

# Desenvolvimento de uma Ferramenta Web para a Execução do Algoritmo Fuzzy ckMeans no Processamento de Imagens

R. Freddo, R. Vargas, C. Galafassi, A. Russini e T. Pinto

**Resumo**—O sensoriamento remoto é um conjunto de métodos de captação de informações dos fenômenos e feições terrestres, por meio de sensores, sem contato direto, sendo indispensáveis as metodologias e técnicas de armazenamento, tratamento, análise e interpretação destas informações. Este trabalho detalha um método, disponibilizado por meio de uma ferramenta online com interação gráfica cliente/servidor, que decompõe uma imagem de bandas espectrais em um vetor coluna. Mostra o modelo matemático formal do algoritmo Fuzzy ckMeans e o processo de sobreposição para o agrupamento de uma imagem dado um número de *clusters*. Para validar a proposta aplicou-se o algoritmo em duas imagens capturadas por satélite, sendo uma área utilizada para o cultivo do arroz e outra área que abrange a confluência de dois rios localizados na fronteira do estado do Rio Grande do Sul. Resultados do agrupamento são discutidos por um especialista em sensoriamento remoto.

**Index Terms**—Agrupamento, ckMeans, Sensoriamento Remoto.

## I. INTRODUÇÃO

**A**GRUPAR para classificar é uma das habilidades mais básicas do ser humano, sendo que o pensamento de agrupar objetos ou características semelhantes em categorias é algo muito antigo, primitivo. Uma vez que o homem primitivo foi capaz de separar, por exemplo, as plantas venenosas das não venenosas, animais perigosos e os não perigosos, entre outras classificações. Cada definição de grupo gera subgrupos que podem ser rotulados, como no exemplo dos animais, em cães, gatos, cavalos, entre outros [2].

Segundo [4], a análise de grupos divide os dados baseado em similaridades de alguma característica. Esse processo desempenha papel importante em vários campos de atuação, tais como, psicologia e outras ciências sociais, biologia, estatística, reconhecimento de padrões, recuperação de informações, aprendizagem de máquinas e mineração de dados. Na biologia, aplica-se agrupamento de dados para analisar grandes quantidades de informação genética, por exemplo, encontrar grupos de genes com funções semelhantes. Na recuperação de informações utilizam-se técnicas de agrupamento para filtrar e agilizar pesquisas na internet, se realizada uma pesquisa por filmes, por exemplo, pode-se retornar páginas agrupadas em categorias, como suspense, terror, humor, entre outros, podendo ainda ser dividida em subcategorias, hierarquizando a pesquisa e facilitando a utilização. Para entender o clima também é necessário reconhecer padrões na atmosfera e nos

oceanos, tais como pressão atmosférica em dadas regiões e áreas oceânicas com maior impacto no clima. Na psicologia e na medicina, as doenças ou condições com muita frequência apresentam número de variações, a análise de grupos pode ser utilizada para identificar essas duas subcategorias, podendo ainda ser usada nessas áreas para identificar padrões na distribuição espacial e temporal de doenças. Para aplicações no meio empresarial, o agrupamento de dados pode ser usado para separar os clientes em subgrupos, permitindo realizar análises adicionais e de marketing.

Outra importante aplicação do agrupamento de dados está centrada no processamento de imagens, onde sua aplicação foi iniciada na década de 60 pela Agência Espacial Norte Americana (NASA). Posteriormente, surgindo muitas áreas que necessitassem da aplicação do processamento de imagens, tais como, a medicina, a microscopia, a meteorologia, o televisionamento e a indústria [5]. O processamento de imagens digitais é uma área de recente crescimento e que se destaca pelas suas várias aplicações, dentre elas a área do processamento de imagens aplicada no sensoriamento remoto.

Segundo [6], o sensoriamento remoto é caracterizado por um conjunto de atividades que torna possível a obtenção de informações de objetos que compõem a superfície terrestre sem que seja necessário o contato direto com o local desejado. Tem como principal objetivo obter informações e codificá-las em tabelas, gráficos ou mapas, tornando a leitura dessas informações mais simples. Em outras palavras, o processamento de imagens aplicado ao sensoriamento remoto procura melhorar o aspecto visual da imagem, proporcionando maior facilidade no reconhecimento de padrões e posterior retirada de informações da imagem processada [7]. Os mesmos autores ressaltam em seu trabalho que a obtenção das informações de sensoriamento remoto necessita de métodos e análises claras e lógicas para que possam ser aplicados a qualquer produto, tornando possível a análise de diferentes dados por diferentes profissionais.

Segundo [8], processar uma imagem é um método de transformações sucessivas para que se obtenham informações presentes nela com maior facilidade. O processamento de imagens está dividido em categorias, que são o realce de imagens (melhora a visualização das imagens), restauração de imagens (objetiva reconstruir imagens danificadas) e análise de imagens (extrair informações das imagens). Geralmente é preciso unir estas três categorias para que um resultado de qualidade seja obtido [9]. Na última etapa do processamento de imagens, a análise da imagem, ocorre a busca de detalhes

que formem padrões, dando rótulo à imagem.

Para a realização do processamento das imagens existem softwares comerciais como o ENVI, TerrSet e Erdas, tendo aplicação na visualização, exploração, análise e apresentação de dados na área do Sensoriamento Remoto/SIG. Contudo, apresentam a dificuldade de aquisição por meio de compra de licença e posterior instalação. Além desses, existem programas gratuitos para o mesmo fim, como no caso do Spring. O Spring é um SIG (Sistema de Informações Geográficas) com funções de processamento de imagens, análise espacial, modelagem numérica de terreno e consulta de banco de dados espaciais. Apesar de ser gratuito, esse programa requer um processo de cadastramento no site do INPE (Instituto Nacional de Pesquisas Espaciais) para que a aquisição seja possível por meio de download, necessitando também de posterior instalação pelo usuário.

Este artigo está assim dividido: Introdução que mostra conceitos sobre agrupamento de dados para classificação e análise de dados, bem como algumas aplicações práticas em torno deste tema. A seção II mostra os trabalhos relacionados na área. A seção III o algoritmo. A seção IV traz a proposta pelos autores. A seção V mostra os resultados e por fim, a última seção, é mostrado as considerações finais.

## II. TRABALHOS RELACIONADOS

A extração das informações de sensoriamento remoto demanda métodos e análises claras e lógicas para que possam ser aplicados a qualquer produto [7]. Uma das ferramentas de análise é a classificação que pode ser compreendida como uma técnica de reconhecimento de padrões representados em uma imagem multiespectral, através do agrupamento de pixels de forma que pixels semelhantes pertençam ao mesmo *clusters* [10].

Zadeh em [11] concluiu em seu trabalho que a lógica *fuzzy* torna-se a mais adequada quando é necessário trabalhar com informações que apresentam possíveis ambiguidades ou imprecisões. Assim, é possível notar o potencial dessa metodologia para ser aplicada no âmbito do sensoriamento remoto.

No trabalho de [12] aplicou-se um algoritmo *fuzzy* no processo de segmentação de imagens em que o caso de estudo foram imagens médicas, onde resultados preliminares foram discutidos. Entretanto, estudos referentes de como deuse a conversão na leitura e na saída da imagem não foram detalhados.

Neste contexto, este trabalho aplica o algoritmo *Fuzzy ckMeans* [13] a uma imagem multiespectral com o objetivo de avaliar os mapas temáticos digitais gerados, bem como ressaltar a potencialidade do método no campo do sensoriamento remoto. Entretanto, para executar uma imagem no algoritmo *ckMeans* torna-se necessário adaptar a imagem em um formato de que o algoritmo possa executá-lo. Então nesse artigo é proposto de uma forma detalhada como se dá a decomposição de uma imagem em bandas espectrais para ser executado no algoritmo *ckMeans* e como se dá a sobreposição da imagem em bandas gerando a imagem resultante.

Em trabalhos como [14], [15] e [16] os autores trabalham com o algoritmo *Fuzzy C-Means* aplicado ao sensoriamento

remoto. Pelo algoritmo *ckMeans* ser mais rápido que o algoritmo *Fuzzy C-Means* decidiu-se aplicar também na área de sensoriamento remoto.

## III. ALGORITMO

### A. *ckMeans*

A ideia basicamente é que o conjunto  $X = \{x_1, x_2, \dots, x_n\}$  seja dividido em  $p$  *clusters*,  $\mu_{ij}$  é o grau de pertinência da amostra  $x_i$  ao  $j$ -ésimo *clusters* e o resultado do agrupamento é expresso pelos graus de pertinência na matriz  $\mu$ .

O algoritmo *ckMeans* tenta 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 entrada do algoritmo são os  $n$  dados, o número de *clusters*  $p$  e o valor de  $m$ . 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 *clusters* 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 o

algoritmo k-means original), mas como esta matriz é apenas auxiliar, não ocasionará qualquer transtorno.

Os passos do algoritmo para calcular o  $\mu_{Crisp_{ij}}$  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  $\mu_{Crisp}$  e zero nas restantes;
- c) Armazenar em um vetor a quantidade de 1's que cada coluna de  $\mu_{Crisp}$  possui.

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

$$c_j = \frac{\sum_{i=1}^n x_i \mu_{Crisp_{ij}}}{\sum_{i=1}^n \mu_{Crisp_{ij}}} \quad (3)$$

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  $\mu_{Crisp}$  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.

Algumas condições de parada possíveis são:

- Um número de iterações pré-fixado for executado;
- 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.

**B. Pré-processamento e Pós-processamento**

O algoritmo para agrupamento de uma imagem dar-se pelo esquema conforme é mostrado na Figura 1.

Os processos desde a leitura da imagem (decomposição) até a imagem resultante (sobreposição), segue os seguintes passos:

- 1) Leitura da imagem;
- 2) A etapa de decomposição da imagem consiste em transformar uma matriz tridimensional de pixels em um vetor coluna  $S$ , conforme mostrado na Figura 2.

Uma vez que o algoritmo ckMeans retorna uma matriz na qual cada linha possui o grau de pertinência do pixel aos *clusters*, faz-se necessário converter a matriz de pixels em um vetor coluna. Sendo assim, deve-se montar a imagem para tornar possível a interpretação dos padrões obtidos pelo algoritmo de agrupamento;

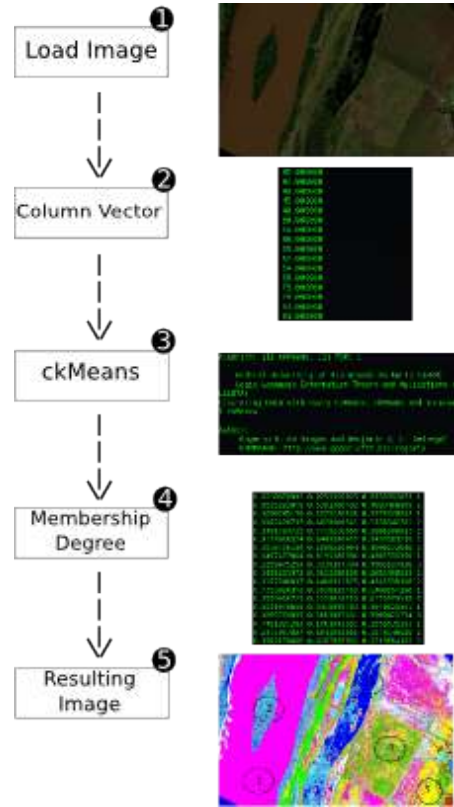


Figura 1. Etapas a serem realizadas para obter o agrupamento no sensoriamento remoto. Fonte: Autor.

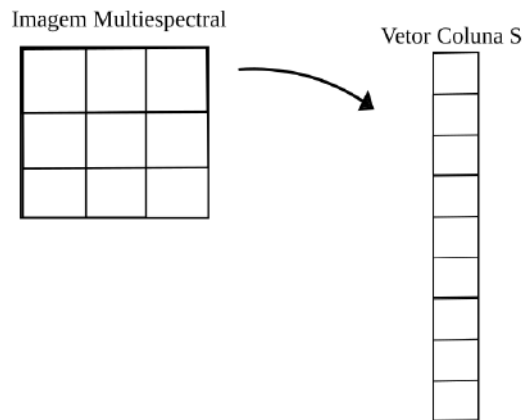


Figura 2. Processo de préprocessamento. Fonte: Autor.

- 3) Executa-se o algoritmo ckMeans conforme descrito na seção 2.1;
- 4) A matriz de grau de pertinência (denominada matriz  $C$ ) é gerada;
- 5) O algoritmo de sobreposição recebe a matriz  $C$  e verifica a qual *cluster* o pixel tem maior grau de pertinência, em cada banda espectral. Os passos do algoritmo são descritos abaixo:

- a) Cria uma matriz nula  $M$  com mesma dimensão da matriz  $C$ ;

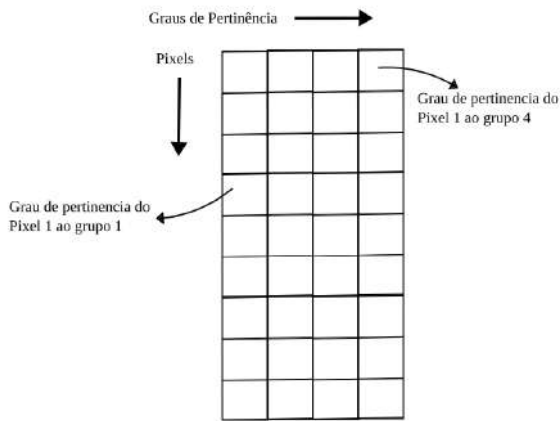


Figura 3. Processo de pós-processamento. Fonte: Autor.

- b) Avalia o conteúdo da matriz  $S$  e procura em cada linha da matriz  $S$  o *cluster* ao qual o pixel possui maior grau de pertinência (em cada banda espectral);
- c) Na matriz  $M$ , atribui valor 1 à posição do maior grau de pertinência e zero às demais posições. Note que assim é obtido uma matriz binária (com valores 0 e 1). Este procedimento é descrito a seguir e para melhor entendimento, a Figura 1 esquematiza todo o processo pelo qual a imagem é submetida, desde o procedimento de decomposição até a sobreposição;
- d) A matriz  $M$  é transformada em três novas matrizes, sendo uma matriz para cada banda espectral (*Red*, *Green* e *Blue*);
- e) As três matrizes são multiplicadas pelo escalar 255, a fim de realçar as diferenças entre os *clusters*, tendo valor máximo (255) ou mínimo (0);
- f) Com as bandas separadas, executa-se um procedimento de sobreposição, ou seja, as bandas são novamente unidas e a imagem é completamente montada.

IV. MODELO PROPOSTO

Para a aplicação do algoritmo citado na seção III, foi desenvolvida uma ferramenta web, ou seja, o algoritmo possui uma interação gráfica a partir do site. A ferramenta funciona partindo do princípio onde umas das partes é o cliente e outra o servidor. Segundo [3], cliente/servidor é um modelo onde ocorre troca de serviços, gerenciados por um provedor de acessos, em que o acesso dos dados pelo usuário se dá por uma interface gráfica.

Atualmente a ferramenta é livre e pode ser acessada através do endereço [labsim.unipampa.edu.br/ckMeans](http://labsim.unipampa.edu.br/ckMeans). O servidor possui um processador AMD A8-6500B APU com Radeon(tm) HD Graphics, 4 Gb de memória RAM, HD de 1 Tb, sistema operacional GNU Linux, distribuição Ubuntu 14.04.

O uso da ferramenta de processamento de imagens é através da interface do site desenvolvido na linguagem de

programação web PHP, versão 5.6.30. O usuário pode acessar pelo domínio na internet e então fazer *upload* da imagem e selecionar os parâmetros com os quais o algoritmo deve executar suas devidas ações. A ferramenta é de uso on-line, ou seja, pode ser acessada diretamente no site. Ao acessar o endereço web, basta apenas selecionar a imagem local. A partir desse processo a imagem é enviada para o servidor onde o algoritmo (desenvolvido na linguagem de programação C++) está hospedado, fazendo então o processamento da imagem. Após o processamento da imagem, o servidor disponibiliza a imagem pronta para o usuário, bastando apenas fazer o download da mesma. A Figura 4 apresenta o processo para a utilização da ferramenta e a geração da imagem. Os itens estão descritos abaixo.

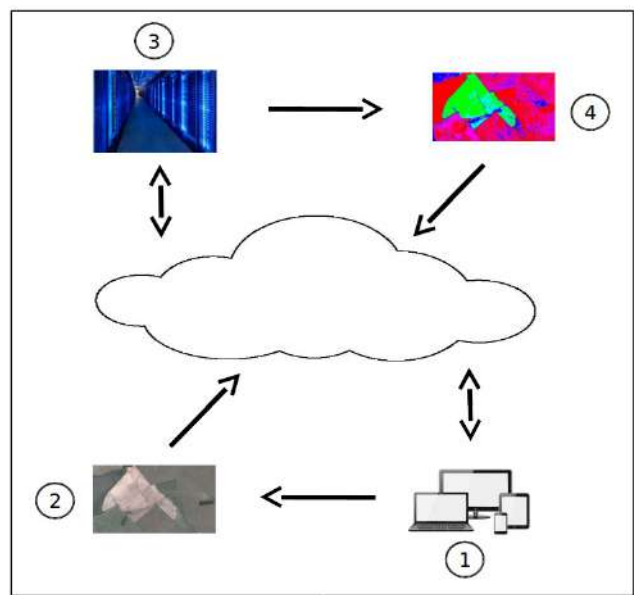


Figura 4. Esquema de funcionamento da ferramenta on-line. Fonte: Autor.

- 1) demonstra o primeiro passo, onde o usuário acessa a ferramenta através do dispositivo que melhor convier.
- 2) representa a imagem original que o usuário escolheu.
- 3) mostra o servidor, onde encontra-se o algoritmo que irá processar a imagem.
- 4) representa a última fase do processamento, ou seja, a imagem já processada, onde, a partir deste ponto ela será retornada para o usuário.

Na atual versão do algoritmo em [12] é possível executar apenas um parâmetro por vez, caso o usuário queira executar diversas inicializações (aleatórias) e alterar os parâmetros, é necessário executar o algoritmo/programa diversas vezes. A grande idealização desta ferramenta web além de não haver a necessidade de instalação de nenhum programa específico é a possibilidade de gerar uma fila de processos permitindo executar o algoritmo com parâmetros e inicializações distintos. No final de todo processo, é listado as imagens processadas de acordo com os parâmetros definidos a priori.

Observando a Figura 5, o usuário deverá preencher a priori os seguintes dados:

- 1) Selecionar a imagem a ser processada;
- 2) Definir a quantidade de *clusters*;
- 3) Informar o parâmetro *fuzziness*<sup>1</sup>;
- 4) Selecionar o critério de parada definido por  $\epsilon^2$ ;
- 5) O botão Run inicia o processamento das imagens;
- 6) O item *download* Mostra as imagens geradas pelo algoritmo.

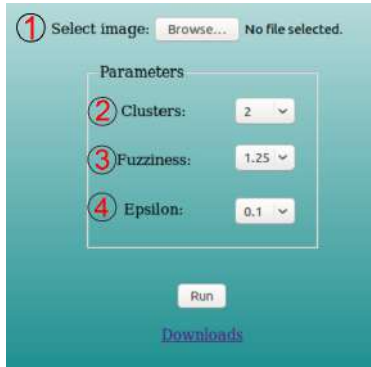


Figura 5. Interface gráfica do site. Fonte: Autor.

V. RESULTADOS

A fronteira oeste do Rio Grande do Sul (Figura 6) concentra as duas cidades de maior produção de arroz do Brasil [1]. Muitas pesquisas nessa área tem se realizado. Distinguir suas áreas em grupos para uma possível medição ou até mesmo verificar novas áreas de cultivo é muito importante para o produtor rural. Para tanto dois experimentos serão discutidos nos resultados.

O experimento A, proposto por [18] iniciou-se com a escolha da imagem a ser utilizada e sua aquisição através do catálogo de imagens do INPE, disponível no endereço <http://www2.dgi.inpe.br/CDSR>. Após o download das bandas da cena selecionada, as mesmas foram importadas no TerrSet, analisadas e foi realizado o recorte para a área de interesse, a qual está representada na Figura 7. O recorte resultou numa imagem com dimensões de 3155x1963x3 pixels, sendo as coordenadas do seu centroide 29° 24' 44,21087”S e 56° 38' 01,14281”W, recobrando uma área de 61932 hectares. A composição colorida 3-4-2 em RGB e as bandas individualizadas foram exportadas para o formato JPG para serem utilizadas como base para a aplicação do algoritmo ckMenas.

Cabe ressaltar que as motivações que levaram a equipe a optar por imagens do satélite CBERS-4 foram as seguintes:

- 1) o satélite foi desenvolvido com parcela de investimentos públicos brasileiros;
- 2) é o primeiro satélite da série com imagens com 10 metros de resolução espacial;
- 3) as imagens são de uso gratuito;
- 4) a pouca disponibilidade, até o momento, de estudos que utilizaram tais imagens.

<sup>1</sup>Valores entre o intervalo [1, 1; 3].

<sup>2</sup>O algoritmo esta definido para executar até 100 iterações caso o critério de parada não seja satisfeito.

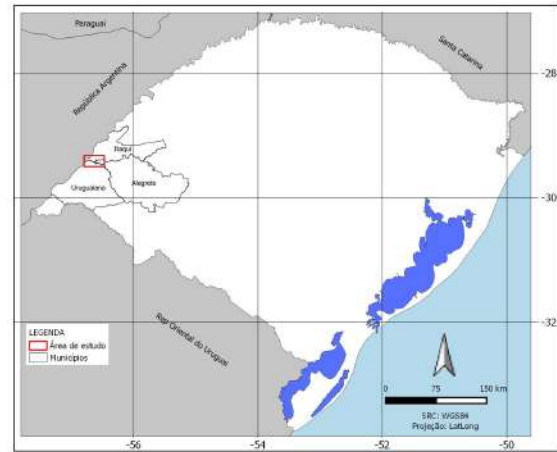


Figura 6. Localização da área de estudo. Fonte: [18]



Figura 7. Imagem com composição colorida 3-4-2 em RGB. Fonte: [18]

Os parâmetros iniciais do algoritmo ckMeans foram 4 *clusters*, a inicialização deu-se de forma aleatória e epsilon ( $\epsilon$ ) em 0,01 como critério de parada. A partir da Figura 7 é possível identificar os leitos dos rios Uruguai e Ibicuí, além de áreas com solo exposto, vegetação ciliar e áreas com cultivos agrícolas. Em função do grande acúmulo de chuvas ocorrido no período, verifica-se a presença de áreas inundadas que são o resultado do processo natural do mútuo represamento dos rios mencionados e do aumento dos reservatórios de água para a irrigação da cultura do arroz e sua interligação com o extravasamento fluvial.

A imagem resultante do algoritmo ckMeans representa quatro classes de usos do solo (*clusters*), que podem ser assim definidas:

- massas de água;
- solo exposto;
- uso agrícola 1;
- uso agrícola 2.

A classe massas de água representa os rios Uruguai e Ibicuí, os reservatórios para fins de irrigação, as áreas inundadas e em alguns casos os pixels nos quais há uma significativa presença de umidade, mesmo que seja possível identificar seu uso como agrícola.

O algoritmo ckMeans, no caso da classe em questão, conseguiu agrupar de maneira precisa os pixels que representam

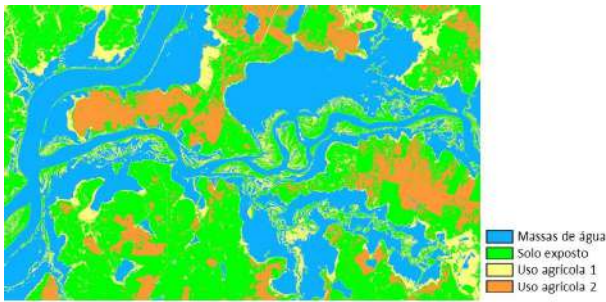


Figura 8. Resultado do processamento com o algoritmo ckMeans. Fonte: [18]



Figura 9. Imagem original. Fonte: Autor.

efetivamente água, incluindo em outras classes aquelas áreas que demonstram presença de umidade, mas que é perceptível a presença de algum tipo de vegetação. Desta forma, considerando a possibilidade do uso do algoritmo para a identificação de áreas inundadas, sua aplicação gera resultados satisfatórios. Ainda, o algoritmo ckMeans agrupou as áreas de solo exposto com as áreas em processo inicial de germinação, além de incluir áreas já em processo mais avançado de desenvolvimento, o que gera certa inconsistência na sua avaliação.

Para a classe uso agrícola 1, o algoritmo ckMeans associou basicamente os pixels que podem ser caracterizados como sendo de alta presença de umidade, mas ainda com possibilidade de identificação da vegetação.

A última classe definida para o experimento é a denominada de uso agrícola 2. Nesta classe devem estar presentes os pixels que representam aquelas áreas nas quais há uma maior presença de áreas agrícolas já mais desenvolvidas e as áreas de vegetação ciliar e eventuais áreas florestadas. O algoritmo ckMeans apresentou um resultado com alto grau de generalização, agrupando, em sua grande maioria, os pixels que representam as características aqui definidas para as duas últimas classes.

O experimento B, teve em sua primeira etapa uma análise realizada por um agrupamento da imagem, os resultados são discutidos a seguir. Obteve-se a imagem (veja Figura 9) a ser processada através do Google Earth®, que disponibiliza imagens de alta resolução espacial (50 cm), o que facilita o reconhecimento de padrões. A imagem é datada do dia 14 de agosto de 2015, possui dimensão de 968x532x3 pixels cujas as coordenadas aproximadas via GPS do centroide da imagem são 29°11'09.6''S e 56°25'15.8''W.

A partir da imagem enviada pelo usuário, o site retorna uma imagem processada, conforme as Figuras 9 e 10. Executou-se o algoritmo ckMeans em um computador com sistema operacional GNU Linux, distribuição Ubuntu 14.04, com 4 GB de memória ram, processador Core i5, 2.50 Ghz.

A figura 9 apresenta claramente as áreas de cultivo com arroz irrigado e pastagem natural predominantes na Fronteira Oeste do Rio Grande do Sul, podendo-se destacar o processo de preparo do solo para implantação da cultura (arroz irrigado), áreas com a cultura já implantada (em fase de desenvolvimento vegetativo). Observa-se também, reservatório de água utilizado para irrigação da cultura, bem como as áreas com pastagem natural, destinadas a pecuária extensiva, juntamente com bosques de eucalipto que servem como abrigo aos animais

e fonte de madeira para uso na propriedade.

Os parâmetros de entrada do algoritmo ckMeans foram:  $\epsilon = 0,001$ ,  $p = 5$ ,  $p = 6$  e  $p = 7$  (*clusters*) e  $m = 1,50$  (*fuzzificação*) por serem os valores mais usuais [19].

As Figuras 10, 11 e 12 mostram as segmentações das imagens obtidas com o algoritmo ckMeans com o número de *clusters* variando em  $p = 5$ ,  $p = 6$  e  $p = 7$ .

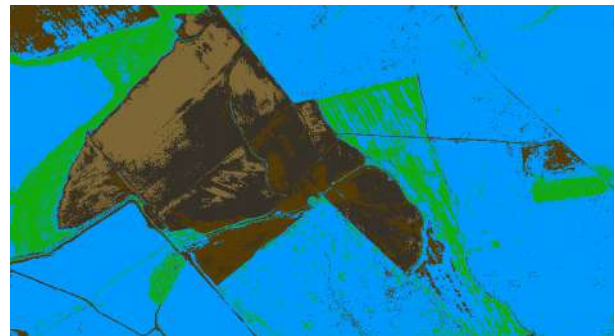


Figura 10. Resultado para  $p = 5$ . Fonte: Autor.

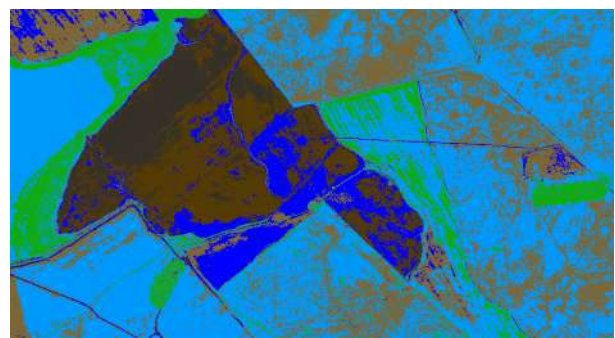


Figura 11. Resultado para  $p = 6$ . Fonte: Autor.

Analisando as imagens, é possível verificar que as imagens geradas com o parâmetro  $p$  variando com valores de 5, 6 e 7 *clusters* apresentam diferença no grau de detalhes provenientes exatamente pelo número de *clusters*.

Através da análise de um especialista, pode-se classificar os *clusters* da seguinte forma para a Figura 10:

- 1) Cor Azul Claro: Área do reservatório de água usada para irrigação da lavoura de arroz. A coloração azul claro

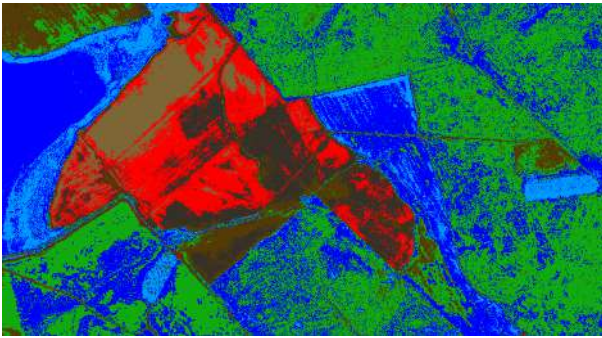


Figura 12. Resultado para  $p = 7$ . Fonte: Autor.

ainda recobre as áreas de arroz irrigado, onde percebe-se claramente o início da irrigação que se dá por inundação, destacando o restante da área não irrigada que apresenta a cultura implantada ainda sem a irrigação. O vermelho ainda nos remete a observar dentro de áreas conhecidas, pontos de maior umidade do solo, ferramenta esta de extrema importância no manejo agrícola;

- 2) Cor Verde: Essa cor representa as áreas com pastagem natural destinadas a pecuária, porém não foi possível distinguir das áreas de arroz irrigado no início do desenvolvimento vegetativo devido a semelhança na altura e arquitetura de plantas bem como pelo predomínio quase absoluto de gramíneas;
- 3) Cor Marron Claro: A coloração marron claro remete a uma área de solo mobilizado no sistema de preparo convencional, manejo este amplamente adotado na lavoura orizícola. Observa-se a mudança nas tonalidades das cores do marron claro para o argila e marron escuro dentro da área com solo mobilizado. Isso permite inferir sobre a textura do solo, onde o marron claro representa um solo de textura mais arenosa, o argila solo de textura média e o aqua textura argilosa e possivelmente maior teor de matéria orgânica. Nas extremidades da imagem aparecem as colorações argila e marron escuro, demonstrando solo exposto com pouca ou nenhuma vegetação. Torna-se interessante como mecanismo para análise a partir de um conhecimento prévio da área e da vegetação ou cultura implantada, por exemplo, falhas de semeadura ou plantio bem como degradação de pastagens oriunda de um pastoreio intensivo, permitindo realizar intervenções em etapas posteriores;
- 4) Cor Argila: Áreas com solo mobilizado e exposto com textura argilosa com possíveis teores de matéria orgânica ou parcialmente exposto demonstrando falhas de semeadura ou áreas de pastagens degradadas;
- 5) Cor Marron Escuro: Solo mobilizado e exposto de textura intermediária.

## VI. CONCLUSÃO

A análise de grupos, em muitos casos, não é um processo realizado em apenas uma execução, dado a estocasticidade dos métodos. Em muitas circunstâncias, é necessário uma série de tentativas, repetições e experimentações, uma vez que não há

um critério universal e efetivo para guiar a seleção de atributos de algoritmos de agrupamentos. Em geral, critérios de validação provêm impressões sobre a qualidade dos *clusters*, mas como escolher este mesmo critério é ainda um problema que requer mais esforços [17]. Este trabalho propôs o uso do algoritmo ckMeans, por meio de uma interface gráfica (site), no processo de segmentação de imagens com uma aplicação no sensoriamento remoto. Os resultados semostam satisfatórios por separar as imagens em grupos (distinguindo através de cores), os quais foram analisados por um especialista da área. A partir do conhecimento da área e da vegetação ou cultura, pode ser utilizado como excelente ferramenta de auxílio na identificação de falhas e problemas no manejo das propriedades rurais, dando-se também aplicabilidade prática do modelo.

A ferramenta web desenvolvida permite processar uma lista de tarefas que o algoritmo original Fuzzy ckMeans não realizava. Essa nova proposta permite que o usuário final execute o algoritmo sem a necessidade de instalar nenhum programa, isso por ser executado via web, e possa deixar em execução as tarefas em uma fila de espera. O tempo de convergência de cada iteração de uma imagem, por exemplo, uma imagem 2D com 1 milhão de pixels, dura em torno de 20 minutos. Assim, a partir desse novo método, as imagens ficam disponíveis para download na web.

Como trabalhos futuros pretende-se comparar a acurácia da metodologia com algoritmos tradicionais como K-Means e Fuzzy C-Means. Também, com os algoritmos de agrupamento atuais como Speed Up Robust Feature (SURF) proposto por [21] e SLIC (Simple Linear Interactive Clustering) proposto por [20] afim de analisar a complexidade assintótica dos algoritmos e consequentemente a performance computacional.

## REFERÊNCIAS

- [1] IRGA. Instituto Rio Grandense do Arroz: *Colheita do arroz se aproxima do final no Estado*. Acesso em maio 2016. Disponível em <http://www.irga.rs.gov.br/conteudo/6315/colheita-do-arroz-se-aproxima-do-final-no-estado>
- [2] B. S. Everitt, S. Landau, M. Leese and D. Stahl. *Cluster analysis*, 3th ed. London, Wiley, 2011.
- [3] J. C. Curti. *Análise de segurança em aplicações que utilizam plataformas UNIX e MS-Windows como Clientes e Servidores*, Campinas, SP, 2004.
- [4] P. Tan, M. Steinbach and V. Kumar. *Introduction To Data Mining*. Addison, Wesley. 1 ed, 2006.
- [5] R. Gonzalez and R. Woods, *Digital Image Processing*, Pearson Prentice Hall. 3 ed, 2008.
- [6] E. M. Novo, *Sensoriamento Remoto. Princípios e Aplicações*, 2 ed. Edgard Blucher, São Paulo, 1993.
- [7] P. R. Menezes and E. E. Sano, *Classificação Pixel a Pixel de Imagens*. In: P. R. Menezes and T. De Almeida (org.), *Introdução ao processamento de imagens de Sensoriamento Remoto*, Brasília, 2012. p. 191-206.
- [8] M. P. de Albuquerque and M. P. de Albuquerque, *Processamento de Imagens: Métodos e Análises*, Centro Brasileiro de Pesquisas Físicas – CBPF/MCT, Coordenação de Atividades Técnicas - CAT, Rio de Janeiro, Brasil.
- [9] G. M. Botelho, *Processamento de Imagens no Auxílio à Detecção de Câncer de Pele utilizando Redes Neurais Artificiais*, Goiás, Catalão, 2007.
- [10] F. Melgani, B. Al Hashemy and S. Taha, *An explicit fuzzy supervised classification method for multispectral remote sensing images*. *Geoscience and Remote Sensing*, IEEE Transactions on, v. 38, n. 1, p. 287-295, 2000.
- [11] L. A. Zadeh, *Fuzzy sets*. *Information and control*, v. 8, n. 3, p. 338-353, 1965.

- [12] R. R. de Vargas, B. R. C. Bedregal and G. Dimuro, *Using ckMeans algorithm in image segmentation process: Preliminary results on mammography analysis*, In Anais do Congresso Nacional de Matemática Aplicada e Computacional (XXXV CNMAC), Natal, Rio Grande do Norte, Brasil, 2014.
- [13] R. R. de Vargas and B. R. C. Bedregal. *A comparative study between fuzzy c-means and ckmeans algorithms*, In Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2010). Toronto, Canada, 2010.
- [14] F. M. Sousa, S. Nascimento, H. Casimiro and D. Boutov. *Identification of upwelling areas on sea surface temperature images using fuzzy clustering*, Remote Sensing of Environment. Volume 112, Issue 6, 16 June 2008.
- [15] Q. Shen, J. Li, F. Zhang, X. Sun, J. Li, W. Li and B. Zhang, *Classification of Several Optically Complex Waters in China Using in Situ Remote Sensing Reflectance*, Remote Sens. 2015.
- [16] J. Oldeland, W. Dorigo, L. Lieckfeld, A. Lucieer and N. Jürgens. *Combining vegetation indices, constrained ordination and fuzzy classification for mapping seminatural vegetation units from hyperspectral imagery*, Remote Sensing of Environment. Volume 114, Issue 6, 15 June 2010, Pages 1155-1166.
- [17] J. Cavalcanti, *Clusterização Baseada em Algoritmos Fuzzy*, Tese de mestrado, Universidade Federal de Pernambuco Recife, Brasil. 2006.
- [18] S. B. Gass, R. Vargas and C. Galafassi, *Comparativo entre os algoritmos K-means e ck Means para mapeamento automatizado de uso do solo*, Anais XVIII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, INPE, Santos-SP, Brasil. 2017.
- [19] E. Cox, *Fuzzy modeling and genetic algorithms for data mining and exploration*, Elsevier/Morgan Kaufmann, 2005. Hardcover. (Morgan Kaufmann series in data management systems).
- [20] Lelkes, D. and Reyzin, Lev. *Interactive Clustering of Linear Classes and Cryptographic Lower Bounds*, Springer-Verlag New York, Inc., New York, NY, USA, 2015.
- [21] Bay, Herbert and Tuytelaars, Tinne and Van Gool, Luc. *SURF: Speeded Up Robust Features*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.