



Pós-Graduação em Ciência da Computação

“Modelo de Mistura de Gaussianas *Fuzzy* Contextual”

Por

Nara Miranda Portela

Tese de Doutorado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE
2015

Nara Miranda Portela

Modelo de Mistura de Gaussianas *Fuzzy* Contextual

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

Universidade Federal de Pernambuco – UFPE

Centro de Informática

Pós-graduação em Ciência da Computação

Orientador: George Darmiton da Cunha Cavalcanti

Coorientador: Tsang Ing Ren

Recife

2015

Catálogo na fonte
Bibliotecário Jefferson Luiz Alves Nazareno CRB4-1758

P843m Portela, Nara Miranda.
 Modelo de mistura de gaussianas fuzzy contextual / Nara Miranda
 Portela. – Recife: O Autor, 2015.
 131 f.: fig., tab.

 Orientador: George Darmiton da Cunha Cavalcanti.
 Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência
 da Computação, 2015.
 Inclui referências.

 1. Inteligência artificial. 2. Sistemas de reconhecimentos de padrões.
 3. Processamento de imagens I. Cavalcanti, George Darmiton da Cunha.
 (Orientador). II. Título.

 005.7 CDD (22. ed.) UFPE-MEI 2015-133

Tese de Doutorado apresentada por **Nara Miranda Portela** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Modelo de Mistura de Gaussianas Fuzzy Contextual**” orientada pelo **Prof. George Darmiton da Cunha Cavalcanti** e aprovada pela Banca Examinadora formada pelos professores:

Profa. Judith Kelner
Centro de Informática/UFPE

Prof. Carlos Alexandre Barros de Mello
Centro de Informática / UFPE

Prof. Getulio Jose Amorim do Amaral
Departamento de Estatística / UFPE

Prof. Herman Martins Gomes
Departamento de Sistemas e Computação / UFCG

Prof. Luiz Eduardo Soares de Oliveira
Departamento de Informática / UFPR

Visto e permitida a impressão.
Recife, 27 de fevereiro de 2015.

Profa. Edna Natividade da Silva Barros

Coordenadora da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.

Agradecimentos

A George, pela estabilidade e calma transmitidas nos momentos difíceis ao longo destes anos do doutorado. Por acreditar na minha capacidade de resolver os problemas que encontramos, mesmo quando os nossos prazos estavam acabando.

A Tsang, por mostrar um novo caminho ao sugerir o uso de sistemas *Fuzzy* em um momento decisivo para progresso deste trabalho.

À minha família, pelo apoio irrestrito e pela confiança que têm depositado em mim. Em especial à Rapha, por me transmitir força e calma ao segurar minha mão.

Aos amigos que fiz nas longas estadias no CIn e que tornaram meu dia-a-dia mais agradável. A Diogo, por me ajudar a trabalhar com a *Berkeley database* e, em especial, a Nadilma, pelas nossas diversas sessões de terapia imprescindíveis para a nossa saúde mental ao longo destes anos.

À CAPES, pelo apoio financeiro concedido para realização deste trabalho.

Success is not final, failure is not fatal: it is the courage to continue that counts.
(Winston Churchill)

Resumo

Métodos de agrupamento identificam uma organização dos padrões existentes em um conjunto de dados, encontram similaridades ou diferenças entre os padrões existentes e assim, derivam conclusões úteis a respeito dos dados. Quando existe dependência entre dados e seus vizinhos, a vizinhança do dado analisado pode fornecer informação valiosa e pouco custosa. Na abordagem contextual do agrupamento, assume-se uma dependência entre os dados e seus vizinhos buscando um resultado mais eficiente. Modelos de Mistura Finita Variante no Espaço (SVFMM) é um método de agrupamento contextual que possibilita a modelagem da correlação espacial dos dados diretamente nos pesos das misturas através da combinação do Modelo de Mistura de Gaussianas (GMM) com Campos Aleatórios Markovianos (MRF). Os parâmetros do SVFMM são estimados via máxima probabilidade *a posteriori* (MAP) usando o algoritmo *Expectation Maximization* (EM). Os primeiros SVFMMs propostos precisavam de uma etapa reparatória adicional no algoritmo EM para garantir que as propriedades de probabilidade dos pesos da mistura fossem conservadas, ou seja, que eles fossem sempre positivos e sua soma, para todos os componentes, fosse sempre igual a 1 (um). Derivações do SVFMM buscaram eliminar essa etapa reparatória, no entanto, foi necessário adotar distribuições de probabilidade que aumentassem o número de parâmetros do modelo ou assumir configurações do MRF cujos parâmetros deveriam ser determinados empiricamente e não calculado em forma fechada a partir dos dados. Nós propomos, neste trabalho, o Modelo de Mistura de Gaussianas *Fuzzy* Contextual (CFGMM), que utiliza o Modelo de Mistura de Gaussianas *Fuzzy* (FGMM) como método de agrupamento e Campos Aleatórios Markovianos Gaussianos como técnica de modelagem da relação espacial entre dados vizinhos. A abordagem *fuzzy* do GMM possibilitou a simplificação matemática da estimação dos parâmetros do modelo em relação à estimativa MAP do SVFMM. No modelo proposto, os parâmetros não são estimados via MAP, e sim pela aplicação do gradiente para minimizar a função objetivo do FGMM. A restrição imposta aos pesos da mistura foi incorporada na derivação das equações dos parâmetros do modelo, eliminando, assim, a necessidade de uma etapa de correção adicional sem aumentar o número de parâmetros do modelo e garantindo que os parâmetros do MRF ainda fossem calculados de forma fechada a partir dos dados. Realizamos a validação do CFGMM como ferramenta de agrupamento contextual de *pixels* na segmentação de imagens baseada em regiões. Estudos de caso usando imagens sintéticas e bases dados de imagens, tais como *BrainWeb* e *Berkeley*, mostraram que a abordagem contextual é capaz de melhorar o desempenho do FGMM na segmentação de imagens. Em comparação com outros métodos de agrupamento, o CFGMM obteve o melhor desempenho na segmentação de imagens com alto nível de ruído. A desvantagem do CFGMM em relação aos métodos de agrupamento pontuais é o maior tempo de processamento devido à incorporação do contexto no cálculo dos parâmetros do modelo. Além disso, o CFGMM apresenta uma

tendência à perda de definição dos detalhes das regiões da imagem ao longo das iterações. No entanto, esse problema pode ser sanado adotando-se um critério de parada precoce.

Palavras-chave: Modelo de Mistura de Gaussianas *Fuzzy*. Agrupamento contextual de dados. Dependência espacial. Segmentação de imagens. Campos Aleatórios Markovianos Gaussianos.

Abstract

Clustering aims to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups according to some criterion. Clustering methods search for patterns in a dataset and discover similarities or differences among the objects in order to draw conclusions about them. When exists dependency between data and its neighbors, the neighborhood of the analyzed data can provide useful and inexpensive information. In the contextual clustering approach, it is assumed correlation between data and its neighbors. Spatially Variant Finite Mixture Models (SVFMM) is a contextual clustering technique that provides the possibility of modeling the pixel spatial information directly in the mixture weights by combining Gaussian Mixture Models (GMM) and Markov Random Fields (MRF). The parameters of the SVFMM are estimated using maximum a posteriori probability (MAP) with the Expectation Maximization algorithm. First SVFMMs needed a reparatory step in the EM algorithm to ensure that mixture weights probability properties were preserved, i.e., they should be always positive and their sum, among all groups, should be equal to one. SVFMM based methods were proposed to eliminate this reparatory step. To accomplish this task, however, they adopted probabilities distributions that improved the number of models parameters or they assumed MRF configurations whose parameters could not be calculated in a closed form from the data but in an empirical way. In this work, we propose the Contextual Fuzzy Gaussian Mixture Model (CFGMM) that use Fuzzy Gaussian Mixture Model (FGMM) as clustering technique and Gaussian Markov Random Field to model neighboring data relationship. The Fuzzy approach of the GMM facilitates the mathematic simplification of the model parameters estimation compared to MAP estimation of the SVFMM. In the proposed model, parameters are not estimated using MAP but using gradient to minimize the objective function of the FGMM. The mixture weights constraints were taken into account in the equations derivation of the model parameters. Thus, we eliminate one extra step without improve the number of parameters and ensuring that MRF parameters were still calculated in a closed form from the data. We perform the CFGMM validation as contextual pixel clustering tool for region based image segmentation. Case studies using synthetic images and real image databases, such as BrainWeb and Berkeley, show that contextual approach is capable of improving the image segmentation performance when compared with the FGMM. Compared with other clustering methods, the proposed model obtains the best performance for images having high level of noise. The disadvantage of CFGMM is that it requires more time to segment an image than techniques that do not take into account contextual information. Beside, CFGMM induce definition loss of the image region details along the algorithm iterations. However, this problem can be rectified by adopting an early stop criterion.

Keywords: Fuzzy Gaussian Mixture Model. Contextual clustering. Spatial dependence. Image segmentation. Gaussian Markov Random Fields.

Lista de ilustrações

Figura 1 – Vizinhança-4, vizinhança-8 e vizinhança-D.	46
Figura 2 – Dois exemplos de vizinhança em um reticulado bidimensional com seus respectivos <i>cliques</i>	51
Figura 3 – Imagem artificial com ruído, imagem natural e imagem de ressonância magnética (acima) e suas respectivas segmentações baseadas em região (abaixo).	64
Figura 4 – Probabilidade de <i>pixels</i> vizinhos pertencerem ao mesmo grupo <i>versus</i> a probabilidade de <i>pixels</i> vizinhos pertencerem a grupos diferentes para as imagens segmentadas: imagem artificial com ruído, imagem natural e imagem de ressonância magnética.	65
Figura 5 – Nível de cinza do <i>pixel</i> <i>versus</i> a média no nível de cinza dos <i>pixels</i> vizinhos para as imagens: imagem artificial com ruído, imagem natural e imagem de ressonância magnética.	66
Figura 6 – Processo de agrupamento dos dados pelo Modelo de Mistura de Gaussianas <i>Fuzzy</i> Contextual.	72
Figura 7 – Imagens sintéticas de 2, 3 e 5 classes	75
Figura 8 – Imagens sintéticas de 3 (acima) e 5 classes (abaixo) com relação sinal ruído 1, 4, 8, 12 e 16 dB (da esquerda para a direita)	75
Figura 9 – Exemplo do modelo no formato binário (a) e sua decomposição nas classes: (b) <i>background</i> , (c) fluido cérebro-espinhal (CSF), (d) matéria cinzenta (GM), (e) matéria branca (WM), (f) gordura, (g) músculo, (h) pele, (i) osso, (j) matéria glial e (k) material conectivo.	76
Figura 10 – Exemplo de imagens ponderadas em T1 (a), T2 (b) e Pd (c)	77
Figura 11 – Imagens originais das fatias 50, 93 e 120 com 3% de ruído e 0% de não-uniformidades (coluna da esquerda), mesmas imagens após o pré-processamento (coluna do centro) e os seus respectivos <i>ground truths</i> (coluna da direita).	78
Figura 12 – Exemplo da fatia 100 de imagens com largura de fatia de 1mm, com 0% de não-uniformidades e níveis de ruído de 0%, 1%, 3%, 5%, 7% e 9% (da esquerda para direita).	78
Figura 13 – Exemplo de imagens da <i>Berkeley image segmentation database</i>	79
Figura 14 – Exemplo de imagens em tons de cinza da <i>Berkeley image segmentation database</i> . Cada imagem foi manualmente segmentada por 3 pessoas diferentes.	79

Figura 15 – Segmentação de imagens sintéticas. A primeira coluna mostra, de cima para baixo, as imagens sintéticas de 2, 3 e 5 classes, a segunda coluna exibe os respectivos <i>ground truths</i> e a terceira coluna apresenta o resultado da segmentação pelo CFGMM.	84
Figura 16 – Gráfico do função objetivo J	84
Figura 17 – Gráfico do tempo de processamento por iteração médio, em segundos, na segmentação de imagens de diferentes dimensões pelo KM, FCM, GMM, FGMM e CFGMM. As curvas ajustadas à cada método foram obtidas pela regressão linear dos respectivos resultados.	85
Figura 18 – Gráfico da função objetivo J e do Índice Rand ao longo das iterações na segmentação das imagens sintéticas de 2, 3 e 5 classes pelo CFGMM. J está representado nos gráficos com a linha sólida enquanto que o Índice Rand está representado com a linha tracejada.	87
Figura 19 – Resultado da segmentação da imagem sintética de 5 classes pelo CFGMM ao longo das iterações.	88
Figura 20 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 2 classes.	89
Figura 21 – Segmentação da imagem sintética de 2 classes pelo KM, FCM, GMM, FGMM e CFGMM.	89
Figura 22 – Segmentação da imagem sintética de 3 classes e SNR 8 pelo KM, FCM, GMM, FGMM e CFGMM.	90
Figura 23 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 3 classes.	90
Figura 24 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 5 classes.	91
Figura 25 – Segmentação da imagem sintética de 5 classes e SNR 8 pelo KM FCM, GMM, FGMM e CFGMM.	91
Figura 26 – Segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	92
Figura 27 – Média do Índice Rand na segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	93
Figura 28 – Gráfico de caixa do Índice Rand na segmentação das imagens sintéticas de 3 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	93

Figura 29 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	94
Figura 30 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das imagens sintéticas de 3 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	95
Figura 31 – Segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	96
Figura 32 – Média do Índice Rand na segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	97
Figura 33 – Gráfico de caixa do Índice Rand na segmentação das imagens sintéticas de 5 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	97
Figura 34 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	98
Figura 35 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das imagens sintéticas de 5 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	99
Figura 36 – Fatia 100 da imagem com nível de ruído 5%.	100
Figura 37 – Gráfico da função objetivo J e do Índice Rand ao longo das iterações na segmentação da fatia 100 da imagem de cérebro com nível de ruído 5% pelo CFGMM. J está representado nos gráficos com a linha sólida enquanto que o Índice Rand estão representados com a linha tracejada.	100
Figura 38 – Resultado da segmentação da imagem pelo CFGMM ao longo das iterações.	101
Figura 39 – Gráfico do Índice Rand em função da variação do valor de m na segmentação das fatias da imagem de cérebro com nível de ruído 5%	102
Figura 40 – Segmentação da fatia 100 da imagem de cérebro com nível de ruído 5% pelo KM, FCM, GMM, FGMM e CFGMM.	102
Figura 41 – Média do Índice Rand na segmentação das fatias das imagens de cérebro com níveis de ruído 0, 1, 3, 5, 7 e 9 pelo KM, FCM, GMM, FGMM e CFGMM-B.	104

Figura 42 – Gráfico de caixa do Índice Rand na segmentação das fatias das imagens de cérebro e todos os níveis de ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.	104
Figura 43 – Segmentação da fatia 100 da imagem de cérebro com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9%, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.	105
Figura 44 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das fatias das imagens de cérebro com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9% pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.	106
Figura 45 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das fatias das imagens de cérebro e todos os níveis de ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.	107
Figura 46 – Média do Índice Rand na segmentação das fatias das imagens de cérebro com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM e CFGMM-B.	108
Figura 47 – Gráfico de caixa do Índice Rand na segmentação das fatias das imagens de cérebro e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.	108
Figura 48 – Segmentação da fatia 100 da imagem de cérebro com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, e CFGMM-B.	109
Figura 49 – Amostra de imagens da <i>Berkeley database</i>	111
Figura 50 – Figura #183055 e seus <i>ground truths</i>	112
Figura 51 – Gráfico da função objetivo J e do Índice PR ao longo das iterações na segmentação da imagem #183055 pelo CFGMM. J está representado no gráfico com a linha sólida enquanto que o Índice PR está representado com a linha tracejada.	112
Figura 52 – Resultado da segmentação da imagem pelo CFGMM ao longo das iterações.	113
Figura 53 – Gráfico do valor médio do Índice PR em função da variação do valor de m na segmentação de imagens da <i>Berkeley database</i>	114
Figura 54 – Segmentação da imagem #183055 pelo KM, FCM, GMM, FGMM e CFGMM.	114
Figura 55 – Gráfico do valor médio do Índice PR na segmentação de imagens da <i>Berkeley database</i>	115

Figura 56 – Média do tempo de processamento e do número de iterações na segmentação de imagens da <i>Berkeley database</i>	115
Figura 57 – Segmentação das imagens #80099, #113040, #135069, #183055 e #238011 pelo FCM, GMM, FGMM e CFGMM-B.	116
Figura 58 – Média do Índice PR na segmentação das imagens com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB pelo FCM, GMM, FGMM e CFGMM-B.	117
Figura 59 – Gráfico de caixa do Índice PR na segmentação das imagens e todas as relações sinal ruído pelo FCM, GMM, FGMM e CFGMM-B.	118
Figura 60 – Segmentação da imagem #183055 com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.	118

Lista de tabelas

Tabela 1 – Pontuação do KM, FCM, GMM, FGMM e CFGMM-B geradas pelo <i>benchmark</i> ao analisar a segmentação baseada em região das 200 imagens de teste da <i>Berkeley database</i>	119
---	-----

Lista de abreviaturas e siglas

BSDS500	<i>Berkeley image segmentation database and benchmark 500</i>
CFGMM	Modelo de Mistura Gaussiana <i>Fuzzy</i> Contextual
EM	<i>Expectation maximization</i>
FCM	<i>Fuzzy C-Means</i>
FGMM	Modelo de Mistura Gaussiana <i>Fuzzy</i>
GMRF	Campos Aleatórios Markovianos Gaussianos
GRF	Campos Aleatórios de Gibbs
KM	<i>K-Means</i>
MAP	Máxima probabilidade <i>a posteriori</i>
ML	Máxima Verossimilhança
MR	Ressonância Magnética
MRF	Campos Aleatórios Markovianos
Pd	Densidade de prótons
PR	Índice Probabilístico Rand
SVFMM	Modelo de Mistura Finita Variante no Espaço

Lista de símbolos

A	Conjunto de parâmetros da distribuição de Dirichlet
α	Parâmetro da distribuição de Dirichlet
β	Parâmetro de suavidade do SVFMM
c	<i>Clique</i>
D	Distribuição de probabilidade
d_{ij}	Função de dissimilaridade do elemento i em relação ao grupo j
J	Função objetivo
k	Número de grupos
l	Rótulo de uma classe
λ	Multiplicador de Lagrange
M	Conjunto de <i>ground truths</i> de uma imagem
m	Expoente <i>fuzzy</i> de ponderação
μ	Vetor de médias
$\mathcal{N}_s$	Sistema de vizinhança do sítio s
n	Número de elementos do conjunto X
p	Dimensão dos dados
Ψ	Espaço paramétrico p -dimensional
ψ	Conjunto paramétrico do Modelo de Mistura de Gaussianas
$\mathbf{Q}$	Matriz de precisão
R	Conjunto de rótulos
$\mathbf{r}$	Vetor de rótulo k -dimensional
S	Reticulado bidimensional
s	Sítio de um reticulado bidimensional

Σ	Matriz de covariância
T	Temperatura
t	Operador de Transposição de uma matriz
θ	Parâmetro da função de densidade de probabilidade condicionada à classe
U	Partição <i>fuzzy</i> de um conjunto
u_{ij}	Grau de pertinência do elemento i ao grupo j
V_i	Vizinhança do elemento i
$\mathcal{V}_c$	Potencial <i>clique</i>
W	Conjunto de pesos da mistura
w	Peso da mistura
X	Conjunto de dados
x	Vetor p -dimensional representando um dado observado
Ξ	Conjunto de vetores de probabilidade parâmetros da distribuição multinomial
ξ	Vetor de probabilidade k -dimensional parâmetro da distribuição multinomial
Y	Conjunto de dados completos
Z	Constante de normalização da função de Gibbs
z	Probabilidade <i>a posteriori</i>

Sumário

1	INTRODUÇÃO	21
1.1	Motivação	23
1.2	Objetivo	25
1.3	Contribuições	25
1.4	Organização do texto	27
2	FUNDAMENTAÇÃO TEÓRICA	28
2.1	Agrupamento	28
2.1.1	<i>K-Means</i> (KM)	30
2.1.2	Modelo de Mistura de Gaussianas (GMM)	31
2.1.2.1	Maximizando a verossimilhança via EM	34
2.1.3	<i>Fuzzy C-Means</i> (FCM)	39
2.1.4	Modelo de Mistura de Gaussianas <i>Fuzzy</i> (FGMM)	41
2.2	Segmentação de imagens	45
2.2.1	Segmentação via agrupamento de <i>pixels</i>	47
2.3	Campos Aleatórios Markovianos	49
2.3.1	Sistema de vizinhança	50
2.3.2	Campos Aleatórios Markovianos	51
2.3.2.1	Campos Aleatórios de Gibbs	52
2.3.2.2	Campos Aleatórios Markovianos Gaussianos	53
3	MODELO DE MISTURA CONTEXTUAL	54
3.1	Modelos de Mistura Finita Variante no Espaço	56
3.1.1	Campo Aleatório de Gibbs	57
3.1.2	Campo Aleatório Markoviano Gaussiano	59
3.1.3	Dirichlet conjugada da distribuição multinomial	61
4	MODELO DE MISTURA DE GAUSSIANAS FUZZY CONTEXTUAL (CFGMM)	64
4.1	Definição do modelo proposto	67
4.2	Considerações finais	73
5	EXPERIMENTOS	74
5.1	Metodologia dos experimentos	74
5.1.1	Bancos de imagens	74
5.1.1.1	Imagens sintéticas	74

5.1.1.2	<i>BrainWeb simulated brain database</i>	76
5.1.1.3	<i>Berkeley image segmentation database and benchmark</i>	78
5.1.2	Medidas de desempenho	80
5.1.2.1	Índice Rand e Índice Probabilístico Rand	80
5.1.2.2	<i>F-measure</i>	81
5.1.2.3	Tempo de processamento	82
5.1.2.4	Intervalo de confiança	82
5.1.2.5	Ambiente de teste	82
5.2	Experimentos e resultados obtidos	83
5.2.1	Análise do modelo com o banco de imagens sintéticas	83
5.2.1.1	Limiar de convergência	86
5.2.1.2	Expoente <i>fuzzy</i>	88
5.2.1.3	Comportamento do modelo em relação ao ruído	92
5.2.2	Análise do modelo com o banco de imagem <i>BrainWeb</i>	99
5.2.2.1	Limiar de convergência	99
5.2.2.2	Expoente <i>fuzzy</i>	101
5.2.2.3	Comportamento do modelo em relação ao ruído	102
5.2.3	Análise do modelo com o banco de imagem de <i>Berkeley</i>	109
5.2.3.1	Limiar de convergência	110
5.2.3.2	Expoente <i>fuzzy</i>	113
5.2.3.3	Comportamento do modelo em relação ao ruído	114
5.2.3.4	Resultado do <i>benchmark</i> BSDS500	118
5.3	Considerações finais	119
5.3.1	Dificuldades encontradas	120
6	CONSIDERAÇÕES FINAIS	123
6.1	Trabalhos futuros	124
	Referências	127

1 Introdução

A organização de objetos em grupos de acordo com a similaridade das propriedades observadas é uma forma fundamental para entendê-los e aprender sobre eles. O agrupamento (XU; II, 2009) é a tarefa de encontrar subconjuntos ou grupos, em um conjunto de elementos, de forma que elementos que pertençam ao mesmo grupo sejam similares entre si, segundo algum critério definido, e elementos de grupos distintos sejam diferentes. Cada grupo pode ser representado por um ou mais protótipos que são os elementos que melhor representam as características do grupo.

A partir de dados não rotulados e sem conhecimento prévio das categorias presentes no conjunto de dados, os métodos de agrupamento compreender automaticamente a organização dos padrões existentes nos dados e derivar conclusões úteis a respeito deles. Esse tipo de aprendizado não supervisionado é adequado a situações nas quais coletar e rotular um grande conjunto de exemplos pode se caracterizar como uma tarefa árdua ou mesmo inviável, mas é necessário entender como aqueles exemplos se comportam.

Modelos de Mistura de Gaussianas (GMM)(MCLACHLAN; PEEL, 2000) pode ser usado como método de agrupamento baseados em distribuições estatísticas específicas. Qualquer função de distribuições de probabilidade pode ser modelada por um Modelo de Mistura usando um número suficiente de componentes (distribuições) básicos, por exemplo, a distribuição Gaussiana no GMM. Ao modelar a função de densidade de probabilidade de um conjunto de dados, o GMM automaticamente realiza um agrupamento do conjunto ao discriminar qual componente da mistura gerou cada elemento. O GMM se baseia na suposição de que cada elemento do conjunto se origina a partir de um componente da mistura com uma determinada probabilidade. Assim, ao inferir os parâmetros da mistura, essa probabilidade pode ser usada para associar cada elemento ao componente com maior probabilidade de o ter gerado.

Os parâmetros que caracterizam cada componente de uma mistura podem ser interpretados como o protótipo do grupo formado pelos dados com maior probabilidade de pertencer àquele componente. No caso do Modelo de Mistura de Gaussianas, o protótipo de cada grupo consiste no conjunto de parâmetros que contém o vetor de médias, as matrizes de covariância e os pesos da mistura. As estimativas dos parâmetros da distribuição de cada componente da mistura são calculadas ao maximizar a verossimilhança utilizando o algoritmo *expectation maximization* (EM) (DEMPSTER; LAIRD; RUBIN, 1977). O algoritmo EM é um método iterativo de encontrar a estimativa de Máxima Verossimilhança dos parâmetros de uma distribuição dado um conjunto de dados. Além disso, a verossimilhança de um Modelo de Mistura é uma medida do desempenho do agrupamento

uma vez que mede quão bem o modelo paramétrico está ajustado aos dados que se quer modelar (MCLACHLAN; PEEL, 2000).

Métodos de agrupamento como o Modelo de Mistura particionam os dados usando a lógica binária ou *hard*, na qual cada objeto só pode pertencer a um único grupo. Da necessidade de tratar dados cujos limites entre grupos são ambíguos, ou não estão bem definidos, pode-se usar a lógica *fuzzy* que admite valores intermediários entre o 0 (falso) e o 1 (verdadeiro) da lógica binária. Nos métodos de agrupamento *fuzzy*, os objetos podem pertencer a vários grupos ao mesmo tempo com diferentes graus de associação. Geralmente, o valor do grau de associação é um percentual que representa o quanto esse dado pertence a um grupo.

O método de agrupamento *Fuzzy C-Means* (FCM), proposto por Bezdek (BEZDEK, 1981), é uma extensão do método *C-Means* proposto por Dunn (DUNN, 1974). A partir da função de soma do quadrado dos erros intragrupos do *C-Means*, uma nova família de funções objetivo *fuzzy* foi criada. Através de um processo iterativo, o FCM reveza entre as etapas de atualizar os protótipos dos grupos (vetores de médias *fuzzy*) e de realizar uma nova partição dos grupos até que uma função objetivo seja otimizada. Ao perceberem que diferentes grupos em um conjunto de observações poderiam ter diferentes formas, Gustafson e Kessel (GUSTAFSON; KESSEL, 1978) propuseram uma modificação no FCM que chamaram de algoritmos de agrupamento com covariância *fuzzy*. Nesta nova família de algoritmos, cada protótipo consiste em um conjunto de parâmetros que contém o vetor de média *fuzzy* e a matriz de covariância *fuzzy*.

Buscando integrar as vantagens do GMM com as vantagens do FCM, Tran e Warner (TRAN; WAGNER, 1998) propuseram uma modificação do FCM baseada no GMM chamada de Modelo de Mistura de Gaussianas *Fuzzy* (FGMM). No FGMM, o grau de associação de um elemento a um grupo é representado pela função de pertinência *fuzzy* do FCM em vez da probabilidade *a posteriori* do GMM. Além do vetor de média *fuzzy* e da matriz de covariância *fuzzy*, os pesos da mistura também passam a ser tratados segundo a lógica *fuzzy*. O FGMM reconhece, então, a existência de grupos de diferentes formas no conjunto de dados, assim como, diferentes densidades de dados (probabilidades *a priori* de grupos). Além disso, ao mesclar as duas técnicas, é possível obter a capacidade de generalização do GMM com a convergência mais rápida que caracteriza os algoritmos da família do FCM (JU; LIU, 2012). Os protótipos no FGMM são definidos como os conjuntos que contêm o vetor de médias *fuzzy*, a matriz de covariância *fuzzy* e o peso da mistura *fuzzy*.

O GMM, o FCM e o FGMM utilizam uma **abordagem pontual** ao particionar os dados. Nela, os elementos de um conjunto de dados são modelados de forma independente, ou seja, desconsiderando as informações dos elementos vizinhos. Na **abordagem contextual**, assume-se uma dependência entre os dados e seus vizinhos. A abordagem

contextual é especialmente interessante em situações nas quais existe uma dependência entre os dados e seus vizinhos, por exemplo, na segmentação de imagens (PORTELA; CAVALCANTI; TSANG, 2013), no estudo de séries temporais (YANG; GUO; JENSEN, 2013), na análise de vídeos (YANG et al., 2014; ZHANG et al., 2014) ou no processamento de voz (JOTHILAKSHMI, 2014).

Tanto a abordagem pontual como a abordagem contextual do agrupamento podem ser usadas na segmentação de imagens. Uma maneira natural de realizar a segmentação é subdividindo a imagem em componentes que se assemelham, lógica usada pelos algoritmos de agrupamento de dados. O critério específico de similaridade a ser adotado depende da aplicação. Grupos de *pixels* podem se assemelhar por terem a mesma cor e/ou mesma textura e/ou por serem vizinhos, etc (FORSYTH; PONCE, 2002; THILAGAMANI; SHANTI, 2011). Na segmentação baseada em região, a imagem é dividida em regiões disjuntas e internamente homogêneas de acordo com um conjunto de critérios predefinidos (NAZ; MAJEED; IRSHAD, 2010; SARASWATHI; ALLIRANI, 2013).

Na segmentação baseada em região utilizando técnicas de agrupamento, cada grupo encontrado nos *pixels*, deve ser tratado como uma região da imagem. O fim do processo de agrupamento se dá quando os *pixels* pertencentes a um grupo/região específico possuem, em geral, um alto grau de similaridade ou *pixels* de diferentes grupos/regiões possuam alto grau de dissimilaridade (VANTARAM; SABER, 2012). A abordagem contextual do agrupamento permite incorporar na segmentação de imagens o conhecimento *a priori* de que *pixels* vizinhos provavelmente pertencem ao mesmo grupo. Isto é, além dos valores de intensidade, a informação espacial do *pixel* e sua relação com a sua vizinhança também é usada para determinar o grupo ao qual cada *pixel* é associado.

1.1 Motivação

Na análise de imagens, o contexto ocupa um importante papel na interpretação de uma cena. No menor nível de abstração, *pixels* isolados fornecem menos informações do que os mesmos *pixels* juntos dando informação de contexto. Sendo as relações espaciais um atributo essencial em uma imagem, restrições espaciais podem ser necessárias para alcançar uma segmentação de imagens mais eficiente. Estudos mostram que a ideia de combinar a informação dos *pixels* vizinhos e algoritmos de agrupamento para realizar a segmentação contextual de imagens são eficientes, especialmente, na segmentação de imagens com ruído (GONG et al., 2013; ZHANG; WU; NGUYEN, 2013).

Modelos de Mistura Finita Variante no Espaço (SVFMM) é uma técnica eficiente de segmentação contextual de imagens, bastante explorada na literatura (BLEKAS et al., 2005; NIKOU; GALATSANOS; LIKAS, 2007; NIKOU; LIKAS; GALATSANOS, 2010; NGUYEN; WU, 2013) e bem fundamentada na teoria de decisão Bayesiana. SVFMM

oferece uma abordagem que possibilita a modelagem da correlação espacial diretamente nos pesos das misturas através do uso de Campos Aleatórios Markovianos (MRF) (NIKOU; GALATSANOS; LIKAS, 2007; SANJAY-GOPAL; HEBERT, 1998). Os parâmetros do modelo são estimados via máxima probabilidade *a posteriori* (MAP) usando o algoritmo EM.

Os primeiros SVFMMs propostos precisavam de uma etapa reparatória adicional no algoritmo EM para garantir que as propriedades de probabilidade dos pesos da mistura fossem conservadas, ou seja, que eles fossem sempre positivos e sua soma, para todos os componentes, fossem sempre igual a 1 (um). Dentre os artifícios propostos para garantir que os pesos da mistura estejam dentro dos requisitos, pode-se citar o algoritmo de Projeção de Gradiente proposto por Sanjay-Gopal *et al.* (SANJAY-GOPAL; HEBERT, 1998) ou a Programação Quadrática proposta por Blekas *et al.* (BLEKAS *et al.*, 2005).

Nguyen e Wu propuseram em (NGUYEN; WU, 2013) um SVFMM que não precisa da etapa reparatória, no entanto o parâmetro do MRF adotado deve ser determinado empiricamente e não calculado em forma fechada¹ a partir dos dados. Nikou *et al.* propuseram em (NIKOU; LIKAS; GALATSANOS, 2010) uma nova abordagem do SVFMM, na qual os pesos da mistura seguem uma distribuição de Dirichlet conjugada da distribuição multinomial (DCM-SVFMM, do inglês, *Dirichlet Compound Multinomial-based Spatially Variant Finite Mixture Model*). A correlação espacial dos dados é modelada, então, nos parâmetros da distribuição de Dirichlet com o uso de GMRFs. Os parâmetros da distribuição multinomial são totalmente integrados ao *framework* Bayesiano e as estimativas MAP dos parâmetros podem ser computadas em forma fechada através do algoritmo EM. Embora a etapa reparatória tenha sido eliminada no DCM-SVFMM, o fato de adotar a distribuição de probabilidade DCM que não é usada no SVFMM padrão aumentou o número de parâmetros do modelo.

A dificuldade na compreensão dos cálculos matemáticos envolvidos na etapa reparatória do SVFMM padrão (SANJAY-GOPAL; HEBERT, 1998; BLEKAS *et al.*, 2005) serviu de motivação para a busca de uma maneira de eliminar essa etapa extra do algoritmo EM, necessária para a correção dos pesos do modelo. Além disso, foi necessário garantir que os parâmetros do modelo ainda pudessem ser calculados de forma fechada a partir dos dados e que a solução encontrada não aumentasse o número de parâmetros do modelo de mistura deixando-o mais complexo.

¹ Em matemática, uma expressão é dita ser uma expressão de forma fechada se, e somente se, pode ser expressa analiticamente em termos de um número delimitado de certas funções bem conhecidas. Tipicamente, estas bem conhecidas funções são definidas como funções elementares - constantes, uma variável x , operações elementares de aritmética ($+$ $-$ $\times$ $\div$), raízes n -ésimas, exponenciais e logaritmos (os quais também incluem funções trigonométricas e funções trigonométricas inversas).

1.2 Objetivo

Esta tese tem por objetivo geral propor um método que realize o agrupamento mais eficiente de dados com ruído ao levar em consideração o contexto dos dados ao particioná-los. A melhoria do resultado do agrupamento só é possível, no entanto, em situações nas quais existe uma dependência entre os dados e seus vizinhos.

O presente trabalho possui 5 objetivos específicos:

- Eliminar a etapa extra do algoritmo EM do SVFMM ao substituir o Modelo de Mistura de Gaussianas pelo Modelo de Mistura de Gaussianas *Fuzzy*;
- Utilizar uma função de densidade *a priori* dos pesos da mistura que modele a correlação dos dados com sua vizinhança e seja incorporada no processo de estimação dos parâmetros.
- Derivar as equações de atualização dos parâmetros do modelo com o uso do cálculo do gradiente da função objetivo.
- Implementar e validar o modelo como ferramenta de agrupamento de *pixels* usando como estudo de caso a segmentação de imagens baseada em regiões;
- Comparar o desempenho do CFGMM com o de outras técnicas de agrupamento na segmentação de imagens de diferentes bancos de imagens.

Duas hipóteses são testadas neste trabalho. A primeira é que a abordagem contextual do CFGMM seja capaz de melhorar significativamente o desempenho da segmentação baseada em região de imagens em relação à abordagem pontual do FGMM. A segunda é que o CFGMM tenha melhor resultado do que os algoritmos de agrupamento KM, FCM, GMM e FGMM na segmentação de imagens baseada em regiões.

1.3 Contribuições

Quando existe dependência entre dados e seus vizinhos, a vizinhança do dado analisado pode fornecer informação valiosa e pouco custosa. Na análise de imagens, por exemplo, a classe à qual cada pixel é associado depende (a) dos seus próprios atributos, (b) dos atributos dos outros pixels e (c) da relação existente entre as classes, logo, classificar cada pixel individualmente não parece ser a estratégia mais eficiente. Tal situação também aparece em outras aplicações como comunicação, séries temporais e reconhecimento de voz. Neste trabalho propomos um novo método de agrupamento contextual de dados, que combina a informação de dados vizinhos para alcançar um agrupamento mais eficiente, em situações nas quais existe uma dependência entre os dados e seus vizinhos.

O Modelo de Mistura *Fuzzy* Contextual (CFGMM) aqui proposto utiliza o Modelo de Mistura de Gaussianas *Fuzzy* como técnica de agrupamento e Campos Aleatórios Markovianos Gaussianos como técnica de modelagem da relação espacial entre dados vizinhos. Além disso, apresentamos duas versões do CFGMM: na primeira o processo de rotulagem de dados se baseia na regra de decisão Bayesiana; na segunda a rotulagem dos dados se baseia na análise da partição *fuzzy*.

O processo de otimização iterativa baseado no gradiente para estimar os parâmetros do FGMM é matematicamente mais simples do que a Máxima Verossimilhança via EM adotada pelo GMM. Utilizando o FGMM em vez do GMM na abordagem contextual do SVFMM, encontramos um método mais simples que a estimativa de máxima probabilidade *a posteriori* para estimar os parâmetros do modelo. Ao simplificar o cálculo da estimativa dos parâmetros, foi possível incorporar as restrições dos pesos da mistura na derivação das equações de atualização dos parâmetros, eliminando, assim, a necessidade da etapa de correção do algoritmo SVFMM padrão para garantir que os pesos da mistura obedecem as propriedades da probabilidade.

O uso do Campos Aleatórios Markovianos Gaussianos, além de ser uma maneira conveniente e consistente de modelar entidades com dependência contextual, permite que o parâmetro que controla o grau de suavidade imposto ao modelo em relação aos dados seja derivado diretamente da função objetivo. Dessa maneira, os parâmetros do modelo podem ser calculados em forma fechada a partir dos dados analisados.

Os estudos de caso com diferentes grupos de imagens mostraram que, ao incorporar a abordagem contextual, o FGMM melhora o seu desempenho na segmentação de imagens, principalmente ruidosas, assim como acontece com os algoritmos FCM e GMM ao adotarem a abordagem contextual (GONG et al., 2013; LI; ZHANG; QU, 2008; NGUYEN; WU, 2011b; NIKOU; GALATSANOS; LIKAS, 2007; ZHANG; WU; NGUYEN, 2013). A desvantagem do CFGMM em relação aos métodos de agrupamento pontuais é o maior tempo de processamento devido à incorporação do contexto no cálculo dos parâmetros do modelo. O gargalo do CFGMM na segmentação de imagens é processo de acessar os atributos da vizinhança de cada *pixel* no cálculo dos parâmetros a cada iteração. Além disso, os experimentos mostraram que ao longo das iterações do CFGMM as regiões da imagem vão ficando mais homogêneas mas as bordas das regiões vão ficando borradas. As quinas vão se tornando curvas e as linhas retas vão perdendo a definição. Havendo, assim, um *trade-off* entre eliminação de ruído no interior das regiões e a perda de definição dos detalhes da imagem.

1.4 Organização do texto

Este trabalho está organizado em seis capítulos. Neste capítulo, é realizada a contextualização do tema proposto e expostos o objetivo e a motivação deste trabalho.

No Capítulo 2, são apresentados os principais conceitos necessários para a compreensão e avaliação do método proposto nesta Tese. São descritos diferentes algoritmos de agrupamento de dados e como esses algoritmos podem ser usados na segmentação de imagens baseada em região. Além disso, são apresentados os Campos Aleatórios Markovianos usados para modelar a correlação espacial dos dados.

O Capítulo 3, apresenta o estado-da-arte sobre Modelo de Mistura Contextual e descreve diferentes tipos de Modelo de Mistura Variante no Espaço.

O Modelo de Mistura de Gaussianas *Fuzzy* Contextual é proposto no Capítulo 4. Inicialmente é feita uma avaliação da importância da informação dos pixels vizinhos na análise de imagens. Em seguida o modelo proposto é formalizado e é demonstrado o passo-a-passo de como os seus parâmetros são estimados.

O Capítulo 5 apresenta os experimentos realizados para validar o método proposto como técnica de segmentação contextual de imagens baseada em regiões. São utilizados um banco de imagens sintéticas, um banco de imagens naturais e um banco de imagens simuladas de ressonância magnética para comparar o desempenho do modelo proposto com o de outras técnicas de agrupamento na segmentação de imagens.

O Capítulo 6 expõe as considerações finais sobre o trabalho desenvolvido.

2 Fundamentação Teórica

Neste capítulo, é realizada a fundamentação teórica necessária para o desenvolvimento do método proposto nesta Tese. Na Seção 2.1 é definido o conceito de agrupamentos de dados e descritas as técnicas de agrupamento de dados baseadas em otimização iterativa: *K-Means*, *Fuzzy C-Means*, Modelo de Mistura de Gaussianas e Modelo de Mistura de Gaussianas *Fuzzy*.

A Seção 2.2 expõe como a segmentação de imagens baseada em região pode ser tratada como um agrupamento dos *pixels* de uma imagem e a relevância da informação espacial na segmentação de imagens.

Os conceitos básicos sobre Campos Aleatórios Markovianos são apresentados na Seção 2.3 pois se tratam de uma importante ferramenta na modelagem de dependência contextual de dados. Em seguida, são definidos Campos Aleatórios de Gibbs e Campos Aleatórios Markovianos Gaussianos.

2.1 Agrupamento

O agrupamento (do inglês, *clustering*) de dados é um processo de decisão que visa encontrar subconjuntos ou grupos de dados (do inglês, *clusters*) que apresentem características semelhantes e separar aqueles com características distintas, a partir de um conjunto de dados não rotulados. Algoritmos de agrupamento aprendem a identificar a organização dos padrões existentes nos dados, encontram similaridades ou diferenças entre os padrões existentes e, assim, derivam conclusões úteis a respeito deles (RUSSELL; NORVIG, 2009). A ideia de agrupar objetos similares para produzir uma categorização é conhecida em diferentes ramos da ciência, e pode ser encontrada sob diferentes nomes em diferentes contextos, por exemplo, taxonomia em biologia ou tipologia em ciências sociais (THEODORIDIS; KOUTROUMBAS, 2009).

O uso de técnicas não-supervisionadas, como o agrupamento de dados, é indicado para os domínios de aplicação nos quais coletar e rotular um grande conjunto de exemplos pode se caracterizar como uma tarefa árdua ou mesmo inviável, mas é necessário entender como aqueles exemplos se comportam (DUDA; HART; STORK, 2001).

Embora não haja uma definição única para agrupamento de dados, uma possível definição considera o particionamento de n elementos de um conjunto de dados não rotulados em k grupos tal que cada elemento pertença apenas a um grupo, tendo como objetivo atribuir os elementos que apresentam características semelhantes a um mesmo grupo. Essa abordagem é conhecida como agrupamento particional (XU; II, 2009).

Seja $X = \{x_1, \dots, x_n\}$ um conjunto de dados p -dimensionais composto de n elementos não rotulados. Um k -agrupamento particional é uma partição de X em k conjuntos $C_1, \dots, C_k$ sem estrutura hierárquica que satisfaçam as seguintes condições (XU; II, 2009):

- $C_j \neq \emptyset, j = 1, \dots, k$;
- $X = \sum_{j=1}^k C_j$;
- $C_j \cap C_c = \emptyset, j \neq c, j, c = 1, \dots, k$.

Os algoritmos de agrupamento se utilizam de medidas de similaridade para quantificar quão similar são dois elementos. A medida de similaridade mais óbvia entre dois elementos é a distância Euclidiana entre eles. Se a distância for uma boa medida de similaridade, espera-se que a distância entre elementos de um mesmo grupo seja significativamente menor que a distância entre elementos de grupos diferentes. Outras medidas de proximidade são a distância de Mahalanobis ou o produto interno (*inner*) (RUSSELL; NORVIG, 2009).

Uma vez que a medida de similaridade é definida, é necessário definir um critério de agrupamento para quantificar a qualidade da partição dos dados. Neste ponto, o problema de agrupamento se transforma na busca de uma partição que otimize esse critério. O critério de agrupamento mais simples e mais utilizado é a soma do quadrado dos erros (DUDA; HART; STORK, 2001).

Os algoritmos de agrupamento mais populares são os baseados em otimização iterativa. Neles, a ideia básica é encontrar uma partição inicial e mover os dados de um grupo para outro a cada iteração se essa troca resultar na melhoria da qualidade da partição, dado um critério de agrupamento (DUDA; HART; STORK, 2001). Essa abordagem garante uma otimização local, mas não garante uma otimização global. As principais desvantagens dessa abordagem é que seu desempenho apresenta alto grau de dependência em relação à escolha inicial dos parâmetros, entretanto a simplicidade computacional desse tipo de algoritmo o deixa bastante atraente. Outro inconveniente desse método de agrupamento é a necessidade de se conhecer *a priori* o número de grupos k no conjunto de dados que está sendo analisado.

A seguir descrevemos os algoritmos de agrupamento baseado em otimização iterativa *K-Means*, *Fuzzy C-Means*, Modelo de Mistura de Gaussianas e o Modelo de Mistura de Gaussianas *Fuzzy*.

2.1.1 *K-Means* (KM)

O algoritmo *K-Means* (KM) (MACQUEEN, 1967) é a técnica mais simples de agrupamento de dados (DUDA; HART; STORK, 2001). No algoritmo KM o espaço de características é particionado em k grupos, cada um representado por um centroide, μ , chamado também de protótipo do grupo. O centroide do grupo nada mais é do que a média dos elementos daquele grupo. O algoritmo KM é um tipo de algoritmo de otimização iterativa pois os centroides tendem a se mover à medida que o critério soma do quadrado dos erros diminui.

No KM, a medida de similaridade é a distância Euclidiana. Sendo os dados vistos como vetores de características de dimensão p , $x = \{x^1, \dots, x^p\}$, a distância Euclidiana entre dois vetores x_1 e x_2 é dada por

$$D(x_1, x_2) = \sqrt{\sum_{c=1}^p (x_1^c - x_2^c)^2} = \|x_1 - x_2\| = (x_1 - x_2)^T (x_1 - x_2). \quad (2.1)$$

O objetivo do KM é a minimização do critério de agrupamento soma do quadrado dos erros. Seja $X = \{x_1, \dots, x_n\}$ um conjunto de dados p -dimensionais composto de n elementos particionado em k grupos, $C_1, \dots, C_k$. Seja μ_j a média dos elementos de um grupo C_j . Então, a soma do quadrado dos erros é dada por

$$J = \sum_{j=1}^k \sum_{x \in C_j} (D(x_i, \mu_j))^2 \quad (2.2)$$

O valor de J depende de como os dados de um grupo estão distribuídos ao redor do vetor que representa esse grupo, o centroide, e de quantos grupos foram formados. A partição ótima é aquela que minimiza J .

A cada iteração do algoritmo, k grupos, $C_1, \dots, C_k$, são formados associando cada elemento do conjunto de dados ao centroide mais próximo:

$$x_i \in C_j \iff D(x_i, \mu_j) < D(x_i, \mu_c), j \neq c, j, c = 1, \dots, k \quad (2.3)$$

para $1 \leq i \leq n$.

Em seguida o valor dos centroides são atualizados pela equação:

$$\mu_j^{novo} = \frac{1}{|C_j|} \sum_{x \in C_j} x \quad (2.4)$$

em que $|C_j|$ é o número de elementos associados ao conjunto C_j .

O KM pode ser resumido no Algoritmo 1:

Algoritmo 1: *K-Means* (KM)

- Inicializar os centroides dos grupos $\mu_1, \dots, \mu_k$.
 - Repetir até a convergência de J , Equação 2.2:
 - Formar k grupos associando cada exemplo ao seu centroide mais próximo seguindo o critério da Equação 2.3.
 - Atualizar o valor dos centroides μ_j , usando a Equação 2.4, para $j = 1, \dots, k$;
-

O critério de convergência adotado pode ser, por exemplo, a porcentagem de mudança da função objetivo J a cada iteração, assim, o algoritmo converge quando $\left| \frac{\Delta J}{J} \right| < Th$, em que ΔJ é a diferença do valor de J em duas iterações consecutivas e Th é um limiar de convergência pré-definido.

A partição dos dados em k grupos se dá pela associação de cada elemento do conjunto de dados a um dos centroides. Após a convergência do Algoritmo 1, é possível rotular cada elemento do conjunto de dados X , associando cada x_i ao centroide mais próximo usando a equação:

$$j^* = \underset{j}{\operatorname{argmin}} D(x_i, \mu_j). \quad (2.5)$$

O dado x_i é associado ao grupo j^* correspondente à distância $D(x_i, \mu_{j^*})$.

2.1.2 Modelo de Mistura de Gaussianas (GMM)

Os Modelos de Mistura são famílias de distribuições formadas pela composição de mais de uma distribuição básica. Esses modelos são formados pela distribuição de probabilidade D_1 com probabilidade w_1 , pela distribuição de probabilidade D_2 com probabilidade w_2 , etc. A densidade resultante é a combinação ponderada das densidades que caracterizam essas distribuições, sendo as respectivas probabilidades os fatores de ponderação da combinação. As distribuições $D_1, D_2, \dots$ são chamadas de componentes da mistura, enquanto que as probabilidades $w_1, w_2, \dots$ são chamadas de pesos da mistura.

O agrupamento baseado no Modelo de Mistura se enquadra na categoria de algoritmos baseados em otimização iterativa. No Modelo de Mistura, a medida de similaridade é a probabilidade *a posteriori* dos dados em relação às classes e o objetivo é a maximização do critério de agrupamento função de verossimilhança. Além de ser uma técnica de agrupamento de dados, o Modelo de Mistura nós dá uma descrição estatística dos dados através dos parâmetros da mistura. No caso de uma mistura de Gaussianas, grupos são descritos estatisticamente pelos parâmetros média, μ , e matriz de covariância, Σ .

No agrupamento baseado no Modelo de Mistura, pressupõe-se que os dados a serem agrupados são gerados a partir de uma distribuição de mistura composta por k componentes, $D_1, \dots, D_k$. Também é pressuposto que cada elemento do conjunto de dados é gerado a partir de um componente e que elementos de diferentes grupos são gerados a partir de diferentes componentes. A formação de agrupamento consiste, então, em recuperar um Modelo de Mistura a partir dos dados a serem agrupados. O problema é que não se conhece nem que componente gerou cada elemento, nem os parâmetros de cada componente e nem, tão pouco, o número de componentes (número de grupos). Utiliza-se, então, a abordagem da Máxima Verossimilhança e o algoritmo *Expectation Maximization* para estimar os parâmetros que maximizam a verossimilhança dos dados (RUSSELL; NORVIG, 2009).

Sendo $X = \{x_1, \dots, x_n\}$ um conjunto de dados p -dimensionais observados composto de n elementos, um Modelo de Mistura vai descrever o comportamento de um elemento do conjunto x_i , $1 \leq i \leq n$, através da função densidade de probabilidade $p(x_i)$ dada por

$$p(x_i) = \sum_{j=1}^k w_j p(x_i | \theta_j) \quad (2.6)$$

em que k é o número de distribuições que formam a mistura, as densidades $p(x|\theta_j)$ são os componentes da mistura, caracterizadas pelo vetor de parâmetros θ_j , e as probabilidades w_j são conhecidas como pesos da mistura e devem satisfazer os critérios: $w_j \geq 0$ e $\sum_{j=1}^k w_j = 1$, para $1 \leq j \leq k$. Em outras palavras, um Modelo de Mistura é caracterizado pelas distribuições D_j , cada uma com densidade $p(x|\theta_j)$, e pelas probabilidades w_j , para $1 \leq j \leq k$. No Modelo de Mistura de Gaussianas, os componentes da mistura são distribuições Gaussianas p -variadas com vetor de médias μ_j e matriz de covariância Σ_j ,

$$\begin{aligned} p(x|\theta_j) &= N(x_i | \mu_j, \Sigma_j) \\ &= \frac{1}{2\pi^{p/2} |\Sigma_j|^{1/2}} \exp \left[-\frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right]. \end{aligned} \quad (2.7)$$

Dois tipos de parâmetros devem ser estimados na mistura: os valores dos pesos, $\{w_1, \dots, w_k\}$, e os parâmetros, $\{\theta_1, \dots, \theta_k\}$, que caracterizam as distribuições, no caso do Modelo de Mistura de Gaussianas, $\theta_j = \{\mu_j, \Sigma_j\}$. O conjunto paramétrico do Modelo de Mistura de Gaussianas é dado por $\psi = \{w_1, \dots, w_k, \mu_1, \dots, \mu_k, \Sigma_1, \dots, \Sigma_k\}$, e a partir de agora, $p(x)$ será tratada como $p(x|\psi)$ para ficar evidente a relação do modelo com os seus parâmetros. No agrupamento, os parâmetros, $\psi_j = \{w_j, \mu_j, \Sigma_j\}$, para $1 \leq j \leq k$, que caracterizam cada componente da mistura podem ser encarados como o protótipo do grupo formado pelos dados com maior probabilidade de pertencer àquele componente.

Uma forma usual de se estimar os parâmetros ψ é através da técnica de Máxima Verossimilhança (ML, do inglês, *Maximum Likelihood*) (MCLACHLAN; PEEL, 2000). Assumindo que os dados X são independentes e identicamente distribuídos¹ (i.i.d.) com

¹ Uma sequência ou um conjunto de variáveis aleatórias são independentes e identicamente distribuídas

função de densidade $p(x|\psi)$, a densidade resultante para todo o conjunto de dados é

$$p(X|\psi) = \prod_{i=1}^n p(x_i|\psi). \quad (2.8)$$

A função $p(X|\psi)$ é chamada de verossimilhança dos parâmetros dado o conjunto de dados, ou somente função de verossimilhança. Analiticamente é sempre mais fácil trabalhar com o logaritmo da verossimilhança, $\ln p(X|\psi)$, do que com a própria verossimilhança (pois o produtório de $p(X|\psi)$ vira um somatório em $\ln p(X|\psi)$). Como o logaritmo da verossimilhança é monotonicamente crescente, o vetor de parâmetros ψ^* que maximiza o logaritmo da verossimilhança também maximiza a verossimilhança.

$$\begin{aligned} \ln p(X|\psi) &= \ln \left(\prod_{i=1}^n p(x_i|\psi) \right) \\ &= \sum_{i=1}^n \ln \{p(x_i|\psi)\}. \end{aligned} \quad (2.9)$$

A verossimilhança é interpretada como uma função dos parâmetros ψ e os dados X estão constantes. Na busca da Máxima Verossimilhança, o objetivo é achar uma boa estimativa de ψ usando as informações fornecidas pelo conjunto X através da maximização de $\ln p(X|\psi)$. Intuitivamente, a melhor estimativa de ψ é aquela que maximiza a probabilidade de se obter os dados observados. A estimativa da Máxima Verossimilhança dos parâmetros ψ^* é dada por

$$\psi^* = \underset{\psi \in \Psi}{\operatorname{argmax}} \ln p(X|\psi) \quad (2.10)$$

sendo Ψ o espaço paramétrico. A estimação da Máxima Verossimilhança quase sempre apresenta uma boa convergência quando o tamanho do conjunto de treinamento aumenta. Além disso, a estimação da Máxima Verossimilhança é quase sempre mais simples do que métodos alternativos como, por exemplo, estimação paramétrica Bayesiana (DUDA; HART; STORK, 2001).

Sempre que possível, a função de logaritmo da verossimilhança é maximizada pela resolução de um sistema de equações $\nabla \ln p(X|\psi^*) = 0$. Frequentemente, quando esta abordagem é aplicada a Modelos de Mistura de distribuições, ela não possui solução analítica explícita. McLachlan e Peel (MCLACHLAN; PEEL, 2000) propuseram o uso do algoritmo *Expectation Maximization* (EM) (DEMPSTER; LAIRD; RUBIN, 1977) para a estimação da Máxima Verossimilhança dos parâmetros de um Modelo de Mistura de Gaussianas.

(i.i.d.) se as variáveis aleatórias tiverem a mesma distribuição de probabilidade e sejam mutuamente independentes.

2.1.2.1 Maximizando a verossimilhança via EM

Em 1977, Dempster, Laird e Rubin introduziram o algoritmo iterativo *Expectation Maximization* (EM) (DEMPSTER; LAIRD; RUBIN, 1977). Esse método é uma solução geral a diversos problemas de estimação de parâmetros com dados incompletos como é o caso da mistura de distribuições.

Quando o método EM é aplicado à mistura de distribuições, tem-se um conjunto de dados completos $Y = \{X, R\}$ composto por um conjunto X de dados observados e um conjunto R de dados não observados. O logaritmo da função de densidade de probabilidade dos dados completos é dada por:

$$\begin{aligned}\ln p(Y|\psi) &= \ln p(X, R|\psi) \\ &= \ln (p(X|R, \psi)p(R|\psi)) \\ &= \ln p(X|R, \psi) + \ln p(R|\psi).\end{aligned}\tag{2.11}$$

O conjunto de valores observados $X = \{x_1, \dots, x_n\}$ é modelado por uma mistura de distribuições $p(x|\psi)$ definida na Equação 2.6. A parte não observada é composta por n rótulos $R = \{\mathbf{r}^1, \dots, \mathbf{r}^n\}$ associados a cada observação $x_1, \dots, x_n$, indicando qual componente da mistura descreverá cada observação. Os rótulos são valores binários k -dimensionais $\mathbf{r}^i = (r_1^i, \dots, r_k^i)$, com

$$r_j^i = \begin{cases} 1 & \text{se } x_i \in D_j \\ 0 & \text{caso contrário} \end{cases}\tag{2.12}$$

em que o j -ésimo componente corresponde à distribuição D_j (MCLACHLAN; PEEL, 2000).

A probabilidade de $x_i \in D_j$, ou seja, de r_j^i ser igual a um, pode ser especificada em termos dos pesos da mistura $w_1, \dots, w_k$, $p(r_j^i = 1) = w_j$. Assim, podemos escrever a distribuição de $\mathbf{r}^i$ como

$$p(\mathbf{r}^i|w) = \prod_{j=1}^k w_j^{r_j^i}\tag{2.13}$$

cujo logaritmo é

$$\ln p(\mathbf{r}^i|w) = \sum_{j=1}^k r_j^i \ln w_j\tag{2.14}$$

para $i = 1, \dots, n$.

Similarmente, a distribuição condicional de x_i dado $\mathbf{r}^i$ é descrita por $p(x_i|r_j^i = 1) = p(x_i|\theta_j)$, que também pode ser escrita na forma

$$p(x_i|\mathbf{r}^i, \theta) = \prod_{j=1}^k p(x_i|\theta_j)^{r_j^i}\tag{2.15}$$

cujo logaritmo é

$$\ln p(x_i|\mathbf{r}^i, \theta) = \sum_{j=1}^k r_j^i \ln p(x_i|\theta_j).\tag{2.16}$$

Pelo teorema de Bayes, o logaritmo da função de densidade de probabilidade dos dados completos foi simplificado por

$$\begin{aligned}\ln p(Y|\psi) &= \sum_{i=1}^n \left\{ \ln p(\mathbf{r}^i|w) + \ln p(x_i|\mathbf{r}^i, \theta) \right\} \\ &= \sum_{i=1}^n \left\{ \sum_{j=1}^k r_j^i \ln w_j + \sum_{j=1}^k r_j^i \ln p(x_i|\theta_j) \right\}\end{aligned}\quad (2.17)$$

e resulta no logaritmo da verossimilhança dos parâmetros ψ dado o conjunto de dados completos Y

$$\ln p(Y|\psi) = \sum_{i=1}^n \sum_{j=1}^k r_j^i \{ \ln w_j + \ln p(x_i|\theta_j) \}. \quad (2.18)$$

No algoritmo EM, os parâmetros são estimados alternando duas etapas: a etapa E (*Expectation*) que calcula a esperança do logaritmo da verossimilhança $l(\psi)$ com respeito às variáveis desconhecidas, condicionada ao dado observado X e aos valores iniciais definidos aos parâmetros ψ^t ; a etapa M (*Maximization*) que maximiza o logaritmo da verossimilhança para estimar os parâmetros em função dos resultados obtidos na etapa E.

Etapa E (*Expectation*):

Na t -ésima iteração, a etapa E calcula a esperança condicional com respeito a R , dadas as observações X e os parâmetros ψ^t , cuja equação é definida por

$$Q(\psi|\psi^t) = E_R \left\{ \ln p(Y|\psi) | X, \psi^t \right\} \quad (2.19)$$

em que $\ln p(Y|\psi)$ é o logaritmo da verossimilhança de ψ em relação à Y (MCLACHLAN; PEEL, 2000).

A função da Equação 2.18 é linear em relação aos dados não observados R :

$$\begin{aligned}Q(\psi|\psi^t) &= E_R \left\{ \sum_{i=1}^n \sum_{j=1}^k r_j^i \{ \ln w_j + \ln p(x_i|\theta_j) \} | X, \psi^t \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^k E_R \left\{ r_j^i | X, \psi^t \right\} \{ \ln w_j + \ln p(x_i|\theta_j) \}.\end{aligned}\quad (2.20)$$

Então, a etapa E se resume a calcular a esperança condicional da variável r_j^i dado x_i , ou seja,

$$E_R \left\{ r_j^i | X, \psi^t \right\} = \Pr(r_j^i = 1 | x_i, \psi^t) = z_j(\psi^t | x_i) \quad (2.21)$$

em que

$$z_j(\psi^t | x_i) = \frac{w_j p(x_i | \theta_j)}{\sum_{g=1}^k w_g p(x_i | \theta_g)}, \quad j = 1, \dots, k \quad (2.22)$$

é a probabilidade *a posteriori* de x_i pertencer a j -ésima componente da mistura com $j = 1, \dots, k$. A esperança condicional do logaritmo da verossimilhança resulta em (MCLACHLAN; PEEL, 2000):

$$\begin{aligned} Q(\psi|\psi^t) &= E_R \left\{ \ln p(Y|\psi) | X, \psi^t \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{ \ln w_j + \ln p(x_i|\theta_j) \}. \end{aligned} \quad (2.23)$$

Etapa M (*Maximization*):

Na t -ésima iteração, a etapa M calcula as estimativas dos parâmetros. Valores são atualizados em função dos estimadores

$$\psi^{t+1} = \psi^* = \underset{\psi \in \Psi}{\operatorname{argmax}} Q(\psi|\psi^t). \quad (2.24)$$

Os parâmetros $\psi^* = \{w^*, \mu^*, \Sigma^*\}$ são obtidos pela resolução do sistema de equações $\nabla_{\psi} Q(\psi|\psi^t) = 0$.

$$\begin{aligned} \nabla_{\psi} Q(\psi|\psi^t) &= 0 \\ \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \nabla_{\psi} \{ \ln w_j^* + \ln p(x_i|\theta_j^*) \} &= 0 \end{aligned} \quad (2.25)$$

com o gradiente ∇ como

$$\nabla_{\psi} = \begin{bmatrix} \frac{\partial}{\partial w} \\ \frac{\partial}{\partial \theta} \end{bmatrix}. \quad (2.26)$$

Para resolver a Equação 2.25 é possível maximizar o termo que contém w e o termo que contém θ separadamente uma vez que eles são independentes. Para achar a expressão de w_j^* é introduzido o multiplicador de Lagrange λ com a restrição que $\sum_j w_j = 1$ (BILMES, 1997):

$$\nabla_{\psi} \left[\sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \ln w_j^* + \lambda (\sum_j w_j - 1) \right] = 0 \quad (2.27)$$

ou²

$$\sum_{i=1}^n \sum_{j=1}^k \frac{1}{w_j^*} z_j(\psi^t|x_i) + \lambda = 0. \quad (2.28)$$

Desenvolvendo o somatório de j tem-se³ $\lambda = -n$, resultando em:

$$w_j^* = \frac{1}{n} \sum_{i=1}^n z_j(\psi^t|x_i) \quad j = 1, \dots, k. \quad (2.29)$$

² $\frac{d \ln(u)}{dx} = \frac{1}{u} \frac{du}{dx}$

³ $\sum_{i=1}^n \sum_{j=1}^k \frac{1}{w_j} z_j(\psi^t|x_i) = \sum_{i=1}^n \sum_{j=1}^k \frac{1}{w_j} \frac{w_j p(x_i|\theta_j)}{p(x_i|\psi^*)} = \sum_{i=1}^n 1 = n$

No Modelo de Mistura de Gaussianas multivariadas, os parâmetros que caracterizam a distribuição Gaussiana definida na Equação 2.7 são $\theta_j = \{\mu_j, \Sigma_j\}$, sendo μ_j o vetor média e Σ_j a matriz de covariância da j -ésima componente. O logaritmo da distribuição Gaussiana é dado por

$$\ln p(x_i|\theta) = -\frac{1}{2} \ln [(2\pi)^p |\Sigma_j|] - \frac{1}{2} (x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T. \quad (2.30)$$

Substituindo $\ln p(x_i|\theta)$ na esperança condicional do logaritmo da verossimilhança da Equação 2.23 tem-se

$$Q(\psi|\psi^t) = \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \left\{ \ln w_j - \frac{1}{2} \ln [(2\pi)^p |\Sigma_j|] - \frac{1}{2} (x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T \right\}. \quad (2.31)$$

Para cada $j = 1, \dots, k$, no cálculo de μ_j^* , é necessário derivar 2.31 em relação à μ_j e igualar a zero:

$$\begin{aligned} \nabla_{\mu} Q(\psi|\psi^t) &= 0 \\ \sum_{i=1}^n z_j(\psi^t|x_i) \left\{ (x_i - \mu_j^*) \Sigma_j^{-1} \right\} &= 0 \\ \sum_{i=1}^n z_j(\psi^t|x_i) x_i \Sigma_j^{-1} - \sum_{i=1}^n z_j(\psi^t|x_i) \mu_j \Sigma_j^{-1} &= 0 \\ \Sigma_j^{-1} \left[\sum_{i=1}^n z_j(\psi^t|x_i) x_i - \mu_j \sum_{i=1}^n z_j(\psi^t|x_i) \right] &= 0. \end{aligned} \quad (2.32)$$

Resolvendo a equação tem-se

$$\mu_j^* = \frac{\sum_{i=1}^n x_i z_j(\psi^t|x_i)}{\sum_{i=1}^n z_j(\psi^t|x_i)} = \frac{\sum_{i=1}^n x_i z_j(\psi^t|x_i)}{nw_j}. \quad (2.33)$$

Para calcular Σ_j^* podemos reescrever a Equação 2.31 ignorando os termos constantes e o parâmetro w da seguinte maneira ⁴(BILMES, 1997):

$$\begin{aligned} Q(\psi|\psi^t) &= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \left\{ -\frac{1}{2} \ln (|\Sigma_j|) - \frac{1}{2} (x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T \right\} \\ &= \sum_{j=1}^k \left[\frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i) \ln (|\Sigma_j^{-1}|) - \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i) (x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T \right] \\ &= \sum_{j=1}^k \left[\frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i) \ln (|\Sigma_j^{-1}|) - \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i) \text{tr} (\Sigma_j^{-1} A_{i,j}) \right] \end{aligned} \quad (2.34)$$

⁴ Para calcular Σ_j são necessários alguns conceitos de álgebra matricial:

- O traço de uma matriz quadrada $\text{tr}(A)$ é igual a soma dos elementos da diagonal de A ;
- O traço de um escalar é igual ao escalar;
- $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$ e $\text{tr}(AB) = \text{tr}(BA)$ que implica em $\sum_i x_i^T A x_i = \text{tr}(AB)$ com $B = \sum_i x_i x_i^T$;
- $|A^{-1}| = 1/|A| \rightarrow -\ln |A| = \ln \frac{1}{|A|} = \ln |A^{-1}|$;
- $\frac{\partial \log |A|}{\partial A} = 2A^{-1} - \text{diag}(A^{-1})$;
- $\frac{\partial \text{tr}[AB]}{\partial A} = B + B^T - \text{diag}(B)$.

com $A_{i,j} = ((x_i - \mu_j)(x_i - \mu_j)^T)$.

Derivando $Q(\psi|\psi^t)$ em relação à Σ_j tem-se

$$\begin{aligned}\nabla_{\Sigma} Q(\psi|\psi^t) &= \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i)(2\Sigma_j - \text{diag}(\Sigma_j)) - \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i)(2A_{i,j} - \text{diag}(A_{i,j})) \\ &= \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i)(2B_{i,j} - \text{diag}(B_{i,j})) \\ &= 2S - \text{diag}(S)\end{aligned}\tag{2.35}$$

com $B_{i,j} = \Sigma_j - A_{i,j}$ e $S = \frac{1}{2} \sum_{i=1}^n z_j(\psi^t|x_i)B_{i,j}$. Igualando a equação a zero, i.e., $2S - \text{diag}(S) = 0$ implica que $S = 0$. O que resulta em

$$\sum_{i=1}^n z_j(\psi^t|x_i)(\Sigma_j^* - A_{i,j}) = 0.\tag{2.36}$$

Resolvendo a equação tem-se

$$\begin{aligned}\Sigma_j^* &= \frac{\sum_{i=1}^n z_j(\psi^t|x_i) A_{i,j}}{\sum_{i=1}^n z_j(\psi^t|x_i)} \\ &= \frac{\sum_{i=1}^n (x_i - \mu_j)(x_i - \mu_j)^T z_j(\psi^t|x_i)}{\sum_{i=1}^n z_j(\psi^t|x_i)} \\ &= \frac{\sum_{i=1}^n (x_i - \mu_j)(x_i - \mu_j)^T z_j(\psi^t|x_i)}{nw_j}\end{aligned}\tag{2.37}$$

para $j = 1, \dots, k$.

As duas etapas do algoritmo EM são repetidas de forma alternada até a convergência do GMM. O critério de convergência adotado pode ser, por exemplo, a porcentagem de mudança de Q a cada iteração, assim, o algoritmo converge quando $\left| \frac{\Delta Q}{Q} \right| < Th$, em que $\Delta Q = Q(\psi|\psi^t) - Q(\psi|\psi^{t-1})$ e Th é um limiar de convergência pré-definido. O GMM pode ser resumido no Algoritmo 2.

O vetor de parâmetros, ψ , descrevem as propriedades globais dos dados mas não mostram como rotular os dados. A partição dos dados em k grupos se dá pela associação de cada elemento do conjunto de dados a um componente da mistura. Uma vez que os parâmetros do modelo, ψ , são estimados, é possível rotular cada elemento do conjunto de dados X associando cada x_i a um dos k grupos usando a regra de decisão Bayesiana.

Pelo teorema de Bayes, a probabilidade *a posteriori* de um elemento x_i pertencer ao j -ésimo grupo, com $j = 1, \dots, k$, é dada pela equação:

$$z_j(\psi^t|x_i) = \frac{w_j p(x_i|\theta_j)}{p(x_i|\psi)}, \quad j = 1, \dots, k\tag{2.38}$$

em que o denominador $p(x_i|\psi)$ é a distribuição da mistura da Equação 2.6. Para cada elemento x_i , $p(x_i|\psi)$ é constante para todas as classes. O classificador Bayesiano pode,

Algoritmo 2: EM para GMM

-
- Inicializar dos parâmetros $\psi^0 = (w^0, \theta^0)$, $t=0$.
 - Fazer até a convergência de $Q(\psi|\psi^t)$, Equação 2.23.
 - Etapa E (*Expectation*)
 - * Calcular a probabilidade *a posteriori*, $z_j(\psi^t|x_i)$, usando a Equação 2.22 para $i = 1, \dots, n$ e $j = 1, \dots, k$.
 - Etapa M (*Maximization*)
 - * Calcular os pesos da mistura, w_j^{t+1} , usando a Equação 2.29 para $j = 1, \dots, k$.
 - * Calcular o vetor de médias, μ_j^{t+1} , usando a Equação 2.33 para $j = 1, \dots, k$.
 - * Calcular as matrizes de covariância, Σ_j^{t+1} , usando a Equação 2.37 para $j = 1, \dots, k$.
 - $t = t + 1$;
-

então, ser usado para agrupar os dados usando a regra:

$$j^* = \underset{j}{\operatorname{argmax}} z_j(\psi^t|x_i). \quad (2.39)$$

O elemento x_i é associado ao grupo j^* ao qual ele tem a maior probabilidade *a posteriori* de pertencer.

2.1.3 Fuzzy C-Means (FCM)

Métodos de agrupamento como o *K-Means* e o GMM particionam os dados usando a lógica binária ou *hard*, na qual cada objeto só pode pertencer a um único grupo. Da necessidade de tratar dados cujos limites entre grupos são ambíguos, ou não estão bem definidos, pode-se usar a lógica *fuzzy* que admite valores intermediários entre o 0 (falso) e o 1 (verdadeiro) da lógica binária. Nos métodos de agrupamento *fuzzy*, os objetos podem pertencer a vários grupos ao mesmo tempo com diferentes graus de associação. Geralmente, o valor do grau de associação é um percentual que representa o quanto esse dado pertence a um grupo.

O algoritmo *Fuzzy C-Means* (FCM) (BEZDEK, 1981) é um algoritmo de agrupamento de dados baseado em otimização iterativa que usa a lógica *fuzzy*. Sendo $X = \{x_1, \dots, x_n\}$ um conjunto de dados p -dimensionais composto de n elementos, o objetivo do FCM é fornecer uma partição *fuzzy* de X em k grupos, com $2 \leq k \leq n$, buscando a minimização de uma função objetivo.

Um critério de agrupamento padrão em um algoritmo de agrupamento *fuzzy* do tipo FCM é a medida de dissimilaridade entre um elemento e o centro do grupo (centroide),

expressa pela função de dissimilaridade. Um algoritmo FCM busca, então, encontrar o centroide que minimiza a função de dissimilaridade. No FCM padrão, a função objetivo *fuzzy* que deve ser minimizada é dada por

$$J(U, \mu|X) = \sum_{i=1}^n \sum_{j=1}^k (u_{ij}^m d_{ij}^2) \quad (2.40)$$

em que u_{ij} representa o grau de pertinência do elemento x_i ao grupo j e é chamado de função de pertinência *fuzzy* (ou grau de pertinência), que deve satisfazer os critérios: $0 \leq u_{ij} \leq 1$ e $\sum_{j=1}^k u_{ij} = 1$, para $1 \leq j \leq k$ e $1 \leq i \leq n$. d_{ij} é a função de dissimilaridade, m é o expoente *fuzzy* de ponderação de cada função de pertinência, com $m > 1$, e $U = \{u_{ij}\}$ é a partição *fuzzy* de X .

A função de dissimilaridade ajuda a focar na busca por uma estrutura interna dos dados e cada tipo de função implica em uma diferente maneira de ver a geometria dos grupos. Ao lidar com distância de Mahalanobis, por exemplo, a forma geométrica buscada nos grupos é a hiperelipsoidal. Já a distância Tchebyshev foca na estrutura em forma de hipercubos (PEDRYCZ, 2005). A função de dissimilaridade mais comum, no entanto, é a distância Euclidiana que busca uma estrutura na forma de hiperesfera nos dados. Assim,

$$d_{ij} = D(x_i, \mu_j) \quad (2.41)$$

sendo $D(x_i, \mu_j)$ dado pela Equação 2.1, x_i é o i -ésimo elemento e μ_j é o centroide do j -ésimo grupo.

Para minimizar a função objetivo J , novas posições dos centroides devem ser determinadas usando as seguintes equações:

$$u_{ij} = \left[\sum_{c=1}^k \left(\frac{d_{ij}}{d_{ic}} \right)^{2/(m-1)} \right]^{-1} \quad (2.42)$$

$$\mu_j^{novo} = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (2.43)$$

em que μ_j^{novo} é a nova posição do j -ésimo grupo. O cálculo do grau de pertinência u_{ij} se baseia nos valores prévios das posições dos centroides.

O FCM itera usando as Equações 2.41-2.43, até que a função objetivo convirja. O critério de convergência adotado pode ser, por exemplo, a porcentagem de mudança da função objetivo J a cada iteração, assim, o algoritmo converge quando $\left| \frac{\Delta J}{J} \right| < Th$, em que ΔJ é a diferença do valor de J em duas iterações consecutivas e Th é um limiar de convergência pré-definido. Assim, o FCM pode ser resumido no Algoritmo 3:

A partição *hard* dos dados em k grupos se dá pela associação de cada elemento do conjunto de dados a um dos centroides. Após a convergência do Algoritmo 3, é possível

Algoritmo 3: *Fuzzy C-Means* (FCM)

-
- Inicializar a partição *fuzzy* U satisfazendo os critérios $0 \leq u_{ij} \leq 1$ e $\sum_{j=1}^k u_{ij} = 1$, para $1 \leq j \leq k$ e $1 \leq i \leq n$.
 - Repetir até a convergência de $J(U, \psi|X)$, Equação 2.40:
 - Calcular o vetor de médias, $\{\mu_j\}$, usando a Equação 2.43, para $j = 1, \dots, k$.
 - Calcular a função de dissimilaridade, d_{ij} , usando a Equação 2.41, para $i = 1, \dots, n$ e $j = 1, \dots, k$. Se $d_{ij} = 0$ para algum j , faça $u_{ij} = 1$, $u_{ig} = 0$, $\forall g \neq j$.
 - Atualizar a partição *fuzzy* U usando a Equação 2.42, para $i = 1, \dots, n$ e $j = 1, \dots, k$.
-

rotular cada elemento do conjunto de dados X associando cada x_i a um dos k grupos usando a partição *fuzzy*.

Sendo $U = \{u_{ij}\}$ a partição *fuzzy* de X , u_{ij} representa o grau de pertinência do elemento x_i ao grupo j . O elemento x_i é associado ao grupo j^* com maior grau de pertinência:

$$j^* = \underset{j}{\operatorname{argmax}} u_{ij}. \quad (2.44)$$

2.1.4 Modelo de Mistura de Gaussianas *Fuzzy* (FGMM)

O GMM e o FCM possuem características similares. Além de ambos serem técnicas de otimização iterativa, os dois métodos possibilitam uma partição *soft* do espaço de características, ou seja, cada observação pode pertencer a mais de um grupo. No GMM, a função de densidade de probabilidade $p(x|\theta_j)$ condicionada a um grupo j indica o grau da associação de uma observação x a um grupo j . Já no FCM, o grau da associação está representado pelo grau de pertinência u_{ij} . Além disso, o somatório tanto de $p(x|\theta_j)$ quanto de u_{ij} em relação a todos os grupos deve ser 1.

Tran *et al.* (TRAN; WAGNER, 1998) propuseram uma modificação do FCM baseada no Modelo de Mistura de Gaussianas chamada de Modelo de Mistura de Gaussianas *Fuzzy* (FGMM, do inglês, *Fuzzy Gaussian Mixture Model*). O FGMM surgiu da necessidade do FCM de tratar a presença de grupos de diferentes formas e densidades no conjunto de dados a ser agrupado. O FGMM explora a abordagem Bayesiana ao considerar a probabilidade *a priori* de cada grupo e a função de densidade de probabilidade condicional de cada grupo. Além disso, Ju e Liu (JU; LIU, 2012) mostraram que ao mesclar as duas técnicas, é possível obter a capacidade de generalização do GMM com uma convergência mais rápida característica dos algoritmos da família do FCM.

Diferente do GMM que usa a Máxima Verossimilhança via *Expectation Maximization*

para o ajuste do modelo aos dados e estimação dos parâmetros do modelo, o FGMM utiliza um processo de otimização iterativa baseado no gradiente, que é matematicamente mais simples, para estimar os valores dos parâmetros.

A medida de dissimilaridade do FCM foi generalizada através do uso do vetor de médias *fuzzy*, da matriz de covariância *fuzzy* e do peso da mistura *fuzzy* com o objetivo de obter um resultado satisfatório para o problema da variedade de formas e densidades dos grupos. A função de dissimilaridade d_{ij} é definida então como

$$d_{ij}^2 = -\log p(x_i, j|\psi) = -\log[w_j N(x_i|\mu_j, \Sigma_j)] \quad (2.45)$$

para $1 \leq j \leq k$ e $1 \leq i \leq n$, em que w_j são os pesos da mistura, também interpretados como a probabilidade *a priori* de um elemento pertencer ao j -ésimo grupo, e deve obedecer às seguintes restrições $0 \leq w_j \leq 1$ e $\sum_{j=1}^k w_j = 1$. $N(x_i|\mu_j, \Sigma_j)$ são as Gaussianas p -dimensionais componentes da mistura com vetor de médias μ_j e matriz de covariância Σ_j . ψ_j consiste no conjunto de parâmetros do modelo $\psi_j = \{w_j, \mu_j, \Sigma_j\}$. Substituindo $N(x_i|\mu_j, \Sigma_j)$, Equação 2.7, na Equação 2.45, obtém-se

$$d_{ij}^2 = -\log w_j + \frac{1}{2} \log(2\pi)^p |\Sigma_j| + \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j). \quad (2.46)$$

Comparada com a Equação 2.1, a função de dissimilaridade definida pela Equação 2.45 considera, não só a distância, mas também a covariância e os pesos da mistura.

Assim como outras técnicas de agrupamento, o processo de agrupamento por FGMM busca a minimização de uma função objetivo $J(U, \psi|X)$, Equação 2.40. Substituindo a Equação 2.46 na Equação 2.40 tem-se

$$J(U, \psi|X) = -\sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log w_j - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log N(x_i|\mu_j, \Sigma_j). \quad (2.47)$$

A minimização de J consiste na busca das estimativas dos parâmetros ótimos $\psi_j^* = \{w_j^*, \mu_j^*, \Sigma_j^*\}$ que minimizam J . A estimativa são obtidas por $\nabla_{\psi} J(U, \psi|X) = 0$ com

$$\nabla_{\psi} = \begin{bmatrix} \frac{\partial}{\partial w} \\ \frac{\partial}{\partial \mu} \\ \frac{\partial}{\partial \Sigma} \end{bmatrix}. \quad (2.48)$$

Para resolver o sistema de equações $\nabla_{\psi} J(U, \psi|X) = 0$ é possível minimizar o termo que contém w e o termo que contém μ e Σ separadamente uma vez que eles são independentes.

No cálculo da estimativa de w^* , minimiza-se o primeiro termo do lado direito da Equação 2.47 respeitando a condição $\sum_{j=1}^k w_j = 1$. É necessário o uso de multiplicadores de Lagrange λ resultando na expressão de $f(w)$ que deve ser maximizada.

$$f(w) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log w_j - \lambda \left(\sum_{j=1}^k w_j - 1 \right). \quad (2.49)$$

Para cada elemento $i = 1, \dots, n$, deriva-se $f(w_s)$ com respeito à w_s e iguala-se o resultado a zero

$$\begin{aligned} \frac{\partial f(w_s)}{\partial w_s} &= 0 \\ \frac{\partial}{\partial w_s} [u_{is}^m \log w_s^* - \lambda (w_s - 1)] &= 0 \\ \frac{u_{is}^m}{w_s^*} - \lambda &= 0. \end{aligned} \quad (2.50)$$

Resolvendo a equação tem-se

$$w_s^* = \frac{u_{is}^m}{\lambda}. \quad (2.51)$$

Levando em consideração a restrição $\sum_{j=1}^k w_j = 1$, tem-se

$$\sum_{j=1}^k \frac{u_{ij}^m}{\lambda} = 1. \quad (2.52)$$

Isolando λ tem-se

$$\lambda = \sum_{j=1}^k u_{ij}^m. \quad (2.53)$$

Substituindo 2.53 em 2.51 tem-se

$$w_s^* = \frac{u_{is}^m}{\sum_{j=1}^k u_{ij}^m} \quad (2.54)$$

obtendo-se a equação de w_j^* definida como peso da mistura *fuzzy*:

$$w_j^* = \frac{\sum_{i=1}^n u_{ij}^m}{\sum_{i=1}^n \sum_{c=1}^k u_{ic}^m} = \frac{1}{n} \sum_{i=1}^n u_{ij}^m \quad (2.55)$$

No cálculo das estimativas μ_j^* e Σ_j^* , minimiza-se o segundo termo do lado direito da Equação 2.47 em relação à cada um dos parâmetros.

$$\begin{aligned} f(\mu, \Sigma) &= - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log N(x_i | \mu_j, \Sigma_j) \\ &= - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \left[\frac{1}{2} \log(2\pi)^p |\Sigma_j| + \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right]. \end{aligned} \quad (2.56)$$

Para cada grupo $j = 1, \dots, k$, derivando-se $f(\mu, \Sigma)$ com respeito à μ tem-se

$$\begin{aligned} \sum_{i=1}^n u_{ij}^m \Sigma_j^{-1} (x_i - \mu_j) &= 0 \\ \sum_{i=1}^n u_{ij}^m x_i \Sigma_j^{-1} - \sum_{i=1}^n u_{ij}^m \mu_j \Sigma_j^{-1} &= 0 \\ \Sigma_j^{-1} \left[\sum_{i=1}^n u_{ij}^m x_i - \mu_j \sum_{i=1}^n u_{ij}^m \right] &= 0. \end{aligned} \quad (2.57)$$

Resolvendo a equação tem-se

$$\mu_j^* = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}. \quad (2.58)$$

O cálculo da estimativa Σ_j^* requer um maior número de manipulações envolvendo álgebra matricial⁵ para chegar ao resultado final. Para cada grupo $j = 1, \dots, k$, tem-se que $f(\mu, \Sigma)$, eliminando os termos constantes, é o seguinte:

$$f(\mu, \Sigma) = \sum_{i=1}^n u_{ij}^m \left[-\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right] \quad (2.59)$$

Visto que $(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$ é um escalar, e conhecendo as propriedades do traço de uma matriz, pode-se dizer que $(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) = \text{tr} \left((x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right) = \text{tr} \left(\Sigma_j^{-1} N^i \right)$ sendo $N^i = (x_i - \mu_j)^T (x_i - \mu_j)$.

$$f(\mu, \Sigma) = \sum_{i=1}^n u_{ij}^m \left[\frac{1}{2} \log |\Sigma_j^{-1}| - \frac{1}{2} \text{tr} \left(\Sigma_j^{-1} N^i \right) \right]. \quad (2.60)$$

Fazendo $\Sigma_j^{-1} = A$ e derivando $f(\mu, \Sigma)$ em relação à A tem-se

$$\frac{\partial}{\partial A} \left\{ \sum_{i=1}^n u_{ij}^m \left[\frac{1}{2} \log |A| - \frac{1}{2} \text{tr} \left(A N^i \right) \right] \right\} = \sum_{i=1}^n u_{ij}^m \left[\frac{1}{2} A^{-1} - \frac{1}{2} N^{iT} \right]. \quad (2.61)$$

Mas sabendo que $(x_i - \mu)(x_i - \mu)^T$ é simétrica, tem-se que $\left((x_i - \mu)(x_i - \mu)^T \right)^T = (x_i - \mu)(x_i - \mu)^T$, em consequência, $N^T = N$. Substituindo A e N^T em $f(\mu, \Sigma)$ e fazendo $\frac{\partial f(\mu, \Sigma)}{\partial A} = 0$ resulta em

$$\begin{aligned} \sum_{i=1}^n u_{ij}^m \left[\frac{1}{2} \left(\Sigma_j^{*-1} \right)^{-1} - \frac{1}{2} \left((x_i - \mu_j)^T (x_i - \mu_j) \right)^T \right] &= 0 \\ \sum_{i=1}^n u_{ij}^m \left[\frac{1}{2} \Sigma_j^* - \frac{1}{2} (x_i - \mu_j)^T (x_i - \mu_j) \right] &= 0 \\ \frac{1}{2} \left[\Sigma_j^* \sum_{i=1}^n u_{ij}^m - \sum_{i=1}^n u_{ij}^m (x_i - \mu_j)^T (x_i - \mu_j) \right] &= 0. \end{aligned} \quad (2.62)$$

Resolvendo a equação tem-se

$$\Sigma_j^* = \frac{\sum_{i=1}^n u_{ij}^m (x_i - \mu_j)^T (x_i - \mu_j)}{\sum_{i=1}^n u_{ij}^m}. \quad (2.63)$$

⁵ Para calcular Σ são necessários alguns conceitos de álgebra matricial:

- O traço de uma matriz quadrada $\text{tr}(A)$ é igual a soma dos elementos da diagonal de A ;
- $\text{tr}(AB) = \text{tr}(BA)$ que implica em $\sum_i x_i^T A x_i = \text{tr}(AB)$ com $B = \sum_i x_i x_i^T$;
- $|A^{-1}| = 1/|A| \rightarrow -\ln |A| = \ln \frac{1}{|A|} = \ln |A^{-1}|$;
- Se A é inversível, $(A^{-1})^{-1} = A$;
- $\frac{\partial \ln |A|}{\partial A} = A^{-1}$;
- $\frac{\partial \text{tr}(AB)}{\partial A} = B^T$.

O FGMM itera atualizando consecutivamente os valores dos parâmetros do modelo, $\psi = \{w, \mu, \Sigma\}$, da função de dissimilaridade, d , e da partição *fuzzy*, U , até que a função objetivo convirja. O critério de convergência adotado pode ser, por exemplo, a porcentagem de mudança da função objetivo J a cada iteração, assim, o algoritmo converge quando $\left| \frac{\Delta J}{J} \right| < Th$, em que ΔJ é a diferença do valor de J em duas iterações consecutivas e Th é um limiar de convergência pré-definido. Assim, o FGMM pode ser resumido no Algoritmo 4:

Algoritmo 4: FGMM

- Inicializar a partição *fuzzy* U .
 - Repetir até a convergência de $J(U, \psi|X)$, Equação 2.40:
 - Calcular os pesos da mistura, $\{w_j\}$, os vetores de médias, $\{\mu_j\}$, e as matrizes de covariância, $\{\Sigma_j\}$, usando, respectivamente as Equações 2.55, 2.58 e 2.63, para $j = 1, \dots, k$.
 - Calcular a função de dissimilaridade, d_{ij} , usando a Equação 2.45, para $i = 1, \dots, n$ e $j = 1, \dots, k$. Se $d_{ij} = 0$ para algum j , faça $u_{ij} = 1$, $u_{ig} = 0$, $\forall g \neq j$.
 - Atualizar a partição *fuzzy* U usando a Equação 2.42, para $i = 1, \dots, n$ e $j = 1, \dots, k$.
-

2.2 Segmentação de imagens

Imagem (do latim, *imago*) significa representação visual de um objeto. Para fins computacionais, imagem é uma representação em 2 dimensões composta de um número finito de elementos, cada um com localização e valor específicos. Estes elementos são chamados de *pixel* (do inglês, *picture element*). Uma imagem digital pode ser definida como uma função bidimensional, $f(x, y)$, onde x e y são coordenadas espaciais e a amplitude de f em qualquer par de coordenadas é chamada de intensidade ou nível de cinza da imagem nesse ponto. Quando x , y e os valores de intensidade de f são quantidades finitas e discretas, chamamos de imagem digital (GONZALEZ; WOODS, 2009).

Uma imagem digital também pode ser representada em forma de uma matriz bidimensional $A \times B$. Nesta matriz, cada elemento $f(x, y)$, $x = 0, 1, \dots, A - 1$ e $y = 0, 1, \dots, B - 1$ é chamado *pixel*. Diz-se então que a imagem tem dimensão A *pixels* na horizontal (eixo x) e B *pixels* na vertical (eixo y).

Imagens multibandas ou multiespectrais, $f(x, y)$ é um vetor, $(f_1(x, y), \dots, f_p(x, y))$, em que cada componente indica a intensidade da cena no ponto (x, y) da correspondente banda espectral e p indica a dimensão do vetor. Nesse caso, é necessária uma sequência

de imagens monocromáticas para formar uma imagem colorida. Em geral, $f(x, y)$ pode representar grandezas diferentes, tais como, temperatura, pressão, frequência, amostradas em pontos (x, y) e com intervalos de valores distintos. Uma imagem multiespectral também pode ser representada como uma sequência de p imagens monocromáticas, tal que cada imagem é conhecida como banda (PEDRINI; SCHWARTZ, 2008).

Uma imagem colorida é uma imagem multiespectral, em que a cor em cada ponto (x, y) é definida por meio de três grandezas: luminância, matiz e saturação. A luminância está associada com o brilho da luz, o matiz com o comprimento de onda dominante e a saturação com o grau de pureza (ou intensidade) do matiz.

O modelo de cor RGB, usado para representar imagens coloridas em dispositivos como monitores e câmeras de vídeo, é adotado frequentemente no processamento no processamento digital de imagens. O modelo RGB é um modelo aditivo no qual cada cor representada é definida pela composição dos componentes espectrais do vermelho (R, red), verde (G, green) e azul (B, blue). Essas são as cores primárias, a partir das quais todas as outras se originam por composição. Imagens representadas no modelo de cor RGB são imagens multiespectrais com três bandas, uma para cada cor primária (GONZALEZ; WOODS, 2009).

Um *pixel* i na coordenada (x, y) tem quatro vizinhos horizontais e verticais cujas coordenadas são dadas por: $(x - 1, y)$, $(x + 1, y)$, $(x, y - 1)$ e $(x, y + 1)$. Esse conjunto de *pixels*, chamados de vizinhança-4 de i é expresso por V_i^4 . Cada *pixel* é uma unidade de distância de (x, y) , e alguns vizinhos de i ficarão para fora da imagem se (x, y) estiver na borda da imagem (GONZALEZ; WOODS, 2009).

Os quatro vizinhos diagonais de i têm coordenadas, $(x - 1, y - 1)$, $(x + 1, y - 1)$, $(x - 1, y + 1)$ e $(x + 1, y + 1)$ que são expressos por V_i^D . Esses pontos, juntos com a vizinhança-4, são chamados de vizinhança-8 de i , expressos por V_i^8 . Como no caso anterior, alguns vizinhos em V_i^D e V_i^8 ficarão para fora da imagem se (x, y) se localizar na borda da imagem. A Figura 1 ilustra os 3 tipos de vizinhança descritos anteriormente.

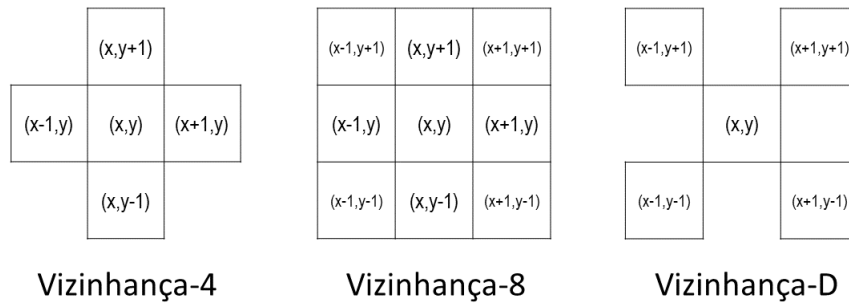


Figura 1 – Vizinhança-4, vizinhança-8 e vizinhança-D.

A segmentação subdivide uma imagem em seus objetos ou regiões constituintes

seguindo algum critério. O nível de detalhamento pelo qual uma subdivisão é realizada depende da aplicação. Isso significa que uma segmentação deve terminar quando os objetos ou regiões constituintes para a determinada aplicação são detectados. A maior dificuldade da segmentação automática é a determinação do critério de parada da segmentação e do critério usado para definir o grau de similaridade ou discrepância entre os elementos analisados. A segmentação de imagens é uma das tarefas mais difíceis no processamento de imagens. A precisão da segmentação determina o sucesso ou o fracasso final nos procedimentos de análise da imagem e auxilia na extração de informações das imagens para uma dada aplicação (GONZALEZ; WOODS, 2009).

Os algoritmos de segmentação podem ser divididos em duas categorias básicas relacionadas às propriedades dos valores de intensidade: descontinuidade e similaridade. Na primeira categoria, a abordagem é dividir uma imagem com base nas mudanças bruscas de intensidade, como as bordas. Na segmentação baseada nas bordas, o pressuposto é que as fronteiras das regiões são suficientemente diferentes entre si em relação ao fundo da imagem para permitir a detecção de limites com base nas descontinuidades locais em intensidade. Na segunda categoria estão as abordagens da segmentação baseada em região que fazem a divisão de uma imagem em regiões disjuntas e internamente homogêneas de acordo com um conjunto de critérios predefinidos. A limiarização, o crescimento de região e a divisão e fusão de regiões são exemplos dos métodos desta categoria (GONZALEZ; WOODS, 2009).

Muitos métodos vêm sendo propostos e estudados para o problema de segmentação de imagens (VANTARAM; SABER, 2012). Dentre eles estão os baseados em contornos ativos (BLAKE; ISARD, 1998) (por exemplo, *snakes* e *level sets*), crescimento de regiões e *split-and-merge* (ADAMS; BISCHOF, 1994), em agrupamento (FAWCETT, 2006), por exemplo, *mean shift*, *K-Means*, *Fuzzy C-Means*, *Modelo de Mistura*, e em energia (BOYKOV; FUNKA-LEA, 2006), por exemplo, *graph cuts*.

Na Seção 2.2.1, é feita uma breve explicação de como algoritmos de agrupamento podem ser usados na segmentação de imagens baseada em regiões.

2.2.1 Segmentação via agrupamento de *pixels*

Assumindo que R representa toda a região espacial ocupada por uma imagem. Podemos ver a segmentação baseada em regiões de uma imagem como um processo que particiona R em k sub-regiões, $R_1, \dots, R_k$ de tal maneira que *pixels* de uma mesma região possuam propriedades semelhantes e que *pixels* de regiões diferentes possuam propriedades diferentes.

- $R_j \neq \emptyset, j = 1, \dots, k;$
- $R = \sum_{j=1}^k R_j;$

- $R_j \cap R_c = \emptyset, j \neq c, j, c = 1, \dots, k.$

Todos os *pixels* da imagem devem fazer parte de alguma sub-região, a união de todas as sub-regiões forma a região R completa e as regiões são disjuntas entre si.

Uma maneira natural de realizar a segmentação é subdividindo os componentes da imagem que mais se assemelham, lógica usada pelos algoritmos de agrupamento de dados. Nessa abordagem, as imagens são representadas em termos de grupos de *pixels* que se assemelham. O critério específico de similaridade a ser adotado depende do domínio do problema. Grupos de *pixels* podem se assemelhar por terem a mesma cor e/ou mesma textura e/ou por serem vizinhos, etc (FORSYTH; PONCE, 2002; THILAGAMANI; SHANTI, 2011).

No agrupamento baseado em partição, um grande conjunto de dados é dividido em sub-conjuntos de maneira que haja uma semelhança entre os elementos de cada grupo formado. A decomposição de uma imagem em regiões cujo interior mantenha coerência de cor e textura ou a decomposição de um vídeo em *shots* (segmentos de vídeo que mostram a mesma coisa ou um mesmo ponto de vista) são exemplos do uso do particionamento no processamento de imagens.

A essência da segmentação de imagens baseada em agrupamento de *pixels* é analisar os atributos de um conjunto de pixels e particioná-los usando funções objetivo pré-definidas que identifiquem agrupamentos significativos de *pixels* (VANTARAM; SABER, 2012; SARASWATHI; ALLIRANI, 2013). Na segmentação baseada em região utilizando técnicas de agrupamento cada grupo encontrado nos *pixels*, deve ser tratado como uma região da imagem. O fim do processo de agrupamento se dá quando os *pixels* pertencentes a um grupo/região específico possuem, em geral, uma alto grau de similaridade e *pixels* de diferentes grupos/regiões possuam alto grau de dissimilaridade. A principal vantagem da abordagem de agrupamento é a sua simplicidade e a facilidade de implementação e seu maior desafio é trabalhar com o grande número de *pixels* presentes na imagem.

Uma ramificação dos métodos de segmentação de imagens baseados em agrupamento são aqueles que consideram, no processo de agrupamento, o contexto em que os *pixels* estão inseridos. Na análise de imagens, o contexto ocupa um importante papel na interpretação de uma cena. No menor nível de abstração, *pixels* isolados fornecem menos informações do que os mesmos *pixels* juntos dando informação de contexto espacial, espectral ou temporal. Contexto espacial é obtido considerando *pixels* adjacentes em uma vizinhança espacial. Informação espectral diz respeito ao uso de diferentes bandas da imagem captadas por diferentes sensores enquanto que o contexto temporal é obtido considerando uma mesma cena em diferentes momentos.

Na abordagem pontual de análise de imagens, assume-se que não existe relação entre as várias classes ou *pixels* da imagem. Em outras palavras, sendo um *pixel* x pertencente à

classe k , o *pixel* vizinho poderia pertencer a qualquer outra classe. Na análise dependente de contexto, classificar cada *pixel* separadamente não tem sentido. A classe à qual cada *pixel* é associado depende (a) dos seus próprios atributos, (b) dos atributos dos outros *pixels* e (c) da relação existente entre as classes. Tal problema aparece em várias aplicações tais como comunicação, séries temporais, reconhecimento de voz e processamento de imagens (THEODORIDIS; KOUTROUMBAS, 2009; KANNAN; RAMATHILAGAM; CHUNG, 2012).

Sendo as relações espaciais um atributo essencial em uma imagem, restrições espaciais podem ser necessárias para alcançar uma segmentação de imagens mais eficiente. A abordagem contextual permite incorporar o conhecimento *a priori* de que *pixels* vizinhos provavelmente pertencem ao mesmo grupo. Isto é, além dos valores de intensidade, a informação espacial do *pixel* e sua relação com a sua vizinhança também é usada para determinar o grupo ao qual cada *pixel* é associado. A ideia de combinar a correlação espacial dos *pixels* vizinhos e algoritmos de agrupamento para realizar a segmentação contextual de imagens pode ser explorada de diferentes maneiras para cada técnica de agrupamento. Essa combinação se destaca, especialmente, na segmentação de imagens com ruído (GONG et al., 2013; ZHANG; WU; NGUYEN, 2013) e de imagens com regiões compostas por diferentes texturas (MIGNOTTE, 2011).

Contexto pode ser considerado no agrupamento pelo *Fuzzy C-Means* adicionando um termo que incorpore a relação espacial de *pixels* vizinhos à função objetivo (GONG et al., 2013), modelando a correlação espacial na função de pertinência *fuzzy* (LI; ZHANG; QU, 2008) ou utilizando as duas abordagens combinadas (NGUYEN; WU, 2011b). A restrição espacial também pode ser incorporada no algoritmo *K-Means* na fase de rotulagem dos dados (MIGNOTTE, 2011; WANG et al., 2012). A informação da vizinhança pode ser incorporado ao Modelo de Mistura de Gaussianas modelando a correlação espacial diretamente nos pesos das misturas com o uso Campos Aleatórios Markovianos (NIKOU; GALATSANOS; LIKAS, 2007) ou aplicando um *template* de médias no cálculo dos componentes da mistura (ZHANG; WU; NGUYEN, 2013).

Modelos de Mistura Contextuais são explicados mais detalhadamente no Capítulo 3 pois são uma peça fundamental no desenvolvimento deste trabalho. A Seção 2.3 é dedicada à descrição de Campos Aleatórios Markovianos, ferramenta adotada neste trabalho para modelar a correlação espacial no modelo proposto.

2.3 Campos Aleatórios Markovianos

O processo de Markov e, em particular, a cadeia de Markov, é um dos tipos de modelos mais populares para representar dados dependentes no tempo. A estrutura unidimensional do tempo simplifica os cálculos de propriedades desses modelos. Os Campos

Aleatórios Markovianos (MRF, do inglês, *Markov Random Field*) são uma generalização das cadeias de Markov substituindo o espaço com índice unidimensional do tempo por um espaço com índice mais genérico, tal como o espaço geográfico ou uma posição em um grafo de vizinhança.

A teoria de MRF fornece uma maneira conveniente e consistente de modelar entidades com dependência espacial assim como os *pixels* em uma imagem. Essa modelagem é alcançada através da caracterização da influência mútua entre as entidades usando distribuições condicionais de MRF. Entretanto, para entender a teoria de MRF é necessário que se tenha em mente a definição de um sistema de vizinhança.

2.3.1 Sistema de vizinhança

Seja S um reticulado (do inglês, *lattice*) bidimensional finito de tamanho $A \times B$.

$$S = \{(x, y) | 1 \leq x \leq A, 1 \leq y \leq B\} \quad (2.64)$$

Cada sítio em S é usado para indexar um atributo relevante do dado que está sendo analisado e vai variar dependendo da aplicação. Na análise de imagens, S pode representar um *pixel* da imagem de tamanho $A \times B$ (LI, 2009). Convenientemente, um *pixel* na posição (x, y) de uma imagem pode ser reindexado como um único índice s , com s assumindo valores entre $\{1, 2, \dots, n\}$ com $n = A \times B$. A relação entre os sítios é mantida ao definirmos o sistema de vizinhança de cada $s \in S$. Para fins práticos, adotaremos essa notação, exceto quando uma elaboração seja necessária.

Associado a cada sítio $s \in S$ existe uma variável aleatória $X = \{x_s | s \in S\}$ representando o estado (ou rótulo) do atributo correspondente. x_s pode ser a intensidade do *pixel* na posição s , $x_s \in \{0, \dots, 255\}$. Na detecção de bordas, x_s pode assumir os valores $\{0, 1\}$ se existir ou não um elemento de borda no sítio s . Na segmentação de imagens em k regiões, x_s pode assumir o rótulo associado a cada região da imagem, $x_s \in \{1, \dots, k\}$.

Um sistema de vizinhança $\mathcal{N}_s$ é definido para representar as relações entre os sítios do reticulado. Para cada sítio $s \in S$ vai existir uma vizinhança $\mathcal{N}_s$ que respeita as seguintes propriedades:

- $s \notin \mathcal{N}_s$,
- $s \in \mathcal{N}_m \Leftrightarrow m \in \mathcal{N}_s$.

Em outras palavras, nenhum sítio s pode pertencer a sua própria vizinhança e se s é vizinho de m , então, m também será vizinho de s (THEODORIDIS; KOUTROUMBAS, 2009). Em um reticulado bidimensional, $\mathcal{N}_s$ pode ser definido como o conjunto de elementos de S cuja distância ao elemento central s é menor ou igual um determinado valor. Se s

representa o *pixel* i de uma imagem, $\mathcal{N}_s$ pode ser, por exemplo, a vizinhança-4 do *pixel* i como descrito na Seção 2.2, nesse caso, $\mathcal{N}_s = V_i^4$.

Um *clique* c , dado S e $\mathcal{N}$, é um subconjunto de sítios de S , no qual quaisquer dois sítios de c são vizinhos com respeito ao tipo de vizinhança escolhido. Por definição, qualquer conjunto unitário formado por um único sítio e o conjunto vazio são considerados um *clique*. Cada tipo de *clique* em um reticulado bidimensional é determinado pelo seu tamanho, forma e orientação. A Figura 2 ilustra dois casos de vizinhança em um reticulado bidimensional, vizinhança-4 e vizinhança-8, e os tipos de *cliques* correspondentes. Ao aumentar o número de sítios em um sistema de vizinhança, o número de *cliques* aumenta rapidamente, e, por consequência, o custo computacional envolvido também aumenta bastante.

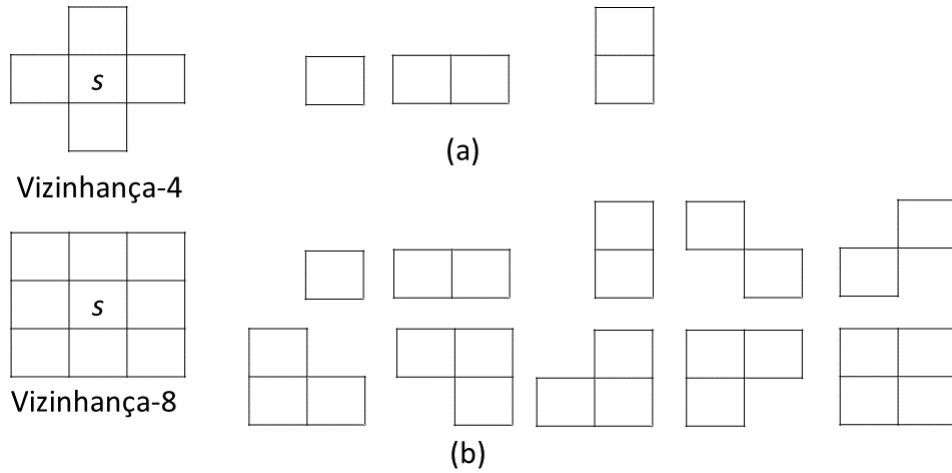


Figura 2 – Dois exemplos de vizinhança em um reticulado bidimensional com seus respectivos *cliques*.

2.3.2 Campos Aleatórios Markovianos

Um campo aleatório X é chamado de Campo Aleatório Markoviano (MRF, do inglês, *Markov Random Field*) em S com respeito a um sistema de vizinhança $\mathcal{N}$ se, e somente se, as seguintes condições forem satisfeitas:

- Markovianidade: $p(x_s | x_{S-\{s\}}) = p(x_s | x_{\mathcal{N}_s})$, $\forall s \in S$, em que $x_{S-\{s\}}$ é o conjunto formado por todos os elementos de S exceto s e $x_{\mathcal{N}_s} = \{x_{s'} | s' \in \mathcal{N}_s\}$.
- Positividade: $p(x) > 0$, $\forall x \in \Omega$, em que Ω é o conjunto de todas as possíveis realizações de X . Em outras palavras, toda a realização do campo aleatório tem uma probabilidade de ocorrer.

Isto é, um MRF definido em um reticulado nada mais é que uma coleção de variáveis aleatórias tal que a probabilidade de uma variável individual, condicionada aos valores

de todas as variáveis pertencentes ao conjunto, é igual à probabilidade dessa variável condicionada a um pequeno subconjunto de elementos, denominados de sistemas de vizinhança (LEVADA; MASCARENHAS; TANNUS, 2008). A forma e a extensão do sistema de vizinhança são aspectos fundamentais na definição e caracterização de um campo aleatório.

Ao assumir dependência entre um elemento e sua vizinhança negamos a suposição de independência estatística desses dados. Em termos probabilísticos, as restrições contextuais podem ser expressas em termos da probabilidade condicional $p(x_s | x_{\mathcal{N}_s})$.

O uso prático de MRF é atribuído ao teorema Hammersley-Clifford (HAMMERSLEY; CLIFFORD, 1971; BESAG, 1974) que estabelece a equivalência entre distribuições de MRF e de Gibbs. A relevância desse teorema se deve ao fato de que a distribuição conjunta, $p(X)$, é usada na maioria das aplicações mas a derivação da distribuição conjunta a partir de distribuições condicionais do MRF é matematicamente complexa. O teorema de equivalência MRFs-Gibbs aponta que a distribuição conjunta do MRF é uma distribuição de Gibbs, assumindo assim, uma forma mais simples.

2.3.2.1 Campos Aleatórios de Gibbs

Um campo aleatório X é dito um Campo Aleatório de Gibbs (GRF, do inglês, *Gibbs Random Field*) em S com respeito à $\mathcal{N}$ se, e somente se, segue a distribuição de Gibbs

$$p(x) = Z^{-1} \times e^{-\frac{1}{T}U(x)} \quad (2.65)$$

em que $Z = \sum_{x \in X} e^{-\frac{1}{T}U(x)}$ é uma constante de normalização, T é uma constante chamada de temperatura, que normalmente assume o valor 1 e $U(x)$ é a função de energia.

A energia

$$U(x) = \sum_c \mathcal{V}_c(x) \quad (2.66)$$

é a soma dos potenciais *clique* $\mathcal{V}_c(x)$ em relação a todos os *cliques* possíveis, conjunto $\mathcal{C}$. O valor de $\mathcal{V}_c(x)$ depende da configuração local do *clique* c . A distribuição Gaussiana é um membro especial dessa família de distribuições de Gibbs.

$p(x)$ mede a probabilidade de ocorrência de uma configuração particular x . A temperatura T controla a forma da distribuição e também pode ser chamada de parâmetro de suavidade. A configuração mais provável é aquela com menos energia. Quanto menor a energia $U(x)$ do sistema numa dada configuração x , maior a probabilidade de observação dessa configuração em um momento qualquer. Quando a temperatura é alta, todas as configurações tendem a ser uniformemente distribuídas. Quando a temperatura se aproxima do zero, a distribuição se concentra em torno da energia global mínima.

Um MRF é caracterizado por sua propriedade local (Markovianidade) enquanto que o GRF é caracterizado por sua propriedade global (a distribuição de Gibbs). O teorema

Hammersley-Clifford (HAMMERSLEY; CLIFFORD, 1971) estabelece a equivalência desses dois tipos de propriedades. O teorema atesta que X é um MRF em S com respeito à $\mathcal{N}$ se, e somente se, X é um GRF em S com respeito à $\mathcal{N}$ (BESAG, 1974). Ou seja,

$$p(x_s | x_{S-\{s\}}) = Z^{-1} \times e^{-\sum_{c \in \mathcal{C}} V_c(x)}. \quad (2.67)$$

em que $\mathcal{C}$ são os *cliques* que contêm s . $p(x_s | x_{S-\{s\}})$ só depende dos potenciais que contêm s , logo, só depende da vizinhança de s .

Embora tenham elaborado o teorema, Hammersley e Clifford (HAMMERSLEY; CLIFFORD, 1971) nunca apresentaram a prova de seu resultado. A prova do teorema foi apresentada pela primeira vez no trabalho de Besag (BESAG, 1974). O valor prático do teorema é que ele fornece uma maneira simples de especificar a probabilidade conjunta. É possível especificar a probabilidade $p(X)$ especificando as funções potenciais $V_c(x)$ e escolhendo funções potenciais adequadas para o comportamento do sistema desejado. Em outras palavras, as formas das funções potenciais determinam a forma da distribuição de Gibbs. Quando todos os parâmetros envolvidos nas funções potenciais são especificados, a distribuição de Gibbs está completamente definida.

2.3.2.2 Campos Aleatórios Markovianos Gaussianos

A abundância de variáveis aleatórias com distribuição normal em aplicações estatísticas motivou o aparecimento dos Campos Aleatórios Markovianos Gaussianos (GMRF, do inglês, *Gaussian Markov Random Fields*). Um GMRF é uma variável aleatória Gaussiana com propriedades Markovianas que são, convenientemente, representadas na matriz de precisão (inversa da matriz de covariância) $\mathbf{Q}$. O elemento $Q_{sm} = 0$ se, e somente se, os sítios s e m não forem vizinhos, ou seja, $Q_{sm} \neq 0 \Leftrightarrow \forall m \in \mathcal{N}_s$. Por exemplo, numa grade regular com n sítios, cada um com 4 vizinhos (ignorando as bordas), teremos uma matriz de dimensão $n \times n$ com uma fração de elementos nulos aproximadamente igual a $(n^2 - 4n)/n^2 = 1 - 4/n$. Se n for grande, isto implica que a quase totalidade da matriz será composta de zeros. Ela será uma matriz esparsa (RUE; MARTINO; CHOPIN, 2009).

Uma variável aleatória $X \in R^n$ é dita um Campo Aleatório Markoviano Gaussiano (GMRF, do inglês, *Gauss-Markov Random Field*) em S com respeito à $\mathcal{N}$, com média μ e matriz de precisão $\mathbf{Q} > 0$ se, e somente se, a distribuição conjunta do vetor é dada por (BESAG, 1974; RUE; MARTINO; CHOPIN, 2009):

$$p(x) = (2\pi)^{-\frac{n}{2}} |\mathbf{Q}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \mathbf{Q} (x - \mu)\right) \quad (2.68)$$

em que $\mathbf{Q}$ tem dimensão $n \times n$ e é simétrica e positivamente definida.

A definição de GMRF implica que qualquer distribuição normal multivariada com matriz de covariância simétrica positivamente definida é um campo de Markov Gaussiano com respeito a um reticulado S e à vizinhança $\mathcal{N}$. Ou seja, usa-se o termo GMRF para designar uma distribuição normal multivariada se a matriz de precisão $\mathbf{Q} = \Sigma^{-1}$ for esparsa.

3 Modelo de Mistura Contextual

O Modelo de Mistura descrito na Seção 2.1.2 utiliza uma abordagem pontual. Nela, os elementos de um conjunto de dados são modelados de forma independente, ou seja, desconsiderando as informações dos elementos vizinhos. Na abordagem contextual, assume-se uma dependência entre os dados e seus vizinhos. A abordagem contextual é especialmente interessante na segmentação de imagens baseada em agrupamento via GMM, pois é possível incorporar no modelo o conhecimento *a priori* de que *pixels* vizinhos possuem maior probabilidade de pertencer ao mesmo grupo do que pertencer a grupos diferentes. Isto é, além dos valores de intensidade, a informação espacial do *pixel* e sua relação com a sua vizinhança também é usada para determinar o grupo ao qual cada *pixel* é associado.

A informação da vizinhança pode ser incorporada ao processo de agrupamento dos dados modelando a correlação espacial diretamente nos pesos das misturas, como é o caso dos Modelos de Mistura Finita Variantes no Espaço (SVFMM) (SANJAY-GOPAL; HEBERT, 1998). Esse modelo permite a estimação da máxima probabilidade *a posteriori* (MAP) dos parâmetros do modelo via o algoritmo *Expectation Maximization* (EM). A dependência contextual dos pesos da mistura é modelada com o uso de Campos Aleatórios Markovianos (MRF) e a distribuição de Gibbs. No entanto, essa abordagem necessita a adição de uma etapa reparatória para garantir que os pesos da mistura sejam positivos e sua soma, para todos os componentes, seja igual a 1 (um). Dentre os artifícios propostos para garantir que os pesos da mistura estejam dentro dos requisitos, pode-se citar o algoritmo de Projeção de Gradiente proposto por Sanjay-Gopal *et al.* em (SANJAY-GOPAL; HEBERT, 1998) ou a Programação Quadrática proposta por Blekas *et al.* (BLEKAS *et al.*, 2005). Ao incorporar essa etapa adicional ao algoritmo EM, os parâmetros de suavidade da distribuição de Gibbs não podem ser calculados diretamente a partir dos dados.

Uma nova maneira de modelar a correlação espacial nos pesos das misturas baseadas em Campos Aleatórios Markovianos Gaussianos (GMRF) foi proposta por Nikou *et al.* (NIKOU; GALATSANOS; LIKAS, 2007). O uso de GMRF possibilita o cálculo do parâmetro de suavidade do GMRF diretamente a partir dos dados. Com essa modificação foi possível definir um parâmetro de suavidade para cada grupo ou, ainda, um parâmetro de suavidade de cada grupo que controlasse o grau de suavidade em diferentes direções espaciais, contribuindo, assim, para a melhor adaptação do modelo ao conjunto de dados que se quer modelar. Sfikas *et al.* (SFIKAS *et al.*, 2010) combinaram o GMRF com processos lineares para propor um SVFMM que garantisse a preservação das bordas das regiões das imagens. Embora o uso do GMRF tenha permitido um avanço na modelagem dos dados, ele manteve a imposição da suavidade nos pesos da mistura sem considerar que

o mesmo é um vetor de probabilidade, logo, deve ser positivo e sua soma, para todos os componentes, deve ser igual a 1 (um).

Focados em eliminar a etapa reparatória necessária para que as propriedades de probabilidade dos pesos da mistura fossem respeitadas, Nikou *et al.* propuseram em (NIKOU; LIKAS; GALATSANOS, 2010) uma nova abordagem do SVFMM, a *Dirichlet Compound Multinomial-based Spatially Variant Finite Mixture Model* (DCM-SVFMM) na qual os pesos da mistura seguem uma distribuição de Dirichlet conjugada da distribuição multinomial (DCM, do inglês, *Dirichlet compound multinomial distribution*). A distribuição DCM é uma distribuição multinomial discreta cujos parâmetros são gerados por uma distribuição de Dirichlet (JOHNSON; KOTZ; BALAKRISHNAN, 1997), assim, os pesos da mistura são vetores de probabilidade. Mais precisamente, a classe a qual um *pixel* pertence é modelada por uma distribuição discreta multinomial cujos parâmetros seguem uma lei de Dirichlet (NIKOU; LIKAS; GALATSANOS, 2010). A correlação espacial dos dados é modelada, então, nos parâmetros da DCM com o uso de GMRFs. Os parâmetros da distribuição multinomial são totalmente integrados ao *framework* Bayesiano e as estimativas MAP dos parâmetros podem ser computadas através do algoritmo EM. Embora a etapa reparatória tenha sido eliminada no DCM-SVFMM, o fato de adotar a distribuição de probabilidade DCM que não é usada no SVFMM padrão aumentou o número de parâmetros do modelo.

Com o mesmo objetivo de Nikou *et al.* em (NIKOU; LIKAS; GALATSANOS, 2010), Nguyen e Wu (NGUYEN; WU, 2013), usaram MRF para modelar a dependência contextual dos pesos da mistura no qual a distribuição de Gibbs adotada possuía uma função de energia dependente tanto das probabilidades *a priori*, representada pelos pesos da mistura, quanto das probabilidades *a posteriori*. A estimativa de Máxima Verossimilhança dos parâmetros foi calculada via EM. No entanto, o parâmetro de suavidade da distribuição de Gibbs foi determinado empiricamente e não calculado automaticamente dos dados.

A ideia de modelar a correlação entre *pixel* vai além do uso de Campos Aleatórios de Gibbs e Campos Aleatórios Markovianos Gaussianos. Um outro tipo de Campo Aleatório Markoviano, o Modelo de Potts, foi usado por Portela *et al.* (PORTELA; CAVALCANTI; TSANG, 2013) para descrever como um elemento individual, no caso, o peso da mistura, modifica seu comportamento conforme o comportamento dos indivíduos na sua vizinhança. A Máxima Pseudo-Verossimilhança foi usada para estimar os parâmetros do modelo uma vez que a Máxima Verossimilhança, no caso do modelo de Potts, é computacionalmente intratável. No trabalho de Nguyen e Wu (NGUYEN; WU, 2011a) os pesos da mistura seguem uma distribuição de Dirichlet. Os parâmetros do modelo proposto não são estimados via MAP, e sim pela aplicação do gradiente para minimizar o negativo da função de verossimilhança.

Além do *framework* estabelecido pelo SVFMM no qual a correlação espacial é

modelada diretamente nos pesos das misturas, existem outras possibilidades de incorporar informação espacial ao Modelo de Mistura. Zhang *et al.* (ZHANG; WU; NGUYEN, 2013) propõem um método que incorpora a informação espacial local aplicando um *template* de médias no cálculo dos componentes da mistura e no cálculo dos pesos da mistura. Como não usa uma distribuição paramétrica, esse modelo possui menos parâmetros pra calcular, no entanto, é necessário definir inicialmente o tamanho da janela adotada no *template*.

Em seguida, são explicados com mais detalhes os principais Modelos de Mistura Finita Variante no Espaço que usam a distribuição Gaussiana como componente da mistura, pois estas são peça fundamental no desenvolvimento deste trabalho. Além disso, fica mais claro como o uso de Campos Aleatórios Markovianos podem ser usados como ferramenta na modelagem da dependência contextual de dados.

3.1 Modelos de Mistura Finita Variante no Espaço

Modelos de Mistura Finita Variante no Espaço (SVFMM, do inglês, *Spatially Variant Finite Mixture Model*) oferecem uma abordagem que facilita a modelagem da correlação espacial diretamente nos pesos das misturas, que passam a ser chamados de pesos da mistura contextuais (SANJAY-GOPAL; HEBERT, 1998). A influência da vizinhança sobre um elemento central é inserida no modelo em forma de probabilidade *a priori*. Há, então, a inclusão do conceito de Campos Aleatórios Markovianos (NIKOU; GALATSANOS; LIKAS, 2007) (SANJAY-GOPAL; HEBERT, 1998). A estimativa dos parâmetros da mistura é calculada via estimação de máxima probabilidade *a posteriori*, estimador MAP, usando o algoritmo EM. Embora a segmentação contextual obtenha um resultado mais robusto, ela pode demandar um maior custo computacional.

Considerando o vetor de parâmetros da mistura ψ uma variável aleatória, a influência mútua entre elementos vizinhos pode ser incorporada no processo de estimação dos parâmetros através da aplicação de funções de densidade *a priori* adequadas $p(\psi) = p(\theta, w)$ que modelem essa correlação local. Dada a densidade *a priori*, a função de densidade *a posteriori* pode ser formada e maximizada através da estimativa MAP dos parâmetros ψ . A forma geral da probabilidade *a posteriori* acumulada considerando todos os n elementos do conjunto é

$$p(\psi|X) = \frac{p(X|\psi)p(\psi)}{p(X)} \quad (3.1)$$

em que $p(X|\psi)$ é a verossimilhança de ψ definida em 2.8. $p(X)$ é constante e não afeta o cálculo então pode ser desconsiderada. O logaritmo da probabilidade *a posteriori* conjunta é dado por

$$\ln p(\psi|X) \propto \sum_{i=1}^n \{\ln p(x_i|\psi) + \ln p(\psi)\}. \quad (3.2)$$

Para derivar o algoritmo EM a fim de encontrar o estimador MAP dos parâmetros ψ utiliza-se também a ideia de dados completos $Y = (X \cup R)$. A esperança do logaritmo da densidade *a posteriori* $p(Y|\psi)p(\psi)$, com respeito às variáveis desconhecidas R , condicionada aos dados observados X e aos valores iniciais definidos dos parâmetros ψ^t , na etapa E do algoritmo EM para o estimador MAP consiste em

$$Q_{MAP}(\psi|\psi^t) = E_R \left\{ \ln p(Y|\psi)|X, \psi^t \right\} + \ln \{p(\psi)\} \quad (3.3)$$

sabendo que $Q(\psi|\psi^t) = E_R \left\{ \ln p(Y|\psi)|X, \psi^t \right\}$, substitui-se $E \left\{ \ln p(Y|\psi)|X, \psi^t \right\}$ pela expressão encontrada em 2.23,

$$Q_{MAP}(\psi|\psi^t) = \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \left\{ \ln w_{ij} + \ln p(x_i|\theta_j) \right\} + \ln \{p(\psi)\}. \quad (3.4)$$

O peso da mistura contextual w_{ij} corresponde à probabilidade *a priori* do i -ésimo *pixel* pertencer ao j -ésimo grupo e deve obedecer às restrições $0 \leq w_{ij} \leq 1$ e $\sum_{j=1}^k w_{ij} = 1 \forall i$.

Assumindo que os parâmetros w e θ são modelados de maneira independente e especificando as respectivas funções de densidade *a priori* como $p(w)$ e $p(\theta)$. A expressão de Q_{MAP} assume a seguinte forma

$$Q_{MAP}(\psi|\psi^t) = \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \left\{ \ln w_{ij} + \ln p(x_i|\theta_j) \right\} + \ln \{p(w)p(\theta)\}. \quad (3.5)$$

Considerando $p(\theta)$ uniformemente distribuída no espaço paramétrico Ψ , a correlação local entre um elemento do conjunto de dados e sua vizinhança pode ser modelada por um Campo Aleatório de Gibbs ou um Campo Aleatório Markoviano Gaussiano.

3.1.1 Campo Aleatório de Gibbs

A função de densidade acumulada de Gibbs (SANJAY-GOPAL; HEBERT, 1998; LI, 2009) é uma maneira de incorporar correlação local em um modelo para os parâmetros $W = \{w_1, \dots, w_n\}$ dada por

$$p(W) = \frac{1}{Z} \exp(-\beta U(W)) \text{ em que } U(W) = \sum_{i=1}^n \mathcal{V}_i(W) \quad (3.6)$$

sendo Z e β constantes. $U(W)$ é a função de energia, β é o inverso da temperatura, T , controla o grau de suavidade imposto ao modelo e influencia na continuidade espacial dos grupos, quanto menor for β , maior a influência da vizinhança sobre o elemento central analisado. A função densidade acumulada deve conferir maior probabilidade para o conjunto de pesos dentro da vizinhança de w_i . A função $\mathcal{V}_i(\cdot)$ denota o potencial *clique*, em relação à vizinhança V_i do peso w_i correspondente ao i -ésimo *pixel* e pode ser calculada pela seguinte expressão

$$\mathcal{V}_i(W) = \sum_{v \in N_i} g(d(i, v)) \quad (3.7)$$

em que $d(i, v)$ especifica a distância ao quadrado entre os pesos w_i e w_v , *i.e.*, $d(i, v) = |w_i - w_v|^2 = \sum_{c=1}^k (w_i^c - w_v^c)^2$. Finalmente, a função $g(u)$ deve ser não negativa e monotonicamente crescente. Blekas *et al.* (BLEKAS et al., 2005) sugeriu a expressão $g(u) = (1 + u^{-1})^{-1}$, em vez da adotada por Sarjay-Gopal, $g(u) = u$, uma vez que a nova expressão é mais robusta a *outliers*.

Assim, na etapa E do cálculo do estimador MAP dos parâmetros ψ , a expressão de Q_{MAP} assume a seguinte forma

$$\begin{aligned} Q_{MAP}(\psi|\psi^t) &= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{\ln w_{ij} + \ln p(x_i|\theta_j)\} + \ln \left\{ \frac{1}{Z} \exp \left(-\beta \sum_{i=1}^n \mathcal{V}_{V_i}(W) \right) p(\theta) \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{\ln w_{ij} + \ln p(x_i|\theta_j)\} - \beta \sum_{i=1}^n \sum_{v \in V_i} g(d(i, v)) - \ln Z + \ln \{p(\theta)\}. \end{aligned} \quad (3.8)$$

Como $\ln Z$ e $\ln p(\theta)$ não vão influenciar na etapa subsequente, na etapa M, podem ser eliminados do Q_{MAP} que, finalmente, assume a forma

$$Q_{MAP}(\psi|\psi^t) = \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{\ln w_{ij} + \ln p(x_i|\theta_j)\} - \beta \sum_{i=1}^n \sum_{v \in V_i} g(d(i, v)). \quad (3.9)$$

Na etapa M, $Q_{MAP}(\psi|\psi^t)$ pode ser maximizado independentemente em relação a cada parâmetro. Assumindo que os componentes da mistura são distribuições Gaussianas com $\theta_j = \{\mu_j, \Sigma_j\}$, as equações de atualização dos parâmetros são dadas por

$$\mu_j^* = \frac{\sum_{i=1}^n x_i z_j(\psi^t|x_i)}{\sum_{i=1}^n z_j(\psi^t|x_i)} \quad (3.10)$$

e

$$\Sigma_j^* = \frac{\sum_{i=1}^n (x_i - \mu_j^{t+1})(x_i - \mu_j^{t+1})^T z_j(\psi^t|x_i)}{n w_j^{t+1}}. \quad (3.11)$$

Derivando $Q_{MAP}(\psi|\psi^t)$ em relação à w_{ij} tem-se a equação de segundo grau

$$\begin{aligned} \nabla_{w_{ij}} Q_{MAP}(\psi|\psi^t) &= 4\beta \left[\sum_{v \in V_i} g'(d(i, v)) \right] (w_{ij}^{(t+1)})^2 - 4\beta \left[\sum_{v \in V_i} g'(d(i, v)) w_{vj} \right] (w_{ij}^{(t+1)}) \\ &\quad - z_j(\psi^t|x_i) \end{aligned} \quad (3.12)$$

cujas raízes são

$$w_{ij}^* = \frac{\sum_{v \in V_i} g'(d(i, v)) w_{vj} \pm \sqrt{\left[\sum_{v \in V_i} g'(d(i, v)) w_{vj} \right]^2 + \frac{1}{\beta} z_j(\psi^t|x_i) \left[\sum_{v \in V_i} g'(d(i, v)) \right]}}{2 \left[\sum_{v \in V_i} g'(d(i, v)) \right]}, \quad (3.13)$$

com g' sendo a derivada de g com respeito a w_{ij} . w_{ij}^* é a raiz da equação com valor não negativo uma vez que $w \geq 0$. Deve-se notar que a vizinhança, V_i , na equação acima pode incluir *pixels* com valores atualizados de parâmetros, w_{ij}^* , assim como *pixels* cujos parâmetros ainda não foram atualizados.

Uma das desvantagens dessa proposta é uma complexidade adicional na etapa M. Para garantir que w_{ij}^* respeite as restrições $0 \leq w_{ij} \leq 1$ e $\sum_{j=1}^k w_{ij} = 1 \forall i$, uma etapa reparatória deve ser adicionada aumentando a complexidade do algoritmo. Dentre os artifícios propostos para garantir que as probabilidades *a priori* atinjam os requisitos, pode-se citar o algoritmo de Projeção de Gradiente proposto por Sanjay-Gopal *et al.* (SANJAY-GOPAL; HEBERT, 1998) ou a Programação Quadrática proposta por Blekas *et al.* (BLEKAS *et al.*, 2005). Outro problema encontrado nessa proposta é que o parâmetro β na Equação 3.9 é o mesmo para todos os grupos e em todas as direções. Além disso, β não pode ser estimado diretamente a partir dos dados.

3.1.2 Campo Aleatório Markoviano Gaussiano

A correlação local também pode ser inserida no modelo com o uso de Campos Aleatórios Markovianos Gaussianos (NIKOU; GALATSANOS; LIKAS, 2007; BESAG, 1974; RUE; MARTINO; CHOPIN, 2009). A densidade *a priori* acumulada $p(W)$, com $W = \{w_1 \dots, w_n\}$, segue a seguinte expressão:

$$p(W) \propto \prod_{j=1}^k \beta_j^{-n} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^n \left(\sum_{v \in V_i} (w_{ij} - w_{vj})^2 \right)}{\beta_j^2} \right] \quad (3.14)$$

em que o parâmetro β_j controla o grau de suavidade imposto ao modelo em relação ao j -ésimo grupo. Valores diferentes de β para grupos diferentes contribui para a melhor adaptação do modelo ao conjunto de dados que se quer modelar. V_i é o sistema de vizinhança adotado em relação ao i -ésimo *pixel*. As suposições estatísticas usadas na definição de $p(W)$ ficam claras ao assumir que a diferença local entre os pesos contextuais da mistura segue uma distribuição Gaussiana:

$$w_{ij} - w_{iv} \sim N(0, \beta^2) \quad \forall i, j, \quad \forall v \in V_i. \quad (3.15)$$

Em outras palavras, a diferença local entre os pesos contextuais da mistura, para todas as posições espaciais e para todos os grupos são variáveis aleatórias Gaussianas independentes e identicamente distribuídas com média zero e variância β^2 .

Na etapa E do cálculo do estimador MAP dos parâmetros ψ , a expressão de Q_{MAP}

assume a seguinte forma

$$\begin{aligned}
Q_{MAP}(\psi|\psi^t) &= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{ \ln w_{ij} + \ln p(x_i|\theta_j) \} \\
&\quad + \ln \left\{ \prod_{j=1}^k \beta_j^{-n} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^n \sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta_j^2} \right] \right\} \\
&= \sum_{i=1}^n \sum_{j=1}^k z_j(\psi^t|x_i) \{ \ln w_{ij} + \ln p(x_i|\theta_j) \} \\
&\quad + \sum_{j=1}^k \left\{ \ln \beta_j^{-n} - \frac{1}{2} \frac{\sum_{i=1}^n \sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta_j^2} \right\}.
\end{aligned} \tag{3.16}$$

A função $Q_{MAP}(\psi|\psi^t)$ pode ser maximizada independentemente em relação a cada parâmetro proporcionando as seguintes equações de atualização (NIKOU; GALATSANOS; LIKAS, 2007):

$$\mu_j^* = \frac{\sum_{i=1}^n x_i z_j(\psi^t|x_i)}{\sum_{i=1}^n z_j(\psi^t|x_i)} \tag{3.17}$$

e

$$\Sigma_j^* = \frac{\sum_{i=1}^n (x_i - \mu_j^{t+1}) (x_i - \mu_j^{t+1})^T z_j(\psi^t|x_i)}{n w_j^{t+1}}. \tag{3.18}$$

O termo *a priori* baseado em Campos Aleatórios Markovianos Gaussianos permite que seus parâmetros sejam calculados em forma fechada¹ via estimação MAP usando o algoritmo EM. O cálculo do parâmetro β_j de cada grupo pode ser calculado fazendo $\nabla_{\beta_j} Q(\psi|\psi^t) = 0$ que resulta na seguinte equação de atualização:

$$\beta_j^* = \frac{1}{n} \sum_{i=1}^n \left(\sum_{v \in V_i} (w_{ij} - w_{vj}) \right)^2. \tag{3.19}$$

Nikou *et al.* (NIKOU; GALATSANOS; LIKAS, 2007) ainda propõem uma segunda versão do algoritmo no qual o parâmetro β_j de cada grupo ainda controla o grau de suavidade em diferentes direções.

Enfim, as probabilidades w_{ij} são calculadas fazendo $\nabla_{w_{ij}} Q(\psi|\psi^t) = 0$ o que resulta na equação de segundo grau com respeito à w_{ij} (NIKOU; GALATSANOS; LIKAS, 2007):

$$|V_i|(w_{ij})^2 - \left[\sum_{v \in V_i} w_{vj} \right] (w_{ij}) - z_j(\psi^t|x_i) \beta_j^2 = 0 \tag{3.20}$$

em que $|V_i|$ é o número de *pixels* na vizinhança do *i*-ésimo *pixel*.

¹ Em matemática, uma expressão é dita ser uma expressão de forma fechada se, e somente se, pode ser expressa analiticamente em termos de um número delimitado de certas funções bem conhecidas. Tipicamente, estas bem conhecidas funções são definidas como funções elementares - constantes, uma variável x , operações elementares de aritmética ($+$ $-$ $\times$ $\div$), raízes n -ésimas, exponenciais e logaritmos (os quais também incluem funções trigonométricas e funções trigonométricas inversas).

A Equação 3.20 possui duas soluções reais não negativas. Entretanto, assim como na proposta anterior, o inconveniente deste método é a imposição da suavidade em w_{ij} sem considerar que o mesmo é um vetor de probabilidade ($0 \leq w_{ij} \leq 1$ e $\sum_{j=1}^k w_{ij} = 1 \forall i$). Para esse propósito, artifícios computacionais foram introduzidos na etapa M para forçar que as variáveis w_{ij} satisfaçam tais restrições (SANJAY-GOPAL; HEBERT, 1998; BLEKAS et al., 2005).

Uma vantagem do uso de Campos Aleatórios Markovianos Gaussianos em relação à função de densidade acumulada de Gibbs para modelar a relação espacial dos *pixels* da imagem é a possibilidade de calcular o parâmetro de suavidade do modelo, β , em forma fechada. Além disso, é possível modificar o modelo de maneira que exista um parâmetro de suavidade para cada grupo. Valores diferentes de β para grupos diferentes podem contribuir para a melhor adaptação do modelo ao conjunto de dados que se quer modelar.

3.1.3 Dirichlet conjugada da distribuição multinomial

No trabalho de Nikou *et al.* (NIKOU; LIKAS; GALATSANOS, 2010), assume-se que os pesos da mistura seguem uma distribuição de Dirichlet conjugada da distribuição multinomial (DCM - *Dirichlet compound multinomial distribution*). A distribuição DCM é um multinomial cujos parâmetros são gerados por uma distribuição de Dirichlet (JOHNSON; KOTZ; BALAKRISHNAN, 1997), assim w_{ij} são vetores de probabilidade. Mais precisamente, a classe a qual um *pixel* pertence é modelada por uma distribuição discreta multinomial cujos parâmetros seguem uma lei de Dirichlet (NIKOU; LIKAS; GALATSANOS, 2010). Além disso, a suavidade espacial é imposta aos parâmetros do Dirichlet assumindo GMRF.

Dado o i -ésimo *pixel*, considera-se que a probabilidade *a posteriori* de x_i , $z(\psi|x_i)$, é uma variável aleatória que segue uma distribuição multinomial com vetor de probabilidade $\xi^i = (\xi_1^i, \xi_2^i, \dots, \xi_k^i)$ sendo k o número de componentes da mistura/grupos na imagem. Sendo $\Xi = \{\xi^1, \xi^2, \dots, \xi^n\}$ o conjunto de parâmetros para toda imagem e adotando a notação z_j^i para a probabilidade *a posteriori* $z_j(\psi|x_i)$, a definição de multinomial afirma que:

$$p(z^i|\xi^i) = \frac{M!}{\prod_{j=1}^k (z_j^i)!} \prod_{j=1}^k (\xi_j^i)^{z_j^i}, \quad i = 1, 2, \dots, n \quad (3.21)$$

com $\xi_j^i \geq 0$ e $\sum_{j=1}^k \xi_j^i = 1 \forall i$. O modelo descrito na equação acima representa a probabilidade ξ_j^i que o pixel i pertença ao componente da mistura j como um dos k possíveis resultados de um processo multinomial com M realizações/repetições. Cada um dos k resultados do processo aparece com probabilidade ξ_j^i , $j = 1, \dots, k$. Genericamente falando, esse é um modelo gerador (do inglês, *generative model*) da imagem.

A distribuição DCM assume que os parâmetros Ξ da multinomial seguem uma distribuição Dirichlet parametrizado por $A = (\alpha^1, \alpha^2, \dots, \alpha^n)$ em que $\alpha^i = (\alpha_1^i, \alpha_2^i, \dots, \alpha_k^i)$,

$i=1,2,\dots,n$, é o vetor de parâmetros de Dirichlet para ξ_j^i :

$$P(\xi_j^i|\alpha^i) = \frac{\Gamma\left(\sum_{j=1}^k \alpha_j^i\right)}{\prod_{j=1}^k \Gamma(\alpha_j^i)} \prod_{i=1}^k \left(\xi_j^i\right)^{\alpha_j^i-1} \quad (3.22)$$

com $\alpha_j^i \geq 0$, $i = 1, \dots, n$, $j = 1, \dots, k$, e $\Gamma(\cdot)$ sendo a função Gama.

No modelo proposto por Nikou *et al.* (NIKOU; LIKAS; GALATSANOS, 2010), a probabilidade de se obter z^i do i -ésimo pixel é obtido pela marginalização dos parâmetros Ξ :

$$p(z^i|\alpha^i) = \int_0^1 p(z^i|\xi^i) p(\xi^i|\alpha^i) d\xi^i, \quad i = 1, \dots, n. \quad (3.23)$$

Substituindo 3.21 e 3.22 em 3.23, após algumas manipulações, obtém-se a seguinte equação para z^i (NIKOU; LIKAS; GALATSANOS, 2010):

$$p(z^i|\alpha^i) = \frac{M!}{\prod_{j=1}^k (z_j^i)!} \frac{\Gamma\left(\prod_{j=1}^k \alpha_j^i\right)}{\Gamma\left(\sum_{j=1}^k \alpha_j^i + z_j^i\right)} \prod_{i=1}^k \frac{\Gamma(\alpha_j^i + z_j^i)}{\Gamma(\alpha_j^i)}, \quad i = 1, \dots, n. \quad (3.24)$$

Supondo um modelo gerador da imagem no qual a geração do *pixel* i pelo componente j é modelado por um processo DCM com uma realização/repetição ($M=1$). Consequentemente, o vetor $z^i = (z_1^i, z_2^i, \dots, z_k^i)$ possui o j -ésimo componente igual a 1 e todos os outros iguais a 0. Isto é também ilustrado nos pesos contextuais da mistura que, nesse caso, são probabilidades *a posteriori*:

$$\begin{aligned} w_{ij} &= p(z_j^i = 1|x^i) = 1 \\ w_{iv} &= p(z_v^i = 1|x^i) = 0, \quad v = 1, \dots, k, v \neq j. \end{aligned} \quad (3.25)$$

Assim, considerando que o processo DCM só tem uma realização/repetição ($M=1$) e que $\Gamma(x+1) = x\Gamma(x)$,

$$w_{ij} = p(z_j^i = 1|x^i) = \frac{\alpha_j^i}{\sum_{l=1}^k \alpha_l^i}, \quad j = 1, 2, \dots, k. \quad (3.26)$$

O novo modelo vai se tornar variante no espaço ao se introduzir o conhecimento espacial *a priori* nos parâmetros A da distribuição de Dirichlet. Mais especificamente, o conhecimento *a priori* assume a forma de GMRF uma vez que os seus parâmetros podem ser estimados em forma fechada (NIKOU; GALATSANOS; LIKAS, 2007):

$$p(A) \propto \prod_{j=1}^k \beta_j^{-n} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^n \left(\sum_{v \in V_i} (\alpha_j^i - \alpha_j^v) \right)^2}{\beta_j^2} \right] \quad (3.27)$$

A principal característica desse modelo é o fato dele impor diferentes graus de suavidade, β , para diferentes classes contribuindo para a melhor adaptação do modelo ao conjunto de dados que se quer modelar.

Nikou *et al.* (NIKOU; LIKAS; GALATSANOS, 2010) nomeou o modelo proposto como *Dirichlet Compound Multinomial-based Spatially Variant Finite Mixture Model* (DCM-SVFMM). De maneira geral, o DCM-SVFMM funciona da seguinte maneira: uma amostra ξ (vector de probabilidade) é inicialmente gerada a partir de uma distribuição Dirichlet com parâmetro α , obtendo-se assim uma distribuição multinomial com parâmetro ξ . A variável z , denotando a probabilidade *a posteriori* observação x pertencer a uma classe, é o resultado de um processo multinomial parametrizado por ξ . Além disso, os parâmetros α da distribuição Dirichlet possuem uma restrição espacialmente modelados como GMRF, assim como no SVFMM padrão.

Uma vez definido o modelo DCM-SVFMM, na etapa E do cálculo do estimador MAP dos parâmetros $\psi = \{\theta, \pi\}$ e A , a expressão de Q_{MAP} assume a seguinte forma

$$Q_{MAP}(\psi, A|\psi^t, A^t) = \sum_{i=1}^n \sum_{j=1}^k z_j^i \left\{ \ln \frac{\alpha_j^i}{A^i} + \ln p(x^i|\theta_j) \right\} + \ln \left\{ \prod_{j=1}^k \beta_j^{-n} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^n \sum_{v \in V_i} (\alpha_j^i - \alpha_j^v)^2}{\beta_j^2} \right] \right\} \quad (3.28)$$

no qual, $A^i = \sum_{j=1}^k \alpha_j^i$.

Para estimar o parâmetro $\alpha_j^{i(t+1)}$, a equação de Q_{MAP} acima deve ser derivada com respeito à α_j^i . Fazendo o gradiente $\nabla_{\alpha_j^i} Q(\psi, A|\psi^t, A^t) = 0$, resulta na equação de terceiro grau com respeito à α_j^i

$$(\alpha_j^i)^3 + \left[A_{-j}^i - \frac{\sum_{v \in V_i} \alpha_j^v}{|V_i|} \right] (\alpha_j^i)^2 + \left[\frac{A_{-j}^i \sum_{v \in V_i} \alpha_j^v}{|V_i|} \right] (\alpha_j^i) - \frac{z_j^i A_{-j}^i \beta_j^2}{2|V_i|} = 0 \quad (3.29)$$

com $A_{-j}^i = \sum_{v=1}^k \alpha_v^i$, $v \neq j$, para $i = 1, \dots, n$ e $j = 1, \dots, k$. Se baseando na teoria polinomial (BIRKHOFF; LANE, 1941), a Equação 3.29 só possui uma solução real não negativa que satisfaça a restrição $\alpha_j^i \geq 0$ e esta solução tem forma fechada.

A solução para a variância das classes é obtida fazendo o gradiente $\nabla_{\beta_j^2} Q(\psi, A|\psi^t, A^t) = 0$, resultando na equação:

$$\beta_j^{2(t+1)} = \frac{1}{n} \sum_{i=1}^n \left(\sum_{v \in V_i} (\alpha_j^i - \alpha_j^v) \right)^2, \quad j = 1, \dots, k. \quad (3.30)$$

Existem duas vantagens nesse método. A primeira é que ele modela explicitamente os pesos da mistura como vetores de probabilidade, garantindo que $0 \leq w_{ij} \leq 1$ e $\sum_{j=1}^k w_{ij} = 1 \forall i$, e, simultaneamente, impõe suavidade espacial ao modelo. A segunda é que o cálculo da estimativa MAP via EM dos parâmetros resulta em equações de atualização de forma fechada descartando a necessidade de uma etapa reparadora (NGUYEN; WU, 2011a; NIKOU; LIKAS; GALATSANOS, 2010). Embora a etapa reparatória tenha sido eliminada no DCM-SVFMM, o fato de adotar a distribuição de probabilidade DCM que não é usada no SVFMM padrão aumentou o número de parâmetros do modelo.

4 Modelo de Mistura de Gaussianas *Fuzzy* Contextual (CFGMM)

O *Fuzzy C-Means* (FCM), o Modelo de Mistura de Gaussianas (GMM) e o Modelo de Mistura de Gaussianas *Fuzzy* (FGMM) apresentados no Capítulo 2 são técnicas de agrupamento nas quais assume-se que os dados são independentes, ou seja, realizam o agrupamento desconsiderando as informações dos elementos vizinhos. Ao contrário da abordagem pontual, na abordagem contextual, assume-se uma dependência entre os dados e seus vizinhos. A abordagem contextual é especialmente interessante em situações nas quais existe uma dependência entre os dados e seus vizinhos, como, por exemplo, na segmentação de imagens (Seção 2.2).

A desvantagem da abordagem pontual para a segmentação de imagens é que a localização dos *pixels* não é levada em consideração no processo de agrupamento, ou seja, não é utilizado o conhecimento prévio de que *pixels* adjacentes têm maior probabilidade de pertencer ao mesmo grupo do que pertencer a grupos distintos. Para ilustrar essa afirmação, fizemos um teste usando três tipos de imagens segmentadas, uma artificial, uma natural e uma de ressonância magnética (MR), exibidas na Figura 3.



Figura 3 – Imagem artificial com ruído, imagem natural e imagem de ressonância magnética (acima) e suas respectivas segmentações baseadas em região (abaixo).

O gráfico na Figura 4 ilustra o resultado esperado de que a probabilidade de *pixels* vizinhos pertencerem a mesma região é maior do que a probabilidade de *pixels* vizinhos pertencerem à regiões diferentes, para cada uma das três imagens. Quanto maior a quantidade de bordas na imagem, menor a probabilidade de *pixels* vizinhos pertençam

ao mesmo grupo. A vizinhança considerada é vizinhança-4 composta pelos 4 *pixels* que são adjacentes, na horizontal e na vertical, ao *pixel* que está sendo analisado. O eixo vertical do gráfico representa uma probabilidade, e como tal varia de 0 a 1. Entretanto, para uma melhor visualização do resultado, a figura só exibe um corte no eixo que varia de 0,88 a 1,00.

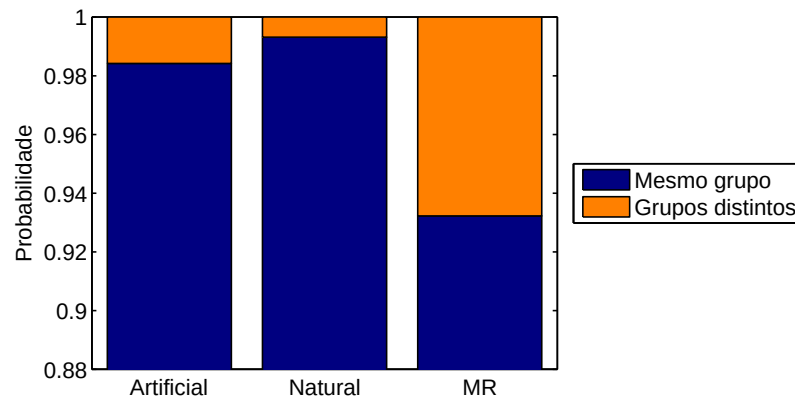


Figura 4 – Probabilidade de *pixels* vizinhos pertencerem ao mesmo grupo *versus* a probabilidade de *pixels* vizinhos pertencerem a grupos diferentes para as imagens segmentadas: imagem artificial com ruído, imagem natural e imagem de ressonância magnética.

Na análise de imagens, o contexto ocupa um importante papel na interpretação de uma cena. *Pixels* isolados fornecem menos informações do que os mesmos *pixels* juntos dando informação de contexto espacial, que é obtido considerando *pixels* adjacentes em uma vizinhança espacial. Na abordagem pontual da análise de imagens, assume-se que não existe relação entre os *pixels* da imagem. Na abordagem contextual, analisar cada *pixels* separadamente não tem sentido. A classe a qual cada *pixel* é associado depende (a) do seu próprio valor, (b) do valor dos outros *pixels* e (c) na relação existente entre as classes. Os gráficos da Figura 5 ilustram a dependência entre o valor de um *pixel* e dos seus vizinhos. Para cada imagem da Figura 3 foi gerado um gráfico cuja abscissa corresponde a intensidade dos *pixels* das imagens, convertidas para tons de cinza, e a ordenada corresponde ao valor da média de intensidade dos *pixels* vizinhos, considerando a vizinhança-4. Em todos os gráficos fica claro a correlação positiva entre as duas grandezas medidas, especialmente no caso da imagem natural.

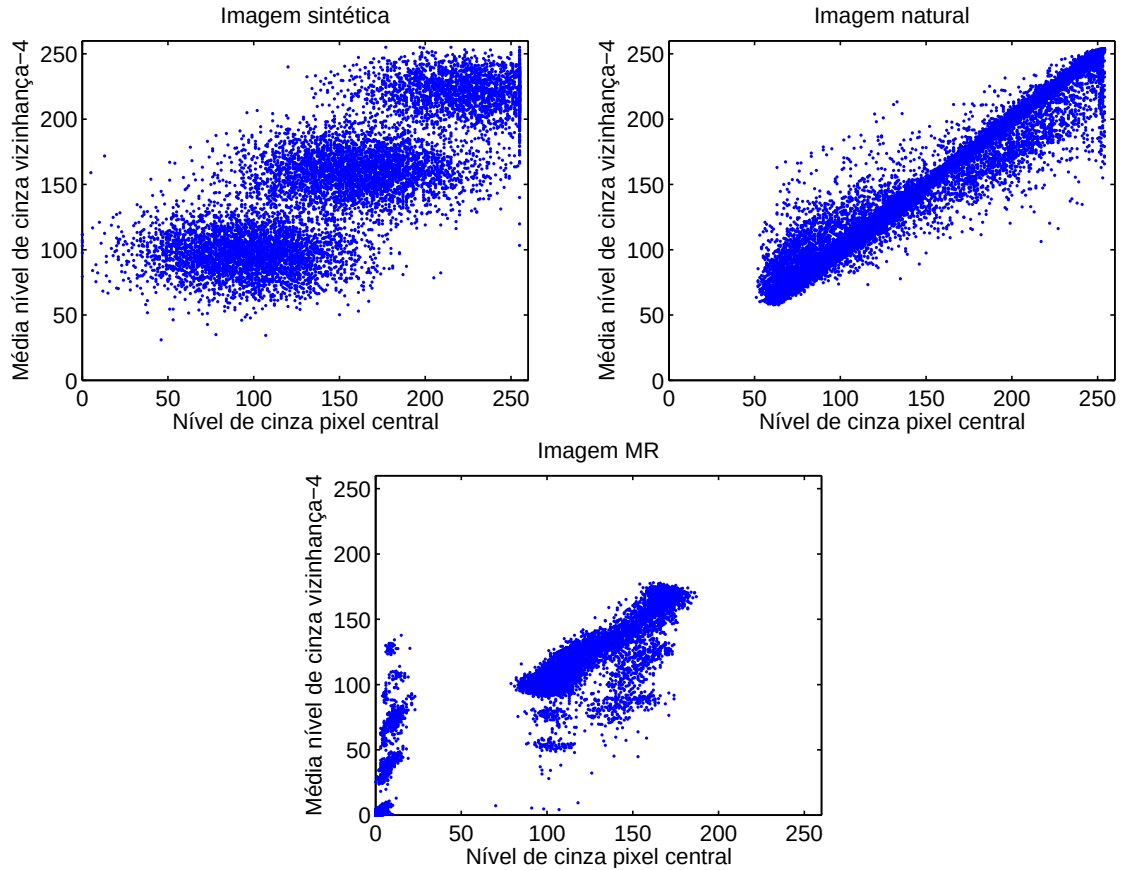


Figura 5 – Nível de cinza do *pixel* versus a média no nível de cinza dos *pixels* vizinhos para as imagens: imagem artificial com ruído, imagem natural e imagem de ressonância magnética.

O Modelo de Mistura de Gaussianas *Fuzzy* Contextual (CFGMM) aqui proposto se inspira no Modelo de Mistura de Gaussianas *Fuzzy* (FGMM) descrito no Capítulo 2 e no Modelo de Mistura Finita Variantes no Espaço (SVFMM) descrito no Capítulo 3.

O CFGMM, assim como o FGMM, é um algoritmo não hierárquico que realiza o agrupamento de dados buscando a minimização de uma função objetivo através de um processo de otimização iterativa baseada no gradiente mais simples do que a combinação *Expectation Maximization* (EM) com máxima probabilidade *a posteriori* (MAP) adotada no SVFMM. A derivação das equações dos parâmetros do modelo via aplicação de gradiente adotada pelo FGMM já é, por si só, matematicamente mais simples do que a abordagem EM-MAP adotada no SVFMM. Além disso, essa simplificação matemática permite que a restrição imposta aos pesos da mistura (que devem ser positivos e sua soma, para todos os componentes, deve ser 1) seja considerado no processo de derivação das equações com o uso de multiplicadores de Lagrange. Inserir a restrição dos pesos da mistura no cálculo da estimativa dos parâmetros eliminaria a necessidade da etapa adicional adotada por diversos algoritmos baseados em SVFMM, simplificando assim, o processo de otimização.

Assim como SVFMM, o CFGMM pode ser usado no agrupamento com restrição

espacial, no qual assume-se uma dependência espacial entre o *pixel* e seus vizinhos, respeitando, assim, o conhecimento *a priori* de que *pixels* vizinhos provavelmente irão pertencer ao mesmo grupo. Entretanto, como o CFGMM tem como base o FCM, é esperado que sua convergência seja mais rápida do que a do SVFMM que se baseia no GMM (JU; LIU, 2012).

Inspirado no SVFMM, o novo método vai inserir a informação de contexto no peso da mistura, w , da função de dissimilaridade descrita na Equação 2.45. No modelo contextual, é introduzida uma distribuição *a priori* $p(\cdot)$ que é dependente do grupo mas também é dependente da posição do *pixel*. Essa distribuição tem como objetivo incorporar a correlação local entre os *pixels* no cálculo da estimativa dos parâmetros. $p(w)$ considera a informação espacial baseada na probabilidade *a priori* com o uso de Campos Aleatórios Markovianos Gaussianos (GMRF) (NIKOU; GALATSANOS; LIKAS, 2007). A teoria de MRF fornece uma maneira conveniente e consistente de modelar entidades com dependência contextual assim como os *pixels* em uma imagem. O uso do GMRF em especial, facilita que o parâmetro que controla o grau de suavidade imposto ao modelo em relação aos dados seja derivado diretamente da função objetivo.

4.1 Definição do modelo proposto

Sendo $X = \{x_1, \dots, x_n\}$ um conjunto de dados p -dimensionais observados composto de n elementos, o objetivo do CFGMM é fornecer uma partição de X em k grupos, com $2 \leq k \leq n$, buscando a minimização da função objetivo (BEZDEK, 1981):

$$J(U, \mu|X) = \sum_{i=1}^n \sum_{j=1}^k (u_{ij}^m d_{ij}^2) \quad (4.1)$$

em que u_{ij} é dado por

$$u_{ij} = \left[\sum_{c=1}^k \left(\frac{d_{ij}}{d_{ic}} \right)^{2/(m-1)} \right]^{-1} \quad (4.2)$$

e representa o grau de pertinência do elemento x_i ao grupo j e é chamado de função de pertinência *fuzzy* (ou grau de pertinência), que deve satisfazer os critérios: $0 \leq u_{ij} \leq 1$ e $\sum_{j=1}^k u_{ij} = 1$, para $1 \leq j \leq k$ e $1 \leq i \leq n$. d_{ij} é a função de dissimilaridade, m é o expoente *fuzzy* de ponderação de cada função de pertinência, com $m > 1$, e $U = \{u_{ij}\}$ é a partição *fuzzy* de X .

No CFGMM proposto, a medida de dissimilaridade do FGMM (TRAN; WAGNER, 1998) foi adaptada através do uso da distribuição *a priori* do peso da mistura *fuzzy* com o objetivo de incorporar a correlação local entre os *pixels* no cálculo da estimativa dos parâmetros. $p(\cdot)$ adotado é dependente do grupo e da posição do *pixel*. Assim, a função de dissimilaridade do FGMM passa a assumir a seguinte forma no CFGMM:

$$d_{ij}^2 = -\log[p(w_{ij})N(x_i|\mu_j, \Sigma_j)] \quad (4.3)$$

em que $N(x_i|\mu_j, \Sigma_j)$ é uma distribuição Gaussiana p -variada com vetor de médias μ_j e matriz de covariância Σ_j . w_{ij} é probabilidade *a priori* do i -ésimo *pixel* pertencer ao j -ésimo grupo e deve obedecer as restrições $0 \leq w_{ij} \leq 1$ e $\sum_{j=1}^k w_{ij} = 1 \forall i$. w_{ij} é chamado de peso contextual da mistura e difere do peso da mistura do GMM por depender tanto do grupo quanto do *pixel*.

A densidade *a priori* $p(w_{ij})$ (NIKOU; GALATSANOS; LIKAS, 2007) é dada por

$$p(w_{ij}) \propto \beta^{-1} \exp \left[-\frac{1}{2} \frac{\left(\sum_{v \in V_i} (w_{ij} - w_{vj}) \right)^2}{\beta^2} \right]. \quad (4.4)$$

O parâmetro β controla o grau de suavidade imposto ao modelo e V_i denota a vizinhança do i -ésimo *pixel*. Visando a obtenção de um modelo matemático tratável na definição de $p(w_{ij})$ foi suposto que a diferença local entre os pesos contextuais da mistura segue uma distribuição Gaussiana (NIKOU; GALATSANOS; LIKAS, 2007):

$$w_{ij} - w_{iv} \sim N(0, \beta^2) \quad \forall i, j, \quad \forall v \in V_i. \quad (4.5)$$

Em outras palavras, a diferença local entre os pesos contextuais da mistura, para todas as posições espaciais e para todos os grupos são variáveis aleatórias Gaussianas independentes e identicamente distribuídas com média zero e variância β^2 .

A densidade *a priori* conjunta $p(W)$ para o conjunto de pesos da mistura, segue, então, a seguinte expressão:

$$p(W) \propto \beta^{-nk} \exp \left[-\frac{1}{2} \frac{\sum_{i=1}^n \sum_{j=1}^k \sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta^2} \right] \quad (4.6)$$

Assim como outras técnicas de agrupamento, o processo de agrupamento por CFGMM busca a minimização de uma função objetivo $J(U, \psi|X)$, sendo $\psi = \{w, \mu, \Sigma, \beta\}$ o conjunto de parâmetros do modelo. Substituindo a nova função de dissimilaridade da Equação 4.3 na função objetivo da Equação 4.1, tem-se a expressão matemática da função objetivo do CFGMM:

$$J(U, \psi|X) = - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log p(w_{ij}) - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log N(x_i|\mu_j, \Sigma_j). \quad (4.7)$$

Substituindo 4.6 em 4.7, tem-se

$$J(U, \psi|X) = - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log \left[\frac{1}{\beta} \exp \left(-\frac{1}{2} \frac{\sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta^2} \right) \right] - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log N(x_i|\mu_j, \Sigma_j). \quad (4.8)$$

A minimização de J consiste na busca das estimativas dos parâmetros ótimos $\psi_j^* = \{w_{ij}^*, \mu_j^*, \Sigma_j^*, \beta^*\}$ que minimizam J . As estimativas são obtidas pela resolução do

sistema de equações $\nabla_{\psi} J(U, \psi^* | X) = 0$ com o gradiente como:

$$\nabla_{\psi} = \begin{bmatrix} \frac{\partial}{\partial w} \\ \frac{\partial}{\partial \mu} \\ \frac{\partial}{\partial \Sigma} \\ \frac{\partial}{\partial \beta} \end{bmatrix}. \quad (4.9)$$

Para resolver a equação $\nabla_{\psi} J(U, \psi^* | X) = 0$ é possível minimizar o termo que contém w e β e o termo que contém μ e Σ separadamente uma vez que eles são independentes. O segundo termo do lado direito da Equação 4.8 não contém w_{ij} nem β logo μ_j^* e Σ_j^* podem ser estimados da mesma forma que no FGMM como descrito na Seção 2.1.4.

$$\mu_j^* = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}. \quad (4.10)$$

$$\Sigma_j^* = \frac{\sum_{i=1}^n u_{ij}^m (x_i - \mu_j)^T (x_i - \mu_j)}{\sum_{i=1}^n u_{ij}^m}. \quad (4.11)$$

No cálculo da estimativa de w^* , minimiza-se o primeiro termo do lado direito da Equação 4.8 respeitando a condição $\sum_{j=1}^k w_{ij} = 1$. É necessário o uso de multiplicadores de Lagrange λ resultando na expressão de $f(w_{ij})$ que deve ser maximizada.

$$\begin{aligned} f(w_{ij}) &= \sum_{i=1}^n \left\{ \sum_{j=1}^k u_{ij}^m \log \left[\frac{1}{\beta} \exp \left(-\frac{1}{2} \frac{\sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta^2} \right) \right] - \lambda \left(\sum_{j=1}^k w_{ij} - 1 \right) \right\} \\ &= \sum_{i=1}^n \left\{ \sum_{j=1}^k u_{ij}^m (\log \beta^{-1}) - \sum_{j=1}^k u_{ij}^m \left(\frac{1}{2} \frac{\sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta^2} \right) - \lambda \left(\sum_{j=1}^k w_{ij} - 1 \right) \right\}. \end{aligned} \quad (4.12)$$

Para cada elemento $i = 1, \dots, n$, deriva-se $f(w_{is})$ com respeito à w_{is} e iguala-se o resultado a zero¹

$$\begin{aligned} \frac{\partial}{\partial w_{is}} \left[u_{is}^m (\log \beta^{-1}) - \frac{u_{is}^m}{2\beta^2} \left(\sum_{v \in V_i} (w_{is}^* - w_{vs})^2 \right) - \lambda (w_{is} - 1) \right] &= 0 \\ -\frac{u_{is}^m}{2\beta^2} \left[\sum_{v \in V_i} \frac{\partial}{\partial w_{is}} ((w_{is}^* - w_{vs})^2) \right] - \lambda &= 0 \\ -\frac{u_{is}^m}{2\beta^2} \left[\sum_{v \in V_i} (2(w_{is}^* - w_{vs})) \right] - \lambda &= 0 \\ -\frac{u_{is}^m}{\beta^2} \left[|V_i| w_{is}^* - \sum_{v \in V_i} w_{vs} \right] &= \lambda \end{aligned} \quad (4.13)$$

¹ Para calcular essa derivada é necessário aplicar a regra da cadeia. Sendo $f(x) = x^2$ e $g(w_i) = (w_i - w_v)$, logo, $f(g(w_i)) = (w_i - w_v)^2$. Para encontrar $\frac{\partial f(g(w_i))}{\partial w_i}$ temos que conhecer $\frac{\partial f(x)}{\partial x}$ e $\frac{\partial g(w_i)}{\partial w_i}$.

A derivada de $f(x)$ é dada por: $\frac{\partial f(x)}{\partial x} = 2x$.

A derivada de $g(x)$ é dada por: $\frac{\partial g(w_i)}{\partial w_i} = \frac{\partial}{\partial w_i} (w_i - w_v) = 1$.

Aplicando a regra da cadeia $\frac{\partial f(g(w_i))}{\partial w_i} = 2(w_i - w_v)$.

Isolando w_{ij}^* na Equação 4.13 temos

$$w_{is}^* = \frac{1}{|V_i|} \sum_{v \in V_i} w_{vs} - \frac{\beta^2}{|V_i| u_{is}^m} \lambda. \quad (4.14)$$

Levando em consideração a restrição $\sum_{j=1}^k w_j = 1$ tem-se

$$\begin{aligned} 1 &= \sum_{j=1}^k \left(\frac{1}{|V_i|} \sum_{v \in V_i} w_{vj} - \frac{\beta^2}{|V_i| u_{ij}^m} \lambda \right) \\ \lambda &= \frac{1}{\beta^2 \sum_{j=1}^k u_{ij}^{-m}} \left(\sum_{j=1}^k \sum_{v \in V_i} w_{vj} - |V_i| \right). \end{aligned} \quad (4.15)$$

Substituindo 4.15 em 4.14 tem-se

$$w_{is}^* = \frac{1}{|V_i|} \sum_{v \in V_i} w_{vs} - \frac{u_{is}^{-m}}{\sum_{j=1}^k u_{ij}^{-m}} \left(\frac{1}{|V_i|} \sum_{j=1}^k \sum_{v \in V_i} w_{vj} - 1 \right) \quad (4.16)$$

que pode ser reescrita da forma

$$w_{ij}^* = \frac{1}{|V_i|} \sum_{v \in V_i} w_{vj} + \frac{u_{ij}^{-m}}{\sum_{c=1}^k u_{ic}^{-m}} \left(1 - \frac{1}{|V_i|} \sum_{c=1}^k \sum_{v \in V_i} w_{vc} \right). \quad (4.17)$$

No cálculo da estimativa de β^* , minimiza-se o primeiro termo do lado direito da Equação 4.8 em relação a β .

$$\begin{aligned} \frac{\partial}{\partial \beta} \left\{ - \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \log \left[\frac{1}{\beta} \exp \left(- \frac{1}{2} \frac{\sum_{v \in V_i} (w_{ij} - w_{vj})^2}{\beta^2} \right) \right] \right\} &= 0 \\ - \frac{1}{\beta} + \sum_{i=1}^n \sum_{j=1}^k \sum_{v \in V_i} (w_{ij} - w_{vj})^2 \frac{1}{\beta^3} &= 0 \end{aligned} \quad (4.18)$$

Resolvendo a Equação 4.18 tem-se

$$\beta^* = \sqrt{\sum_{i=1}^n \sum_{j=1}^k \sum_{v \in V_i} (w_{ij} - w_{vj})^2}. \quad (4.19)$$

O algoritmo CFGMM itera atualizando consecutivamente os valores dos parâmetros do modelo, $\psi = \{w, \mu, \Sigma, \beta\}$, da função de dissimilaridade, d , e da partição *fuzzy*, U , até que a função objetivo convirja para um valor mínimo, ou seja, até que a diferença entre o

J de duas iterações consecutivas seja menor do que um limiar pré-determinado. Assim, o CFGMM pode ser resumido no Algoritmo 5:

Algoritmo 5: CFGMM

- Inicializar a partição *fuzzy* U e os pesos contextuais da mistura W .
 - Repetir até a convergência de $J(U, \psi|X)$, Equação 4.8:
 - Calcular os vetores de médias, $\{\mu_j\}$, e as matrizes de covariância, $\{\Sigma_j\}$, usando, respectivamente as Equações 4.10 e 4.11, para $j = 1, \dots, k$.
 - Calcular os pesos da mistura, $\{w_{ij}\}$, usando a Equação 4.17 para $i = 1, \dots, n$ e $j = 1, \dots, k$.
 - Calcular o parâmetro de suavidade do modelo, β , usando a Equação 4.19.
 - Calcular a função de dissimilaridade, d_{ij} , usando a Equação 4.3, para $i = 1, \dots, n$ e $j = 1, \dots, k$. Se $d_{ij} = 0$ para algum j , faça $u_{ij} = 1$, $u_{ig} = 0$, $\forall g \neq j$.
 - Atualizar a partição *fuzzy* U usando a Equação 4.2, para $i = 1, \dots, n$ e $j = 1, \dots, k$.
-

A Figura 6 mostra o esquema em blocos do processo de agrupamento dos dados pelo Modelo de Mistura Finita *Fuzzy* Contextual. O método recebe como entrada os dados a serem agrupados, X , e o conjunto de parâmetros de entrada $\{k, V_i, Th, m\}$, composto, respectivamente, pelo número de grupos, tipo de vizinhança, limiar de convergência de J e expoente *fuzzy*. Após a inicialização da partição *fuzzy*, U , e dos pesos contextuais da mistura, W , o modelo é ajustado aos dados pelo processo iterativo descrito no Algoritmo 5 e simplificado na Figura 6. No processo de convergência do algoritmo, os parâmetros do modelo são otimizados em relação à função objetivo, J .

Os parâmetros de cada componente de uma mistura, $\{w_j, \mu_j, \Sigma_j\}$, para $1 \leq j \leq k$, podem ser interpretados como o protótipo do grupo formado pelos dados associados àquele componente. No caso do modelo proposto, o protótipo de cada grupo consiste no conjunto de parâmetros que contém os pesos da mistura, o vetor de médias e as matrizes de covariância. Os protótipos de cada grupo descrevem o comportamento dos dados mas não representam o agrupamento dos dados.

O agrupamento dos dados em k grupos se dá pela associação de cada elemento do conjunto de dados a um componente da mistura. Assim, é possível derivar duas abordagens para o Modelo de Mistura de Gaussianas *Fuzzy* Contextual: na primeira o processo de rotulagem de dados se baseia no Modelo de Mistura de Gaussianas; na segunda a rotulagem dos dados se baseia no processo adotado pelo *Fuzzy C-Means*.

Uma vez que os parâmetros do modelo, ψ , são estimados, é possível rotular cada

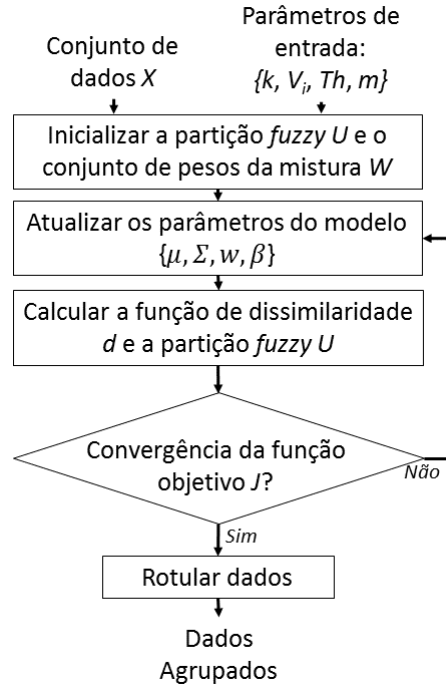


Figura 6 – Processo de agrupamento dos dados pelo Modelo de Mistura de Gaussianas Fuzzy Contextual.

elemento do conjunto de dados X associando cada x_i a um dos k grupos usando a regra de decisão Bayesiana ou a partição *fuzzy*.

CFGMM-B:

Na abordagem do CFGMM baseada no GMM, o processo de rotulagem dos dados se dá pela aplicação da regra de Bayes. Pelo teorema de Bayes, a probabilidade *a posteriori* de um elemento x_i pertencer ao j -ésimo grupo, com $j = 1, \dots, k$, é dada pela equação:

$$z_j(\psi^t | x_i) = \frac{w_j p(x_i | \theta_j)}{p(x_i | \psi)}, \quad j = 1, \dots, k \quad (4.20)$$

em que o denominador $p(x_i | \psi)$ é a distribuição da mistura da Equação 2.6. Para cada elemento x_i , $p(x_i | \psi)$ é constante para todas as classes. O classificador Bayesiano pode, então, ser usado para agrupar os dados usando a regra:

$$j^* = \underset{j}{\operatorname{argmax}} z_j(\psi^t | x_i). \quad (4.21)$$

O elemento x_i é associado ao grupo j^* que ele tem a maior probabilidade *a posteriori* de pertencer.

CFGMM-F:

Na abordagem do CFGMM baseada no FCM, o processo de rotulagem dos dados se dá pelo uso da partição *fuzzy*. Sendo $U = \{u_{ij}\}$ a partição *fuzzy* de X , u_{ij} representa o

grau de pertinência do elemento x_i ao grupo j ,

$$j^* = \underset{j}{\operatorname{argmax}} u_{ij} \quad (4.22)$$

o elemento x_i é associado ao grupo j^* com maior grau de pertinência:

4.2 Considerações finais

Neste capítulo foram propostas duas abordagens para um método de agrupamento contextual de dados baseado na combinação do Modelo de Mistura de Gaussianas *Fuzzy*. A inserção do contexto no FGMM foi realizada ao considerar uma densidade *a priori* para os pesos da mistura baseada em Campos Aleatórios Markovianos Gaussianos que são usados para representar a relação espacial entre os dados. Uma vez definida a função objetivo do CFGMM, as equações de atualização dos parâmetros do modelo foram derivadas.

Uma vantagem do uso de Campos Aleatórios Markovianos Gaussianos em relação à função de densidade acumulada de Gibbs para modelar a relação espacial dos dados é a possibilidade de calcular o parâmetro de suavidade do modelo, β , a partir dos dados com o uso de uma equação de atualização derivada da função objetivo especificada na Equação 4.7. Além disso, é possível modificar o modelo de maneira que exista um parâmetro de suavidade para cada grupo. Valores diferentes de β para grupos diferentes podem contribuir para a melhor adaptação do modelo ao conjunto de dados que se quer modelar. Nikou *et al.* (NIKOU; GALATSANOS; LIKAS, 2007) propõem uma versão do modelo na qual o parâmetro β de cada grupo controla o grau de suavidade em diferentes direções espaciais. O CFGMM proposto pode ser adaptado para diferentes sistemas de vizinhança definidas no GMRF.

Uma vez que a expressão da função objetivo J do Modelo de Mistura de Gaussianas *Fuzzy* Contextual é matematicamente menos complexa do que a expressão do Q_{MAP} do Modelo de Mistura Finita Variante no Espaço, a abordagem *fuzzy* possibilita que as propriedades da probabilidade, $0 \leq w_j \leq 1$ e $\sum_{j=1}^k w_j = 1$, sejam respeitadas no processo de otimização da função objetivo. Nesse caso, a etapa de correção dos valores dos pesos da mistura necessária nos algoritmos SVFMM é eliminada no processo de agrupamento simplificando o processo de otimização e acelerando a convergência do algoritmo.

O agrupamento de dados pode ser resumido em três etapas. Primeiro, formalizar o modelo dos dados usando o Modelo de Mistura de Gaussianas *Fuzzy* Contextual. Segundo, ajustar o modelo aos dados a serem agrupados estimando os parâmetros do modelo. Por fim, associar cada elemento do conjunto de dados a um grupo usando a regra de Bayes (CFGMM-B) ou a partição *fuzzy* (CFGMM-F).

5 Experimentos

Neste capítulo são apresentados os bancos de imagens usados nos experimentos realizados: um banco de imagens sintéticas, um banco de imagens naturais e um banco de imagens simuladas de ressonância magnética. Além disso, são descritas as medidas de desempenho adotadas para avaliar as técnicas de agrupamento, são elas: o Índice Rand, o Índice Probabilístico Rand, *F-measure* e o tempo de processamento.

Um estudo dos parâmetros de entrada do Modelo de Mistura de Gaussianas *Fuzzy* Contextual (CFGMM) é, inicialmente, conduzido para determinar uma configuração padrão que será adotada nos experimentos seguintes. Nesse estudo, o resultado da segmentação das imagens sintéticas pelo CFGMM adotando diferentes parâmetros são analisados para definir a melhor configuração do modelo. Além disso, a análise experimental da complexidade matemática do CFGMM é realizada.

O desempenho do CFGMM é, então, comparado a outros métodos de agrupamento, o *K-Means*, o *Fuzzy C-Means*, o Modelo de Mistura de Gaussianas e o Modelo de Mistura de Gaussianas *Fuzzy*, usando o conjunto de imagens sintéticas. Na análise dos resultados, buscamos evidências de que as equações de atualização dos parâmetros do modelo foram corretamente derivadas, que forneçam suporte à hipótese de que a abordagem contextual é capaz de melhorar significativamente o desempenho do agrupamento na segmentação de imagens quando comparada à abordagem pontual e que forneçam suporte à hipótese de que a segmentação das imagens pelo método proposto seja significativamente melhor do que a segmentação por outras técnicas de agrupamento.

5.1 Metodologia dos experimentos

Imagens sintéticas, a *BrainWeb simulated brain database* (COCOSCO et al., 1997) e a *Berkeley image segmentation database and benchmark* (MARTIN et al., 2001) são adotadas para realizar a validação. Uma comparação quantitativa entre o desempenho do algoritmo desenvolvido e o dos encontrados na literatura é realizada usando as medidas: Índice Rand, o Índice Probabilístico Rand, *F-measure* e o tempo de processamento.

5.1.1 Bancos de imagens

5.1.1.1 Imagens sintéticas

O banco de imagens sintéticas é formado por imagens em tons de cinza desenvolvidas para ajustar os parâmetros do modelo e avaliar a robustez do método proposto em relação à variação de ruído. Três imagens de tamanho 100×100 *pixels* foram criadas. A primeira

imagem, exibida na Figura 7 (esquerda), é composta por duas classes com mesma média, 125, e diferentes níveis de ruído. A segunda, exposta na Figura 7 (centro), possui 3 classes com intensidades de nível de cinza, 96, 160 e 224. A terceira imagem sintética, exposta na Figura 7 (direita), possui 5 classes com intensidades de nível de cinza, 25, 76, 127, 178 e 229. Níveis de cinza extremos, como os próximos a 0 ou a 255, foram evitados porque, ao adicionar o ruído, os *pixels* dessas classes não poderiam receber valores menores que 0 ou maiores que 255. Desse modo, a distribuição do ruído iria se comportar de maneira diferente nas classes com níveis de cinza 0 e 255, em relação às outras classes com níveis de cinza intermediários.



Figura 7 – Imagens sintéticas de 2, 3 e 5 classes

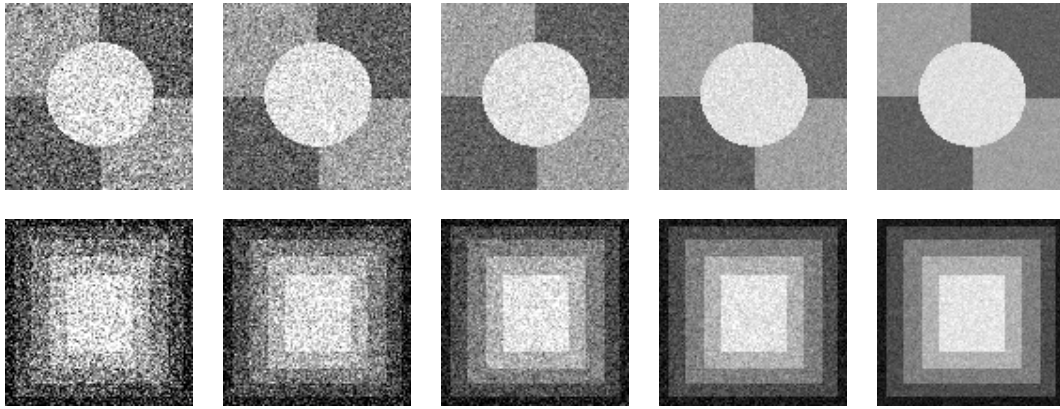


Figura 8 – Imagens sintéticas de 3 (acima) e 5 classes (abaixo) com relação sinal ruído 1, 4, 8, 12 e 16 dB (da esquerda para a direita)

O ruído Gaussiano com relação sinal ruído 1, 4, 8, 12 e 16 dB foi adicionado às imagens de 3 e 5 classes, Figura 8. A relação sinal ruído (SNR) em dB de um sinal é dada por

$$SNR = 10 \log_{10} \frac{\sigma_{\text{sinal}}^2}{\sigma_{\text{ruído}}^2} \quad (5.1)$$

em que σ_{sinal}^2 é a variância do sinal e $\sigma_{\text{ruído}}^2$ é a variância do ruído.

5.1.1.2 BrainWeb simulated brain database

A *BrainWeb simulated brain database* é composto por modelos anatômicos de cérebros normais que podem servir de *ground truth* para avaliar o desempenho de métodos de análise de imagem. O modelo anatômico utilizado para gerar simulações de imagens tri-dimensionais de ressonância magnética (MR, do inglês, *Magnetic Resonance*) consiste em um volume cuja intensidade dos *voxels* refletem a porção de tecido presente em cada *voxel*. Os modelos estão disponíveis tanto em um formato binário, no qual cada *voxel* possui o rótulo da classe de tecido a qual pertence, como em um formato *fuzzy*, no qual cada *voxel* pode pertencer a mais de uma classe com diferentes graus de pertinência. A Figura 9 apresenta algumas imagens para ilustração das classe de tecido que compõem o modelo.

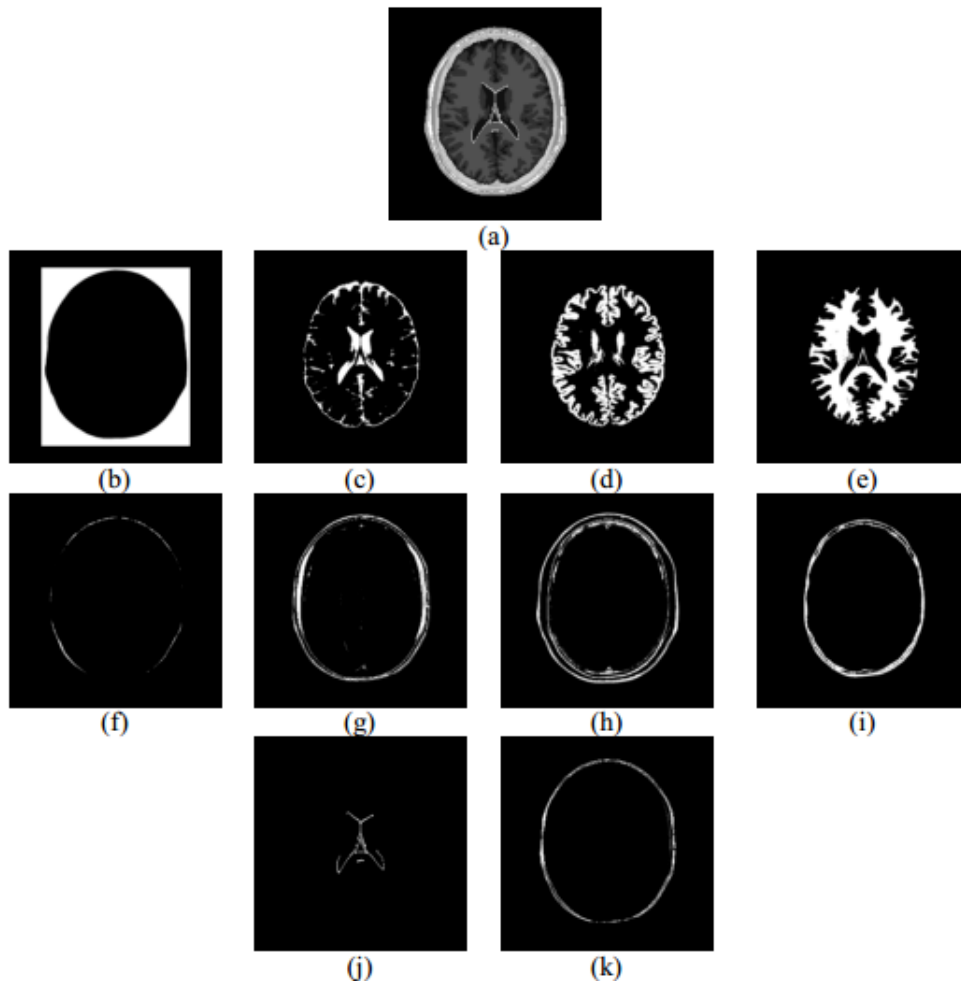


Figura 9 – Exemplo do modelo no formato binário (a) e sua decomposição nas classes: (b) *background*, (c) fluido cérebro-espinhal (CSF), (d) matéria cinzenta (GM), (e) matéria branca (WM), (f) gordura, (g) músculo, (h) pele, (i) osso, (j) matéria glial e (k) material conectivo.

Atualmente, a *BrainWeb* contém um banco de dados de imagens de MR do cérebro simulado em dois modelos anatômicos: normal e esclerose múltipla (*multiple sclerosis*). As simulações de imagens de MR estão disponíveis em três modalidades, ponderadas em T1, T2 e densidade de prótons (Pd). Além disso, permite-se a definição dos seguintes parâmetros nas simulações das imagens: largura da fatia da imagem (*slice*), nível de ruído (*noise levels*) e intensidade de não-uniformidades (*intensity non-uniformity*). A percentagem de ruído representa a razão entre o desvio padrão do ruído branco Gaussiano adicionado e o sinal do tecido com maior brilho na imagem. A Figura 10 ilustra as imagens ponderadas em T1, T2 e Pd.



Figura 10 – Exemplo de imagens ponderadas em T1 (a), T2 (b) e Pd (c)

Apenas 4 classes são consideradas nos experimentos: matéria branca (WM), matéria cinzenta (GM), fluido cérebro-espinhal (CSF) e *background*. Uma etapa de pré-processamento é, então, necessária para converter em *background* as classes correspondentes à gordura, músculo, pele, osso, matéria glial e material conectivo. A Figura 11 exibe, na coluna da esquerda, imagens originais das fatias 50, 93 e 120 com 3% de ruído e 0% de não-uniformidades. As mesmas imagens são mostradas, na coluna do centro, após o pré-processamento. A coluna da direita corresponde aos respectivos *ground truths*. As imagens coloridas são a versão RGB da imagem multiespectral composta pelas imagens ponderadas em T1, T2 e densidade de prótons (Pd).

Nos experimentos são usadas fatias de uma imagem tri-dimensional simulada de MR baseada em um modelo anatômico de cérebro normal. As fatias têm largura de 1mm, com 0% de não-uniformidades e níveis de ruído variando de 0%, 1%, 3%, 5%, 7% e 9%. A Figura 12 exibe a fatia 100 da imagem com diferentes níveis de ruído após o pré-processamento. As imagens coloridas são a versão RGB da imagem multiespectral composta pelas imagens ponderadas em T1, T2 e densidade de prótons (Pd).

Entre as 181 fatias sagitais, selecionamos somente aquelas com um mínimo de 10 *voxels* de cada classe, resultando em um total de 133 fatias. Essa seleção foi necessária para garantir que a matriz de covariância seja não-singular (ou seja, inversível) e os parâmetros da função de densidade Gaussiana possam ser calculados.

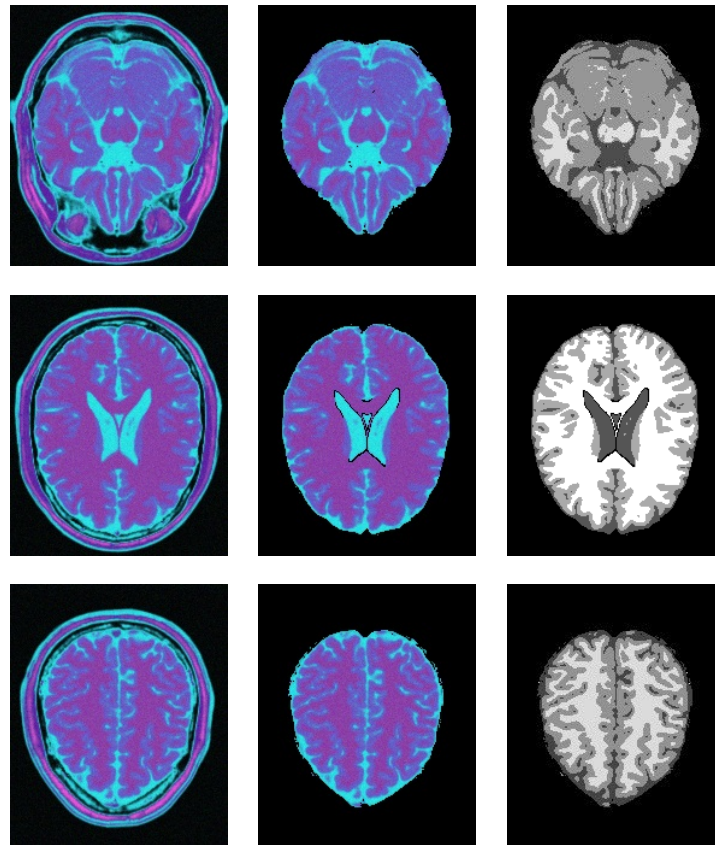


Figura 11 – Imagens originais das fatias 50, 93 e 120 com 3% de ruído e 0% de não-uniformidades (coluna da esquerda), mesmas imagens após o pré-processamento (coluna do centro) e os seus respectivos *ground truths* (coluna da direita).

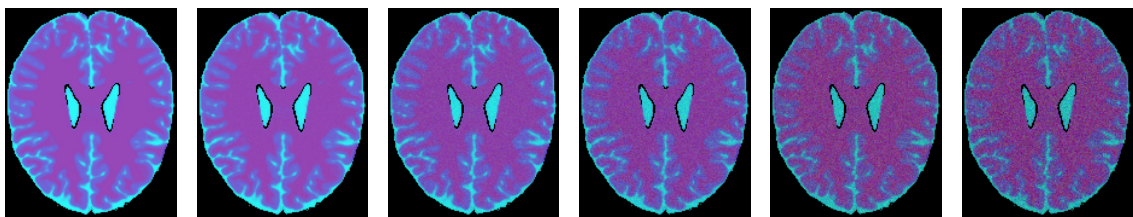


Figura 12 – Exemplo da fatia 100 de imagens com largura de fatia de 1mm, com 0% de não-uniformidades e níveis de ruído de 0%, 1%, 3%, 5%, 7% e 9% (da esquerda para direita).

5.1.1.3 Berkeley image segmentation database and benchmark

O *Berkeley image segmentation database and benchmark* (BSDS500) ([ARBELAEZ et al., 2011](#)) contém um conjunto de 500 imagens naturais com os respectivos *ground truths*, obtidos através da segmentação manual de cada imagem por diferentes indivíduos. Estas imagens têm dimensão de 481×321 *pixels* e estão disponíveis em tons de cinza e em RGB no formato .jpg. A Figura 13 apresenta algumas imagens da *Berkeley database*.

A segmentação manual da imagem exibida na Figura 14 mostra como a segmentação realizada por humanos pode gerar diferentes resultados.

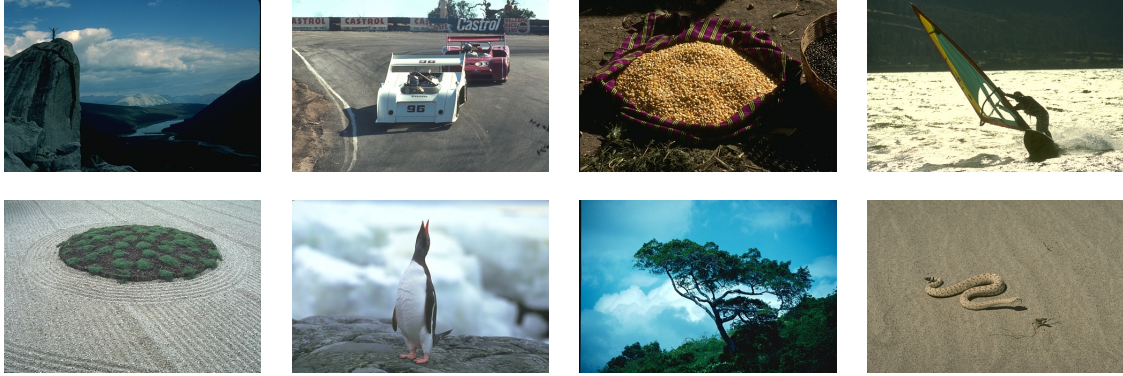


Figura 13 – Exemplo de imagens da *Berkeley image segmentation database*

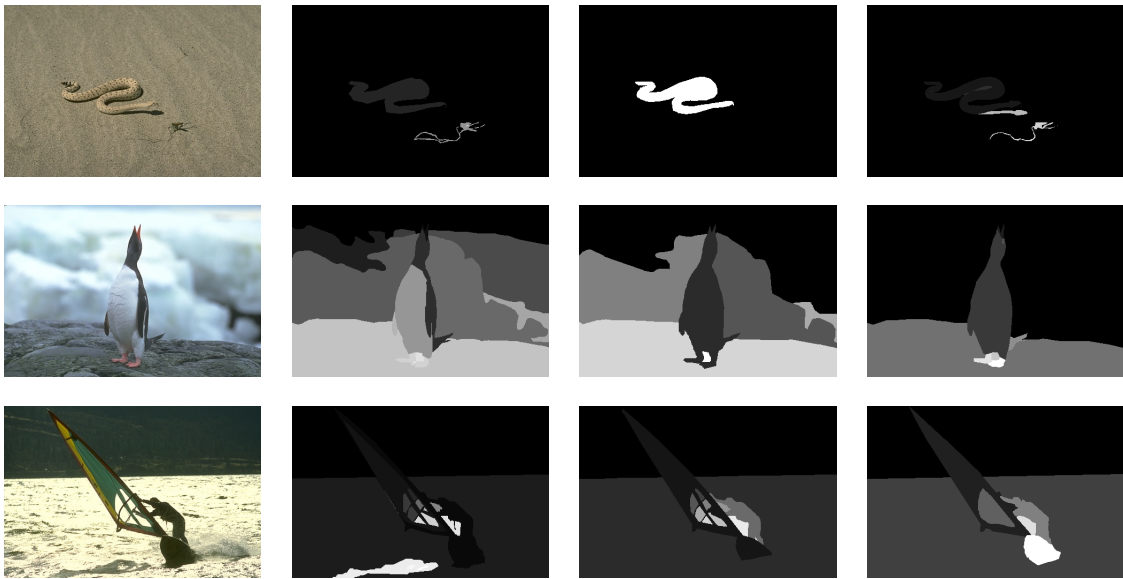


Figura 14 – Exemplo de imagens em tons de cinza da *Berkeley image segmentation database*. Cada imagem foi manualmente segmentada por 3 pessoas diferentes.

Além de fornecer o conjunto de imagens que podem servir de base de comparação para diferentes algoritmos de segmentação baseada em região ou em bordas, o BSDS500 também calcula medidas de desempenho para os algoritmos. O *F-measure* (MARTIN et al., 2001) é usado na análise de detecção de borda das regiões das imagens e a Variação da Informação (*Variation of Information*), o Índice de Rand (*Rand Index*) e a Cobertura da Segmentação (*Segmentation Covering*) (ARBELAEZ et al., 2011) são usados como medida de desempenho da segmentação baseada em região.

5.1.2 Medidas de desempenho

5.1.2.1 Índice Rand e Índice Probabilístico Rand

Originalmente, o Índice Rand (R) (RAND, 1971) foi proposto para a avaliação de agrupamento de dados. Ele funciona comparando a compatibilidade de rótulos entre pares de elementos de um grupo. O Índice Rand entre o resultado de uma segmentação, S_{test} , e o *ground truth*, S_g , de uma imagem é dado pela soma dos pares de *pixel* que possuem o mesmo rótulo em S_{test} e S_g com aqueles que possuem rótulos diferentes em ambas segmentações, dividido pelo número total de pares de *pixels*. O índice R é dado por

$$R(S_{test}, S_g) = \frac{2}{n(n-1)} \sum_{i,j,j>i} [c_{ij} + d_{ij}], \quad (5.2)$$

em que $c_{ij} = 1$ se os *pixels* i e j pertencerem ao mesmo grupo em S_{test} e pertencerem ao mesmo grupo em S_g , $c_{ij} = 0$ caso contrário. $d_{ij} = 1$ se os *pixels* i e j pertencerem à grupos diferentes em S_{test} e pertencerem à grupos diferentes em S_g , $d_{ij} = 0$ caso contrário. O termo de normalização é o inverso do número de todos os possíveis pares nos n *pixels* da imagem.

O Índice Probabilístico Rand (PR) (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005) é uma derivação do Índice Rand proposta para lidar com casos nos quais há disponibilidade de vários *ground truths* de segmentação para cada imagem, como é o caso da *Berkeley database*. O índice PR conta a fração de pares de *pixels* que recebem o mesmo rótulo tanto no resultado da segmentação quanto no *ground truth*, fazendo uma média dos vários *ground truths* para considerar a variação da percepção humana no processo de segmentação. Sendo $S_g = \{S_1, S_2, \dots, S_M\}$ o conjunto de M *ground truths* de uma imagem e S_{test} o resultado da segmentação desta imagem a ser avaliado, o índice PR é dado por

$$PR(S_{test}, S_g) = \frac{2}{n(n-1)} \sum_{i,j,j>i} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})], \quad (5.3)$$

em que $c_{ij} = 1$ se os *pixels* i e j pertencerem ao mesmo grupo em S_{test} e $c_{ij} = 0$ caso contrário. O termo de normalização é o inverso do número de todos os possíveis pares nos n *pixels* da imagem e p_{ij} é a probabilidade que os *pixels* i e j pertençam ao mesmo grupo dado S_g , que é calculada como a média do relacionamento de paridade de *pixels* considerando todos os *ground truths*. O termo relacionamento de paridade de *pixels* consiste em um elemento binário que recebe o valor verdadeiro (1) quando um par de *pixels* pertence a um mesmo grupo e o valor falso (0) caso contrário.

O índice PR assume valores no intervalo entre 0 e 1. O valor 0 indica que cada par de *pixel* no resultado da segmentação de uma imagem, S_{test} , tem um relacionamento oposto a cada par de *pixel* nos *ground truths* enquanto que o valor 1 indica que cada par de *pixel* em S_{test} tem o mesmo relacionamento de cada par de *pixel* dos *ground truths*. Um par de *pixels* que está na mesma região na maioria dos *ground truths*, será penalizado se não pertencer à mesma região no resultado da segmentação.

5.1.2.2 *F-measure*

Além de disponibilizar o banco de imagens contendo 500 imagens naturais com os respectivos *ground truths*, o *Berkeley image segmentation database and benchmark* (BSDS500) (ARBELAEZ et al., 2011) disponibiliza o *benchmark* que avalia a qualidade dos resultados produzidos por um algoritmo de segmentação para comparação padronizada com outras técnicas de segmentação. Neste trabalho, além do Índice PR apresentado na Seção 5.1.2.1, foi usada a medida de desempenho *F-measure* calculada pelo *benchmark* usada na análise de detecção de borda das regiões das imagens segmentadas da *Berkeley database* nos experimentos realizados. Embora o método proposto não tenha sido validado na tarefa de segmentação baseada em bordas, achamos importante calcular o *F-measure* do método proposto para possibilitar a comparação padronizada do seu desempenho com o desempenho de outras técnicas encontradas na literatura.

O *F-measure* é a média harmônica das medidas *precision* e *recall* usado por Martin (MARTIN; FOWLKES; MALIK, 2004) para avaliar os resultados de segmentações de imagens. *Precision* é definido como a proporção de *pixels* do resultado da segmentação que foram corretamente classificados de acordo com o *ground truth*. O *Recall* é definido como a proporção de *pixels* do *ground truth* que foram classificados corretamente no resultado da segmentação. O *F-measure* é a média harmônica entre o *Precision* e o *Recall*.

Considerando a segmentação de imagens em duas classes, *foreground* e *background* e sendo o resultado de uma segmentação, S_{test} , e o *ground truth*, S_g , de uma imagem, cada *pixel* analisado vai se encaixar em uma de 4 categorias (FAWCETT, 2006): Verdadeiro Positivo, Falso Positivo, Falso Negativo ou Verdadeiro Negativo. Se o *pixel* é considerado *foreground* em S_{test} e S_g , ele é contado como Verdadeiro Positivo (TP, do inglês, *True Positive*); se o *pixel* é considerado *foreground* em S_{test} e *background* em S_g , ele é contado como Falso Positivo (FP, do inglês, *False Positive*); se o *pixel* é considerado *background* em S_{test} e *foreground* em S_g , ele é contado como Falso Negativo (FN, do inglês, *False Negative*); se o *pixel* é considerado *background* em S_{test} e S_g , ele é contado como Verdadeiro Negativo (TN, do inglês, *True Negative*).

O *Precision* é calculado como

$$Precision = \frac{TP}{TP + FP}. \quad (5.4)$$

O *Recall* é dado por

$$Recall = \frac{TP}{TP + FN}. \quad (5.5)$$

O *F-measure* pode ser usado como medida única de desempenho e é dado por

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (5.6)$$

Os valores de *Precision*, *Recall* e *F-measure* estão compreendidos entre 0 e 1. Quanto mais próximos de 1 os valores estiverem, maior o número de *pixels* considerados da mesma classe pelo resultado da segmentação e pelo *ground truth*.

5.1.2.3 Tempo de processamento

O custo computacional de um método pode ser avaliado pelo seu tempo de processamento para executar uma determinada tarefa. A média do tempo, em segundos, usado por cada método para analisar todas as imagens de cada um dos bancos de imagens são medidos para comparar suas performances.

5.1.2.4 Intervalo de confiança

Na análise dos resultados obtidos nos experimentos adotaremos o intervalo de confiança para dar ideia da dispersão ou variabilidade das estimativas. Um intervalo muito grande indica que a estimativa calculada não é tão acurada quanto outra com intervalo menor, ou seja, quanto maior a amplitude do intervalo menor a confiabilidade da estimativa.

Sempre que um experimento foi executado mais de uma vez, o seu resultado é expresso considerando a média e o intervalo da confiança da média dos experimentos. Seja R_i o resultado do i -ésimo experimento, com $i = 1, \dots, n$. Assumindo que os resultados $R_1, \dots, R_n$, formam uma amostra aleatória de uma população com distribuição normal com média desconhecida μ e variância desconhecida σ . Assumindo que a variância populacional é igual a variância amostral, o intervalo de confiança da média μ com nível de confiança de 95% é dado por (DEGROOT; SCHERVISH, 2011):

$$\left[\bar{R} - 1,95 \frac{s}{\sqrt{n}}, \bar{R} + 1,95 \frac{s}{\sqrt{n}} \right] \quad (5.7)$$

em que $\bar{R}$ é a média e s é a variância dos resultados obtidos (amostral). É possível, então, afirmar que se um grande número de amostras é retirado da população, 95% dos intervalos irão conter o valor verdadeiro do parâmetro μ .

5.1.2.5 Ambiente de teste

Para realização dos experimentos, é utilizado um computador com sistema operacional Windows®8 de 64 bits, com 8GB de memória RAM e processador Intel(R) Core(TM) i7 com 2,4GHz de frequência, além do MATLAB®(MATLAB, 2010) como ambiente de execução. Para execução do *benchmark* de segmentação BSDS500 de Berkeley (ARBELAEZ et al., 2011) foi necessário o uso de emulador do sistema operacional Linux.

5.2 Experimentos e resultados obtidos

O método de agrupamento de dados baseado no Modelo de Mistura de Gaussianas *Fuzzy* Contextual proposto neste trabalho recebe, como entrada, os dados a serem agrupados, X , e o conjunto de parâmetros de entrada $\{k, V_i, Th, m\}$, composto, respectivamente, pelo número de grupos, tipo de vizinhança, limiar de convergência de J e expoente *fuzzy* (Figura 6). Após a inicialização da partição *fuzzy*, U , e dos pesos contextuais da mistura, W , o conjunto de parâmetros do CFGMM composto pelos pesos da mistura, pelo vetor de médias, pelas matrizes de covariância e pelo parâmetro de suavidade são otimizados em relação à função objetivo no processo iterativo descrito no Algoritmo 5. Por fim, cada elemento de X é associado a um dos k grupos usando a regra de Bayes (CFGMM-B) ou a partição *fuzzy* (CFGMM-F).

O desempenho do método proposto vai depender, não só, do processo de otimização dos parâmetros do modelo, como também, dos valores dos parâmetros de entrada e da maneira que a partição *fuzzy* e os pesos contextuais são inicializados. Aqui, iremos avaliar como algumas dessas variáveis influenciam o desempenho do método para determinar uma configuração padrão a ser adotada nos experimentos seguintes.

Visando a fase de validação do modelo, o Modelo de Mistura de Gaussianas *Fuzzy* (FGMM) é usado na inicialização da partição *fuzzy* e pesos contextuais da mistura. O algoritmo FGMM foi escolhido pois pode ser usado para inicializar igualmente os parâmetros dos métodos clássicos de agrupamento, o *K-Means* (KM), o *Fuzzy C-Means* (FCM) e o Modelo de Mistura de Gaussianas (GMM) uma vez que ele tem como parâmetros os vetores de médias, $\{\mu_j\}$ do KM, partição *fuzzy*, U , do FCM, e os pesos da mistura, $\{w_j\}$, os vetores de médias, $\{\mu_j\}$, as matrizes de covariância, $\{\Sigma_j\}$ do GMM.

Tendo em vista que o desempenho de algoritmos de agrupamento iterativo apresenta alto grau de dependência em relação à escolha inicial dos parâmetros, como é o caso do KM, do FCM, do GMM, do FGMM e do CFGMM, os valores iniciais de U e de W são determinados rodando o FGMM 30 vezes e escolhendo os parâmetros associados ao menor valor da função objetivo J . Como configuração inicial dos parâmetros de entrada, foi adotado o expoente *fuzzy*, $m = 1,6$, o limiar de convergência, $Th < 10^{-4}$, e a vizinhança V_i^4 . O número de grupos, k , é determinado de acordo com a imagem analisada. O processo de rotulagem dos dados se dará pela aplicação da regra de Bayes como explicado na Seção 4.1.

5.2.1 Análise do modelo com o banco de imagens sintéticas

Inicialmente, o comportamento do Modelo de Mistura de Gaussianas *Fuzzy* Contextual na segmentação de imagens baseada em regiões foi avaliado usando as imagens sintéticas de 2 classes, de 3 classes com SNR 8 dB e de 5 classes com SNR 8 dB.

A Figura 15 mostra as imagens originais, os respectivos *ground truths* e o resultado da segmentação. É possível observar que o CFGMM foi capaz de segmentar as imagens, tanto no caso em que as classes tinham a mesma média e níveis de ruído diferentes quanto no caso em que as classes tinham diferentes médias com o mesmo nível de ruído.

O gráfico da Figura 16 mostra a função objetivo, J , ao longo das iterações do CFGMM na segmentação das imagens exibidas na Figura 15. É possível verificar que J diminui ao longo das iterações atestando a convergência do algoritmo e corroborando o cálculo de otimização de J apresentado na Seção 4.1.

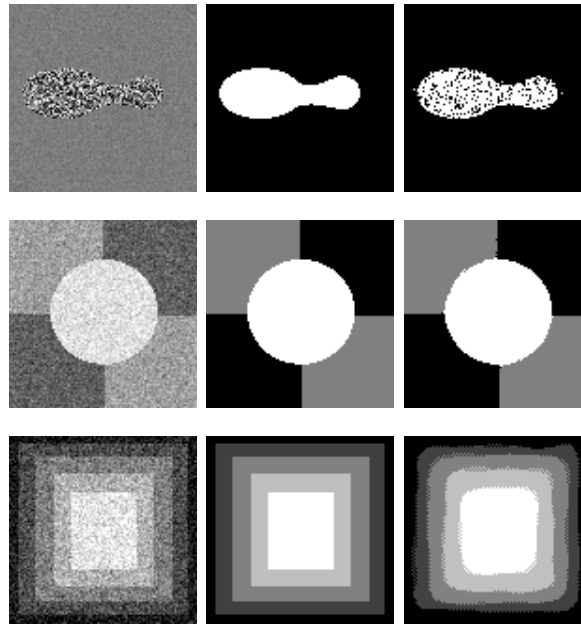


Figura 15 – Segmentação de imagens sintéticas. A primeira coluna mostra, de cima para baixo, as imagens sintéticas de 2, 3 e 5 classes, a segunda coluna exhibe os respectivos *ground truths* e a terceira coluna apresenta o resultado da segmentação pelo CFGMM.

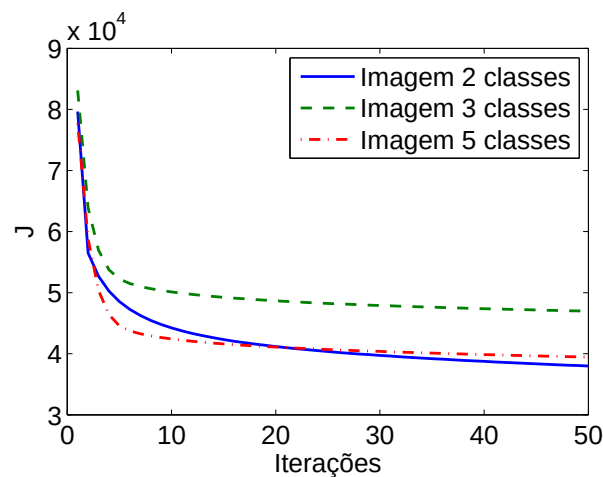


Figura 16 – Gráfico do função objetivo J .

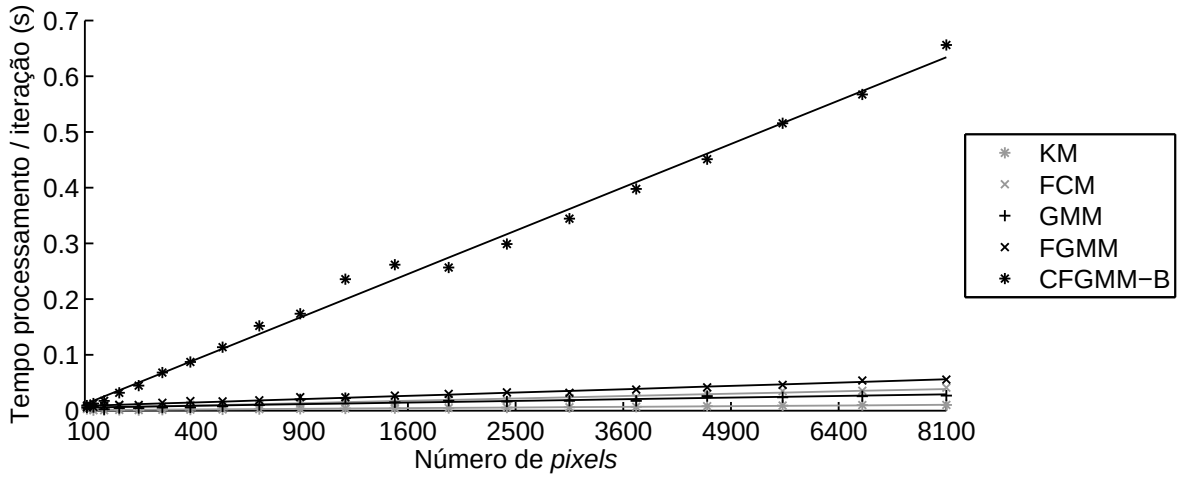


Figura 17 – Gráfico do tempo de processamento por iteração médio, em segundos, na segmentação de imagens de diferentes dimensões pelo KM, FCM, GMM, FGMM e CFGMM. As curvas ajustadas à cada método foram obtidas pela regressão linear dos respectivos resultados.

A complexidade matemática de um método de agrupamento pode ser determinada experimentalmente calculando o tempo de processamento de cada iteração do método para diferentes tamanhos de conjuntos de dados a serem agrupados. Com o objetivo de avaliar a complexidade matemática do CFGMM, a imagem sintética de 5 classes com SNR 8 dB foi redimensionada em diferentes tamanhos e o tempo de processamento por iteração médio para os métodos segmentarem cada imagem foi calculado.

O resultado obtido está exibido no gráfico da Figura 17 junto com as curvas que mais se aproximam ao resultado de cada método. O tamanho do conjunto de dados corresponde ao número de *pixels* de cada imagem. A curva que mais se aproxima ao resultado de cada método foi determinada usando o método dos Quadrados Mínimos Ordinários (WOLBERG; WOLBERG, 2006), uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados. As equações das curvas do KM, do FCM, do GMM, do FGMM e do CFGMM são descritas, respectivamente, nas Equações 5.8, 5.9, 5.10, 5.11 e 5.12, abaixo:

$$f(x) = 2,38 \times 10^{-7} x + 8,62 \times 10^{-4} \quad (5.8)$$

$$f(x) = 8,49 \times 10^{-7} x + 4,81 \times 10^{-3} \quad (5.9)$$

$$f(x) = 5,93 \times 10^{-7} x + 5,60 \times 10^{-3} \quad (5.10)$$

$$f(x) = 1,19 \times 10^{-6} x + 8,56 \times 10^{-3} \quad (5.11)$$

$$f(x) = 1,56 \times 10^{-5} x + 1,15 \times 10^{-2} \quad (5.12)$$

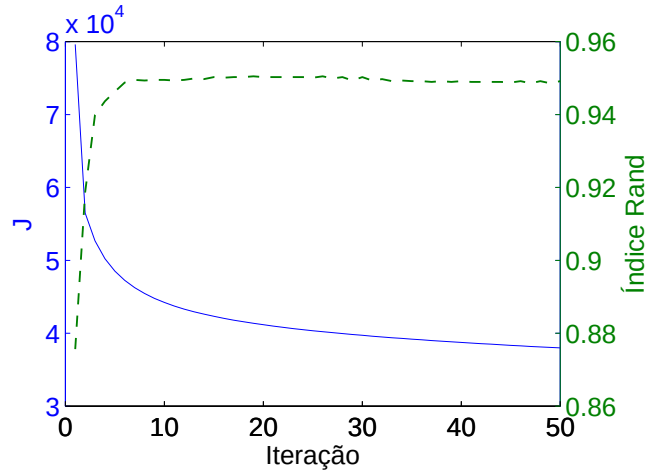
Observando o gráfico da Figura 17 percebe-se que o tempo de processamento por iteração do CFGMM aumenta linearmente com o aumento do tamanho do conjunto de dados agrupados (número de *pixels* da imagem). Esse comportamento é reproduzido pelo KM, pelo FCM, pelo GMM e pelo FGMM. No entanto, o coeficiente angular da curva do CFGMM, Eq. 5.12 é muito maior do que os coeficientes das curvas das outras técnicas, Eq. 5.8 à 5.11.

5.2.1.1 Limiar de convergência

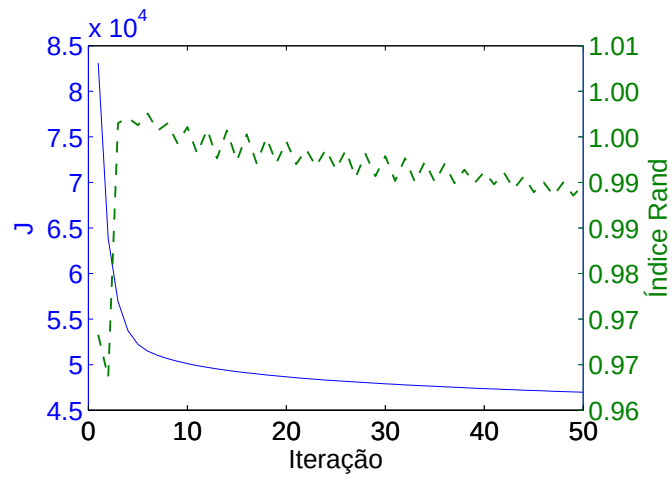
O critério de convergência adotado é a porcentagem de mudança da função objetivo J a cada iteração, assim, o algoritmo converge quando atinge um número máximo de iterações ou $\left| \frac{\Delta J}{J} \right| < Th$, em que ΔJ é a diferença do valor de J em duas iterações consecutivas e Th é um limiar de convergência pré-definido. A convergência do algoritmo e a qualidade da segmentação ao longo das iterações são exibidas na Figura 18. O comportamento do Índice Rand é exibido junto com o valor da função objetivo ao longo das iterações no processo de segmentação das imagens sintéticas de 2 classes, de 3 classes com SNR 8 dB e de 5 classes com SNR 8 dB. O critério de parada usado foi o limiar de convergência, $Th < 10^{-4}$, com número máximo de 50 iterações.

Observa-se na Figura 18 (a) e (b) que a qualidade da segmentação das imagens sintéticas de 2 e 3 classes, avaliada pela medida Índice Rand, são mantidas estáveis após o valor de J estabilizar. Já na imagem sintética de 5 classes, Figura 18 (c), a qualidade de segmentação atinge um valor ótimo por volta da iteração 10 e, em seguida, decai continuamente ao longo das iterações, mesmo com o valor de J estabilizado.

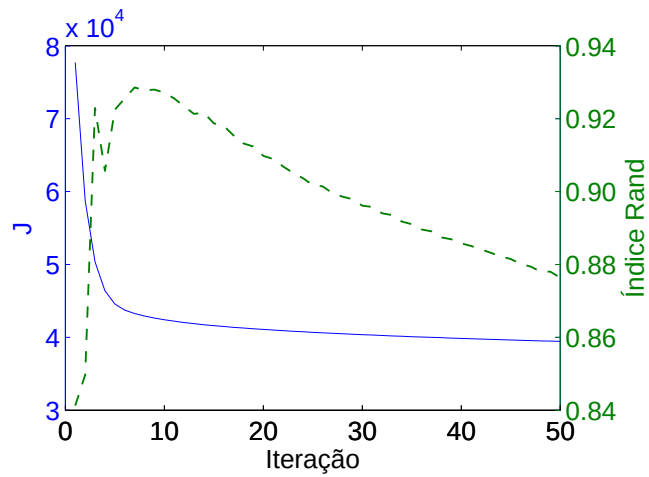
A Figura 19 mostra o resultado da segmentação da imagem sintética de 5 classes a cada 5 iterações até o algoritmo atingir o critério de parada. Ao longo das iterações, as regiões da imagem vão ficando mais homogêneas (com menos ruído) mas as bordas das regiões vão ficando borradas. As quinas vão se tornando curvas e as linhas retas vão perdendo a definição. Para evitar a degradação dos detalhes da imagem, definimos, então, como critério de parada padrão nos experimentos seguintes o valor de $Th = 10^{-2}$, com número máximo de 15 iterações.



(a) Imagem 2 classes



(b) Imagem 3 classes



(c) Imagem 5 classes

Figura 18 – Gráfico da função objetivo J e do Índice Rand ao longo das iterações na segmentação das imagens sintéticas de 2, 3 e 5 classes pelo CFGMM. J está representado nos gráficos com a linha sólida enquanto que o Índice Rand está representado com a linha tracejada.

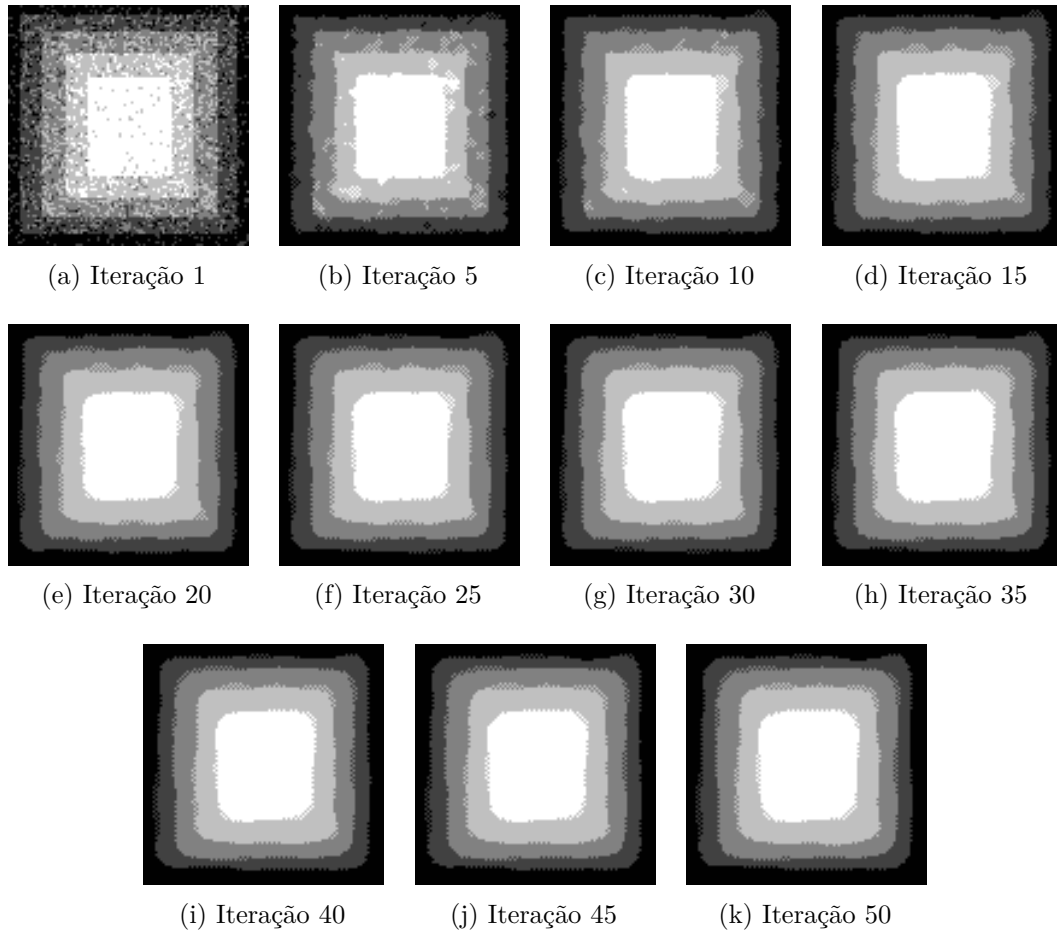


Figura 19 – Resultado da segmentação da imagem sintética de 5 classes pelo CFGMM ao longo das iterações.

5.2.1.2 Expoente *fuzzy*

O expoente *fuzzy* influencia na forma da partição *fuzzy* resultante. Como descrito na Seção 4.1, o expoente *fuzzy* possui a restrição de só poder assumir valores maiores que 1, $m > 1$. Quando o valor de m se aproxima de 1, a partição se torna "Booleana" e os graus de pertinência estão bem próximos de 0 ou 1. A medida que $m \rightarrow \infty$, a partição se torna completamente *fuzzy*, ou seja, o grau de pertinência é igual para todas as classes ($u_{ij} = 1/k$) (PEDRYCZ, 2005).

O intervalo de valores aceitáveis para m no FCM está entre 1 e 30 (BEZDEK; EHRLICH; FULL, 1984). Tipicamente, o valor de m é ajustado para 2 (XU; II, 2009). Entretanto, se existe um conjunto de dados de teste disponível, a melhor estratégia para selecionar m é a experimental. Para a maioria dos dados, o intervalo de $1,5 \leq m \leq 3,0$ gera bons resultados (BEZDEK; EHRLICH; FULL, 1984). Na falta de estudos sobre a influência do expoente *fuzzy* no comportamento do FGMM e do CFGMM, assumimos o intervalo $1,3 \leq m \leq 3,2$ para executar uma primeira análise.

A Figura 20 exibe o gráfico do Índice Rand em função de m para o resultado da segmentação da imagem sintética com 2 classes considerando o intervalo de $1,3 \leq m \leq 3,2$. O gráfico mostra que o desempenho do FCM e do CFGMM, em relação à medida Índice Rand, decaiu em função do aumento do valor de m para a imagem com 2 classes. O desempenho do FGMM permaneceu constantes em função da variação do valor de m .

O FCM obteve o valor máximo do Índice Rand de 0,84 em $m = 2,3$, o FGMM obteve o valor máximo do Índice Rand de 0,93 em $m = 1,8$ e o CFGMM obteve o valor máximo do Índice Rand de 0,95 em $m = 1,3$. O resultado da segmentação da imagem pelo KM e pelo GMM resultou nos valores do Índice Rand de 0,83 e 0,81, respectivamente. Como o KM e o GMM não têm o parâmetro m , os seus resultados não constam nos gráficos. O CFGMM apresentou o melhor resultado, seguido do FGMM, do FCM, do KM e do GMM, considerando a medida Índice Rand.

A Figura 21 mostra a imagem sintética com 2 classes e exemplos dos resultados da sua segmentação pelo KM, FCM com $m = 2,3$, GMM, FGMM com $m = 1,8$ e CFGMM com $m = 1,3$.

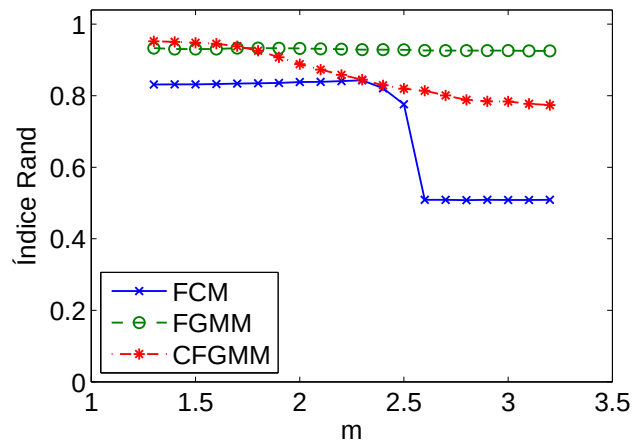


Figura 20 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 2 classes.

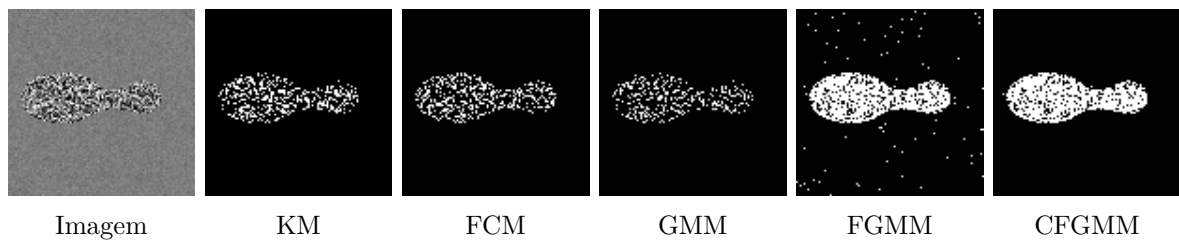


Figura 21 – Segmentação da imagem sintética de 2 classes pelo KM, FCM, GMM, FGMM e CFGMM.

A Figura 23 exibe o gráfico do Índice Rand em função de m para o resultado da segmentação da imagem sintética com 3 classes e SNR 8 considerando o intervalo de $1,3 \leq m \leq 3,2$. O gráfico mostra que o desempenho do FCM e do CFGMM, em relação à medida Índice Rand, também permaneceu constante em função da variação do valor de m para a imagem com 3 classes. O desempenho do FGMM apresentou um leve pico em $m = 1,3$ e permaneceu constante em relação aos outros valores de m .

O FCM obteve o valor máximo do Índice Rand de 0,90 em $m = 1,7$, o FGMM obteve o valor máximo do Índice Rand de 0,86 em $m = 1,3$ e o CFGMM obteve o valor máximo do Índice Rand de 0,99 em $m = 1,3$. O valor do Índice Rand para o resultado da segmentação da imagem pelo KM e GMM foi de 0,90 e 0,90, respectivamente. O CFGMM também apresentou o melhor resultado, seguido do GMM e do FGMM, empatados, e do FCM, considerando a medida Índice Rand. A Figura 22 mostra a imagem sintética com 3 classes e com relação sinal ruído (SNR) 8 e exemplos dos resultados da sua segmentação pelo KM, FCM com $m = 1,7$, GMM, FGMM com $m = 1,3$ e CFGMM com $m = 1,3$.

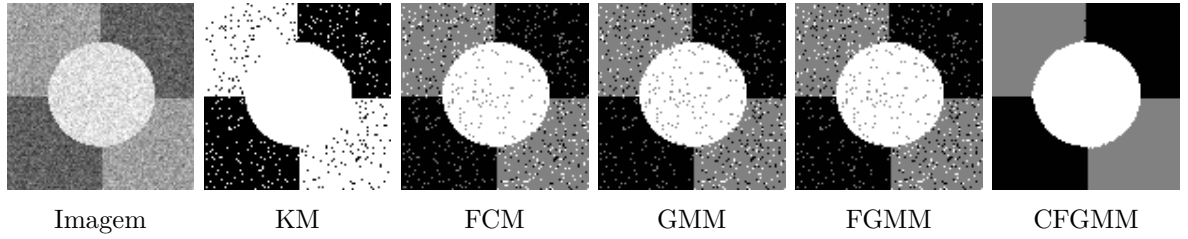


Figura 22 – Segmentação da imagem sintética de 3 classes e SNR 8 pelo KM, FCM, GMM, FGMM e CFGMM.

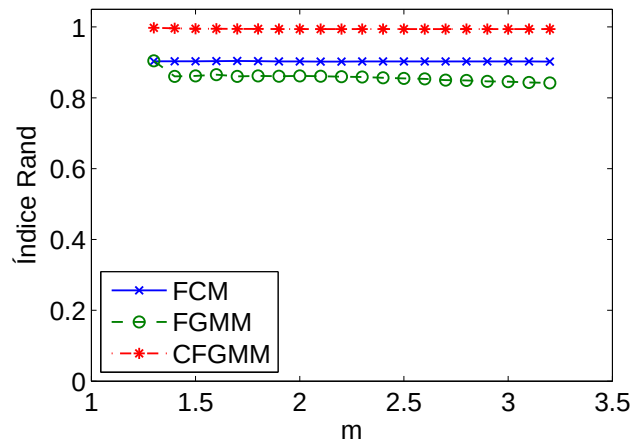


Figura 23 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 3 classes.

A Figura 24 exibe o gráfico do Índice Rand em função de m para o resultado da segmentação da imagem sintética com 5 classes e SNR 8 considerando o intervalo de

$1,3 \leq m \leq 3,2$. O gráfico mostra que houve uma pequena variação em função de m no desempenho do FGMM e do CFGMM, em relação ao Índice Rand. O CFGMM apresentou discretos picos em $m = 1,8, 2,0, 2,2$ e $2,4$. O desempenho de FGMM cresceu rapidamente no intervalo $1,3 \leq m \leq 2,5$ e se estabilizou com $m \leq 2,5$ apresentando um discreto pico em $m = 2,7$. O desempenho do FCM permaneceu constante.

O FCM obteve o valor máximo do Índice Rand de 0,81 em $m = 1,4$, o FGMM obteve o valor máximo do Índice Rand de 0,81 em $m = 2,9$ e o CFGMM obteve o valor máximo do Índice Rand de 0,93 em $m = 2,0$. O resultado da segmentação da imagem pelo KM e GMM resultou nos valores de Índice Rand, 0,81 e 0,81, respectivamente. O CFGMM apresentou o melhor resultado seguido do FCM, do GMM e do FGMM, considerando a medida Índice Rand.

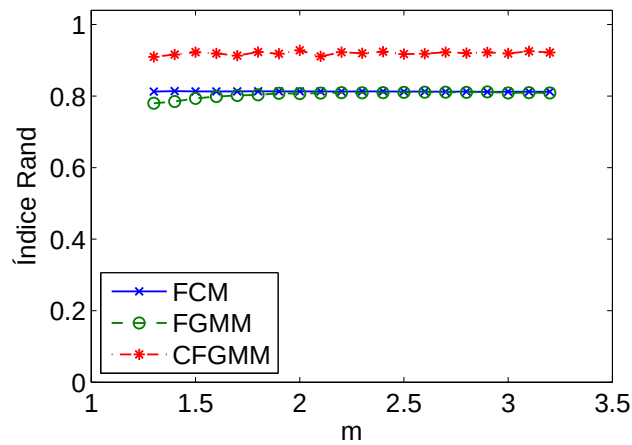


Figura 24 – Gráfico do Índice Rand em função da variação do valor de m na segmentação da imagem sintética de 5 classes.

A Figura 25 mostra a imagem sintética com 5 classes e com relação sinal ruído (SNR) 8 e exemplos dos resultados da sua segmentação pelo KM, FCM com $m = 1,4$, GMM, FGMM com $m = 2,9$ e CFGMM com $m = 2,0$.

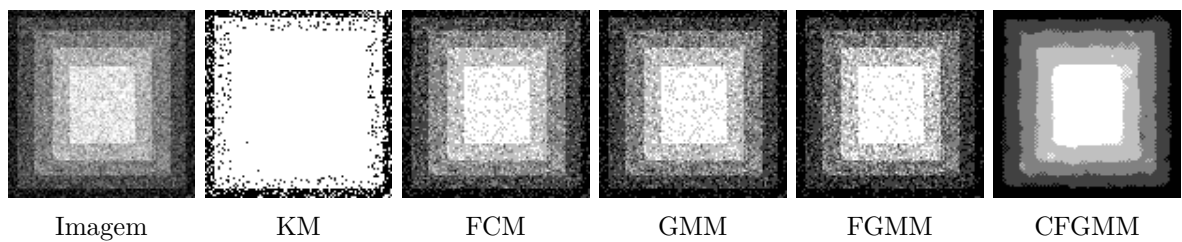


Figura 25 – Segmentação da imagem sintética de 5 classes e SNR 8 pelo KM FCM, GMM, FGMM e CFGMM.

5.2.1.3 Comportamento do modelo em relação ao ruído

O comportamento das duas versões do modelo proposto, o CFGMM-B e o CFGMM-F, e dos métodos clássicos de agrupamento, KM, FCM, GMM e FGMM foi avaliado na segmentação das imagens sintéticas de 3 e 5 classes com diferentes relações sinal ruído (SNR) da Figura 8. O limiar de convergência adotado para todos os algoritmos foi $Th < 10^{-2}$ e a vizinhança V_i^4 foi adotada pelo CFGMM. Os algoritmos foram executados 30 vezes para cada imagem do conjunto e o FGMM foi usado na inicialização dos parâmetros.

A Figura 26 mostra as imagens sintéticas com 3 classes e com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB e exemplos dos respectivos resultados da segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F. Os expoentes *fuzzy* adotados no FCM, no FGMM, no CFGMM-B e no CFGMM-F foram todos iguais, $m = 1,3$.

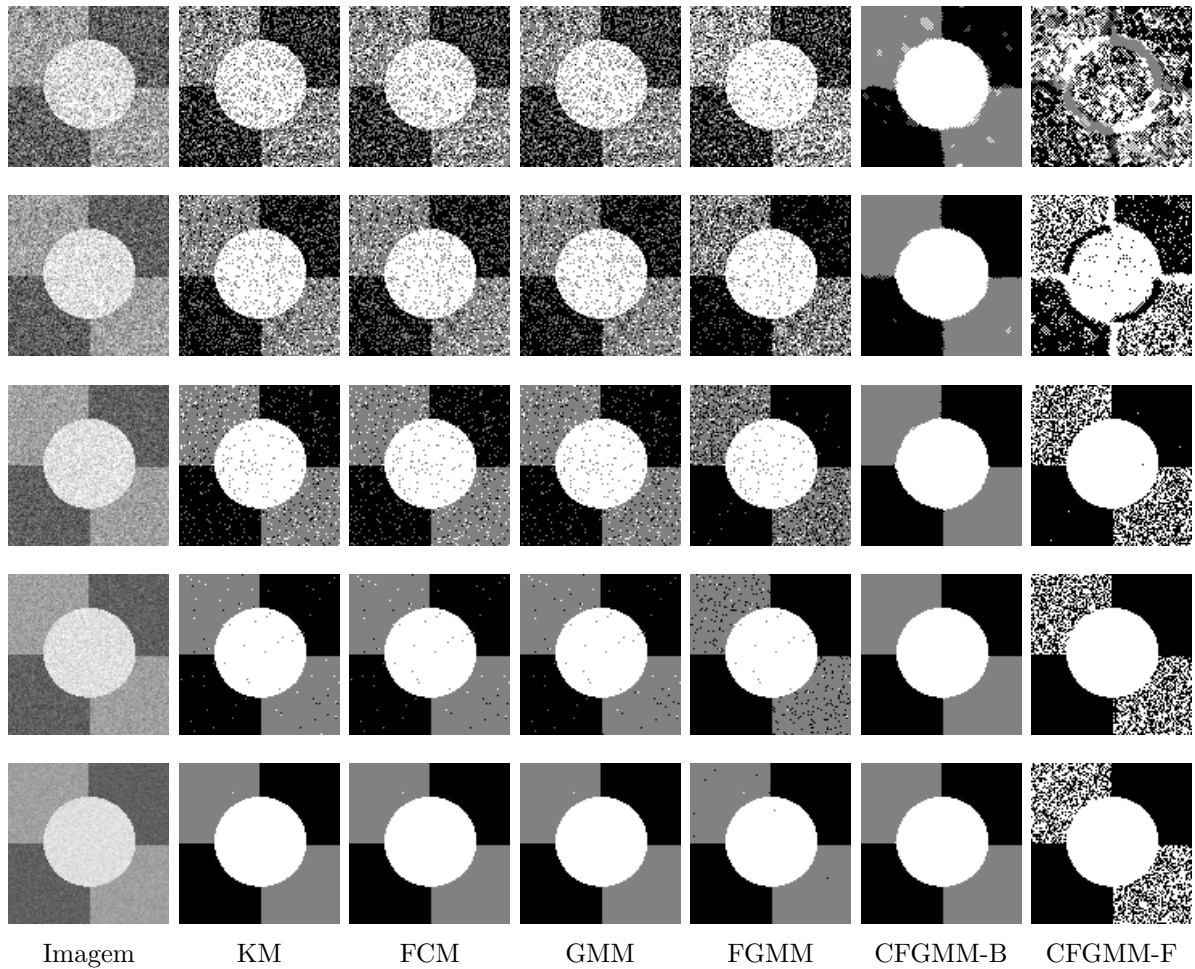


Figura 26 – Segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

O desempenho de todos os métodos piora com o aumento do nível de ruído da imagem. Embora o FGMM tenha apresentado pior desempenho entre todos os métodos pontuais de agrupamento, o CFGMM-B apresentou melhor resultado para imagens com alto nível de ruído, SNR=1, 4, 8 e 12, se equiparando ao KM, ao FCM, ao GMM e ao FGMM na segmentação da imagem com menor nível de ruído, SNR = 16. O FGMM obteve melhor resultado que o KM, o FCM e o GMM apenas na segmentação da imagem com SNR=1. O KM, o FCM e o GMM obtiveram resultados próximos em todos cenários, com o GMM se destacando nos cenários com nível de ruído intermediário, SNR = 8 e 12. O CFGMM-F apresentou o pior resultado em todos os cenários.

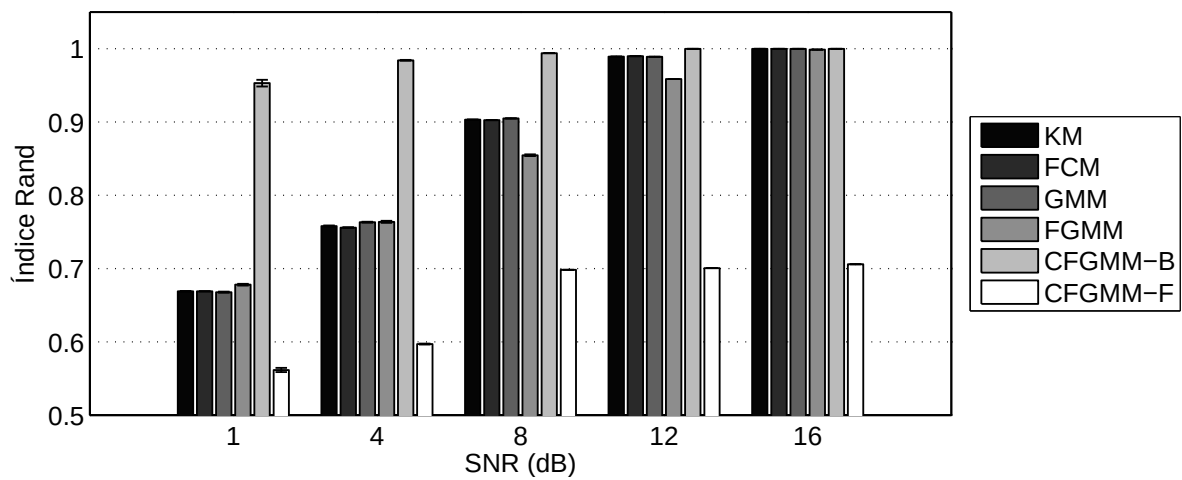


Figura 27 – Média do Índice Rand na segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

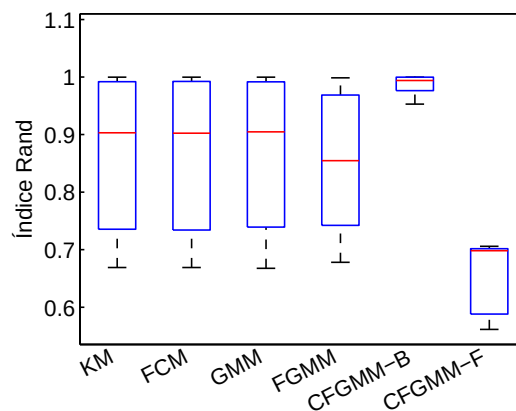


Figura 28 – Gráfico de caixa do Índice Rand na segmentação das imagens sintéticas de 3 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

O gráfico de caixa (*boxplot*) do resultado do Índice Rand na segmentação das imagens sintéticas com 3 classes sobre todas as relações sinal ruído, que está exibidos na Figura 28, ilustra o desempenho geral dos métodos avaliados. A marca central na caixa é a mediana, as bordas inferior e superior da caixa correspondem ao primeiro e terceiro quartis, respectivamente, as barras horizontais superior e inferior mostram os dados mais extremos sem considerar os *outliers*. Os *outliers*, quando presentes são mostrados individualmente.

A CFGMM-B obteve o melhor resultado geral e teve o desempenho mais constante na segmentação das imagens com diferentes níveis de ruído, indicado no gráfico pela menor distância entre o primeiro e o terceiro quartil. O KM, FCM e GMM tiveram resultados gerais similares, seguido do FGMM. O CFGMM-F obteve o pior resultado geral.

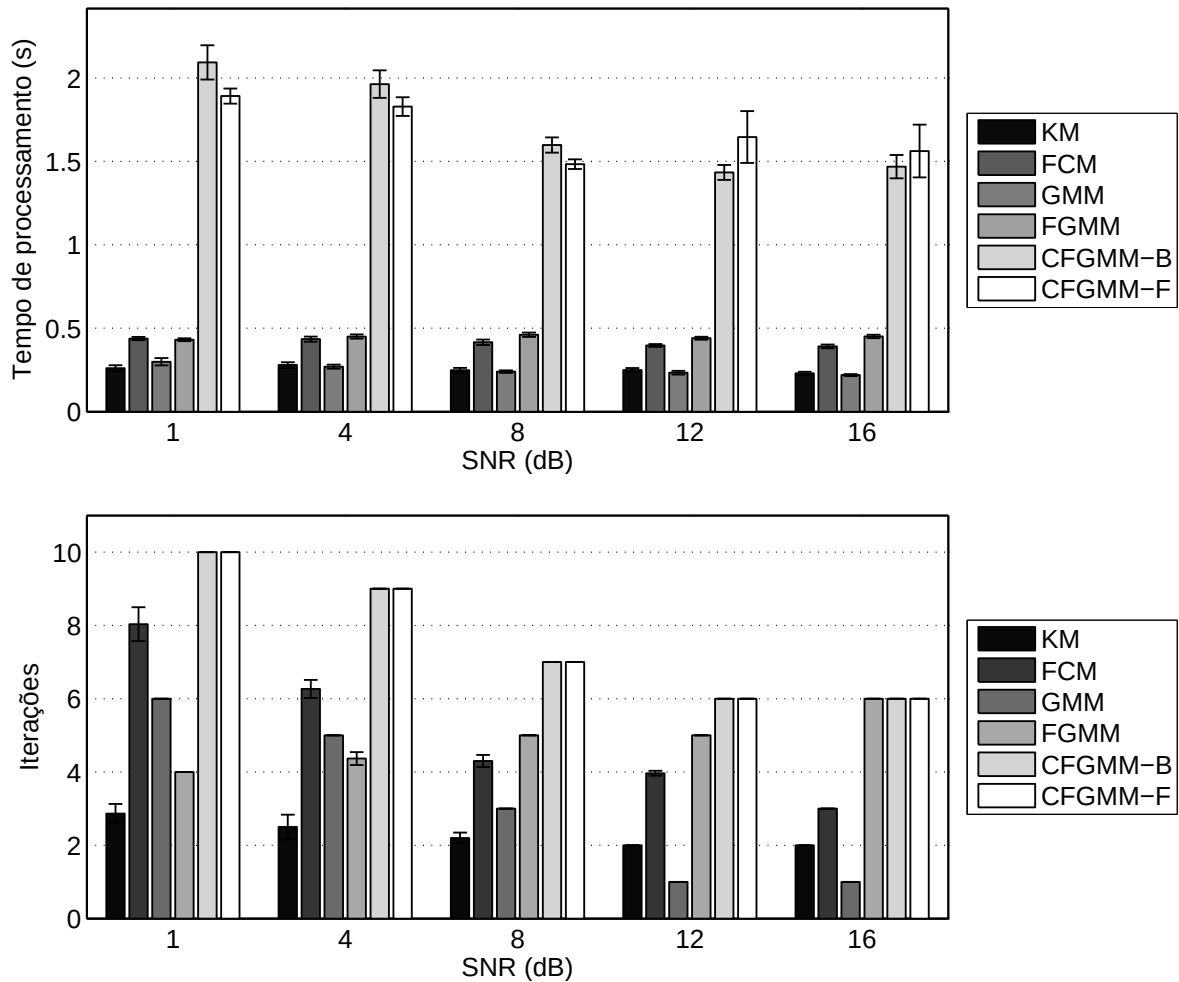


Figura 29 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das imagens sintéticas de 3 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

Os gráficos da Figura 29 exibem a média do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as imagens sintéticas com 3 classes e todas as relações sinal ruído. Todos os métodos foram executados no mesmo computador

sob as mesmas condições. O gráfico mostra uma tendência de diminuição do tempo de processamento do CFGMM ao diminuir o nível de ruído. O KM obteve menor tempo de processamento em todos os cenários, seguido do GMM, FCM e do FGMM. O menor tempo de processamento do KM em relação aos outros métodos está de acordo com o conhecimento comum de que ele é o método mais básico entre os testados por usar o menor número de parâmetros para representar cada grupo, somente o vetor de médias. Em todos os cenários o CFGMM-B e o CFGMM-F tiveram os maiores tempo de processamento. Entretanto, esse resultado já era esperado pois ambos possuem mais parâmetros do que os outros métodos. Além disso, há a incorporação do contexto no cálculo dos parâmetros do modelo.

Os gráficos de caixa das médias do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as imagens sintéticas com 3 classes sobre todas as relações sinal ruído, que estão exibidos na Figura 30, ilustram o desempenho geral dos métodos avaliados. No geral, o KM obteve o menor tempo de processamento e número de iterações, seguido do GMM, do FCM, do FGMM, sendo, este último, o com menor variação na segmentação das imagens com diferentes níveis de ruído. O CFGMM-B e o CFGMM-F obtiveram os maiores tempo de processamento e número de iterações.

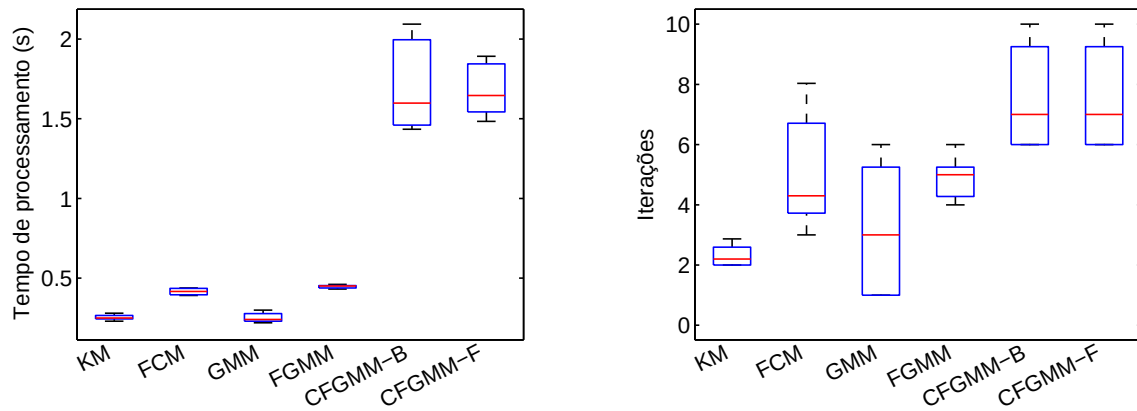


Figura 30 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das imagens sintéticas de 3 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

A Figura 31 mostra as imagens sintéticas com 5 classes e com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB e exemplos dos respectivos resultados da segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F. Os expoentes *fuzzy* adotados no FCM, no FGMM e CFGMM foram $m=1,4$, $m=2,9$ e $m=2,0$, respectivamente.

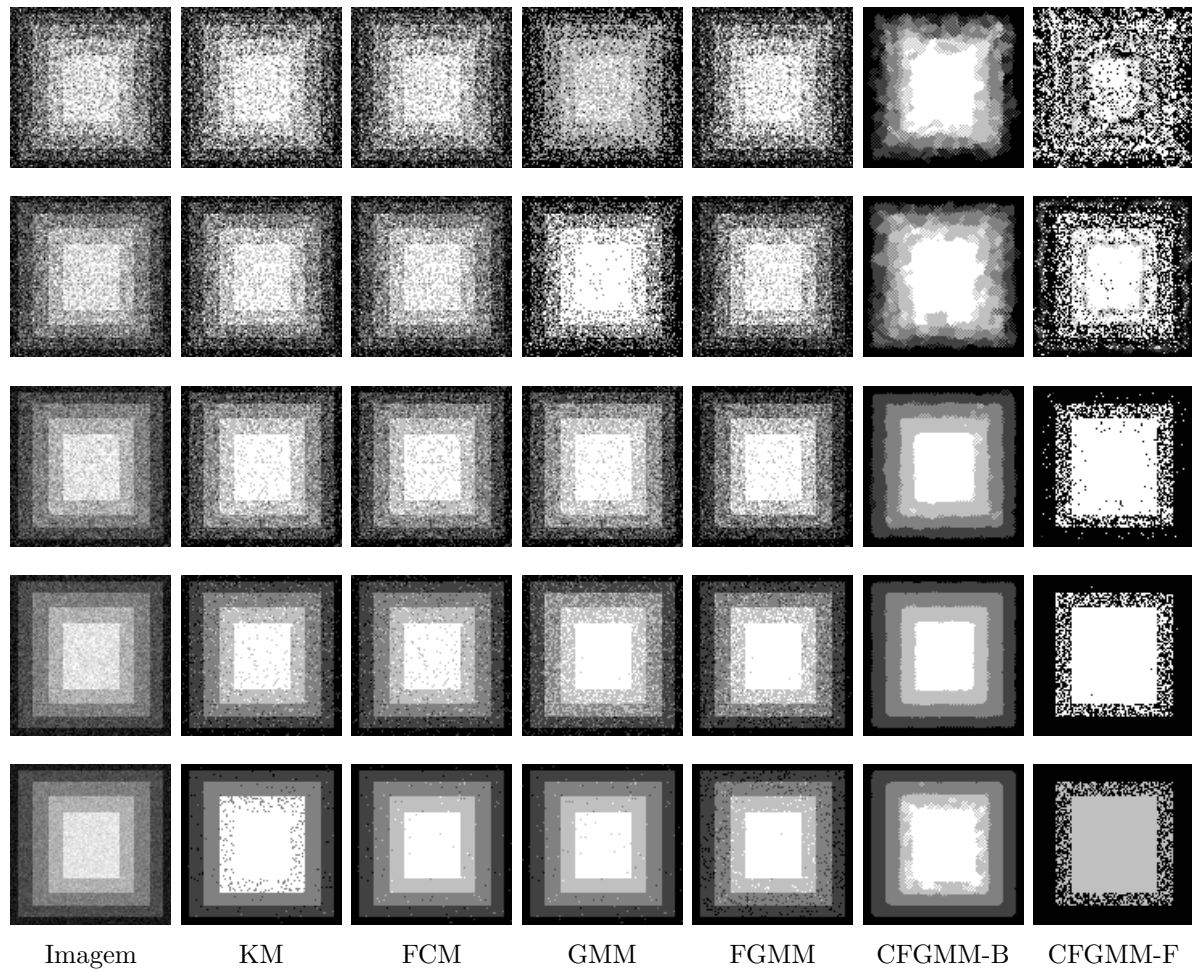


Figura 31 – Segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

O gráfico da Figura 32 exibe a média do Índice Rand para o resultado da segmentação das imagens sintéticas com 5 classes e com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB. O CFGMM-B apresentou melhor resultado para imagens com alto nível de ruído, SNR=1, 4, 8 e 12, se equiparando ao KM, ao FCM e ao GMM na segmentação da imagem com menor nível de ruído, SNR = 16. O KM e o FCM obtiveram resultados próximos em todos os cenários. O FCM e o FGMM apresentaram resultados próximo nas imagens com maiores ruídos mas o FCM se destacou em relação ao FGMM em imagens com pouco ruído. O CFGMM-F apresentou o pior resultado em todos os cenários.

O gráfico de caixa dos resultados do Índice Rand na segmentação das imagens sintéticas com 5 classes e todas as relações sinal ruído, que está exibido na Figura 33, ilustra o desempenho geral dos métodos avaliados. A CFGMM-B obteve o melhor resultado geral seguido do FCM. O KM, o GMM e o FGMM tiveram resultados gerais similares, no

entanto o FGMM teve desempenho mais constante que o GMM e o KM na segmentação das imagens com diferentes níveis de ruído. O CFGMM-F obteve o pior resultado geral mas teve o desempenho mais constante entre todas as técnicas.

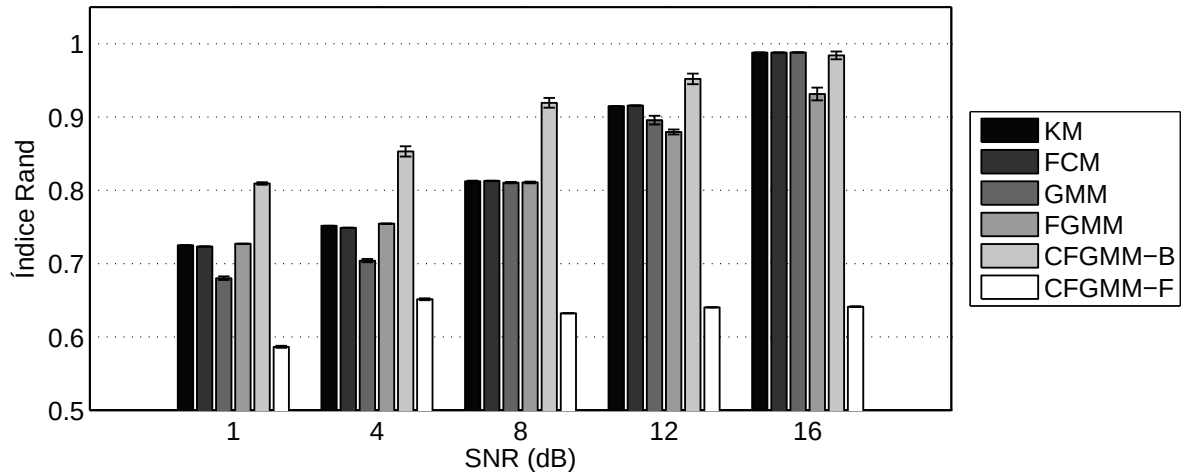


Figura 32 – Média do Índice Rand na segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

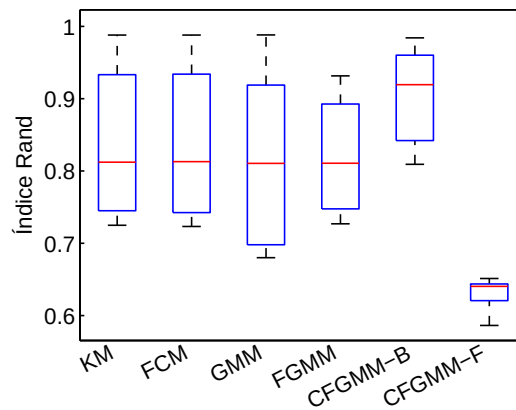


Figura 33 – Gráfico de caixa do Índice Rand na segmentação das imagens sintéticas de 5 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

Os gráficos da Figura 34 exibem a média do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as imagens sintéticas com 5 classes e todas as relações sinal ruído. Todos os métodos foram executados no mesmo computador sob as mesmas condições. O gráfico mostra uma tendência de diminuição do tempo de processamento ao diminuir o nível de ruído. O KM obteve o menor tempo de processamento médio na segmentação de imagens com maiores níveis de ruído, SNR = 1 e 4, enquanto que

o GMM obteve menor tempo de processamento na segmentação de imagens com menores níveis de ruído, $\text{SNR} = 8, 12$ e 16 . Em geral, o FGMM teve o terceiro melhor desempenho, seguido do FCM. Em todos os cenários o CFGMM-B e o CFGMM-F tiveram os maiores tempo de processamento.

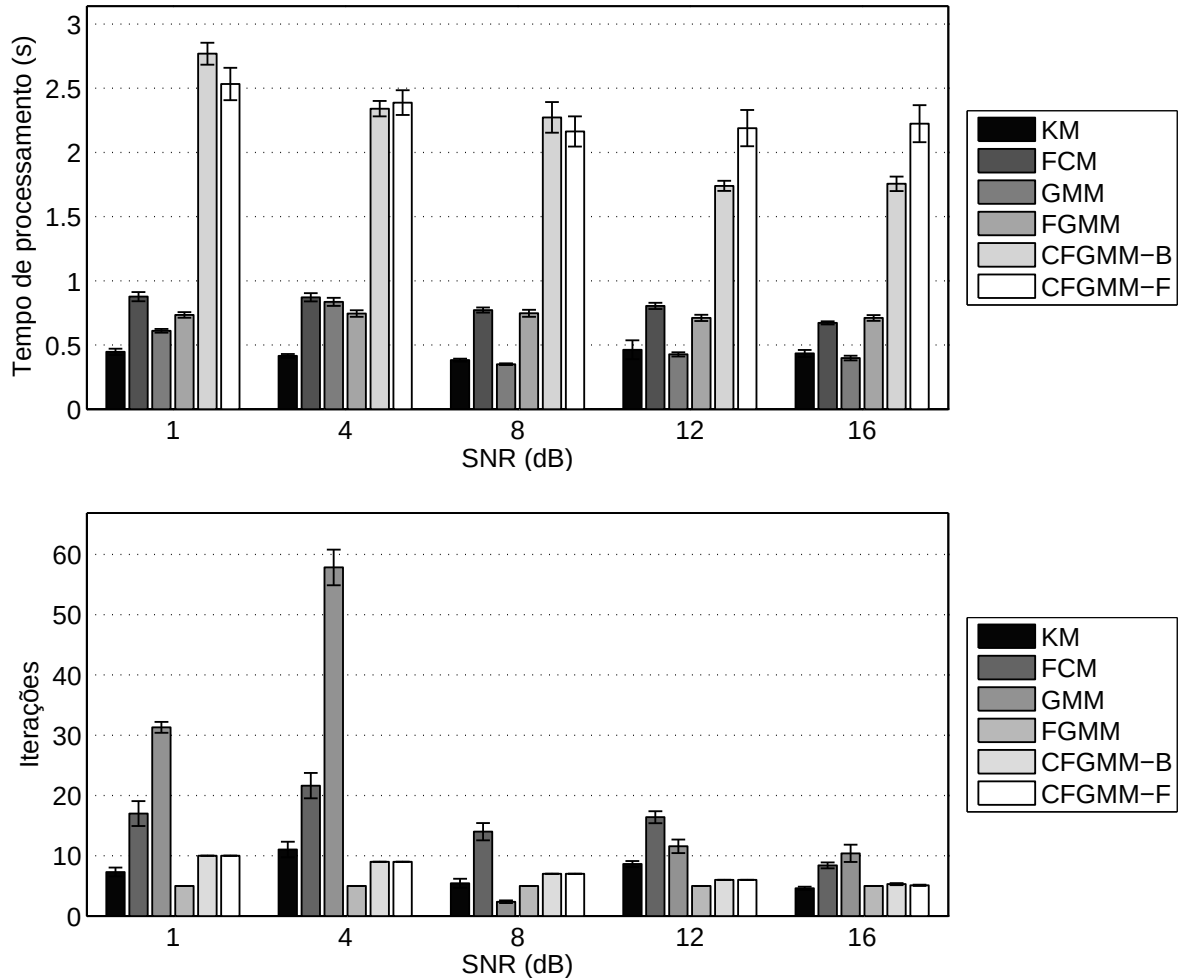


Figura 34 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das imagens sintéticas de 5 classes e SNR 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

Os gráficos de caixa das médias do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as imagens sintéticas com 5 classes sobre todas as relações sinal ruído, que estão exibidos na Figura 35, ilustram o desempenho geral dos métodos avaliados. No geral, embora a mediana do tempo de processamento do KM tenha ficado próxima da do GMM, o KM teve o resultado mais constante e o menor número de iterações. Embora o FGMM tenha apresentado um número de iterações menor e com menos variação do que o FCM, a mediana do tempo de processamento dos dois métodos ficaram muito próximas. Este fato é justificado pelo FCMM ter menos parâmetros do que FGMM para calcular a cada iteração. O CFGMM-B e o CFGMM-F

obtiveram os maiores tempo de processamento, mas tiveram baixo número de iterações quando comparados ao FCM e ao GMM.

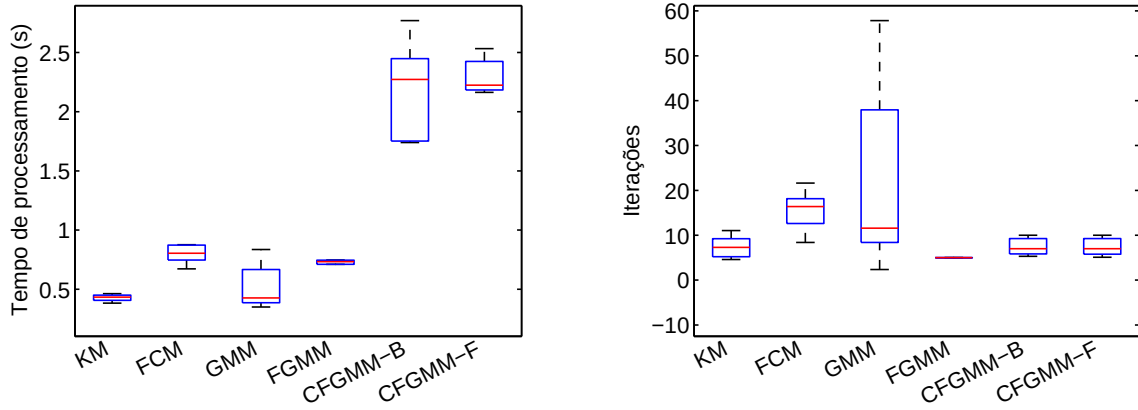


Figura 35 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das imagens sintéticas de 5 classes e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

5.2.2 Análise do modelo com o banco de imagem *BrainWeb*

O Modelo de Mistura de Gaussianas *Fuzzy* Contextual também foi analisado na segmentação de imagens de cérebro da *BrainWeb simulated brain database*. As imagens utilizadas são fatias de uma imagem tri-dimensional simulada de ressonância magnética baseada em um modelo anatômico de cérebro normal.

Os experimentos com imagens sintéticas mostraram que o processo de rotulagem de dados baseado na regra de Bayes, adotada na versão CFGMM-B, se mostrou muito mais adequado à segmentação das imagens, segundo o Índice Rand, do que o uso da partição *fuzzy*, adotada na versão CFGMM-F. Logo, só estudamos nesta seção o desempenho do CFGMM-B.

Inicialmente, o comportamento do CFGMM, com processo de rotulagem baseado na regra de Bayes, em função do limiar de convergência e do expoente *fuzzy* foi avaliado. Em seguida, foi realizado um estudo comparativo com o CFGMM-B e as técnicas clássicas KM, FCM, GMM e FGMM.

5.2.2.1 Limiar de convergência

A convergência do algoritmo e a qualidade da segmentação ao longo das iterações foram testadas na fatia 100 da imagem com nível de ruído 5% (Figura 36). A Figura 37 exibe o comportamento do Índice Rand junto com o valor da função objetivo ao longo das iterações no processo de segmentação da imagem. O critério de parada usado foi o limiar de convergência, $Th < 10^{-4}$, com número máximo de 50 iterações.

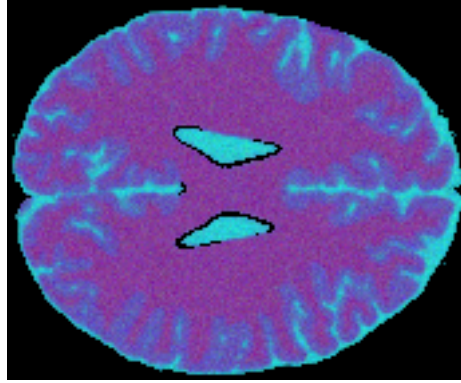


Figura 36 – Fatia 100 da imagem com nível de ruído 5%.

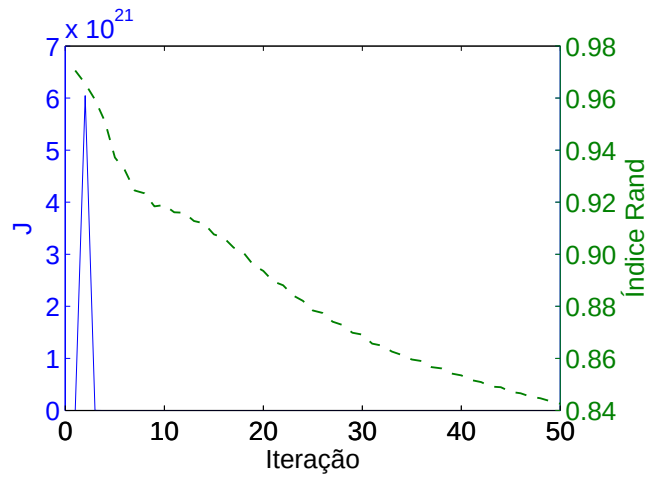


Figura 37 – Gráfico da função objetivo J e do Índice Rand ao longo das iterações na segmentação da fatia 100 da imagem de cérebro com nível de ruído 5% pelo CFGMM. J está representado nos gráficos com a linha sólida enquanto que o Índice Rand estão representados com a linha tracejada.

A qualidade da segmentação da imagem, avaliada pelo Índice Rand, é decrescente ao longo das iterações mesmo após o valor de J estabilizado. A Figura 38 mostra o resultado da segmentação da imagem a cada 5 iterações até o algoritmo atingir o critério de parada. Ao longo das iterações as regiões da imagem vão ficando mais homogêneas (com menos ruído) mas as bordas das regiões vão ficando borradas. Os sulcos cerebrais, preenchidos pelo fluido cérebro-espinhal, passam a ser eliminados da imagem segmentada.

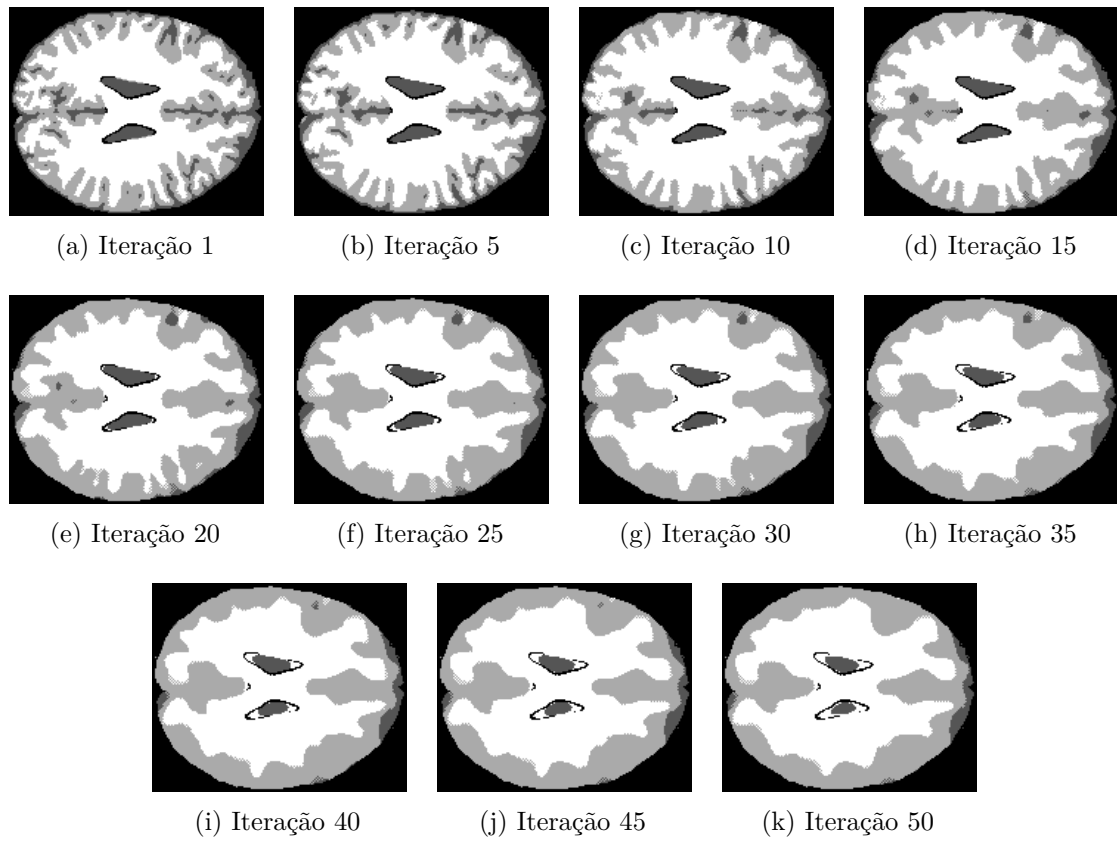


Figura 38 – Resultado da segmentação da imagem pelo CFGMM ao longo das iterações.

5.2.2.2 Expoente *fuzzy*

O valor médio do Índice Rand, em função de m , para o resultado das segmentações das 133 fatias da imagem com nível de ruído 5% considerando o intervalo de $1,3 \leq m \leq 3,2$, estão exibidos nos gráficos da Figura 39.

O gráfico mostra que o valor médio do Índice Rand apresentou pouca variação em função da variação do valor de m no caso do FCM e do CFGMM. Em relação ao FGMM, os valores do Índice Rand apresentam uma queda inicial e depois se estabilizam com o aumento de m . O resultado da segmentação da imagem pelo GMM resultou no valor de Índice Rand 0,94 e foi o melhor desempenho entre as 4 técnicas. Como o KM e o GMM não têm o parâmetro m , os seus resultados não constam nos gráficos. O FCM obteve o valor máximo do Índice Rand de 0,90 em $m = 1,3$, FGMM obteve o valor máximo do Índice Rand de 0,84 em $m = 1,3$ e o CFGMM obteve o valor máximo do Índice Rand de 0,89 em $m = 1,7$.

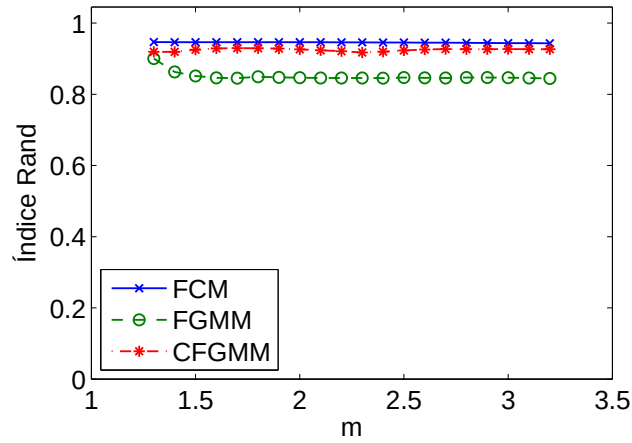


Figura 39 – Gráfico do Índice Rand em função da variação do valor de m na segmentação das fatias da imagem de cérebro com nível de ruído 5% .

A Figura 40 mostra a fatia 100 da imagem com nível de ruído 5% e exemplos dos resultados da sua segmentação pelo KM, FCM com $m = 1,3$, GMM, FGMM com $m = 1,3$ e CFGMM com $m = 1,7$.

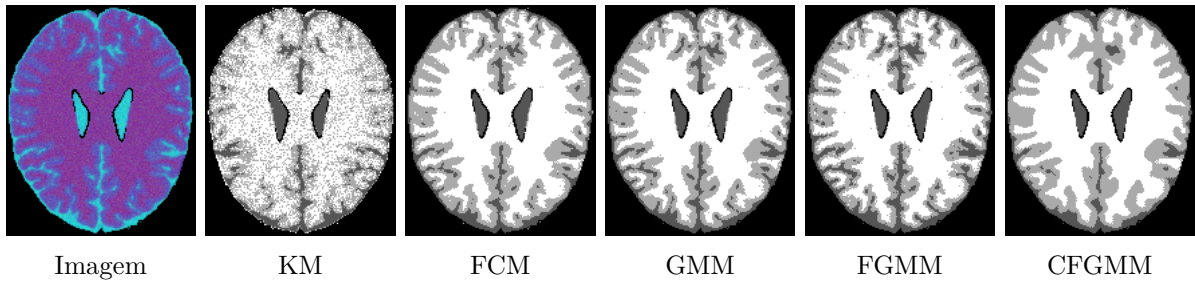


Figura 40 – Segmentação da fatia 100 da imagem de cérebro com nível de ruído 5% pelo KM, FCM, GMM, FGMM e CFGMM.

5.2.2.3 Comportamento do modelo em relação ao ruído

O comportamento do CFGMM-B e dos métodos clássicos de agrupamento, KM, FCM, GMM e FGMM foi avaliado na segmentação das imagens do banco de imagens *BrainWeb* com diferentes níveis de ruído. Devido ao seu ruim desempenho na segmentação das imagens sintéticas, o CFGMM-F não é avaliado nesta seção.

Um critério de parada precoce foi adotado para evitar a eliminação de detalhes da imagem, como as regiões correspondentes aos sulcos cerebrais, preenchidos pelo fluido cérebro-espinhal, na imagem segmentada. O limiar de convergência adotado para todos os algoritmos foi $Th < 10^{-2}$, o expoente *fuzzy* adotado no FCM, no FGMM e no CFGMM-B foi, respectivamente $m = 1,3$, $m = 1,3$ e $m = 1,7$ e a vizinhança V_i^4 foi adotada pelo CFGMM. Os

algoritmos foram executados 30 vezes para cada imagem do conjunto e o FGMM foi usado na inicialização dos parâmetros.

A média do Índice Rand para o resultado da segmentação das imagens com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9% está exibida no gráfico da Figura 41. Além da média do Índice Rand alcançada por cada método estudado, o gráfico exibe o intervalo de confiança da média com nível de confiança de 95% assumindo que a variância populacional é igual a variância amostral (DEGROOT; SCHERVISH, 2011). Os valores do Índice Rand (eixo vertical do gráfico) variam de 0 a 1. Entretanto, para uma melhor visualização do resultado a figura só exibe um corte no eixo que varia de 0,5 a 1.

A qualidade da segmentação da imagem pelo FCM, avaliada pelo Índice Rand, apresentou melhor resultado para imagens com baixo nível de ruído, 0%, 1% e 3%, enquanto que o GMM teve um melhor desempenho na segmentação das imagens com nível de ruído maiores, 5%, 7% e 9%. O CFGMM-B obteve melhor resultado que o FGMM em 4 dos 6 cenários.

O gráfico de caixa (*boxplot*) das médias do Índice de Rand das segmentações das imagens sobre todos os níveis de ruído, que está exibido na Figura 42, ilustra o desempenho geral dos métodos avaliados. A marca central na caixa é a mediana, as bordas inferior e superior da caixa correspondem ao primeiro e terceiro quartis, respectivamente, as barras horizontais superior e inferior mostram os dados mais extremos sem considerar os *outliers*. Os *outliers*, quando presentes, são mostrados individualmente.

O FCM obteve o melhor resultado geral. As medianas do FCM e do CFGMM-B estão próximas, no entanto o terceiro quartil do FCM está bem acima do CFGMM-B indicando um melhor resultado. Por outro lado, o CFGMM-B teve o desempenho mais constante na segmentação das imagens com diferentes níveis de ruído, indicado no gráfico da Figura 42 pela menor distância entre o primeiro e o terceiro quartil, embora não tenha obtido o melhor desempenho em nenhum cenário específico. O KM obteve o pior resultado geral.

A Figura 43 mostra a fatia 100 da imagem com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9% e exemplos dos respectivos resultados da segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B. Analisando o resultado do CFGMM na segmentação de cada região da imagem do cérebro detalhadamente, é possível perceber que a matéria branca e o *background* foram bem segmentados mas que o fluido cérebro-espinhal foi confundido com matéria cinzenta. Esse resultado é decorrência do fato do fluido cérebro-espinhal ocupar uma região de menor área na imagem e, na sua maioria, corresponder às reentrâncias na matéria cinzenta que são longas mas estreitas, aumentando a proporção de número de *pixels* de borda em relação às outras regiões da imagem. Por consequência, o CFGMM acaba eliminando a região de fluido pois ele tende a borrar as bordas das regiões e eliminar os detalhes finos das regiões nas imagens, como foi visto nos experimentos da Seção 5.2.1.

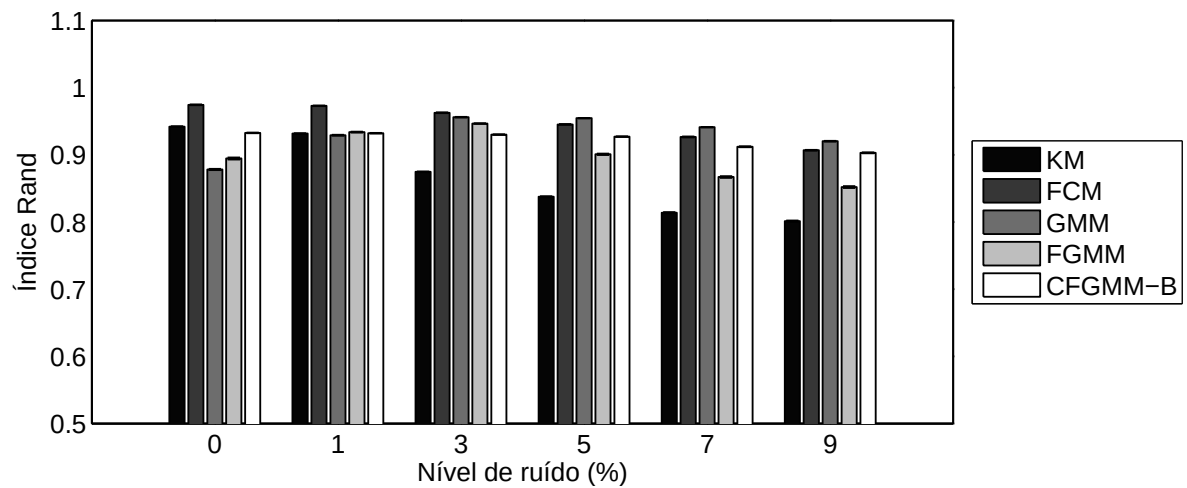


Figura 41 – Média do Índice Rand na segmentação das fatias das imagens de cérebro com níveis de ruído 0, 1, 3, 5, 7 e 9 pelo KM, FCM, GMM, FGMM e CFGMM-B.

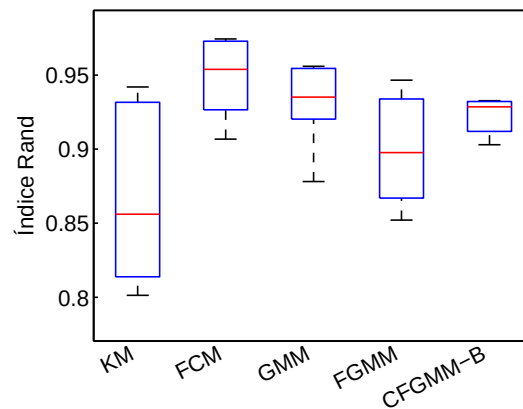


Figura 42 – Gráfico de caixa do Índice Rand na segmentação das fatias das imagens de cérebro e todos os níveis de ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.

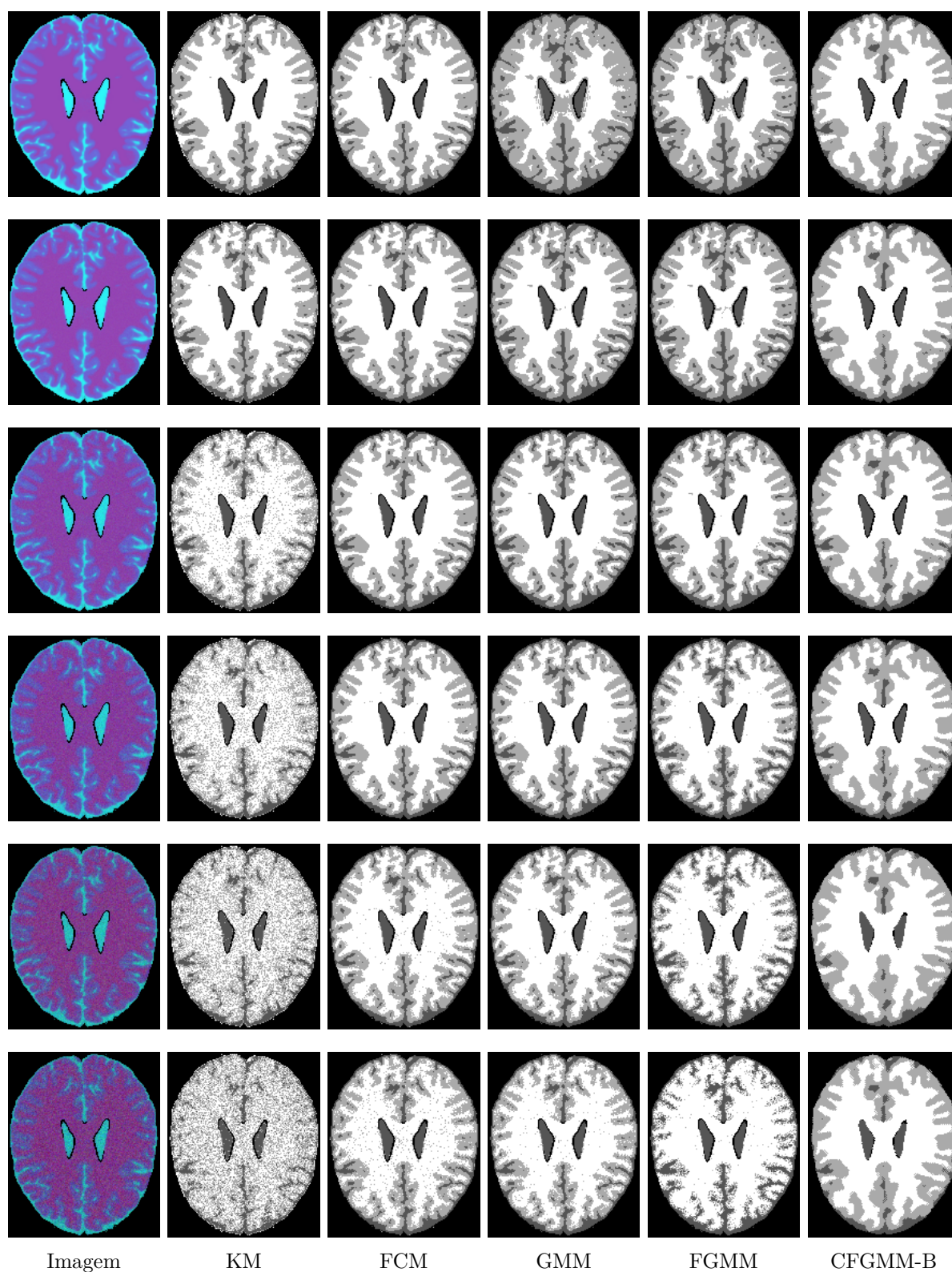


Figura 43 – Segmentação da fatia 100 da imagem de cérebro com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9%, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.

Os gráficos da Figura 44 exibem a média do tempo de processamento, em segundos, e o número de iterações necessários para segmentar as imagens de cérebro da *BrainWeb* com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9%. Todos os métodos foram executados no mesmo computador sob as mesmas condições. O KM e o FCM obtiveram os menores tempo de processamento em todos os cenários, mesmo quando o número de iterações do FCM foi alto, níveis de ruído 0% e 1%. O GMM e o FGMM tiveram o segundo e terceiro menor tempo de processamento. Em todos os cenários o CFGMM-B teve o maior tempo de processamento.

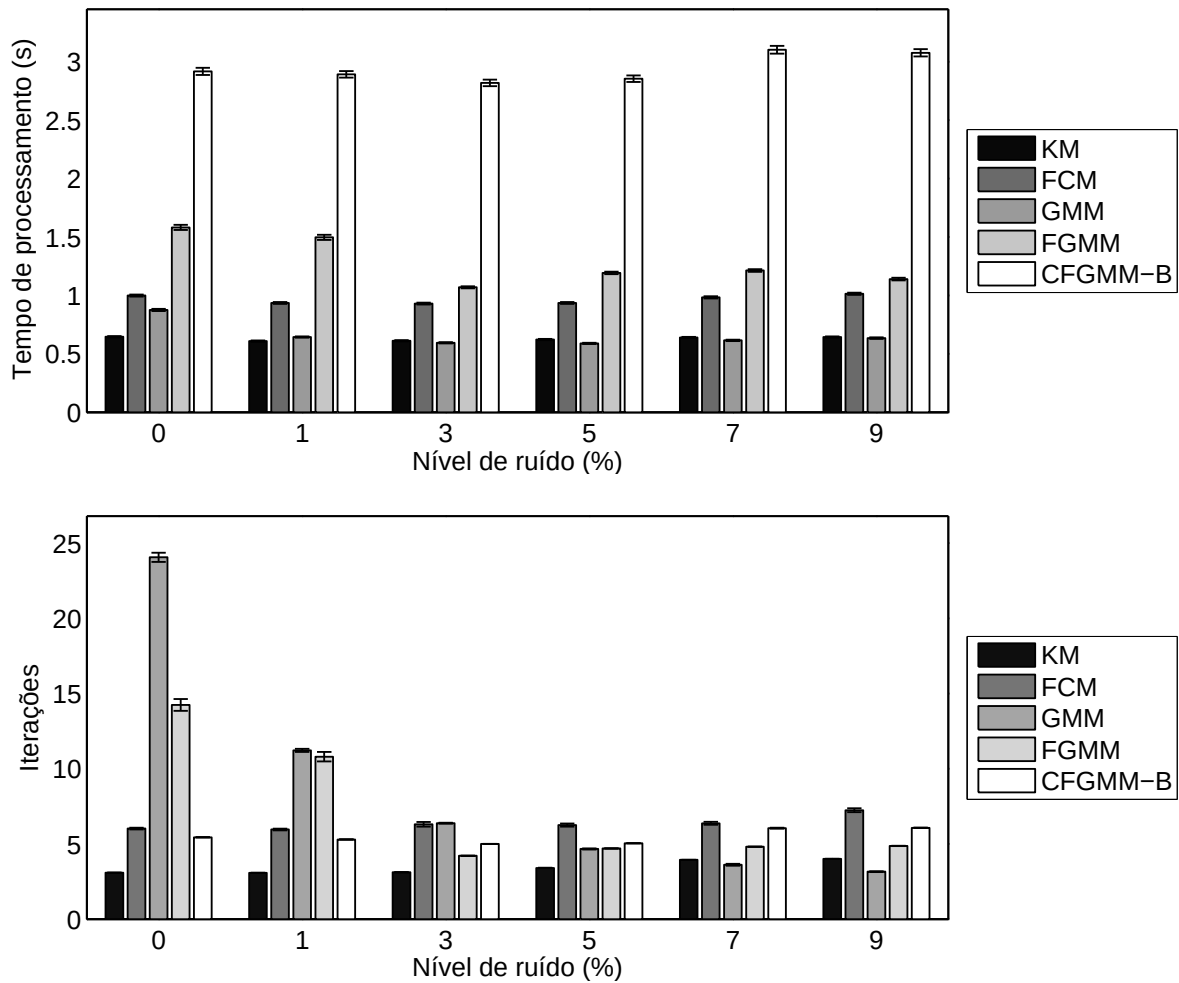


Figura 44 – Média do tempo de processamento (acima) e do número de iterações (abaixo) na segmentação das fatias das imagens de cérebro com níveis de ruído 0%, 1%, 3%, 5%, 7% e 9% pelo KM, FCM, GMM, FGMM, CFGMM-B e CFGMM-F.

Os gráficos de caixa das médias do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as fatias das imagens de cérebro sobre todos os níveis de ruído, que estão exibidos na Figura 45, ilustram o desempenho geral dos métodos avaliados.

O KM obteve o menor tempo de processamento, seguido do GMM, do FCM, do FGMM e CFGMM-B. A mediana do número de iterações do KM foi a menor e o resto das técnicas ficaram próximas, no entanto, o GMM e o FGMM tiveram uma maior variação do resultado indicado pelas maiores distâncias entre o primeiro e o terceiro quartis.

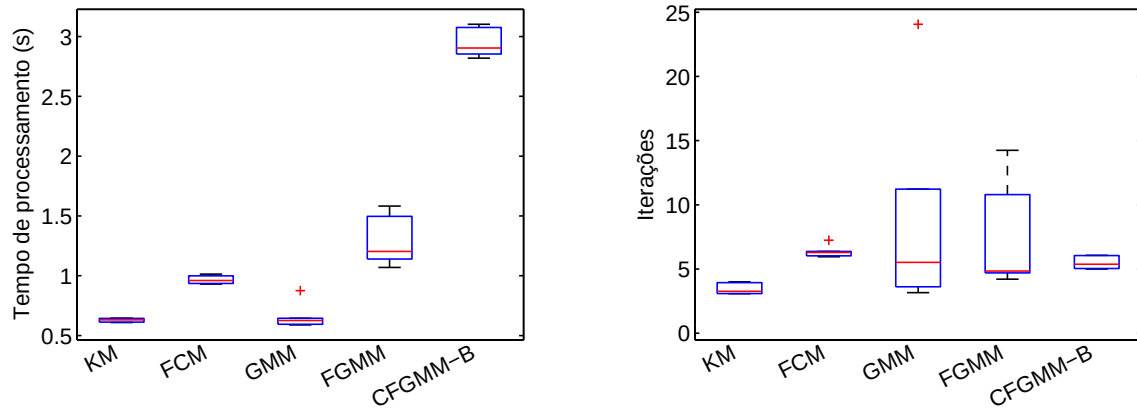


Figura 45 – Gráfico de caixa do tempo de processamento (esquerda) e do número de iterações (direita) na segmentação das fatias das imagens de cérebro e todos os níveis de ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.

Com o objetivo de estudar o comportamento das técnicas na segmentação de imagens sob efeito de ruídos mais significativos, o ruído Gaussiano com relação sinal ruído 1, 4, 8, 12 e 16 dB foi adicionado às fatias da imagem do banco de imagens *BrainWeb* assim como descrito da Subseção 5.1.1.1.

A média do Índice Rand para o resultado da segmentação das imagens com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB está exibida no gráfico da Figura 46. O desempenho de todos os métodos, avaliado pelo Índice PR, piora com o aumento do nível de ruído da imagem. O CFGMM-B apresentou melhor resultado para imagens com alto nível de ruído, SNR = 1, 4 e 8, e o GMM obteve o melhor desempenho para imagens com menor nível de ruído, SNR = 12 e 16. O CFGMM-B apresentou melhor resultado que o FGMM em todos os cenários experimentados. O FGMM apresentou o pior desempenho em todos os cenários com exceção do SNR = 1.

O gráfico de caixa dos resultados médios do Índice Rand na segmentação das fatias das imagens de cérebro e todas as relações sinal ruído, que está exibido na Figura 47, ilustra o desempenho geral dos métodos avaliados.

A mediana do CFGMM-B está ligeiramente acima do FCM e GMM embora o terceiro percentil destes esteja acima do CFGMM-F. O CFGMM-B teve o desempenho mais constante na segmentação das imagens com diferentes níveis de ruído, indicado no gráfico pela menor distância entre o primeiro e o terceiro quartil. O FGMM obteve o pior resultado geral.

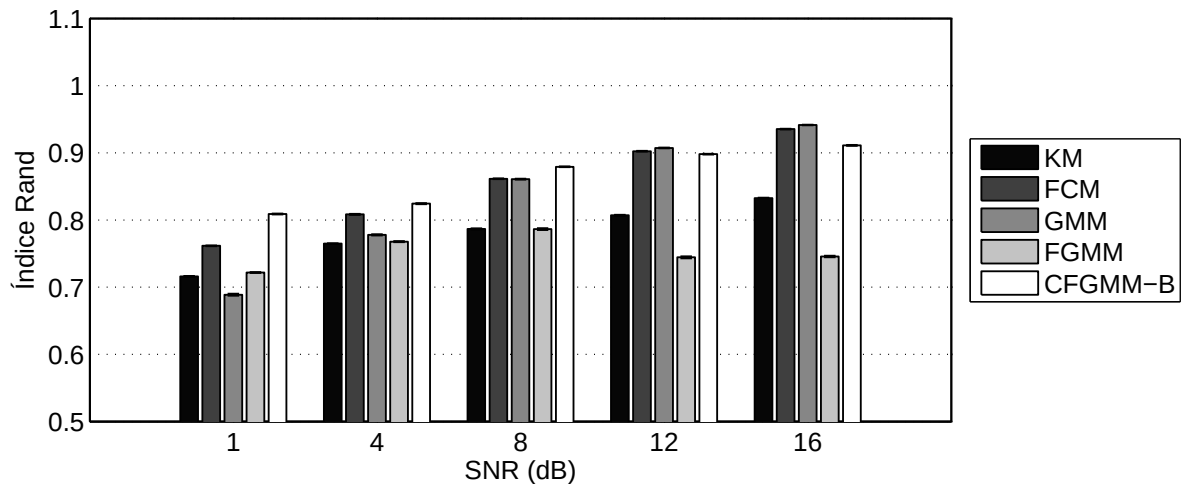


Figura 46 – Média do Índice Rand na segmentação das fatias das imagens de cérebro com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB pelo KM, FCM, GMM, FGMM e CFGMM-B.

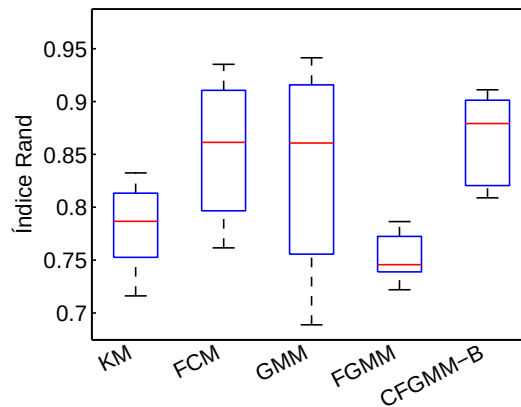


Figura 47 – Gráfico de caixa do Índice Rand na segmentação das fatias das imagens de cérebro e todas as relações sinal ruído pelo KM, FCM, GMM, FGMM e CFGMM-B.

A Figura 48 mostra a fatia 100 da imagem com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB e exemplos dos respectivos resultados da segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.

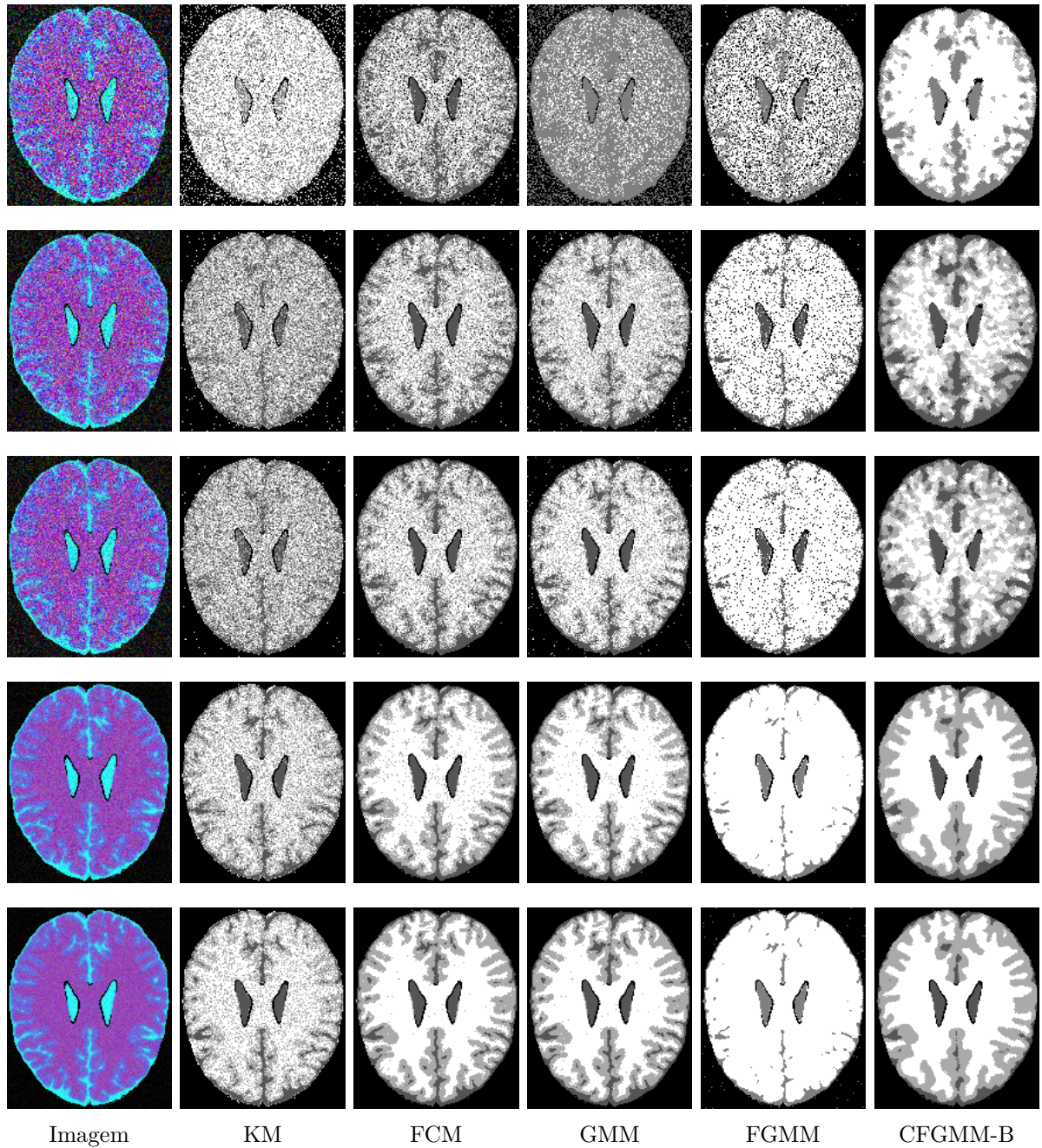


Figura 48 – Segmentação da fatia 100 da imagem de cérebro com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM, e CFGMM-B.

5.2.3 Análise do modelo com o banco de imagem de *Berkeley*

Inicialmente, uma amostra de imagens da *Berkeley database* foi selecionada para estudarmos o comportamento do CFGMM na segmentação baseada em regiões de imagens naturais. Imagens com diferentes texturas foram escolhidas para a avaliação. A Figura 49 é composta pelas 30 imagens selecionadas e seus respectivos números de classes, que foram

definidos por Nikou *et al.* (NIKOU; GALATSANOS; LIKAS, 2007) com o objetivo de maximizar o Índice PR.

Na Seção 5.2.3.1 foi realizada a análise do comportamento do Índice PR junto com o valor da função objetivo ao longo das iterações no processo de segmentação baseada em região da imagem pelo CFGMM-B com o objetivo de avaliar o melhor critério de parada a ser adotado. Na Seção 5.2.3.2 foi a vez de avaliar a influência do expoente *fuzzy* na segmentação das imagens da Figura 49 pelo CFGMM-B.

Em seguida, foi realizado um estudo comparativo com o CFGMM-B e as técnicas clássicas KM, FCM, GMM e FGMM. Como os experimentos com imagens sintéticas mostraram que o processo de rotulagem de dados baseado na regra de Bayes, adotada na versão CFGMM-B, se mostrou muito mais adequado à segmentação das imagens do que o uso da partição *fuzzy*, adotada na versão CFGMM-F. A medida de desempenho adotada nesse estudo é o Índice PR, que é a derivação do Índice Rand adequado a situações nas quais existem mais de um *ground truth*.

Por fim, foi conduzido um experimento com todas as 200 imagens de teste da *Berkeley database* e o *benchmark* de segmentação BSDS500 da Berkeley foi utilizado para calcular as medidas de desempenho *F-measure* e Índice PR para todas as técnicas testadas. O *F-measure* é usado pelo BSDS na análise de detecção de borda das regiões das imagens e o Índice de PR é usado como medida de desempenho da segmentação baseada em região.

5.2.3.1 Limiar de convergência

A convergência do algoritmo e a qualidade da segmentação ao longo das iterações foram testadas com a imagem #183055 exibida da Figura 50 junto com seus *ground truths*, obtidos através da segmentação manual por diferentes indivíduos.

A Figura 51 exibe o comportamento do Índice PR junto com o valor da função objetivo ao longo das iterações no processo de segmentação da imagem. O critério de parada usado foi o limiar de convergência, $Th < 10^{-4}$, com número máximo de 50 iterações.

A qualidade da segmentação da imagem #183055, avaliada pelo Índice PR, é crescente ao longo das iterações apresentando uma aceleração nas primeiras iterações e mantendo um crescimento constante após o valor de J estabilizado.

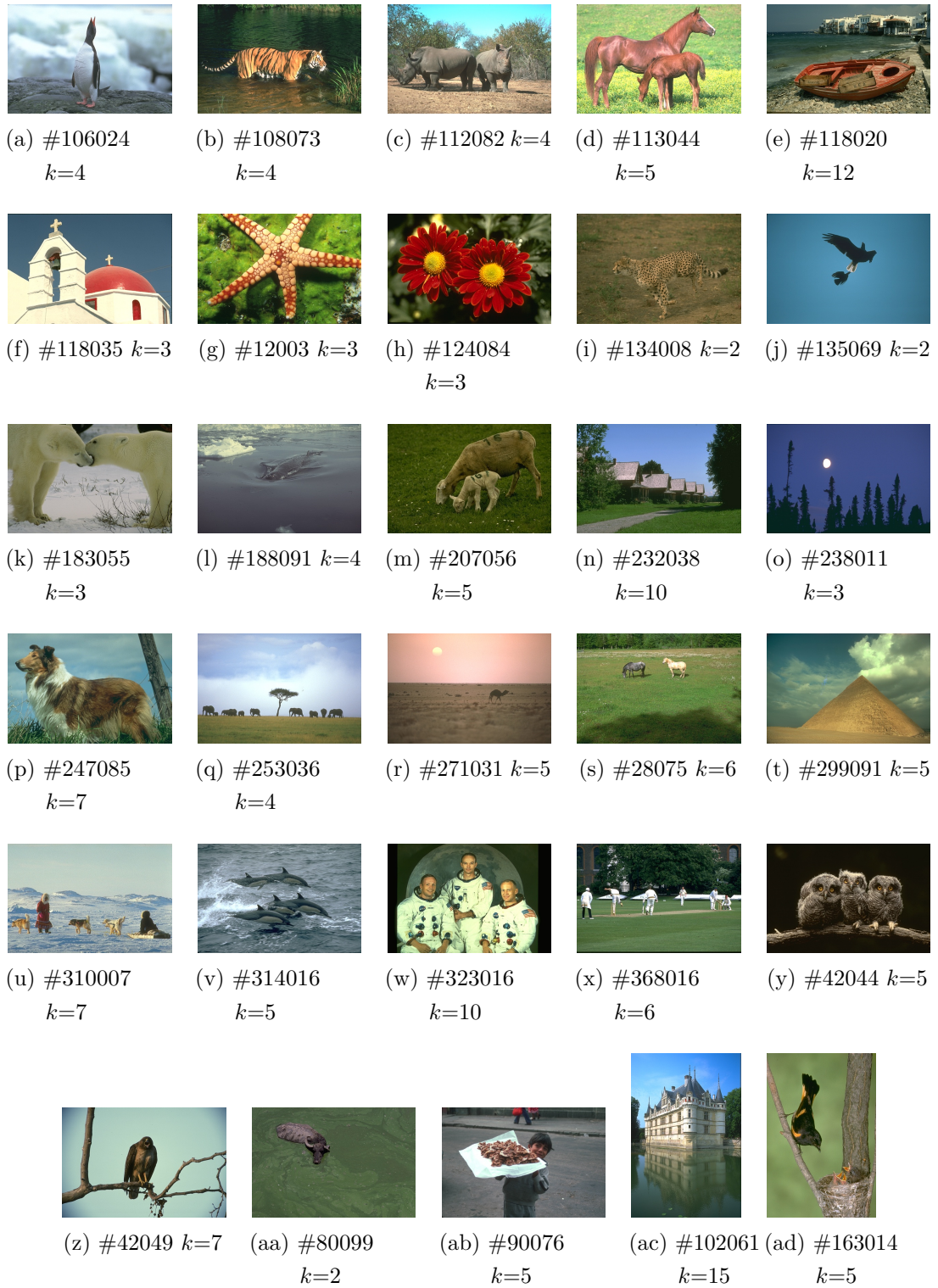
Figura 49 – Amostra de imagens da *Berkeley database*.



Figura 50 – Figura #183055 e seus *ground truths*.

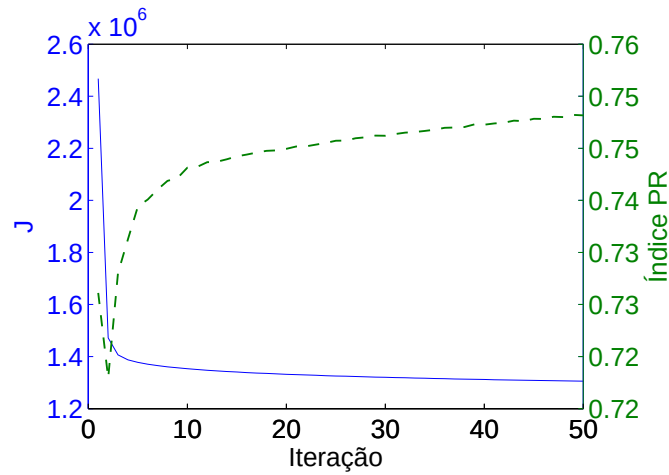


Figura 51 – Gráfico da função objetivo J e do Índice PR ao longo das iterações na segmentação da imagem #183055 pelo CFGMM. J está representado no gráfico com a linha sólida enquanto que o Índice PR está representado com a linha tracejada.

A Figura 52 mostra o resultado da segmentação da imagem a cada 5 iterações até o algoritmo atingir o critério de parada. Ao longo das iterações as regiões da imagem vão ficando mais homogêneas (com menos ruído) mas as bordas das regiões vão ficando borradas.

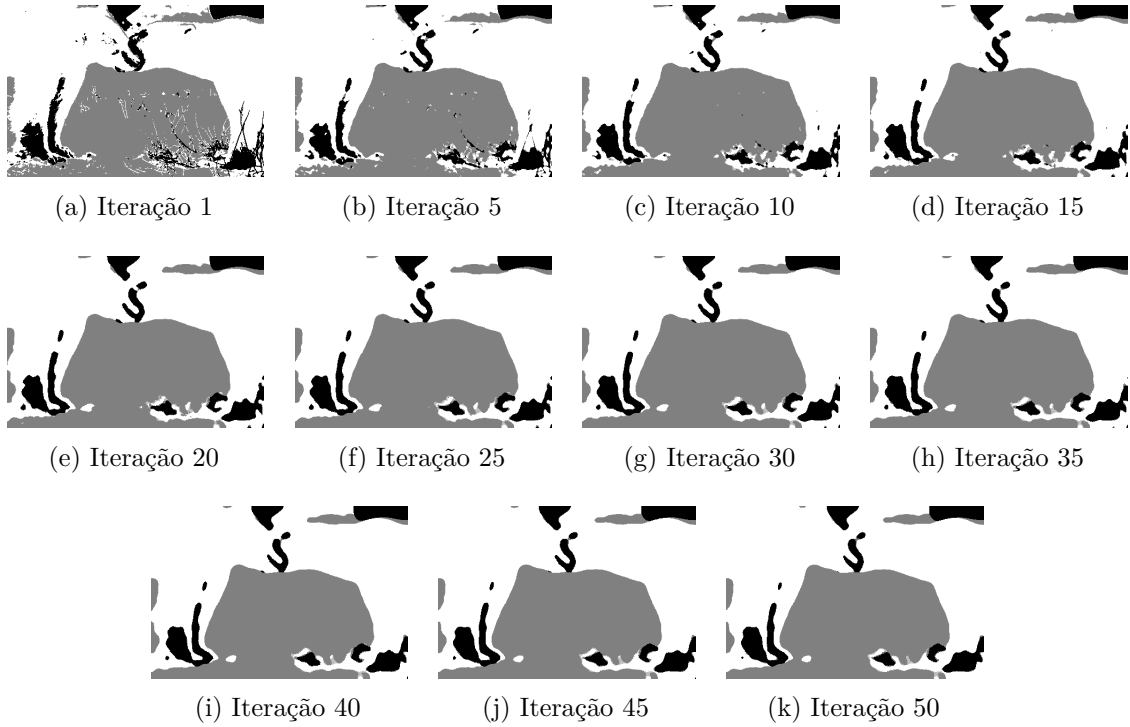


Figura 52 – Resultado da segmentação da imagem pelo CFGMM ao longo das iterações.

5.2.3.2 Expoente *fuzzy*

A Figura 53 exibe os gráficos do valor médio do Índice PR, em função de m , para o resultado das segmentações das 30 imagens da *Berkeley database* da Figura 49 considerando o intervalo de $1, 3 \leq m \leq 3, 2$.

O Gráfico 53 mostra que o desempenho do FCM, do FGMM e do CFGMM, em relação ao Índice PR, varia bastante em função de m . O resultado da segmentação da imagem pelo KM e pelo GMM resultou nos valores de Índice PR, 0,67 e 0,70, respectivamente. Como o KM e o GMM não têm o parâmetro m , os seus resultados não constam no gráfico. O FCM obteve o valor máximo do Índice PR de 0,68 em $m = 1,3$, o FGMM obteve o valor máximo do Índice PR de 0,66 em $m = 1,4$ e o CFGMM obteve o valor máximo do Índice PR de 0,69 em $m = 1,9$, sendo, este último, o melhor resultado entre as 3 técnicas.

A Figura 54 mostra a imagem #183055 e exemplos dos resultados da sua segmentação pelo KM, FCM com $m = 1,3$, GMM, FGMM com $m = 1,4$ e CFGMM com $m = 1,9$.

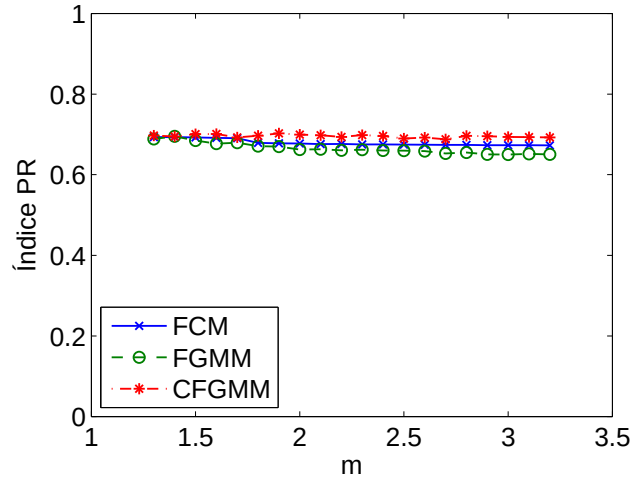


Figura 53 – Gráfico do valor médio do Índice PR em função da variação do valor de m na segmentação de imagens da *Berkeley database*.

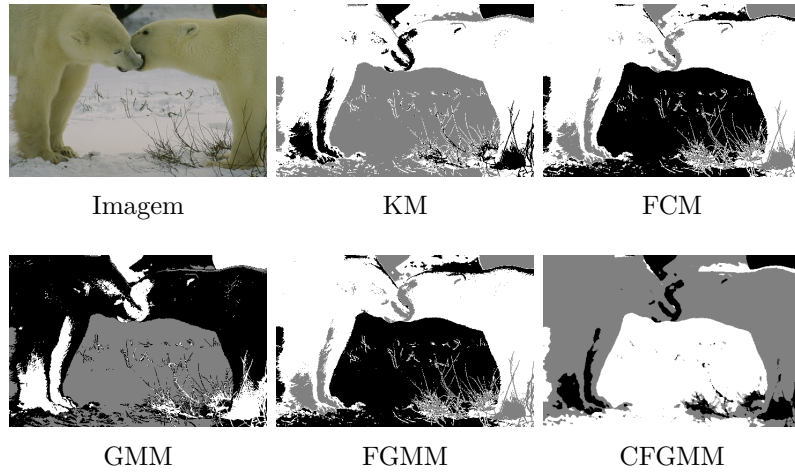


Figura 54 – Segmentação da imagem #183055 pelo KM, FCM, GMM, FGMM e CFGMM.

5.2.3.3 Comportamento do modelo em relação ao ruído

O desempenho do CFGMM-B foi comparado ao dos métodos clássicos de agrupamento, o FCM, o GMM e o FGMM, usando as 30 imagens da *Berkeley database* listadas na Figura 49. Os algoritmos foram executados 30 vezes para cada imagem do conjunto. Devido ao desempenho ruim, segundo o Índice Rand, do CFGMM-F na segmentação das imagens sintéticas, o mesmo não foi avaliado nesta seção.

A média do Índice PR está exibido no gráfico da Figura 55. O limiar de convergência adotado para todos os algoritmos foi $Th < 10^{-2}$ e os expoentes *fuzzy* adotados no FCM, no FGMM e no CFGMM-B foram $m = 1,3$, $m = 1,4$ e $m = 1,9$, respectivamente, e a vizinhança V_i^4 foi adotada pelo CFGMM. Além da média do Índice PR alcançada por cada método estudado, o gráfico exibe o intervalo de confiança da média com nível de confiança de

95% assumindo que a variância populacional é igual a variância amostral (DEGROOT; SCHERVISH, 2011). Os valores do Índice PR (eixo vertical do gráfico) variam de 0 a 1. Entretanto, para uma melhor visualização do resultado, a figura só exibe um corte no eixo que varia de 0,6 a 0,8. Embora o CFGMM-B tenha apresentado melhor resultado em relação ao FGMM, ele não apresentou diferença significativa em relação ao FCM e ao GMM. O KM apresentou o pior resultado.

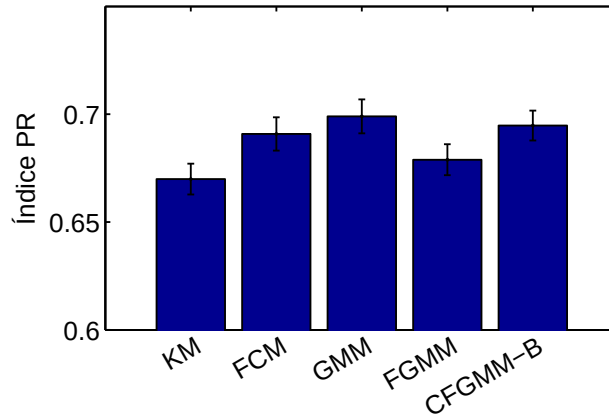


Figura 55 – Gráfico do valor médio do Índice PR na segmentação de imagens da *Berkeley database*.

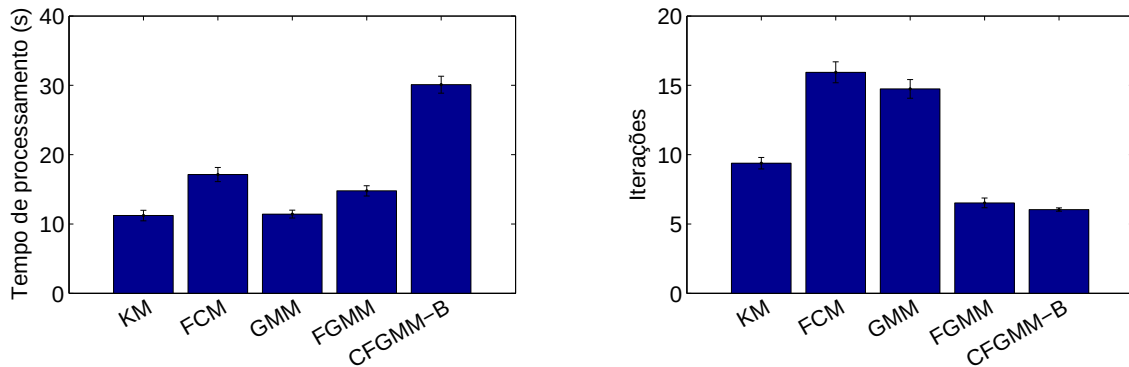


Figura 56 – Média do tempo de processamento e do número de iterações na segmentação de imagens da *Berkeley database*.

Os gráficos da Figura 56 exibem a média do tempo de processamento, em segundos, e do número de iterações necessários para segmentar as imagens da *Berkeley database* listadas na Figura 49. Todos os métodos foram executados no mesmo computador sob as mesmas condições. O KM obteve menor tempo de processamento. O GMM teve o segundo melhor desempenho, seguido do FGMM e do FCM. O CFGMM-B teve o maior tempo de processamento mesmo necessitando de menos iterações para a sua convergência.

A Figura 57 mostra as imagens #80099, #113040, #135069, #183055 e #238011 com exemplos dos resultados das suas segmentações pelo KM, FCM, GMM, FGMM e CFGMM-B.

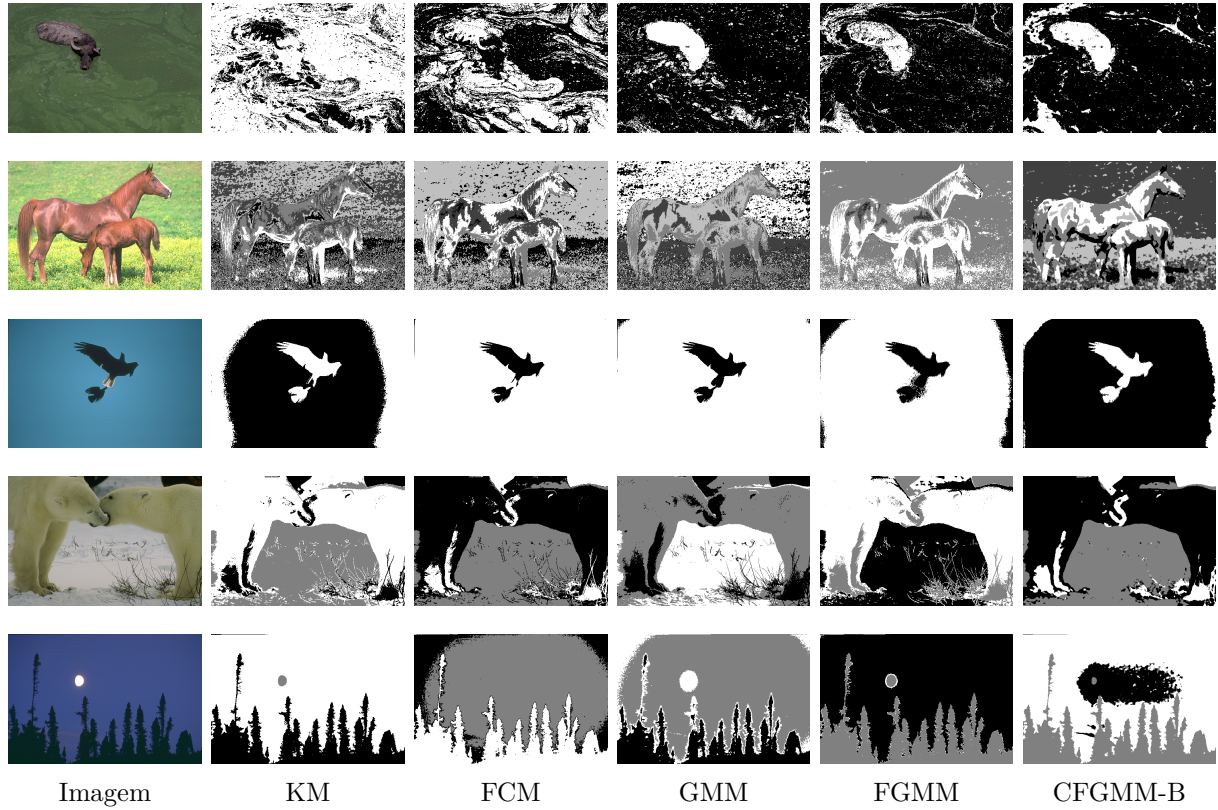


Figura 57 – Segmentação das imagens #80099, #113040, #135069, #183055 e #238011 pelo FCM, GMM, FGMM e CFGMM-B.

Com o objetivo de estudar o comportamento das técnicas na segmentação de imagens naturais sob efeito de ruídos, o ruído Gaussiano com relação sinal ruído 1, 4, 8, 12 e 16 dB foi adicionado às 30 imagens da amostra do do *Berkeley database* assim como descrito da Subseção 5.1.1.1.

A média do Índice PR para o resultado da segmentação das imagens com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB está exibida no gráfico da Figura 58. O desempenho de todos os métodos, avaliado pelo Índice PR, piora com o aumento do nível de ruído da imagem. O CFGMM-B apresentou melhor resultado para imagens com o maior nível de ruído, $SNR = 1$, e não apresentou diferença estatisticamente relevante em relação ao GMM nos níveis de ruído mais baixos 4, 8, 12 e 16. O CFGMM-B apresentou melhor resultado que o FGMM em todos os cenários experimentados. O FGMM apresentou o pior desempenho em todos os cenários.

O gráfico de caixa (*boxplot*) das médias do Índice PR das segmentações das imagens sobre todas as relações sinal ruído, que está exibido na Figura 59, ilustra o desempenho

geral dos métodos avaliados. A marca central na caixa é a mediana, as bordas inferior e superior da caixa correspondem ao primeiro e terceiro quartis, respectivamente, as barras horizontais superior e inferior mostram os dados mais extremos sem considerar os *outliers*. Os *outliers*, quando presentes, são mostrados individualmente.

Embora a mediana do desempenho geral do GMM e do CFGMM-B estejam muito próximas, o CFGMM obteve um desempenho mais constante na segmentação das imagens com diferentes níveis de ruído indicado no gráfico pela menor distância entre o primeiro e o terceiro quartil. O gráfico deixa clara a superioridade do CFGMM-B em relação ao FGMM. O FCM e o FGMM obtiveram os piores resultados gerais, sendo o FGMM o pior deles.

A Figura 60 mostra a imagem #183055 com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB e exemplos dos respectivos resultados da segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.

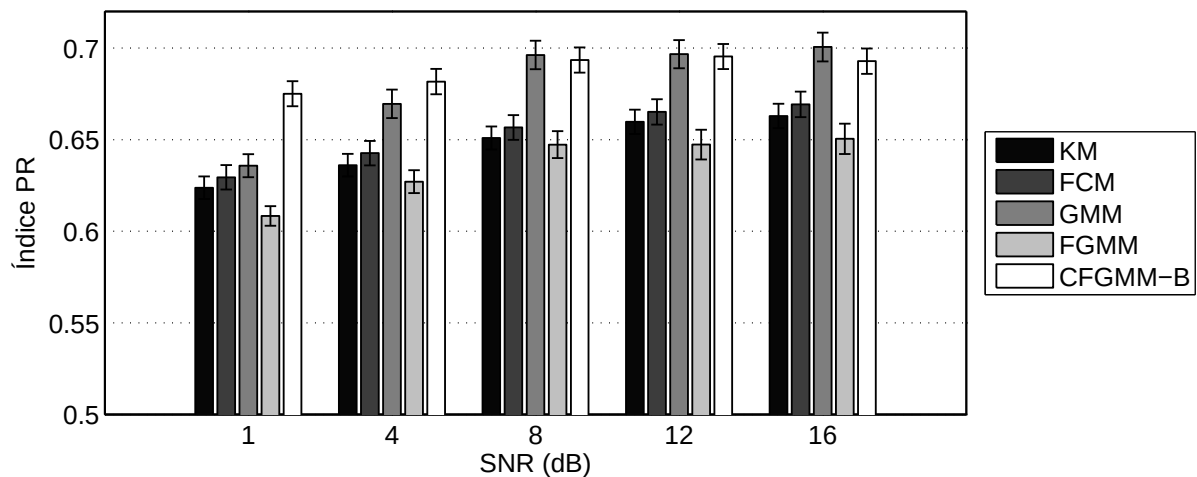


Figura 58 – Média do Índice PR na segmentação das imagens com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB pelo FCM, GMM, FGMM e CFGMM-B.

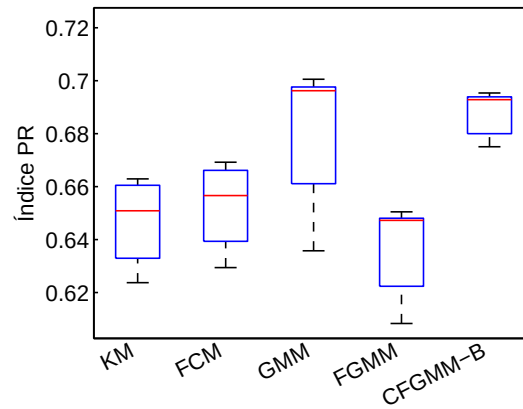


Figura 59 – Gráfico de caixa do Índice PR na segmentação das imagens e todas as relações sinal ruído pelo FCM, GMM, FGMM e CFGMM-B.

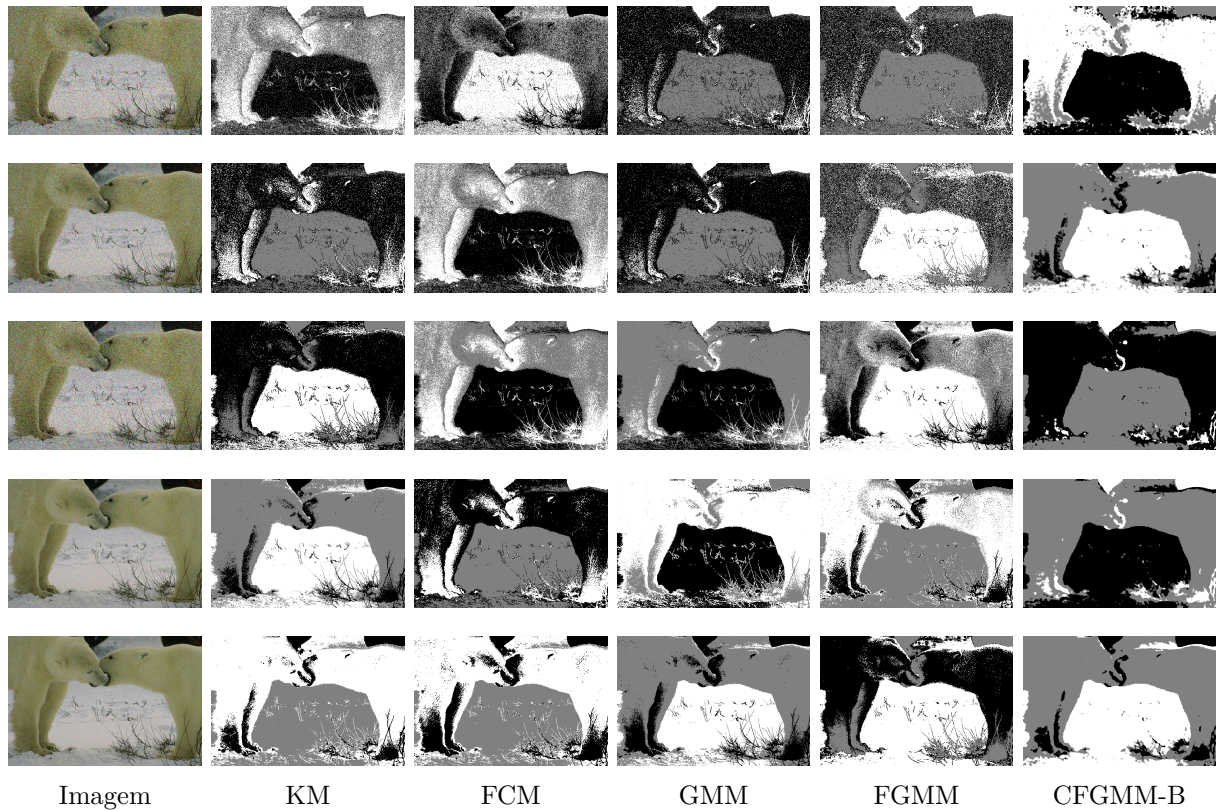


Figura 60 – Segmentação da imagem #183055 com relação sinal ruído (SNR) 1, 4, 8, 12 e 16 dB, de cima para baixo. Da esquerda para a direita, as colunas exibem as imagens originais, e exemplos segmentação pelo KM, FCM, GMM, FGMM e CFGMM-B.

5.2.3.4 Resultado do *benchmark* BSDS500

Para complementar a avaliação da qualidade das segmentações realizadas pelo algoritmo proposto, foi utilizado o *benchmark* de segmentação BSDS500 de Berkeley. O

resultado da segmentação de imagens naturais baseada em regiões pelo CFGMM-B foi comparado ao dos métodos clássicos de agrupamento, KM, FCM, GMM e FGMM, usando as 200 imagens do conjunto de teste da *Berkeley database*. As medidas de desempenho Índice PR e *F-measure* calculadas pelo *benchmark* para todas as técnicas está exibida na Tabela 1.

Usando o Índice de PR como medida de desempenho da segmentação baseada em região, o CFGMM-B apresentou o melhor resultado seguido do FCM e do GMM, empatados, do KM e, por fim, do FGMM. Segundo o *F-measure* calculado pelo *benchmark* como uma medida de qualidade de segmentação baseada em bordas, o CFGMM-B apresentou o melhor resultado seguido do FGMM, FCM, e, empatados o KM e o GMM. O destaque do CFGMM-B é devido a sua tendência à homogeneizar as regiões das imagens resultando em uma segmentação com menos ruído, logo com menos bordas, em relação às técnicas não contextuais. Mesmo obtendo o melhor valor do *F-measure* entre as técnicas selecionadas, o CFGMM-B não obteve um bom resultado na detecção de bordas, tendo em vista que o valor do *F-measure* vai de 0 a 1 e a nota do CFGMM-B foi 0,51. Esse resultado é justificado pelo fato do CFGMM ser um algoritmo de segmentação baseado em regiões e não se preocupar com a preservação das bordas no processo de agrupamento dos dados.

Tabela 1 – Pontuação do KM, FCM, GMM, FGMM e CFGMM-B geradas pelo *benchmark* ao analisar a segmentação baseada em região das 200 imagens de teste da *Berkeley database*.

	KM	FCM	GMM	FGMM	CFGMM-B
Índice PR	0.71	0.72	0.72	0.65	0.73
<i>F-measure</i>	0.35	0.36	0.35	0.40	0.51

5.3 Considerações finais

Os experimentos realizados demonstraram que o modelo proposto pode ser usado como ferramenta de segmentação de imagens baseada em região via agrupamento dos *pixels* de imagens sintéticas, imagens naturais e imagens simuladas de ressonância magnética. Os experimentos realizados com imagens sintéticas demonstraram que o modelo de agrupamento proposto foi capaz de segmentar imagens, tanto no caso em que as classes da imagem têm mesma média e níveis de ruído diferentes quanto no caso em que as classes têm diferentes médias com o mesmo nível de ruído.

Foi possível verificar a convergência do modelo ao longo das iterações corroborando o cálculo de otimização da função objetivo J apresentado na Seção 4.1. Além disso, a análise experimental da complexidade matemática mostrou que o tempo de processamento por iteração na segmentação de imagens aumenta linearmente com o aumento do número de *pixels* da imagem.

O desempenho do CFGMM não sofreu muita influência do valor do expoente *fuzzy*, m , mas se mostrou sensível à variação do critério de parada. Todos os experimentos mostraram que existe um *trade-off* entre o critério de parada e a qualidade da segmentação da imagem pelo CFGMM. Iterar demais causa problemas de eliminação dos detalhes da imagem. Poucas iterações não são suficientes para eliminar os ruídos no interior das regiões das imagens. Um critério de parada precoce foi adotado para evitar a degradação dos detalhes da imagem.

O processo de rotulagem de dados baseado na regra de Bayes, adotada na versão CFGMM-B, se mostrou muito mais adequado à segmentação das imagens sintéticas do que o uso da partição *fuzzy*, adotada na versão CFGMM-F.

A abordagem contextual do CFGMM-B proporcionou uma melhora de desempenho na segmentação de imagens com ruído quando comparada à abordagem pontual do FGMM em todos os cenários experimentados com exceção da imagem sintética de 3 classes com SNR 16 dB cujas medidas de qualidade de segmentação pelas duas técnicas não apresentaram diferenças significativas.

Na comparação com métodos clássicos de agrupamento, ao considerar a informação da vizinhança, o CFGMM-B apresentou um melhor desempenho na segmentação de imagens com maiores níveis de ruído e se equiparou às outras técnicas na segmentação de imagens com menor nível de ruído.

O CFGMM-B e o CFGMM-F mostraram ter uma maior custo computacional ao levar um maior tempo de processamento para segmentar as mesmas imagens, mesmo nos casos em que necessitou de menos iterações para alcançar o resultado final. esse resultado já era esperado pois ambos possuem mais parâmetros do que os outros métodos. Além disso, há a incorporação do contexto no cálculo dos parâmetros do modelo.

5.3.1 Dificuldades encontradas

Durante o desenvolvimento do CFGMM e ao longo dos experimentos documentados neste trabalho encontramos, naturalmente, algumas dificuldades. As principais estão listadas a seguir:

- A principal dificuldade encontrada ao longo deste trabalho foi na definição correta da função de dissimilaridade do FGMM. Além da função de dissimilaridade ser usada no cálculo da função de pertinência *fuzzy*, a forma como ela é definida influencia diretamente na derivação das equações de atualização dos parâmetros do FGMM e, por consequência, do CFGMM. Enquanto Tran e Wagner (TRAN; WAGNER, 1998) usavam a Equação 2.45, reproduzida abaixo na Equação 5.13, Ju e Liu (JU; LIU,

2012) usavam a Equação 5.14:

$$d_{ij}^2 = -\log[w_j N(x_i|\mu_j, \Sigma_j)] \quad (5.13)$$

$$d_{ij}^2 = \frac{1}{[w_j N(x_i|\mu_j, \Sigma_j)]}. \quad (5.14)$$

Embora a definição das funções estejam diferentes, em ambos os trabalhos, as equações de atualização dos parâmetros estavam iguais, o que nos levou a acreditar que as definições eram equivalentes no processo de otimização da função objetivo do FGMM. Inicialmente adotamos a função de dissimilaridade da Equação 5.14 pelo fato do trabalho de Ju e Liu (JU; LIU, 2012) ser mais recente e mais completo do que o de Tran e Wagner (TRAN; WAGNER, 1998). No entanto, na implementação, o FGMM apresentava problemas de convergência. Ao realizar uma avaliação mais cuidadosa, percebemos que a definição da função de dissimilaridade e as equações dos parâmetros do FGMM de Tran e Wagner (TRAN; WAGNER, 1998) estava de acordo com as equações que nós encontramos. Além disso, ao substituir a Equação 5.13 pela Equação 5.14 o FGMM parou de apresentar erro na convergência. Assim, adotamos a definição da função de dissimilaridade da Equação 5.13 de Ju e Liu (JU; LIU, 2012) na proposta do CFGMM.

- Nos experimentos realizados por Ju e Liu (JU; LIU, 2012), o FGMM e o GMM obtiveram a mesma capacidade de se ajustar aos dados e a combinação do algoritmo EM com a lógica fuzzy reduziu custo computacional do FGMM. Usando FGMM na abordagem SVFMM seria possível obter a capacidade de generalização do GMM com uma convergência mais rápida característica dos algoritmos da família do FCM. Os experimentos aqui realizados, no entanto, mostraram que o desempenho geral do FGMM na segmentação de imagens foi inferior ao do GMM e o do FCM. Embora o número de iterações necessário para a convergência do FGMM em diversos cenários tenha sido inferior ao do GMM, o tempo de processamento do FGMM foi, em geral, superior ao do GMM. Esse fato vai contra o resultado obtido por Ju em (JU; LIU, 2012).
- Durante a execução dos experimentos tivemos dificuldades em reproduzir o cálculo do *F-measure* usando por Martin (MARTIN; FOWLKES; MALIK, 2004) no *benchmark* de segmentação BSDS500 de Berkeley (ARBELAEZ et al., 2011). Embora o *benchmark* BSDS500 esteja disponível no site do Grupo de Visão Computacional da Universidade de Berkeley (BERKELEY,), ele só roda no sistema operacional Linux. Como o ambiente de teste adotado nos experimentos usava o sistema operacional Windows®, foi necessário o uso de um o uso de emulador do sistema operacional Linux. Como o *benchmark* BSDS500 foi usado somente para calcular medidas de

desempenho, a utilização do emulador não influenciou o resultado dos métodos avaliados.

6 Considerações Finais

O foco deste trabalho foi a elaboração do Modelo de Mistura de Gaussianas *Fuzzy* Contextual (CFGMM) e a derivação das equações para o cálculo da estimativa dos parâmetros do modelo. Campos Aleatórios Markovianos Gaussianos (GMRF) foram usados para modelar a dependência contextual diretamente nos pesos das misturas.

O CFGMM realiza o agrupamento de dados buscando a minimização de uma função objetivo através de um processo de otimização iterativa baseado no gradiente que é mais simples do que a estimativa MAP via *Expectation Maximization* usado do Modelo de Mistura Finita Variante no Espaço. Ao simplificar o processo de otimização foi possível incorporar a restrição imposta aos pesos da mistura na derivação das equações dos parâmetros do modelo, eliminando, assim, a necessidade de uma etapa de correção adicional.

Duas versões do CFGMM foram propostas: na primeira o processo de rotulagem de dados, após o cálculo dos parâmetros, baseia-se na regra de decisão Bayesiana (CFGMM-B); na segunda, a rotulagem dos dados se baseia na análise da partição *fuzzy* (CFGMM-F). Por fim, foi realizada a implementação e validação do modelo como ferramenta de agrupamento de *pixels* na segmentação de imagens baseada em regiões.

A equações de atualização dos parâmetros do modelo foram derivadas através da minimização da função objetivo J . Os experimentos mostraram que o valor de J diminui ao longo das iterações do processo de segmentação de diferentes imagens, verificando, assim, a convergência do modelo ao longo das iterações e corroborando com o cálculo de otimização da função objetivo J . Na análise experimental da complexidade matemática do CFGMM foi verificado que o tempo de processamento por iteração do CFGMM aumenta linearmente com o aumento do tamanho do conjunto de dados processados.

Inicialmente, a influência dos parâmetros de entrada do CFGMM na qualidade da segmentação de diferentes tipos de imagem foi explorada nos experimentos realizados. O desempenho do CFGMM se mostrou sensível à variação do critério de parada, mas não sofreu muita influência do valor do expoente *fuzzy*, m . Não houve uma unanimidade entre os experimentos para o melhor valor de m , mas para a maioria dos dados, os melhores resultados foram obtidos com o intervalo $1,3 \leq m \leq 2,0$.

O critério de convergência adotado foi a porcentagem de mudança da função objetivo J a cada iteração, assim, o algoritmo converge quando $\left| \frac{\Delta J}{J} \right| < Th$ ou atinge um número máximo de iterações. Os experimentos mostraram que ao longo das iterações as regiões da imagem vão ficando mais homogêneas mas as bordas das regiões vão ficando borradas. As quinas vão se tornando curvas e as linhas retas vão perdendo a definição.

Havendo, assim, um *trade-off* entre eliminação de ruído no interior das regiões e a perda de definição dos detalhes da imagem. Ao medir a qualidade da segmentação ao longo das iterações percebe-se que a qualidade da segmentação atinge um valor ótimo muito rápido, no intervalo de 5 a 10 iterações. Para evitar a degradação dos detalhes da imagem, definiu-se um critério de parada padrão precoce com valor de $Th = 10^{-2}$ e com número máximo de 15 iterações.

A abordagem contextual do CFGMM-B apresentou melhor desempenho na segmentação de imagens com ruído do que a abordagem pontual do FGMM na maioria dos cenários analisados. O processo de rotulagem de dados baseado na regra de Bayes, adotada na versão CFGMM-B, se mostrou muito mais adequado à segmentação das imagens sintéticas do que o uso da partição *fuzzy*, adotada na versão CFGMM-F.

Os experimentos com as imagens sintéticas e com as imagens naturais da *Berkeley database* mostraram que, na comparação com métodos clássicos de agrupamento, o Modelo de Mistura de Gaussianas, o *Fuzzy C-Means* e o Modelo de Mistura de Gaussianas *Fuzzy*, ao considerar a informação da vizinhança, o CFGMM-B apresentou um melhor desempenho na segmentação de imagens com maiores níveis de ruído e se equiparou às outras técnicas na segmentação de imagens com menor nível de ruído.

O resultado do experimento com as fatias da imagem tri-dimensional simulada de ressonância magnética do *BrainWeb* mostrou que a qualidade da segmentação da imagem por todos os métodos, avaliada pelo Índice Rand, piora com o aumento do nível de ruído da imagem. O FCM obteve melhores resultados na segmentação de imagens com os menores níveis de ruído avaliados, o GMM obteve melhores resultados na segmentação de imagens com níveis intermediários de ruído avaliados e o CFGMM-B obteve melhores resultados com os maiores níveis de ruído avaliados. Analisando o resultado do CFGMM na segmentação de cada região da imagem do cérebro detalhadamente, é possível perceber que a matéria branca e o *background* foram bem segmentados mas que o fluido cérebro-espinhal, que corresponde à região de menor área na imagem, foi confundido com matéria cinzenta. Esse resultado é decorrência da degradação dos detalhes da imagem pelo CFGMM.

O CFGMM-B e o CFGMM-F mostraram ter um maior custo computacional que os métodos clássicos de agrupamento ao levar maior tempo de processamento para segmentar as mesmas imagens, mesmo nos casos em que necessitou de menos iterações para alcançar o resultado final. Entretanto, esse resultado já era esperado devido à incorporação do contexto no cálculo dos parâmetros do modelo.

6.1 Trabalhos futuros

Os próximos passos para dar continuidade a este trabalho devem incluir a investigação de algumas ideias que surgiram no decorrer da pesquisa e a realização de análises

mais detalhadas de alguns pontos específicos relacionados ao modelo proposto.

Pretendemos explorar o uso de diferentes funções de dissimilaridade do CFGMM para buscar por geometrias nos grupos de dados que sejam mais adequadas à forma real dos dados analisados. Tensionamos testar, com o CFGMM, modificações da função de dissimilaridade aplicadas ao FCM clássico, como as propostas por Hathaway *et al.* em (HATHAWAY; BEZDEK; HU, 2000), por Groenen e Jajuga em (GROENEN; JAJUGA, 2001), por Lu *et al.* em (LU; S.XIAO; GU, 2014) ou por Ferreira e De Carvalho em (FERREIRA; CARVALHO, 2014).

Tendo em vista que a etapa de rotulagem dos dados teve muita influência no desempenho do CFGMM, pretendemos investigar diferentes maneiras realizar essa etapa buscando a melhoria da qualidade do agrupamento dos dados. Estimativas de densidade não paramétricas das funções de distribuições de probabilidade podem ser usadas no cálculo da probabilidade *a posteriori* do CFGMM-B, adotando o método do histograma, o estimador ingênuo ou do estimador Kernel (SILVERMAN, 1986).

Procurando uma melhor adaptação do modelo proposto aos dados analisados, e consequentemente uma melhor segmentação da imagem, pretendemos explorar diferentes sistemas de vizinhança definidas no GMRF adotado para representação estatística da relação espacial entre *pixels* vizinhos. Tencionamos modificar o modelo de maneira que o parâmetro de suavidade, β , possa assumir diferentes valores para diferentes classes, ou ainda, possa controlar o grau de suavidade em diferentes direções espaciais. Além disso, pretendemos explorar maneiras de preservar os detalhes da imagem no processo de segmentação incorporando, por exemplo, um modelo de relação espacial que garanta a preservação das bordas das regiões das imagens.

Com relação aos parâmetros de entrada do CFGMM proposto, pretendemos automatizar a definição do número de componentes da mistura, que está relacionada ao número de classes da imagem, e o cálculo do valor do expoente *fuzzy*, que pode variar com a natureza das imagens. Adicionalmente, buscaremos um novo critério de parada que evite a degradação dos detalhes da imagem.

Buscando um melhor desempenho computacional, pretendemos explorar maneiras de reduzir o tempo de processamento do CFGMM para torna-lo mais competitivo em relação aos métodos de agrupamento de dados com abordagem pontual. Esse objetivo pode ser alcançado, por exemplo, adotando uma maneira mais eficiente de acessar os atributos da vizinhança de cada dado analisado no cálculo dos parâmetros a cada iteração do CFGMM.

Como um trabalho mais amplo, é necessário explorar o uso do CFGMM em outras aplicações nas quais exista dependência entre dados vizinhos, tais como, comunicação, reconhecimento de voz e processamento de vídeos nos quais o contexto temporal é obtido

considerando os dados em diferentes momentos ou no processamento de imagens espectrais nos quais o contexto espectral é obtido com uso de diferentes bandas da imagem captadas por diferentes sensores. Pretendemos ainda aplicar o CFGMM no *framework* semi-supervisionado proposto em (PORTELA; CAVALCANTI; TSANG, 2014).

Referências

- ADAMS, R.; BISCHOF, L. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 16, n. 6, p. 641 –647, 1994. Citado na página 47.
- ARBELAEZ, P. et al. Contour detection and hierarchical image segmentation. *IEEE Pattern Analysis and Machine Intelligence*, v. 33, n. 5, p. 898 –916, 2011. Citado 5 vezes nas páginas 78, 79, 81, 82 e 121.
- BERKELEY. *Berkeley image segmentation database and benchmark*. [Online; último acesso em 21 de maio de 2015]. Disponível em: <<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>>. Citado na página 121.
- BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 36, n. 2, p. 192 –236, 1974. Citado 3 vezes nas páginas 52, 53 e 59.
- BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1.. ed. Norwell, MA: Kluwer Academic Publishers, 1981. Citado 3 vezes nas páginas 22, 39 e 67.
- BEZDEK, J. C.; EHRLICH, R.; FULL, W. FCM: The fuzzy C-Means clustering algorithm. *Computers and Geosciences*, v. 10, n. 2-3, p. 191 –203, 1984. Citado na página 88.
- BILMES, J. *A Gentle Tutorial of the EM algorithm and its application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. [S.l.], 1997. Citado 2 vezes nas páginas 36 e 37.
- BIRKHOFF, G.; LANE, S. M. *A Survey of Modern Algebra*. 4.. ed. New York, NY: MacMillan, 1941. Citado na página 63.
- BLAKE, A.; ISARD, M. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. 1.. ed. New York, NY: Springer, 1998. Citado na página 47.
- BLEKAS, K. et al. A spatially-constrained mixture model for image segmentation. *IEEE Transactions on Neural Networks*, v. 16, n. 2, p. 494 –498, 2005. Citado 6 vezes nas páginas 23, 24, 54, 58, 59 e 61.
- BOYKOV, Y.; FUNKA-LEA, G. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, v. 70, n. 2, p. 109 –131, 2006. Citado na página 47.
- COCOSCO, C. A. et al. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage*, v. 5, p. 425, 1997. Citado na página 74.
- DEGROOT, M. H.; SCHERVISH, M. J. *Probability and Statistics*. 4.. ed. Upper Saddle River, NJ: Pearson, 2011. Citado 3 vezes nas páginas 82, 103 e 115.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 39, n. 1, p. 1 –38, 1977. Citado 3 vezes nas páginas 21, 33 e 34.

- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2.. ed. Hoboken, NJ: Wiley, 2001. Citado 4 vezes nas páginas [28](#), [29](#), [30](#) e [33](#).
- DUNN, J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, v. 3, n. 3, p. 32 –57, 1974. Citado na página [22](#).
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861 –874, 2006. Citado 2 vezes nas páginas [47](#) e [81](#).
- FERREIRA, M. R. P.; CARVALHO, F. A. T. D. Kernel fuzzy C-means with automatic variable weighting. *Fuzzy Sets and Systems*, v. 237, n. 0, p. 1 –46, 2014. Citado na página [125](#).
- FORSYTH, D. A.; PONCE, J. *Computer Vision: A Modern Approach*. 2.. ed. Upper Saddle River, NJ: Prentice Hall PTR, 2002. Citado 2 vezes nas páginas [23](#) e [48](#).
- GONG, M. et al. Fuzzy C-Means clustering with local information and kernel metric for image segmentation. *IEEE Transactions on Image Processing*, v. 22, n. 2, p. 573 –584, 2013. Citado 3 vezes nas páginas [23](#), [26](#) e [49](#).
- GONZALEZ, R. C.; WOODS, R. E. *Processamento Digital de Imagens*. 3.. ed. São Paulo, SP: Pearson Prentice Hall, 2009. Citado 3 vezes nas páginas [45](#), [46](#) e [47](#).
- GROENEN, P. J. F.; JAJUGA, K. Fuzzy clustering with squared minkowski distances. *Fuzzy Sets and Systems*, v. 120, n. 2, p. 227 –237, 2001. Citado na página [125](#).
- GUSTAFSON, D. E.; KESSEL, W. C. Fuzzy clustering with a fuzzy covariance matrix. In: *IEEE Conference on Decision and Control*. San Diego, CA: [s.n.], 1978. v. 17, p. 761 –766. Citado na página [22](#).
- HAMMERSLEY, J. M.; CLIFFORD, P. Markov field on finite graphs and lattices. *unpublished*, 1971. Citado 2 vezes nas páginas [52](#) e [53](#).
- HATHAWAY, R. J.; BEZDEK, J. C.; HU, Y. Generalized fuzzy c-means clustering strategies using l p norm distances. *IEEE Transactions on Fuzzy Systems*, v. 8, n. 5, p. 576 –582, 2000. Citado na página [125](#).
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Discrete Multivariate Distributions*. 1.. ed. Hoboken, NJ: Wiley, 1997. Citado 2 vezes nas páginas [55](#) e [61](#).
- JOTHILAKSHMI, S. Automatic system to detect the type of voice pathology. *Applied Soft Computing*, v. 21, n. 0, p. 244 –249, 2014. Citado na página [23](#).
- JU, Z.; LIU, H. Fuzzy gaussian mixture models. *Pattern Recognition*, v. 45, n. 3, p. 1146 –1158, 2012. Citado 4 vezes nas páginas [22](#), [41](#), [67](#) e [121](#).
- KANNAN, S. R.; RAMATHILAGAM, S.; CHUNG, P. C. Effective fuzzy C-Means clustering algorithms for data clustering problems. *Expert Systems with Applications*, v. 39, n. 7, p. 6292 –6300, 2012. Citado na página [49](#).
- LEVADA, A.; MASCARENHAS, N.; TANNUS, A. Classificação contextual de imagens utilizando campos aleatórios Markovianos e teoria dos jogos. In: *Anais do Workshop de Visão Computacional*. Bauru, SP: [s.n.], 2008. p. 167 –172. Citado na página [52](#).

- LI, S. Z. *Markov Random Field Modeling in Image Analysis*. 3.. ed. New York, NY: Springer, 2009. Citado 2 vezes nas páginas 50 e 57.
- LI, X.; ZHANG, T.; QU, Z. Image segmentation using fuzzy clustering with spatial constraints based on Markov random field via Bayesian theory. *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, E91-A, n. 3, p. 723–729, 2008. Citado 2 vezes nas páginas 26 e 49.
- LU, C.; S.XIAO; GU, X. Improving fuzzy C-means clustering algorithm based on a density-induced distance measure. *The Journal of Engineering*, v. 1, n. 1, p. 0–3, 2014. Citado na página 125.
- MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: [s.n.], 1967. v. 1, p. 281–297. Citado na página 30.
- MARTIN, D. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *International Conference on Computer Vision*. Vancouver, BC: [s.n.], 2001. v. 2, p. 416–423. Citado 2 vezes nas páginas 74 e 79.
- MARTIN, D. R.; FOWLKES, C.; MALIK, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 5, p. 530–549, 2004. Citado 2 vezes nas páginas 81 e 121.
- MATLAB. *version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc., 2010. Citado na página 82.
- MCLACHLAN, G. J.; PEEL, D. *Finite Mixture Models*. 1.. ed. Hoboken, NJ: Wiley, 2000. Citado 7 vezes nas páginas 21, 22, 32, 33, 34, 35 e 36.
- MIGNOTTE, M. A de-texturing and spatially constrained K-Means approach for image segmentation. *Pattern Recognition Letters*, v. 32, n. 2, p. 359–367, 2011. Citado na página 49.
- NAZ, S.; MAJEED, H.; IRSHAD, H. Image segmentation using fuzzy clustering: A survey. In: *International Conference on Emerging Technologies*. Islamabad, PK: [s.n.], 2010. p. 181–186. Citado na página 23.
- NGUYEN, T. M.; WU, Q. M. J. Dirichlet gaussian mixture model: Application to image segmentation. *Image and Vision Computing*, v. 29, n. 12, p. 818–828, 2011. Citado 2 vezes nas páginas 55 e 63.
- NGUYEN, T. M.; WU, Q. M. J. A fuzzy C-Means based spatial pixel and membership relationships for image segmentation. In: *Canadian Conference on Computer and Robot Vision*. st. john's, NFLD: [s.n.], 2011. p. 278–284. Citado 2 vezes nas páginas 26 e 49.
- NGUYEN, T. M.; WU, Q. M. J. Fast and robust spatially constrained gaussian mixture model for image segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 23, n. 4, p. 621–635, 2013. Citado 3 vezes nas páginas 23, 24 e 55.

- NIKOU, C.; GALATSANOS, N. P.; LIKAS, A. C. A class-adaptive spatially variant mixture model for image segmentation. *IEEE Transactions on Image Processing*, v. 16, p. 1121 –1130, 2007. Citado 13 vezes nas páginas 23, 24, 26, 49, 54, 56, 59, 60, 62, 67, 68, 73 e 110.
- NIKOU, C.; LIKAS, A. C.; GALATSANOS, N. P. A Bayesian framework for image segmentation with spatially varying mixtures. *IEEE Transactions on Image Processing*, v. 19, p. 2278 –2289, 2010. Citado 6 vezes nas páginas 23, 24, 55, 61, 62 e 63.
- PEDRINI, H.; SCHWARTZ, W. R. *Análises de Imagens Digitais: Princípios, Algoritmos e Aplicações*. 1.. ed. São Paulo, SP: Thomson Learning, 2008. Citado na página 46.
- PEDRYCZ, W. *Knowledge-Based Clustering: From Data to Information Granules*. 1.. ed. Hoboken, NJ: Wiley, 2005. Citado 2 vezes nas páginas 40 e 88.
- PORTELA, N. M.; CAVALCANTI, G. D. C.; TSANG, I. R. Contextual image segmentation based on the Potts model. In: *IEEE International Conference on Tools with Artificial Intelligence*. Herndon, VA: [s.n.], 2013. p. 256 –261. Citado 2 vezes nas páginas 23 e 55.
- PORTELA, N. M.; CAVALCANTI, G. D. C.; TSANG, I. R. Semi-supervised clustering for MR brain image segmentation. *Expert Systems with Application*, v. 41, n. 4, p. 1492 –1497, 2014. Citado na página 126.
- RAND, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, v. 66, n. 336, p. 846 –850, 1971. Citado na página 80.
- RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 71, n. 2, p. 319 –392, 2009. Citado 2 vezes nas páginas 53 e 59.
- RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 3.. ed. Upper Saddle River, NJ: Prentice Hall, 2009. Citado 3 vezes nas páginas 28, 29 e 32.
- SANJAY-GOPAL, S.; HEBERT, T. J. Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Transactions on Image Processing*, v. 7, n. 7, p. 1014 –1028, 1998. Citado 6 vezes nas páginas 24, 54, 56, 57, 59 e 61.
- SARASWATHI, S.; ALLIRANI, A. Survey on image segmentation via clustering. In: *International Conference on Information Communication and Embedded Systems*. Chennai, TN: [s.n.], 2013. p. 331 –335. Citado 2 vezes nas páginas 23 e 48.
- SFIKAS, G. et al. Spatially varying mixtures incorporating line processes for image segmentation. *Journal of Mathematical Imaging and Vision*, v. 36, n. 2, p. 91 –110, 2010. Citado na página 54.
- SILVERMAN, B. W. *Density estimation for statistics and data analysis*. London, UK: Chapman and Hall, 1986. Citado na página 125.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 4.. ed. Boston, MA: Academic Press, 2009. Citado 3 vezes nas páginas 28, 49 e 50.

- THILAGAMANI, S.; SHANTI, N. A survey on image segmentation through clustering. *International Journal of Research and Reviews in Information Sciences*, v. 1, n. 1, p. 14–17, 2011. Citado 2 vezes nas páginas 23 e 48.
- TRAN, D.; WAGNER, M. Fuzzy gaussian mixture models for speaker recognition. In: *International Conference on Spoken Language Processing*. Sydney ,NSW: [s.n.], 1998. p. 759–762. Citado 5 vezes nas páginas 22, 41, 67, 120 e 121.
- UNNIKRISHNAN, R.; PANTOFARU, C.; HEBERT, M. A measure for objective evaluation of image segmentation algorithms. In: *International Conference on Computer Vision and Pattern Recognition*. San Diego, CA: [s.n.], 2005. p. 34–41. Citado na página 80.
- VANTARAM, S. R.; SABER, E. Survey of contemporary trends in color image segmentation. *Journal of Electronic Imaging*, v. 21, n. 4, p. 040901–1–040901–28, 2012. Citado 3 vezes nas páginas 23, 47 e 48.
- WANG, J. et al. Fast approximate K-Means via cluster closures. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI: [s.n.], 2012. p. 3037–3044. Citado na página 49.
- WOLBERG, J. R.; WOLBERG, J. *Data analysis using the method of least squares: extracting the most information from experiments*. 1.. ed. Berlin, DE: Springer, 2006. Citado na página 85.
- XU, R.; II, D. C. W. *Clustering*. 1.. ed. Hoboken, NJ: Wiley, 2009. Citado 4 vezes nas páginas 21, 28, 29 e 88.
- YANG, B.; GUO, C.; JENSEN, C. S. Travel cost inference from sparse, spatio temporally correlated time series using Markov models. *Proceedings of the VLDB Endowment*, VLDB Endowment, v. 6, n. 9, p. 769–780, 2013. Citado na página 23.
- YANG, K. et al. Efficient Lagrange multiplier selection for scalable video coding using temporal inter-layer dependency. *Signal, Image and Video Processing*, v. 1, n. 1, p. 1–11, 2014. Citado na página 23.
- ZHANG, H.; WU, Q. M. J.; NGUYEN, T. M. Incorporating Mean template into finite mixture model for image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, v. 24, n. 2, p. 328–335, 2013. Citado 4 vezes nas páginas 23, 26, 49 e 56.
- ZHANG, L. et al. Context-based person identification framework for smart video surveillance. *Machine Vision and Applications*, v. 25, n. 7, p. 1711–1725, 2014. Citado na página 23.