma melhor classificação dos dados originais. Neste trabalho utilizamos os índices internos Dunn e Dunn modificado. Neste caso, o algoritmo de clusterização é executado para vários valores diferentes para o parâmetro k , que representa o número de clusters. Em seguida, os valores dos índices obtidos a partir das partições geradas são plotados em função de k . O melhor número de clusters é dado pelo mínimo dessa função.

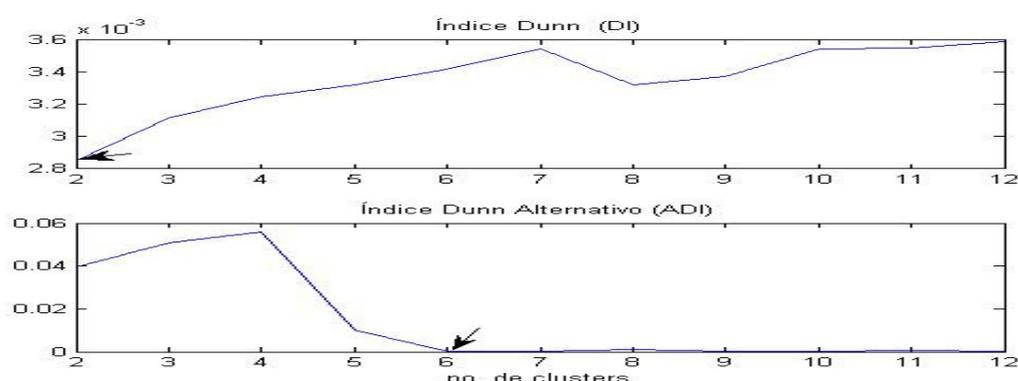


Figura 4: Índices de validação do número de clusters definidos para o algoritmo FCM.

A **Figura 4** mostra o gráfico dos índices de Dunn e Dunn alternativo para análise de agrupamentos, referentes aos dados do recorte (**Figura 1.(b)**). A análise destas funções aponta para um número máximo de clusters igual a seis (Dunn ADI). Entretanto, o mínimo da função Índice Dunn aponta para dois. Como a qualidade do agrupamento melhora com a diminuição do número de clusters, escolhemos $k=2$, i.é, o FCM será rodado com dois clusters.

3.2.3 Implementação do Sistema Fuzzy a partir da extração de classes do FCM.

Definidos os clusters, observamos que estes possuem formas geométricas similares as de elipses, cujas origens são os centros dos clusters. Como exemplo, a projeção do eixo maior sobre a horizontal nos fornece o suporte das funções de pertinência da primeira variável de entrada, i.é, a componente vermelha do sistema fuzzy de classificação. A projeção sobre a vertical corresponde ao limite das funções para a segunda variável de entrada, i.é, a componente infravermelho próximo.

Na **Figura 5.(b)**, temos como exemplo as funções de pertinência resultante do algoritmo FCM para a componente do vermelho. O diagrama da **Figura 5.(a)** mostra como funciona o sistema classificador. As entradas ψ_1 e ψ_2 são a primeira e segunda componente das imagens escolhidas. ψ_r é a imagem resultante após a classificação. O sistema Fuzzy, mostrado no diagrama, possui duas variáveis de entrada, que correspondem as imagens a serem processadas (bandas 3 e 4 para o exemplo).

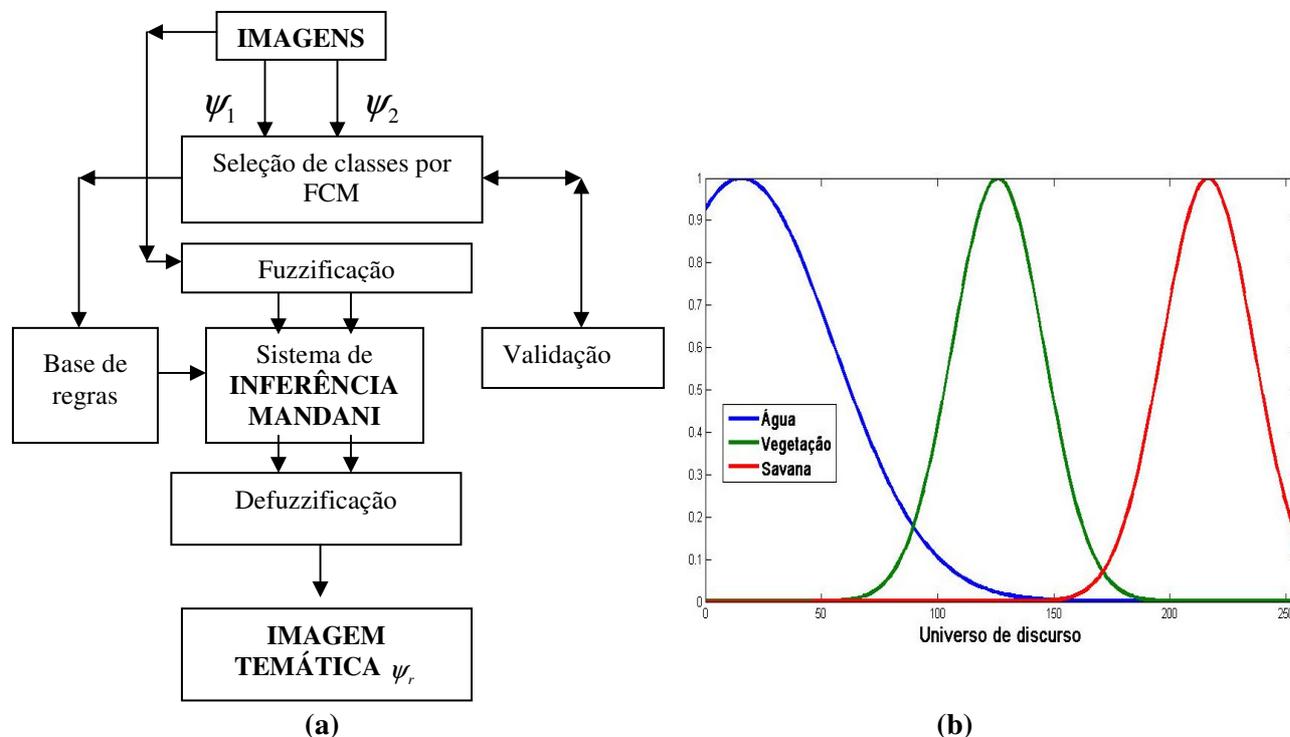


Figura 5: Sistema de Classificação Fuzzy (a) e Funções de pertinência (b) com suportes obtidos dos intervalos de classes calculados pelo algoritmo FCM.

O universo de discurso para cada uma das variáveis é o intervalo [0, 255], que corresponde aos valores de níveis de cinza. Para a imagem da **Figura 1.(a)**, o estudo das bandas 3 e 4 revelam a existência de três classes distintas: água, vegetação densa e as savanas. Após a aplicação do algoritmo FCM e a validação para três classes, obtém-se 3 intervalos que servirão de suporte para as funções de pertinência de cada variável de entrada. Na **Tabela 1.(a)**, estão os valores correspondentes aos intervalos de classe e o centro dos clusters. Para a segunda imagem, **Figura 1.(b)**, as componentes espectrais evidenciam apenas duas classes distintas: vegetação densa e savanas. O resultado obtido pelo FCM são dois intervalos suporte, que podem ser visualizados na **Tabela 1.(b)**.

Tabela 1-Intervalos e centro de classes para a banda 3.

(a)

CLASSE	SUPORTE	CENTRO (B3, B4)
ÁGUA	0-75	(18, 13)
VEGETAÇÃO	60-190	(141, 208)
SAVANA	125-255	(208, 94)

(b)

CLASSE	SUPORTE	CENTRO (B3, B4)
VEGETAÇÃO	0-140	(86, 198)
SAVANA	95-255	(200, 41)

4. Resultados

Para melhor verificação do bom funcionamento do algoritmo FCM, sobrepomos os pontos de cada cluster em uma das bandas da imagem original (**Figura 1.(a) e 1.(b)**), o que pode ser visto na **Figura 6**.

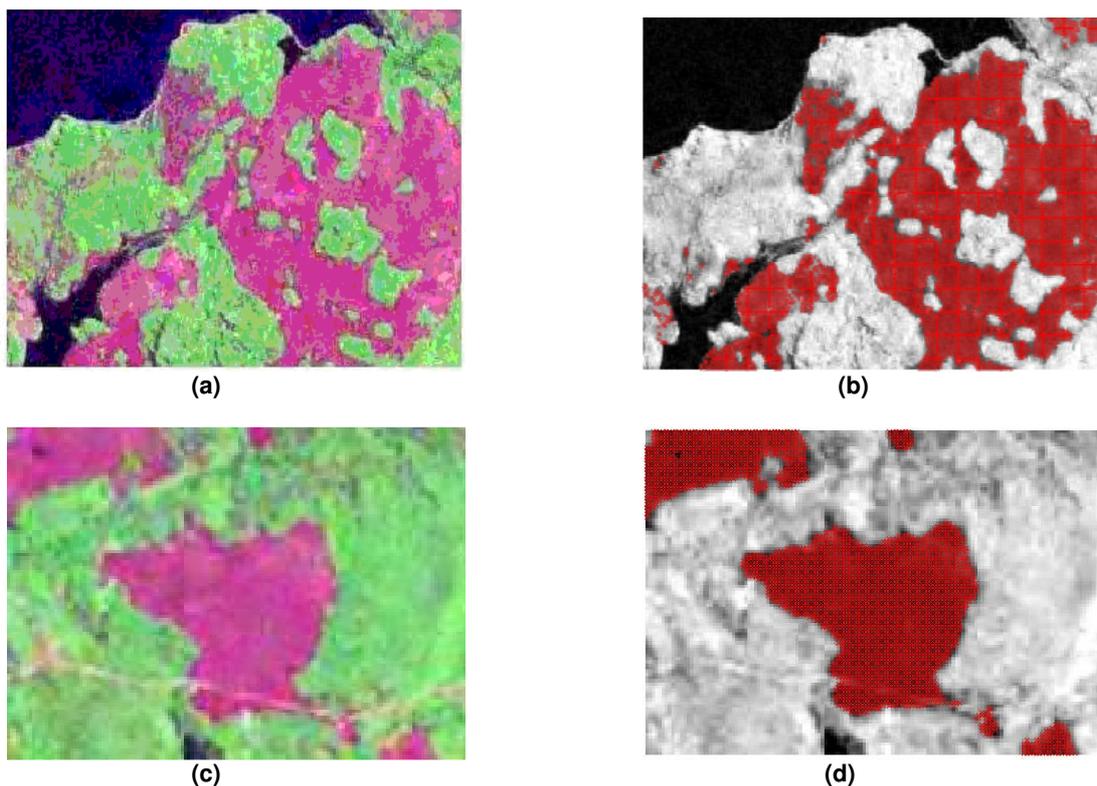


Figura 6: Na figura (a) a composição colorida TM. Em (b), os pontos do cluster, referente à zona de savana, sobreposto sobre a banda do infravermelho próximo. Em (c) e (d), um recorte de savana e o cluster associado.

O algoritmo mostrou-se eficiente para separar as zonas de savanas. Nota-se também, a presença de alguns conjuntos de pontos não que não fazem parte do alvo savanas. Estas zonas apresentam comportamento espectral semelhantes às savanas para a banda do infravermelho próximo, i.é., zonas de baixa refletância. Na **Figura 6.(d)** temos a visualização do cluster "savana" referente a imagem da **Figura 6.(c)**. Observe que a sombra da nuvem na parte inferior da imagem e a estrada que corta a savana (**Figura 6.(c)**) não fazem parte desta classe. Como dito anteriormente, a utilização do FCM foi a base para montar um sistema Fuzzy de classificação (**Figura 5.(a)**), em que as funções de pertinências (**Figura 5.(b)**) foram construídas através de intervalos extraídos dos clusters obtidos. Assim, podemos considerar o classificador como não supervisionado. Na **Figura 7.(b)** temos o resultado de classificação Fuzzy para 3 classes, referente à imagem da **Figura 1.(a)**.

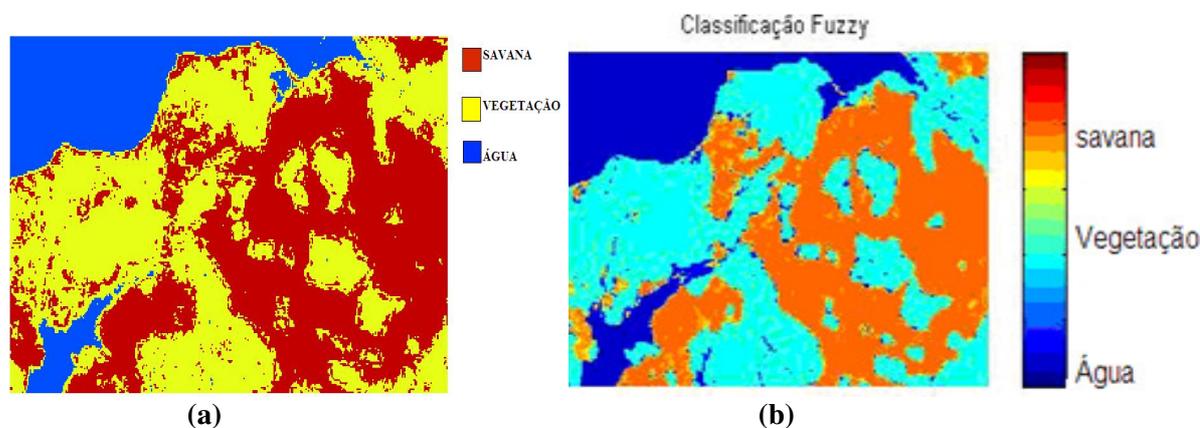


Figura 7: Classificação Binária (a) e Fuzzy (b) para três classes, referente a imagem da **Figura 1.(a)**.

Comparando este resultado com a classificação binária (**Figura 7.(a)**) podemos claramente observar a transição dos biomas floresta e savanas, através das bordas amarelas que acompanham as feições em laranja. Na classificação binária, a inserção dos pixels em cada classe é rígida como pode ser observado na (**Figura 7.(a)**).

5. Conclusões

De uma forma geral, o sistema classificador funcionou satisfatoriamente. As classes majoritárias, (óbvias), validades pela utilização dos índices Dunn's foram detectadas facilmente pelo algoritmo FCM. Quanto aos alvos savanas, o sistema obteve resultados satisfatórios, pois elas eram feições bem diferenciadas na imagem. O FCM foi capaz de extrair automaticamente os intervalos de classe, entretanto, o sistema Fuzzy de classificação apresenta uma dificuldade para tratar algumas feições que geram incertezas adicionais (sombras e nuvens). Este problema poderá ser resolvido, em trabalhos futuros, onde passaremos a trabalhar com regiões segmentadas, e não apenas com pixel's individuais. Comparado ao classificador binário, o sistema fuzzy proposto é muito mais robusto e isto se deve à possibilidade de agregar incertezas. Como estamos trabalhando com dados reais, faz-se necessário um método mais flexível a variações, que leve em consideração mudanças não bruscas entre biomas.

6. Referências

- Meirelles, M. S. P.; Câmara, G.; Almeida, C. M, **Geomática - Modelos e Aplicações Ambientais**. Brasília, DF: EMBRAPA Informação tecnológica, 2007. 593 p.
- Gomide, F.A. Campos; Gudwin, Ricardo R., 1994, **Modelagem, Controle, Sistemas e Lógica Difusa**; SBA Controle & Automação/ vol.4.
- Aguiar, H., Oliveira, JR., 1999. **Lógica Difusa – Aspectos Práticos e Aplicações**; Editora Interciência, 1a Edição.
- Barron, J. J., 1993, **A Lógica Fuzzy em Foco**, Byte No 06, pág. 94 – 100.
- Florenzano, Tereza Galloti. **Imagens de Satélites para Estudos Ambientais**. Oficina de Textos, São Paulo, 2002.
- Gonzalez, R. C. E Wints, P. **Digital image processing**. 2ed., Massachusetts, Addison Wesley, 1987.
- Fonseca, E.L. **Métodos para Classificação de Imagens Digitais Multiespectrais: Classificadores supervisionados e não supervisionados**. 2008.
- Holanda, A.S; Guerra, C.E Suemitsu,C. Uso de imagens NDVI e Modelo de Mistura Espectral para o monitoramento de Savanas no município de Santarém-PA. Anais, SBGF 2009, Salvador, Bahia, 2009.
- Caldas, N.S & Guerra, C.E. Limiarização Fuzzy aplicada a Imagens de Satélite. Anais, Siges, 2009, Santarém, Pará.
- Ponzoni, F.J. & Y.E. Shimabukuro. **Sensoriamento Remoto no Estudo de Vegetação**. Oficina de Textos, São Paulo. 2007.
- Florenzano, Tereza Galloti. **Imagens de Satélites para Estudos Ambientais**. Oficina de Textos, São Paulo, 2002.
- Gonzalez, Rafee C.; Woods, Richard E. e Eddins, Steven L. **Digital Image Processing Using MATLAB**. Prentice Hall, 2003.