

# Identificação de áreas agrícolas para manejo diferenciado utilizando o algoritmo ckMeansImage

Rogério R. de Vargas<sup>1</sup>, Ricardo Freddo<sup>1</sup>, Alexandre Russini<sup>1</sup>, Natália C. de Amorim<sup>1</sup>, and Cristiano Galafassi<sup>1</sup>

Laboratório de Sistemas Inteligentes e Modelagem (LabSIM), Federal University of Pampa, Itaqui-RS/Brazil  
[www.labsim.unipampa.edu.br](http://www.labsim.unipampa.edu.br)

**Resumo** Os avanços tecnológicos produziram várias ferramentas para auxiliar a tomada de decisões para profissionais de ciências agrícolas. As tecnologias que apresentam grande potencial para o setor agrícola são o Sensoriamento Remoto, que possui diversas ferramentas para extração e análise de informações, sendo uma delas o procedimento já realizado. A alta complexidade dos componentes antrópicos e naturais faz com que as rotações computacionais fiquem confusas e se tornem ferramentas com grande potencial em suas aplicações. Desta forma, o trabalho aplica o algoritmo ckMeansImage em uma imagem multiespectral com o objetivo de detectar áreas para cultivo. O algoritmo processou uma imagem da região de estudo em 4 grupos, sendo os seguintes: Gramíneas e leguminosas, remanescentes de arroz irrigado, pós-colheita de soja e acúmulo de água. Por meio dos resultados conclui-se que o algoritmo apresenta resultados satisfatórios, permitindo uma interpretação da representação de cada grupo.

**Palavras-chave:** Agrupamento, ckMeansImage, Fuzzy, Sensoriamento Remoto.

## 1 Introdução

A agricultura tem passado por inúmeras transformações tecnológicas nas últimas décadas, desde a introdução da eletrônica embarcada em máquinas agrícolas à Agricultura de Precisão.

Esta última caracterizada como sendo um conjunto de ferramentas que permite aos agricultores terem um diagnóstico da sua propriedade dos mais diferentes fatores a campo inerentes ao processo produtivo e intervir de forma precisa caso seja necessário. Isso demanda o conhecimento das diferentes áreas, desde o solo, plantas, irrigação, mecanização, tecnologia da informação, dentre outras, para que as variáveis do sistema produtivo possam ser identificadas, analisadas e intervenções sejam propostas.

No entanto, para que isto seja possível, são necessários softwares específicos, exigindo habilidades dos profissionais referentes a programação e

interdisciplinaridade de conhecimento. Na pesquisa, temos inúmeros trabalhos sendo desenvolvidos pelas mais diferentes instituições de pesquisa no País, na qual buscam o desenvolvimento de tecnologias digitais para auxílio a tomada de decisão nas atividades agropecuárias, uma destas tecnologias tem sido os produtos oriundos do Sensoriamento Remoto.

O sensoriamento remoto para aplicação na agricultura tem seus estudos datados desde a década de 1970. Para uma cultura agrícola específica ou para um conjunto de culturas numa determinada área de interesse, podendo variar em nível de municipal, estadual, nacional ou continental [1].

A tecnologia de sensoriamento remoto denota um vasto potencial para ser aplicada na agricultura. Com as técnicas de sensoriamento remoto, é possível adquirir informações sobre estimativa de área plantada, produção agrícola, vigor vegetativo das culturas, além de fornecer subsídios para o manejo agrícola em vários níveis geográficos. Ainda, os satélites imageadores possuem uma periodicidade na captura de imagens de uma determinada área (resolução temporal), proporcionando a obtenção de dados de uma mesma área agrícola em diferentes estágios do cultivo, acompanhando o ciclo de crescimento e desenvolvimento, permitindo gerar dados multitemporais [2].

O Sensoriamento Remoto pode ter sua atuação como ciência interdisciplinar largamente expandida quando empregado juntamente com as soluções computacionais, conforme apresentado no trabalho de [3]. Inovações em ferramentas para extração e análise de informações do grande volume de dados produzidos pelos sensores orbitais são contribuições determinantes para as ciências da terra [4].

Uma das ferramentas utilizadas para o estudo de fenômenos e objetos através das imagens multiespectrais, conforme [5], é o procedimento conhecido como Classificação de imagens.

Menezes e Sano definem a Classificação de imagens como o processo de atribuir uma ou mais classes a cada pixel de uma imagem, sendo que estas classes representam objetos ou fenômenos do mundo real.

Porém, a natureza e seus fenômenos são complexos, o que torna difícil seu estudo no âmbito da lógica booleana, por este motivo, a aplicação de rotinas computacionais em lógica fuzzy apresentam grande potencial na classificação de imagens multiespectrais [5].

Neste contexto, este trabalho aplica o algoritmo `ckMeansImage` [6] em uma imagem multiespectral com o objetivo de detectar áreas para cultivo, bem como demonstrar a potencialidade que o método apresenta quando utilizado como ferramenta de apoio ao processo de produção agrícola.

A Seção 2 descreve o algoritmo Fuzzy `ckMeans`; a metodologia para este trabalho é mostrado na Seção 3; a proposta na Seção 4 mostra o uso do algoritmo fuzzy para estabelecer diretrizes para o cultivo; a Seção 4 discute o uso do algoritmo em um estudo de caso; e por fim, as conclusões e trabalhos futuros são mostrados na Seção 5.

## 2 Algoritmo Fuzzy ckMeans

Nesta seção nós apresentamos o algoritmo fuzzy ckMeans, proposto por [7]. Este algoritmo é uma hibridização entre os algoritmos K-Means [8] e o Fuzzy C-Means [9], sendo alterado a forma de calcular o centro dos grupos.

### 2.1 Algoritmo ckMeans

O objetivo do algoritmo é dividir o conjunto  $X = \{x_1, x_2, \dots, x_n\}$  de  $n$  dados em  $p$  clusters,  $\mu_{ij}$  é o grau de pertinência da amostra  $x_i$  ao  $j$ -ésimo cluster e o resultado do agrupamento é expresso pelos graus de pertinência na matriz  $\mu$ .

O algoritmo ckMeans busca encontrar conjuntos nos dados, minimizando a função objetivo mostrada na Equação (1):

$$J = \sum_{i=1}^n \sum_{j=1}^p \mu_{ij}^m d(x_i; c_j)^2 \quad (1)$$

onde:

- $n$  é o número de dados;
- $p$  é o número de clusters considerados no algoritmo o qual deve ser decidido antes da execução;
- $m > 1$  é o parâmetro da fuzzificação;
- $x_i$  um vetor de dados de treinamento, onde  $i = 1, 2, \dots, n$ . Cada posição no vetor representa um atributo do dado;
- $c_j$  é o centro de um agrupamento fuzzy ( $j = 1, 2, \dots, p$ );
- $d(x_i; c_j)$  é a distância.

A seguir mostra-se os passos do algoritmo ckMeans que possui como entrada os dados  $n$ , o número de clusters  $p$  e o valor de fuzzificação  $m$ . Os passos são:

1. Inicialize  $\mu$  com um valor aleatório contínuo entre zero (nenhuma pertinência) e um (pertinência total), onde a soma das pertinências deve ser um;
2. Calcule o centro do cluster  $j$  da seguinte maneira: Cria-se uma nova matriz  $\mu$ , chamada de  $\mu_{Crisp}$  contendo valores 1 ou 0. Cada linha dessa nova matriz tem 1 na posição do maior valor dessa linha na matriz  $\mu$  e zero nas demais posições da linha. Quando uma coluna da matriz  $\mu_{Crisp}$ , for toda com zeros, é atribuído o valor 1 na posição que corresponde ao maior valor dessa mesma coluna na matriz  $\mu$ .

O algoritmo ckMeans retorna uma matriz  $\mu_{Crisp}$  com valores em  $\{0, 1\}$  conforme é mostrado na Equação (2). Ou seja,  $\mu_{Crisp}$  é a matriz enquanto  $\mu_{Crisp_{ij}}$  é o conteúdo dessa matriz na posição  $(ij)$ .

$$\mu_{Crisp_{ij}} = \max \left( \left\lfloor \frac{\mu_{ij}}{\max_{l=1}^p \mu_{il}} \right\rfloor, \left\lfloor \frac{\mu_{ij}}{\max_{l=1}^n \mu_{lj}} \right\rfloor \right) \quad (2)$$

O primeiro argumento do *max* tem que cada dado tenha o valor 1 no *clusters* ao qual pertence com maior grau de pertinência e grau de pertinência zero nos demais. O segundo argumento tem por objetivo que o maior grau de cada coluna (*clusters*) seja 1. Para assim garantir que todo *cluster* tenha pelo menos um elemento. Dessa forma, em raras ocasiões, pode acontecer que uma linha tenha mais de uma valor 1 (o que não ocorre no algoritmo k-means original), mas como esta matriz é apenas auxiliar, não ocasionará qualquer transtorno.

Os passos do algoritmo para calcular o  $\mu\text{Crisp}_{ij}$  é realizado da seguinte forma:

- (a) Leia  $\mu$ ;
- (b) Em cada linha encontrar o maior valor da matriz  $\mu$  e atribuir 1 a essa mesma posição em  $\mu\text{Crisp}$  e zero nas restantes;
- (c) Armazenar em um vetor a quantidade de 1's que cada coluna de  $\mu\text{Crisp}$  possui.

Se uma coluna não tiver 1's marque sumariamente com 1 a posição onde está o maior valor. Após calculada a matriz  $\mu\text{Crisp}$  calculam-se os novos centros dos *clusters* conforme a Equação (3).

$$c_j = \frac{\sum_{i=1}^n x_i \mu\text{Crisp}_{ij}}{\sum_{i=1}^n \mu\text{Crisp}_{ij}} \quad (3)$$

O  $c_j$  é calculado pelo somatório dos dados que pertencem ao *cluster* (de forma crisp) e dividido pela quantidade de objetos classificados como 1 na matriz  $\mu\text{Crisp}$  deste *cluster*.

3. Calcule um valor inicial para  $J$  usando a Equação (1);
4. Calcule a tabela da função de pertinência *fuzzy*  $\mu$  conforme mostrado na Equação (4)

$$\mu_{ij} = \frac{\left(\frac{1}{d(x_i; c_j)}\right)^{\frac{2}{m-1}}}{\sum_{k=1}^p \left(\frac{1}{d(x_i; c_k)}\right)^{\frac{2}{m-1}}} \quad (4)$$

5. Retornar a etapa 2 até que uma condição de parada seja alcançada.

Condição de parada:

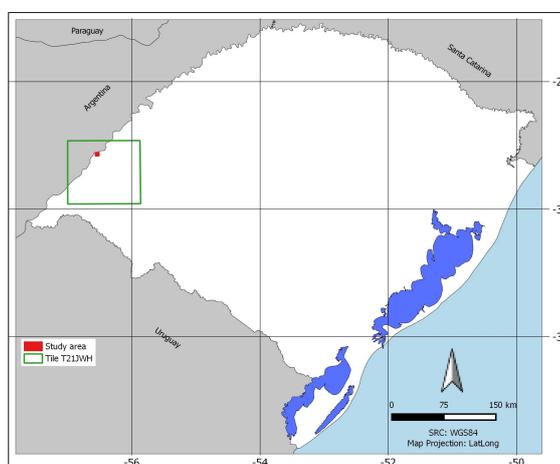
- O usuário informa um valor de parada  $\epsilon > 0$ , e se

$$d(J_U; J_A) \leq \epsilon$$

então para, onde  $J_A$  é a função objetivo (Equação (1)) calculada na iteração anterior e  $J_U$  é a função objetivo da última iteração.

### 3 Metodologia

A área coberta pela imagem (obtida no Google Earth, disponível em <http://www.google.com/earth> é datada em 14/03/2016) selecionada para o presente experimento, representa a cidade de Itaqui, localizado na fronteira oeste do Rio Grande do Sul, Brasil (Figura 1). A região é caracterizada por um relevo plano, em que a cidade está situada entre os Rio Uruguai ao norte, a corrente do Cambaí a leste, e o fluxo de olarias a oeste. O ponto central aproximado da imagem de estudo é definido com as seguintes coordenadas: 29° 10' 09.43" S e 56° 34' 34.76" W.



**Fig. 1.** Área de estudo.

Em um trabalho de campo, com um receptor GNSS (Global Navigation Satellite System) foram rastreadas as seguintes coordenadas, conforme é mostrado na Tabela 1 e Figura 5.

**Table 1.** Coordenadas geodésicas.

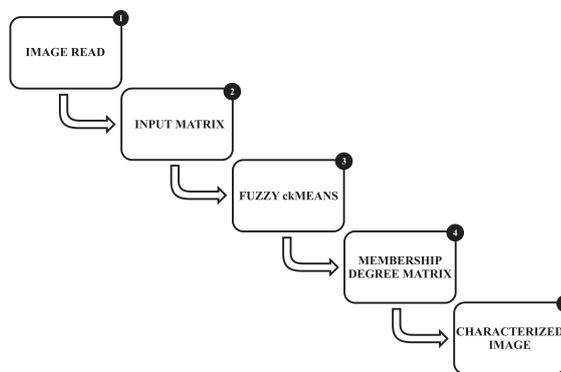
Latitude ( $\phi$ )	Longitude ( $\lambda$ )
29° 10' 07.73" S	56° 34' 36.44" W
29° 10' 02.63" S	56° 34' 32.25" W
29° 10' 05.64" S	56° 34' 20.87" W
29° 10' 09.39" S	56° 34' 19.85" W
29° 10' 11.32" S	56° 34' 21.53" W

As coordenadas geodésicas formam um polígono que engloba uma área de cultivo de soja. Para fins da aplicação da metodologia proposta, o polígono deverá estar compreendido em um único grupo.

Após o levantamento em campo, processado a imagem, um especialista do conhecimento realiza a interpretação da imagem discutindo as características dos grupos formados.

## 4 Proposta

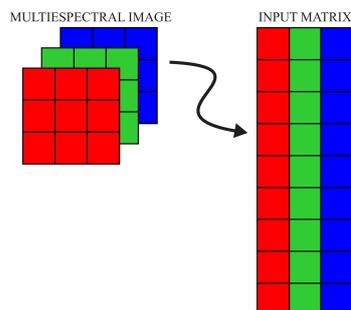
O processo, desde a leitura da imagem até a imagem caracterizada, tem os seguintes passos:



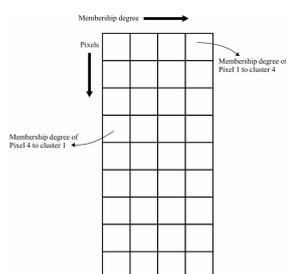
**Fig. 2.** Passos do algoritmo Fuzzy ckMeansImage

1. Image Read - extrair o número digital dos pixels da imagem (RGB);
2. Input Matrix - converter o número digital de cada pixel para três colunas, formando uma matriz, desse modo cada coluna representa uma banda da imagem (RGB). O processo é mostrado na Figura 3. É importante mencionar que a saída do algoritmo Fuzzy ckMeansImage possui um grau de pertinência para cada linha na matriz de entrada;
3. Fuzzy ckMeans - executa o Fuzzy ckMeans com a matriz de entrada;
4. Membership Degree Matrix - matriz resultante com o grau de pertinência de cada pixel em cada linha;
5. Characterized Image - é gerada uma matriz com grau de pertinência para cada banda (RGB), desse modo é possível executar o processo de recomposição, atribuindo as novas cores para cada grupo. No final do processo, o algoritmo Fuzzy ckMeansImage retorna uma imagem agrupada.

A ferramenta utilizada que contém o Algoritmo Fuzzy ckMeansImage pode ser acessada no seguinte endereço *web*: <http://labsim.unipampa.edu.br/labsim/ckMeansImage/>.



**Fig. 3.** Processo de decomposição.



**Fig. 4.** Processo de recomposição.

## 5 Resultados

O experimento consiste em processar uma imagem onde observa-se áreas para o cultivo de soja e arroz, conforme pode ser visto na Figura 5.

A imagem processada, foi gerada usando o algoritmo Fuzzy ckMeansImage com 4 grupos, conforme pode ser observada na Figura 6.

Por meio de uma análise detalhada da imagem pode-se identificar, através da diferenciação de cores, diversas superfícies compostas desde campo nativo, áreas de arroz irrigado, solo exposto e resteva da cultura da soja.

Nas áreas verde-escura destacamos a superfície composta de gramíneas e leguminosas, composição esta característica de campo nativo do bioma pampa, predominante na Fronteira Oeste do Rio Grande do Sul.

A coloração em destaque verde-claro, caracteriza-se pela resteva da cultura do arroz irrigado, em processo inicial de rebrote, não tendo mais valor econômico neste estágio, sendo apenas utilizada como pastagem para gado de corte e/ou cobertura de solo.

A coloração marrom claro representa áreas de soja após colheita (resteva), podendo-se desta forma delimitar áreas e as diferencia-las em função da superfície e/ou cobertura. A partir de uma análise mais detalhada da imagem se permite identificar áreas cultivadas com soja, com depressões superficiais de solo, onde ocorre acumulação de água, destacada na imagem pela coloração marrom-escuro.



**Fig. 5.** Área de estudo.



**Fig. 6.** Imagem processada com o algoritmo Fuzzy ckMeansImage.

Estas representações foram observadas a campo, em visita em loco, também observada na área cultivada com arroz e no campo nativo, tendo-se desta forma precisão nas inferências. O acúmulo de água prejudica o desenvolvimento da cultura da soja, podendo-se desta forma através do processamento de uma imagem de satélite por meio da utilização do algoritmo c-means, realizar uma pré-identificação destes pontos e realizar um planejamento, bem como uma intervenção localizada.

## 6 Conclusão

O presente trabalho mostra uma aplicação do algoritmo Fuzzy ckMeansImage na identificação em áreas de plantio. O algoritmo mostrou resultados satisfatórios, por mapear áreas em grupos, permitindo ao especialista de conhecimento inferir o que cada grupo representa pela sua coloração.

A intervenção localizada reduz de forma considerável o custo operacional para movimentação de solo e sistematização, que quando não realizada uma pré-identificação, acaba sendo realizada em área total, operação está não sendo mais necessária em todos os casos. Isso também permite realizar um planejamento, caso não seja realizada a intervenção, para que esses pontos com excesso de umidade não sejam cultivados, economizando insumos, sementes dentre outros produtos, pois são áreas que terão nenhuma, ou em alguns casos, produção de grãos que se quer cobrem os custos.

Como estudo futuro, pretende-se aplicar o algoritmo Fuzzy ckMeansImage utilizando um novo parâmetro como entrada (conhecido como  $\alpha$ -corte) para detectar os *pixels* imprecisos no processo do agrupamento, possibilitando identificar com maior cautela pontos em que o algoritmo não classificou com determinado grau de certeza.

## References

1. Formaggio, A., Sanches, I.: Sensoriamento remoto em Agricultura:. Oficina de Textos (2017)
2. RUDORFF, B. F. T.; MOREIRA, M.A.: 9. In: Sensoriamento remoto aplicado À agricultura. (2002) 20
3. Honda, B., Jorge, L.A.C.: Computação aplicada a agricultura de precisão. Revista Científica Eletrônica UNISEB. Ribeira Preto 1(1) (2013)
4. Plaza, A., Du, Q., Chang, Y.L., King, R.L.: High performance computing for hyperspectral remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 4(3) (2011) 528–544
5. de Vargas, R., Galafassi, C., de Amorim, N.C., Freddo, R.: Algoritmo ckmeans aplicado ao sensoriamento remoto. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics 5(1) (2017)
6. Neto, R.F., De Vargas, R.R., Gass, S.L.B., Galafassi, C., Russini, A., de Amorim, N.C.: Identificação de uma área inundada do rio uruguai por um algoritmo fuzzy. Anais do Salão Internacional de Ensino, Pesquisa e Extensão 9(7) (2018)
7. de Vargas, R., Bedregal, B.: A comparative study between Fuzzy C-Means and ckMeans algorithms. In: Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2010), Toronto, Canada (2010)
8. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, University of California Press (1967) 281–297
9. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA (1981)