Universidade de São Paulo
Instituto de Física de São Carlos
Departamento de Física e Informática

Alexandre Luis Magalhães Levada

Combinação de Modelos de Campos Aleatórios Markovianos para Classificação Contextual de Imagens Multiespectrais

São Carlos
2010

Alexandre Luis Magalhães Levada

Combinação de Modelos de Campos Aleatórios Markovianos para Classificação Contextual de Imagens Multiespectrais

Tese apresentada ao Programa de Pós-graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências.

Área de Concentração: Física Aplicada
Opção: Física Computacional
Orientador: Prof. Dr. Nelson D. A. Mascarenhas
Co-orientador: Prof. Dr. Alberto Tannús

São Carlos
2010

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Levada, Alexandre Luis Magalhães.
Combinação de Modelos de Campos Aleatórios Markovianos para Classificação Contextual de Imagens Multiespectrais. / Alexandre Luis Magalhães Levada ; orientador Nelson D. A. Mascarenhas; co-orientador Alberto Tannús. -- São Carlos, 2010.
146 p.

Tese (Doutorado – Programa de Pós-Graduação em Física – Área de concentração: Física Aplicada – Opção Física Computacional) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Campos Aleatórios Markovianos. 2. Classificação Contextual. 3. Máxima Pseudo Verossimilhança. 4. Inferência Bayesiana. 5. Análise Assintótica. I. Título.

FOLHA DE APROVAÇÃO

Alexandre Luis Magalhães Levada

Tese apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências.
Área de Concentração: Física Aplicada – Opção: Física Computacional.

Aprovado em: 05/05/2010

Comissão Julgadora

Prof. Dr. Max Henrique Machado Costa

Instituição: UNICAMP Assinatura _____

Prof. Dr. Ricardo da Silva Torres

Instituição: UNICAMP Assinatura Ricardo Torres

Prof. Dr. Francisco Aparecido Rodrigues

Instituição: ICMC/USP Assinatura _____

Prof. Dr. João do Espírito Santo Batista Neto

Instituição: ICMC/USP Assinatura _____

Prof. Dr. Nelson Delfino D'Ávila Mascarenhas

Instituição: UFSCar Assinatura _____

Agradecimentos

À toda minha família, em especial aos meus pais, Miriam e Celso, e minha irmã, Vivian, pelo amor, apoio incondicionais e incentivos para que eu pudesse alcançar mais este objetivo, mas principalmente pela companhia ao longo de todos esses anos de vida. Agradeço também à minha irmãzinha Karen, que mesmo já não estando mais entre nós, é meu anjo da guarda e me protege e abençoa todos os dias.

À minha querida Débora, companheira de todas as horas, por toda atenção, amizade e amor, por todo o tempo que passamos estudando e escrevendo artigos juntos, pelas inesquecíveis viagens que fizemos (que saudade!), enfim, por existir e ser assim tão especial para mim.

Ao Prof. Dr. Nelson D. A. Mascarenhas, meu orientador, pela confiança e amizade, por todos os ensinamentos, as opiniões, as colaborações, as correções, a companhia ao longo desses anos de aprendizado, as viagens aos congressos, mas principalmente, por compartilhar tanto conhecimento e sabedoria. Será sempre um exemplo de vida e conduta.

Ao Prof. Dr. Alberto Tannús, por ter viabilizado tanto o doutorado quanto o projeto de pesquisa no IFSC, pela co-orientação ao longo de todo esse período e também pelo apoio, juntamente com a aluna Hilde Buzzá, durante a etapa de aquisição das imagens multiespectrais de ressonância magnética, extremamente importantes para o desenvolvimento do trabalho, principalmente para a obtenção e validação dos resultados obtidos.

A meus amigos e companheiros do GAPIS (Ana Luisa, Michelle, Márcia, Ana Paula, Cristiane, Érica, Denis, Eduardo, Moacir, Murillo, Rafael, Renato, Chofhi, Léo, John, Zorzan, Gustavo, Luis, João Paulo), as meninas da Rep. Pipoca (Giovana, Fernanda, Natália e Mônica) e a todos os outros amigos de São Carlos com quem tive a oportunidade de conviver ao longo desses anos, pela amizade, companhia e discussões sobre os mais variados temas e assuntos, o que certamente contribuiu muito para que eu me tornasse uma pessoa melhor. Jamais irei esquecê-los pois levarei comigo todo aprendizado e todas as lembranças que marcaram esses bons e velhos tempos.

À todos os professores que tive ao longo da vida, pelos ensinamentos valiosos que me transmitiram, essenciais para que eu pudesse chegar até aqui, em especial aos Prof(s).

Dr(s). Luciano da Fontoura Costa, Mário de Castro Andrade Filho e Rodrigo Capobianco Guido pelo aprendizado através das disciplinas cursadas durante o doutorado.

À todo pessoal do IFSC (professores, funcionários e alunos) por toda ajuda e atenção quando precisei.

À Embrapa, em especial ao Prof. Dr. Paulo E. Cruvinel e ao Prof. Dr. Clóvis Isberto Biscegli, pelas imagens multiespectrais de tomografia computadorizada de ciência do solo e imagens multiespectrais de ressonância magnética de frutas.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão da bolsa de doutorado e pelo apoio financeiro para a realização desta pesquisa.

Em suma, agradeço a todos aqueles que, direta ou indiretamente, contribuíram de alguma maneira para que tudo isso se tornasse realidade.

À Deus.

Resumo

LEVADA, A. L. M. **Combinação de modelos de campos aleatórios markovianos para classificação contextual de imagens multiespectrais.** 146p. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

Este projeto de doutorado apresenta uma nova abordagem MAP-MRF para a classificação contextual de imagens multiespectrais utilizando combinação de modelos de Campos Aleatórios Markovianos definidos em sistemas de ordens superiores. A modelagem estatística para o problema de classificação segue o paradigma Bayesiano, com a definição de um modelo Markoviano para os dados observados (*Gaussian Markov Random Field* multiespectral) e outro modelo para representar o conhecimento *a priori* (*Potts*). Nesse cenário, o parâmetro β do modelo de Potts atua como um parâmetro de regularização, tendo papel fundamental no compromisso entre as observações e o conhecimento *a priori*, de modo que seu correto ajuste é necessário para a obtenção de bons resultados. A introdução de sistemas de vizinhança de ordens superiores requer a definição de novos métodos para a estimação dos parâmetros dos modelos Markovianos. Uma das contribuições desse trabalho é justamente propor novas equações de pseudo-verossimilhança para a estimação desses parâmetros no modelo de *Potts* em sistemas de segunda e terceira ordens. Apesar da abordagem por máxima pseudo-verossimilhança ser amplamente utilizada e conhecida na literatura de campos aleatórios, pouco se conhece acerca da acurácia dessa estimação. Foram derivadas aproximações para a variância assintótica dos estimadores propostos, caracterizando-os completamente no caso limite, com o intuito de realizar inferências e análises quantitativas sobre os parâmetros dos modelos Markovianos. A partir da definição dos modelos e do conhecimento dos parâmetros, o próximo estágio é a classificação das imagens multiespectrais. A solução para esse problema de inferência Bayesiana é dada pelo critério de estimação `MAP`, onde a solução ótima é determinada maximizando a probabilidade *a posteriori*, o que define um problema de otimização. Como não há solução analítica para esse problema no caso de prioris Markovianas, algoritmos iterativos de otimização combinatória foram empregados para aproximar a solução ótima. Nesse trabalho, adotam-se três métodos sub-ótimos: *Iterated Conditional Modes*, *Maximizer of the Posterior Marginals* e *Game Strategy Approach*. Porém, é demonstrado na literatura que tais métodos convergem para máximos locais e não globais, pois são altamente dependentes de sua condição inicial. Isto motivou o desenvolvimento de uma

nova abordagem para combinação de classificadores contextuais, que utiliza múltiplas inicializações simultâneas providas por diferentes classificadores estatísticos pontuais. A metodologia proposta define um *framework* MAP-MRF bastante robusto para solução de problemas inversos, pois permite a utilização e a integração de diferentes condições iniciais em aplicações como classificação, filtragem e restauração de imagens. Como medidas quantitativas de desempenho, são adotados o coeficiente *Kappa* de Cohen e o coeficiente *Tau* de Kendall para verificar a concordância entre as saídas dos classificadores e a "verdade terrestre" (amostras pré-rotuladas). Resultados obtidos mostram que a inclusão de sistemas de vizinhança de ordens superiores é de fato capaz de melhorar significativamente não apenas o desempenho da classificação como também a estimação dos parâmetros dos modelos Markovianos, reduzindo tanto o erro de estimação quanto a variância assintótica. Além disso, a combinação de classificadores contextuais através da utilização de múltiplas inicializações simultâneas melhora significativamente o desempenho da classificação se comparada com a abordagem tradicional com apenas uma inicialização.

***Palavras-chave*:** Campos Aleatórios Markovianos, Classificação Contextual, Máxima Pseudo Verossimilhança, Inferência Bayesiana, Análise Assintótica, Imagens Multiespectrais.

Abstract

LEVADA, A. L. M. **Combining markov random field models for multispectral image contextual classification.** 146p. Thesis (Doctoral) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

This work presents a novel MAP-MRF approach for multispectral image contextual classification by combining higher-order Markov Random Field models. The statistical modeling follows the Bayesian paradigm, with the definition of a multispectral *Gaussian Markov Random Field* model for the observations and a Potts MRF model to represent the *a priori* knowledge. In this scenario, the Potts MRF model parameter (β) plays the role of a regularization parameter by controlling the tradeoff between the likelihood and the prior knowledge, in a way that a suitable tunning for this parameter is required for a good performance in contextual classification. The introduction of higher-order MRF models requires the specification of novel parameter estimation methods. One of the contributions of this work is the definition of novel pseudo-likelihood equations for the estimation of these MRF parameters in second and third order neighborhood systems. Despite its widely usage in practical MRF applications, little is known about the accuracy of maximum pseudo-likelihood approach. Approximations for the asymptotic variance of the proposed MPL estimators were derived, completely characterizing their behavior in the limiting case, allowing statistical inference and quantitative analysis. From the statistical modeling and having the model parameters estimated, the next step is the multispectral image classification. The solution for this Bayesian inference problem is given by the MAP criterion, where the optimal solution is obtained by maximizing the *a posteriori* distribution, defining an optimization problem. As there is no analytical solution for this problem in case of Markovian priors, combinatorial optimization algorithms are required to approximate the optimal solution. In this work, we use three suboptimal methods: *Iterated Conditional Modes*, *Maximizer of the Posterior Marginals* and *Game Strategy Approach*, a variant approach based on non-cooperative game theory. However, it has been shown that these methods converge to local maxima solutions, since they are extremelly dependent on the initial condition. This fact motivated the development of a novel approach for combination of contextual classifiers, by making use of multiple initializations at the same time, where each one of these initial conditions is provided by different

pointwise pattern classifiers. The proposed methodology defines a robust MAP-MRF framework for the solution of general inverse problems since it allows the use and integration of several initial conditions in a variety of applications as image classification, denoising and restoration. To evaluate the performance of the classification results, two statistical measures are used to verify the agreement between the classifiers output and the ground truth: Cohen's *Kappa* and Kendall's *Tau* coefficient. The obtained results show that the use of higher-order neighborhood systems is capable of significantly improve not only the classification performance, but also the MRF parameter estimation by reducing both the estimation error and the asymptotic variance. Additionally, the combination of contextual classifiers through the use of multiple initializations also improves the classificatoin performance, when compared to the traditional single initialization approach.

***Keywords*:** Markov Random Fields, Contextual Classification, Maximum Pseudo-Likelihood, Bayesian Inference, Asymptotic analysis, Multispectral Images.

Lista de Figuras

Figura 1 - Diagrama de blocos do sistema proposto para classificação contextual de imagens multiespectrais utilizando combinação de algoritmos subótimos de otimização combinatória. 26

Figura 2 - O uso de múltiplas condições iniciais torna possível cobrir uma região muito maior do espaço de estados, possibilitando convergência a melhores resultados. 27

Figura 3 - Funções de custo para estimação Bayesiana: a.) função de custo quadrática, b.) de valor absoluto e c.) uniforme. 34

Figura 4 - Sistema de vizinhança para um GMRF multiespectral de primeira ordem. 40

Figura 5 - Sistemas de vizinhanças não-causais de primeira a quinta ordem. . . 56

Figura 6 - Padrões de configuração espacial para um modelo de Potts de primeira ordem: cada padrão oferece uma contribuição distinta para a função de pseudo-verossimilhança. 59

Figura 7 - Imagens suaves apresentam padrões de configuração mais homogêneos. 61

Figura 8 - Imagens ruidosas apresentam predominância de padrões de configuração heterogêneos. 62

Figura 9 - Comparação entre a distribuição de padrões de configuração espacial para sistemas de segunda ordem para a imagem Lena suave e ruidosa (k_0 denota concordância total e k_{22} concordância nula). 62

Figura 10 - Imagens sintéticas geradas a partir do algoritmo Metropolis-Hastings representando diversas ocorrências do modelo GMRF. 69

Figura 11 - Interpretação do resultado de um teste de hipóteses em termos de probabilidades de significância. 73

Figura 12 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de segunda ordem com três classes ($M = 3$). 73

Figura 13 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de terceira ordem com três classes ($M = 3$). 74

Figura 14 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de segunda ordem com quatro classes ($M = 4$). 74

Figura 15 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de terceira ordem com quatro classes ($M = 4$). 74

Figura 16 - Abordagens para solução de problemas em reconhecimento de padrões estatístico: conforme se percorre a árvore do topo até as extremidades inferiores, da esquerda para a direita, menos informação se encontra disponível. 79

Figura 17 - Superfícies de separação geradas por um classificador de árvore de decisão para um problema de duas classes bidimensional 85

Figura 18 - Estrutura de um perfil de decisão. 89

Figura 19 - Diagrama de construção do *phantom* utilizado para imageamento e estudo de materiais presentes no solo. 91

Figura 20 - Bandas de imagem multiespectral de ciência do solo, adquirida através de raios X e γ utilizando múltiplas energias. 91

Figura 21 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma maçã. 92

Figura 22 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma pera. 92

Figura 23 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma goiaba. 92

Figura 24 - Bandas de uma imagem multiespectral de ressonância magnética nuclear do cérebro de um *marmocet*. 93

Figura 25 - Mapas de classes para classificação pontual da imagem goiaba. . . . 99

Figura 26 - Mapas de classes para classificação contextual da imagem goiaba. . . 99

Figura 27 - Mapas de classes para classificação pontual da imagem maçã. 99

Figura 28 - Mapas de classes para classificação contextual da imagem maçã. . . . 100

Figura 29 - Mapas de classes para classificação pontual e contextual do phantom de tomografia computadorizada de ciência do solo. 100

Figura 30 - Mapas de classes para classificação pontual e contextual da imagem de cérebro de um *marmocet*. 100

Figura 31 - Mapas de classes para classificação contextual da imagem maça para diferentes inicializações. 101

Figura 32 - Mapas de classes para classificação contextual da imagem maça utilizando sistemas de vizinhança de segunda ordem. 103

Figura 33 - Mapas de classes para classificação contextual da imagem maça utilizando sistemas de vizinhança de terceira ordem. 103

Figura 34 - Mapas de classes para classificação contextual da imagem goiaba utilizando sistemas de vizinhança de segunda ordem. 104

Figura 35 - Mapas de classes para classificação contextual da imagem goiaba utilizando sistemas de vizinhança de terceira ordem. 104

Figura 36 - Mapas de classes para classificação contextual da imagem de cérebro utilizando sistemas de vizinhança de segunda ordem. 104

Figura 37 - Mapas de classes para classificação contextual da imagem de cérebro utilizando sistemas de vizinhança de terceira ordem. 105

Figura 38 - Mapas de classes para classificação contextual da imagem pêra utilizando a mesma inicialização (PARZENC) em algoritmos de otimização combinatória distintos. 106

Figura 39 - Mapas de classes para classificação contextual da imagem pêra utilizando a mesma inicialização (LDC) em algoritmos de otimização combinatória distintos. 107

Figura 40 - Mapas de classes para classificação contextual da imagem de cérebro de um *marmocet* utilizando a mesma inicialização (NMC) em algoritmos de otimização combinatória distintos. 107

Figura 41 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo ICM com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15). 109

Figura 42 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo GSA com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15). 109

Figura 43 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo MPM com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15). 110

Figura 44 - Mapas de classes para o melhor desempenho individual dentro do esquema proposto (KNNC+MPM) e para o melhor desempenho utilizando a combinação dos três algoritmos iterativos considerando a mesma inicialização (KNNC). 111

Figura 45 - Mapas de classes para o melhor desempenho individual dentro do esquema proposto (KNNC+MPM) e para o melhor desempenho utilizando a combinação dos três algoritmos iterativos considerando a diferentes inicializações (PARZENC, QDC e NMC). 111

Figura 46 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo ICM em sistemas de vizinhança de segunda e terceira ordens (Regra do Máximo). 112

Figura 47 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo GSA em sistemas de vizinhança de segunda (Regra do Máximo) e terceira ordens (Regra da Soma). 112

Figura 48 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo MPM em sistemas de vizinhança de segunda e terceira ordens (Regra da Soma). 113

Lista de Tabelas

Tabela 1 - Equivalência entre conceitos de Teoria dos Jogos e o problema de classificação contextual modelado por Campos Aleatórios Markovianos 50

Tabela 2 - Número de padrões de configuração espacial para diversos sistemas de vizinhança de ordens superiores. 60

Tabela 3 - Vetores soluções representando os possíveis padrões de configuração espacial em sistemas de segunda ordem. 60

Tabela 4 - Estimadores de MPV, variâncias assintóticas e intervalos de confiança de 90% para os parâmetros do modelo GMRF em imagens simuladas (Fig. 10a) 69

Tabela 5 - Estimadores de MPV, variâncias assintóticas e intervalos de confiança de 90% para os parâmetros do modelo GMRF em imagens simuladas (Fig. 10b) 69

Tabela 6 - Estimadores de MPV, informação de Fisher observada, variâncias assintóticas, estatísticas de teste e probabilidades de significância para as imagens sintéticas geradas a partir de simulação de Monte Carlo usando sistemas de vizinhança de segunda ordem 75

Tabela 7 - Estimadores de MPV, informação de Fisher observada, variâncias assintóticas, estatísticas de teste e probabilidades de significância para as imagens sintéticas geradas a partir de simulação de Monte Carlo usando sistemas de vizinhança de terceira ordem 75

Tabela 8 - Desempenho da classificação em função do Kappa. 87

Tabela 9 - Regras de combinação de classificadores para fusão de decisões. . . . 90

Tabela 10 - Resultados referentes ao teste da hipótese A 98

Tabela 11 - Resultados referentes ao teste de hipótese B: Kappas obtidos para um mesmo algoritmo iterativo são significativamente distintos 101

Tabela 12 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (`PARZENC`) em diferentes algoritmos de classificação contextual para a imagem de pêra. 105

Tabela 13 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (LDC) em diferentes algoritmos de classificação contextual para a imagem de pêra. 105

Tabela 14 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (NMC) em diferentes algoritmos de classificação contextual para a imagem de cérebro de um *marmocet*. 106

Tabela 15 - Comparação entre os melhores desempenhos individuais e o melhor desempenho obtido com a combinação de todas as inicializações para os algoritmos ICM, GSA e MPM em sistemas de vizinhança de segunda ordens. 108

Tabela 16 - Comparação entre os desempenhos médios ao se utilizar uma única inicialização e múltiplas inicializações para os algoritmos ICM, GSA e MPM. 110

Tabela 17 - Comparação entre o desempenho da classificação e o tempo de execução ao se variar o número de inicializações para os algoritmos de otimização combinatória ICM, GSA e MPM. 113

Tabela 18 - Resultados da classificação pontual da imagem de ciência do solo . . 135

Tabela 19 - Resultados da classificação MPM para o phantom de ciência do solo 135

Tabela 20 - Resultados da classificação ICM para o phantom de ciência do solo . 136

Tabela 21 - Resultados da classificação GSA para o phantom de ciência do solo . 136

Tabela 22 - Resultados da classificação pontual da imagem RMN de maçã 137

Tabela 23 - Resultados da classificação MPM para a imagem RMN de maçã . . . 137

Tabela 24 - Resultados da classificação ICM para a imagem RMN de maçã . . . 138

Tabela 25 - Resultados da classificação GSA para a imagem RMN de maçã . . . 138

Tabela 26 - Resultados da classificação pontual da imagem RMN de pêra 139

Tabela 27 - Resultados da classificação MPM para a imagem RMN de pêra . . . 139

Tabela 28 - Resultados da classificação ICM para a imagem RMN de pêra 140

Tabela 29 - Resultados da classificação GSA para a imagem RMN de pêra 140

Tabela 30 - Resultados da classificação pontual da imagem RMN de goiaba . . . 141

Tabela 31 - Resultados da classificação MPM para a imagem RMN de goiaba . . 141

Tabela 32 - Resultados da classificação ICM para a imagem RMN de goiaba . . . 142

Tabela 33 - Resultados da classificação GSA para a imagem RMN de goiaba . . 142

Tabela 34 - Resultados da classificação pontual da imagem RMN de cérebro . . . 143

Tabela 35 - Resultados da classificação MPM para a imagem RMN de cérebro . 143

Tabela 36 - Resultados da classificação ICM para a imagem RMN de cérebro . . 144

Tabela 37 - Resultados da classificação GSA para a imagem RMN de cérebro . . 144

Tabela 38 - Resultados da classificação MPM da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores 145

Tabela 39 - Resultados da classificação ICM da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores 146

Tabela 40 - Resultados da classificação GSA da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores 146

Lista de Abreviaturas

MRF	Markov Random Field
ICM	Iterated Conditional Modes
MPM	Maximizer of the Posterior Marginals
GSA	Game Strategy Approach
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
PARZENC	Parzen Windows Classifier
KNNC	K-Nearest-Neighbor Classifier
LOGLC	Logistic Classifier
LDC	Linear Discriminant Classifier
QDC	Quadratic Discriminant Classifier
NMC	Nearest Mean Classifier
TREEC	Decision Tree Classifier
HMM	Hidden Markov Model
MLP	Multi-Layer Perceptron
RNA	Rede Neural Artificial
GNC	Graduated Non-Convexity
HCF	Highest Confidence First
SA	Simulated Annealing
GMRF	Gaussian Markov Random Field
MAP	Maximum a Posteriori

NMR	Nuclear Magnetic Ressonance
CT	Computerized Tomography
CART	Classification and Regression Tree
MCMC	Markov Chain Monte Carlo
GRF	Gibbs Random Field
LOOCV	Leave-One-Out Cross-Validation
PWI	Pairwise Interaction Model
LCDF	Local Conditional Density Function
MV	Máxima Verossimilhança
MPV	Máxima Pseudo-Verossimilhança
SVM	Support Vector Machines

Sumário

1 Introdução **23**

1.1 Contexto, motivação e definição do problema 23

1.2 Uma visão geral da metodologia proposta 25

1.3 Trabalhos relacionados 28

1.4 Organização do trabalho 31

2 Classificação Contextual e Estimação Bayesiana **33**

2.1 Modelos Markovianos para Classificação 37

2.2 Algoritmos Iterativos de Otimização Combinatória 43

3 Inferência Estatística em Modelos Markovianos **55**

3.1 Estimação por Máxima Pseudo-Verossimilhança 56

3.2 Aproximação para a Variância Assintótica dos Estimadores de MPV 64

3.3 Considerações Finais 75

4 Metodologia Proposta **77**

4.1 Classificação Supervisionada 77

4.2 Medidas Quantitativas de Desempenho 85

4.3 Regras de Combinação de Classificadores 89

4.4 Experimentos e Resultados Obtidos 90

5 Considerações Finais **115**

5.1 Trabalhos Publicados 116

Referências **119**

Apêndice A – Informação de Fisher no modelo de Potts **131**

A.1 Informação de Fisher observada para o modelo de Potts usando a primeira derivada 131

A.2 Informação de Fisher observada para o modelo de Potts usando a segunda derivada 132

Apêndice B – Resultados Quantitativos **135**

B.1 Coeficientes Kappa e Tau referentes à imagem de ciência do solo 135

Apêndice C – Resultados Quantitativos **137**

C.1 Coeficientes Kappa e Tau referentes à imagem RMN de maçã 137

Apêndice D – Resultados Quantitativos **139**

D.1 Coeficientes Kappa e Tau referentes à imagem RMN de pêra 139

Apêndice E – Resultados Quantitativos **141**

E.1 Coeficientes Kappa e Tau referentes à imagem RMN de goiaba 141

Apêndice F – Resultados Quantitativos **143**

F.1 Coeficientes Kappa e Tau referentes à imagem RMN de cérebro 143

Apêndice G – Resultados Quantitativos **145**

G.1 Coeficientes Kappa e Tau referentes à combinação de múltiplas inicializações . 145

Capítulo 1

Introdução

1.1 Contexto, motivação e definição do problema

Os primeiros trabalhos envolvendo modelos de campos aleatórios (MRF) datam do início da segunda metade do século XX, mais especificamente das décadas de 50, 60 e 70 (1–6). Porém, foi a partir de resultados avançados em probabilidade e estatística aplicadas, como o Teorema de Hammersley-Clifford (7), e posteriormente o memorável trabalho dos irmãos Geman (8), que essa teoria, antes restrita apenas a física estatística e a matemática, foi introduzida na área de processamento de imagens e recentemente a inúmeras aplicações em visão computacional e reconhecimento de padrões, dentre elas uma de fundamental importância: a classificação de imagens.

O problema de se classificar/segmentar imagens é uma das tarefas mais desafiadoras em aprendizado de máquina (tanto supervisionado quanto não supervisionado), pois se trata de um procedimento extremamente complexo. Parte dessa complexidade advém da enorme variedade de imagens existentes, tais como imagens de cenas naturais, imagens tomográficas, de ressonância magnética, de ultra-sonografia, dentre outras, o que acaba dificultando muito a definição de métodos robustos o suficiente para classificar os tipos de dados encontrados em imagens. Em geral, cada tipo de imagem possui dados com características próprias, sendo adequadamente representado por um modelo estatístico específico. Outro sério problema encontrado na classificação de imagens é a presença de ruídos, que distorcem os dados observados. Freqüentemente, ruídos são inerentes ao processo de aquisição de imagens reais. Por exemplo, em tomografia computadorizada, baixos tempos de exposição causam o aparecimento de ruído Poisson nas projeções. Além disso, existe ainda o problema do ruído gaussiano, que aparece praticamente em todos os sistemas eletrônicos. Na grande maioria dos casos, os métodos tradicionais de classificação

não fornecem resultados satisfatórios para imagens ruidosas.

Diante do exposto, a proposta desse trabalho consiste em combinar modelos de campos Markovianos em sistemas de vizinhança de ordens superiores para modelar tanto as observações (GMRF) quanto o conhecimento *a priori* (Modelo de Potts) seguindo o paradigma Bayesiano (abordagem MAP-MRF), com o intuito de superar as limitações encontradas nos métodos de classificação tradicionais. A grande vantagem da abordagem MAP-MRF é que ela permite o desenvolvimento de técnicas e algoritmos para solução de problemas inversos de maneira sistemática, seguindo princípios formais e teoricamente embasados ao invés de simples heurísticas.

Do ponto de vista computacional, a classificação contextual de imagens através da abordagem de campos aleatórios Markovianos consiste basicamente em, dada uma solução inicial, obtida por um classificador pontual (supervisionado ou não supervisionado), melhorar essa solução iterativamente e de maneira local através de algortimos de otimização combinatória, até que um determinado critério de convergência seja satisfeito. Dentre os diversos algoritmos sub-ótimos para classificação contextual de imagens podemos citar o ICM (*Iterated Conditional Modes*) (9), MPM (*Maximizer of the Posterior Marginals*) (10), GSA (*Game Strategy Approach*) (11), GNC (*Graduated Non-Convexity*) (12–13) e o HCF (*Highest Confidence First*) (14).

A maior vantagem desses métodos sub-ótimos é a velocidade de convergência, principalmente quando comparados ao método ótimo (que obtém o máximo global), conhecido como Simulated Annealing (SA) (15), um algoritmo de relaxação estocástica, que consegue escapar de máximos locais, através de pequenas perturbações aleatórias na função objetivo (energia do sistema). Resultados encontrados na literatura comparando o custo computacional dos diversos métodos de otimização combinatória mostram significativas diferenças no tempo de execução gasto em cada um deles. Por exemplo, em (16), é reportado que para imagens em tons de cinza, em termos relativos, o algoritmo MPM consegue ser da ordem de cem vezes mais rápido que o SA, enquanto o eficiente ICM chega a ser da ordem de mil vezes mais rápido. Em imagens com dimensões espaciais um pouco maiores, o método chega a ser computacionalmente inviável. Portanto, uma das grandes motivações para esse projeto é justamente explorar essa imensa diferença entre a velocidade de convergência do método ótimo (SA) e dos métodos sub-ótimos, com o intuito de propor novas metodologias para classificação contextual de imagens multiespectrais via campos aleatórios Markovianos, utlilizando múltiplas inicializações, sistemas de vizinhança de ordens superiores e combinação de algoritmos de otimização combinatória.

Mas para que a classificação contextual seja possível, primeiramente é necessário se estimar os parâmetros dos modelos Markovianos de maneira adequada e precisa. Novas equações para estimação de parâmetros no modelo de Potts utilizando máxima pseudo-

verossimilhança em sistemas de vizinhança de ordens superiores foram propostas. Estudos acerca da acurácia dos métodos de estimação propostos, através da derivação de aproximações para a variância assintótica dos estimadores também foram desenvolvidos, fornecendo um forte embasamento teórico para os resultados experimentais obtidos.

1.2 Uma visão geral da metodologia proposta

Basicamente, a metodologia proposta para classificação contextual de imagens multiespectrais segue as etapas ilustradas pelo diagrama de blocos da Figura 1. Dada uma imagem multiespectral de entrada, o objetivo é classificar cada pixel, mais precisamente cada vetor de padrões N dimensional, em uma das possíveis C classes. O primeiro passo do método proposto é definir os modelos estatísticos, tanto espectrais (para os dados multivariados) quanto contextuais (Markovianos). A seguir, os conjuntos de treinamento e teste são obtidos a partir das regiões de interesse (abordagem supervisionada, pois existe um subconjunto de amostras pré-rotuladas). A partir dos dados de treinamento, os parâmetros das distribuições condicionais de cada classe são estimados. Tipicamente, sob hipótese de dados gaussianos multivariados ou ainda assumindo um modelo `GMRF` para as observações, como é o caso nesse trabalho, estimam-se os vetores média e as matrizes de covariância de cada classe. Eventualmente, pode ser necessária a utilização de técnicas de extração de atributos, tais como Análise dos Componentes Principais (`PCA`) (17–19), Análise Discriminante Linear (`LDA`) (19–20) e Análise de Componentes Independentes (`ICA`) (21–22), visando a obtenção um subconjunto de características que melhor represente os dados observados do ponto de vista de discriminação entre classes.

A essa altura, chega-se à etapa de classificação supervisionada, onde diferentes classificadores pontuais são utilizados para gerar diversas inicializações (mapas de rótulos) como entrada para os algoritmos iterativos de otimização combinatória, encarregados de efetivamente realizar a classificação contextual utilizando os modelos de campos aleatórios Markovianos. Dentre as diversas técnicas de classificação existentes na literatura, foram selecionados sete classificadores dentro da abordagem estatística de reconhecimento de padrões: Bayesiano Linear (`LDC`) e Quadrático (`QDC`), K-Vizinhos mais próximos (`KNN`), janelas de Parzen (`PARZENC`), Logístico (`LOGLC`), Mínima Distância (`NMC`) e Árvore de Decisão (`TREEC`). Convém ressaltar que, para a obtenção das diversas inicializações (mapas de rótulos iniciais), pode-se variar também outras etapas anteriores, fixando um mesmo classificador pontual. Por exemplo, é possível modificar as condições iniciais variando também as amostras de treinamento, e a etapa de extração de atributos.

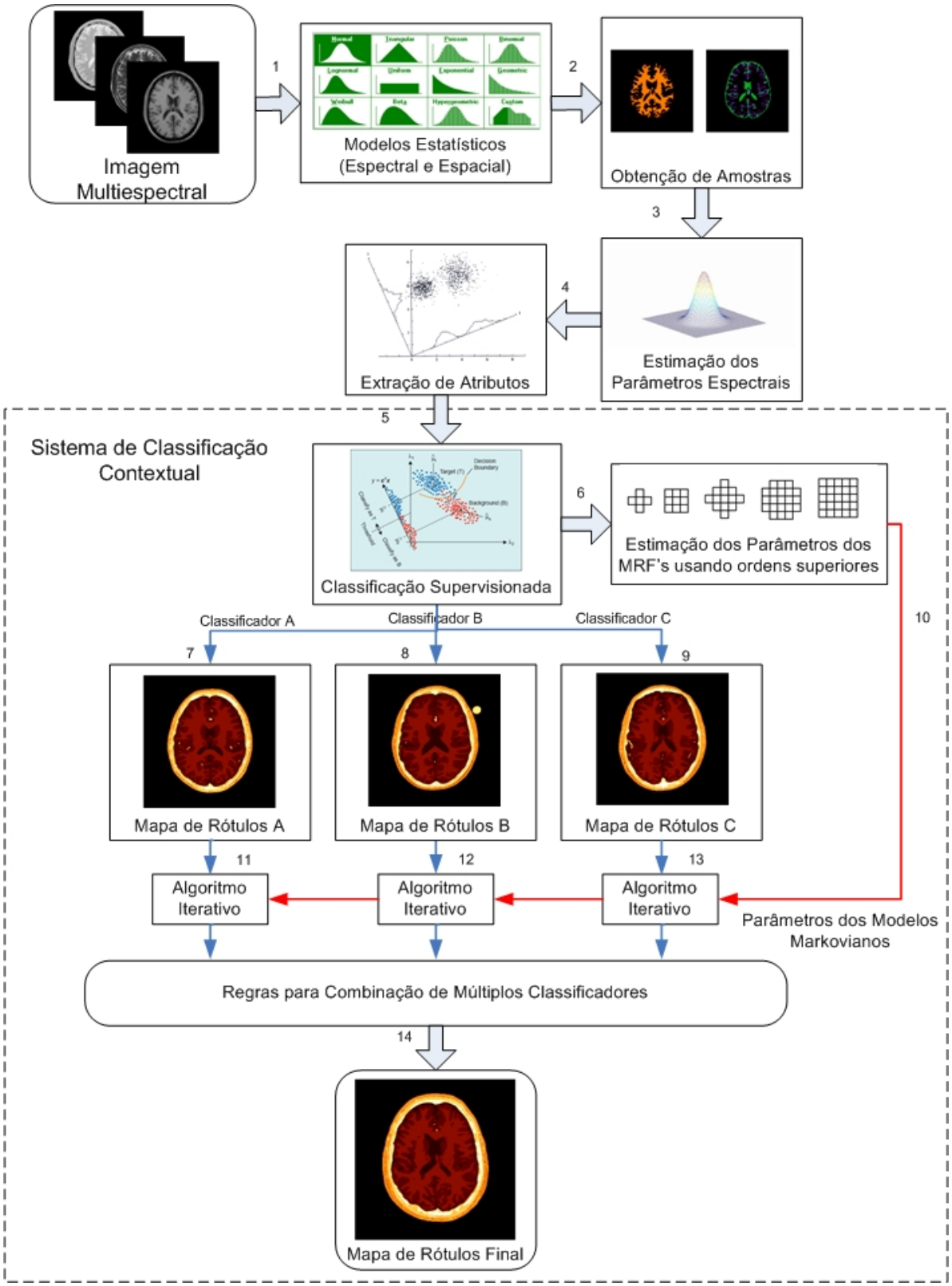


Figura 1 - Diagrama de blocos do sistema proposto para classificação contextual de imagens multiespectrais utilizando combinação de algoritmos sub-ótimos de otimização combinatória.

Deve-se notar a importância de uma metodologia de estimação de parâmetros adequada no processo de classificação contextual, uma vez que toda a evolução de cada um dos campos de rótulos iniciais depende diretamente dos valores dos parâmetros utilizados no cálculo das probabilidades. Como as inicializações são diferentes, cada uma delas evolui de maneira própria para um ótimo local, seguindo trajetórias distintas ao longo do espaço de soluções. Finalmente, a etapa de combinação de classificadores incorpora as informações provenientes de diversos contextos no processo de tomada de decisão, através da utilização de combinadores baseados em regras. É possível fazer uma analogia entre essa etapa e uma reunião de especialistas. Numa situação como essa, a decisão final não depende apenas do parecer de um único indivíduo, mas sim de uma função das decisões de cada indivíduo, que, no caso mais comum, é uma simples votação por maioria. Outras estratégias de combinação são as regras do produto, da soma, do máximo, do mínimo e da mediana (23–24), além de outros esquemas de votação, como o método *Borda Count*, por exemplo.

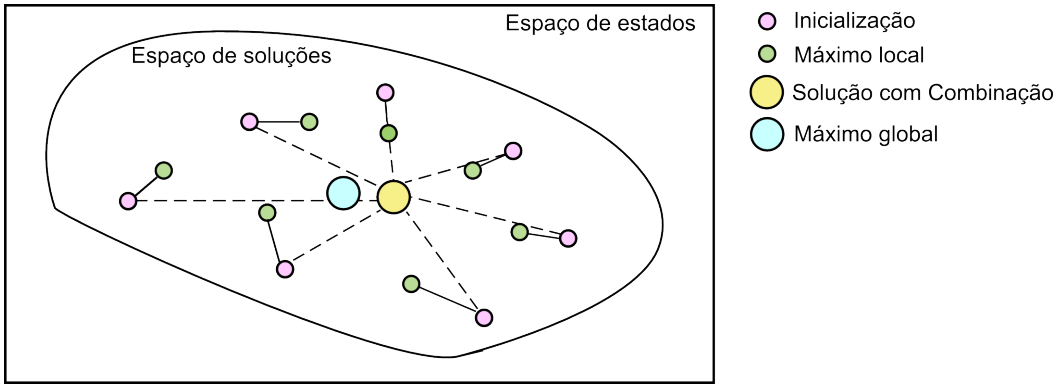


Figura 2 - O uso de múltiplas condições iniciais torna possível cobrir uma região muito maior do espaço de estados, possibilitando convergência a melhores resultados.

Basicamente, o objetivo principal da metodologia proposta consiste em investigar o efeito da utilização de múltiplas inicializações em paralelo em algoritmos de otimização combinatória sub-ótimos (`ICM`, `MPM`, `GSA`) responsáveis pela maximização da probabilidade *a posteriori* na estimação Bayesiana. Trata-se de um cenário recorrente em problemas MAP-MRF. Espera-se que com tal abordagem seja possível conseguir boas soluções como alternativas ao algoritmo *Simulated Annealing*, extremamente custoso do ponto de vista computacional, em diversas aplicações. A Figura 2 ilustra como a utilização de múltiplas inicializações pode afetar o desempenho final fazendo com que se consiga convergência a melhores soluções, evitando mínimos locais ao cobrir uma região mais ampla do espaço de soluções.

Todo restante do documento está organizado de maneira a descrever as etapas envolvidas na metodologia proposta, desde a formulação inicial do problema até a abordagem

de combinação de classificadores contextuais, com ênfase para a etapa de estimação dos parâmetros dos modelos Markovianos, fundamental dentro do *framework* desenvolvido.

1.3 Trabalhos relacionados

Esta seção apresenta o panorama atual da área de classificação contextual de imagens, com ênfase na estimação de parâmetros em modelos Markovianos. São apresentadas algumas vantagens e desvantagens das abordagens existentes que podem ser consideradas o estado da arte em classificação contextual.

1.3.1 Estimação de parâmetros em campos Markovianos

Sem dúvida, uma das maiores dificuldades na abordagem MAP-MRF para solução de problemas inversos é justamente a estimação dos parâmetros dos modelos de campos aleatórios Markovianos. Em parte, isso ocorre porque o método mais utilizado na prática, Máxima Verossimilhança, não pode ser aplicado devido a existência da função de partição na distribuição de Gibbs conjunta, sendo computacionalmente intratável. Por esse motivo, atualmente, na maioria das aplicações esses parâmetros ainda são ajustados manualmente por tentativa e erro (25–26). Por essa razão, foi constatada a existência de uma grande lacuna na literatura contendo diversos problemas em aberto relacionados a inferência em modelos Markovianos. Embora diversas abordagens para estimação de parâmetros em modelos Markovianos existam, dentre as quais pode-se destacar *máxima pseudo-verossimilhança* (27–28), a abordagem conhecida como *coding method* (6), aproximações *mean field* (29–30), o método *least square fit* (31) e métodos baseados em algoritmos *Markov Chain Monte Carlo* (MCMC) (32–35), pouco se conhece acerca da acurácia de tais métodos.

Embora cada um dos métodos citados acima tenha suas próprias vantagens e desvantagens, um estudo mais completo reportando uma comparação objetiva entre eles ainda não existe, e certamente representa um desafio em aberto. Seus desempenhos e custos computacionais tem sido pontualmente estudados em problemas envolvendo campos Markovianos, mas apenas em casos isolados. Outro fator limitante é que a grande maioria dos métodos de estimação utiliza sistemas de vizinhança muito restritivos, o que pode ser

insuficiente para uma modelagem contextual adequada dependendo do problema. Nessa seção serão discutidos alguns aspectos das várias abordagens de estimação de parâmetros MRF's, justificando a opção adotada.

Uma possível alternativa para a estimação por máxima pseudo-verossimilhança é a técnica conhecida como *coding method* (6). A idéia básica dessa abordagem é dividir a imagem numa série de subconjuntos disjuntos, chamados de *codings*, de tal forma que quaisquer dois elementos pertencentes a subconjntos distintos não sejam vizinhos. Segundo (36), essa condição permite a definição de uma expressão genuina para a função de verossimilhança. Entretanto, a desvantagem dessa abordagem é que, como um estimador é obtido para cada subconjunto independentemente, não é claro como combiná-los de maneira ótima. Além disso, pouco se conhece a respeito de suas propriedades estatísticas.

Outra técnica de estimação existente é diretamente derivada da física estatística, sendo conhecida como aproximação de campo médio (*Mean Field Approximation*) (29–30, 37–38), que tenta aproximar o comportamento de sistemas físicos de partículas em condições de equilíbrio. Nessas situações, a aproximação para a distribuição conjunta torna-se o produtório das probabilidades locais, fazendo com que essa aproximação assuma uma forma bastante parecida com a função de máxima pseudo-verossimilhança (27). Pode ser mostrado que, sob certas condições, as duas abordagens coincidem (36), indicando que ambas estão de fato fortemente relacionadas.

A estimação de parâmetros em modelos Markovianos também pode ser realizada pela técnica de mínimos quadrados conhecida como *Least Squares Fit*, originalmente proposta por (31) na literatura de processamento de imagens. Basicamente, essa abordagem utiliza técnicas de histograma para estimar a distribuição conjunta de pixels de uma imagem. Existem trabalhos mostrando a relação entre técnicas de histograma e a estimação por máxima-verossimilhança (39–40). De acordo com (36), a abordagem *Least Squares Fit* possui algumas vantagens como não requerer otimizações numéricas, nem a solução de equações não-lineares. Entretanto, uma séria desvantagem é que o número de equações lineares a serem resolvidas aumenta exponencialmente com o tamanho da imagem e com o conjunto de rótulos, uma forte restrição a problemas de classificação com muitas classes. Além disso, pouco se sabe acerca das propriedades estatísticas desses estimadores.

Métodos de simulação de Monte Carlo (8, 41–45), mais precisamente algoritmos MCMC (*Markov Chain Monte Carlo*) tem se mostrado úteis em diversos problemas de estimação. Basicamente, essa abordagem permite o cálculo de valores esperados de funções de várias variáveis através da geração de amostras aleatórias das distribuições conjuntas. A grande desvantagem dessa abordagem é o custo computacional e o tempo gasto para a obtenção das amostras, o que a torna inadequada a algumas aplicações reais.

Adicionalmente a boas propriedades estatísticas, sólida fundamentação matemática e

tratabilidade computacional, outra vantagem da estimação por máxima pseudo verossimilhança se refere ao uso de sistemas de vizinhanças de ordens superiores, o que permite uma melhor modelagem espacial. O conceito de ordem adotado ao longo desse trabalho refere-se à extensão da vizinhança ao redor do elemento (pixel) observado. Recentemente, equações de pseudo-verossimilhança para o modelo de Potts em sistemas de segunda e terceira ordens foram propostas (46). Avaliações assintóticas revelaram que o aumento na vizinhança torna a estimação mais precisa, reduzindo tanto o erro (viés) quanto a variância assintótica dos estimadores (47–48).

1.3.2 Classificação contextual de imagens

Definitivamente, a inclusão de informação contextual na tomada de decisão traz benefícios a classificação de dados espaciais. Freqüentemente, campos aleatórios Markovianos vem sendo utilizados com sucesso na classificação de imagens (49–55). Como a estimação de parâmetros é uma tarefa árdua em problemas MAP-MRF, novas técnicas baseadas em inferência Bayesiana variacional (56) tem sido propostas para estimação conjunta dos parâmetros e do mapa de classes num *framework* integrado (57–60). Entretando, a abordagem variacional é de difícil tratabilidade matemática em modelos não gaussianos, o que de certa forma ainda restringe um pouco sua aplicação.

Abordagens hierárquicas para análise multi-resolução (52–53), aprendizado semi supervisionado (61–62) e mistura de distribuições (57, 63) também são algumas alternativas interessantes e que estão se difundindo rapidamente na literatura de reconhecimento de padrões. Um atenuante relacionado a esse tipo de de abordagem é o aumento tanto no custo quanto no tempo de computação, devido a utilização do algoritmo EM (*Expectation-Maximization*), que possui convergência relativamente lenta.

A utilização de métodos baseados na teoria dos grafos para segmentação/classificação de imagens (64–65) é uma técnica amplamente adotada na solução de problemas MAP-MRF, principalmente devido a sua velocidade e eficiência computacional. Nesse cenário, pode ser mostrado que o problema de estimação MAP é equivalente ao problema de fluxo máximo e corte mínimo (*max-flow/min-cut*) no grafo correspondente. Entretanto, a desvantagem de tais métodos é a restrição existente nas funções de energia que podem ser minimizadas: apenas funções satisfazendo certas condições específicas podem ser minimizadas através do problema de corte mínimo (66).

Dessa forma, a grande vantagem do método proposto em relação as abordagens atuais

existentes na literatura é a possibilidade de se incorporar uma quantidade muito maior de informação na classificação de uma amostra, tanto pelo fato de se adotar sistemas de vizinhança de ordens superiores quanto pelo fato de ser possível utilizar múltiplas inicializações simultaneamente. Além disso, o método proposto utiliza algoritmos de rápida convergência, com baixo custo computacional e nenhum pouco restritivos em termos da modelagem estatística adotada, ou seja, é viável para uma ampla variedade de modelos.

1.4 Organização do trabalho

Além deste capítulo introdutório, contendo a motivação e os objetivos do projeto em questão, essa monografia apresenta outros cinco capítulos, divididos em: Classificação Contextual de Imagens via Estimação Bayesiana, Inferência Estatística e Estimação de Parametros em Modelos Markovianos, Metodologia Proposta, Conclusões e Referências Bibliográficas.

O capítulo 2 descreve a formulação matemática do problema de classificação contextual dentro do paradigma de estimação Bayesiana. Diferentes funções de custo levam a diferentes estimadores Bayesianos. Nesse projeto, os objetos de estudo são dois tipos de estimadores em particular: o estimador `MAP` (*Maximum a Posteriori*) e o estimador `MPM` (*Maximizer of the Posterior Marginals*), pois podem ser aproximados por métodos computacionais viáveis como os algoritmos `ICM` (*Iterated Conditional Modes*), `GSA` (*Game Strategy Approach*) e o próprio `MPM`. A seção 2.1 introduz os modelos Markovianos adotados para classificação contextual - o modelo `GMRF` para as observações e modelo de Potts para os dados - mostrando a regra de decisão obtida combinando esses dois modelos. A seção 2.2 descreve os algoritmos iterativos de otimização combinatória utlizados nesse trabalho.

O capítulo 3 fornece uma descrição detalhada sobre inferência estatística e estimação de parâmetros nos modelos Markovianos adotados. São apresentadas novas equações de pseudo-verossimilhança para o modelo de Potts para estimação de parâmetros em sistemas de segunda e terceira ordens. A acurácia da estimação dos parâmetros dos modelos `GMRF` e de Potts é verificada através da derivação de aproximações para a variância assintótica dos estimadores e da simulação computacional por métodos de Monte Carlo.

O capítulo 4 descreve em detalhes a metodologia proposta, com ênfase nos resultados de classificação obtidos através da combinação dos classificadores contextuais. A seção 4.1 descreve os classificadores estatísticos utlizados para gerar diferentes inicializações

aos algoritmos de otimização combinatória. Na seção 4.2 são discutidas duas medidas quantitativas para a avaliação do desempenho da classificação: o coeficiente *Kappa* de Cohen e o coeficiente *Tau* de Kendall. Experimentos utlizando imagens multiespectrais de tomografia computadorizada e ressonância magnética foram desenvolvidos. Análises estatísticas dos resultados fornecem evidências acerca de diversas hipóteses levantadas sobre a metodologia proposta, mostrando que o método proposto é capaz de melhorar significativamente o desempenho da classificação contextual.

Finalmente, o capítulo 5 apresenta as conclusões e considerações finais, bem como as perspectivas futuras para a continuidade das pesquisas relacionadas ao projeto.

Capítulo 2

Classificação Contextual e Estimação Bayesiana

Diversos problemas nas áreas de processamento de imagens e visão computacional são mal condicionados por natureza simplesmente porque o espaço de soluções é extremamente vasto (67). Mais formalmente, o mal condicionamento surge em decorrência de problemas que não apresentam uma das seguintes condições: existência/unicidade da solução, continuidade e estabilidade (68–69). Uma das maneiras de se reduzir o espaço de soluções, e como conseqüência atenuar o efeito do mal condicionamento, é a incorporação de conhecimento *a priori* sob a forma de restrições aplicadas à solução desejada. Na classificação contextual de imagens, uma maneira de introduzir conhecimento a priori na formulação do problema é através da abordagem Bayesiana, utilizando restrições de suavidade na forma de probabilidades *a priori*, que levam em consideração o contexto espacial dos dados. Nesse trabalho, o termo contexto se refere ao conjunto de elementos que pertencem a vizinhança espacial de uma determinada amostra a ser classificada (vetor de padrões).

Na teoria de estimação Bayesiana, o objetivo consiste em escolher o estimador $\hat{x} = \hat{x}(y)$ que minimiza o risco de Bayes (valor esperado da função de perda), onde a função de perda $C(x, \hat{x})$ determina o custo de se estimar $\hat{x}$ quando o verdadeiro valor é x, sendo estritamente positiva. Algumas das funções de custo mais utilizadas em estimação Bayesiana são: função de custo quadrática, que penaliza severamente erros relativamente altos, função de custo de valor absoluto, que penaliza de modo linearmente crescente os erros, e função de custo uniforme, que não penaliza o erro, apenas classifica como erro ou acerto, tolerando uma pequena margem ao redor do verdadeiro valor. A Figura 3, adaptada de (70), ilustra os gráficos das funções de custo descritas acima, onde At denota a diferença entre o valor estimado e o valor real.

Para o caso de uma função de custo quadrática, o estimador resultante é o conhecido

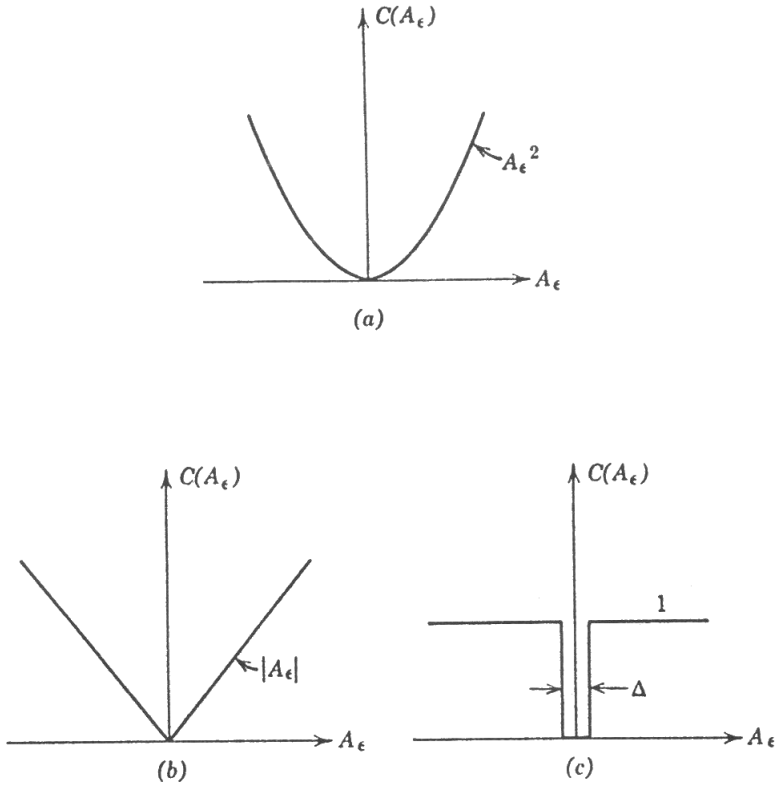


Figura 3 - Funções de custo para estimação Bayesiana: a.) função de custo quadrática, b.) de valor absoluto e c.) uniforme.

estimador de mínimo erro médio quadrático, ou `MMSE` (*Minimum Mean Square Estimation*), que é obtido pelo valor esperado da distribuição *a posteriori*. Porém, no caso de prioris Markovianas em classificação de imagens, as duas principais funções de custo adotadas na literatura são a função de custo uniforme e a função de custo denominada de distorção de *Hamming*, ou ainda distância de *Hamming*, definida como o número de erros de classificação cometidos. No caso de uma função de custo uniforme, quando todos os erros são tratados igualmente (tem o mesmo peso), pode-se definir:

$$C\left(x,\hat{x}\right)=1-\delta\left(x-\hat{x}\right)=\begin{cases}1, & \text{se} \quad x\neq\hat{x}\\ 0, & \text{se} \quad x=\hat{x}\end{cases} \tag{2.1}$$

Uma interpretação para essa função de custo é a seguinte: se $x\neq\hat{x}\left(y\right)$, não importa se a desigualdade é provocada por uma diferença apenas razoável ou gigantesca, o estimador deve ser idêntico ao ideal, obviamente permitindo uma pequena variação ao redor do verdadeiro valor, para ser uma função de custo nula. Trata-se de uma função de custo considerada conservadora (10).

O risco de Bayes é expresso como o valor esperado da função de perda. Aplicando a definição de valor esperado, fatorando a distribuição conjunta e agrupando os termos de

forma a reescrever o risco em termos do valor esperado condicional, tem-se (67):

$$\begin{aligned} E\left[C\left(X,\hat{x}\left(Y\right)\right)\right] &= \sum_{x}\sum_{y} C\left(x,\hat{x}\left(y\right)\right) P\left(x,y\right) \\ &= \sum_{x}\sum_{y} C\left(x,\hat{x}\left(y\right)\right) P\left(x|y\right) P\left(y\right) \\ &= \sum_{y} E\left[C\left(X,\hat{x}\left(Y\right)\right)|Y=y\right] P\left(y\right) \end{aligned} \tag{2.2}$$

Dessa forma, o estimador Bayesiano ótimo $\hat{x}$ é definido como a função que, para todo y minimiza o valor esperado da função de perda condicional, que no caso de função de custo uniforme, é definido como (67):

$$\begin{aligned} E\left[C\left(X,\hat{x}\left(Y\right)\right)|Y=y\right] &= \sum_{x} C\left(X=x,\hat{x}\left(y\right)\right) P\left(x|y\right) \\ &= \sum_{x}\left[\left(1-\delta\left(x-\hat{x}\left(y\right)\right)\right) P\left(x|y\right)\right] \\ &= \sum_{x} P\left(x|y\right) - \sum_{x} P\left(x|y\right)\delta\left(x-\hat{x}\left(y\right)\right) \\ &= 1 - P\left(\hat{x}\left(y\right)|Y=y\right) \end{aligned} \tag{2.3}$$

Assim, para minimizar o valor esperado da função de perda, no caso de uma função de custo uniforme, o estimador $\hat{x}\left(y\right)$ deve ser aquele que maximiza a probabilidade *a posteriori* $P\left(\hat{x}\left(y\right)|y\right)$, ou seja, deve satisfazer o critério `MAP` (*Maximum a Posteriori*). Embora esse critério seja considerado conservador, por tratar todos os erros identicamente, é o mais popular na solução de problemas em estimação Bayesiana. Isso acontece devido a algumas vantagens do estimador `MAP`, sendo a principal delas a fácil tratabilidade matemática. O processo de maximização da probabilidade *a posteriori* é equivalente à maximização de $P\left(y|x\right)P\left(x\right)$, ou ainda:

$$\hat{x}_{MAP} = \arg\max_{x} P\left(x|y\right) = \arg\max_{x}\left[logP\left(y|x\right) + logP\left(x\right)\right] \tag{2.4}$$

Ou seja, no critério `MAP` é fácil se incorporar restrições através do conhecimento *a priori*, na forma da distribuição $P\left(x\right)$, o que o torna aplicável a uma vasta gama de problemas. Porém, em diversas ocasiões, é mais razoável associar um custo proporcional ao número de erros cometidos (10, 71), como por exemplo na função de perda definida como a distância de *Hamming*, dada por:

$$C\left(x,\hat{x}\right) = \sum_{s\in S}\left[1-\delta\left(x_s-\hat{x}_s\right)\right] \tag{2.5}$$

onde x_s denota um elemento do campo aleatório e S o conjunto formado por todos os elementos do campo em questão. De acordo com essa função de perda, paga-se um preço muito maior por mil erros do que por apenas um único erro isolado. Definindo o valor esperado condicional da função de perda, exatamente como no caso anterior, e após algumas manipulações algébricas, tem-se a seguinte expressão para o risco de Bayes (67):

$$\begin{aligned} E\left[C\left(X,\hat{x}\left(Y\right)\right)|Y=y\right] &= \sum_{x} C\left(x,\hat{x}\left(y\right)\right) P\left(x|y\right) \\ &= \sum_{x}\sum_{s\in S}\left[1-\delta\left(x_s-\hat{x}_s\right)\right] P\left(x|y\right) \\ &= \sum_{s\in S}\sum_{x} P\left(x|y\right) - \sum_{s\in S}\left[\sum_{x}\delta\left[x_s-\hat{x}_s\right] P\left(x|y\right)\right] \\ &= N_1N_2 - \sum_{s\in S}\left[\sum_{x_s^-}\left(\sum_{x_s}\delta\left[x_s-\hat{x}_s\right] P\left(x|y\right)\right)\right] \\ &= N_1N_2 - \sum_{s\in S}\left[\sum_{x_s^-} P\left(x_s=\hat{x}_s,x_s^-|y\right)\right] \end{aligned} \tag{2.6}$$

onde N_1 e N_2 são as dimensões espaciais da da imagem S, $x_s^- = \{x_t : t \in S, t \neq s\}$ é o conjunto de todos os pixels da imagem, com exceção do pixel central x_s e $\sum_{x_s^-} P\left(x_s=\hat{x}_s,x_s^-|y\right)$ é a probabilidade *a posteriori* marginal referente a $P\left(x|y\right)$. Portanto, este estimador é obtido maximizando-se a probabilidade *a posteriori* marginal de cada elemento $s \in S$, e por isso é conhecido como *Maximizer of the Posterior Marginals*. Para cada $s \in S$, o estimador `MPM` $\hat{x}_s$ é obtido como:

$$\hat{x}_s\left(y\right) = \arg\max_{\hat{x}_s}\left[\sum_{x_s^-} P\left(x_s=\hat{x}_s,x_s^-|y\right)\right] = \arg\max_{x_s}\left[P\left(x_s|y\right)\right], \forall s \in S \tag{2.7}$$

A principal dificuldade de se utilizar o estimador `MPM` é a computação das probabilidades marginais *a posteriori*. Uma das possíveis soluções, que será a metodologia adotada nesse trabalho, é a utilização de métodos de simulação `MCMC` (*Markov Chain Monte Carlo*) (8, 41, 45, 72–74). A seção 2.2 descreve com detalhes todos os algoritmos de otimização combinatória utilizados na aproximação computacional tanto dos estimadores `MAP` quanto `MPM`.

2.1 Modelos Markovianos para Classificação

Os Campos Aleatórios Markovianos representam uma generalização para duas ou mais dimensões dos processos de Markov unidimensionais, conhecidos como Cadeias de Markov. Tradicionalmente, duas abordagens distintas foram desenvolvidas. A primeira delas conhecida como causal, caracterizada por sistemas de vizinhanças não simétricos, é uma extensão direta dos modelos ocultos de Markov, ou `HMM` (*Hidden Markov Models*) 1-D para o plano. Nessa abordagem, é preciso se impor um ordenamento artificial para compensar a falta da noção natural de causalidade dos dados espaciais em relação a séries temporais, o que freqüentemente causa o aparecimento de artefatos direcionais nos campos aleatórios (67). Porém, a vantagem da abordagem causal é a grande variedade de ferramentas matemáticas disponíveis para problemas envolvendo cadeias de Markov unidimensionais (75–77).

Na outra abordagem, o ponto de partida foram as idéias e ferramentas utilizadas na mecânica estatística para caracterizar a energia de sistemas físicos de partículas organizadas num reticulado bidimensional (78–79). Nesses modelos, a contribuição de cada partícula para a energia total do sistema depende da interação de cada elemento com seus vizinhos. Esse desenvolvimento expressa a natureza Markoviana do campo aleatório de maneira não-causal, sendo o principal foco desse trabalho. A seguir são apresentadas definições sobre os modelos de campos aleatórios utilizados na classificação contextual de imagens multiespectrais, bem como suas componentes fundamentais e a metodologia adotada.

A modelagem MAP-MRF para o problema de classificação contextual definido pela metodologia proposta segue o paradigma Bayesiano, com a definição de um modelo `GMRF` (*Gaussian Markov Random Field*) multiespectral para os dados (observações), além de um modelo para representar o conhecimento *a priori*, nesse caso o modelo de Potts. A solução pode então ser encontrada maximizando a probabilidade *a posteriori*, o que define um problema de otimização. A grande vantagem da inferência Bayesiana em relação a abordagem paramétrica clássica é que, dentro desse novo paradigma, todos os parâmetros são considerados variáveis aleatórias, sendo possível a definição de modelos estatísticos para todas as variáveis envolvidas no problema.

Tal modelagem foi originalmente proposta em (80–81), mas além de ser restrita apenas a sistemas de vizinhança de primeira ordem (definidos pelos quatro vizinhos mais próximos), os autores utilizaram apenas o algoritmo `ICM` (determinístico) para aproximar o estimador `MAP` do campo de rótulos resultante. Nossa metodologia utiliza sistemas de vizinhança de ordens superiores (2ª e 3ª ordens), bem como a inclusão e combinação de

algoritmos não determinísticos, como o MPM e o GSA na aproximação do estimador resultante. Vale ressaltar que no contexto desse trabalho, a ordem de um sistema de vizinhança refere-se a sua extensão, ou seja, quanto maior a ordem, maior a quantidade de informação contextual observada ao redor de uma amostra.

As seções a seguir descrevem os modelos Markovianos adotados, bem como os algoritmos iterativos utilizados.

2.1.1 Gaussian Markov Random Field

A abundância de variáveis aleatórias com distribuição normal em aplicações estatísticas motivou o aparecimento dos campos aleatórios Markovianos Gaussianos. A grande diferença desse modelo para os demais, é que se trata de um modelo para dados contínuos, ou seja, os elementos do reticulado bidimensional não se restringem a valores inteiros. Porém, uma limitação de se adotar a densidade Gaussiana na modelagem de pixels de imagens consiste no fato de que seu domínio engloba todo o conjunto dos reais, ao passo que os valores observados dos pixels são proporcionais à intensidade luminosa, sendo estritamente positivos. Mesmo assim, trata-se de um modelo amplamente utilizado na prática. Nesse projeto, a definição dos modelos Markovianos será realizada a partir das funções densidade condicionais locais, visto que do ponto de vista da aplicação a ser desenvolvida (classificação de imagens), é a representação mais adequada, pois permite associar uma probabilidade para cada vetor observado. A função densidade condicional local (LCDF) do modelo GMRF multivariado referente ao elemento (i, j) de uma imagem multiespectral pertencente à classe m, dada a vizinhança $y_{\eta_{ij}}$ ao seu redor, é dada pela seguinte expressão (81):

$$p\left(\vec{y}_{ij}|y_{\eta_{ij}}, x_{ij} = m\right) = \frac{1}{(2\pi)^{K/2}\left|\Sigma\right|^{1/2}} \exp\left\{-\frac{1}{2}\left[\vec{y}_{ij} - \vec{\mu}_m - \left(\theta_m^T \vec{y}_{\eta_{ij}} - 2\left(\sum_{ct}\theta^{ct}\right)\vec{\mu}_m\right)\right]^T \times \Sigma_m^{-1}\left[\vec{y}_{ij} - \vec{\mu}_m - \left(\theta_m^T \vec{y}_{\eta_{ij}} - 2\left(\sum_{ct}\theta^{ct}\right)\vec{\mu}_m\right)\right]\right\} \tag{2.8}$$

onde θ^{ct} é uma matriz diagonal cujos elementos são os parâmetros de dependência espacial nas direções horizontais, verticais e diagonais (4×4), $ct = 1, \ldots, K$, onde K é o número de bandas, θ^T é uma matriz definida empilhando-se as matrizes θ^{ct} referentes a cada uma das bandas da imagem $(4 \times 4K)$, ou seja, $\theta^T = \left[\theta^{ct1}, \theta^{ct2}, \ldots, \theta^{ctK}\right]$. A grande

vantagem desse modelo em relação a um modelo não Markoviano é a capacidade de modelar interações entre elementos vizinhos presentes no conjunto de dados observados, que nesse caso são bandas de imagens multiespectrais. Note que no caso de elementos independentes, ou seja, quando não existe dependência espacial entre elementos vizinhos (condição em que θ^{ct} são todos nulos), a expressão acima é simplificada para uma simples função densidade gaussiana multivariada. Para facilitar a compreensão, definem-se dois vetores de parâmetros separados $\vec{\Phi} = \{\vec{\mu}_i, \Sigma_i\}, i = 1, \ldots, C$, referente aos parâmetros espectrais (vetores médias e matrizes de covariância) e $\Theta = \{\theta^{ct}, ct = 1, \ldots, K\}$, referente aos parâmetros de interação entre pixels nas direções horizontais e verticais e diagonais em cada uma das bandas, sendo que C é o número de classes do problema de classificação. Assim, o modelo `GMRF` contém tanto parâmetros estatísticos correspondentes à informação espectral quanto à informação espacial, para modelagem de contexto.

Um exemplo ilustrativo para o caso de dados multivariados bidimensionais ($K = 2$) e sistema de vizinhança de primeira ordem é apresentado em (81). Nessa situação, as variáveis definidas na equação 2.8 são representadas como indica a Figura 4. Nesse caso, o vetor de padrões observado (amostra) é definida pelo vetor 2-D $\vec{y}_{ij} = [y^1_{ij} y^2_{ij}]$. Como se trata de um sistema de primeira ordem, cada elemento do vetor observado possui quatro vizinhos, dois na direção horizontal e outros dois na direção vertical. A interação entre o elemento central e seus vizinhos horizontais na primeira banda é modelada pelo parâmetro θ^{h1} e na segunda banda pelo parâmetro θ^{h2}. Analogamente, a interação entre o elemento observado e seus vizinhos verticais é modelado pelos parâmetros θ^{v1} e θ^{v2} em cada uma das respectivas bandas. Note que um elemento pertencente a uma determinada banda depende exclusivamente dos elementos vizinhos pertencentes aquela mesma banda. As matrizes e vetores referenciados pela equação 2.8 são especificados nas equações 2.9, 2.10, 2.11 e 2.12.

$$\vec{y}_{ij} = \begin{pmatrix} y^1_{ij} \\ y^2_{ij} \end{pmatrix}, \vec{y}_{\eta_{ij}} = \begin{pmatrix} y^1_{(i-1)j} + y^1_{(i+1)j} \\ y^2_{(i-1)j} + y^2_{(i+1)j} \\ y^1_{i(j-1)} + y^1_{i(j+1)} \\ y^2_{i(j-1)} + y^2_{i(j+1)} \end{pmatrix} \tag{2.9}$$

$$\vec{\mu}_m = \begin{pmatrix} \mu^1_m \\ \mu^2_m \end{pmatrix}, \Sigma_m = \begin{pmatrix} \sigma^{11}_m & \sigma^{12}_m \\ \sigma^{21}_m & \sigma^{22}_m \end{pmatrix} \tag{2.10}$$

$$\theta^{ct1} = \theta^h = \begin{pmatrix} \theta^{h1} & 0 \\ 0 & \theta^{h2} \end{pmatrix}, \theta^{ct2} = \theta^v = \begin{pmatrix} \theta^{v1} & 0 \\ 0 & \theta^{v2} \end{pmatrix} \tag{2.11}$$

$$\theta^T = \begin{pmatrix} \theta^h & \theta^v \end{pmatrix} = \begin{pmatrix} \theta^{h1} & 0 & \theta^{v1} & 0 \\ 0 & \theta^{h2} & 0 & \theta^{v2} \end{pmatrix} \tag{2.12}$$

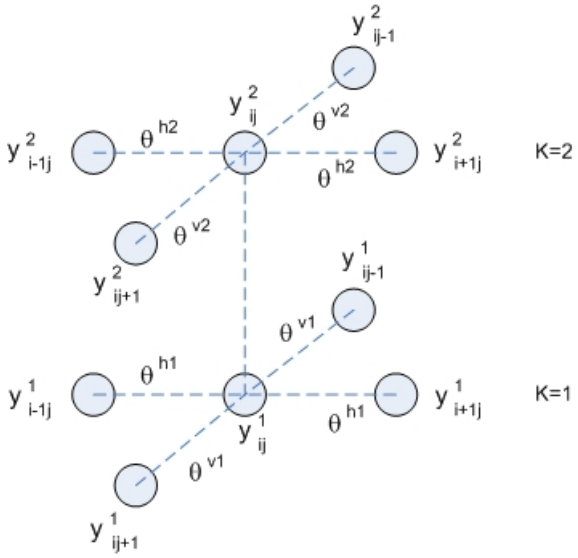


Figura 4 - Sistema de vizinhança para um GMRF multiespectral de primeira ordem.

Os vetores $\vec{y}_{ij}$ e $\vec{y}_{\eta_{ij}}$ representam, respectivamente, o vetor de padrões observado (a partir da imagem multiespectral) e o vetor formado pela soma dos elementos vizinhos em cada uma das direções (para cada uma das K bandas). O vetor $\vec{\mu}_m$ é o vetor média da classe m e Σ_m é a matriz de covariância da classe m. θ^h e θ^v são matrizes que contém os parâmetros de interação horizontal e vertical para cada banda e θ^T é a matriz resultante definida pela concatenação de θ^h e θ^v.

2.1.2 Modelo de Potts

O modelo Potts é um modelo Markoviano discreto, no sentido de que as variáveis pertencentes ao campo aleatório podem assumir apenas valores discretos pertencentes a um intervalo finito. Trata-se de um modelo que tenta imitar a maneira com que elementos individuais dos mais diversos tipos (i.e., átomos, animais, estruturas biológicas, indivíduos de uma sociedade, etc.) modificam seu comportamento conforme o comportamento de outros elementos pertencentes ao contexto em que se encontram. Em outras palavras, é um modelo muito utilizado para se estudar efeitos coletivos como conseqüência de interações locais. É certamente o modelo Markoviano mais utilizado até hoje. Registros na literatura indicam que o modelo de Potts tem um papel fundamental em pesquisas das mais variadas áreas, como a matemática (82–85) a física (79, 86–87), a biologia (88–89), a computação (90–91) e até a sociologia (92).

Na utilização de modelos estocásticos em processamento de imagens e reconhecimento

de padrões muitas vezes é preciso se definir uma distribuição de probabilidade sobre o conjunto de possíveis imagens que reflita adequadamente o conhecimento a priori desejado. Um dos modelos mais utilizados para esse fim (modelar o conhecimento a priori na forma de restrição de suavidade) é o modelo Markoviano de Potts. Imagens em geral são suaves, no sentido de que seus pixels possuem uma alta correlação espacial, devido à presença de regiões homogêneas. Nesse trabalho será adotado um modelo de interação por pares (`PWI`), cuja idéia consiste em se considerar apenas as interações entre um pixel central e seus vizinhos.

De acordo com o Teorema de Hammersley-Clifford (7), o modelo de Potts pode ser equivalentemente definido de duas formas: por uma distribuição conjunta (*Gibbs*) ou pelas funções densidade condicionais locais (`LCDF`). Para os propósitos desse trabalho, a representação por modelos locais é mais adequada pois permite o cálculo direto das probabilidades para cada elemento do campo de maneira individual e não para toda uma ocorrência de um campo aleatório como seria o caso a partir da distribuição de Gibbs e da energia global do sistema. Considerando um sistema de vizinhança geral, define-se a função densidade condicional local de um modelo de Potts como:

$$p\left(x_{ij} = m | x_{\eta_{ij}}, \beta\right) = \frac{\exp\left\{\beta U_{ij}\left(m\right)\right\}}{\sum_{l=1}^{C} \exp\left\{\beta U_{ij}\left(l\right)\right\}} \tag{2.13}$$

onde η_{ij} representa a vizinhança do pixel x_{ij}, U_{ij} denota o número de pixels pertencentes a vizinhança cujo rótulo é igual a l, $l \in G = \{1, 2, \ldots, C\}$, com C denotando o número de rótulos (classes), β é um parâmetro de dependência espacial entre elementos vizinhos que corresponde ao inverso da temperatura ($\beta = 1/T$) e m é o valor observado para o pixel central x_{ij}.

Convém ressaltar que, quanto maior o valor de β, maior a dependência espacial entre os elementos do campo pois maior a probabilidade do elemento central pertencer a classe dominante. Note porém que, quando $\beta = 0$, ou seja, não há dependência espacial (pixels estatisticamente independentes), a função distribuição do modelo se degenera para a função massa de probabilidade de uma variável discreta com distribuição uniforme (todos os valores são equiprováveis) e a informação contextual é completamente desprezada. Um caso particular do modelo de Potts, quando $C = 2$, ou seja, quando as variáveis podem assumir apenas dois valores distintos, é o modelo de Ising (78), utilizado em mecânica estatística para modelar o comportamento de sistemas físicos de partículas através da interação entre cada elemento e seus vizinhos. Basicamente, nesse modelo o sistema pode apresentar dois tipos de comportamento bem definidos: um comportamento ferromagnético, situação na qual prevalece o alinhamento dos *spins* (baixas temperaturas e maior probabilidade para configurações de baixa energia), ou anti-ferromagnético,

quando prevalece o desalinhamento de *spins* (altas temperaturas).

O modelo de Ising foi proposto em 1925 como uma tentativa de se explicar certos comportamentos observados empiricamente em materiais ferromagnéticos. Convém ressaltar que a solução analítica para o modelo de Ising em um reticulado bidimensional, sem a presença de um campo magnético externo, foi proposta por (93). Porém, o modelo de Ising tridimensional ainda não tem solução analítica, sendo computacionalmente intratável (94), e permanece como um dos mais desafiadores problemas em aberto da atualidade.

Para fins de classificação contextual de imagens, uma das maiores dificuldades de se adotar esse modelo é a etapa de estimação de parâmetros. Por esse motivo, atualmente, na maioria das aplicações esses parâmetros ainda são ajustados manualmente, muitas vezes por técnicas de tentativa e erro (25–26). No capítulo 3 são apresentadas novas equações de pseudo-verossimilhança para a estimação do parâmetro β em sistemas de ordens superiores.

2.1.3 Modelagem Bayesiana para Classificação Contextual

Adotando a combinação de modelos Markovianos (`GMRF` + *Potts*), deve-se classificar um pixel (ou vetor de padrões) (i, j) de acordo com a seguinte regra de decisão:

$$x_{ij}^{(p+1)} = \arg\left[\max_m Q\left(x_{ij} = m | x_{S/(i,j)}^{(p)}, \mathbf{y}, \hat{\Phi}, \hat{\Theta}, \hat{\beta}\right)\right] \tag{2.14}$$

onde $\hat{\Phi}$, $\hat{\Theta}$ e $\hat{\beta}$ são os vetores com os parâmetros estimados, $\mathbf{y}$ é a imagem multiespctral observada, S é o reticulado que define o campo aleatório, x^p denota o campo de rótulos na p-ésima iteração e o funcional $Q\left(x_{ij} = m | x_{S/(i,j)}^{(p)}, \mathbf{y}, \hat{\Phi}, \hat{\Theta}, \hat{\beta}\right)$ é obtido calculando-se a probabilidade *a posteriori* assumindo um `GMRF` para verossimilhança (dados) e uma *priori* dada pelo modelo de Potts. A expressão para esse funcional é (81):

$$\begin{aligned} Q\left(x_{ij} = m | x_{S/(i,j)}^{(p)}, \mathbf{y}, \hat{\Phi}, \hat{\Theta}, \hat{\beta}\right) = & \\ -\frac{1}{2}\ln\left|\hat{\Sigma}_m\right| - \frac{1}{2}\left[\vec{y}_{ij} - \hat{\vec{\mu}}_m - \left(\hat{\theta}^T y_{n_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\vec{\mu}}_m\right)\right]^T & \\ \times\hat{\Sigma}_m^{-1}\left[\vec{y}_{ij} - \hat{\vec{\mu}}_m - \left(\hat{\theta}^T y_{n_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\vec{\mu}}_m\right)\right] + \hat{\beta}U_{ij}(m) & \end{aligned} \tag{2.15}$$

Observando a expressão acima, fica claro que a regra de decisão adotada na classificação considera tanto a informação espectral (note que primeiro termo do funcional vem do modelo GMRF) quanto o contexto em que o vetor de padrões está inserido, devido a modelagem espacial (segundo termo do funcional vem do modelo de Potts). Nesse cenário, o papel do parâmetro β do modelo de Potts é o mesmo de um parâmetro de regularização, pois define o compromisso entre as observações (verossimilhança) e o conhecimento *a priori*. Considere a situação em que o parâmetro β do modelo de Potts é proximo a zero. Nessas situações, a informação contextual é pouco informativa pois trata-se um campo de rótulos com aspecto ruidoso (a dependência espacial entre elementos vizinhos é praticamente inexistente). Nesse caso, o funcional acima nos diz que é recomendável desconsiderar a informação contextual e realizar a classificação com base na verossimilhança apenas. Por outro lado, na situação em que o campo de rótulos é suave, a informação contextual é informativa, sendo refletida num valor significativo para o parâmetro β, o que faz com que haja um balanço entre verossimilhança e *priori*, incorporando de maneira significativa o contexto espacial na classificação.

2.2 Algoritmos Iterativos de Otimização Combinatória

Como exposto anteriormente, o estimador `MAP` é obtido maximizando a probabilidade *a posteriori*. É conhecido que o único algoritmo computacional que garante a obtenção do estimador `MAP` propriamente dito é o método estocástico conhecido como *Simulated Annealing* (`SA`) (15), pois, sob condições específicas e bem definidas, pode-se mostrar que esse algoritmo converge para o máximo global (8). Maiores detalhes sobre esse algoritmo são discutidos na seção 2.2.2.

Porém, em problemas envolvendo análise de imagens, encontrar o estimador `MAP` requer a otimização de funções extremamente complexas, definidas em domínios de milhares de dimensões. Ademais, problemas modelados de acordo com a abordagem MAP-MRF não admitem soluções fechadas, sendo necessário a utilização de métodos numéricos. Por esse motivo, diversos algoritmos sub-ótimos de otimização combinatória foram propostos na literatura com o intuito de obter aproximações para o estimador `MAP` de maneiras computacionalmente viáveis. Basicamente, métodos determinísticos como os baseados na descida do gradiente, caminham estritamente no sentido de minimizar a função objetivo, o que geralmente fornece rápida convergência para um mínimo local. Por outro lado, métodos estocásticos permitem atualizações da função objetivo também no sentido oposto, fazendo com que o método escape de mínimos locais. Outra característica inerente aos

métodos sub-ótimos é a grande dependência às condições iniciais, ou seja, a inicialização tem grande influência na solução final obtida após um número finito de iterações.

Diversos algoritmos determinísticos de otimização combinatória podem ser utilizados para se obter aproximações para o estimador MAP. Basicamente, a diferença entre cada um dos algoritmos é a dinâmica de interação adotada na resolução do problema, ou seja, como o campo aleatório evolui através da interação dos elementos com seus vizinhos. Dentre os algoritmos mais conhecidos podemos citar o ICM (*Iterated Conditional Modes*) (9), o MPM (*Maximizer of the Posterior Marginals*) (10) e o GSA (*Game Strategy Approach*) (11), baseado na teoria dos jogos não cooperativos (95–96).

2.2.1 Métodos para Amostragem em Campos Aleatórios

Amostragem é o processo de se gerar ocorrências de um campo aleatório Markoviano, dado um modelo e os parâmetros especificados. Suponha que se deseja gerar amostras a partir de uma distribuição de Gibbs $P(X|Y)$. Esse não é um problema trivial, pois a distribuição de Gibbs é definida em termos da função de partição, uma constante de normalização computacionalmente intratável. Uma maneira de se resolver esse problema é através dos métodos de simulação de Monte Carlo, mais precisamente algoritmos MCMC (*Markov Chain Monte Carlo*). Basicamente, métodos MCMC são uma classe de algoritmos para amostragem baseados na definição de uma cadeia de Markov cuja distribuição no estado de equilíbrio (distribuição estacionária) é a distribuição desejada, ou seja, da qual se deseja obter as amostras. Um dos primeiros métodos propostos para esse fim foi o algoritmo *Metropolis*, que elimina a necessidade de se computar a função de partição (41). O Algoritmo 1 descreve o método *Metropolis* para amostragem de um MRF, conforme descrito em (16).

A razão definida no passo (7) do algoritmo (*Metropolis-Hastings*) pode ser computada na prática pois não depende da função de partição (apenas os pixels ao redor do pixel corrente precisam ser incluídos). Um segundo algoritmo, proposto em (8), é conhecido como *Gibbs Sampler*, ou Amostrador de Gibbs. Outras variações de métodos para simulação de Monte Carlo que apresentam rápida convergência ao estado de equilíbrio, sendo especialmente adequados para simulações utilizando temperaturas próximas a crítica, são os algoritmos de *cluster spin-flip* conhecidos como *Swendsen-Wang* (42) e *Wolff* (43). Porém, ambos são restritos a simulações envolvendo o modelo de Potts para um número arbitrário de estados. Os Algoritmos 2 e 3 descrevem os métodos *Gibbs Sampler* e *Swendsen-Wang* respectivamente.

Algoritmo 1: Metropolis-Hastings

Input : Um campo de rótulos $S_x \times S_y$, onde cada elemento é gerado aleatoriamente e independentemente dos demais; O número máximo de iterações N_{iter}.

Output: Uma ocorrência do modelo Markoviano definido pela distribuição de Gibbs $P(X)$.

Definir um modelo Markoviano para o campo de rótulos X (Modelo de Potts).
Inicializar um campo de rótulos $S_x \times S_y$ atribuindo um rótulo aleatório pertencente ao conjunto $G = \{1, 2, \ldots, C\}$ a cada pixel. Chamar essa configuração inicial de x.
while $n \leq N_{iter}$ **do**
  for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
      Escolher um rótulo $g \in G$ aleatoriamente e definir $y_{ij} = g$.
      Seja $p = min\left\{1, \frac{P(X=y)}{P(X=x)}\right\}$.
      Substituir x por y com probabilidade p.
    end
  end
end

Algoritmo 2: Gibbs Sampler

Input : Um campo de rótulos $S_x \times S_y$, onde cada elemento é gerado aleatoriamente e independentemente dos demais;
O número máximo de iterações N_{iter}.

Output: Uma ocorrência do modelo Markoviano definido pela distribuição de Gibbs $P(X)$.

Definir um modelo Markoviano para o campo de rótulos X (Modelo de Potts).
Inicializar um campo de rótulos $S_x \times S_y$ atribuindo um rótulo aleatório pertencente ao conjunto $G = \{1, 2, \ldots, C\}$ a cada pixel. Chamar essa configuração inicial de x.
while $n \leq N_{iter}$ **do**
  for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
      Computar o conjunto de probabilidades $\{p_g\}$, $\forall g \in G$,onde $p_g = p\left(x_{ij} = g | x_{\eta_{ij}}\right)$ com $x_{\eta_{ij}}$ denotando a vizinhança do pixel x_{ij}.
      Atribuir o rótulo g ao elemento x_{ij} com probabilidade p_g.
    end
  end
end

2.2.2 Simulated Annealing

Simulated Annealing (SA) é um método não-determinístico utilizado na otimização de funções de muitas variáveis, que tenta evitar mínimos locais, através da introdução de pequenas perturbações na função objetivo (16), e que foi popularizado na comunidade científica após o trabalho de (15). Apesar de não fazer assumir nenhuma restrição acerca da suavidade das funções, este método impõe severos requisitos computacionais. Trata-se de um algoritmo pertencente a classe dos métodos de relaxação estocástica baseados no algoritmo Metropolis para simulação de sistemas contendo uma grande quantidade de partículas arrajandas num reticulado, em geral bidimensional. No contexto de processamento de imagens, tornou-se popular após o trabalho de (8), onde o algoritmo SA foi aplicado na segmentação de imagens.

No algoritmo SA, cada ponto no espaço de estados é analogo a um estado de um sistema físico e a função a ser minimizada é análoga a energia interna desse sistema. O objetivo desse algoritmo é trazer o sistema, de um estado inicial arbitrário, para o estado de mínima energia. A cada iteração, o método considera algum dos estados vizinhos do estado atual e, de maneira estocástica, decide entre se mover para o próximo estado ou permanecer no estado atual. As probabilidades de transição são escolhidas de modo que o sistema tende a convergir para estados de baixa energia e a temperatura do sistema é lentamente diminuída a cada iteração.

A escolha da função que define o resfriamento do sistema, a temperatura inicial, o número de iterações máximo e o esquema de perturbação do sistema são todos aspectos fundamentais e devem ser escolhidos de acordo com a aplicação. Em (8) é sugerido a seguinte função de resfriamento:

$$T_{k+1} = \frac{ln(1+k)}{ln(2+k)} T_k \tag{2.16}$$

Ainda, em aplicações de processamento de imagens, valores iniciais típicos para T pertencem ao intervalo $[2, 4]$ e o esquema de perturbação involve duas escolhas aleatórias: tanto a do próximo elemento a ser visitado, em geral um dos vizinhos, quanto o rótulo a ser atribuído ao elemento em questão. O Algoritmo 4 descreve o método SA.

Algoritmo 3: Swendsen-Wang

Input : Um campo de rótulos $S_x \times S_y$, onde cada elemento é gerado aleatoriamente e independentemente dos demais;
O número máximo de iterações N_{iter}.
Output: Uma ocorrência do modelo de Potts.

Inicializar um campo de rótulos $S_x \times S_y$ atribuindo um rótulo aleatório pertencente ao conjunto $G = \{1, 2, \ldots, C\}$ a cada pixel. Chamar essa configuração inicial de x.
while $n \leq N_{iter}$ **do**
  for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
      Criar uma ligação entre o elemento x_{ij} e cada um de seus vizinhos que possuem o mesmo rótulo com probabilidade $p = 1 - e^{-\beta}$.
    end
  end
  Identificar todos os $N_{cluster}$ clusters formados.
  for $k \leftarrow 1$ **to** $N_{cluster}$ **do**
    Escolher um rótulo $g \in G = \{1, 2, \ldots, C\}$ aleatoriamente.
    Atribuir o novo rótulo g a todos os elementos do k-ésimo cluster.
  end
end

Algoritmo 4: Simulated Annealing

Input : Um campo de rótulos $S_x \times S_y$, onde cada elemento é gerado aleatoriamente e independentemente dos demais;
O número máximo de iterações N_{iter};
A temperatura inicial T.
Output: O estimador MAP do mapa de classes (solução ótima).

Definir um modelo Markoviano para o campo de rótulos X (Modelo de Potts).
Inicializar um campo de rótulos $S_x \times S_y$ atribuindo um rótulo aleatório pertencente ao conjunto $G = \{1, 2, \ldots, C\}$ a cada pixel. Chamar essa configuração inicial de x.
while $T > 0$ **do**
  while $n \leq N_{iter}$ **do**
    Perturbar x em z.
    Seja $\Delta = U(z) - U(x)$.
    if $\Delta > 0$ **then**
      Substituir x por z.
    else
      Substituir x por z com probabilidade $e^{\frac{\Delta}{T}}$.
    end
  end
  Resfriar o sistema de através de uma função monotonicamente decrescente como por exemplo a dada pela equação (2.16)
end

2.2.3 ICM

O algoritmo ICM foi proposto por (9) como uma alternativa computacionalmente viável na busca de se superar as dificuldades na obtenção do estimador MAP, que em outras palavras nada mais é que a moda da distribuição *a posteriori*. O algoritmo ICM é iterativo e determinístico. Basicamente, a idéia consiste em, para cada pixel visitado, atualizar seu valor através do rótulo com maior probabilidade *a posteriori*. Observando que $P(x|y) = P(x_s|x_s^-, y)\, P(x_s^-|y)$, onde x_s^- denota todo o campo aleatório com exceção do pixel corrente x_s, a subseqüente maximização de $P(x_s|x_s^-, y)$ para cada pixel caminha sempre na direção de crescimento da função $P(x|y)$. Assim, o algoritmo ICM converge rapidamente para um máximo local, pois os resultados são dependentes da inicialização. Segundo (67), o ICM é equivalente ao SA com congelamento instantâneo.

A grande diferença entre ICM e SA é que o algoritmo ICM sempre escolhe como próximo rótulo o valor x_s^* que maximiza a probabilidade condicional dentre todos os possíveis valores de x_s, gerando uma regra determinística. Por outro lado, a atualização dos rótulos no SA é realizada através do cálculo de probabilidades de transição, gerando uma regra estocástica que permite ao método escapar de eventuais máximos locais, por vezes caminhando no sentido contrário ao crescimento da função. O Algoritmo 5 descreve o método ICM, conforme descrito em (16).

Algoritmo 5: Iterated Conditional Modes

Input : Um campo de rótulos inicial de dimensões $S_x \times S_y$, obtido por um classificador pontual;
O número máximo de iterações N_{iter}.

Output: Uma aproximação para o estimador MAP do mapa de classes (classificação contextual).

Definir um modelo Markoviano para o campo de rótulos X (Modelo de Potts).
Inicializar $\hat{x}$ escolhendo o rótulo $\hat{x}_{ij}$ que maximiza $P(Y_{ij} = y_{ij}|X_{ij} = x_{ij})$ para cada pixel (Classificação por Máxima Verossimilhança).*
while $n \leq N_{iter}$ **do**
  for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
      Atualizar o rótulo $\hat{x}_{ij}$ com o valor x_{ij} que maximiza:
      $P\left(X_{ij} = x_{ij}|x_{ij}^-, y\right) \propto P(Y_{ij} = y_{ij}|X_{ij} = x_{ij})\, P\left(X_{ij} = x_{ij}|x_{\eta_{ij}}\right)$
    end
  end
end

Note que devido aos interesses desse projeto, o passo (2) do algoritmo ICM é generalizado, utilizando diferentes estratégias de inicialização. Resultados obtidos na literatura (9, 16, 25, 97) demonstram que poucas iterações são necessárias para a convergência do algoritmo ICM (tipicamente 5 iterações são suficientes).

2.2.4 GSA

Esse método, que também pode ser utilizado para aproximar o estimador MAP, é baseado na teoria dos jogos não-cooperativos (11). Em um jogo de n-jogadores (*n-person game*), $I = \{1, 2, \ldots, n\}$ denota o conjunto de todos os jogadores. Cada jogador i tem um conjunto de estratégias puras (*pure strategies*) S_i. O processo do jogo consiste em um determinado instante cada um dos jogadores escolher uma determinada estratégia $s_i \in S_i$. Assim, uma situação (ou jogada) $s = (s_1, s_2, \ldots, s_n)$ é obtida e a cada jogador é associado um ganho (*payoff*) $H_i(s)$. Na abordagem proposta em (11), define-se $H_i(s)$ de tal forma que o ganho de cada jogador depende apenas de sua própria estratégia e do conjunto de estratégias de seus vizinhos (note a semelhança com um MRF). Na abordagem não-cooperativa, cada jogador deve tentar maximizar o seu ganho escolhendo sua própria estratégia independentemente. Em resumo, trata-se do problema de otimizar o ganho global, a partir de decisões independentes e locais (note a semelhança com o conceito de independência condicional, que caracteriza os MRF's). Uma estratégia mista (*mixed strategy*) para um jogador é definida como uma distribuição de probabilidade definida no conjunto das estratégias puras. Assume-se a hipótese de que cada jogador conhece todas as estratégias e o ganho obtido sob cada possível situação.

As soluções para um jogo de n-jogadores não-cooperativo são dadas pelo conjunto de pontos que satisfazem o equilíbrio de Nash (*Nash points*) (95). Pode ser mostrado que o equilíbrio de Nash sempre existe em jogos não-cooperativos de n-jogadores (95). Formalmente, uma jogada ou solução $t_* = (t_1^*, t_2^*, \ldots, t_n^*)$ satisfaz a condição de equilíbrio de Nash se nenhum dos jogadores pode melhorar seu ganho esperado alterando sua estratégia unilateralmente, ou, em termos matemáticos:

$$\forall i : H_i(t^*) = max_{s_i \in S_i} H_i(t^* || t) \tag{2.17}$$

onde $t^*||t$ representa a jogada obtida substituindo a configuração t^* por t. A conexão entre a teoria dos jogos e os campos aleatórios é demonstrada em (11), onde se prova que o conjunto de pontos que satisfazem o Equilíbrio de Nash em um jogo não-cooperativo de n-jogadores é de fato idêntico ao conjunto de máximos locais da probabilidade *a posteriori*

em um problema de classificação contextual de imagens. Baseado nesse e outros resultados, é proposto um algoritmo iterativo que converge para o Equilíbrio de Nash, ou seja, para um máximo local, dada uma situação inicial. Em outras palavras, dada uma jogada inicial, após sucessivas iterações do algoritmo `GSA`, é possível se atingir uma situação que satisfaz o equilíbrio de Nash. Os fundamentos do algoritmo `GSA` estão baseados em dois resultados (11):

Teorema 1. *O conjunto dos pontos de máximos locais da probabilidade a posteriori em um problema de classificação contextual por MRF's é idêntico ao conjunto dos pontos que satisfazem o Equilíbrio de Nash (Nash Points) do correspondente jogo de n-jogadores não cooperativo*

Teorema 2. *O algoritmo de otimização combinatória `GSA` converge para o equilíbrio de Nash quando o número de iterações cresce, independentemente da inicialização.*

Para melhor ilustrar a relação entre a abordagem as duas abordagens, a Tabela 1 mostra a equivalência entre alguns conceitos da teoria dos jogos e do problema de classificação através de campos aleatórios.

Tabela 1 - Equivalência entre conceitos de Teoria dos Jogos e o problema de classificação contextual modelado por Campos Aleatórios Markovianos

Classificação contextual via MRF's	**Teoria dos Jogos**
reticulado/grade	jogo de n-jogadores
pixels	jogadores
rótulos	estratégias puras
campo de rótulos	jogada/situação
função de energia	ganho
funções densidades condicionais locais	estratégias mistas
pontos de máximos locais (MAP)	Equilíbrio de Nash

Sejam $x^{(k)} = \left(x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}\right)$ o conjunto de rótulos (situação) na k-ésima iteração, K o número máximo de iterações, $\alpha \in (0,1)$ um número real representando a probabilidade de aceitação de um novo rótulo (ou estratégia) e $G = \{1, 2, \ldots, C\}$ o conjunto de possíveis rótulos (estratégias puras). O Algoritmo 6 descreve em detalhes o método `GSA`, conforme descrito em (97).

Basicamente, o algoritmo funciona da seguinte maneira: o método parte de uma inicialização (situação) qualquer e a cada iteração, para cada jogador (pixel), seleciona a estratégia (rótulo) mais provável. Calcula-se o ganho local do jogador tanto para a estratégia original quanto para a nova estratégia. Se o novo ganho obtido for menor que o anterior, a estratégia do jogador não deve ser atualizada (pois no Equilíbrio de Nash nenhum dos jogadores pode melhorar seu ganho alterando sua estratégia unilateralmente).

Algoritmo 6: Game Strategy Approach

Input : Um campo de rótulos inicial $L^{(0)}$ de dimensões $S_x \times S_y$;
O número máximo de iterações N_{iter};
A probabilidade de aceitação de novas estratégias α.

Output: Uma aproximação para o estimador MAP do mapa de classes (classificação contextual).

Inicializar $L^{(0)} = \left(l_1^{(0)}, l_2^{(0)}, \ldots, l_{S_x S_y}^{(0)}\right)$ *utilizando um classificador pontual e definir* $n = 0$.
while $n \leq N_{iter}$ **do**
  for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
      Escolher um rótulo (ou estratégia) differente da atual $l'_{ij} \neq l_{ij}^{(n)}$ *que maximiza o ganho local, ou seja:*
        $H_{ij}\left(l_{ij}^{(n)} \,||\, l'_{ij}\right) = max_{l_{ij}}\left[H_{ij}\left(l_{ij}^{(n)} \,||\, l_{ij}\right)\right]$
      if $H_{ij}\left(l_{ij}^{(n)} \,||\, l'_{ij}\right) \leq H_{ij}\left(l_{ij}^{(n)}\right)$ **then** $l_{ij}^{(n+1)} = l_{ij}^{(n)}$;
      else
        Aceitar l'_{ij} *com probabilidade* α, *i.e.,*
          $l_{ij}^{(n+1)} = l'_{ij}$ *com probabilidade* α *ou,*
          $l_{ij}^{(n+1)} = l_{ij}^{(n)}$ *com probabilidade* $1 - \alpha$.
      end
      $L^{(n+1)} = \left(l_1^{(n+1)}, l_2^{(n+1)}, \ldots, l_{S_x S_y}^{(n+1)}\right)$
      if $L^{(n+1)}$ *satisfaz o Equilíbio de Nash* **then** *break;*
      else
        n = n + 1;
      end
    end
  end
end

Caso contrário, o jogador em questão tem probabilidade α de alterar sua estratégia e probabilidade $1 - \alpha$ de manter a estratégia.

Conforme examinado, o esquema de atualização dos rótulos depende do parâmetro α. Para valores de $0 < \alpha < 1$, tem-se um esquema não-determinístico, que garante a convergência de $L^{(k)}$ para o Equilíbrio de Nash conforme o número de iterações aumenta. Para $\alpha = 1$, o algoritmo pode não convergir, dependendo da inicialização. Entretanto, uma versão determinística do algoritmo ($\alpha = 1$) pode ser gerada excluindo-se a atualização simultânea de elementos vizinhos. Por exemplo, os rótulos podem ser atualizados seqüencialmente. Nesse caso, o esquema obtido é similar aos algoritmos `ICM` (9) ou o `HCF` (14), dependendo da maneira como os rótulos são substituídos. Vale notar que o caráter não-determinístico presente no algoritmo `GSA` é diferente do existente no `SA`. No `GSA`, apenas a aceitação ou não dos novos candidatos é aleatória, enquanto no `SA` a seleção de candidatos também segue um padrão não-determinístico. Além disso, o algoritmo `SA` introduz perturbações no campo de rótulos e modifica a temperatura do sistema através de uma função de resfriamento com o intuito de escapar de mínimos locais.

2.2.5 MPM

Conforme discutido nas seções anteriores, a diferença básica entre o estimador `MAP` e o estimador `MPM` é a função de perda adotada (função de custo uniforme versus função de custo distância de Hamming). O ponto de partida para o algoritmo `MPM` é se calcular as distribuições marginais condicionais. A abordagem definida em (10) propõe utilizar métodos `MCMC`, como os algoritmos Gibbs Sampler (amostrador de Gibbs) e Metropolis para a simulação de campos aleatórios na obtenção das probabilidades marginais. Em resumo, esses algoritmos simulam uma cadeia de Markov sobre os $C^{N \times N}$ estados que representam todas as possíveis configurações do campo aleatório. A idéia é que cada pixel seja visitado e seu rótulo seja atualizado ao longo de diversas iterações. Na verdade, quando o número de visitas para cada pixel se aproxima de infinito, a cadeia de Markov resultante gera uma amostra (campo de rótulos) da distribuição *a posteriori* $P(x|y)$, independente da configuração inicial $x(0)$ (8). Como resultado, é gerada uma seqüência de configurações $x(0) \rightarrow x(1) \rightarrow \cdots \rightarrow x(n) \rightarrow \cdots$ que corresponde a uma cadeia de Markov que atinge o estado de equilíbrio. Uma vez que a cadeia de Markov atinge o equilíbrio, pode-se considerar que todas as configurações obtidas a partir desse momento são amostras da distribuição a posteriori $P(x|y)$. Além disso, no estado de equilíbrio, o valor esperado de uma função $g(.)$ de uma variável aleatória X pode ser calculado

assumindo-se a hipótese de ergodicidade (98), que diz que a média no *ensemble* converge para a média populacional. Por exemplo, a distribuição marginal *a posteriori* para o estimador `MPM` pode ser aproximada pela seguinte média amostral (16):

$$P\left(x_s = g|y\right) \approx \frac{1}{(n-k)} \sum_{i=k+1}^{n} \delta\left(x_s^i - g\right) \tag{2.18}$$

onde g é o valor observado no i-ésimo passo da simulação, δ é a função que retorna o valor unitário quando o argumento é nulo, k é o número de iterações necessárias para a seqüência se estabilizar (*burn-in*) e n é um valor razoavelmente grande para que a estimação seja precisa, a um custo computacional razoável. Em outras palavras, a equação acima apenas conta o número de vezes que cada rótulo aparece nos n passos da simulação para estimar as probabilidades marginais como sendo a freqüência relativa de cada um dos possíveis valores $g \in G$. Um problema desse algoritmo é que ambos k e n (conhecidos como *magic numbers*) são escolhidos empiricamente e podem depender tanto da aplicação quanto das dimensões das imagens em questão. É comprovado que o algoritmo `MPM` é computacionalmente mais caro que o `ICM`, mas mesmo assim, ainda é consideravelmente menos custoso que o método `SA`. Resultados encontrados na literatura comparando o custo computacional dos diversos métodos de otimização combinatória mostram significativas diferenças no tempo de computação gasto em cada um deles (16). O Algoritmo 7 descreve o método `MPM` para a classificação de imagens.

2.2.6 Considerações Finais

Este capítulo apresentou a classificação contextual de imagens multiespectrais como um problema de estimação Bayesiana, através de sua fundamentação matemática. Foram discutidos como diferentes funções de custo geram diferentes estimadores, bem como algoritmos computacionais que aproximam esses estimadores e modelos Markovianos para classificação. O capítulo seguinte apresenta novas técnicas para inferência estatística e estimação de parâmetros nos modelos Markovianos adotados, duas etapas importantes que compõem a metodologia proposta.

Algoritmo 7: Maximizer of the Posterior Marginals

Input : Um campo de rótulos inicial de dimensões $S_x \times S_y$ dado por um classificador pontual;
O número de iterações de "burn-in" K;
O número de iterações da simulação MCMC N_{iter}.

Output: Uma aproximação para o estimador MAP do mapa de classes (classificação contextual).

Definir um modelo Markoviano para o campo de rótulos X (Modelo de Potts).
Inicializar $\hat{x}$ escolhendo o rótulo $\hat{x}_{ij}$ que maximiza $P(Y_{ij} = y_{ij}|X_{ij} = x_{ij})$ para cada pixel (Classificação por Máxima Verossimilhança).*
while $n \leq N$ **do**
    for $i \leftarrow 1$ **to** S_x **do**
        for $j \leftarrow 1$ **to** S_y **do**
            Seja $p = min\left\{1, \frac{P(X=z|Y=y)}{P(X=x|Y=y)}\right\}$
            Substituir x por z com probabilidade p
        end
    end
    if $n > K$ **then**
        Armazena a configuração atual se a iteração está fora da janela de "burn-in", ou seja
        $w^{(n-k)} \leftarrow x^{(n)}$.
    end
end
Calcular $P(x_{ij} = g|y)$, $g \in G$, para todo elemento do campo, como
    $P(x_{ij} = g|y) \approx \frac{1}{N_{iter}-K} \sum_{l=1}^{N_{iter}-K} \delta\left(w_{ij}^{(l)} - g\right)$
for $i \leftarrow 1$ **to** S_x **do**
    for $j \leftarrow 1$ **to** S_y **do**
        Escolher o rótulo $\hat{x}_{ij}$ entre todos os valores $g \in G$ que maximiza a probabilidade a posteriori marginal
        $P(X_{ij} = \hat{x}_{ij}|Y = y) > P(X_{ij} = g|Y = y)$
    end
end

Capítulo 3

Inferência Estatística em Modelos Markovianos

Tradicionalmente, dois ramos independentes de pesquisa desenvolveram extensões de processos Markovianos 1-D para dados espaciais. A primeira abordagem, que é a utilizada nesse projeto, adota o paradigma e todo o ferramental existente na Mecânica Estatística, expressando a natureza Markoviana de um campo aleatório de maneira não-causal. Já a segunda abordagem, estende os modelos ocultos de Markov (`HMM`) para modelos 2-D causais. Porém, a grande dificuldade com esses métodos é justamente a falta de ordenação natural existente em dados espaciais, de forma que esquemas artificiais de ordenação devem ser assumidos.

A noção fundamental associada à propriedade Markoviana é a independência condicional, visto que o conhecimento de uma região local torna um elemento estatisticamente independente do restante do campo aleatório. Em outras palavras, um Campo Aleatório Markoviano é uma coleção de variáveis aleatórias para as quais a probabilidade de ocorrência de um elemento dado todo o conhecimento de todo o restante do campo aleatório é igual a probabilidade de ocorrência desse elemento dado uma região de suporte finita ao seu redor, conhecida como vizinhança (99), ou em termos matemáticos:

$$p(x_{ij}|\{x_{kl},(k,l)\in S\backslash(i,j)\}) = p(x_{ij}|\{x_{kl},(k,l)\in \eta_{ij}\}) \tag{3.1}$$

onde x_{ij} denota um elemento do campo aleatório, x_{kl} denota um elemento qualquer, $S\backslash(i,j)$ representa todo o campo aleatório com exceção do elemento (i,j) e η_{ij} denote a vizinhança do elemento (i,j).

Diante do exposto, pode-se perceber que tanto a forma (causal ou não-causal) quanto a extensão (ordem) de um sistema de vizinhança são as duas caraterísticas primordiais na definição de qualquer campo aleatório Markoviano.

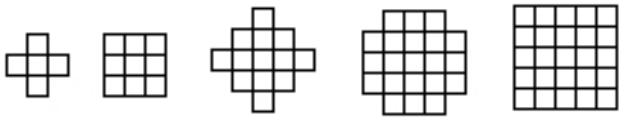


Figura 5 - Sistemas de vizinhanças não-causais de primeira a quinta ordem.

Um sistema de vizinhança é dito ser causal se todos os elementos da região de suporte pertencem a metade "anterior" do plano (assimétrico), ou seja, o campo pode ser rearranjado em um vetor aleatório 1-D (notação lexicográfica) que satisfaça a propriedade Markoviana (100). Já no caso de regiões de suporte simétricas em relação ao elemento central, tem-se um sistema de vizinhança não-causal. A ordem de um sistema de vizinhança pode ser definida como sendo o tamanho da região de suporte que define a janela de observações ao redor de um dado elemento. Sistemas não-causais podem ser classificados de acordo com sua extensão como sendo de primeira ordem (quatro vizinhos), segunda ordem (oito vizinhos), terceira ordem (doze vizinhos) e assim sucessivamente. A Figura 5 ilustra vários sistemas de vizinhanças não-causais de primeira a quinta ordem, respectivamente.

3.1 Estimação por Máxima Pseudo-Verossimilhança

Uma das maiores dificuldades na utilização de Campos Aleatórios Markovianos em aplicações de processamento de imagens é justamente a etapa de estimação dos parâmetros. Isso se deve ao fato de que o principal método de estimação, conhecido como Máxima Verossimilhança, não pode ser aplicado devido à presença da função de partição na distribuição de Gibbs conjunta. Sabe-se que tal função é computacionalmente intratável (16), o que inviabiliza a definição da função de verossimilhança.

Uma solução alternativa, proposta por (9) consiste na utilização das funções densidades condicionais locais (`LCDF`) na estimação por Máxima Pseudo Verossimilhança (`MPV`). Basicamente, as principais motivações para o uso dessa abordagem são:

- É um método computacionalmente viável
- Do ponto de vista estatístico, estimadores de `MPV` possuem uma série de propriedades desejáveis, como por exemplo, consistência e normalidade assintótica (45, 101). Assim, é possível se caracterizar completamente o seu comportamento no caso limite.

Entretanto, uma das maiores limitações dessa abordagem tem sido a utilização de sistemas de vizinhanças extremamente restritos, muitas vezes insuficientes para uma modelagem contextual adequada. Observa-se que na maioria das aplicações, os sistemas de vizinhanças adotados são de primeira ordem. Outro aspecto pouco investigado na literatura de MRF's é análise da acurácia da estimação dos parâmetros. Apesar de amplamente difundida, pouco se conhece acerca da precisão que se tem na estimação por máxima pseudo-verossimilhança. Um dos objetivos desse trabalho é justamente propor novas equações de pseudo-verossimilhança em sistemas de ordens superiores, verificando e comparando a acurácia da estimação através da derivação de aproximações para a variância assintótica dos estimadores de máxima pseudo-verossimilhança.

O objetivo desse capítulo é apresentar novas equações de pseudo-verossimilhança para o modelo de Potts para estimação de parâmetros em sistemas de segunda e terceira ordens (46–47), bem como verificar a acurácia da estimação nos modelos `GMRF` e de Potts através da derivação de aproximações para a variância assintótica dos estimadores de máxima pseudo-verossimilhança e de simulação computacional por métodos de Monte Carlo (102–103).

3.1.1 GMRF

Tendo em vista o modelo `GMRF` multiespectral adotado para a classificação contextual, é possível observar que não há dependência entre os parâmetros espaciais pertencentes a diferentes bandas espectrais. Portanto, é razoável admitir que a estimação dos parâmetros de cada banda seja realizada independentemente. Assumindo esta hipótese (independência entre os parâmetros de diferentes bandas) e considerando um sistema de vizinhança de segunda ordem, a função de máxima pseudo-verossimilhança é dada por:

$$\begin{aligned} \log PL\left(\theta,\mu,\sigma^2\right) &= \log \prod_{(i,j)\in S} p\left(y_{ij}|y_{\eta_{ij}},\theta,\mu,\sigma^2\right) = \\ &= \sum_{(i,j)\in S}\left\{-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left[y_{ij} - \vec{\theta^T}\vec{\psi}_{ij} - \mu\left(1-2\vec{\theta^T}I\right)\right]^2\right\} \end{aligned} \tag{3.2}$$

onde y_{ij} representa um elemento do campo, μ é a média, σ^2 é a variância, $\vec{\theta^T} = [\theta_1,\theta_2,\theta_3,\theta_4]$ é o vetor de parâmetros de dependência espacial do modelo `GMRF`, $\vec{\psi}_{ij} = [(y_{i+1j}+y_{i-1j}),(y_{ij+1}+y_{ij-1}),(y_{i+1j-1}+y_{i-1j+1}),(y_{i+1j+1}+y_{i-1j-1})]^T$ e I denota uma vetor identidade 4 x 1. Nessa notação, θ_1 representa o parâmetro de dependência espacial na direção horizontal, θ_2 é o

parâmetro para a direção vertical, θ_3 e θ_4 para as direções diagonais.

Alguns comentários acerca da equação merecem destaque. Convém notar que, como a estimação é realizada em cada banda separadamente, cada amostra observada y_{ij} (elemento do campo aleatório) é um simples escalar pertencente ao intervalo $[0, 255]$ visto que se trata de uma imagem em tons de cinza. Outro detalhe diz respeito a estimação dos parâmetros μ e σ^2. Ambos são estimados pelas simples médias e variâncias amostrais.

Assim como a distribuição gaussiana, o modelo `GMRF` possui algumas características peculiares. Uma delas é que dentre os modelos Markovianos conhecidos ele é o único que admite uma solução fechada para a equação de pseudo-verossimilhança. Assim, pode-se calcular esse estimador diretamente a partir da seguinte expressão (67):

$$\hat{\vec{\theta}} = \left\{ \left[\sum_{(i,j)\in S} (y_{ij} - \hat{\mu})\, \tilde{\vec{\psi}}_{ij}^{T} \right] \left[\sum_{(i,j)\in S} \tilde{\vec{\psi}}_{ij} \tilde{\vec{\psi}}_{ij}^{T} \right]^{-1} \right\} \tag{3.3}$$

onde $\hat{\mu}$ é a média amostral dos elementos do campo, $\tilde{\vec{\psi}}_{ij}$ é a versão centralizada do vetor $\vec{\psi}_{ij}$, ou seja, $\tilde{\vec{\psi}}_{ij} = \vec{\psi}_{ij} - \frac{1}{N} \sum_{(k,l)\in S} \vec{\psi}_{ij}$ (subtraído do vetor médio) e N é o número de elementos do campo.

3.1.2 Modelo de Potts

A introdução de sistemas de vizinhança de ordens superiores requer a definição de novos métodos para a estimação dos parâmetros dos modelos Markovianos. O objetivo dessa seção é justamente propor novas equações de pseudo-verossimilhança para a estimação do parâmetro do modelo de Potts em sistemas de segunda e terceira ordens, inspirado na metodologia inicialmente desenvolvida em (104) para sistemas de primeira ordem. A função de pseudo-verossimilhança para o modelo de Potts definido num sistema de vizinhança de ordem k é definida como:

$$PL(\beta) = \prod_{s\in S} p\left(x_s = m \mid \eta_s^k\right) = \prod_{s\in S} \frac{\exp\left\{\beta U_s\left(m\right)\right\}}{\sum_{\ell=1}^{M} \exp\left\{\beta U_s\left(\ell\right)\right\}}. \tag{3.4}$$

Aplicando o logaritmo, diferenciando em relação ao parâmetro e igualando o resultado a zero chega-se à seguinte expressão, que é a base para a derivação das equações propostas:

$$\frac{\partial}{\partial\beta} \log PL(\beta) = \sum_{s\in\Omega} U_s\left(m_s\right) - \sum_{s\in\Omega} \left[\frac{\sum_{\ell=1}^{M} U_s\left(\ell\right) \exp\left\{\beta U_s\left(\ell\right)\right\}}{\sum_{\ell=1}^{M} \exp\left\{\beta U_s\left(\ell\right)\right\}} \right] = 0, \tag{3.5}$$

Observando a expressão 3.5 é possível notar que o primeiro termo independe do parâmetro, definindo uma constante. O segundo termo, porém, pode ser expandido mapeando-se todos os possíveis padrões de configuração espacial envolvendo o sistema de vizinhança adotado. Por exemplo, considerando um sistema de vizinhança de primeira ordem, a enumeração de todos os possíveis padrões de configuração é bastante simples e direta. A análise da equação 3.5 revela que existem apenas 5 padrões espaciais que oferecem contribuições distintas a função de pseudo-verossimilhança em sistemas de primeira ordem. Esses padrões variam desde a condição definida como concordância-nula (todos os elementos da vizinhança são distintos), até a condição de concordância-total (todos os elementos pertencentes a vizinhança são iguais), como ilustra a Figura 6.

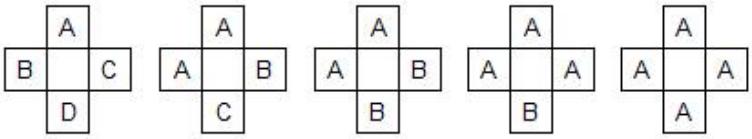


Figura 6 - Padrões de configuração espacial para um modelo de Potts de primeira ordem: cada padrão oferece uma contribuição distinta para a função de pseudo-verossimilhança.

Esses padrões de configuração espaciais podem ser representados por um conjunto de vetores N-D, onde N é o número de vizinhos presente no sistema de vizinhança. A idéia é que cada vetor indique a distribuição dos rótulos ao redor do elemento central. A equação 3.6 mostra a representação vetorial para os padrões de configuração espacial dados pela Figura 6.

$$\vec{v}_0 = [1,1,1,1] \qquad \vec{v}_1 = [2,1,1,0] \qquad \vec{v}_2 = [2,2,0,0] \qquad (3.6)$$
$$\vec{v}_3 = [3,1,0,0] \qquad \vec{v}_4 = [4,0,0,0]$$

3.1.2.1 Geração dos padrões de configurações espaciais

Seja π o número de elementos de um sistema de vizinhança qualquer (i.e., $\pi = 4, 8, 12, ...$). Para cada inteiro $L = 1, 2, ..., \pi$, seja ainda

$$A_\pi(L) = \left\{ (a_1, \ldots, a_L) / a_i \in \{1, 2, \ldots, \pi\}, a_1 \leq a_2 \leq \cdots \leq a_L, \sum_{i=1}^{L} a_i = \pi \right\} \qquad (3.7)$$

e $n_\pi(L)$ a cardinalidade do conjunto $A_\pi(L)$. Então, o número de possíveis padrões de configuração espacial (λ) para um sistema de vizinhança pode calculado por $\lambda = n_\pi(1) + \cdots + n_\pi(L)$. Dessa forma, o problema consiste basicamente em encontrar a solução do sistema de equações dado por:

$$\sum_{i=1}^{L} a_i = \pi, \qquad L = 1, 2, \ldots, N \tag{3.8}$$

onde a_i é um inteiro pertencente ao intervalo $[0, \pi]$. Os vetores solução são encontrados através de uma busca pelo espaço de soluções. No entanto, para reduzir o custo computacional devido à dimensionalidade do problema, foi adotada uma heurística para restringir a busca ao primeiro quadrante do espaço, tendo em vista que vetores simétricos representam o mesmo padrão de configuração espacial (i.e., $[7,1,0,0,0,0,0,0] \equiv [1,7,0,0,0,0,0,0]$). A Tabela 2 apresenta o número de padrões de configuração espacial (λ) e o tempo gasto para a geração dos vetores solução utilizando um método de busca exaustiva (força bruta) para diversos sistemas de vizinhança, mais precisamente de primeira a quinta ordens. A Tabela 3 ilustra o conjunto de vetores solução para um sistema de vizinhança de segunda ordem. Vale ressaltar que o procedimento de mapear todos os possíveis padrões de configuração espacial é realizado apenas uma única vez, por isso a escolha do ineficiente método de busca exaustiva. Todos os experimentos foram executados num computador com processador Athlon X2 Dual Core 2.21Ghz contendo 2GB de memória RAM.

Tabela 2 - Número de padrões de configuração espacial para diversos sistemas de vizinhança de ordens superiores.

Sistema de vizinhança	Número de padrões (λ)	Tempo (s)
1ª Ordem	5	0,1
2ª Ordem	22	0,2
3ª Ordem	77	4,5
4ª Ordem	637	30,0
5ª Ordem	1575	1517,0

Tabela 3 - Vetores soluções representando os possíveis padrões de configuração espacial em sistemas de segunda ordem.

$[1,1,1,1,1,1,1,1]$	$[2,1,1,1,1,1,1,0]$	$[3,1,1,1,1,1,0,0]$	$[2,2,1,1,1,1,0,0]$
$[4,1,1,1,1,0,0,0]$	$[3,2,1,1,1,0,0,0]$	$[2,2,2,1,1,0,0,0]$	$[5,1,1,1,0,0,0,0]$
$[4,2,1,1,0,0,0,0]$	$[3,3,1,1,0,0,0,0]$	$[3,2,2,1,0,0,0,0]$	$[2,2,2,2,0,0,0,0]$
$[6,1,1,0,0,0,0,0]$	$[5,2,1,0,0,0,0,0]$	$[4,3,1,0,0,0,0,0]$	$[4,2,2,0,0,0,0,0]$
$[3,3,2,0,0,0,0,0]$	$[4,4,0,0,0,0,0,0]$	$[5,3,0,0,0,0,0,0]$	$[6,2,0,0,0,0,0,0]$
	$[7,1,0,0,0,0,0,0]$	$[8,0,0,0,0,0,0,0]$	

3.1.2.2 Expansão da função de pseudo-verossimilhança

Dado o conjunto completo de todos os possíveis padrões de configuração espaciais para um sistema de vizinhança, pode-se expandir o segundo termo da equação de pseudo-verosimilhança. Encarando o numerador de 3.5 como um simples produto interno entre dois vetores $\vec{U_s}$ e $\vec{w_s}$, onde $\vec{U_s}$ representa o vetor de configuração espacial referente ao elemento corrente (i.e., $\vec{U_s} = [5, 2, 1, 0, 0, 0, 0, 0]$ no caso de um sistema de segunda ordem) e $\vec{w_s}$ é um vetor tal que cada elemento é calculado como $w_s[n] = \exp\{\beta U_s[n]\}$, e o denominador de 3.5 como o produto interno entre $\vec{w_s}$ e um vetor coluna identidade $\vec{r} = [1, 1, \ldots, 1]$, o segundo termo pode ser expandido em um somatório de λ termos, cada um associado a um possível padrão de configuração espacial.

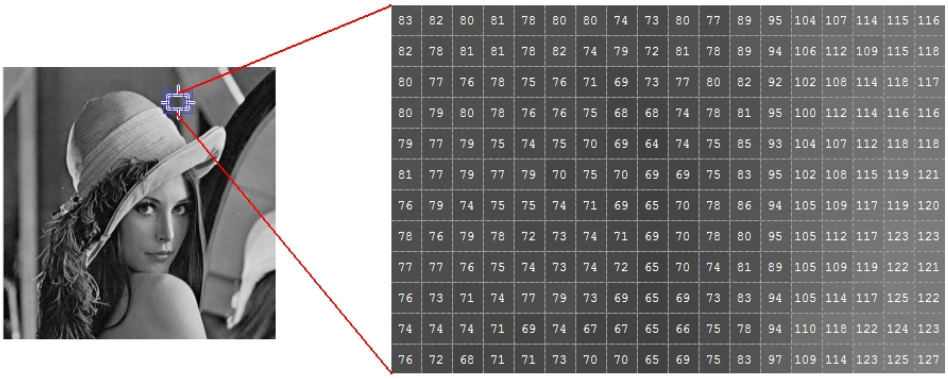
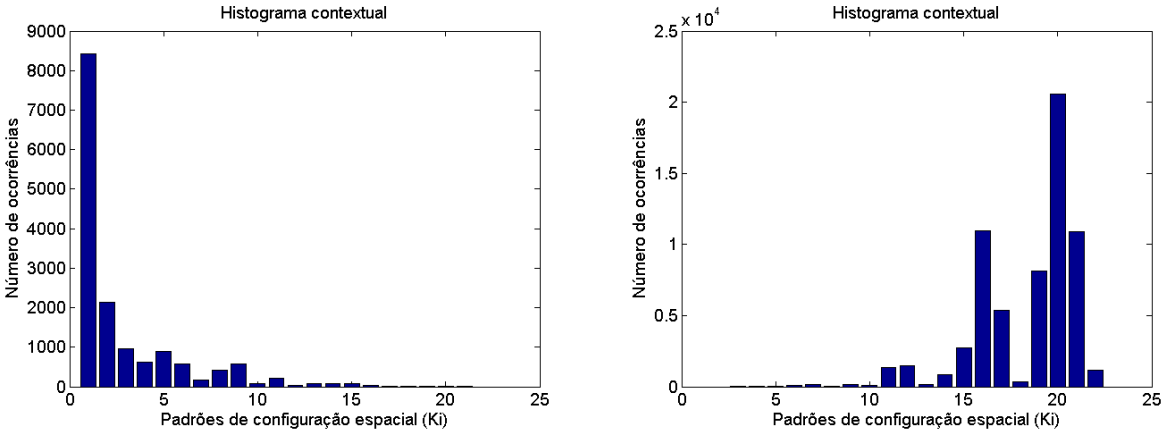


Figura 7 - Imagens suaves apresentam padrões de configuração mais homogêneos.

Entretanto, como a equação envolve a soma para todos os elementos do campo aleatório, é necessária a definição de constantes K_i, $i = 1, 2, \ldots, \lambda$, representando o número de ocorrências de cada possível padrão de configuração ao longo do campo. Basicamente, a idéia é que o conjunto dos coeficientes K_i, $i = 1, 2, \ldots, \lambda$, defina um histograma contextual, ou seja, ao invés de indicar a distribuição dos níveis de cinza da imagem, este conjunto fornece a distribuição de padrões espaciais definidos em termos de um sistema de vizinhança. Por exemplo, em aplicações de análise de imagens, existem tanto regiões suaves quanto regiões ruidosas nas imagens de interesse. Áreas homogêneas tendem a apresentar padrões contextuais onde se observa predominância de elementos semelhantes, devido à alta correlação existente entre pixels vizinhos, enquanto áreas mais ruidosas apresentam maior variabilidade em função da maior dispersão dos dados. Por exemplo, as Figuras 7, 8, 9a e 9b ilustram a diferença existente entre a distribuição de padrões de configuração espaciais na imagem Lena com e sem ruído Gaussiano.

Figura 8 - Imagens ruidosas apresentam predominância de padrões de configuração heterogêneos.



(a) Imagem sem ruído (b) Imagem com ruído

Figura 9 - Comparação entre a distribuição de padrões de configuração espacial para sistemas de segunda ordem para a imagem Lena suave e ruidosa (k_0 denota concordância total e k_{22} concordância nula).

Convém notar que padrões de configuração simétricos oferecem a mesma contribuição à equação de pseudo-verossimilhança, uma vez que o produto interno entre dois vetores não depende da ordem dos elementos. Em outras palavras, o que efetivamente contribui para a equação de pseudo-verossimilhança é apenas a configuração dos elementos vizinhos, independentemente do valor do elemento central. A expressão analítica completa para a equação geral de pseudo-verossimilhança do modelo de Potts para sistemas de segunda ordem pode ser escrita como (47):

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\log PL(\beta) &= \sum_{s\in\Omega} U_s(m_s) - \frac{8e^{8\hat{\beta}}}{e^{8\hat{\beta}}+M-1}K_1 - \frac{7e^{7\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2 - \frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3 \\
&- \frac{6e^{6\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_4 - \frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_5 - \frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_6 \\
&- \frac{5e^{5\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{\hat{\beta}}+M-4}K_7 - \frac{8e^{4\hat{\beta}}}{2e^{4\hat{\beta}}+M-2}K_8 - \frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_9 \\
&- \frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{10} - \frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{11} \\
&- \frac{4e^{4\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{12} - \frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{13} - \frac{6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{14} \\
&- \frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{15} - \frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{16} \\
&- \frac{3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{17} - \frac{8e^{2\hat{\beta}}}{4e^{2\hat{\beta}}+M-4}K_{18} - \frac{6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{19} \\
&- \frac{4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{20} - \frac{2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{21} - \frac{8e^{\hat{\beta}}}{8e^{\hat{\beta}}+M-8}K_{22} = 0 \quad (3.9)
\end{aligned}
$$

Como pode ser visto, trata-se de uma equação transcendental, pois não admite uma solução fechada. Existem diversos aspectos a serem observados referentes a equação proposta. É interessante notar que a equação é composta por 22 termos e que cada termo é definido como um produto entre dois fatores, sendo uma fração e um valor de K_i. O primeiro fator (fração) é a contribuição referente ao padrão de configuração à equação de pseudo-verossimilhança. O segundo fator nada mais é que o número de vezes que o padrão em questão ocorre ao longo do campo aleatório. Outro aspecto interessante diz respeito a validade da equação para um número arbitrário de rótulos/classes/estados M. Para valores pequenos de M, tipicamente $M = 2$, $M = 3$ e $M = 4$, a equação é bastante simplificada, pelo simples fato de que diversos padrões de configuração não são realizáveis, e conseqüentemente, vários coeficientes K_i, $i = 1, 2, \ldots, \lambda$, tornam-se nulos. Por exemplo, no caso de apenas dois rótulos (*spins*, caracterizando o modelo de Ising), a equação 3.9 tem apenas cinco termos não-nulos dos vinte e dois termos totais. Portanto, em termos práticos, redução no número de classes no problema de classificação implica em redução do custo computacional na estimação do parâmetro do modelo de Potts por máxima pseudo-verossimilhança.

Também é interessante notar a interpretação física do problema em questão. Em termos físicos, a equação proposta permite estimar uma quantidade inversamente proporcional à temperatura em um sistema de partículas, dispostas num reticulado 2-D, a partir da observação da configuração das partículas em um dado instante, considerando apenas interações por pares.

Atualmente, métodos para estimação de parâmetros em modelos Markovianos utilizando máxima pseudo-verossimilhança consideram, em sua maioria, sistemas de vizi-

nhança de primeira ordem. Resultados recentes baseados na expansão do logaritmo da função de pseudo-verossimilhança adotando sistemas de vizinhança de segunda ordem mostram uma equação analítica de 67 termos (105) (cerca de 3 vezes mais termos do que a equação proposta neste trabalho). Portanto, a equação proposta neste trabalho permite uma considerável redução na complexidade da estimação de parâmetros do modelo de Potts. Tal fato possibilitou a definição de uma equação de pseudo-verossimilhança analítica para sistemas de vizinhança de terceira ordem (ou seja, considerando os 12 vizinhos mais próximos).

Adotando a mesma metodologia, uma equação de pseudo-verossimilhança para sistemas de vizinhança de terceira ordem foi derivada pela expansão da equação 3.5 nos setenta e sete padrões de configuração obtidos resolvendo a equação 3.8 para $N = 12$. A expressão analítica completa para a equação geral de pseudo-verossimilhança do modelo de Potts para sistemas de terceira ordem é dada pela equação 3.10 (103).

Em todos os experimentos nesse trabalho a solução para as equações de máxima pseudo-verossimilhança propostas é obtida numericamente pelo algoritmo de Brent (106), que além de não necessitar do cálculo de derivadas e gradientes da função objetivo, possui taxa de convergência superlinear. Esse algoritmo é uma combinação de três métodos: o método da bissetriz, o método da secante e o método de interpolação quadrática inversa.

3.2 Aproximação para a Variância Assintótica dos Estimadores de MPV

A obtenção de estimadores não-viesados não é garantida para a estimação por máxima-verossimilhança nem para máxima pseudo-verossimilhança. Na verdade, não há nenhum método estatístico que garanta a existência de estimadores não-viesados para uma amostra finita de tamanho fixo igual a N. O que ocorre geralmente é que estimadores de máxima-verossimilhança coincidem com estimadores não-viesados de variância uniformemente mínima (ENVVUM) porque eles são funções de estatísticas suficientes completas (se um estimador de máxima verossimilhança é único, então necessariamente é função de estatísticas suficientes). Porém, a teoria assintótica traz à tona as propriedades fundamentais de procedimentos matemáticos, fornecendo um ferramental extremamente poderoso para a análise estatística. Há diversos motivos que fazem da estimação por máxima verossimilhança um método de referência (107–109). Por exemplo, ao se fazer o número de amostras crescer indefinidamente ($N \to \infty$), tais estimadores tornam-se assintoticamente

não-viesados (consistência), eficientes (possuem mínima variância) e normais. Infelizmente, essa última propriedade ainda não pode ser verificada no caso da estimação por máxima pseudo-verossimilhança. O objetivo dessa seção é justamente investigar o comportamento da estimação por máxima pseudo-verossimilhança em modelos Markovianos, mais precisamente o modelos `GMRF` e Potts, através da derivação de aproximações para a variância assintótica utilizando a informação de Fisher observada (47, 102–103).

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\log PL(\beta) \;=\; & \sum_{s\in\Omega} U_s\left(m_s\right) - \frac{12e^{12\hat{\beta}}}{e^{12\hat{\beta}}+M-1}K_1 - \frac{11e^{11\hat{\beta}}+e^{\hat{\beta}}}{e^{11\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2 - \frac{10e^{10\hat{\beta}}+2e^{2\hat{\beta}}}{e^{10\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3 - \frac{9e^{9\hat{\beta}}+3e^{3\hat{\beta}}}{e^{9\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_4 \\
- & \frac{8e^{8\hat{\beta}}+4e^{4\hat{\beta}}}{e^{8\hat{\beta}}+e^{4\hat{\beta}}+M-2}K_5 - \frac{7e^{7\hat{\beta}}+5e^{5\hat{\beta}}}{e^{7\hat{\beta}}+e^{5\hat{\beta}}+M-2}K_6 - \frac{12e^{6\hat{\beta}}}{2e^{6\hat{\beta}}+M-2}K_7 - \frac{12e^{4\hat{\beta}}}{3e^{4\hat{\beta}}+M-3}K_8 \\
- & \frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+e^{3\hat{\beta}}+M-3}K_9 - \frac{10e^{5\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{5\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{10} - \frac{6e^{6\hat{\beta}}+6e^{3\hat{\beta}}}{e^{6\hat{\beta}}+2e^{3\hat{\beta}}+M-3}K_{11} \\
- & \frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{12}\frac{6e^{6\hat{\beta}}+5e^{5\hat{\beta}}+e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{5\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{13} - \frac{7e^{7\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{7\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{14} \\
- & \frac{7e^{7\hat{\beta}}+4e^{4\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{4\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{15} - \frac{8e^{8\hat{\beta}}+4e^{2\hat{\beta}}}{e^{8\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{16} - \frac{8e^{8\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{8\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{17} \\
- & \frac{9e^{9\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{9\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_{18} - \frac{10e^{10\hat{\beta}}+2e^{\hat{\beta}}}{e^{10\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_{19} - \frac{12e^{3\hat{\beta}}}{4e^{3\hat{\beta}}+M-4}K_{20} - \frac{4e^{4\hat{\beta}}+6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-4}K_{21} \\
- & \frac{8e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{2e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{22} - \frac{8e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{2e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{23} - \frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{2\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+2e^{2\hat{\beta}}+M-4}K_{24} \\
- & \frac{5e^{5\hat{\beta}}+6e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+2e^{3\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{25} - \frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}+e^{\hat{\beta}}+M-4}}K_{26} - \frac{10e^{5\hat{\beta}}+2e^{\hat{\beta}}}{2e^{5\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{27} \\
- & \frac{6e^{6\hat{\beta}}+6e^{2\hat{\beta}}}{e^{6\hat{\beta}}+3e^{2\hat{\beta}}+M-4}K_{28} - \frac{6e^{6\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}+e^{\hat{\beta}}+M-4}}K_{29} - \frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{4\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{30} \\
- & \frac{7e^{7\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{31} - \frac{7e^{7\hat{\beta}}+3e^{3\hat{\beta}}+2e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{32} - \frac{8e^{8\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{8\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{33} \\
- & \frac{9e^{9\hat{\beta}}+3e^{\hat{\beta}}}{e^{9\hat{\beta}}+3e^{\hat{\beta}}+M-4}K_{34} - \frac{6e^{3\hat{\beta}}+6e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+3e^{2\hat{\beta}}+M-5}K_{35} - \frac{9e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{3e^{3\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-5}K_{36} - \frac{4e^{4\hat{\beta}}+8e^{2\hat{\beta}}}{e^{4\hat{\beta}}+4e^{2\hat{\beta}}+M-5}K_{37} \\
- & \frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+2e^{2\hat{\beta}+e^{\hat{\beta}}+M-5}}K_{38} - \frac{4e^{4\hat{\beta}}+6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{39} - \frac{8e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{2e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{40} \\
- & \frac{5e^{5\hat{\beta}}+6e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{2\hat{\beta}}+e^{\hat{\beta}}+M-5}K_{41} - \frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}+2e^{\hat{\beta}}+M-5}}K_{42} - \frac{5e^{5\hat{\beta}}+4e^{4\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{4\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{43} \\
- & \frac{6e^{6\hat{\beta}}+4e^{4\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{44} - \frac{6e^{6\hat{\beta}}+3e^{3\hat{\beta}}+3e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{3\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{45}\frac{7e^{7\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{46} \\
- & \frac{8e^{8\hat{\beta}}+4e^{\hat{\beta}}}{e^{8\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{47} - \frac{12e^{2\hat{\beta}}}{6e^{2\hat{\beta}}+M-6}K_{48} - \frac{3e^{3\hat{\beta}}+8e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}+M-6}K_{49} - \frac{6e^{3\hat{\beta}}+4e^{2\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-6}K_{50} \\
- & \frac{9e^{3\hat{\beta}}+3e^{\hat{\beta}}}{3e^{3\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{51} - \frac{4e^{4\hat{\beta}}+6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-6}K_{52} - \frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}+3e^{\hat{\beta}}+M-6}}K_{53} \\
- & \frac{8e^{4\hat{\beta}}+4e^{\hat{\beta}}}{2e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{54} - \frac{5e^{5\hat{\beta}}+4e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-6}K_{55} - \frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}+4e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{56} \\
- & \frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{57} - \frac{7e^{7\hat{\beta}}+5e^{\hat{\beta}}}{e^{7\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{58} - \frac{10e^{2\hat{\beta}}+2e^{\hat{\beta}}}{5e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-7}K_{59} \\
- & \frac{3e^{3\hat{\beta}}+6e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+3e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-7}K_{60} - \frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{61} - \frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-7}K_{62} \\
- & \frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{63} - \frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-7}K_{64} - \frac{6e^{6\hat{\beta}}+6e^{\hat{\beta}}}{e^{6\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{65} \\
- & \frac{8e^{2\hat{\beta}}+4e^{\hat{\beta}}}{4e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-8}K_{66} - \frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+5e^{\hat{\beta}}+M-8}K_{67} - \frac{6e^{3\hat{\beta}}+6e^{\hat{\beta}}}{2e^{3\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{68} \\
- & \frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-8}K_{69} - \frac{5e^{5\hat{\beta}}+7e^{\hat{\beta}}}{e^{5\hat{\beta}}+7e^{\hat{\beta}}+M-8}K_{70} - \frac{6e^{2\hat{\beta}}+6e^{\hat{\beta}}}{3e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-9}K_{71} \\
- & \frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+7e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+7e^{\hat{\beta}}+M-9}K_{72} - \frac{4e^{4\hat{\beta}}+8e^{\hat{\beta}}}{e^{4\hat{\beta}}+8e^{\hat{\beta}}+M-9}K_{73} - \frac{4e^{2\hat{\beta}}+8e^{\hat{\beta}}}{2e^{2\hat{\beta}}+8e^{\hat{\beta}}+M-10}K_{74} \\
- & \frac{3e^{3\hat{\beta}}+9e^{\hat{\beta}}}{e^{3\hat{\beta}}+9e^{\hat{\beta}}+M-10}K_{75} - \frac{2e^{2\hat{\beta}}+10e^{\hat{\beta}}}{e^{2\hat{\beta}}+10e^{\hat{\beta}}+M-11}K_{76} - \frac{12e^{\hat{\beta}}}{12e^{\hat{\beta}}+M-12}K_{77} = 0
\end{aligned}
\tag{3.10}
$$

A matriz de covariância assintótica para estimadores de máxima pseudo-verossimilhança é dada pela seguinte equação (110):

$$C(\vec{\theta}) = H^{-1}(\vec{\theta})J(\vec{\theta})H^{-1}(\vec{\theta}) \tag{3.11}$$

onde $H(\vec{\theta})$ é o valor esperado da matriz Hessiana (matriz das derivadas parciais de segunda ordem) e $J(\vec{\theta})$ é a variância da matriz Jacobiana (matriz das derivadas parciais de primeira ordem), dadas por:

$$H(\vec{\theta}) = E\left[\nabla^2 log\ PL(\vec{\theta})\right] \tag{3.12}$$

$$J(\vec{\theta}) = Var\left[\nabla log\ PL(\vec{\theta})\right] \tag{3.13}$$

3.2.1 Informação de Fisher Observada

Freqüentemente, na prática, não é possível calcular valores esperados como os indicados na equação 3.13, também conhecidos em estatística como informação de Fisher esperada. Nessas situações, pode-se adotar uma alternativa denominada informação de Fisher observada. Inclusive, pode ser verificado na literatura de inferência estatística que o uso da informação de Fisher observada em diversos casos é mais indicado que o próprio valor obtido pela informação esperada (111). A informação de Fisher observada, calculada utilizando a função de pseudo-verossimilhança é dada por:

$$I_{obs}(\theta) = \left[\frac{\partial}{\partial\theta} log\ PL(\theta)\right]^2 \tag{3.14}$$

e pode ser estimada da seguinte maneira, justificada pela Lei dos Grandes Números:

$$\hat{I}^1_{obs}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial\theta} log\ p(y_i|y_{\eta_i},\theta)\right]^2\Bigg|_{\theta=\hat{\theta}} \tag{3.15}$$

onde y_{η_i} denota a vizinhança do elemento y_i. Dessa forma, $\hat{I}^1_{obs}(\theta)$ é um estimador não-viesado da informação de Fisher observada, ou seja, $I_{obs}(\theta) = E\left[\hat{I}^1_{obs}(\theta)\right]$, fazendo com que $\hat{I}^1_{obs}(\theta) \approx I_{obs}(\theta)$. De maneira similar, $I_{obs}(\theta)$ pode ser estimada utilizando-se a segunda derivada da função de pseudo-verossimilhança:

$$\hat{I}^2_{obs}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial\theta^2} log\ p(y_i|y_{\eta_i},\theta)\Big|_{\theta=\hat{\theta}} \tag{3.16}$$

Convém notar que, a partir das definições acima, pode-se verificar que o valor esperado da derivada do logaritmo da função de pseudo-verossimilhança é zero, visto que $\hat{\theta}$ é um ponto crítico:

$$\begin{aligned} E\left[\frac{\partial}{\partial\theta} log\ PL(\theta)\right] &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial\theta} log\ p(y_i|y_{\eta_i},\theta)\right]^2\Big|_{\theta=\hat{\theta}} &(3.17)\\ &= \frac{1}{N}\frac{\partial}{\partial\theta} log\prod_{i=1}^{N} p(y_i|y_{\eta_i},\theta)\Big|_{\theta=\hat{\theta}} &(3.18)\\ &= \frac{1}{N}\frac{\partial}{\partial\theta} log\ PL(\theta)\Big|_{\theta=\hat{\theta}} = 0 \end{aligned}$$

A seguir expressões para a variância assintótica dos estimadores de `MPV` são derivadas para os modelos de Potts e `GMRF`.

3.2.2 GMRF

Assumindo a hipótese de independência entre os parâmetros espaciais do modelo `GMRF` ($\vec{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4]$), o que é bastante intuitivo visto que a dependência em uma direção não depende de vizinhos em outras direções, a matriz de covariância torna-se diagonal (correlação nula):

$$C(\vec{\theta}) = \begin{bmatrix} c_{11} & 0 & 0 & 0 \\ 0 & c_{22} & 0 & 0 \\ 0 & 0 & c_{33} & 0 \\ 0 & 0 & 0 & c_{44} \end{bmatrix} \tag{3.19}$$

com as variâncias assintóticas (elementos da diagonal) dadas por:

$$c_{kk}(\vec{\theta}) = \frac{Var\left[\frac{\partial}{\partial\theta_k} log\ PL(\vec{\theta})\right]}{E^2\left[\frac{\partial^2}{\partial\theta_k^2} log\ PL(\vec{\theta})\right]}, \qquad k = 1, ..., 4 \tag{3.20}$$

Expandindo o numerador, tem-se:

$$Var\left[\frac{\partial}{\partial\theta_k} log\ PL(\vec{\theta})\right] = E\left[\left(\frac{\partial}{\partial\theta_k} log\ PL(\vec{\theta})\right)^2\right] - E^2\left[\frac{\partial}{\partial\theta_k} log\ PL(\vec{\theta})\right] \tag{3.21}$$

Diante do exposto, as variâncias assintóticas $c_{kk}(\vec{\theta})$ são simplificadas para:

$$c_{kk}(\vec{\theta}) = \frac{E\left[\left(\frac{\partial}{\partial\theta_k} log\ PL(\vec{\theta})\right)^2\right]}{E^2\left[\frac{\partial^2}{\partial\theta_k^2} log\ PL(\vec{\theta})\right]} \approx \frac{\hat{I}^1_{obs}}{\left[\hat{I}^2_{obs}\right]^2} \qquad k = 1, ..., 4 \tag{3.22}$$

Portanto, a partir das funções densidade condicionais locais do modelo `GMRF` e após manipulações algébricas, chega-se às seguintes expressões para $\hat{I}^1_{obs}(\vec{\theta})$ e $\hat{I}^2_{obs}(\vec{\theta})$ (102):

$$\hat{I}^1_{obs}(\vec{\theta}) = \frac{1}{N\sigma^4}\sum_{i=1}^{N}\left\{\left[y_{ij} - \vec{\theta}^T\vec{\psi}_{ij} - \mu\left(1 - 2\vec{\theta}^T I\right)\right]\left[\psi^k_{ij} - 2\mu\right]\right\}^2 \tag{3.23}$$

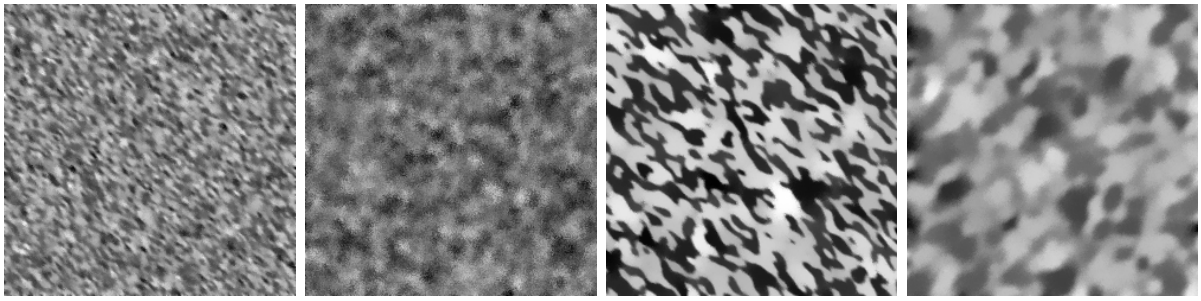
$$\hat{I}^2_{obs}(\vec{\theta}) = -\frac{1}{N\sigma^2}\sum_{i=1}^{N}\left[\psi^k_{ij} - 2\mu\right]^2 \tag{3.24}$$

onde ψ^k_{ij} representa o k-ésimo elemento de $\vec{\psi}_{ij}$, $k = 1, \ldots, 4$.

A aproximação proposta permite o cálculo da variância assintótica dos estimadores de máxima pseudo-verossimilhança do modelo `GMRF` de uma maneira matematicamente e computacionalmente viável. É conhecido que tais estimadores são assintoticamente normais e, portanto, com a aproximação proposta é possível caracterizar completamente o comportamento limite destes estimadores, bem como realizar estimação intervalar, teste de hipóteses e análises quantitativas sobre os parâmetros do modelo `GMRF` numa variedade aplicações em processamento de imagens e reconhecimento de padrões.

3.2.2.1 Análise Assintótica da Estimação por MPV no modelo GMRF

Para ilustrar a aplicação do método desenvolvido para aproximação da variância assintótica foi proposto um conjunto de experimentos utilizando imagens sintéticas geradas a partir de algoritmos de simulação de Monte Carlo (`MCMC`), mais precisamente o algoritmo *Metropolis-Hastings* (16, 41). A Figura 10 ilustra algumas imagens sintéticas que representam ocorrências de um modelo `GMRF` geradas a partir do algoritmo Metropolis variando-se os parâmetros de dependência espacial.



(a) (b) (c) (d)

Figura 10 - Imagens sintéticas geradas a partir do algoritmo Metropolis-Hastings representando diversas ocorrências do modelo GMRF.

Tabela 4 - Estimadores de MPV, variâncias assintóticas e intervalos de confiança de 90% para os parâmetros do modelo GMRF em imagens simuladas (Fig. 10a)

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% IC
1	0,25	0,2217	0,0390	[0,1799 0,3077]
2	0,3	0,2758	0,0387	[0,2398 0,3667]
3	-0,1	-0,1145	0,0394	[-0,1771 -0,0479]
4	0,2	0,1743	0,0386	[0,1150 0,2416]

Os parâmetros estimados por `MPV` foram comparados com os verdadeiros valores dos parâmetros. Em todas as simulações, os parâmetros μ e σ adotados foram zero e cinco, respectivamente. Uma das maiores dificuldades com o modelo `GMRF` é a escolha de parâmetros $\vec{\theta}$ para os quais a matriz de correlação é positiva definida. Com campos aleatórios discretos não se observa esse problema, pois quaisquer valores de parâmetros levam à construção de um modelo matematicamente válido. Entretanto, no caso do modelo GMRF, apenas uma porção do espaço paramétrico gera modelos válidos. As regiões de validade são conhecidas apenas em modelos de primeira ordem, mas não em modelos de ordens superiores (16). Na verdade, mesmo parâmetros estimados pelas abordagens conhecidas podem não pertencer à região de validade e as simulações eventualmente não funcionarem corretamente. A Tabela 4 mostra os verdadeiros valores dos parâmetros utilizados na simulação da Figura (10a), os estimadores de máxima pseudo-verossimilhança e intervalos de confiança de 90% para os parâmetros. A Tabela 5 mostra os mesmos resultados para a imagem sintética ilustrada pela Figura (10b).

Tabela 5 - Estimadores de MPV, variâncias assintóticas e intervalos de confiança de 90% para os parâmetros do modelo GMRF em imagens simuladas (Fig. 10b)

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% IC
1	0,2	0,1908	0,0506	[0,1079 0,2738]
2	0,15	0,1605	0,0524	[0,0746 0,2464]
3	0,07	0,0716	0,0482	[-0,0074 0,1506]
4	0,05	0,0523	0,0418	[-0,0146 0,1192]

É possível verificar, de acordo com nossa percepção visual, que as imagens 10a e 10b são distintas. Para ilustrar uma possível aplicação do método proposto, pode-se definir um teste de hipóteses para verificar se uma dada ocorrência do campo aleatório é isotrópica ou não, ou seja, se os parâmetros de dependência espacial são estatisticamente idênticos ou não. Considerando a imagem da Figura 10a, para um nível de significância de 5%, tem-se que a diferença entre o maior e o menor parâmetro é significativa $(0,2758 \pm 0,075; -0,1145 \pm 0,077)$ e, portanto, trata-se de uma ocorrência de um modelo Markoviao não-isotrópico. Já para a imagem da Figura 10b, isso não ocorre $(0,1908 \pm 0,099; 0,0523 \pm 0,081)$, o que nos permite concluir que se trata de uma ocorrência de um Markoviano modelo isotrópico. Outro aspecto importante observado nas simulações, é que em todos os casos, os intervalos de confiança de 90% obtidos contem o verdadeiro valor do parâmetro, mostrando a acurácia da estimação por `MPV`.

3.2.3 Modelo de Potts

Analogamente ao modelo `GMRF`, foram derivadas expressões para uma aproximação da variância assintótica desses estimadores utilizando a Informação de Fisher observada (47, 103). A idéia é a mesma do caso anterior, estudar a acurácia da estimação por máxima-verossimilhança em modelos Markovianos. Nesse caso, por se tratar de um modelo uniparamétrico, a variância assintótica é dada por:

$$c(\beta) = \frac{E\left[\left(\frac{\partial}{\partial\beta} log\ PL(\beta)\right)^2\right]}{E^2\left[\frac{\partial^2}{\partial\beta^2} log\ PL(\beta)\right]} \approx \frac{\hat{I}^1_{obs}}{\left[\hat{I}^2_{obs}\right]^2} \qquad k = 1, ..., 4 \tag{3.25}$$

Dessa forma, pode-se aproximar $\hat{I}^1_{obs}$ como:

$$\hat{I}^1_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\left[U_i(m_i) - \frac{\sum_{\ell=1}^{M} U_i(\ell) e^{\beta U_i(\ell)}}{\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}}\right]^2\right\}, \tag{3.26}$$

que, após algumas manipulações algébricas, torna-se:

$$\hat{I}^1_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{\ell=1}^{M}\left[\sum_{k=1}^{M}\left(U_i(m_i) - U_i(\ell)\right)\left(U_i(m_i) - U_i(k)\right) e^{\beta(U_i(\ell)+U_i(k))}\right]}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\} \tag{3.27}$$

O Apêndice A contém maiores detalhes sobre o desenvolvimento matemático das

equações propostas.

Analogamente, $\hat{I}^2_{obs}$ pode ser aproximada por:

$$\hat{I}^2_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\left[\sum_{\ell=1}^{M} U_i(\ell)^2 e^{\beta U_i(l)}\right]\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right] - \left[\sum_{\ell=1}^{M} U_i(\ell) e^{\beta U_i(\ell)}\right]^2}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\} \tag{3.28}$$

Com o intuito de comparar as expressões para a Informação de Fisher utilizando a primeira e a segunda derivadas, a equação 3.28 é reescrita. Após algumas manipulações algébricas, detalhadas no Apêndice A, tem-se:

$$\hat{I}^2_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{\ell=1}^{M-1}\left[\sum_{k=\ell+1}^{M}(U_i(\ell) - U_i(k))(U_i(\ell) - U_i(k)) e^{\beta(U_i(\ell)+U_i(k))}\right]}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\} \tag{3.29}$$

Note que, embora bastante similares, as equações 3.27 e 3.29 diferem quanto aos numeradores, contrariando o comportamento assintótico dos estimadores de máxima verossimilhança (variáveis independentes e identicamente distribuídas), onde prevalece a igualdade da informação, ou em outras palavras, a informação de Fisher calculada através da primeira derivada é igual a informação de Fisher obtida através da segunda derivada, o que produz estimadores eficientes no sentido estatístico, ou seja, cuja variância atinge um limiar inferior conhecido como Limite de Cramèr-Ráo (112–113). Examinando as equações propostas, é fácil notar que uma condição trivial para a igualdade da informação no modelo de Potts ocorre quando:

$$U_i(m_i) = U_i(\ell) = U_i(k), \qquad \forall\beta \tag{3.30}$$

o que significa dizer que a igualdade da informação é trivialmente satisfeita para padrões de configuração espacial onde não há dominância de nenhuma classe na vizinhança (i.e, numa condição de empate, quando a informação contextual torna-se totalmente desprezível). Porém, em tais situações a informação de Fisher observada é nula. Por exemplo, considerando sistemas de vizinhança de segunda ordem, um caso trivial que satisfaz a condição de igualdade de informação é dado por uma ocorrência do modelo de Ising (apenas 2 estados possíveis, zero e um) que simula um padrão idêntico a um tabuleiro de xadrez, onde, para todo elemento do campo, o número de vizinhos que assumem o valor zero é sempre igual ao número de vizinhos que assumem o valor um. Os autores acreditam que a verificação de condições para o surgimento do equilíbrio da informação tanto no modelo de Potts de q estados quanto em outros modelos Markovianos é um problema matemático em aberto que merece uma atenção especial. Trabalhos futuros nessa direção serão desenvolvidos

com o intuito de estudar o impacto do equilíbrio da informação no comportamento global de campos aleatórios.

A aproximação proposta permite o cálculo da variância assintótica dos estimadores dados pelas equações de pseudo-verossimilhança derivadas nesse trabalho. Para testar a acurácia da estimação a partir das equações propostas, foram geradas diversas imagens sintéticas representando ocorrências de modelos de Potts, com valores de parâmetros conhecidos, utilizando algoritmos de simulação de MCMC (16, 44–45, 114).

Basicamente, o objetivo dessa simulação é validar a seguinte hipótese:

- H: As equações de pseudo-verossimilhança propostas fornecem resultados estatisticamente equivalentes aos verdadeiros valores dos parâmetros, ou:

$$H : \beta = \hat{\beta}_{MPL} \tag{3.31}$$

Sabe-se que a distribuição assintótica de uma seqüência de estimadores de máxima pseudo-verossimilhança é normal (101). Portanto, usando a propriedade de consistência desses estimadores e adotando a aproximação proposta é possível se caracterizar completamente a distribuição assintótica do estimador em questão e definir a seguinte estatística:

$$Z_n = \frac{\beta_n - \hat{\beta}_{MPL}}{Var_n(\hat{\beta}_{MPL})} \approx N(0,1) \tag{3.32}$$

criando a seguinte regra: rejeitar a hipótese H se $|Z_n| > c$. Considerando um nível de significância $\alpha = 0,1$ (ou seja, a máxima probabilidade de se rejeitar a hipótese H dado que ela é verdadeira), tem-se que $c = 1,64$. Entretanto, deseja-se quantificar a evidência contra ou a favor da hipótese H. Por isso, foi realizada uma análise estatística completa em termos da estatística de teste (Z_n), do nível de significância α e das probabilidades de significância, ou *p-values*, calculados como $P(|Z_n| > z_{obs})$, para medir a veracidade da hipótese sendo testada. Quanto maior o valor dessa probabilidade, mais evidência a favor de H. E no caso de valores pequenos dessa probabilidade, deve-se duvidar da hipótese sendo testada. Em outras palavras, para se rejeitar H, deve-se ter um valor de α significativamente maior do que a probabilidade de significância. Essa abordagem fornece uma maneira estatisticamente correta de se reportar o resultado de um teste de hipóteses, definindo um modelo robusto para análise quantitativa dos resultados obtidos. A Figura 11 mostra um diagrama ilustrando a interpretação dos resultados de um teste de hipóteses estatístico em termos de probabilidades de significância.

Para os experimentos a seguir, foram adotados tanto algoritmos `MCMC` locais, *Gibbs*

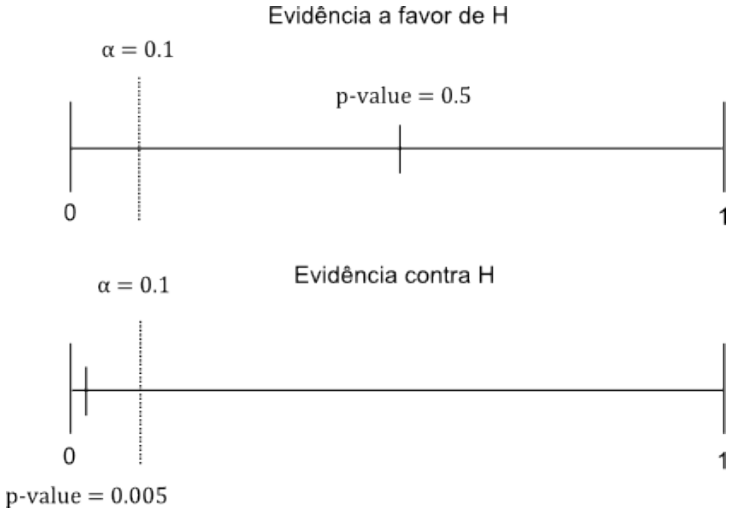
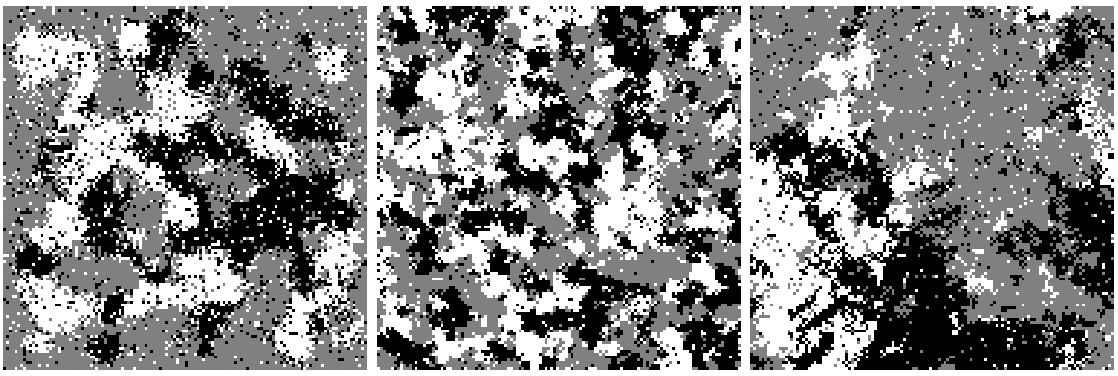


Figura 11 - Interpretação do resultado de um teste de hipóteses em termos de probabilidades de significância.

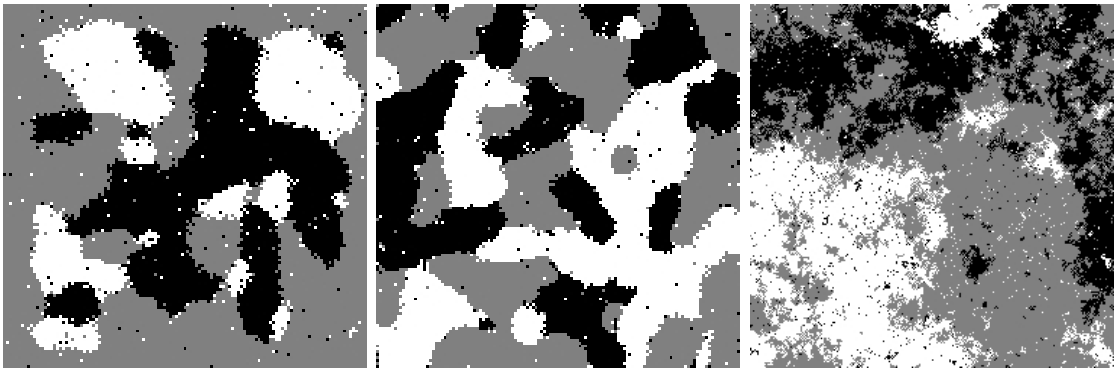
Sampler (8) e *Metropolis* (41), quanto globais, *Swendsen-Wang* (42), (44) para a geração de ocorrências do modelo de Potts. Imagens sintéticas geradas utilizando diversos valores de parâmetros, bem como vizinhanças de segunda e terceira ordens são mostradas nas Figuras 12, 13, 14 e 15.



(a) Gibbs Sampler ($\beta = 0,45$) (b) Metropolis ($\beta = 0,5$) (c) Swendsen-Wang ($\beta = 0,4$)

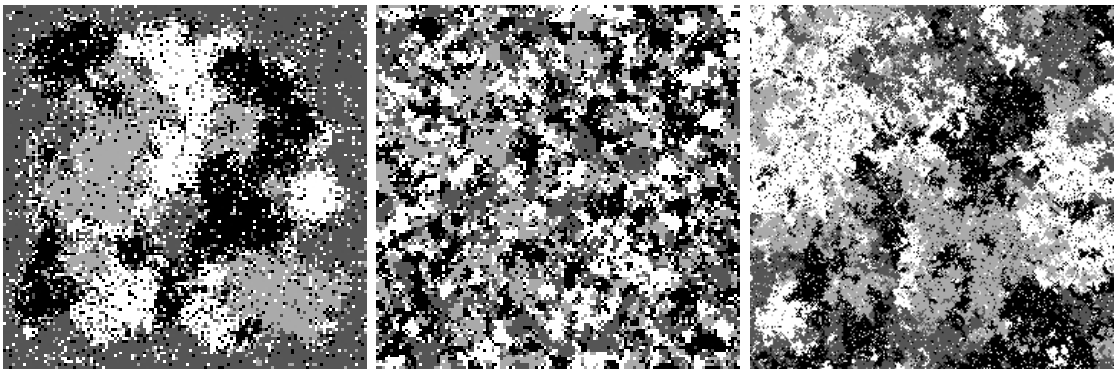
Figura 12 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de segunda ordem com três classes ($M = 3$).

Os resultados obtidos, ilustrados pelas Tabelas 6 e 7, mostram que a variância assintótica é reduzida em sistemas de vizinhança de ordens superiores, indicando que quanto maior a vizinhança, maior é a precisão da estimação por máxima pseudo-verossimilhança. Essa conclusão pode ser comprovada observando-se os valores dos erros de estimação, que diminuem sensivelmente na vizinhança de terceira ordem. Também é interessante notar que em todas as situações a estatística de teste Z_n situa-se muito abaixo do limiar ($c = 1,64$) e as probabilidades de significância muito acima de ($\alpha = 0,1$). Baseado em



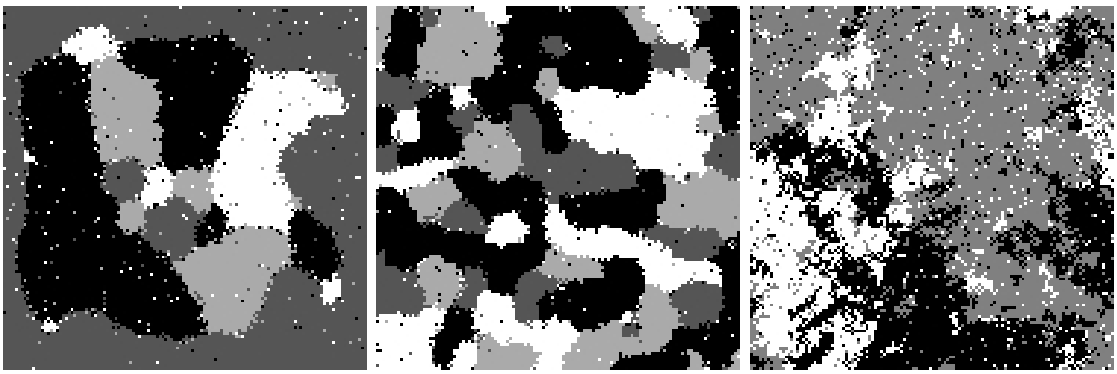
(a) Gibbs Sampler ($\beta = 0,45$) (b) Metropolis ($\beta = 0,5$) (c) Swendsen-Wang ($\beta = 0,4$)

Figura 13 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de terceira ordem com três classes ($M = 3$).



(a) Gibbs Sampler ($\beta = 0,45$) (b) Metropolis ($\beta = 0,5$) (c) Swendsen-Wang ($\beta = 0,4$)

Figura 14 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de segunda ordem com quatro classes ($M = 4$).



(a) Gibbs Sampler ($\beta = 0,45$) (b) Metropolis ($\beta = 0,5$) (c) Swendsen-Wang ($\beta = 0,4$)

Figura 15 - Imagens sintéticas geradas por algoritmos de simulação de Monte Carlo para sistemas de vizinhanças de terceira ordem com quatro classes ($M = 4$).

evidências estatísticas, é fortemente recomendado que a hipótese H seja aceita.

Portanto, considerando os dados observados na simulação, pode-se concluir que não existem diferenças significativas entre os verdadeiros valores dos parâmetros e os valores

Tabela 6 - Estimadores de MPV, informação de Fisher observada, variâncias assintóticas, estatísticas de teste e probabilidades de significância para as imagens sintéticas geradas a partir de simulação de Monte Carlo usando sistemas de vizinhança de segunda ordem

	Swendsen-Wang		*GibbsSampler*		*Metropolis*	
M	3	4	3	4	3	4
β	0,4	0,4	0,45	0,45	0,5	0,5
$\hat{\beta}_{MPL}$	0,4460	0,4878	0,3849	0,4064	0,4814	0,4889
$\left\|\beta - \hat{\beta}_{MPL}\right\|$	0,0460	0,0878	0,0651	0,0436	0,0186	0,0111
$\hat{I}^1_{obs}$	0,4694	0,6825	0,8450	1,3106	0,3908	0,8258
$\hat{I}^2_{obs}$	3,0080	3,3181	3,8248	4,5387	2,2935	2,6436
$\hat{Var}_n(\hat{\beta}_{MPL})$	0,0519	0,0620	0,0578	0,0636	0,0743	0,1182
Z_n	0,2458	0,3571	0,2707	0,1729	0,0682	0,0322
p-values	0,8104	0,7264	0,7872	0,8650	0,9520	0,9760

Tabela 7 - Estimadores de MPV, informação de Fisher observada, variâncias assintóticas, estatísticas de teste e probabilidades de significância para as imagens sintéticas geradas a partir de simulação de Monte Carlo usando sistemas de vizinhança de terceira ordem

	Swendsen-Wang		*GibbsSampler*		*Metropolis*	
M	3	4	3	4	3	4
β	0,4	0,4	0,45	0,45	0,5	0,5
$\hat{\beta}_{MPL}$	0,3602	0,3772	0,4185	0,4309	0,4896	0,4988
$\left\|\beta - \hat{\beta}_{MPL}\right\|$	0,0398	0,0228	0,0315	0,0191	0,0104	0,0012
$\hat{I}^1_{obs}$	0,2738	0,5372	0,1104	0,1433	0,0981	0,1269
$\hat{I}^2_{obs}$	3,5691	4,6800	1,8703	2,3416	1,4165	1,4547
$\hat{Var}_n(\hat{\beta}_{MPL})$	0,0215	0,0245	0,0316	0,0261	0,0489	0,0600
Z_n	0,2510	0,1456	0,1772	0,1182	0,0470	0,0049
p-values	0,8036	0,8886	0,8572	0,9044	0,9602	0,9940

estimados através das equações propostas, indicando a acurácia do método proposto.

3.3 Considerações Finais

Este capítulo apresentou de forma detalhada o desenvolvimento de novas técnicas de estimação de parâmetros e inferência estatística em campos aleatórios Markovianos, focando nos modelos de *Potts* e `GMRF`. Foram derivadas novas equações de pseudo-verossimilhança para sistemas de vizinhança de ordens superiores no modelo de *Potts* e aproximações para a variância assintótica dos estimadores de `MPV`. Análises assintóticas

através de métodos de simulação mostraram a acurácia do método de estimação proposto. De posse dos modelos e de seus respectivos parâmetros, o capítulo seguinte apresenta a metodologia proposta para a classificação contextual de imagens multiespectrais, descrevendo os classificadores, as medidas de desempenho e a análise dos resultados.

Capítulo 4

Metodologia Proposta

Este capítulo descreve em detalhes a metodologia proposta, com ênfase nos resultados da classificação contextual de imagens multiespectrais de tomografia computadorizada e ressonância magnética nuclear obtidos através dos algoritmos de otimização combinatória `ICM`, `MPM` e `GSA` utilizando sistemas de ordens superiores e combinação de múltiplas inicializações.

4.1 Classificação Supervisionada

Basicamente, o principal objetivo da área de reconhecimento de padrões é a classificação de objetos. Nesse caso, objetos são representados em termos de seus atributos, através dos vetores de padrões. Como e quantos atributos devem ser utilizados na classificação é um problema complicado de ser resolvido e, geralmente, depende muito da aplicação. Neste trabalho, os objetos a serem classificados são vetores de pixels que constituem imagens multiespectrais. A dimensionalidade dos dados depende do número de bandas de cada imagem, sendo que nos experimentos realizados, utilizam-se imagens tanto com três quanto quatro bandas, definindo espaços de atributos no $\Re^3$ e $\Re^4$.

O problema de reconhecimento de padrões pode ser visto como uma tarefa de categorização, ou seja, atribuir cada amostra a uma de C possíveis classes, sendo que o conjunto de classes pode ser previamente definido pelo analista/desenvolvedor do sistema (amostras pré-rótuladas), caracterizando o modelo de aprendizado supervisionado, ou podem ser formadas criando aglomerados por meio de padrões de similaridade (amostras sem rótulos), no caso do modelo de aprendizado não-supervisionado. Ao fim do processo de classificação, tem-se uma partição do espaço de atributos de forma que cada classe define

uma região de influência.

Existem diversas abordagens utilizadas em reconhecimento de padrões, sendo que as mais relevantes são: *template matching*, a abordagem estatística, a abordagem sintática ou estrutural e as redes neurais. Na literatura, a abordagem estatística se destaca principalmente devido ao seu forte embasamento teórico e matemático. As redes neurais também são amplamente utilizadas atualmente, principalmente pela grande capacidade de generalização. A base teórica para o reconhecimento de padrões estatístico é a Teoria de Decisão Bayesiana, cujo principal objetivo é o particionamento do espaço de atributos em regiões ótimas de decisão, no sentido de minimizar o risco de Bayes. Pode ser mostrado que, sob certas condições, a minimização do risco de Bayes é equivalente à minimização da probabilidade de erro, fazendo com que o classificador Bayesiano seja considerado ótimo. Uma extensa literatura sobre os princípios do reconhecimento de padrões estatístico pode ser encontrada em (17–19, 22, 115).

Ainda na abordagem estatística, cada padrão é representado em termos de seus D atributos e pode ser visualizado como um vetor num espaço D-dimensional. Nesse contexto, o cenário ideal consiste em se ter atributos que permitam aos vetores de padrões pertencentes a diferentes classes ocupar regiões compactas e disjuntas do espaço de atributos. A eficiência da representação do conjunto de atributos é determinada através da separabilidade entre as classes.

Diversas estratégias podem ser utilizadas no desenvolvimento de um sistema de reconhecimento de padrões, dependendo do tipo de informação estatística disponível. Todas essas abordagens são mostradas em estrutura de árvore na Figura 16, adaptada de (116). Conforme se percorre a árvore do topo até as extremidades inferiores, no sentido da esquerda para a direita, menos informação se encontra disponível e a dificuldade no problema de classificação aumenta. O foco desse trabalho é a abordagem supervisionada, onde se conhece tanto o número de classes quanto amostras de treinamento de cada uma dessas classes.

A abordagem paramétrica, como o próprio nome diz, assume um modelo paramétrico conhecido para os dados (i.e., forma da distribuição dos dados é conhecida *a priori*). O problema aqui consiste na determinação dos parâmetros das distribuições condicionais utilizando a informação existente no conjunto de treinamento (*plug-in rule*). Funções discriminantes podem ser construídas utilizando-se a regra de Bayes. Métodos não paramétricos não assumem qualquer conhecimento prévio acerca da distribuição dos dados e estimam as densidades ou os limites de decisão diretamente a partir dos dados.

Nesse trabalho foram adotados sete classificadores estatísticos para gerar diferentes inicializações aos algoritmos de otimização combinatória utilizados na classificação contextual. São eles os classificadores Bayesiano linear (`LDC`) e quadrático (`QDC`) (sob hipótese

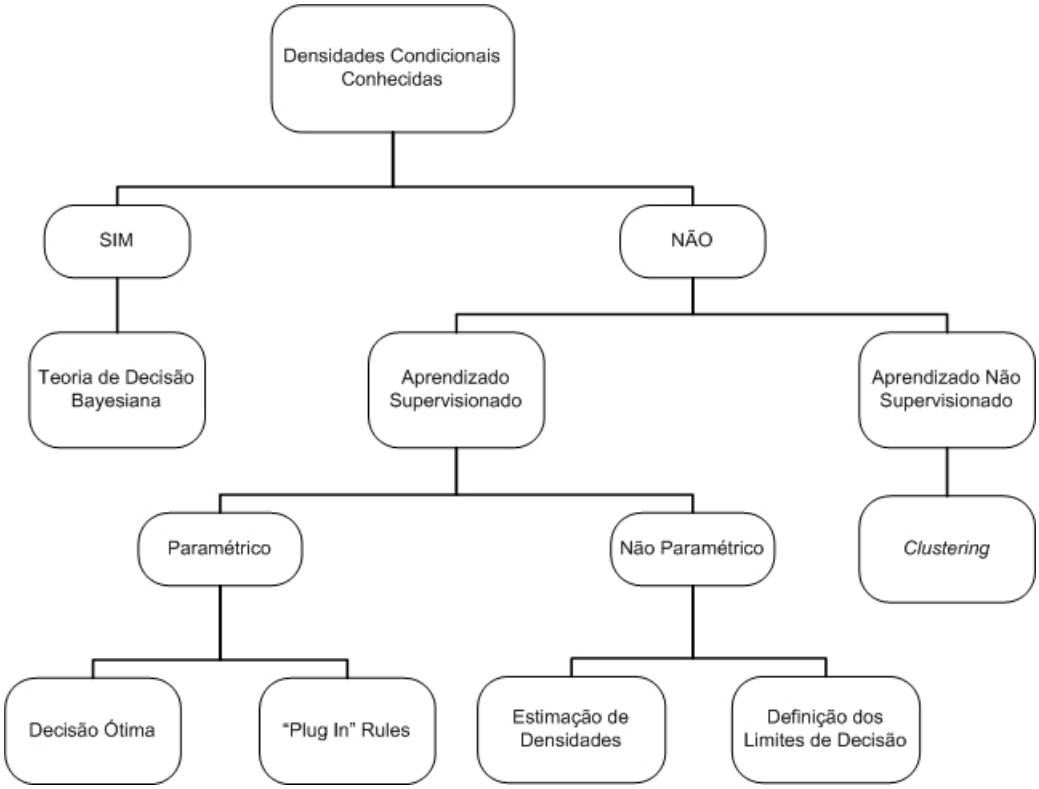


Figura 16 - Abordagens para solução de problemas em reconhecimento de padrões estatístico: conforme se percorre a árvore do topo até as extremidades inferiores, da esquerda para a direita, menos informação se encontra disponível.

gaussiana), o classificador de janelas de Parzen (`PARZENC`), o classificador K-vizinhos mais próximos ou *K-Nearest-Neighbors* (`KNNC`), o classificador logístico (`LOGLC`), o classificador de mínima distância (`NMC`) e o classificador de árvore de decisão (`TREEC`). Uma breve descrição de cada um deles é dada nas seções subseqüentes.

4.1.1 Funções Discriminantes Lineares e Quadráticas

Sob hipótese gaussiana, ou seja, considerando as densidades condicionais das classes como normais, tem-se:

$$p\left(\vec{x}|\omega_j\right) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_j\right|^{1/2}} exp\left\{-\frac{1}{2}\left(\vec{x}-\vec{\mu_j}\right)^T \Sigma_j^{-1}\left(\vec{x}-\vec{\mu_j}\right)\right\} \tag{4.1}$$

Pode-se mostrar que os componentes do vetor de parâmetros da j-ésima classe $\vec{\theta_j} = \{\vec{\mu}_j, \Sigma_j\}$, onde $\vec{\mu}_j$ é o vetor média da classe j e Σ_j é a matriz de covariância da classe j,

são estimados por máxima verossimilhança através das equações abaixo:

$$\hat{\vec{\mu}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\vec{x}_i \tag{4.2}$$

$$\hat{\Sigma}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\vec{x}_i - \vec{\mu}_j\right)\left(\vec{x}_i - \vec{\mu}_j\right)^T \tag{4.3}$$

Dessa forma, o processo de classificação pode ser realizado através do cálculo de funções discriminantes g_j. Atribui-se o vetor de padrões observado $\vec{x}_i$ à classe ω_j que fornece o máximo valor da função discriminante. Usando a regra de Bayes, desconsiderando termos constantes e utilizando os parâmetros estimados, pode-se definir a seguinte regra de decisão: atribuir o objeto $\vec{x}_i$ a classe ω_j se $g_j > g_i$ para todo $i \neq j$, onde:

$$g_i\left(\vec{x}\right) = log\left(p\left(\omega_i\right)\right) - \frac{1}{2}log\left(\left|\hat{\Sigma}_i\right|\right) - \frac{1}{2}\left(\vec{x} - \vec{\mu_i}\right)^T\Sigma_i^{-1}\left(\vec{x} - \vec{\mu_i}\right) \tag{4.4}$$

Essa regra de decisão define o classificador conhecido como Bayesiano Quadrático, ou *Quadratic Discriminant Classifier* (`QDC`).

A probabilidade *a priori* $p\left(\omega_i\right)$ pode ser estimada por $n_i/\sum_j n_j$, onde n_i denota o número de elementos da classe ω_i. Em problemas de classificação multivariados com diferentes matrizes de covariância (C classes), pode haver dados insuficientes para se obter bons estimadores para as matrizes Σ_i, $i = 1, 2, \ldots, C$. Uma alternativa para esse problema, que também diminui o custo computacional, consiste em considerar as matrizes de covariâncias de todas as classes idênticas, ou seja, $\Sigma = \Sigma_1 = \cdots = \Sigma_C$. Nesse caso, a função discriminante torna-se linear em $\vec{x}$ e simplifica-se para:

$$g_i\left(\vec{x}\right) = log\left(p\left(\omega_i\right)\right) - \frac{1}{2}\vec{\mu}_i^T\Sigma^{-1}\vec{\mu}_i + \vec{x}_i^T\Sigma^{-1}\vec{\mu}_i \tag{4.5}$$

onde Σ é a matriz de covariância comum a todas as classes. A esse classificador dá-se o nome de Bayesiano Linear, ou ainda *Linear Discriminant Classifier* (`LDC`).

4.1.2 Classificador de Janelas de *Parzen*

Trata-se de um método não paramétrico para a estimação de funções densidades a partir de um conjunto de treinamento. Basicamente, esse método tenta estimar a densidade dividindo o espaço em pequenas regiões e contando o número de amostras

localizadas em cada região. Suponha que $\Re_n$ seja um hipercubo d-dimensional (i.e, um intervalo na reta, um quadrado no plano, um cubo no espaço, e assim sucessivamente). Seja ainda h_n o comprimento da aresta deste hipercubo (i.e, o tamanho do intervalo na reta, as dimensões do quadrado no plano, etc.). Então, o volume deste hipercubo é dado por (19):

$$V_n = h_n^d \tag{4.6}$$

Pode-se derivar uma expressão analítica para o número de amostras localizadas no hipercubo, ou k_n, através da definição de uma função janela:

$$\varphi(\vec{u}) = \begin{cases} 0 & \text{se } |u_j| \leq 1/2 \\ 1 & \text{se } |u_j| > 1/2 \end{cases} \tag{4.7}$$

para $j = 1, \ldots, d$. Assim, $\varphi(\vec{u})$ define um hipercubo unitário centrado na origem. Segue diretamente da definição acima que $\varphi\left(\left(\vec{x} - \vec{x}_i\right)/h_n\right)$ é igual a um se amostra $\vec{x}_i$ cai dentro do hipercubo de volume V_n centrado em $\vec{x}$, e zero caso contrário. Portanto, o número de amostras localizadas dentro deste hipercubo é calculado por:

$$k_n = \sum_{i=1}^{n} \varphi\left(\frac{(\vec{x} - \vec{x}_i)}{h_n}\right) \tag{4.8}$$

e a densidade pode ser estimada por:

$$p_n(\vec{x}) = \frac{k_n}{nV_n} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{V_n}\varphi\left(\frac{(\vec{x} - \vec{x}_i)}{h_n}\right) \tag{4.9}$$

A análise da equação acima sugere que maneiras mais gerais de se estimar a densidade podem ser conseguidas considerando outros *kernels* ou funções janela $\varphi(\vec{x})$ ao invés da simples função uniforme. A equação 4.9 expressa a estimativa para $p_n(\vec{x})$ como uma média de funções que dependem de $\vec{x}$ e das amostras $\vec{x}_i$. Basicamente, esse *kernel* é utilizado para interpolação, suavizando a densidade resultante. *Kernels* comumente adotados são o retangular, o triangular, *biweight*, gaussiano e o de Bartlett-Epanechnikov (115). Um problema com essa abordagem é a escolha do parâmetro de suavização h_n. Se ele é muito pequeno, a densidade estimada é uma coleção de picos ruidosos, já se ele é muito grande, a densidade é demasiadamente suavizada. Maiores detalhes sobre a escolha desse parâmetro podem ser obtidos em (115). É interessante notar que se trata de um classificador não linear.

4.1.3 Classificador K-Vizinhos mais próximos

Basicamente, a motivação para essa abordagem foi o problema de se encontrar a janela ótima para a estimação da densidade, fazendo o volume de cada região ser uma função dos dados de treinamento. A regra de decisão conhecida como k-vizinhos mais próximos, ou `KNN`, é um método não paramétrico que, como o próprio nome diz, classifica um vetor de padrões $\vec{x}$ como sendo pertencente à classe com maior representatividade dentre as k amostras mais similares. Em outras palavras, a decisão é tomada examinando-se os rótulos das amostras mais próximas e realizando-se uma votação por maioria. Cada amostra define uma região de influência no espaço de atributos, de forma que a superfície de decisão é formada pela união das bordas das regiões que possuem rótulos distintos, caracterizando um comportamento não linear. As funções densidade condicionais de cada classe podem ser estimadas considerando uma hiperesfera de raio R (no caso Euclidiano) composta pelos k protótipos rotulados mais próximos a $\vec{x}$, através da seguinte expressão (115):

$$p\left(\vec{x}|\omega_j\right) \approx \frac{k_j}{N_j V_R} \tag{4.10}$$

onde k_j é o número de ocorrências de elementos da classe ω_j na região definida pela hiperesfera, N_j é o número de protótipos da classe ω_j no conjunto de treinamento e V_R é o volume da hiperesfera de raio R. O denominador da equação acima pode ser entendido como um simples fator de normalização. A probabilidade *a priori*, $p\left(\omega_j\right)$ é definida como:

$$p\left(\omega_j\right) = \frac{N_j}{N} \tag{4.11}$$

onde N é o número total de amostras de treinamento. Então, a regra de decisão consiste em atribuir $\vec{x}$ à classe ω_j que maximiza a probabilidade *a posteriori*, ou seja:

$$p\left(\omega_j|\vec{x}\right) \geq p\left(\omega_i|\vec{x}\right), \qquad i = 1, \ldots, C \tag{4.12}$$

ou ainda, aplicando a regra de Bayes:

$$\frac{k_j}{N_j V_R}\frac{N_j}{N} \geq \frac{k_i}{N_i V_R}\frac{N_i}{N}, \qquad i = 1, \ldots, C \tag{4.13}$$

ou finalmente, simplificando a equação acima:

$$k_j \geq k_i, \qquad i = 1, \ldots, C \tag{4.14}$$

É interessante observar a relação entre o limite do erro de classificação da regra do vizinho mais próximo, P_{NN} e o erro de Bayes, P^*. No caso teórico de se ter infinitas amostras de treinamento, P_{NN} é sempre menor ou igual a $2P^*$ (19).

4.1.4 Classificador de Mínima Distância

O classificador de mínima distância, ou *Nearest Mean Classifier* (`NMC`), é uma das abordagens mais simples existentes em reconhecimento de padrões. A idéia básica consiste em atribuir a amostra $\vec{x}$ para a classe que minimiza a distância entre a amostra e a média da classe. Em outras palavras:

$$c^* = argmin_c \left\{ \|\vec{x} - \vec{\mu}_c\|^2 \right\}, \qquad c = 1, \ldots, C \tag{4.15}$$

É um classificador linear, visto que as superfícies de decisão são hiperplanos, pois

$$\|\vec{x} - \vec{\mu}_c\|^2 = \vec{x}^T\vec{x} - 2\vec{x}^T\vec{\mu}_c + \vec{\mu}_c^T\vec{\mu}_c \tag{4.16}$$

e a decisão é feita tomando-se o valor que maximiza o seguinte funcional linear:

$$g_j(\vec{x}) = \vec{x}^T\vec{\mu}_j - \frac{1}{2}\vec{\mu}_j^T\vec{\mu}_j, \qquad j = 1, \ldots, C \tag{4.17}$$

4.1.5 Classificador Logístico

A hipótese básica da discriminação logística é que a diferença entre qualquer par de logaritmos de funções densidades condicionais é linear, ou seja:

$$log\left(\frac{p(\vec{x}|\omega_s)}{p(\vec{x}|\omega_c)}\right) = \beta_{s0} + \vec{\beta}_s^T\vec{x}, \qquad s = 1, \ldots, C-1 \tag{4.18}$$

Segundo (117) esse modelo oferece uma gama de vantagens, incluindo:

- É apropriado tanto para variáveis contínuas quanto discretas.

- É relativamente simples de utilizar.
- É aplicável a uma grande variedade de distribuições.
- Tem relativamente poucos parâmetros (em comparação a métodos não lineares).

A discriminação logística já foi amplamente utilizada em diversos problemas práticos com sucesso, inclusive para situações em que os dados se afastam de um comportamento gaussiano. Pode ser mostrado que as distribuições *a posteriori* têm a forma:

$$p\left(\omega_s|\vec{x}\right) = \frac{exp\left\{\beta'_{s0} + \vec{\beta}_s^T\vec{x}\right\}}{1 + \sum_{s=1}^{C-1} exp\left\{\beta'_{s0} + \vec{\beta}_s^T\vec{x}\right\}}, \qquad s = 1, \ldots, C-1 \tag{4.19}$$

$$p\left(\omega_c|\vec{x}\right) = \frac{1}{1 + \sum_{s=1}^{C-1} exp\left\{\beta'_{s0} + \vec{\beta}_s^T\vec{x}\right\}} \tag{4.20}$$

onde $\beta'_{s0} = \beta_{s0} + log\left(p\left(\omega_s\right)/p\left(\omega_c\right)\right)$. Assim, a regra para decisão depende apenas das funções lineares (numerador). Pode-se definir a seguinte regra: atribuir $\vec{x}$ à classe ω_j se:

$$max\left\{\beta'_{s0} + \vec{\beta}_s^T\vec{x}\right\} = \beta'_{j0} + \vec{\beta}_j^T\vec{x} > 0, \qquad s = 1, \ldots, C-1 \tag{4.21}$$

Os parâmetros do classificador logístico podem ser trivialmente estimados por máxima verossimilhança (118–119), através de um esquema de otimização não linear e iterativo.

4.1.6 Classificador de Árvore de Decisão

A principal vantagem dos classificadores de árvore de decisão é que eles são métodos não paramétricos e não lineares, capazes de modelar limites e superfícies de decisão extremamente complexos. Em geral, essa abordagem é recomendada para problemas em que o conjunto de dados é complexo, com classes não linearmente separáveis, além de problemas onde o conjunto de dados é composto de variáveis de tipo misto (variáveis contínuas, ordinais, nominais, etc.) A idéia básica das árvores binárias de decisão consiste em sucessivamente particionar o espaço de atributos em duas regiões até que se consiga separar os dados em classes. Um exemplo de um problema de classificação em duas classes bidimensional é ilustrado na Figura 17, adaptada de (115). Note que esse classificador é capaz de produzir limites de decisão extremamente complexos e não lineares.

A construção de uma árvore de decisão envolve basicamente três etapas (120):

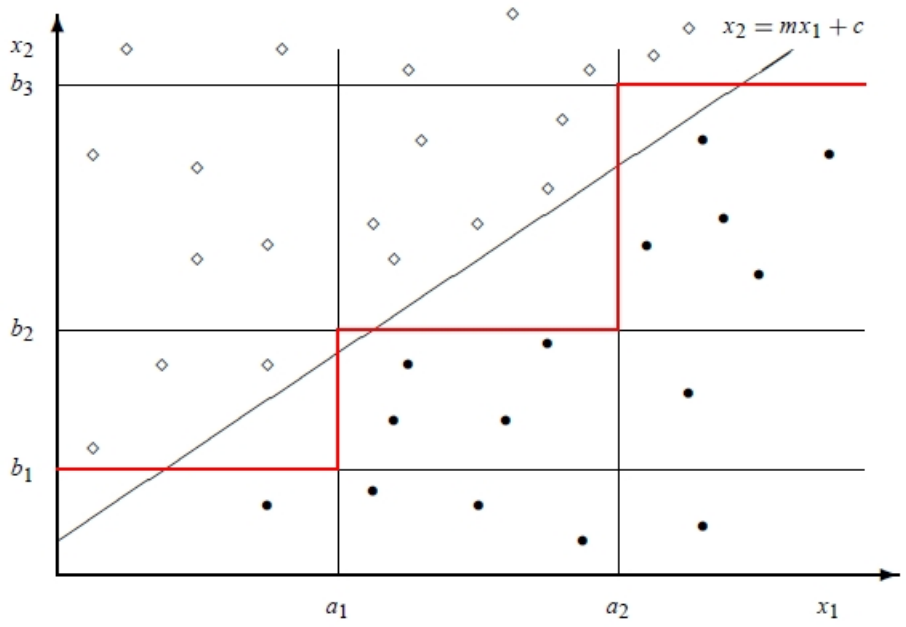


Figura 17 - Superfícies de separação geradas por um classificador de árvore de decisão para um problema de duas classes bidimensional

- Especificar uma regra de partição (splitting rule) para cada nó da árvore, ou seja, determinar os atributos e limiares que serão utilizados para particionar os dados.
- Determinação de quais serão os nós terminais (stopping rules). É preciso determinar um critério que decida quando um nó deve parar de crescer e tornar-se um nó terminal.
- Associar classes a cada um dos nós terminais. Em geral, essa associação é realizada minimizando-se uma taxa de erro de classificação estimada.

O algoritmo mais amplamente utilizado para implementar árvores de decisão é conhecido como `CART` (*Classification and Regression Tree*), (121). Entretanto, uma possível desvantagem desse método é que o tempo de treinamento pode ser excessivamente longo no caso de um grande conjunto de dados.

4.2 Medidas Quantitativas de Desempenho

Para avaliar o desempenho da classificação supervisionada de maneira objetiva é necessário utilizar critérios quantitativos. Os critérios mais utilizados são o erro estimado

de classificação e a taxa de acerto ou acurácia global. Porém, tais medidas não permitem uma análise estatística robusta, nem inferências acerca dos resultados obtidos. Com o intuito de contornar esses problemas, nesse trabalho foram adotadas duas medidas quantitativas para aferir o desempenho dos métodos propostos: o coeficiente *Kappa* de Cohen (122) e o coeficiente *Tau* de Kendall (123). Uma comparação detalhada acerca de medidas de desempenho existentes para classificação pode ser encontrada em (124). A grande motivação para a utilização desses critérios é que ambos possuem boas propriedades estatísticas, como por exemplo, o fato de serem assintoticamente normais, com expressões conhecidas e bem definidas para o cálculo de suas variâncias. Além disso, existem fórmulas para ambos serem calculados diretamente a partir da matriz de confusão (125–126), o que representa uma facilidade imensa em se tratando de problemas de classificação supervisionada. A matriz de confusão para um problema de c classes é definida como:

$$C = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1c} \\ \epsilon_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ \epsilon_{c1} & \cdots & & \epsilon_{cc} \end{bmatrix}_{c \times c} \tag{4.22}$$

onde cada elemento ϵ_{ij} representa o número de elementos da classe i classificados como sendo da classe j. Dessa forma, os elementos da diagonal indicam o número de acertos.

4.2.1 O Coeficiente Kappa

O coeficiente *Kappa* foi originalmente proposto por (122) como um método para determinar a concordância entre opiniões ou rankings expressos por diferentes especialistas. No contexto da classificação supervisionada, esse coeficiente determina um grau de concordância *a posteriori*, ou seja, dadas amostras previamente rotuladas (verdade terrestre), ele mede qual é a concordância entre esses rótulos e o resultado fornecido pelo classificador. Quanto melhor o desempenho da classificação, maior o grau de concordância e conseqüentemente, maior o valor do *Kappa*. A expressão para o cálculo do coeficiente *Kappa* a partir da matriz de confusão é dada por (125):

$$\hat{K} = \frac{N \sum_{i=1}^{C} c_{ii} - \sum_{i=1}^{C} x_{i+} x_{+i}}{N^2 - \sum_{i=1}^{C} x_{i+} x_{+i}} \tag{4.23}$$

onde x_{i+} é a soma dos elementos da linha i , x_{+i} é a soma dos elementos da coluna i, C é o número de classes e N é o número total de observações. A variância desse estimador pode ser calculada como:

$$\hat{\sigma}_k^2 = \frac{1}{N}\left[\frac{\theta_1\left(1-\theta_1\right)}{\left(1-\theta_2\right)^2} + \frac{2\left(1-\theta_1\right)\left(2\theta_1\theta_2-\theta_3\right)}{\left(1-\theta_2\right)^3} + \frac{\left(1-\theta_1\right)^2\left(\theta_4-4\theta_2^2\right)}{\left(1-\theta_2\right)^4}\right] \tag{4.24}$$

onde

$$\theta_1 = \frac{1}{N}\sum_{i=1}^{C} x_{ii} \qquad \theta_2 = \frac{1}{N^2}\sum_{i=1}^{C} x_{i+}x_{+i} \tag{4.25}$$

$$\theta_3 = \frac{1}{N^2}\sum_{i=1}^{C} x_{ii}\left(x_{i+}+x_{+i}\right) \qquad \theta_4 = \frac{1}{N^3}\sum_{i=1}^{C}\sum_{j=1}^{C} x_{ij}\left(x_{j+}+x_{+i}\right)^2 \tag{4.26}$$

A análise do valor dessa estatística indica que para valores menores que zero não existe concordância nenhuma, e para um *Kappa* igual a um a concordância é total. Uma possível interpretação para esse coeficiente é que ele expressa a "quantidade ou proporção" de erros que o método de classificação aplicado evita cometer se comparado com um resultado puramente aleatório. Alguns autores definem possíveis interpretações para o desempenho da classificação em função do valor assumido pelo coeficiente *Kappa*. A 8 ilustra uma das possíveis interpretações, sugerida por (125).

Tabela 8 - Desempenho da classificação em função do Kappa.

Coeficiente *Kappa*	**Desempenho da Classificação**
$\hat{k} \leq 0$	Péssimo
$0 < \hat{k} \leq 0,2$	Ruim
$0,2 < \hat{k} \leq 0,4$	Razoável
$0,4 < \hat{k} \leq 0,6$	Bom
$0,6 < \hat{k} \leq 0,8$	Muito Bom
$0,8 < \hat{k} \leq 1,0$	Excelente

Para testar a significância de um resultado em relação a outro, utiliza-se a estatística Z_n, para testar a hipótese de que não há diferenças significativas entre dois valores pontuais de coeficientes *Kappa*. O teste de hipóteses da desigualdade de *Kappas*

$$\begin{aligned} H_0 &: \hat{k}_1 = \hat{k}_2 \\ H_1 &: \hat{k}_1 \neq \hat{k}_2 \end{aligned} \tag{4.27}$$

pode ser efetuado calculando-se essa estatística, dada por:

$$Z_n = \frac{\left|\hat{k}_1 - \hat{k}_2\right|}{\sqrt{\hat{\sigma}^2_{k_1} + \hat{\sigma}^2_{k_2}}} \tag{4.28}$$

A diferença pode ser considerada significativa, se o valor da variável Z_n for maior que um valor crítico. Para um nível de confiança de 95%, a hipótese nula é rejeitada se $Z_n > 1,96$.

4.2.2 O Coeficiente Tau

O coeficiente *Tau* foi originalmente proposto por (123) como uma estatística não paramétrica para medir o grau de correspondência entre dois rankings ou especialistas. Assim como o coeficiente *Kappa*, posteriormente foram desenvolvidas expressões para o cálculo do coeficiente *Tau* a partir da matriz de confusão (126). A interpretação do coeficiente *Tau* é ligeiramente diferente do *Kappa*: ele indica a percentagem de elementos (pixels) a mais que foram classificados corretamente em relação ao que seria esperado apenas pela sorte.

A motivação principal para a busca de novas medidas de desempenho foi que, em relação ao coeficiente *Kappa*, observou-se que o grau de concordância por chance poderia estar sendo superestimado pelo fato de também incluir a concordância real, e por causa disso a magnitude do *Kappa* não refletiria corretamente a concordância presente na classificação (127). Analogamente ao coeficiente *Kappa*, quanto melhor o desempenho da classificação, maior o grau de concordância e conseqüentemente, maior o valor do coeficiente *Tau*. A expressão para o cálculo do coeficiente Tau a partir da matriz de confusão é dada por (126):

$$\hat{T} = \frac{P_0 - P_r}{1 - P_r} \tag{4.29}$$

onde

$$P_0 = \frac{1}{N}\sum_{i=1}^{C} x_{ii} \qquad\qquad P_r = \frac{1}{N^2}\sum_{i=1}^{C} x_{i+}x_{ii} \tag{4.30}$$

A variância desse estimador pode ser calculada através da seguinte expressão:

$$\hat{\sigma}_T^2 = \frac{P_0\,(1-P_0)}{N\,(1-P_r)^2} \tag{4.31}$$

O teste de significância para os coeficientes *Tau* também é realizado utilizando-se a estatística Z_n, análogo ao caso anterior.

4.3 Regras de Combinação de Classificadores

Seja $\vec{x} \in \Re^n$ um vetor de atributos e $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ o conjuntos de possíveis rótulos de cada classe. Cada classificador D_i do conjunto $D = \{D_1, \ldots, D_L\}$ fornece c graus de suporte, neste caso dados pelas probabilidades *a posteriori* de cada pixel observado pertencer a cada classe, calculadas através da equação 2.15. Sem perda de generalidade, pode-se assumir que todos os c graus de suporte pertençam ao intervalo $[0, 1]$, ou seja, $D_i : \Re^n \longrightarrow [0, 1]^c$. Seja ainda $d_i(\vec{x})$ o suporte que o classificador D_i fornece à hipótese de que $\vec{x}$ pertence a classe ω_j. Quanto maior esse suporte, mais provável que o verdadeiro rótulo dessa amostra seja ω_j. Supondo que existam L classificadores distintos, as L saídas para uma amostra $\vec{x}$ em particular podem ser organizadas na forma de uma matriz, denominada Perfil de Decisão (*Decision Profile*) (24). A Figura 18 ilustra a estrutura de um perfil de decisão.

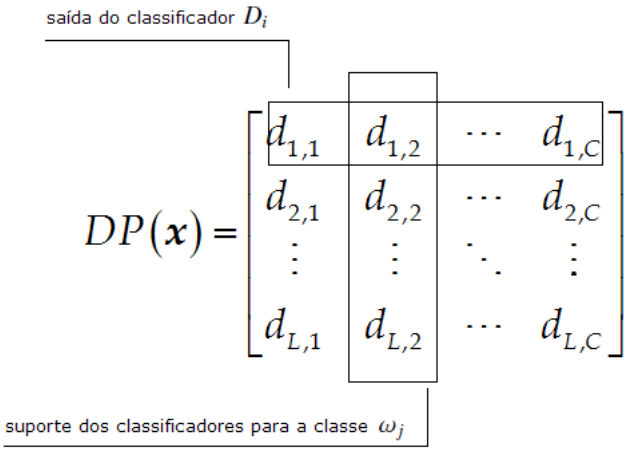


Figura 18 - Estrutura de um perfil de decisão.

Combinadores não treináveis calculam o suporte para a classe ω_j utilizando apenas a j-ésima coluna do perfil de decisão $DP(\vec{x})$ como:

$$\mu_j(\vec{x}) = \Im\left[d_{i,j}, \ldots, d_{L,j}\right] \tag{4.32}$$

onde $\Im$ é uma função de combinação. O rótulo da amostra $\vec{x}$ é então definido como sendo o índice do máximo $\mu_j(\vec{x})$. Neste trabalho, além da regra de votação por maioria, cinco outras regras (funções) estão sendo utilizadas para a combinação dos classificadores contextuais. Essas regras encontram-se definidas na Tabela 9.

Tabela 9 - Regras de combinação de classificadores para fusão de decisões.

Regra	*Função*
SOMA	$\mu_j(\vec{x}) = \sum_{i=1}^{L} d_{i,j}(\vec{x})$
PRODUTO	$\mu_j(\vec{x}) = \prod_{i=1}^{L} d_{i,j}(\vec{x})$
MÍNIMO	$\mu_j(\vec{x}) = min_i\{d_{i,j}(\vec{x})\}$
MÁXIMO	$\mu_j(\vec{x}) = max_i\{d_{i,j}(\vec{x})\}$
MEDIANA	$\mu_j(\vec{x}) = median_i\{d_{i,j}(\vec{x})\}$

4.4 Experimentos e Resultados Obtidos

Para testar a metodologia proposta de combinação de dois modelos Markovianos para classificação contextual (`GMRF` + *Potts*) adotando sistemas de vizinhanças de ordens superiores, foi proposto um conjunto de experimentos que envolvem tanto imagens multiespectrais de tomografia computadorizada (`CT`) de ciência do solo quanto imagens de ressonância magnética nuclear (`RMN`) (de frutas e de cérebro de primatas). As imagens de `CT` e `RMN` de frutas foram adquiridas utilizando os tomógrafos da Embrapa Instrumentação Agropecuária. As imagens de `RMN` de cérebro de primatas, mais precisamente de *marmocets*, uma espécie de macaco comum no Brasil e amplamente utilizado em pesquisas médicas, foram adquiridas no Laboratório de Imagens de `RMN` do Instituto de Física de São Carlos, coordenado pelo Prof. Dr. Alberto Tannús. O propósito de se realizar testes com uma grande variabilidade de imagens é analisar a robustez da metodologia proposta frente a diferentes classes de imagens multiespectrais.

A imagem de CT foi adquirida pelo minitomógrafo de raios-X e raios-γ desenvolvido pela Embrapa Instrumentação Agropecuária com o objetivo de explorar aplicações em ciência do solo (128). Para a obtenção da imagem, foram utilizadas duas fontes de raios-X e duas fontes de raios-γ (Césio e Amerício). As energias dos raios-X foram de 40keV e 85keV. Para os raios-γ foram de 60keV (Amerício) e 662keV (Césio). Foi construído um *phantom* com suporte em *plexiglass*, contendo quatro cilindros com materiais encontrados

no solo: água, alumínio, cálcio e fósforo, conforme mostra a Figura 19. As imagens foram obtidas com um baixo tempo de exposição, por isso há grande quantidade de ruído nos dados, como mostra a Figura 20.

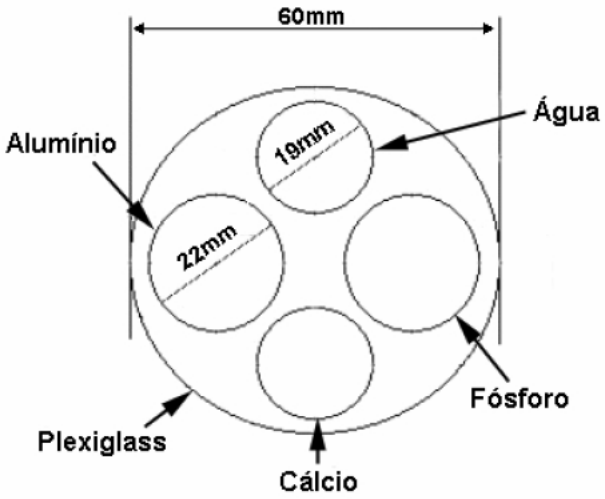
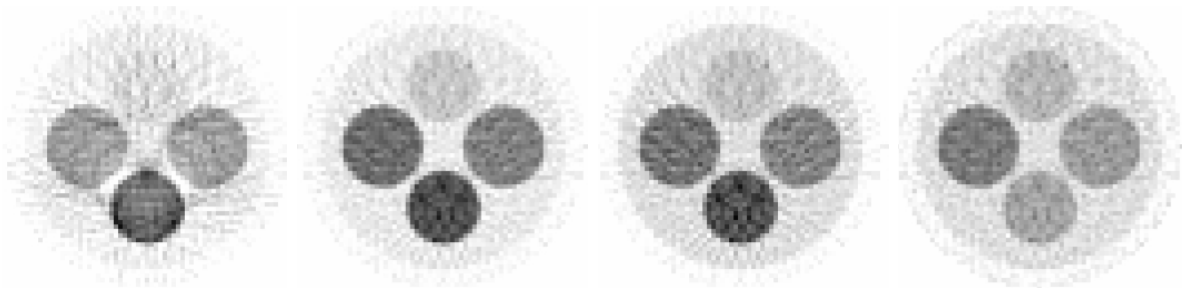


Figura 19 - Diagrama de construção do *phantom* utilizado para imageamento e estudo de materiais presentes no solo.

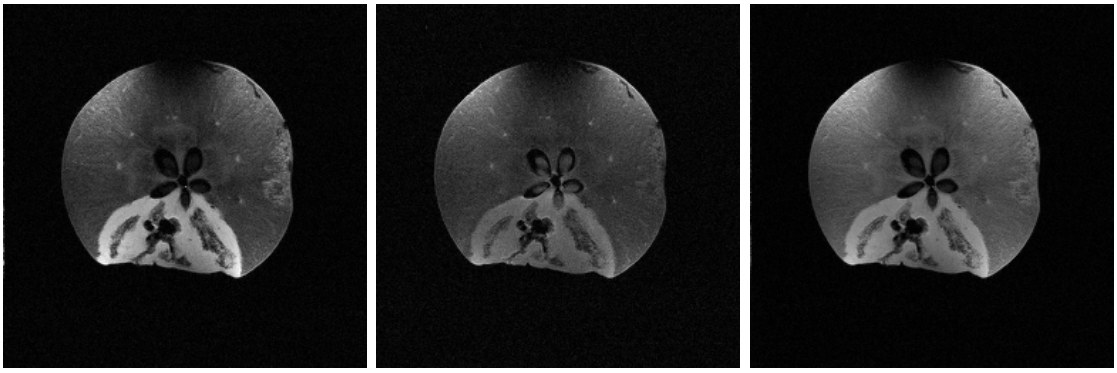


(a) 40KeV (b) 60KeV (c) 85KeV (d) 662Kev

Figura 20 - Bandas de imagem multiespectral de ciência do solo, adquirida através de raios X e γ utilizando múltiplas energias.

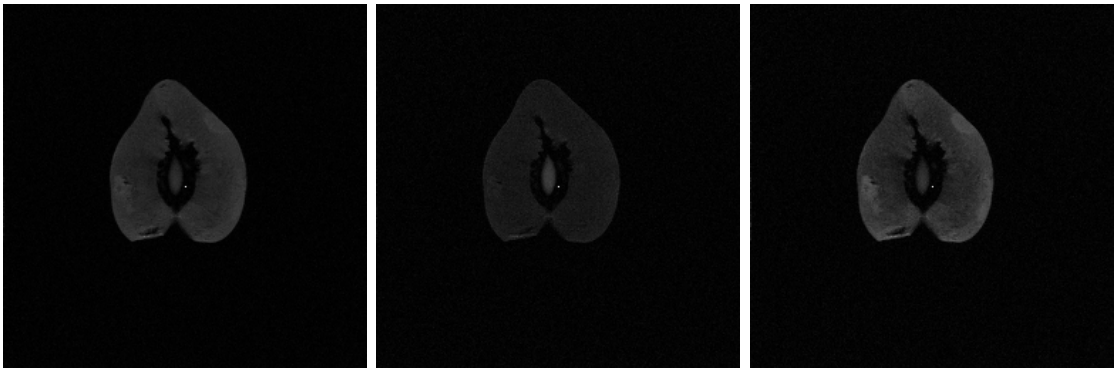
As imagens de ressonância magnética de frutas são parte de outro projeto desenvolvido pela Embrapa Instrumentação Agropecuária visando a implementação de um sistema não-invasivo de controle de qualidade de frutas para a exportação. Para esse trabalho, foram selecionadas imagens de três variedades de frutas: maça, pêra e goiaba. As Figuras 21, 22 e 23 mostram as bandas PD, T1 e T2 referentes a essas imagens.

Por fim, as imagens de `RNM` do cérebro de primatas (Figura 24) são objetos de estudo do projeto CInAPCe (Cooperação Interinstitucional de Apoio a Pesquisas sobre o Cérebro), que tem como principal objetivo o estabelecimento de uma rede científica que busca o desenvolvimento da neurociência através de abordagens multidisciplinares, através da definição de novos métodos e técnicas que visam melhorar o entendimento



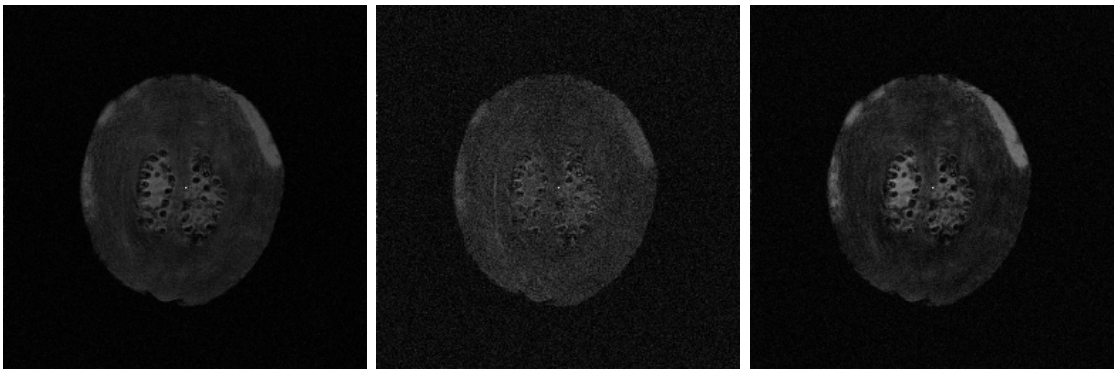
(a) PD (b) T1 (c) T2

Figura 21 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma maçã.



(a) PD (b) T1 (c) T2

Figura 22 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma pera.



(a) PD (b) T1 (c) T2

Figura 23 - Bandas de uma imagem multiespectral de ressonância magnética nuclear de uma goiaba.

dos mecanismos de dano, plasticidade e reparo em epilepsia. Trata-se de um projeto que envolve diversas instituições do estado de São Paulo (USP, UNICAMP e UNIFESP) e conta com o apoio de pesquisadores de diversas áreas do conhecimento como medicina,

física, neurociência, ciência da computação, dentre outras. O Instituto de Física de São Carlos é um dos integrantes do projeto, pois mantém um centro de excelência dedicado a pesquisa e desenvolvimento de técnicas de imageamento por ressonância magnética com amplo reconhecimento na comunidade científica a mais de 30 anos.

(a) PD (b) T1 (c) T2

Figura 24 - Bandas de uma imagem multiespectral de ressonância magnética nuclear do cérebro de um *marmocet*.

Observando as imagens multiespectrais, é possível notar que elas apresentam características próprias como diferentes níveis de ruído, diferentes intensidades e diferentes contrastes. Por exemplo, algumas bandas das imagens são fortemente afetadas por ruído, o que certamente é um grande desafio em termos de classificação.

Todo o sistema está sendo desenvolvido na plataforma `MATLAB`, devido à enorme quantidade de processamento vetorial e matricial exigida. Os algoritmos de classificação pontuais são implementados com a ajuda do pacote `PRTOOLS` versão 4.1*, desenvolvido pela universidade de Delft na Holanda e disponível gratuitamente para fins acadêmicos no site do projeto. Todos os classificadores pontuais descritos na seção 4.1 são suportados no `PRTOOLS v4.1`. Os algoritmos iterativos de otimização combinatória, bem como os algoritmos para estimação de parâmetros em modelos de campos Markovianos, foram todos desenvolvidos pelo autor.

4.4.1 Experimentos e Resultados Obtidos

Os experimentos foram conduzidos de modo a esclarecer diversas hipóteses e conjecturas levantadas acerca da metodologia proposta. Foram definidas as seguintes hipóteses:

- A) A classificação contextual é capaz de melhorar significativamente o desempenho

*Disponível gratuitamente para fins acadêmicos no site [http://www.prtools.org]

da classificação pontual.

- B) Inicializações distintas podem levar a resultados finais significativamente diferentes.

- C) O uso de sistemas de ordens superiores é capaz de provocar uma melhora significativa no desempenho da classificação contextual.

- D) Diferentes algoritmos de classificação contextual são capazes de produzir resultados significativamente distintos.

- E) A utilização de múltiplas inicializações e a combinação de algoritmos de otimização combinatória é capaz de melhorar o desempenho da classificação contextual.

A partir dos resultados quantitativos obtidos, é possível verificar cada uma das hipóteses e conjecturas definidas acima. Primeiramente, para cada imagem foram realizadas sete classificações pontuais utilizando os métodos descritos anteriormente, que são os classificadores: `LDC`, `QDC`, `KNNC`, `PARZENC`, `LOGLC`, `NMC` e `TREEC`. Para cada classificador foram calculados os coeficientes *Kappa* e *Tau*, além de suas respectivas variâncias.

Para o treinamento dos classificadores foram utilizadas 64 amostras de treinamento de cada classe para a imagem de tomografia computadorizada, seguindo as definições encontradas em (129), totalizando 384 (6 x 64) amostras, visto que se trata de um problema de classificação de seis classes, pois é preciso discriminar água, alumínio, cálcio, fósforo, *plexiglass* (suporte) e o fundo da imagem. As imagens de frutas são divididas em 2 categorias: frutas boas (ameixa e manga) e frutas com injúrias (maça, goiaba e pêra). Para as imagens de frutas boas, define-se um problema de classificação de três classes (fruta, semente e fundo), ao passo que para as imagens de frutas com injúrias, tem-se um problema de classificação de quatro classes (fruta boa, injúria, semente e fundo). Foram utilizadas 96 amostras de cada classe, seguindo as definições encontradas em (130). A matriz de confusão e o erro de classificação para os classificadores pontuais são estimados utilizando-se o método conhecido como *leave-one-out cross-validation* (`LOOCV`) (18, 131–132), uma vez que tais classificadores são supervisionados.

A seguir, a partir de cada inicialização (classificador pontual) foram executados os algoritmos de classificação contextual `ICM`, `MPM` e `GSA`, para sistemas de vizinhança de segunda e terceira ordens. Para cada resultado, os coeficientes Kappa, Tau e suas variâncias foram calculados. Para os resultados da classificação contextual, a matriz de confusão foi estimada a partir do mapa de classes resultante, comparando as regiões de interesse obtidas com a verdade terrestre.

O critério de parada adotado para os algoritmos iterativos `ICM` e `GSA` foi o mesmo para todos os experimentos ao longo desse trabalho. Os algoritmos param ao atingir uma de duas condições: convergência (quando menos de 0,1% dos pixels são atualizados de uma iteração para outra), ou quando o número máximo de 5 iterações é alcançado. No caso do algoritmo `MPM`, os parâmetros de "*burn-in*" (k) e o número de passos da simulação MCMC (n) utilizados são 10 e 50. Uma justificativa para essa escolha é a utilização de excelentes estados iniciais para a cadeia de Markov que controla a simulação. Cada uma das possíveis inicializações é gerada a partir de processos de aprendizagem supervisionados aplicados aos dados observados e não apenas de forma aleatoria como tradicionalmente é feito, de forma que toda a supervisão do processo está embutida na inicialização da cadeia de Markov. Assim, é interessante que o estado final na simulação MCMC seja de fato correlacionado com o estado inicial, uma vez que, no *framework* MAP-MRF proposto, deseja-se justamente verificar o efeito da combinação de diferentes inicializações no resultado final da classificação como uma maneira de reduzir o custo computacional. Nesse contexto, estados fortemente descorrelacionados do estado inicial pouco contribuem para a discriminação das regiões da imagem.

Como o volume de informações referente a cada experimento é consideravel, foram feitas análises estatísticas focadas nas hipóteses e conjecturas definidas anteriormente como uma maneira de sumarizar e destacar aspectos importantes presentes nos resultados. Essa análise é feita na seção subseqüente. Foi observado que os coeficientes *Kappa* e *Tau* forneceram resultados quantitativos altamente correlacionados. Por esse motivo, toda análise estatística feita na verificação das hipóteses levantadas é realizada sobre o coeficiente *Kappa*. Os resultados quantitativos (coeficientes Kappa e Tau) referentes às classificações das imagens descritas anteriormente encontram-se nos Apêndices B a F.

É interessante notar uma situação recorrente nos experimentos desenvolvidos: em diversas ocasiões foi observado que nem sempre a melhor inicialização fornece os melhores resultados finais, o que é uma evidência de que mesmo classificadores pontuais de fraco desempenho podem contribuir para melhorar a classificação contextual. Os resultados obtidos também indicam uma característica interessante dos algoritmos de otimização combinatória sub-ótimos: embora um método leve certa vantagem sobre os demais em casos específicos, não existe um algoritmo que seja melhor ou mais adequado para todas as situações. Esta é certamente uma evidência a favor da combinação, visto que não há um método que forneça desempenho uniformemente superior a outro. Por fim, para a classificação das imagens `RMN` de cérebro de primatas foram utilizadas 300 amostras de treinamento para cada uma das 3 classes: massa branca, massa cinzenta e fundo da imagem, o que representa apenas cerca de 1% do total de pixels observados.

No que diz respeito à combinação de algoritmos de otimização combinatória, três formas distintas foram consideradas. Os experimentos foram conduzidos de maneira a se es-

clarecer algumas conjecturas acerca da abordagem MAP-MRF proposta para classificação contextual. Basicamente, o objetivo é verificar experimentalmente algumas hipóteses acerca do método proposto, de acordo com as seguintes diretivas:

- Combinação de um único algoritmo de otimização combinatória utilizando múltiplas inicializações.
- Combinação de diferentes algoritmos de otimização combinatória utilizando as mesmas inicializações.
- Combinação de diferentes algoritmos de otimização combinatória utilizando inicializações distintas.

4.4.2 Análise Estatística

Para testar hipóteses e validar os resultados obtidos com a metodologia MAP-MRF proposta foram realizadas análises estatísticas tanto locais quanto globais. Basicamente, a diferença entre dois valores de coeficiente Kappa pode ser considerada significante se, para um determinado nível de significância α, a estatística de teste Z, dada pela equação 4.28, é superior a um valor crítico z_c. Para testar se os desempenhos médios obtidos em dois experimentos são significativamente distintos, utiliza-se a estatística T. Definindo $\bar{k} = \bar{k}_1 - \bar{k}_2$, onde $\bar{k}_1$ e $\bar{k}_2$ denotam os coeficientes Kappa médios antes e depois da aplicação de uma determinada técnica, deseja-se testar a hipótese $H_0 : \bar{k} = 0$ (não houve mudança) versus $H_1 : \bar{k} \neq 0$ (houve melhora significativa). Duas situações distintas podem ocorrer: os grupos contém o mesmo número de amostras, $N_1 = N_2 = N$ (teste pareado) ou o número de amostras em cada grupo é distinto, $N_1 \neq N_2$ (teste não pareado). No caso pareado, a estatística T tem distribuição *t-student* com $N - 1$ graus de liberdade sendo definida como segue:

$$T = \frac{\bar{k}}{DV_d/\sqrt{N}} \tag{4.33}$$

onde DV_d denota o desvio padrão das diferenças pontuais entre os elementos correspondentes de cada um dos grupos. Note que, a decisão baseada na variável T é bastante intuitiva, pois quanto maior for a diferença entre as médias amostrais, maior a chance de se ter dois grupos distintos (captado pelo numerador de T). Por outro lado, quanto maior for a variabilidade dos resultados observados, maior será a dificuldade de se detectar diferenças entre as médias dos grupos (captado pelo denominador de T).

No caso não pareado, a estatística T tem distribuição *t-student* com ρ graus de liberdade, sendo dada por:

$$T = \frac{\bar{k}_1 - \bar{k}_2}{\sqrt{\frac{\sigma_{k_1}^2}{N_1} + \frac{\sigma_{k_2}^2}{N_2}}} \tag{4.34}$$

onde $\sigma_{k_1}^2$ e $\sigma_{k_2}^2$ são as variâncias amostrais de cada grupo de medidas e N_1 e N_2 denotam o número de elementos em cada grupo. Nessa situação, a equação de *Welch-Satterthwaite* (133–134) é empregada no cálculo de uma aproximação para o número efetivo de graus de liberdade da distribuição da estatística de teste:

$$\rho = \frac{\left(\frac{\sigma_{k_1}^2}{N_1} + \frac{\sigma_{k_2}^2}{N_2}\right)^2}{\frac{\left(\sigma_{k_1}^2 \big/ N_1\right)^2}{N_1 - 1} + \frac{\left(\sigma_{k_2}^2 \big/ N_2\right)^2}{N_2 - 1}} \tag{4.35}$$

Dessa forma, adotando um nível de significância α (i.e., $\alpha = 0,05$ ou $\alpha = 0,01$), deve-se rejeitar a hipótese nula se T for menor que um valor crítico t_c. Essas informações, juntamente com a probabilidade de significância (*p-values*), dada por $P\left(T < t_c\right)$, permitem uma análise robusta dos resultados, bem como a inferência de conclusões relevantes acerca do método proposto.

4.4.3 Discussão e Análise dos Resultados

Para destacar alguns dos aspectos mais relevantes presentes nos resultados obtidos, essa seção tem como objetivo testar diversas hipóteses e conjecturas feitas acerca da metodologia proposta. Para isso, foram utilizados testes T (análise global), que verificam se a diferença entre as médias de duas amostras normais é significativa, bem como a estatística Z e construção de intervalos de confiança (análise pontual), para aferir a significância estatística de alguns resultados. As tabelas referenciadas nessa seção que remetem ao desempenho da classificação (*Kappas*) encontram-se todas nos Apêndices B a F. Toda análise e discussão foi realizada com base nesses dados. A primeira hipótese (**A**) levantada diz respeito à classificação contextual e pode ser enunciada como:

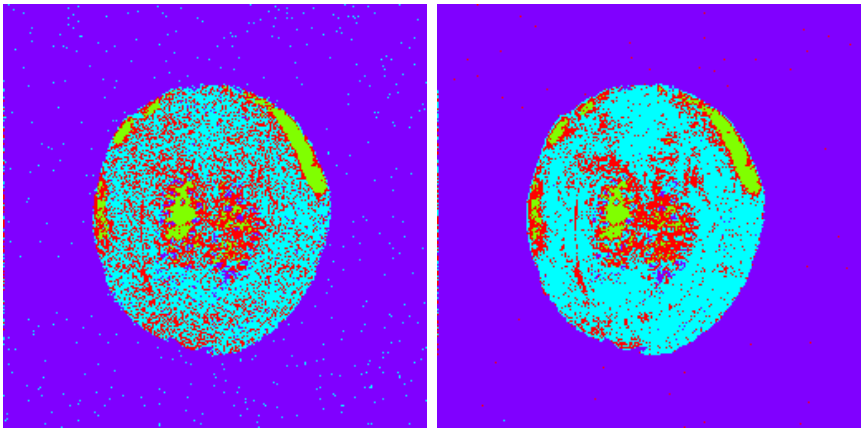
- **A**: *O uso de algoritmos iterativos de classificação contextual é capaz de melhorar significativamente o desempenho da classificação pontual* ($\bar{k} < 0$)

Para verificar essa hipótese, quatro casos distintos foram analisados: a imagem da goiaba (classificação pontual x `MPM` com vizinhança de segunda ordem), conforme dados das Tabelas 30 e 31, a imagem da maçã (classificação pontual x `ICM` com vizinhança de segunda ordem), conforme dados das Tabelas 22 e 24, a imagem de tomografia computadorizada (classificação pontual x `GSA` com vizinhança de segunda ordem), conforme dados das Tabelas 18 e 21 e a imagem do cérebro do *marmocet* (classificação pontual x `ICM` com vizinhança de segunda ordem), conforme dados das Tabelas 34 e 36. Considerando um nível de significância $\alpha = 0,05$, tem-se um $t_c = -1,943$ (vide tabela da distribuição t_6).

Tabela 10 - Resultados referentes ao teste da hipótese A

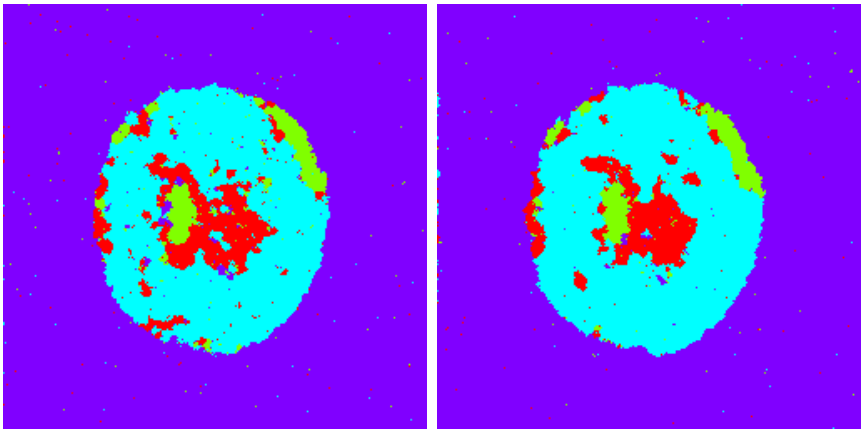
	Goiaba (RMN) Pontual x `MPM`	Maça (RMN) Pontual x `ICM`	*Phantom* (CT) Pontual x `GSA`	*Marmocet* (RMN) Pontual x `ICM`
$\bar{k}$	$-0,1774$	$-0,0719$	$-0,0914$	$-0,2098$
T	$-6,6377$	$-3,6649$	$-2,8483$	$-8,7741$
p-values	$< 0,0005$	$< 0,007$	$< 0,015$	$< 0,0005$

Note que, nos quatro casos analisados o valor da estatística de teste é sempre inferior ao valor crítico t_c. Além disso, as probabilidades de significância são todas pequenas e menores que α caracterizando fortes evidências contra a hipótese em questão. Portanto, deve-se **rejeitar** a hipótese de que não houve mudanças ($\bar{k} = 0$), o que significa que o uso de algoritmos iterativos para classificação contextual é de fato capaz de melhorar significativamente o desempenho médio da classificação pontual. Para ilustrar as conclusões obtidas, as Figuras 25, 26, 27, 28, 29, e 30 mostram alguns resultados visuais (mapas de classe) obtidos nos experimentos em questão. A legenda para as cores nas imagens de `RMN` de frutas é a seguinte: roxo representa o que foi classificado como fundo da imagem, verde-água como fruta, vermelho como semente e verde-claro como injúria (parte estragada). Para a imagem de ciência do solo, a legenda é: rosa representa o fundo da imagem, azul-escuro o suporte de plexiglass, vermelho a água, amarelo o alumínio, verde o cálcio e azul-claro o fósforo. Finalmente, para a imagem do cérebro, vermelho denota massa cinzenta, verde representa massa branca e azul denota o fundo da imagem. Note o aspecto ruidoso dos resultados obtidos através de classificadores pontuais.



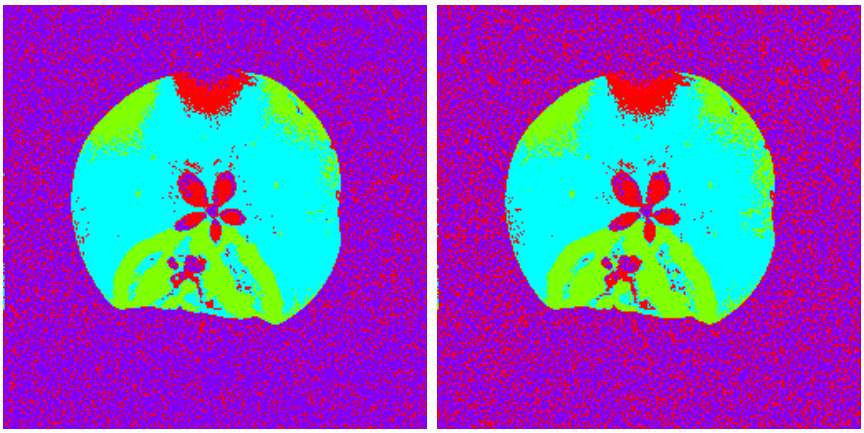
(a) LDC (b) PARZENC

Figura 25 - Mapas de classes para classificação pontual da imagem goiaba.



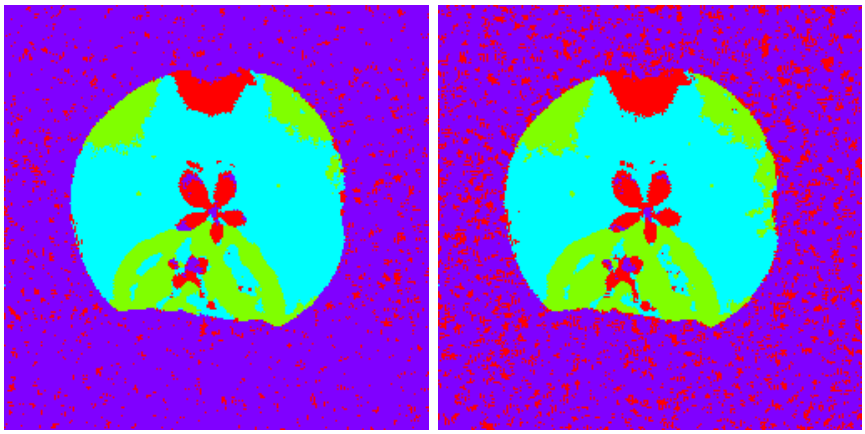
(a) LDC+MPM (b) PARZENC+MPM

Figura 26 - Mapas de classes para classificação contextual da imagem goiaba.



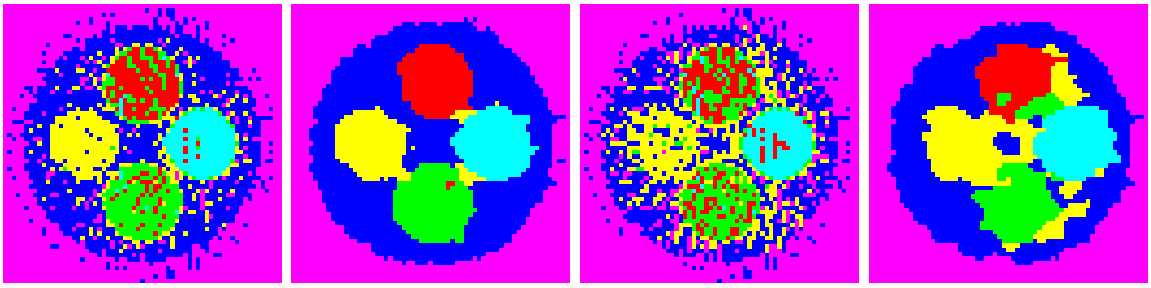
(a) KNNC (b) TREEC

Figura 27 - Mapas de classes para classificação pontual da imagem maçã.



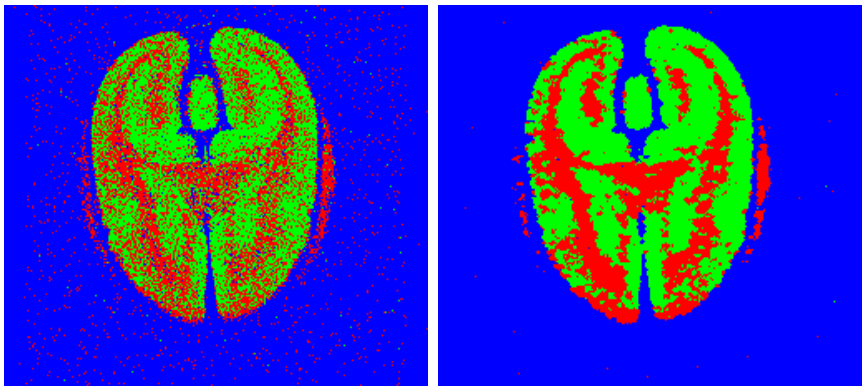
(a) KNNC+ICM (b) TREEC+ICM

Figura 28 - Mapas de classes para classificação contextual da imagem maçã.



(a) PARZENC (b) PARZENC+GSA (b) NMC (b) NMC+GSA

Figura 29 - Mapas de classes para classificação pontual e contextual do phantom de tomografia computadorizada de ciência do solo.



(a) LOGLC (b) LOGLC+GSA

Figura 30 - Mapas de classes para classificação pontual e contextual da imagem de cérebro de um *marmocet*.

A segunda hipótese (**B**) levantada diz respeito à classificação contextual e pode ser enunciada como:

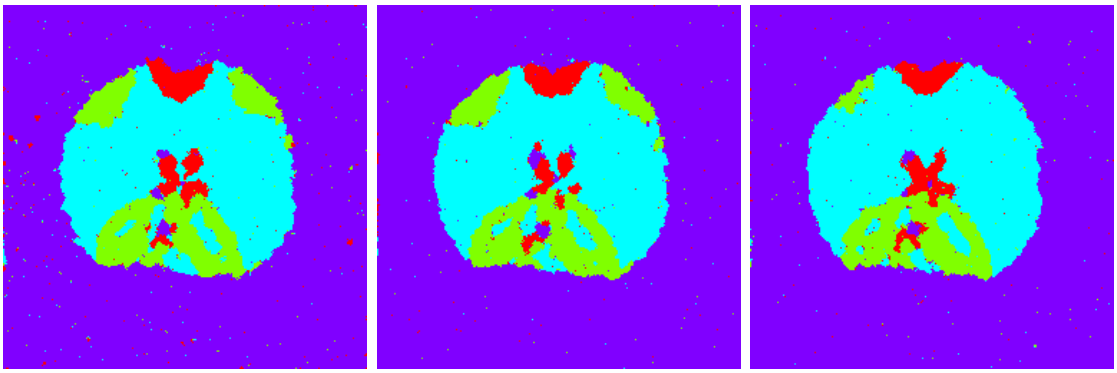
- **B**: *Inicializações distintas podem levar a resultados finais significativamente diferentes (considerando um único algoritmo iterativo)*

Para verificar experimentalmente essa hipótese, outros três casos pontuais e distintos foram analisados utilizando a imagem maçã (`RMN`) e o algoritmo iterativo de classificação contextual `MPM` com sistema de vizinhança de segunda ordem, conforme dados da Tabela 19. É possível notar que um mesmo algoritmo (`MPM`) leva a resultados significativamente diferentes, variando-se apenas a inicialização. Portanto, a hipótese **B** é válida, o que de certa maneira já era esperado, tendo em vista que os métodos de otimização combinatória adotados são sub-ótimos, ou seja, convergem para mínimos/máximos locais, sendo assim fortemente afetados pela condição inicial.

Tabela 11 - Resultados referentes ao teste de hipótese B: Kappas obtidos para um mesmo algoritmo iterativo são significativamente distintos

	`KNNC+MPM`	`NMC+MPM`	`LDC+MPM`
$\hat{k}$	$0,7687$	$0,9031$	$0,9750$
$Var(\hat{k})$	$5,451 \times 10^{-4}$	$2,775 \times 10^{-4}$	$7,648 \times 10^{-5}$
IC de 95%	$0,7687 \pm 0,0452$ [0,7235, 0,8139]	$0,9031 \pm 0,0319$ [0,8712, 0,9350]	$0,9750 \pm 0,0169$ [0,9581, 0,9919]

A Figura 31 mostra os resultados visuais referentes a cada uma das três classificações mostradas na Tabela 11. É visualmente perceptível que existem diferenças notáveis entre cada um dos casos.



(a) KNNC+MPM (b) NMC+MPM (c) LDC+MPM

Figura 31 - Mapas de classes para classificação contextual da imagem maça para diferentes inicializações.

A terceira hipótese (**C**) levantada acerca do método proposto pode ser enunciada como:

- **C**: *O uso de sistemas de vizinhança de ordens superiores é capaz de causar uma melhora significativa no desempenho da classificação contextual.* ($\bar{k} < 0$)

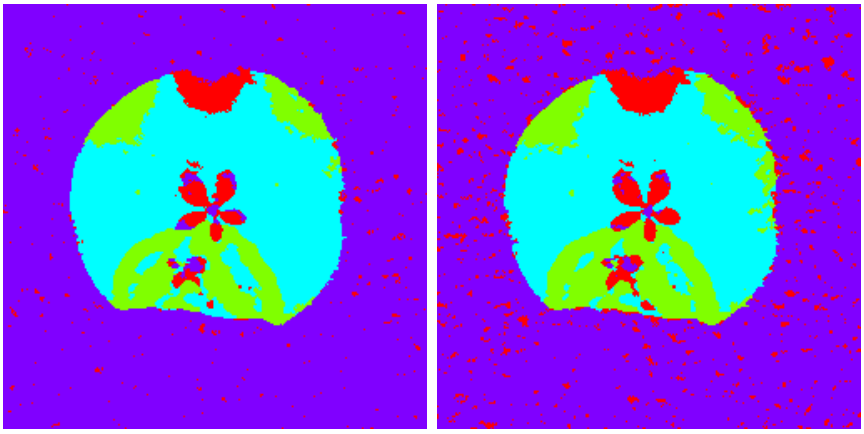
Para verificar essa hipótese, três casos específicos foram analisados: a classificação contextual das imagens da maçã, goiaba e da imagem de cérebro utilizando o algoritmo iterativo `GSA` tanto com sistema de vizinhança de segunda quanto de terceira ordem foi examinado, conforme os dados das Tabelas 25, 33, e 37. Considerando um nível de significância $\alpha = 0,05$, tem-se um $t_c = -1,943$ (vide tabela da distribuição t_6). O cálculo da diferença média e de seu respectivo desvio padrão fornece, para a imagem da maçã (Tabela 25), os valores $\bar{k} = -0,0238$ e $DV_d = 0,0112$, o que, para um $n = 7$, resulta num valor de $T = -5,6033$, que é significativamente menor que t_c. A probabilidade de significância, ou $P(T < -5,6033)$, nesse caso é menor que 0,001, o que quantifica uma forte evidência contra a suposição de que não houve alterações. Para a imagem da goiaba (Tabela 33), os valores obtidos são $\bar{k} = -0,05804$ e $DV_d = 0,0471$, resultando num valor de $T = -3,2603$. A probabilidade de significância é 0,01, ou seja, menor que α, o que também quantifica evidência contra a hipótese nula. Por fim, para a imagem de cérebro (Tabela 37), os valores obtidos foram $\bar{k} = -0,0229$ e $DV_d = 0,0122$, resultando num valor de $T = -4,9655$.

Portanto, considerando os dados observados, deve-se rejeitar a hipótese de que $\bar{k} = 0$, o que implica em dizer que nesse caso a utilização de sistemas de vizinhança de terceira ordem provocou uma melhora significativa no desempenho da classificação contextual. Vale ressaltar a força desses resultados, comentando que essas conclusões seriam suportadas mesmo para valores de α menores de 0,01, ou seja, com menos de 1% de chance de rejeitar H_0 incorretamente, o que quantifica de forma veemente a veracidade da hipótese **C**.

Para ilustrar as diferenças visuais entre as classificações contextuais, as Figuras 32, 33, 34, 35, 36 e 37 comparam alguns mapas de classes obtidos utilizando `GSA` em vizinhanças de segunda e terceira ordens. A aparência visual das imagens classificadas utilizando a vizinhança de ordem superior é consideravelmente melhor. Note, inclusive, que é possível eliminar diversos artefatos presentes na classificação realizada com vizinhança de segunda ordem, como por exemplo, as manchas causadas por erros devido a presença de ruído na imagem.

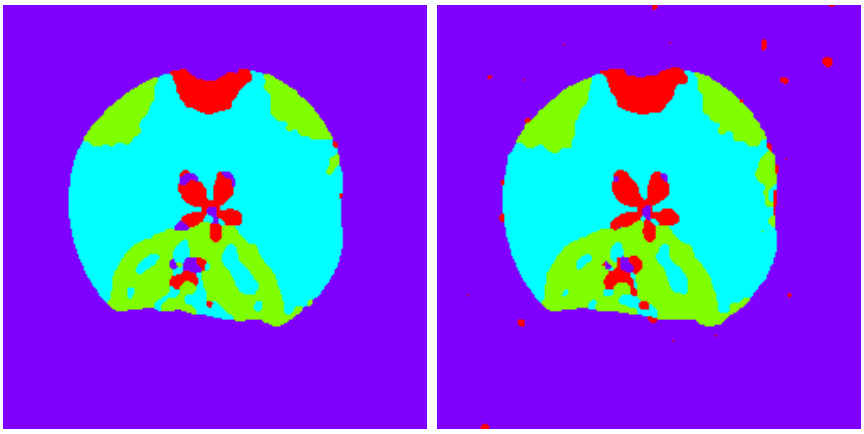
Finalmente, a quarta hipótese (**D**) levantada acerca do método proposto pode ser enunciada como:

- **D**: *Diferentes algoritmos de classificação contextual são capazes de gerar resultados estatisticamente distintos, mesmo para inicializações idênticas.* ($\bar{k} < 0$)



(a) KNNC+GSA (b) TREEC+GSA

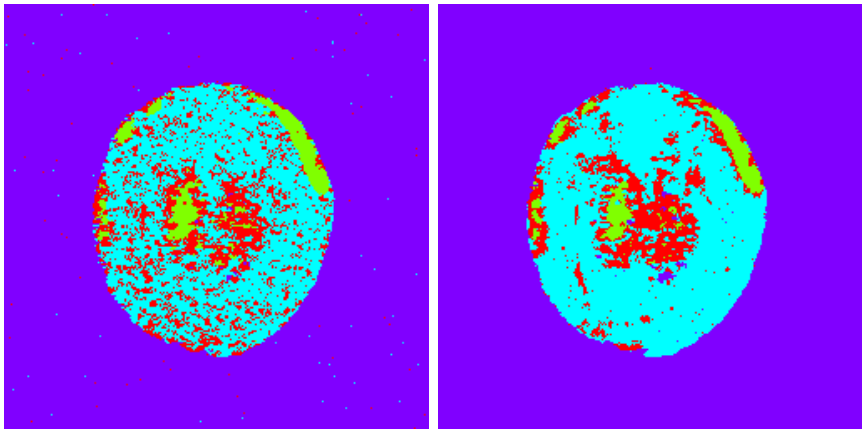
Figura 32 - Mapas de classes para classificação contextual da imagem maça utilizando sistemas de vizinhança de segunda ordem.



(a) KNNC+GSA (b) TREEC+GSA

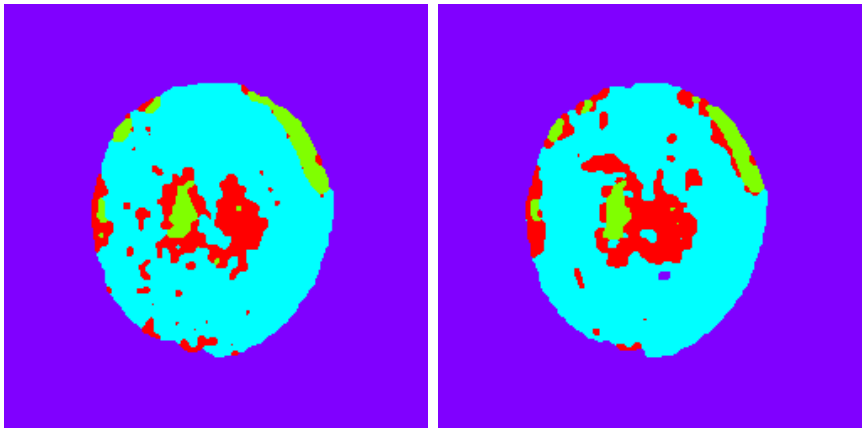
Figura 33 - Mapas de classes para classificação contextual da imagem maça utilizando sistemas de vizinhança de terceira ordem.

Para verificar essa hipótese, os casos das classificações contextuais da imagem da pêra e da imagem de cérebro utilizando tanto o algoritmo iterativo `MPM`, quanto o algoritmo `ICM` foram examinados, conforme os dados das Tabelas 27, 28 e 35. Primeiramente, foi realizado um teste global (T), considerando um nível de significância $\alpha = 0,05$. Para esse valor de T tem-se um $t_c = -1,943$. Para a imagem da pêra, o cálculo da variável T, para $n = 7$, fornece um valor de $T = -3,7116$, que é menor que T_c. A probabilidade de significância, ou $P(T < -3,7116)$, nesse caso é igual a 0,006, o que provê fortes evidências contra a hipótese de não haver mudanças (H_0), o que implica que se deve aceitar a hipótese **D** como válida para os dados observados em questão. Analogamente, para a imagem do cérebro, tem-se $T = -4,0451$, o que fornece uma probabilidade de significância menor que 0,0005, quantificando evidências fortes contra H_0 (e a favor de **D**). Também foi testada a significância dos valores dos coeficientes *Kappa* obtidos em três situações pontuais. As Tabelas 12, 13 e 14 ilustram casos em que, utilizando a mesma inicialização, algoritmos



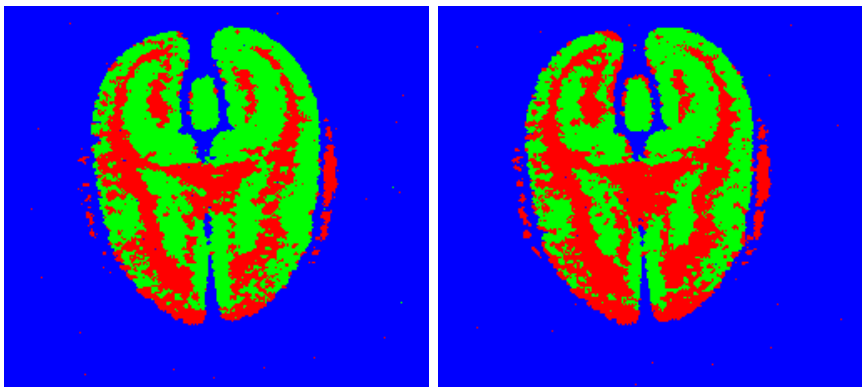
(a) LOGLC+GSA (b) PARZENC+GSA

Figura 34 - Mapas de classes para classificação contextual da imagem goiaba utilizando sistemas de vizinhança de segunda ordem.



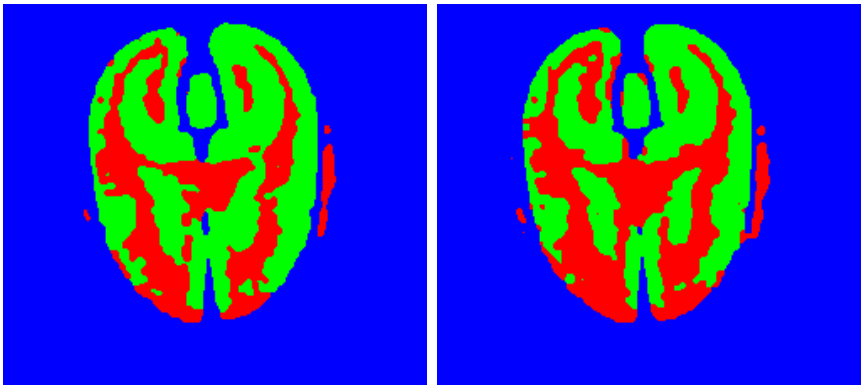
(a) LOGLC+GSA (b) PARZENC+GSA

Figura 35 - Mapas de classes para classificação contextual da imagem goiaba utilizando sistemas de vizinhança de terceira ordem.



(a) LOGLC+GSA (b) QDC+GSA

Figura 36 - Mapas de classes para classificação contextual da imagem de cérebro utilizando sistemas de vizinhança de segunda ordem.



(a) LOGLC+GSA (b) QDC+GSA

Figura 37 - Mapas de classes para classificação contextual da imagem de cérebro utilizando sistemas de vizinhança de terceira ordem.

de classificação contextual distintos fornecem resultados significativamente diferentes, o que é uma evidência positiva em se tratando de combinação de classificadores. Esse fato mostra que as dinâmicas de interação entre os elementos do campo aleatório para cada algoritmo fazem as soluções iniciais evoluírem de maneira própria através do espaço de estado das soluções, o que é interessante do ponto de vista de combinação dos métodos de otimização combinatória. As Figuras 38, 39 e 40 mostram mapas de classes referentes a esses resultados.

Tabela 12 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (`PARZENC`) em diferentes algoritmos de classificação contextual para a imagem de pêra.

	`PARZENC+MPM`	`PARZENC+ICM`
$\hat{k}$	$0,8506$	$0,9513$
$Var(\hat{k})$	$4,514 \times 10^{-4}$	$1,623 \times 10^{-4}$
IC de 95%	$0,8506 \pm 0,0413$ [0,8093, 0,8919]	$0,9513 \pm 0,0247$ [0,9266, 0,9760]

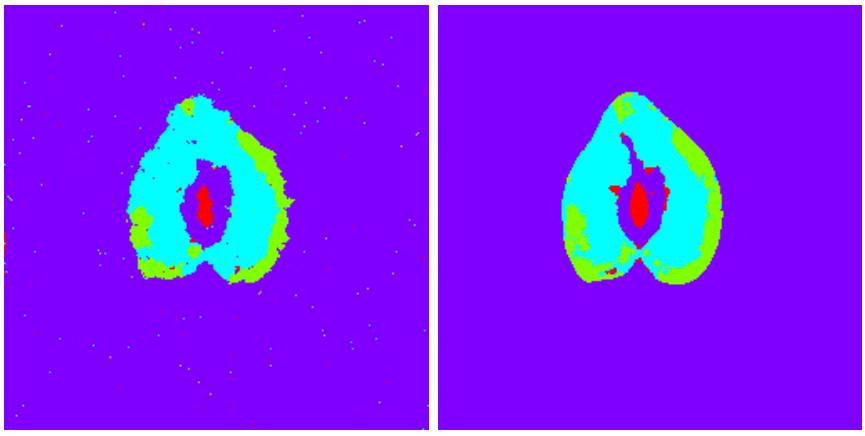
Tabela 13 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (`LDC`) em diferentes algoritmos de classificação contextual para a imagem de pêra.

	`LDC+MPM`	`LDC+ICM`
$\hat{k}$	$0,8333$	$0,9270$
$Var(\hat{k})$	$4,952 \times 10^{-4}$	$2,382 \times 10^{-4}$
IC de 95%	$0,8333 \pm 0,0430$ [0,7903, 0,8763]	$0,9270 \pm 0,0303$ [0,8968, 0,9573]

Tabela 14 - Resultados referentes ao teste de hipótese D: intervalos de confiança de 95% para *Kappas* utilizando a mesma inicialização (NMC) em diferentes algoritmos de classificação contextual para a imagem de cérebro de um *marmocet*.

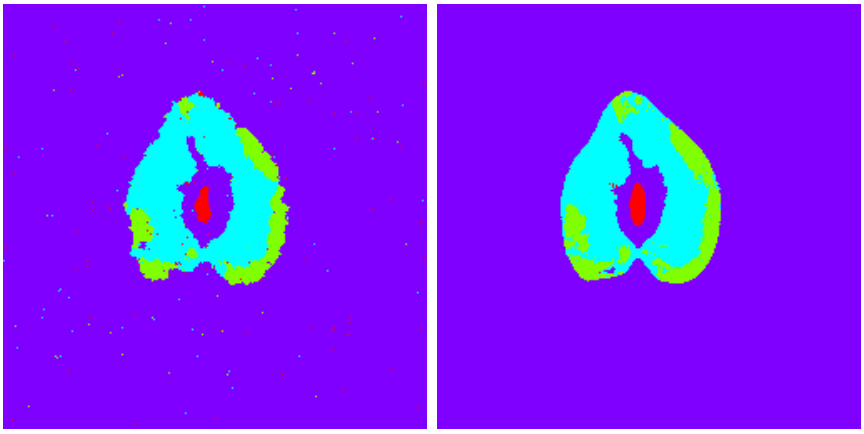
	NMC+MPM	NMC+ICM
$\hat{k}$	$0,9966$	$0,9583$
$Var(\hat{k})$	$5,543 \times 10^{-6}$	$6,744 \times 10^{-5}$
IC de 95%	$0,9966 \pm 0,0046$ [0,9920, 1,0]	$0,9583 \pm 0,0160$ [0,9423, 0,9743]

Além das hipóteses e conjecturas analisadas nos experimentos, foi observada uma constatação interessante nos resultados obtidos: a de que *nem sempre a melhor inicialização fornece o melhor resultado de classificação contextual*. Por exemplo, dados das Tabelas 22 e 23 (maçã pontual x maçã MPM), bem como das Tabelas 26 e 28 (pêra pontual x pêra ICM) ilustram esse comportamento. No caso das Tabelas 22 e 23 o melhor resultado contextual é conseguido com a inicialização TREEC (classificador de árvore de decisão), que é apenas a quarta melhor em termos de *Kappa*. Já no caso das Tabelas 26 e 28, ocorre uma situação ainda mais inusitada: a melhor classificação contextual é obtida também com o classificador TREEC, porém este é a pior de todas as inicializações em termos de coeficiente *Kappa*. Porém, deve-se ter em mente que um alto coeficiente *Kappa* não corresponde necessariamente a um mapa de classes visualmente adequado, visto que ele é calculado apenas em cima do conjunto de teste, não na imagem toda, embora de fato exista grande correlação entre valores de *Kappa* e resultados visuais. Em outras palavras, na maioria das vezes, bons valores de *Kappa* realmente significam bons mapas de classes.



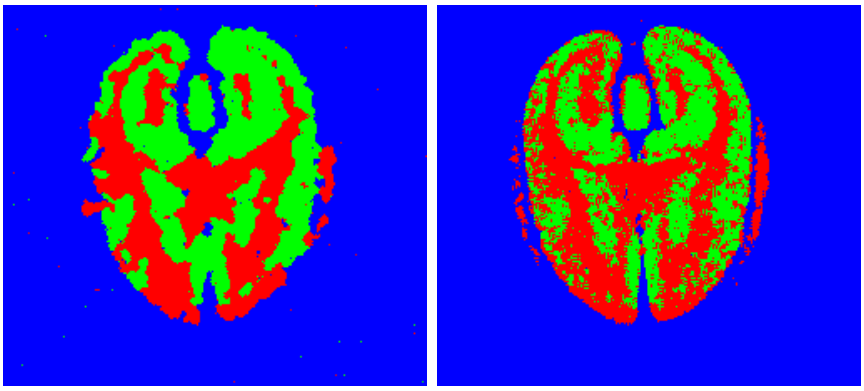
(a) PARZENC+MPM (b) PARZENC+ICM

Figura 38 - Mapas de classes para classificação contextual da imagem pêra utilizando a mesma inicialização (PARZENC) em algoritmos de otimização combinatória distintos.



(a) LDC+MPM (b) LDC+ICM

Figura 39 - Mapas de classes para classificação contextual da imagem pêra utilizando a mesma inicialização (LDC) em algoritmos de otimização combinatória distintos.



(a) NMC+MPM (b) NMC+ICM

Figura 40 - Mapas de classes para classificação contextual da imagem de cérebro de um *marmocet* utilizando a mesma inicialização (NMC) em algoritmos de otimização combinatória distintos.

- **E**: *A utilização de múltiplas inicializações e combinação de algoritmos de otimização combinatória é capaz de melhorar o desempenho da classificação contextual.*

Para verificar essa hipótese, três casos distintos foram considerados:

- E1) Combinação de um único algoritmo de otimização combinatória utilizando múltiplas inicializações.
- E2) Combinação de diferentes algoritmos de otimização combinatória utilizando as mesmas inicializações.
- E3) Combinação de diferentes algoritmos de otimização combinatória utilizando inicializações distintas.

Para verificar a hipótese E1, foram examinados os desempenhos das classificações contextuais da imagem de cérebro utilizando as seis regras de combinação possíveis sobre o algoritmo `ICM` em sistemas de segunda ordem contra os desempenhos da classificação da imagem de cérebro utilizando o algoritmo `ICM` trivialmente, com apenas uma única inicialização, conforme os resultados mostrados nas Tabelas 36 e 39. Comparando o melhor desempenho com uma única inicialização (LDC+ICM) com o melhor desempenho usando múltiplas inicializações (Regra do Máximo), tem-se um $Z = 2,4995$, mostrando que o desempenho é significativamente superior. A probabilidade de significância, ou $P(Z > 2,4995)$, nesse caso é igual a 0,0064, o que fornece fortes evidências a favor da utilização de múltiplas inicializações.

Ainda em relação a hipótese E1, a Tabela 15 sumariza as comparações entre os melhores desempenhos para cada um dos algoritmos iterativos considerando uma única inicialização e o melhor desempenho considerando todas as sete possíveis inicializações. Como pode ser observado, em todos os casos a melhora foi significativa.

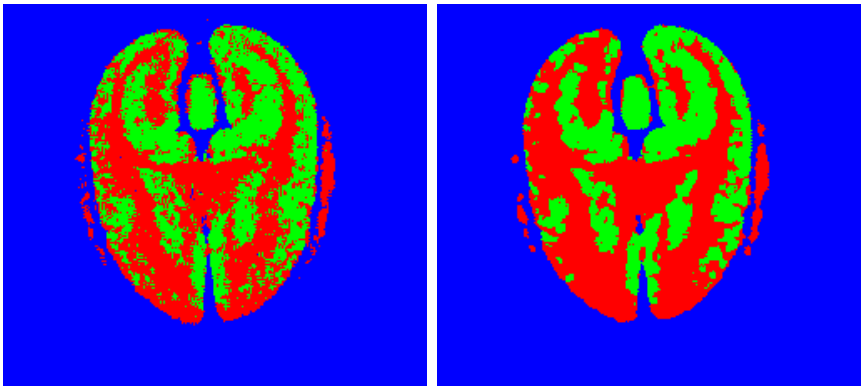
Tabela 15 - Comparação entre os melhores desempenhos individuais e o melhor desempenho obtido com a combinação de todas as inicializações para os algoritmos ICM, GSA e MPM em sistemas de vizinhança de segunda ordens.

Algoritmo de Otimização	**ICM**	**GSA**	**MPM**
	Kappa	Kappa	Kappa
Melhor individual	$0,9617$ (LDC)	$0,9367$ (LDC)	$0,9833$ (LDC)
Melhor combinação	$0,9850$ (*Max*)	$0,9950$ (*Max*)	$0,9983$ (*Sum*)
Estatística Z	$2,4988$ ($> 1,96$)	$5,5745$ ($> 1,96$)	$2,7278$ ($> 1,96$)

As Figuras 41, 42 e 43 comparam os mapas de classes para os melhores desempenhos individuais para cada um dos algoritmos iterativos e os mapas de classes obtidos pela combinação descrita em E1, utilizando todas as sete inicializações simultaneamente (resultados visuais referentes a tabela 15).

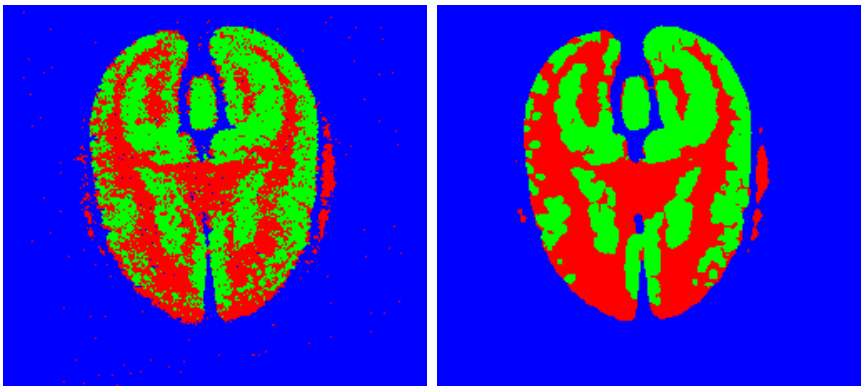
A Tabela 16 mostra uma comparação entre o desempenho médio ao se utilizar uma única inicialização e o desempenho médio obtido quando se considera todas as possíveis inicializações simultâneamente para os algoritmos ICM, GSA e MPM. O *Kappa* médio e o valor de σ^2 que aparecem na Tabela 16 são, respectivamente, as médias amostrais e variâncias calculados a partir das Tabelas 35, 36, 37, 38, 39 e 40. Como pode ser observado, em todos os casos há evidências estatísticas a favor da utilização de múltiplas inicializações, mostrando a viabilidade do método proposto.

O próximo experimento compara o desempenho da classificação contextual usual e o o desempenho obtido pela combinação dos três algoritmos de otimização combinatória utilizando a mesma inicialização. Considerando a condição inicial fornecida pelo classificador pontual `NMC`, e a regra do Máximo, o desempenho da classificação contextual foi dado por um valor de *Kappa* igual a $0,9817$, com uma variância ssintótica de $3,0175 \times 10^{-5}$, o



(a) LDC+ICM (b) Múltiplas inicializações (7)

Figura 41 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo ICM com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15).

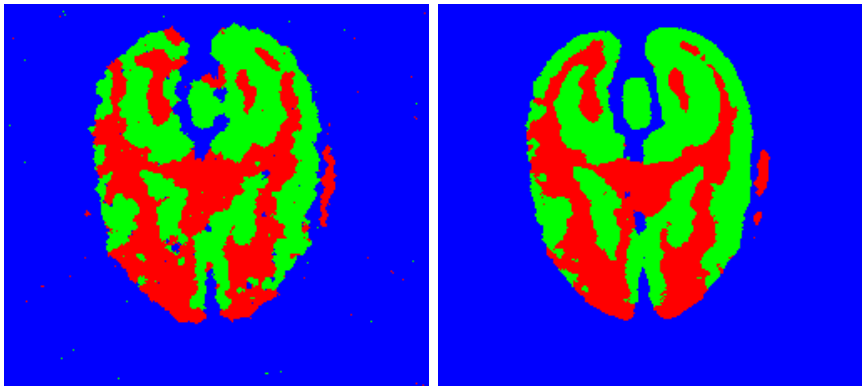


(a) LDC+GSA (b) Múltiplas inicializações (7)

Figura 42 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo GSA com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15).

que resulta em $Z = 5,3085$ ($> 1,96$), se comparado com o melhor desempenho indivídual dentro do esquema (`NMC+ICM`). Similarmente, ao se considerar a inicialização gerada pelo classificador $KNNC$ e a regra do Máximo, se obtém $k = 0,9750$, com uma variância de $4,0960 \times 10^{-5}$, gerando um $Z = 5,1062$ ($> 1,96$) em comparação com o melhor desempenho individual dentro do esquema (`KNNC+MPM`). Esses resultados mostram que, em certas situações, a combinação de diferentes algoritmos de otimização combinatória pode melhorar significativamente o desempenho da classificação contextual de imagens. A Figura 44 mostra o mapa de classes para o melhor resultado individual do esquema e o mapa de classes obtido pela combinação descrita em E2.

O experimento a seguir ilustra a combinação de diferentes algoritmos de otimização combinatória utilizando inicializações distintas. Considerando a inicialização fornecida



(a) LDC+MPM (b) Múltiplas inicializações (7)

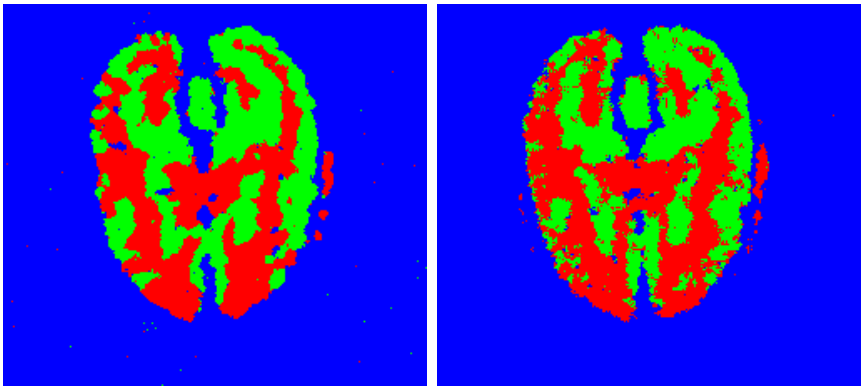
Figura 43 - Mapas de classes para o melhor desempenho individual para a classificação contextual utilizando o algoritmo MPM com uma única inicialização e para o melhor desempenho utilizando combinação de múltiplas inicializações (resultados da tabela 15).

Tabela 16 - Comparação entre os desempenhos médios ao se utilizar uma única inicialização e múltiplas inicializações para os algoritmos ICM, GSA e MPM.

Algoritmo Iterativo	**Metodologia**	**Kappa médio**
ICM	Única inicialização	$0,9021$ ($\sigma^2 = 0,0025$)
	Múltiplas inicializações	$0,9767$ ($\sigma^2 = 5,5779 \times 10^{-5}$)
	Graus de liberdade	6 ($t_c = 1,943$)
	Estatística T	$3,8971 > t_c$
GSA	Única inicialização	$0,8922$ ($\sigma^2 = 0,0026$)
	Múltiplas inicializações	$0,9601$ ($\sigma^2 = 1,8239 \times 10^{-4}$)
	Graus de liberdade	7 ($t_c = 1,894$)
	Estatística T	$3,3873 > t_c$
MPM	Única inicialização	$0,9269$ ($\sigma^2 = 0,0042$)
	Múltiplas inicializações	$0,9897$ ($\sigma^2 = 7,9194 \times 10^{-5}$)
	Graus de liberdade	6 ($t_c = 1,943$)
	Estatística T	$2,5361 > t_c$

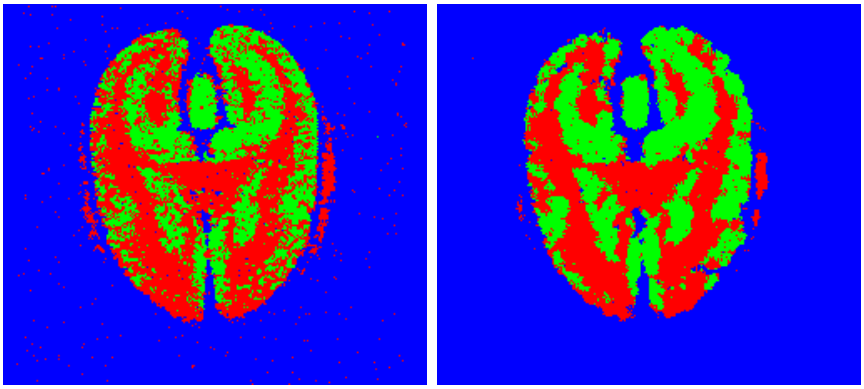
pelo classificador `PARZENC` para o algoritmo `ICM`, a inicialização `QDC` para o algoritmo `GSA` e a inicialização `NMC` para o algoritmo `MPM`, se obtém um *Kappa* de $0,9817$ com uma variância de $3,0180 \times 10^{-5}$, levando a um $Z = 4,2484$ ($> 1,96$), em comparação com o melhor desempenho individual no esquema (`QDC+GSA`). A Figura 45 mostra o mapa de classes para o melhor resultado individual do esquema e o mapa de classes obtido pela combinação descrita em E3.

A seguir, o desempenho da classificação contextual considerando-se múltiplas inicializações em sistemas de segunda e terceira ordens foi comparado para cada um dos algoritmos iterativos. Os resultados mostraram que para todos os casos, houve um aumento significativo no desempenho médio ao se adotar sistemas de ordens superiores. No caso do algoritmo `ICM`, segundo os dados da Tabela 39, foi obtido o valor de $T = -26,8337$, o



(a) KNNC+MPM (b) Combinação

Figura 44 - Mapas de classes para o melhor desempenho individual dentro do esquema proposto (KNNC+MPM) e para o melhor desempenho utilizando a combinação dos três algoritmos iterativos considerando a mesma inicialização (KNNC).

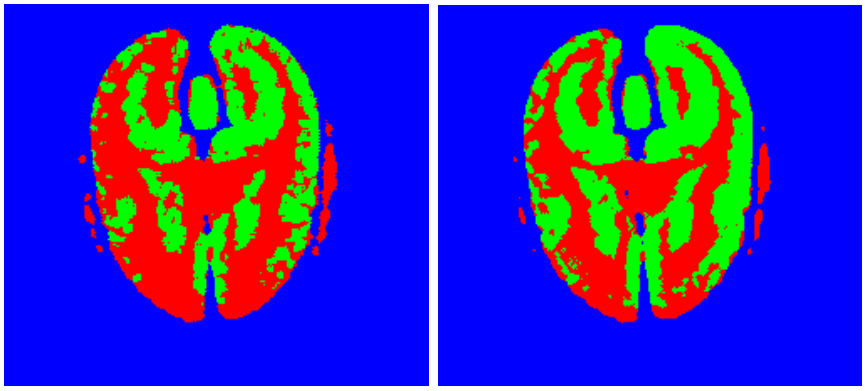


(a) QDC+GSA (b) Combinação

Figura 45 - Mapas de classes para o melhor desempenho individual dentro do esquema proposto (KNNC+MPM) e para o melhor desempenho utilizando a combinação dos três algoritmos iterativos considerando a diferentes inicializações (PARZENC, QDC e NMC).

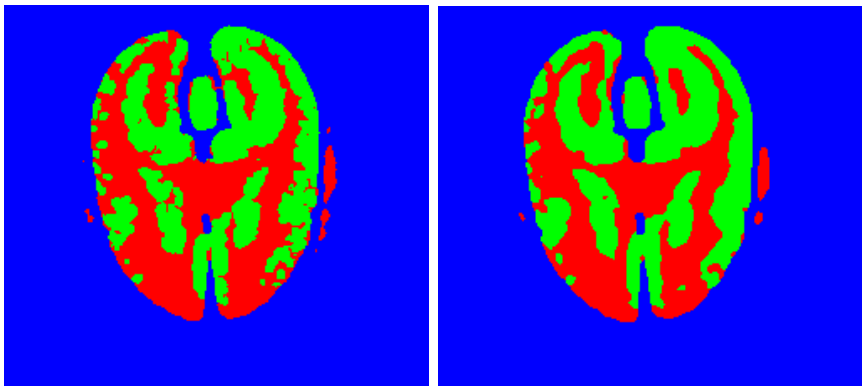
que, para uma distribuição *t-student* com 5 graus de liberdade ($t_c = -2,0105$), demonstra a significância dos resultados obtidos. Para os algoritmos `GSA` e `MPM`, as conclusões foram as mesmas, com $T = -4,6967$ e $T = -2,8305$, respectivamente. As Figuras 46, 47 e 48 ilustram uma comparação entre os melhores resultados obtidos utilizando múltiplas inicializações em sistemas de vizinhança de segunda e terceira ordens. Os resultados mostram claramente e de forma objetiva que a combinação de múltiplas inicializações, aliada a utilização de sistemas de ordens superiores, provê uma melhora significativa no desempenho da classificação supervisionada, mais uma vez assegurando a validade do método proposto.

Por fim, um experimento comparativo foi realizado para medir os tempos de execução de cada um dos algoritmos sub-ótimos variando o número de condições iniciais de um



(a) Segunda ordem (b) Terceira ordem

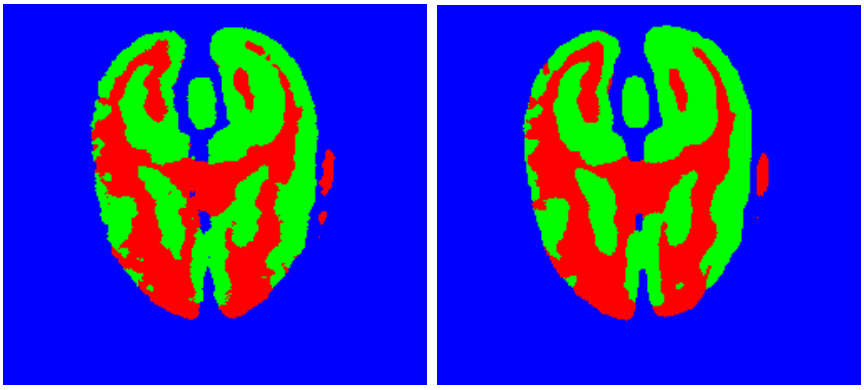
Figura 46 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo ICM em sistemas de vizinhança de segunda e terceira ordens (Regra do Máximo).



(a) Segunda Ordem (b) Terceira Ordem

Figura 47 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo GSA em sistemas de vizinhança de segunda (Regra do Máximo) e terceira ordens (Regra da Soma).

até sete. A Tabela 17 compara tanto o desempenho quanto o tempo gasto na obtenção dos resultados em cada um dos casos. A estratégia para a combinação foi ordenar as condições iniciais em ordem crescente de desempenho, medido pelo coeficiente *Kappa*. Ou seja, a cada etapa, a pior inicialização em termos de coeficiente *Kappa* era adicionada ao esquema de combinação. Por exemplo, de acordo com a Tabela 36 do Apêndice F, para o algoritmo `ICM`, deve-se iniciar com a inicialização `LOGLC` e em seguida adicionar, respectivamente, as inicializações `PARZENC`, `KNNC`, `NMC`, `TREEC`, `QDC` e `LDC`. Os resultados indicam que o algoritmo `GSA`, embora apresente um desempenho levemente inferior ao `MPM` (mas não estatisticamente inferior), no caso de se utilizar todas as inicializações, é o mais eficiente em termos de tempo de processamento.



(a) Segunda Ordem (b) Terceira Ordem

Figura 48 - Mapas de classes comparando a utilização de múltiplas inicializações no algoritmo MPM em sistemas de vizinhança de segunda e terceira ordens (Regra da Soma).

Tabela 17 - Comparação entre o desempenho da classificação e o tempo de execução ao se variar o número de inicializações para os algoritmos de otimização combinatória ICM, GSA e MPM.

Número de Inicializações		1	2	3	4	5	6	7
ICM	*Kappa*	0,8000	0,9867	0,9867	0,9867	0,9850	0,9850	0,9850
	Tempo (s)	16	99	151	215	263	308	370
GSA	*Kappa*	0,7850	0,9450	0,9617	0,9783	0,9800	0,9917	0,9950
	Tempo (s)	15	70	108	143	189	230	298
MPM	*Kappa*	0,8100	0,8833	0,9633	0,9733	0,9733	0,9867	0,9983
	Tempo (s)	503	1020	1557	2123	2717	3344	3991

Capítulo 5

Considerações Finais

Basicamente, nesse trabalho foram discutidos três problemas principais envolvendo a classificação contextual de imagens multiespectrais utilizando a modelagem Bayesiana MAP-MRF. Primeiramente, a derivação de equações para a estimação do parâmetro do modelo Markoviano de *Potts* em sistemas de vizinhanças de ordens superiores. A seguir, a verificação da acurácia da estimação de parâmetros em modelos Markovianos, mais precisamente os modelos de *Potts* e `GMRF`, através da derivação de expressões que aproximam as variâncias assintóticas dos estimadores de máxima pseudo-verossimilhança. Finalmente, foi discutido o problema da classificação contextual de imagens multiespectrais utilizando uma regra de decisão derivada a partir os modelos `GMRF` e *Potts* através da combinação de diferentes algoritmos de otimização combinatória sub-ótimos em sistemas de vizinhança de ordens superiores (segunda e terceira ordens).

Com relação ao problema de inferência estatística referente à estimação dos parâmetros dos modelos Markovianos por máxima pseudo-verossimilhança, a grande contribuição foi o desenvolvimento de novas equações de pseudo-verossimilhança para o modelo de *Potts*, bem como expressões que aproximam a variância assintótica dos estimadores de `MPV` dos modelos de `Potts` e `GMRF`. Desconhecem-se resultados análogos na literatura. Os resultados obtidos mostraram que as equações propostas são válidas, pois produzem resultados estatisticamente equivalentes aos verdadeiros valores dos parâmetros. Além disso, os resultados mostraram que o uso de sistemas de vizinhança de ordens superiores melhora a acurária da estimação, reduzindo tanto a variância assintótica do estimador quanto o erro de estimação.

No que diz respeito a classificação contextual propriamente dita, a maior contribuição foi o desenvolvimento de uma nova abordagem MAP-MRF que utiliza sistemas de vizinhança de ordens superiores, múltiplas inicializações e combinação de algoritmos sub-ótimos de otimização. Evidências estatísticas nos permitem concluir que a metodologia proposta é capaz de melhorar significativamente a performance da classificação contex-

tual. Dessa forma, a principal vantagem do método proposto em relação as abordagens existentes na literatura é a possibilidade de se incorporar uma quantidade muito maior de informação na classificação de uma amostra, tanto pelo fato de se adotar sistemas de vizinhança de ordens superiores quanto pelo fato de ser possível utilizar múltiplas inicializações simultaneamente. Além disso, o método proposto utiliza algoritmos de rápida convergência, com baixo custo computacional e nenhum pouco restritivos em termos da modelagem estatística adotada, ou seja, é viável para uma ampla variedade de modelos.

Por fim, em trabalhos futuros, os autores pretendem estender o modelo proposto para que seja possível sua aplicação na solução de diversos problemas inversos em processamento de imagens. Exemplos de problemas inversos comumente encontrados na prática e cujas soluções são relevantes para diversas áreas do conhecimento são a filltragem e a restauração de imagens. Tendo em vista a freqüente necessidade de se regularizar soluções em problemas inversos, devido a presença de degradações nos dados observados, o modelo MAP-MRF proposto define um podoroso *framework* para esse fim. Dentro desse contexto, pretende-se investigar a filtragem de imagens no domínio wavelet utilizando modelagem contextual. Outro aspecto a ser considerado é a remoção de degradações presentes em imagens (borramentos e ruídos) através de algoritmos iterativos. Também pretende-se investigar a viabilidade de se desenvolver um algoritmo iterativo para realizar filtragem (ou restauração) e classificação contextual de imagens de maneira simultânea, ou seja, estimar tanto as imagens originais (sem ruído e/ou degradações) quanto a verdade terrestre (mapa de classes), a partir do mesmo conjunto de observações degradadas, tudo dentro de um *framework* MAP-MRF integrado.

5.1 Trabalhos Publicados

5.1.1 Artigos em Periódicos e Conferências

O presente projeto de doutorado gerou diversos artigos científicos em conferências e revistas da área. A seguir encontra-se a lista dos principais trabalhos publicados:

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Pseudolikelihood Equations for Potts MRF Model Parameter Estimation on Higher Order Neighborhood Systems. *IEEE Geoscience and Remote Sensing Letters*, v. 5, n. 3, p. 522-526, 2008.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A.; SALVADEO, D. H. P. Spatially Non-Homogeneous Potts Model Parameter Estimation On Higher-Order Neighborhood Systems by Maximum Pseudo-Likelihood. In: Annual ACM Symposium on Applied Computing, ACM SAC, 23., 2008, Fortaleza. *Proceedings...* Fortaleza: ACM, 2008, v.3, p. 1733-1737.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Improving Potts MRF Model Parameter Estimation using Higher-Order Neighborhood Systems on Stochastic Image Modeling. In: International Conference on Systems, Signals and Image Processing, IWSSIP, 15., 2008, Bratislava. *Proceedings...* Bratislava: IEEE, 2008, p. 385-388.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. A Novel Pseudo-Likelihood Equation for Potts MRF Model Parameter Estimation in Image Analysis. In: International Conference on Image Processing, ICIP, 15., 2008, San Diego. *Proceedings...* San Diego: IEEE, 2008, p. 1828-1831.

- *LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Classificação Contextual de Imagens utilizando Campos Aleatórios Markovianos e Teoria dos Jogos. In: Workshop de Visão Computacional, WVC, 4., 2008, Bauru. *Anais...*, 2008, pp. 167-172.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. On the Asymptotic Variances of Gaussian Markov Random Field Model Hyperparameters in Stochastic Image Modeling. In: International Conference on Pattern Recognition, ICPR, 19., 2008, Tampa. *Proceedings...* Tampa, IEEE, 2008, pp. 1-4.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Pseudo-Likelihood Equations for Potts Model on Higher-Order Neighborhood Systems: A Quantitative Approach for Parameter Estimation in Image Analysis. *Brazilian Journal of Probability and Statistics*, v. 23, n.2, pp. 120-140, 2009.

- MARTINS, A. L. D.; LEVADA, A. L. M.; HOMEM, M. R. P.; MASCARENHAS, N. D. A. Super-Resolution Image Reconstruction using maximum pseudo-likelihood parameter estimation. In: International Conference on Image Processing, ICIP, 20., 2009, Cairo. *Proceedings...* Cairo, IEEE, 2009.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. GSAShrink: A Novel Iterative Apporach for Wavelet-Baed Image Denoising. In: *Brazilian Symposium on Computer Graphics and Image Processing*, SIBGRAPI, 22., 2009, Rio de Janeiro. *Proceedings...* Rio de Janeiro, IEEE, 2009, pp. 156-163.

*O trabalho apresentado pelo bolsista, intitulado "Classificação Contextual de Imagens utilizando Campos Aleatórios Markovianos e Teoria dos Jogos" ganhou o prêmio de melhor pôster no WVC2008.

5.1.2 Capítulos de Livros

Além dos trabalhos já publicados citados acima, o aluno conta com a publicação de um capítulo de livro.

- LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Statistical Inference in Markov Random Fields: Parameter Estimation, Asymptotic Evaluation and Contextual Classification of NMR Multispectral Images. In: YIN, P. Y. (Ed.). *Pattern Recognition*. Vukovar/HR: IN-TECH, 2009. cap. 12, p. 223-248.

Referências

1 WHITTLE, P. On stationary processes on a plane. *Biometrika*, v. 41, n. 1/2, p. 434–449, 1954.

2 ABEND, K.; HARLEY, T. J.; KANAL, L. N. Classification of binary random patterns. *IEEE Transactions on Information Theory*, v. 11, n. 4, p. 538–544, 1965.

3 DOBRUSHIN, R. L. The description of random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability and its Applications*, v. 16, n. 2, p. 131–164, 1968.

4 WOODS, J. W. Two-dimensional discrete markovian fields. *IEEE Transactions on Information Theory*, v. 18, n. 4, p. 232–240, 1972.

5 MORAN, P. A. P. A gaussian markovian process on a square lattice. *Journal of Applied Probability*, v. 10, n. 1, p. 54–62, 1973.

6 BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society - Series B*, v. 36, n. 2, p. 192–236, 1974.

7 HAMMERSLEY, J. M.; CLIFFORD, P. Markov field on finite graphs and lattices. 1971. Disponivel em: <http://www.statslab.cam.ac.uk/∼grg/books/hammfest/hamm–cliff.pdf>. Acesso em: jan. 2010.

8 GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis Machine Intelligence*, v. 6, n. 6, p. 721–741, 1984.

9 BESAG, J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society - Series B*, v. 48, n. 3, p. 259–302, 1986.

10 MARROQUIN, J. L.; MITTER, S. K.; POGGIO, T. A. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Society*, v. 82, n. 397, p. 76–89, 1987.

11 YU, S.; BERTHOD, M. A game strategy approach for image labeling. *Computer Vision and Image Understanding*, v. 61, n. 1, p. 32–37, 1995.

12 BLAKE, A.; ZISSERMAN, A. *Visual Reconstruction*. London, England: MIT Press, 1987.

13 NIKILOVA, M.; IDIER, J.; MOHAMAD-DJAFAR, A. Inversion of large-support ill-posed linear operators using a piecewise gaussian mrf. *IEEE Transactions on Image Processing*, v. 7, n. 4, p. 571–585, 1998.

14 CHOU, P. B.; BROWN, C. M. The theory and practice of bayesian image labeling. *International Journal of Computer Vision*, v. 4, n. 3, p. 185–210, 1990.

15 KIRKPATRICK, S.; GELLAT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *Science*, v. 220, n. 4598, p. 671–680, 1983.

16 DUBES, R.; JAIN, A. Random field models in image analysis. *Journal of Applied Statistics*, v. 16, n. 2, p. 131–164, 1989.

17 YOUNG, T. Y.; CALVERT, T. W. *Classification, estimation and pattern recognition*. New York: Elsevier, 1974.

18 FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. 2. ed. New York: Academic Press, 1990.

19 DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. 2. ed. New York: John Wiley & Sons, 2001. 654 p.

20 YANG, M.; AHUJA, N. *Face detection and gesture recognition for human-computer interaction*. New York: Kluwer Academic Publishers, 2001.

21 HYVÄRINEN, A.; KARHUNEN, J.; OJA, E. *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

22 THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 3. ed. New York: Academic Press, 2006. 856 p.

23 KITTLER, J.; HATEF, M.; DUIN, R.; MATAS, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 34, n. 2, p. 299–314, 1998.

24 KUNCHEVA, L. I. *Combining Pattern Classifiers: Methods and Algorithms*. New York: Wiley Interscience, 2004.

25 SOLBERG, A. H. S. Flexible nonlinear contextual classification. *Pattern Recognition Letters*, v. 25, n. 13, p. 1501–1508, 2004.

26 WU, J.; CHUNG, A. C. S. A segmentation model using compound Markov random fields based on a boundary model. *IEEE Transactions on Image Processing*, v. 16, n. 1, p. 241–252, 2007.

27 BESAG, J. Statistical analysis of non-lattice data. *The Statistician*, v. 24, n. 3, p. 179–195, 1975.

28 BESAG, J. Efficiency of pseudo-likelihood estimation for simple gaussian fields. *Biometrika*, v. 64, n. 3, p. 616–618, 1977.

29 CHANDLER, D. *Introduction to modern statistical mechanics*. Oxford: Oxford University Press, 1987.

30 PARISI, G. *Statistical field theory*. Redwood City, CA: Addison-Wesley, 1988.

31 DERIN, H.; ELLIOTT, H. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 9, n. 1, p. 39–55, 1987.

32 SMITH, A. F. M.; ROBERT, G. O. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society, Series B*, v. 55, n. 1, p. 3–23, 1993.

33 BESAG, J.; GREEN, P. J. Spatial statistics and beyesian computation. *Journal of the Royal Statistical Society, Series B*, v. 55, n. 1, p. 25–37, 1993.

34 GILKS, W. R.; CLAYTON, D. G.; SPIEGELHALTER, D. J.; BEST, N. G.; MCNEIL, A. J.; SHARPLES, L. D.; KIRBY, A. J. Modeling complexity: Applications of gibbs sampling in medicine. *Journal of the Royal Statistical Society, Series B*, v. 55, n. 1, p. 39–52, 1993.

35 BESAG, J.; GREEN, P.; HIGDON, D.; MENGERSEN, K. Bayesian computation and stochastic systems. *Statistical Science*, v. 10, n. 1, p. 3–41, 1995.

36 LI, S. Z. *Markov Random Field Modeling in Image Analysis*. 3. ed. London: Springer-Verlag, Inc., 2009.

37 ZHANG, J. The mean field theory in em procedures for markov random fields. *IEEE Transactions on Image Processing*, v. 40, n. 10, p. 2570–2583, 1992.

38 ZHANG, J. The mean field theory in em procedures for blind markov random field image restoration. *IEEE Transactions on Image Processing*, v. 2, n. 1, p. 27–40, 1992.

39 GURELLI, M. I.; ONURAL, L. On a parameter estimation method for gibbs-markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 4, n. 16, p. 424–430, 1994.

40 GIMEL'FARB, G. L. Texture modelling by multiple pairwise pixel interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 18, n. 11, p. 1110–1114, 1996.

41 METROPOLIS, N.; ROSENBLUTH, A.; ROSENBLUTH, M.; TELLER, A.; TELLER, E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, v. 21, n. 6, p. 1087–1092, 1953.

42 SWENDSEN, R.; WANG, J. Nonuniversal critical dynamics in monte carlo simulations. *Physical Review Letters*, v. 58, n. 2, p. 86–88, 1987.

43 WOLFF, U. Collective monte carlo updating for spin systems. *Physical Review Letters*, v. 62, n. 4, p. 361–364, 1989.

44 LANDAU, D. P.; BINDER, K. *A guide to monte carlo simulations in statistical physics*. Cambridge: Cambridge University Press, 2000.

45 WINKLER, G. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Berlin: Springer-Verlag, Inc., 2006.

46 LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Pseudolikelihood equations for potts mrf model parameter estimation on higher order neighborhood systems. *IEEE Geoscience and Remote Sensing Letters*, v. 5, n. 3, p. 522–526, 2008.

47 LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. A novel pseudo-likelihood equation for potts mrf model parameter estimation in image analysis. In: International Conference on Image Processing, ICIP, 15., 2008, San Diego. *Proceedings...* San Diego/CA: IEEE, 2008. p. 1828–1831.

48 LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Improving Potts MRF model parameter estimation using higher-order neighborhood systems on stochastic image modeling. In: International Conference on Systems, Signals and Image Processing, IWSSIP, 15., 2008, Bratislava. *Proceedings...* Bratislava, Eslováquia: IEEE, 2008. p. 385–388.

49 BLANCHET, J.; FORBES, F. Triplet markov fields for the classification of complex structure data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 30, n. 6, p. 1055–1067, 2008.

50 CHEN, C.; HO, P. Statistical pattern recognition in remote sensing. *Pattern Recognition*, v. 41, n. 9, p. 2731–2741, 2008.

51 KATO, Z. Segmentation of color images via reversible jump mcmc sampling. *Image and Vision Computing*, v. 26, n. 3, p. 361–371, 2008.

52 SCARPA, G.; GAETANO, R.; HAINDL, M.; ZERUBIA, J. Hierarchical multiple markov chain model for unsupervised texture segmentation. *IEEE Transactions on Image Processing*, v. 18, n. 8, p. 1830–1843, 2009.

53 GAETANO, R.; SCARPA, G.; POGGI, G. Hierarchical texture-based segmentation of multiresolution remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, v. 47, n. 7, p. 2129–2141, 2009.

54 CECH, J.; SÁRA, R. Languages for constrained binary segmentation based on maximum a posteriori probability labeling. *International Journal of Imaging Systems and Technology*, v. 19, n. 2, p. 69–79, 2009.

55 TOLPEKIN, V.; STEIN, A. Quantification of the effects of land-cover-class spectral separability on the accuracy of markov-random-field-based superresolution mapping. *IEEE Transactions on Geoscience and Remote Sensing*, v. 47, n. 9, p. 3283–3297, 2009.

56 TZIKAS, D. G.; LIKAS, A. C.; GALATSANOS, N. P. The variational approximation for bayesian inference. *IEEE Signal Processing Magazine*, v. 25, n. 6, p. 131–146, 2008.

57 WOOLRICH, M. W.; BEHRENS, T. E. Variational bayes inference of spatial mixture models for segmentation. *IEEE Transactions on Medical Imaging*, v. 25, n. 10, p. 1380–1391, 2006.

58 YUAN, C.; NEUBAUER, C. A variational bayesian approach for classification with corrupted inputs. In: International Conference on Computer Vision and Pattern Recognition, CVPR, 7., 2007, Minneapolis. *Proceedings...* Minneapolis/MN: IEEE, 2007. p. 1–8.

59 BESBES, O.; BELHADJ, Z.; BOUJEMAA, N. A variational framework for adaptive satellite images segmentation. In: International Conference on Scale Space and Variational Methods in Computer Vision, SSVM, 2007, Ischia, Italy. *Proceedings...* Heidelberg, Berlin: Springer, 2010. p. 675–686. (Lecture Notes in Computer Science, v. 4485).

60 LI, Z.; LIU, Q.; CHENG, J.; LU, H. A variational inference based approach for image segmentation. In: International Conference on Pattern Recognition, ICPR, 19., 2008, Tampa. *Proceedings...* Tampa/FL: IEEE, 2008. p. 1–4.

61 ZHU, X.; GOLDBERG, A. *Introduction to Semi-Supervised Learning*. Princeton/NJ: Morgan & Claypool, 2009.

62 YIN, X.; CHEN, S.; HU, E.; ZHANG, D. Semi-supervised clustering with metric learning: An adaptive kernel method. *Pattern Recognition*, v. 43, n. 4, p. 1320–1333, 2010.

63 LI, M.; WU, Y.; ZHANG, Q. Sar image segmentation based on mixture context and wavelet hidden-class-label markov random field. *Computers & Mathematics with Applications*, v. 57, n. 6, p. 961–969, 2009.

64 NIKOS, K.; TZIRITAS, G.; PARAGIOS, N. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Computer Vision and Image Understanding*, v. 112, n. 1, p. 14–29, 2008.

65 ZENG, Y.; SAMARAS, D.; CHEN, W.; PENG, Q. Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in n d images. *Computer Vision and Image Understanding*, v. 112, n. 1, p. 81–90, 2008.

66 KOLMOGOROV, V.; ZABIH, R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 2, p. 147–159, 2004.

67 WON, C. S.; GRAY, R. M. *Stochastic Image Processing*. New York: Kluwer Academic/Plenum Publishers, 2004.

68 BERTERO, M.; BOCCACCI, P. *Introduction to Inverse Problems in Imaging*. Bristol, UK: Taylor & Francis, 1998.

69 ASTER, R.; BORCHERS, B.; THURBER, C. *Parameter Estimation and Inverse Problems*. San Diego, CA: Academic Press, 2005.

70 TREES, H. L. V. *Detecton, Estimation and Modulation Theory - Part I*. Bristol, UK: John Wiley & Sons, 1968.

71 MARROQUIN, J. L.; VELASCO, F. A.; RIVERA, R.; NAKAMURA, M. Gauss-markov measure field models for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 4, p. 337–348, 2001.

72 HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, v. 57, n. 1, p. 97–109, 1970.

73 BESAG, J. Discussion: Markov chain for exploring posterior distributions. *Annals of Statistics*, v. 22, n. 4, p. 1734–1741, 1994.

74 MORGAN, B. J. T. *Applied stochastic modeling*. London, England: Arnold, 2000.

75 BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state markov chains. *Annals of Mathematical Statistics*, v. 37, n. 6, p. 1554–1563, 1966.

76 BAUM, L. E.; PETRIE, T.; SOULES, G.; WEISS, N. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chain. *Annals of Mathematical Statistics*, v. 41, n. 1, p. 164–171, 1970.

77 RABINER, L. R. A tutorial in hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, v. 77, n. 2, p. 257–286, 1989.

78 ISING, E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A: Hadrons and Nuclei*, v. 31, n. 1, p. 253–258, 1925.

79 HEISENBERG, W. Zur theorie des ferromagnetismus. *Zeitschrift für Physik A: Hadrons and Nuclei*, v. 49, n. 9-10, p. 619–636, 1928.

80 YAMAZAKI, T.; GINGRAS, D. Image classification using spectral and spatial information based on mrf models. *IEEE Transactions on Image Processing*, v. 4, n. 9, p. 1333–1339, 1995.

81 YAMAZAKI, T.; GINGRAS, D. A contextual classification system for remote sensing using a multivariate gaussian mrf model. In: International Symposium on Circuits and Systems, ISCAS, 3., 1996, Atlanta. *Proceedings...* Atlanta: IEEE, 1996. v. 2, p. 648–651.

82 WU, F. Y. Jones polynomial as a potts model partition function. *Journal of Knot Theory and Ramifications*, v. 1, n. 1, p. 47–57, 1992.

83 GE, M. L.; HU, L.; WANG, Y. Knot theory, partition function and fractals. *Journal of Knot Theory and Ramifications*, v. 5, n. 1, p. 37–54, 1996.

84 ADAMS, C. C. *The Knot Book*. New York: W. H. Freeman, 1994.

85 JIN, X.; ZHANG, F. Jones polynomials and their zeros for a family of links. *Physica A: Statistical and Theoretical Physics*, v. 333, n. 1-4, p. 183–196, 2004.

86 MONTROLL, E. Statistical mechanics of nearest neighbor systems. *Journal of Chemical Physics*, v. 9, n. 9, p. 706–721, 1941.

87 ENTING, I. G.; GUTTMANN, A. J. Susceptibility amplitudes for the three-and four-state potts models. *Physica A: Statistical Mechanics and its Applications*, v. 321, n. 1-2, p. 90–107, 2003.

88 OUCHI, N. B.; GLAZIER, J. A.; RIEU, J. P.; UPADYAYA, A.; SAWADA, Y. Improving the realism of the cellular potts model in simulations of biological cells. *Physica A: Statistical Mechanics and its Applications*, v. 329, n. 3-4, p. 451–458, 2003.

89 MERKS, R. M. H.; GLAZIER, J. A. A cell-centered approach to developmental biology. *Physica A: Statistical Mechanics and its Applications*, v. 352, n. 1, p. 113–130, 2005.

90 NISHII, R. A markov random field-based approach to decision-level fusion for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 41, n. 10, p. 2316–2319, 2003.

91 FARAG, A. A.; MOHAMED, R. M.; EL-BAZ, A. A unified framework for map estimation in remote sensing image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, n. 7, p. 1617–1634, 2005.

92 LIU, Z.; LUO, J.; SHAO, C. Potts model for exaggeration of a simple rumor transmitted by recreant rumormongers. *Physical Review E*, v. 64, n. 4, p. 046134 [9 páginas], 2001.

93 ONSAGER, L. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, v. 65, n. 3-4, p. 117–149, 1944.

94 ISTRAIL, S. Statistical mechanics, three-dimensionality and NP-completeness. I. Universality of intractability for the partition function of the Ising model across non-planar surfaces (extended abstract). In: Annual ACM Symposium on Theory of Computing, STOC, 32., 2000, New York. *Proceedings...* New York: ACM, 2000. p. 87–96.

95 NASH, J. F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, v. 36, n. 1, p. 48–49, 1950.

96 WEIBULL, J. W. *Evolutionary Game Theory*. London: MIT Press, 1997.

97 BERTHOD, M.; KATO, Z.; YU, S.; ZERUBIA, J. Bayesian image classification using markov random fields. *Image and Vision Computing*, v. 14, n. 4, p. 285–295, 1996.

98 ROBERTS, G. O. Markov chain concepts related to sampling algorithms. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (Eds.). *Markov chain monte carlo in practice*. Boca Raton/FL: Chapman & Hal/CRC, 1996. p. 45–57.

99 WAKS, A.; TRETIAK, O.; GREGORIOU, G. Restoration of noisy regions modeled by noncausal markov random fields of unknown parameters. In: International Conference on Pattern Recognition, ICPR, 3., 1990, Atlantic City/NJ. *Proceedings...* Atantlic City/NJ: IEEE, 1990. v. 12, p. 563–566.

100 ZHANG, J.; FIEGUTH, P.; WANG, D. Random Field Models. In: BOVIK, A. (Ed.). *Handbook of Image and Video Processing*. San Diego/CA: Academic Press, 2000. p. 301–312.

101 JENSEN, J. L.; KÜNCH, H. R. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, v. 46, n. 3, p. 475–486, 1994.

102 LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. On the asymptotic variances of gaussian markov random field model hyperparameters in stochastic image modeling. In: International Conference on Pattern Recognition, ICPR, 19., 2008, Tampa. *Proceedings...* Tampa/FL: IEEE, 2008. p. 1–4.

103 LEVADA, A. L. M.; MASCARENHAS, N. D. A.; TANNÚS, A. Pseudo-likelihood equations for potts model on higher-order neighborhood systems: A quantitative approach for parameter estimation in image analysis. *Brazilian Journal of Probability and Statistics*, v. 23, n. 2, p. 120–140, 2009.

104 FRERY, A. C. *Algumas Ferramentas Estatísticas na Síntese, Processamento e Análise de Imagens de Radar de Abertura Sintética*. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brasil, 1993.

105 FRERY, A. C.; CORREIA, A. H.; FREITAS, C. C. Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. *IEEE Transactions on Geoscience and Remote Sensing*, v. 45, n. 10, p. 3098–3109, 2007.

106 BRENT, R. P. *Algorithms for Minimization without Derivatives*. New York: Prentice Hall, 1973.

107 LEHMANN, E. L. *Theory of Point Estimation*. New York: Wiley, 1983.

108 BICKEL, P. J. *Mathematical Statistics*. New York: Holden Day, 1991.

109 CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. New York: Duxbury, 2002.

110 LIANG, G.; YU, B. Maximum pseudo likelihood estimation in network tomography. *IEEE Transactions on Signal Processing*, v. 51, n. 8, p. 2043–2053, 2003.

111 EFRON, B. F.; HINKLEY, D. V. Assessing the accuracy of the ml estimator: observed versus expected fisher information. *Biometrika*, v. 65, n. 3, p. 457–487, 1978.

112 RAO, C. R. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, v. 37, p. 81–89, 1945.

113 CRAMÉR, H. *Mathematical methods of statistics*. Princeton: Princeton University Press, 1946.

114 CHIB, S. Makov Chain Monte Carlo Technology. In: GENTLE, J. E.; HÄRDLE, W.; MORI, Y. (Eds.). *Handbook of Computational Statistics*. Berlin: Springer-Verlag, 2004. p. 72–98.

115 WEBB, A. *Statistical Pattern Recognition*. 2. ed. London: Arnold, 2002.

116 JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 4–37, 2000.

117 ANDERSON, J. A. Separate sample logistic discrimination. *Biometrika*, v. 59, n. 1, p. 19–35, 1972.

118 ANDERSON, J. A. Logistic Discrimination. In: KRISHNAIAH, P. R.; KANAL L. N. (Eds.). *Handbook of Statistics 2*. Amsterdan: North-Holland, 1982. p. 169–191.

119 DAY, N. E.; KERRIDGE, D. F. A general maximum likelihood discriminant. *Biometrics*, v. 23, n. 2, p. 313–323, 1967.

120 SAFAVIAN, S. R.; LANDGREBE, D. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 21, n. 3, p. 660–674, 1991.

121 BREIMAN, L.; FREIDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall/CRC, 1984.

122 COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37–46, 1960.

123 KENDALL, M. A new measure of rank correlation. *Biometrika*, v. 30, n. 1-2, p. 81–93, 1938.

124 FERRI, C.; J., H.-O.; MODROIU, R. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, v. 30, n. 1, p. 27–38, 2009.

125 CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Enviroment*, v. 37, n. 1, p. 35–46, 1991.

126 MA, Z.; REDMOND, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, v. 61, n. 4, p. 439–453, 1995.

127 FOODY, G. M. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, v. 58, n. 10, p. 1459–1460, 1992.

128 CRUVINEL, P. E.; CESAREO, R.; CRESTANA, S.; MASCARENHAS, S. X and γ-rays computerized minitomograph scanner for soil science. *IEEE Transactions on Instrumentation and Measurement*, v. 39, n. 5, p. 745–750, 1990.

129 PONTI, M. P. *Combinação de Múltiplos Classificadores para Identificação de Materiais em Imagens Ruidosas*. Dissertação (Mestrado) — Programa de Pós Graduação em Ciência da Computação, Departamento de Computação, Universidade Federal de São Carlos, 2004.

130 ALONSO, R. L. *Combinação de Classificadores Atuando em Diferentes Atributos de Imagens Multiespectrais de Ressonância Magnética Nuclear de Frutas com Avaliação do Desempenho Pelo Coeficiente Kappa*. Dissertação (Mestrado) — Programa de Pós Graduação em Ciência da Computação, Departamento de Computação, Universidade Federal de São Carlos, 2002.

131 STONE, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, v. 36, n. 2, p. 111–147, 1974.

132 DEVIJVER, P. A.; KITTLER, J. *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall, 1982.

133 SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, v. 2, n. 6, p. 110–114, 1946.

134 WELCH, B. L. The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, v. 34, n. 1-2, p. 28–35, 1947.

APÊNDICE A – Informação de Fisher no modelo de Potts

A.1 Informação de Fisher observada para o modelo de Potts usando a primeira derivada

A partir da equação (3.26), pode-se escrever:

$$\hat{I}^1_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{U_i\left(m_i\right)\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right] - \left[\sum_{\ell=1}^{M} U_i\left(\ell\right) e^{\beta U_i(\ell)}\right]}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]}\right]^2\right\} \tag{A.1}$$

Inserindo o termo $U_i\left(m_i\right)$ dentro do somatório e colocando a exponencial em evidência, tem-se:

$$\hat{I}^1_{obs} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\left[\sum_{\ell=1}^{M}\left(U_i\left(m_i\right) - U_i\left(\ell\right)\right) e^{\beta U_i(\ell)}\right]^2}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\} \tag{A.2}$$

Expandindo o somatório presente no numerador da equação anterior:

$$\begin{aligned}\left\{\sum_{\ell=1}^{M}\Big(U_i\left(m_i\right) - U_i\left(\ell\right)\Big)e^{\beta U_i(\ell)}\right\}^2 = \\ \left\{\Big(U_i\left(m_i\right) - U_i\left(1\right)\Big)e^{\beta U_i(1)} + \cdots + \Big(U_i\left(m_i\right) - U_i\left(M\right)\Big)e^{\beta U_i(M)}\right\}\times \\ \left\{\Big(U_i\left(m_i\right) - U_i\left(1\right)\Big)e^{\beta U_i(1)} + \cdots + \Big(U_i\left(m_i\right) - U_i\left(M\right)\Big)e^{\beta U_i(M)}\right\}\end{aligned} \tag{A.3}$$

Aplicando a propriedade distributiva e agrupando os termos, a expressão anterior torna-se:

$$\Bigg\{\Big(U_i(m_i)-U_i(1)\Big)e^{\beta U_i(1)}\Big[\Big(U_i(m_i)-U_i(1)\Big)e^{\beta U_i(1)}+\cdots+\Big(U_i(m_i)-U_i(M)\Big)e^{\beta U_i(M)}\Big]+$$
$$\Big(U_i(m_i)-U_i(2)\Big)e^{\beta U_i(2)}\Big[\Big(U_i(m_i)-U_i(1)\Big)e^{\beta U_i(1)}+\cdots+\Big(U_i(m_i)-U_i(M)\Big)e^{\beta U_i(M)}\Big]+$$
$$+\cdots+$$
$$\Big(U_i(m_i)-U_i(M)\Big)e^{\beta U_i(M)}\Big[\Big(U_i(m_i)-U_i(1)\Big)e^{\beta U_i(1)}+\cdots+\Big(U_i(m_i)-U_i(M)\Big)e^{\beta U_i(M)}\Big]\Bigg\}$$

Resultando no numerador da equação para a informação de Fisher usando a primeira derivada (3.27):

$$\Bigg\{\Big(U_i(m_i)-U_i(1)\Big)e^{\beta U_i(1)}\Big[\sum_{k=1}^{M}\Big(U_i(m_i)-U_i(k)\Big)e^{\beta U_i(k)}\Big]+$$
$$+\cdots+\Big(U_i(m_i)-U_i(M)\Big)e^{\beta U_i(M)}\Big[\sum_{k=1}^{M}\Big(U_i(m_i)-U_i(k)\Big)e^{\beta U_i(k)}\Big]\Bigg\}=$$
$$=\sum_{\ell=1}^{M}\Bigg\{\Big(U_i(m_i)-U_i(\ell)\Big)e^{\beta U_i(\ell)}\Big[\sum_{k=1}^{M}\Big(U_i(m_i)-U_i(k)\Big)e^{\beta U_i(k)}\Big]\Bigg\}=$$
$$=\sum_{\ell=1}^{M}\Bigg\{\Big[\sum_{k=1}^{M}\Big(U_i(m_i)-U_i(\ell)\Big)\Big(U_i(m_i)-U_i(k)\Big)e^{\beta(U_i(\ell)+U_i(k))}\Big]\Bigg\}$$

A.2 Informação de Fisher observada para o modelo de Potts usando a segunda derivada

Partindo de (3.28), o primeiro termo do numerador pode ser escrito como:

$$\Big[\sum_{\ell=1}^{M}U_i(\ell)^2e^{\beta U_i(\ell)}\Big]\Big[\sum_{\ell=1}^{M}e^{\beta U_i(\ell)}\Big]=$$
$$\sum_{\ell=1}^{M}\Big[U_i(\ell)e^{\beta U_i(\ell)}\Big]^2+\sum_{\ell=1}^{M-1}\Bigg\{\sum_{k=\ell+1}^{M}\Big[\Big(U_i(\ell)^2+U_i(k)^2\Big)e^{\beta(U_i(\ell)+U_i(k))}\Big]\Bigg\}. \quad \text{(A.4)}$$

Expandindo o quadrado no segundo termo do numerador de (3.28), fornece:

$$\left[\sum_{\ell=1}^{M} U_i(\ell)e^{\beta U_i(\ell)}\right]^2 = \sum_{\ell=1}^{M}\left[U_i(\ell)e^{\beta U_i(\ell)}\right]^2 + 2\sum_{\ell=1}^{M-1}\left\{\sum_{k=\ell+1}^{M}\left[U_i(\ell)U_i(k)e^{\beta(U_i(\ell)+U_i(k))}\right]\right\}. \quad \text{(A.5)}$$

Finalmente, o numerador de (3.28) é reduzido a:

$$\sum_{\ell=1}^{M-1}\left\{\sum_{k=\ell+1}^{M}\left[\Big(U_i^2(\ell) - 2U_i(\ell)U_i(k) + U_k^2(k)\Big)e^{\beta(U_i(\ell)+U_i(k))}\right]\right\}, \quad \text{(A.6)}$$

o que leva a expressão final para a informação de Fisher usando a segunda derivada.

APÊNDICE B – Resultados Quantitativos

B.1 Coeficientes Kappa e Tau referentes à imagem de ciência do solo

As Tabelas 18 a 21 mostram os resultados obtidos para a imagem multiespectral de ciência do solo (Figura 20).

Tabela 18 - Resultados da classificação pontual da imagem de ciência do solo

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
PARZENC	$0,8250$	$4,666 \times 10^{-4}$	$0,8298$	$4,417 \times 10^{-4}$
KNNC	$0,7187$	$6,395 \times 10^{-4}$	$0,7246$	$6,449 \times 10^{-4}$
LOGLC	$0,9812$	$5,767 \times 10^{-5}$	$0,9813$	$5,731 \times 10^{-5}$
LDC	$0,9843$	$4,818 \times 10^{-5}$	$0,9844$	$4,794 \times 10^{-5}$
QDC	$0,9750$	$7,647 \times 10^{-5}$	$0,9751$	$7,588 \times 10^{-5}$
NMC	$0,7187$	$6,700 \times 10^{-4}$	$0,7308$	$6,163 \times 10^{-4}$
TREEC	$0,8979$	$3,809 \times 10^{-4}$	$0,8996$	$3,685 \times 10^{-4}$

Tabela 19 - Resultados da classificação MPM para o phantom de ciência do solo

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9562$	$1,316 \times 10^{-4}$	$0,9565$	$1,299 \times 10^{-4}$
	KNNC	$0,7687$	$5,451 \times 10^{-4}$	$0,7699$	$5,771 \times 10^{-4}$
	LOGLC	$0,9687$	$9,504 \times 10^{-5}$	$0,9689$	$9,419 \times 10^{-5}$
	LDC	$0,9750$	$7,648 \times 10^{-5}$	$0,9751$	$7,587 \times 10^{-5}$
	QDC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	NMC	$0,9031$	$2,775 \times 10^{-4}$	$0,9045$	$2,702 \times 10^{-4}$
	TREEC	$0,9714$	$1,138 \times 10^{-4}$	$0,9715$	$1,127 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,9843$	$4,819 \times 10^{-5}$	$0,9844$	$4,794 \times 10^{-5}$
	KNNC	$0,7875$	$5,118 \times 10^{-4}$	$0,7881$	$5,431 \times 10^{-4}$
	LOGLC	$0,9875$	$3,865 \times 10^{-5}$	$0,9875$	$3,849 \times 10^{-5}$
	LDC	$0,9812$	$5,766 \times 10^{-5}$	$0,9813$	$5,733 \times 10^{-5}$
	QDC	$0,9781$	$6,708 \times 10^{-5}$	$0,9782$	$6,665 \times 10^{-5}$
	NMC	$0,9656$	$1,043 \times 10^{-4}$	$0,9658$	$1,032 \times 10^{-4}$
	TREEC	$0,9714$	$1,137 \times 10^{-4}$	$0,9715$	$1,128 \times 10^{-4}$

Tabela 20 - Resultados da classificação ICM para o phantom de ciência do solo

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	KNNC	$0,8031$	$4,810 \times 10^{-4}$	$0,8032$	$5,137 \times 10^{-4}$
	LOGLC	$0,9968$	$9,74 \times 10^{-6}$	$0,9968$	$9,730 \times 10^{-6}$
	LDC	$0,9968$	$9,74 \times 10^{-6}$	$0,9968$	$9,730 \times 10^{-6}$
	QDC	$1,0$	0	$1,0$	0
	NMC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	TREEC	$1,0$	0	$1,0$	0
Terceira Ordem	PARZENC	$1,0$	0	$1,0$	0
	KNNC	$0,8031$	$4,810 \times 10^{-4}$	$0,8032$	$5,137 \times 10^{-4}$
	LOGLC	$1,0$	0	$1,0$	0
	LDC	$0,9968$	$9,74 \times 10^{-6}$	$0,9968$	$9,730 \times 10^{-6}$
	QDC	$1,0$	0	$1,0$	0
	NMC	$0,9906$	$2,906 \times 10^{-5}$	$0,9906$	$2,897 \times 10^{-4}$
	TREEC	$1,0$	0	$1,0$	0

Tabela 21 - Resultados da classificação GSA para o phantom de ciência do solo

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$1,0$	0	$1,0$	0
	KNNC	$0,8156$	$4,600 \times 10^{-4}$	$0,8160$	$4,853 \times 10^{-4}$
	LOGLC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	LDC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	QDC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	NMC	$0,9437$	$1,669 \times 10^{-4}$	$0,9441$	$1,650 \times 10^{-4}$
	TREEC	$1,0$	0	$1,0$	0
Terceira Ordem	PARZENC	$1,0$	0	$1,0$	0
	KNNC	$0,8031$	$4,810 \times 10^{-4}$	$0,8032$	$5,137 \times 10^{-4}$
	LOGLC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	LDC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	QDC	$0,9937$	$1,942 \times 10^{-5}$	$0,9937$	$1,939 \times 10^{-5}$
	NMC	$0,9656$	$1,042 \times 10^{-4}$	$0,9658$	$1,032 \times 10^{-4}$
	TREEC	$1,0$	0	$1,0$	0

APÊNDICE C – Resultados Quantitativos

C.1 Coeficientes Kappa e Tau referentes à imagem RMN de maçã

As Tabelas 22 a 25 mostram os resultados obtidos para a classificação da imagem multiespectral de ressonância magnética nuclear da maçã (Figura 21).

Tabela 22 - Resultados da classificação pontual da imagem RMN de maçã

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
PARZENC	$0,8055$	$5,647 \times 10^{-4}$	$0,8124$	$5,363 \times 10^{-4}$
KNNC	$0,7847$	$5,555 \times 10^{-4}$	$0,7946$	$5,144 \times 10^{-4}$
LOGLC	$0,8020$	$5,758 \times 10^{-4}$	$0,8097$	$5,405 \times 10^{-4}$
LDC	$0,8298$	$4,959 \times 10^{-4}$	$0,8334$	$4,937 \times 10^{-4}$
QDC	$0,8368$	$4,861 \times 10^{-4}$	$0,8413$	$4,701 \times 10^{-4}$
NMC	$0,8263$	$5,070 \times 10^{-4}$	$0,8306$	$4,989 \times 10^{-4}$
TREEC	$0,8125$	$5,594 \times 10^{-4}$	$0,8208$	$5,106 \times 10^{-4}$

Tabela 23 - Resultados da classificação MPM para a imagem RMN de maçã

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9166$	$2,685 \times 10^{-4}$	$0,9179$	$2,629 \times 10^{-4}$
	KNNC	$0,8888$	$3,477 \times 10^{-4}$	$0,8909$	$3,404 \times 10^{-4}$
	LOGLC	$0,7881$	$5,831 \times 10^{-4}$	$0,7926$	$5,929 \times 10^{-4}$
	LDC	$0,8680$	$4,026 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	QDC	$0,7986$	$5,611 \times 10^{-4}$	$0,8027$	$5,695 \times 10^{-4}$
	NMC	$0,7777$	$6,064 \times 10^{-4}$	$0,7827$	$6,144 \times 10^{-4}$
	TREEC	$0,9687$	$1,058 \times 10^{-4}$	$0,9689$	$1,044 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,8159$	$5,283 \times 10^{-4}$	$0,8201$	$5,258 \times 10^{-4}$
	KNNC	$0,9479$	$1,731 \times 10^{-4}$	$0,9485$	$1,699 \times 10^{-4}$
	LOGLC	$0,8680$	$4,027 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	LDC	$0,8020$	$5,5554 \times 10^{-4}$	$0,8061$	$5,612 \times 10^{-4}$
	QDC	$0,8611$	$4,203 \times 10^{-4}$	$0,8638$	$4,150 \times 10^{-4}$
	NMC	$0,8159$	$5,277 \times 10^{-4}$	$0,8201$	$5,263 \times 10^{-4}$
	TREEC	$1,0$	0	$1,0$	0

Tabela 24 - Resultados da classificação ICM para a imagem RMN de maçã

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8715$	$3,938 \times 10^{-4}$	$0,8740$	$3,876 \times 10^{-4}$
	KNNC	$0,9340$	$2,170 \times 10^{-4}$	$0,9350$	$2,113 \times 10^{-4}$
	LOGLC	$0,8645$	$4,117 \times 10^{-4}$	$0,8672$	$4,059 \times 10^{-4}$
	LDC	$0,8611$	$4,204 \times 10^{-4}$	$0,8638$	$4,150 \times 10^{-4}$
	QDC	$0,8680$	$4,028 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	NMC	$0,8506$	$4,462 \times 10^{-4}$	$0,8537$	$4,419 \times 10^{-4}$
	TREEC	$0,9513$	$1,625 \times 10^{-4}$	$0,9519$	$1,588 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,8680$	$4,028 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	KNNC	$0,9479$	$1,730 \times 10^{-4}$	$0,9484$	$1,700 \times 10^{-4}$
	LOGLC	$0,8784$	$3,755 \times 10^{-4}$	$0,8807$	$3,691 \times 10^{-4}$
	LDC	$0,8715$	$3,938 \times 10^{-4}$	$0,8740$	$3,876 \times 10^{-4}$
	QDC	$0,8715$	$3,938 \times 10^{-4}$	$0,8740$	$3,876 \times 10^{-4}$
	NMC	$0,8680$	$4,028 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	TREEC	$1,0$	0	$1,0$	0

Tabela 25 - Resultados da classificação GSA para a imagem RMN de maçã

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8437$	$4,629 \times 10^{-4}$	$0,8469$	$4,596 \times 10^{-4}$
	KNNC	$0,9131$	$2,787 \times 10^{-4}$	$0,9145$	$2,729 \times 10^{-4}$
	LOGLC	$0,8472$	$4,546 \times 10^{-4}$	$0,8503$	$4,508 \times 10^{-4}$
	LDC	$0,8298$	$4,951 \times 10^{-4}$	$0,8333$	$4,945 \times 10^{-4}$
	QDC	$0,8437$	$4,643 \times 10^{-4}$	$0,8471$	$4,582 \times 10^{-4}$
	NMC	$0,8333$	$4,872 \times 10^{-4}$	$0,8367$	$4,859 \times 10^{-4}$
	TREEC	$0,9479$	$1,730 \times 10^{-4}$	$0,9484$	$1,700 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,8541$	$4,377 \times 10^{-4}$	$0,8571$	$4,330 \times 10^{-4}$
	KNNC	$0,9375$	$2,056 \times 10^{-4}$	$0,9382$	$2,016 \times 10^{-4}$
	LOGLC	$0,8680$	$4,028 \times 10^{-4}$	$0,8706$	$3,968 \times 10^{-4}$
	LDC	$0,8576$	$4,291 \times 10^{-4}$	$0,8604$	$4,240 \times 10^{-4}$
	QDC	$0,8576$	$4,291 \times 10^{-4}$	$0,8604$	$4,240 \times 10^{-4}$
	NMC	$0,8576$	$4,291 \times 10^{-4}$	$0,8604$	$4,240 \times 10^{-4}$
	TREEC	$0,9930$	$2,3985 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$

APÊNDICE D – Resultados Quantitativos

D.1 Coeficientes Kappa e Tau referentes à imagem RMN de pêra

As Tabelas 26 a 29 mostram os resultados obtidos para a classificação da imagem multiespectral de ressonância magnética nuclear da fruta pêra (Figura 22).

Tabela 26 - Resultados da classificação pontual da imagem RMN de pêra

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
PARZENC	$0,9305$	$2,278 \times 10^{-4}$	$0,9316$	$2,215 \times 10^{-4}$
KNNC	$0,9305$	$2,280 \times 10^{-4}$	$0,9316$	$2,213 \times 10^{-4}$
LOGLC	$0,9166$	$2,709 \times 10^{-4}$	$0,9183$	$2,606 \times 10^{-4}$
LDC	$0,9097$	$2,904 \times 10^{-4}$	$0,9114$	$2,812 \times 10^{-4}$
QDC	$0,9409$	$1,956 \times 10^{-4}$	$0,9418$	$1,904 \times 10^{-4}$
NMC	$0,8993$	$3,211 \times 10^{-4}$	$0,9014$	$3,095 \times 10^{-4}$
TREEC	$0,8993$	$3,231 \times 10^{-4}$	$0,9017$	$3,076 \times 10^{-4}$

Tabela 27 - Resultados da classificação MPM para a imagem RMN de pêra

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8507$	$4,514 \times 10^{-4}$	$0,8545$	$4,368 \times 10^{-4}$
	KNNC	$0,9201$	$2,594 \times 10^{-4}$	$0,9215$	$2,517 \times 10^{-4}$
	LOGLC	$0,9166$	$2,697 \times 10^{-4}$	$0,9181$	$2,617 \times 10^{-4}$
	LDC	$0,8333$	$4,952 \times 10^{-4}$	$0,8381$	$4,778 \times 10^{-4}$
	QDC	$0,9131$	$2,803 \times 10^{-4}$	$0,9148$	$2,713 \times 10^{-4}$
	NMC	$0,9236$	$2,490 \times 10^{-4}$	$0,9249$	$2,417 \times 10^{-4}$
	TREEC	$0,9756$	$8,284 \times 10^{-5}$	$0,9758$	$8,187 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9270$	$2,384 \times 10^{-4}$	$0,9282$	$2,317 \times 10^{-4}$
	KNNC	$0,9305$	$2,277 \times 10^{-4}$	$0,9316$	$2,216 \times 10^{-4}$
	LOGLC	$0,9340$	$2,165 \times 10^{-4}$	$0,9349$	$2,119 \times 10^{-4}$
	LDC	$0,9062$	$3,014 \times 10^{-4}$	$0,9082$	$2,901 \times 10^{-4}$
	QDC	$0,9131$	$2,800 \times 10^{-4}$	$0,9147$	$2,717 \times 10^{-4}$
	NMC	$0,8194$	$5,289 \times 10^{-4}$	$0,8249$	$5,094 \times 10^{-4}$
	TREEC	$0,9722$	$9,437 \times 10^{-5}$	$0,9724$	$9,321 \times 10^{-5}$

Tabela 28 - Resultados da classificação ICM para a imagem RMN de pêra

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)	
Segunda Ordem	PARZENC	$0,9513$	$1,623 \times 10^{-4}$	$0,9519$	$1,590 \times 10^{-4}$
	KNNC	$0,9583$	$1,399 \times 10^{-4}$	$0,9587$	$1,374 \times 10^{-4}$
	LOGLC	$0,9722$	$9,437 \times 10^{-5}$	$0,9724$	$9,321 \times 10^{-5}$
	LDC	$0,9270$	$2,382 \times 10^{-4}$	$0,9282$	$2,319 \times 10^{-4}$
	QDC	$0,9722$	$9,437 \times 10^{-5}$	$0,9724$	$9,321 \times 10^{-5}$
	NMC	$0,9236$	$2,487 \times 10^{-4}$	$0,9248$	$2,420 \times 10^{-4}$
	TREEC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9375$	$2,062 \times 10^{-4}$	$0,9383$	$2,011 \times 10^{-4}$
	KNNC	$0,9479$	$1,734 \times 10^{-4}$	$0,9485$	$1,696 \times 10^{-4}$
	LOGLC	$0,9722$	$9,437 \times 10^{-5}$	$0,9724$	$9,321 \times 10^{-5}$
	LDC	$0,9166$	$2,697 \times 10^{-4}$	$0,9181$	$2,617 \times 10^{-4}$
	QDC	$0,9652$	$1,172 \times 10^{-4}$	$0,9655$	$1,155 \times 10^{-4}$
	NMC	$0,9097$	$2,903 \times 10^{-4}$	$0,9114$	$2,814 \times 10^{-4}$
	TREEC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$

Tabela 29 - Resultados da classificação GSA para a imagem RMN de pêra

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9375$	$2,062 \times 10^{-4}$	$0,9383$	$2,011 \times 10^{-4}$
	KNNC	$0,9444$	$1,844 \times 10^{-4}$	$0,9451$	$1,802 \times 10^{-4}$
	LOGLC	$0,9618$	$1,286 \times 10^{-4}$	$0,9621$	$1,265 \times 10^{-4}$
	LDC	$0,9166$	$2,697 \times 10^{-4}$	$0,9181$	$2,617 \times 10^{-4}$
	QDC	$0,9444$	$1,844 \times 10^{-4}$	$0,9451$	$1,802 \times 10^{-4}$
	NMC	$0,9097$	$2,903 \times 10^{-4}$	$0,9114$	$2,814 \times 10^{-4}$
	TREEC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9375$	$2,062 \times 10^{-4}$	$0,9383$	$2,011 \times 10^{-4}$
	KNNC	$0,9479$	$1,734 \times 10^{-4}$	$0,9485$	$1,696 \times 10^{-4}$
	LOGLC	$0,9687$	$1,058 \times 10^{-4}$	$0,9689$	$1,044 \times 10^{-4}$
	LDC	$0,9166$	$2,697 \times 10^{-4}$	$0,9181$	$2,617 \times 10^{-4}$
	QDC	$0,9652$	$1,172 \times 10^{-4}$	$0,9655$	$1,155 \times 10^{-4}$
	NMC	$0,9097$	$2,903 \times 10^{-4}$	$0,9114$	$2,814 \times 10^{-4}$
	TREEC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$

APÊNDICE E – Resultados Quantitativos

E.1 Coeficientes Kappa e Tau referentes à imagem RMN de goiaba

As Tabelas 30 a 33 mostram os resultados obtidos para a classificação da imagem multiespectral de ressonância magnética nuclear da goiaba (Figura 23).

Tabela 30 - Resultados da classificação pontual da imagem RMN de goiaba

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
PARZENC	$0,8090$	$5,504 \times 10^{-4}$	$0,8149$	$5,332 \times 10^{-4}$
KNNC	$0,8159$	$5,432 \times 10^{-4}$	$0,8229$	$5,098 \times 10^{-4}$
LOGLC	$0,7048$	$7,824 \times 10^{-4}$	$0,7229$	$7,032 \times 10^{-4}$
LDC	$0,7256$	$7,376 \times 10^{-4}$	$0,7404$	$6,774 \times 10^{-4}$
QDC	$0,8055$	$5,733 \times 10^{-4}$	$0,8140$	$5,272 \times 10^{-4}$
NMC	$0,7361$	$7,017 \times 10^{-4}$	$0,7472$	$6,740 \times 10^{-4}$
TREEC	$0,6736$	$8,569 \times 10^{-4}$	$0,6983$	$7,310 \times 10^{-4}$

Tabela 31 - Resultados da classificação MPM para a imagem RMN de goiaba

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9063$	$1,948 \times 10^{-4}$	$0,9416$	$1,912 \times 10^{-4}$
	KNNC	$0,9722$	$4,771 \times 10^{-5}$	$0,9861$	$4,739 \times 10^{-5}$
	LOGLC	$0,9375$	$1,622 \times 10^{-4}$	$0,9519$	$1,591 \times 10^{-4}$
	LDC	$0,9409$	$1,952 \times 10^{-4}$	$0,9417$	$1,908 \times 10^{-4}$
	QDC	$0,9364$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$
	NMC	$0,8472$	$6,202 \times 10^{-4}$	$0,7859$	$5,966 \times 10^{-4}$
	TREEC	$0,9722$	$1,202 \times 10^{-5}$	$0,9965$	$1,200 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9687$	$1,058 \times 10^{-4}$	$0,9689$	$1,044 \times 10^{-4}$
	KNNC	$0,9548$	$1,512 \times 10^{-4}$	$0,9553$	$1,482 \times 10^{-4}$
	LOGLC	$0,9583$	$1,398 \times 10^{-4}$	$0,9587$	$1,376 \times 10^{-4}$
	LDC	$0,9826$	$5,948 \times 10^{-5}$	$0,9827$	$5,899 \times 10^{-5}$
	QDC	$0,9791$	$7,119 \times 10^{-5}$	$0,9792$	$7,048 \times 10^{-5}$
	NMC	$0,9027$	$3,100 \times 10^{-4}$	$0,9046$	$3,010 \times 10^{-4}$
	TREEC	$0,9965$	$1,202 \times 10^{-5}$	$0,9965$	$1,200 \times 10^{-5}$

Tabela 32 - Resultados da classificação ICM para a imagem RMN de goiaba

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8923$	$3,405 \times 10^{-4}$	$0,8947$	$3,286 \times 10^{-4}$
	KNNC	$0,9305$	$2,281 \times 10^{-4}$	$0,9316$	$2,212 \times 10^{-4}$
	LOGLC	$0,8715$	$4,018 \times 10^{-4}$	$0,8753$	$3,794 \times 10^{-4}$
	LDC	$0,8506$	$4,559 \times 10^{-4}$	$0,8554$	$4,317 \times 10^{-4}$
	QDC	$0,9305$	$2,282 \times 10^{-4}$	$0,9316$	$2,211 \times 10^{-5}$
	NMC	$0,8125$	$5,435 \times 10^{-4}$	$0,8184$	$5,244 \times 10^{-4}$
	TREEC	$0,9965$	$1,202 \times 10^{-5}$	$0,9965$	$1,200 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9375$	$2,062 \times 10^{-4}$	$0,9383$	$2,011 \times 10^{-4}$
	KNNC	$0,9583$	$1,400 \times 10^{-4}$	$0,9587$	$1,373 \times 10^{-4}$
	LOGLC	$1,0$	0	$1,0$	0
	LDC	$0,9618$	$1,287 \times 10^{-4}$	$0,9621$	$1,265 \times 10^{-4}$
	QDC	$0,9409$	$1,956 \times 10^{-4}$	$0,9417$	$1,904 \times 10^{-4}$
	NMC	$0,9166$	$2,702 \times 10^{-4}$	$0,9182$	$2,612 \times 10^{-4}$
	TREEC	$0,9965$	$1,202 \times 10^{-5}$	$0,9965$	$1,200 \times 10^{-5}$

Tabela 33 - Resultados da classificação GSA para a imagem RMN de goiaba

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8784$	$3,782 \times 10^{-4}$	$0,8812$	$3,660 \times 10^{-4}$
	KNNC	$0,9305$	$2,282 \times 10^{-4}$	$0,9317$	$2,211 \times 10^{-4}$
	LOGLC	$0,8472$	$4,661 \times 10^{-4}$	$0,8523$	$4,388 \times 10^{-4}$
	LDC	$0,8541$	$4,440 \times 10^{-4}$	$0,8582$	$4,261 \times 10^{-4}$
	QDC	$0,9131$	$2,813 \times 10^{-4}$	$0,9149$	$2,703 \times 10^{-4}$
	NMC	$0,8194$	$5,258 \times 10^{-4}$	$0,8247$	$5,106 \times 10^{-4}$
	TREEC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,3906 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9375$	$2,062 \times 10^{-4}$	$0,9383$	$2,011 \times 10^{-4}$
	KNNC	$0,9583$	$1,400 \times 10^{-4}$	$0,9587$	$1,373 \times 10^{-4}$
	LOGLC	$0,9930$	$2,398 \times 10^{-5}$	$0,9930$	$2,390 \times 10^{-5}$
	LDC	$0,9270$	$2,381 \times 10^{-4}$	$0,9282$	$2,319 \times 10^{-4}$
	QDC	$0,9479$	$1,735 \times 10^{-4}$	$0,9485$	$1,695 \times 10^{-4}$
	NMC	$0,8888$	$3,503 \times 10^{-4}$	$0,8914$	$3,378 \times 10^{-4}$
	TREEC	$0,9965$	$1,202 \times 10^{-5}$	$0,9965$	$1,200 \times 10^{-5}$

APÊNDICE F – Resultados Quantitativos

F.1 Coeficientes Kappa e Tau referentes à imagem RMN de cérebro

As Tabelas 34 a 37 mostram os resultados obtidos para a classificação da imagem multiespectral de ressonância magnética nuclear do cérebro de um *marmocet* (Figura 24).

Tabela 34 - Resultados da classificação pontual da imagem RMN de cérebro

Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
PARZENC	$0,7816$	$3,106 \times 10^{-4}$	$0,7963$	$2,703 \times 10^{-4}$
KNNC	$0,7550$	$3,410 \times 10^{-4}$	$0,7733$	$2,924 \times 10^{-4}$
LOGLC	$0,7583$	$3,350 \times 10^{-4}$	$0,7754$	$2,917 \times 10^{-4}$
LDC	$0,7716$	$3,217 \times 10^{-4}$	$0,7875$	$2,792 \times 10^{-4}$
QDC	$0,7866$	$3,051 \times 10^{-4}$	$0,8008$	$2,658 \times 10^{-4}$
NMC	$0,7850$	$3,062 \times 10^{-4}$	$0,7991$	$2,678 \times 10^{-4}$
TREEC	$0,6500$	$4,473 \times 10^{-4}$	$0,6865$	$3,586 \times 10^{-4}$

Tabela 35 - Resultados da classificação MPM para a imagem RMN de cérebro

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9700$	$4,8972 \times 10^{-5}$	$0,9703$	$4,8063 \times 10^{-5}$
	KNNC	$0,9050$	$1,4697 \times 10^{-4}$	$0,9075$	$1,4065 \times 10^{-4}$
	LOGLC	$0,8100$	$2,6512 \times 10^{-4}$	$0,8172$	$2,5605 \times 10^{-4}$
	LDC	$0,9833$	$2,7464 \times 10^{-5}$	$0,9834$	$2,7172 \times 10^{-5}$
	QDC	$0,9767$	$3,8273 \times 10^{-5}$	$0,9768$	$3,7707 \times 10^{-5}$
	NMC	$0,8783$	$1,7483 \times 10^{-4}$	$0,8821$	$1,7483 \times 10^{-4}$
	TREEC	$0,9650$	$5,6969 \times 10^{-5}$	$0,9654$	$5,5669 \times 10^{-5}$
Terceira Ordem	PARZENC	$0,9900$	$1,6553 \times 10^{-5}$	$0,9900$	$1,6448 \times 10^{-5}$
	KNNC	$0,9933$	$1,1061 \times 10^{-5}$	$0,9933$	$1,1013 \times 10^{-5}$
	LOGLC	$0,8083$	$2,6660 \times 10^{-4}$	$0,8156$	$2,5790 \times 10^{-4}$
	LDC	$0,9917$	$1,3811 \times 10^{-5}$	$0,9917$	$1,3736 \times 10^{-5}$
	QDC	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$
	NMC	$0,9850$	$2,4746 \times 10^{-5}$	$0,9851$	$2,4509 \times 10^{-5}$
	TREEC	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$

Tabela 36 - Resultados da classificação ICM para a imagem RMN de cérebro

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,8983$	$1,5707 \times 10^{-4}$	$0,9014$	$1,4868 \times 10^{-4}$
	KNNC	$0,9033$	$1,5033 \times 10^{-4}$	$0,9062$	$1,4186 \times 10^{-4}$
	LOGLC	$0,8000$	$2,7614 \times 10^{-4}$	$0,8080$	$2,6632 \times 10^{-4}$
	LDC	$0,9617$	$6,2206 \times 10^{-5}$	$0,9621$	$6,0745 \times 10^{-5}$
	QDC	$0,9250$	$1,1821 \times 10^{-4}$	$0,9267$	$1,1357 \times 10^{-4}$
	NMC	$0,9133$	$1,3585 \times 10^{-4}$	$0,9157$	$1,2880 \times 10^{-4}$
	TREEC	$0,9133$	$1,3585 \times 10^{-4}$	$0,9157$	$1,2880 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,9733$	$4,3614 \times 10^{-5}$	$0,9736$	$4,2929 \times 10^{-5}$
	KNNC	$0,9667$	$5,4302 \times 10^{-5}$	$0,9670$	$5,3153 \times 10^{-4}$
	LOGLC	$0,8617$	$2,0444 \times 10^{-4}$	$0,8661$	$1,9600 \times 10^{-4}$
	LDC	$0,9950$	$8,3053 \times 10^{-6}$	$0,99501$	$8,2782 \times 10^{-6}$
	QDC	$0,9850$	$2,4743 \times 10^{-5}$	$0,9851$	$2,4512 \times 10^{-5}$
	NMC	$0,9817$	$3,0177 \times 10^{-5}$	$0,9818$	$2,9822 \times 10^{-5}$
	TREEC	$0,9850$	$2,4743 \times 10^{-5}$	$0,9851$	$2,4512 \times 10^{-5}$

Tabela 37 - Resultados da classificação GSA para a imagem RMN de cérebro

	Classificador	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	PARZENC	$0,9067$	$1,4537 \times 10^{-4}$	$0,9093$	$1,3772 \times 10^{-4}$
	KNNC	$0,8817$	$1,8093 \times 10^{-4}$	$0,8859$	$1,6888 \times 10^{-4}$
	LOGLC	$0,7850$	$2,9099 \times 10^{-4}$	$0,7936$	$2,8290 \times 10^{-4}$
	LDC	$0,9367$	$1,0833 \times 10^{-4}$	$0,9380$	$9,6992 \times 10^{-5}$
	QDC	$0,9317$	$1,0833 \times 10^{-4}$	$0,9331$	$1,0429 \times 10^{-4}$
	NMC	$0,8967$	$1,5992 \times 10^{-4}$	$0,9000$	$1,5028 \times 10^{-4}$
	TREEC	$0,9067$	$1,4539 \times 10^{-4}$	$0,9093$	$1,3766 \times 10^{-4}$
Terceira Ordem	PARZENC	$0,9833$	$2,7463 \times 10^{-5}$	$0,9834$	$2,7173 \times 10^{-5}$
	KNNC	$0,9567$	$7,0067 \times 10^{-5}$	$0,9573$	$6,8220 \times 10^{-5}$
	LOGLC	$0,8167$	$2,5773 \times 10^{-4}$	$0,8235$	$2,4859 \times 10^{-4}$
	LDC	$0,9917$	$1,3811 \times 10^{-5}$	$0,9917$	$1,3736 \times 10^{-5}$
	QDC	$0,9933$	$1,1061 \times 10^{-5}$	$0,9933$	$1,1013 \times 10^{-5}$
	NMC	$0,9750$	$4,0964 \times 10^{-5}$	$0,9752$	$4,0306 \times 10^{-5}$
	TREEC	$0,9783$	$3,5584 \times 10^{-5}$	$0,9785$	$3,5086 \times 10^{-5}$

APÊNDICE G – Resultados Quantitativos

G.1 Coeficientes Kappa e Tau referentes à combinação de múltiplas inicializações

As Tabelas 38 a 40 mostram os resultados obtidos para a classificação da imagem multiespectral de ressonância magnética nuclear do cérebro de um *marmocet* (Figura 24) utilizando múltiplas inicializações (`PARZENC`, `KNNC`, `LOGLC`, `LDC`, `QDC`, `NMC` e `TREEC`) e diversas regras de combinação de classificadores.

Tabela 38 - Resultados da classificação MPM da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores

	Combinação	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	MÁXIMO	$0,9817$	$3,0169 \times 10^{-5}$	$0,9818$	$2,9830 \times 10^{-5}$
	MÍNIMO	$0,9833$	$2,7463 \times 10^{-5}$	$0,9834$	$2,7173 \times 10^{-5}$
	SOMA	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$
	PRODUTO	$0,9800$	$3,2872 \times 10^{-5}$	$0,9801$	$3,2472 \times 10^{-5}$
	VOTAÇÃO	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$
	MEDIANA	$0,9967$	$5,5431 \times 10^{-6}$	$0,9967$	$5,5310 \times 10^{-6}$
Terceira Ordem	MÁXIMO	$1,0$	0	$1,0$	0
	MÍNIMO	$1,0$	0	$1,0$	0
	SOMA	$1,0$	0	$1,0$	0
	PRODUTO	$1,0$	0	$1,0$	0
	VOTAÇÃO	$1,0$	0	$1,0$	0
	MEDIANA	$1,0$	0	$1,0$	0

Tabela 39 - Resultados da classificação ICM da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores

	Combinação	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	MÁXIMO	$0,9850$	$2,4743 \times 10^{-5}$	$0,9851$	$2,4512 \times 10^{-5}$
	MÍNIMO	$0,9850$	$2,4743 \times 10^{-5}$	$0,9851$	$2,4512 \times 10^{-5}$
	SOMA	$0,9767$	$3,8257 \times 10^{-5}$	$0,9768$	$3,7722 \times 10^{-5}$
	PRODUTO	$0,9767$	$3,8257 \times 10^{-5}$	$0,9768$	$3,7722 \times 10^{-5}$
	VOTAÇÃO	$0,9683$	$5,1597 \times 10^{-5}$	$0,9686$	$5,0657 \times 10^{-5}$
	MEDIANA	$0,9683$	$5,1597 \times 10^{-5}$	$0,9686$	$5,0657 \times 10^{-5}$
Terceira Ordem	MÁXIMO	$1,0$	0	$1,0$	0
	MÍNIMO	$1,0$	0	$1,0$	0
	SOMA	$0,9933$	$1,1061 \times 10^{-5}$	$0,9933$	$1,1013 \times 10^{-5}$
	PRODUTO	$0,9933$	$1,1061 \times 10^{-5}$	$0,9933$	$1,1013 \times 10^{-5}$
	VOTAÇÃO	$0,9867$	$2,2020 \times 10^{-5}$	$0,9867$	$2,1835 \times 10^{-5}$
	MEDIANA	$0,9867$	$2,2020 \times 10^{-5}$	$0,9867$	$2,1835 \times 10^{-5}$

Tabela 40 - Resultados da classificação GSA da imagem RMN do cérebro de um *marmocet* utilizando múltiplas inicializações e regras de combinação de classificadores

	Combinação	Kappa	Variância (Kappa)	Tau	Variância (Tau)
Segunda Ordem	MÁXIMO	$0,9950$	$8,3055 \times 10^{-6}$	$0,9950$	$8,2780 \times 10^{-6}$
	MÍNIMO	$0,9757$	$8,2813 \times 10^{-5}$	$0,9758$	$8,1901 \times 10^{-5}$
	SOMA	$0,9479$	$1,7344 \times 10^{-4}$	$0,9485$	$1,6968 \times 10^{-4}$
	PRODUTO	$0,9618$	$1,2868 \times 10^{-4}$	$0,9621$	$1,2653 \times 10^{-4}$
	VOTAÇÃO	$0,9444$	$1,8445 \times 10^{-4}$	$0,9451$	$1,8024 \times 10^{-4}$
	MEDIANA	$0,9549$	$1,5120 \times 10^{-4}$	$0,9553$	$1,4829 \times 10^{-4}$
Terceira Ordem	MÁXIMO	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$
	MÍNIMO	$0,9983$	$2,7747 \times 10^{-6}$	$0,9983$	$2,7716 \times 10^{-6}$
	SOMA	$1,0$	0	$1,0$	0
	PRODUTO	$1,0$	0	$1,0$	0
	VOTAÇÃO	$1,0$	0	$1,0$	0
	MEDIANA	$1,0$	0	$1,0$	0