

TESE DE DOUTORADO

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM

CIÊNCIA DA COMPUTAÇÃO

**“Explorando Abordagens de Classificação
Contextual para Floresta de Caminhos Ótimos”**

ALUNO: Daniel Osaku

ORIENTADOR: Prof. Dr. João Paulo Papa

CO-ORIENTADOR: Prof. Dr. Alexandre L. M. Levada

**São Carlos
Junho/2016**

CAIXA POSTAL 676
FONE/FAX: (16) 3351-8233
13565-905 - SÃO CARLOS - SP
BRASIL

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**EXPLORANDO ABORDAGENS DE CLASSIFICAÇÃO
CONTEXTUAL PARA FLORESTA DE CAMINHOS
ÓTIMOS**

DANIEL OSAKU

ORIENTADOR: PROF. DR. JOÃO PAULO PAPA

CO-ORIENTADOR: PROF. DR. ALEXANDRE LUÍS MAGALHÃES LEVADA

São Carlos - SP
Junho/2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**EXPLORANDO ABORDAGENS DE CLASSIFICAÇÃO
CONTEXTUAL PARA FLORESTA DE CAMINHOS
ÓTIMOS**

DANIEL OSAKU

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação, área de concentração: Metodologias e Técnicas de Computação

Orientador: Prof. Dr. João Paulo Papa

São Carlos - SP
Junho/2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

Osaku, Daniel
081e Explorando abordagens de classificação contextual
para floresta de caminhos ótimos / Daniel Osaku. --
São Carlos : UFSCar, 2016.
124 p.

Tese (Doutorado) -- Universidade Federal de São
Carlos, 2016.

1. Classificação contextual de imagens. 2.
Sensoriamento remoto. 3. Ressonância magnética. I.
Título.



UNIVERSIDADE FEDERAL DE SAO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Daniel Osaku, realizada em 24/06/2016:

Prof. Dr. Emerson Carlos Pedrino
UFSCar

Prof. Dr. José Hiroki Saito
UFSCar

Prof. Dr. Alexandre Xavier Falcão
UNICAMP

Prof. Dr. Fernando Vieira Paulovich
USP

Prof. Dr. João Paulo Papa
UFSCar

AGRADECIMENTO

À Deus, por me dar a sabedoria necessária e iluminar o meu caminho ao longo dessa jornada.

Ao orientador, prof. João Paulo Papa, e ao co-orientador, Alexandre Luís Magalhães Levada, pessoas quem tenho a enorme admiração, manifesto a mais profunda gratidão pelos inúmeros conselhos, pela dedicação, atenção, paciência e pelo conhecimento a mim transmitido.

À minha família, em especial à minha esposa Luana, pelo amor, incentivo e por muitas vezes sonharem junto comigo.

Ao professor Rodrigo José Pisani, por disponibilizar as imagens de sensoriamento remoto e de radar, e outras informações que foram imprescindíveis para a qualidade dos experimentos.

Aos professores e funcionários do Departamento de Computação da UFSCar, e a todas as pessoas que de alguma forma contribuíram para a realização deste trabalho.

Aos amigos de pós-graduação, Odair Moreira de Souza, Mauro Jesus Pereira, Amandia Oliveira Sá, Rogério Colpani, Silas Fernandes, Luís Augusto Martins, Clayton Pereira e Leandro Ravaró.

Finalmente, agradeço à FAPESP pelo apoio financeiro.

Resumo

Técnicas de reconhecimento de padrões em imagens foram amplamente estudadas e difundidas com o intuito de desenvolver maneiras para melhorar a eficácia dos classificadores de padrões utilizando amostras rotuladas. Contudo, muitas dessas técnicas classificam seguindo a premissa de que as instâncias são independentes e identicamente distribuídas no espaço de características, levando-se em consideração apenas as propriedades locais da imagem e nenhuma informação sobre as correlações entre *pixels* vizinhos é utilizada. O classificador Floresta de Caminhos Ótimos modela as instâncias como sendo os nós de um grafo, sendo que o problema agora é reduzido para um particionamento desse grafo. Embora existam abordagens que levam em consideração o contexto no processo de reconhecimento de padrões, ainda não existe nenhuma versão do classificador Floresta de Caminhos Ótimos nesse sentido. Assim sendo, um dos objetivos principais da presente tese de doutorado é propor uma versão contextual para a técnica Floresta de Caminhos Ótimos, a qual faria uso então de informações contextuais para auxiliar na tarefa de classificação de dados utilizando métodos baseados em Teoria da Informação e Campos Aleatórios Markovianos para tal finalidade. Uma vez que os modelos Markovianos são dependentes de parâmetros e não é possível encontrar o valor ótimo, pois podem assumir infinitas soluções, uma outra principal contribuição deste trabalho é propor uma abordagem para modelar o processo de encontrar tais parâmetros como sendo um problema de otimização, em que a função de aptidão a ser maximizada é a acurácia da técnica Floresta de Caminhos Ótimos sobre um conjunto rotulado. Os resultados obtidos foram melhores para classificação contextual do que para o método de classificação tradicional, bem como também os métodos de otimização aplicados demonstraram ser uma boa alternativa para a definição dos parâmetros dos modelos Markovianos.

Abstract

Pattern recognition techniques have been widely studied and disseminated in order to develop ways to improve the effectiveness of the pattern classifiers using labeled samples. However, such techniques usually work following the premise that the samples are independent and identically distributed in the feature space, taking into account only the local properties of the image and no information about the correlations between neighboring pixels are employed. The Optimum-Path Forest (OPF) classifier models the instances as the nodes of a graph, being the problem now is reduced to a partition of this graph. Although there are approaches that consider the context in the pattern recognition process, there is no such version for Optimum-Path Forest up to date. Thus, one of the main goal of the presented thesis is to propose a contextual version for the OPF classifier, which would employ contextual informations to support the data classification task using methods based on information theory and Markov Random Fields for such purpose. Since the Markov models are parameter-dependent and it is not a straightforward task to find out the optimal values for such parameters because can assume infinite solutions, another contribution of this work is to propose an approach for modeling the process of find out the parameters as a optimization problem, being the fitness function to be maximized is the OPF accuracy over a labeled set. The results obtained by contextual classification were better than traditional classification results, as well as the optimization methods applied seemed to be a good alternative to fine-tune parameters of the Markov models as well.

Lista de Tabelas

6.1	Descrição dos rótulos para cada cobertura de área.	48
6.2	Cores associadas com cada classe de cobertura do solo.	49
6.3	Número de amostras para a cobertura de área de Itatinga.	50
6.4	Número de amostras para a cobertura de área de Duque de Caxias.	50
6.5	Taxa de acurácia média usando diferentes modelos Markovianos com vetor de 5 características. A melhor acurácia de cada imagem está em negrito.	57
6.6	Taxa de acurácia média usando diferentes modelos Markovianos com vetor de 26 características. A melhor acurácia de cada imagem está em negrito.	57
6.7	Parâmetros utilizados para o PSO e HS, sendo tais valores empiricamente escolhidos.	60
6.8	Acurácia média (\bar{x}) e desvio padrão (σ) sobre as imagens empregadas. As melhores acurácias estão em negrito.	62
6.9	Tempo de execução médio sobre as imagens. As técnicas mais rápidas estão em negrito.	63
6.10	Parâmetros utilizados para o PSO, HS, IHS e GHS: Tais parâmetros foram escolhidos empiricamente.	65
6.11	Tempo de execução média para estimar β	68
6.12	Análise estatística considerando os classificadores baseados em OPF sobre a imagem CBERS-2B. Técnicas similares estão em negrito.	69
6.13	Análise estatística considerando os classificadores baseados em SVM sobre a imagem CBERS-2B. Técnicas similares estão em negrito.	69
6.14	Análise estatística com relação aos classificadores baseados em OPF sobre a imagem Landsat 5 TM. Técnicas similares estão em negrito.	74
6.15	Análise estatística com relação aos classificadores baseados em SVM sobre a imagem Landsat 5 TM. Técnicas similares estão em negrito.	74
6.16	Análise estatística com relação aos classificadores baseados em OPF sobre a imagem Ikonos-2 MS. Técnicas similares estão em negrito.	80
6.17	Análise estatística com relação aos classificadores baseados em SVM sobre a imagem Ikonos-2 MS. Técnicas similares estão em negrito.	80

6.18	Análise estatística em relação aos classificadores baseados em OPF sobre a imagem Geoeye. Técnicas similares estão em negrito.	80
6.19	Análise estatística em relação aos classificadores baseados em SVM sobre a imagem Geoeye. Técnicas similares estão em negrito.	82
6.20	Análise estatística em relação aos classificadores baseados em OPF sobre a imagem ALOS-PALSAR. Técnicas similares estão em negrito.	82
6.21	Análise estatística em relação aos classificadores baseados em SVM sobre a imagem ALOS-PALSAR. Técnicas similares estão em negrito.	86
6.22	Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II).	90
6.23	Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II).	90
6.24	Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.	91
6.25	Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.	92
6.26	Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.	92
6.27	Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.	93
6.28	Acurácias médias obtidas pelo OPF-MRF considerando diferentes modelos Markovianos sobre a imagem CBERS-2B com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.	94
6.29	Acurácias médias obtidas pelo OPF-MRF considerando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.	95

6.30	Acurácias médias obtidas pelo SVM-MRF considerando diferentes modelos Markovianos sobre a imagem CBERS-2 B com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.	96
6.31	Acurácias médias obtidas pelo SVM-MRF considerando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.	97
6.32	Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com OPF-MRF sobre a imagem CBERS-2B. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.	97
6.33	Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com SVM-MRF sobre a imagem CBERS-2B. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.	98
6.34	Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com OPF-MRF sobre a imagem Ikonos-2 MS. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.	98
6.35	Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com SVM-MRF sobre a imagem Ikonos-2 MS. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) está em negrito.	101
6.36	Acurácia média sobre as imagens CBERS-2B (I), Geosyde (II), Ikonos-2 MS (III) e Landsat-5 TM (IV) particionadas em 4, 8 e 16 blocos com o parâmetro β otimizado por BF, HS e PSO. Note que o símbolo ‘-’ significa que não houve particionamento da imagem, isto é, naïve OPF-MRF. O melhor resultado para cada imagem está em negrito.	102
6.37	Tempo de execução médio usando 4, 8 e 16 blocos com o parâmetro β otimizado por BF, HS e PSO. Note que o símbolo ‘-’ significa que não houve particionamento da imagem, isto é, naïve OPF-MRF. Os resultados são mostrados no seguinte formato $h:m:s$, onde h , m e s denotam horas, minutos e segundos, respectivamente.	103

Lista de Figuras

3.1	(a) Grafo completo ponderado nas arestas para um determinado conjunto de treinamento e inicialização dos custos dos nós do grafo. (b) MST do grafo completo. (c) Protótipos escolhidos como sendo os elementos adjacentes de classes diferentes na MST (nós circulados).	16
3.2	(a) Floresta de caminhos ótimos resultante para a Figura 3.1a utilizando a função de valor de caminho f_{max} e dois protótipos. Os identificadores (x, y) acima dos nós são, respectivamente, o custo e o rótulo dos mesmos. A seta indica o nó predecessor no caminho ótimo. (b) Uma instância de teste (triângulo) da classe 2 e suas conexões (linhas pontilhadas) com os nós do conjunto de treinamento. (c) O caminho ótimo do protótipo mais fortemente conexo, seu rótulo 2 e o custo de classificação 0.4 são associados a instância de teste. Note que, mesmo a instância de teste estando mais próxima de um nó da classe 1, ela foi classificada como sendo da classe 2. .	17
4.1	Ordem de sistemas de vizinhança de um <i>pixel</i> s : (a) primeira ordem com 4 vizinhos, (b) segunda ordem com 8 vizinhos e (c) terceira ordem com 12 vizinhos.	22
5.1	Fluxograma do OPF-MRF.	29
5.2	Metodologia proposta para encontrar β^*	32
5.3	Metodologia proposta para classificação contextual com pós-processamento.	41
5.4	Metodologia para classificação contextual por divisão da imagem em blocos.	42

6.1	Imagens de satélite utilizadas nos experimentos: cobrindo a área de Itatinga-SP, Brasil por (a) CBERS-2B CCD (20m) sensor (R2G3B4) e (b) Landsat 5 TM (30m) sensor (R4G3B5), e cobrindo a área de Duque de Caxias-RJ, Brasil por (c) Ikonos-2 MS sensor (R4G3B2) e (d) Geoeye sensor (R5G4B3). As imagens CBERS-2B e Landsat 5 TM possuem 526×492 <i>pixels</i> , e as imagens Ikonos-2 MS e Geoeye possuem 258×250 e 268×250 <i>pixels</i> , respectivamente. Note que as imagens Ikonos-2 MS e Geoeye foram obtidas através de um processo de fusão entre imagens correspondentes dos sensores MS (4m) e PAN (1m) usando o método <i>pan-sharpening</i> . A imagem final tem uma resolução espacial de 1m.	45
6.2	Imagem do radar ALOS-PALSAR utilizada nos experimentos com polarização (a) HH e (b) HV. Tal sensor possui capacidade de imageamento com múltipla polarização (HH, VV, HV e VH; H = polarização horizontal, V = polarização vertical, 1ª letra = radiação emitida, 2ª letra = radiação recebida) [111].	46
6.3	Imagens rotuladas utilizadas nos experimentos: (a), (b), (c) e (d) referem-se às imagens apresentadas nas Figuras 6.1a, 6.1b, 6.1c e 6.1d, respectivamente, e (e) refere-se às imagens apresentadas nas Figuras 6.2a e 6.2b. . .	47
6.4	(a) IBSR-01 fatia 58, (b) dados de referência, (c) <i>baseline</i> da classificação (OPF tradicional), (d) classificação com OPF-MRF, (e) Abordagem proposta em [113] e (f) OPF-MRF aplicado usando técnicas de pré-processamento proposta em [113].	51
6.5	Valores globais de κ do Cohen obtida através de validação cruzada para os conjuntos de dados IBSR-01 e IBSR-02.	52
6.6	Valores globais de κ do Cohen obtida através de validação cruzada para os conjuntos de dados IBSR-03 e IBSR-04.	52
6.7	Imagem CBERS-2B classificada considerando o OPF tradicional em (a), e OPF-MRF com os modelos GIMLL-L1 (b), GIMLL-L2 (c), GMRF (d) e de Potts (e). Imagem Ikonos-2 classificada considerando o OPF tradicional OPF em (f), e OPF-MRF com os modelos GIMLL-L1 (g), GIMLL-L2 (h), GMRF (i) e de Potts (j).	55
6.8	Curva de acurácia média obtida ao longo das iterações através de validação cruzada para as imagens (a) CBERS-2B e (b) Ikonos-2.	56
6.9	Imagens CBERS-2B, Ikonos-2 MS e Geoeye classificadas por OPF (<i>baseline</i>) utilizando: vetor de 5 características em (a), (b) e (c); e vetor de 26 características em (d), (e) e (f).	58

6.10	Imagens CBERS-2B, Ikonos-2 MS e Geoeye classificadas por OPF-MRF utilizando vetor de 5 características com os modelos: Potts/Fisher em (a), (b) e (c), respectivamente; GIMLL-L1/Fisher em (d), (e) e (f); e GIMLL-L2 em (g), (h) e (i).	59
6.11	Imagens CBERS-2B, Ikonos-2 MS e Geoeye classificadas por OPF-MRF utilizando vetor de 26 características com os modelos: Potts/Fisher em (a), (b) e (c), respectivamente; GIMLL-L1/Fisher em (d), (e) e (f); e GIMLL-L2 em (g), (h) e (i).	60
6.12	Acurácia média sobre o conjunto de teste para diferentes valores de β usando BF.	62
6.13	Classificação de imagens MRI-T1: (a) IBSR-04 fatia 58, (b) dados de referência, (c) linha de base, (d), (e) e (f) são os resultados obtidos usando os métodos PSO, HS e BF, respectivamente.	64
6.14	Curva de acurácia para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem CBERS-2B.	66
6.15	Curva de acurácia para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem CBERS-2B.	67
6.16	Imagem CBERS-2B classsificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.	70
6.17	Imagem CBERS-2B classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.	71
6.18	Curva de acurácia para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem Landsat 5 TM.	72
6.19	Curva de acurácia para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem Landsat 5 TM.	72
6.20	Imagem Landsat 5 TM classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.	73
6.21	Imagem Landsat 5 TM classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.	75

6.22	Curva de acurácia média para o OPF-MRF usando β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Ikonos-2 MS.	76
6.23	Curva de acurácia média para o SVM-MRF usando β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Ikonos-2 MS.	77
6.24	Imagem Ikonos-2 MS classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.	78
6.25	Imagem Ikonos-2 MS classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.	79
6.26	Curva de acurácia média para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Geoeeye.	81
6.27	Curva de acurácia média para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Geoeeye.	81
6.28	Imagem Geoeeye classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.	83
6.29	Imagem Geoeeye classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.	84
6.30	Curva de acurácia média para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem ALOS-PALSAR.	85
6.31	Curva de acurácia média para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem ALOS-PALSAR.	85
6.32	Imagem ALOS-PALSAR classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.	87
6.33	Imagem ALOS-PALSAR classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.	88

- 6.34 Resultados de Classificação sobre a imagem CBERS2-B obtidos pelo OPF padrão usando as abordagens de extração de características (I) e (II) em (a) e (b), respectivamente; resultados de classificação obtidos pelo OPF-MRF sem a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) em (c) e (d), respectivamente, e resultados de classificação obtidos pelo OPF-MRF com a etapa de pós-processamento usando as abordagens de extração de características (I) (valor da moda igual a 5) e (II) (valor da moda igual a 4) em (e) e (f), respectivamente. Os resultados de classificação considerando ambas abordagens de extração de características para o OPF-MRF foram obtidos através do modelo de Potts otimizado por meio do método BF. 99
- 6.35 Resultados de Classificação sobre a imagem Ikonos-2 MS obtidos pelo OPF padrão usando as abordagens de extração de características (I) e (II) em (a) e (b), respectivamente; resultados de classificação obtidos pelo OPF-MRF sem a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) em (c) e (d), respectivamente, e resultados de classificação obtidos pelo OPF-MRF com a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) (valor da moda igual a 5) em (e) e (f), respectivamente. Os resultados de classificação considerando ambas abordagens de extração de características para o OPF-MRF foram obtidos através do modelo de Potts otimizado por meio do método BF. 100
- 6.36 Imagem CBERS-2B classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos.104
- 6.37 Imagem Landsat-5 TM classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos. 105
- 6.38 Imagem IKONOS-2 MS classificada por: (a) OPF, (b) naïve OPF-MRF, e OPF-MRF baseado em blocos otimizado por BF com (c) 4, (d) 8 and (e) 16 blocos. 106
- 6.39 Imagem Geoeye classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos.107

Nomenclatura

BW Bandwidth

k-NN k-Nearest Neighbors

a-MRF adaptive Markov Random Field

ANN Artificial Neural Networks

BF Brute Force

CRF Conditional Random Fields

DRF Discriminative Random Fields

GD Gradiente Descendente

GHS Global-best Harmony Search

GIMLL Generalized Isotropic Multi-Level Logistic

GMRF Gaussian Markov Random Fields

HMCR Harmony Memory Consideration Rate

HMM Hidden Markov Models

HS Harmony Search

IBSR Internet Brain Segmentation Repository

ICM Iterated Conditional Modes

IHS Improved Harmony Search

LCDFs Local Conditional Density Functions

MAP Maximum a posteriori

MH *Memória Harmônica*

MRF *Markov Random Fields*

MRI *Magnetic Resonance Image*

MST *Minimum Spanning Tree*

NM *Nelder-Mead*

OPF *Optimum-Path Forest*

OPF-MRF *Optimum-Path Forest with Markov Random Fields*

PAR *Pitch Adjustment Rate*

PSO *Particle Swarm Optimization*

RF *Random Forest*

SOM *Self Organizing Maps*

SVM *Support Vector Machines*

SVM-MRF *Support Vector Machines with Markov Random Fields*

Sumário

Resumo	i
Abstract	ii
1 Introdução	2
1.1 Contextualização e Motivação	2
1.2 Objetivos e Contribuições	5
1.3 Organização do Trabalho	6
2 Revisão Bibliográfica	7
2.1 Reconhecimento de Padrões	7
2.2 Métodos de Classificação Tradicionais	8
2.3 Métodos de classificação Contextual	10
2.4 Métodos de Otimização	12
3 Classificação por Floresta de Caminhos Ótimos	15
3.1 Treinamento	18
3.2 Classificação	19
4 Campos Aleatórios de Markov	20
4.1 Fundamentação Teórica	21
4.2 Modelos de Campos Aleatórios Markovianos	22
4.3 Modo Condicional Iterativo (ICM)	24
4.4 Teoria da Informação de Fisher	25
5 Classificação Contextual com Floresta de Caminhos Ótimos	27
5.1 OPF-MRF: Uma Abordagem Contextual Baseada em Floresta de Caminhos Ótimos	27
5.2 Estimando a Dependência Espacial Utilizando Algoritmos Evolutivos	31
5.2.1 Otimização por Enxame de Partículas	33
5.2.2 Busca Harmônica	35

5.3	Estimando a Dependência Espacial com Métodos de Otimização Numéricos	36
5.3.1	Método Nelder-Mead	36
5.3.2	Método do Gradiente Descendente	38
5.4	Aplicando uma etapa de pós-processamento	39
5.5	Classificação contextual com estimativa de parâmetros em uma imagem dividida em blocos	39
6	Simulações e Resultados	44
6.1	Classificação Contextual com OPF	50
6.1.1	Classificação contextual de imagens de ressonância magnética utilizando o modelo de Potts	50
6.1.2	Sob a Influência de Modelos Markovianos para Classificação usando Floresta de Caminhos Ótimos Baseado em Contexto	53
6.1.3	Um modelo Markoviano baseado em teoria da informação	55
6.2	Estimando a Dependência Espacial Utilizando Algoritmos Evolutivos	58
6.2.1	Melhorando a Classificação da Cobertura do Solo Utilizando OPF-MRF	63
6.2.2	Avaliando a Classificação da Cobertura do Solo através de Floresta de Caminhos Ótimos Baseada em Contexto sob a Influência de Modelos Markovianos Otimizados	86
6.2.3	Classificação contextual em uma imagem dividida em blocos	101
7	Conclusões	108
7.1	Considerações finais	108
7.2	Contribuições	110
7.3	Trabalhos Futuros	111
8	Trabalhos Publicados	112
	Bibliografia	114

Capítulo 1

Introdução

1.1 Contextualização e Motivação

O reconhecimento de padrões é uma das ações mais bem realizadas pelo ser humano. Características intrínsecas dos objetos, tais como sua forma, cor e textura são habilmente utilizadas por nós com o intuito de descrevê-los, associá-los e identificá-los. Para tal, os olhos dos animais são posicionados estrategicamente com o intuito de facilitar essa tarefa de reconhecimento, dado que a sobrevivência dos mesmos está intimamente ligada ao seu processo de visão associado à cognição.

Com o passar dos anos, a evolução dos computadores criou necessidades e desafios para toda a comunidade científica. Tarefas que antes eram realizadas manualmente, passaram a ser conduzidas por máquinas e, muitas vezes, totalmente controladas por elas. Na área de inspeção industrial, por exemplo, a utilização de sensores vem sendo bastante difundida em conjunto com técnicas de processamento e análise de imagens. Em Medicina, o diagnóstico por meio de técnicas de visão computacional tem auxiliado e agilizado o trabalho de muitos profissionais dessas áreas. Desta forma, essas e outras aplicações têm contribuído com o avanço na área de visão computacional e, especificamente, aprendizado de máquina e reconhecimento de padrões.

Com o intuito de imitar o processo de aprendizado do ser humano, as técnicas de reconhecimento de padrões têm sido estudadas e difundidas por um grupo cada vez maior de pesquisadores e profissionais. Por exemplo, uma criança aprende apenas observando e percebendo o comportamento dos adultos e, com isso, reconhecem sinais que indiquem se estão seguros ou em perigo. Muitas técnicas de classificação têm sido propostas, que vão desde técnicas mais tradicionais como Redes Neurais Artificiais (*Artificial Neural Networks* - ANN) [1], Máquinas de Vetores de Suporte (*Support Vector Machines* - SVM) [2] e classificadores Bayesianos [3], por exemplo, à variantes das mesmas, como SVM transdutivas [4], as quais tentam melhorar a eficácia de classificação utilizando amostras não

rotuladas.

Um outro conjunto de classificadores modela as amostras como sendo um grafo, onde os nós representam as instâncias e os arcos são definidos por uma relação de adjacência, sendo que o problema agora é reduzido para um particionamento desse grafo. Um dos trabalhos pioneiros nesse contexto foi o de Zahn [5], o qual propôs computar uma Árvore Geradora Mínima (*Minimum Spanning Tree* - MST) no conjunto de dados, sendo as arestas ponderadas pelas distâncias entre os nós. Em seguida, arestas “pesadas”, ou seja, arestas cujo peso era maior do que um determinado limiar, eram removidas dessa árvore, fazendo com que o grafo ficasse então particionado.

Partindo da ideia acima, outras abordagens foram sendo desenvolvidas na literatura nos últimos anos. Dentre elas, pode-se citar o classificador Floresta de Caminhos Ótimos (*Optimum-Path Forest* - OPF) [6, 7], o qual vem sendo bastante utilizado em diversas aplicações. O classificador OPF modela o problema de reconhecimento de padrões como sendo uma competição entre nós protótipos, os quais são previamente escolhidos e disputam as outras instâncias do grafo entre si. Cada protótipo define uma região de influência discreta, a qual é comumente chamada de Árvore de Caminhos Ótimos (*Optimum-Path Tree* - OPT). A ideia do classificador OPF é bastante semelhante a um processo de segmentação de imagens utilizando crescimento de regiões ou a própria Transformada Watershed [8], porém essa segmentação é realizada no espaço de características, não em uma imagem propriamente dita. Tais técnicas de classificação utilizam um conjunto de instâncias para efetuar o treinamento e, posteriormente, classificar as amostras remanescentes. Segundo Monard et al. [9], este treinamento pode ser realizado utilizando o aprendizado não-supervisionado, onde a amostra é analisada para tentar determinar se algumas instâncias podem ser agrupadas de alguma maneira, formando agrupamentos ou *clusters*. Após determinar os agrupamentos, normalmente é necessária uma análise para determinar o que cada agrupamento representa dentro do contexto do problema. Já no aprendizado supervisionado, tem-se o conhecimento dos rótulos de todas as instâncias do conjunto de treinamento, permitindo, assim, o emprego de abordagens mais complexas e, eventualmente, com melhor eficácia do que a versão não-supervisionada.

Contudo, a maior parte das técnicas assumem que as instâncias são independentes e identicamente distribuídas no espaço de características, levando-se em consideração apenas as propriedades locais da imagem e nenhuma informação sobre as correlações entre elas é utilizada. Em outras palavras, os *pixels* adjacentes provavelmente pertencem à mesma classe [10], sendo crucial explorar também a informação contextual espacial/temporal para uma classificação mais eficaz.

Partindo da premissa que “a arte imita a vida”, houve um grande avanço no sentido de pesquisar e desenvolver maneiras com o intuito de melhorar a eficácia dos classificadores de padrões. Como o ser humano identifica, muitas vezes, objetos e ações utilizando tam-

bém o contexto, é natural que as técnicas de reconhecimento de padrões também possam fazer uso dessa informação adicional. Alguns trabalhos têm sido propostos nessa direção utilizando Campos Aleatórios Markovianos (*Markov Random Fields* - MRF), sendo que a ideia consiste em modelar a imagem rotulada como sendo um MRF e então aplicar modelos estatísticos e/ou baseados em Teoria da Informação com o intuito de melhorar a sensibilidade dos classificadores de padrões. Por exemplo, Tarabalka et al. [11] propuseram uma abordagem híbrida composta por Máquinas de Vetores de Suporte e Campos Aleatórios Markovianos para classificação de imagens de sensoriamento remoto chamada SVM-MRF. A ideia era adicionar informação contextual em um segundo estágio, precedida de uma classificação de *pixel* tradicional. Neste segundo passo, uma regularização por MRF é então executada no mapa de rótulos pelo MAP-MRF.

Moser e Serpico [12] também propuseram uma abordagem semelhante para classificação contextual apresentando uma única formulação para ambos SVM e MRF que executa classificação contextual em um simples passo. Wu et al. [13] também propuseram uma versão diferente do classificador SVM-MRF para segmentação de imagens do cérebro. Portanto, a ideia principal por trás desses classificadores contextuais consiste em modelar algum conhecimento *a priori* como um MRF localmente dependente, objetivando manter a suposição de suavidade espacial dos rótulos dos *pixels* [14, 15, 16].

Kittler et al. [17] propuseram uma abordagem composta por um classificador Bayesiano e MRF gaussiano (*Gaussian Markov Random Fields* - GMRF) que combina a informação espacial e espectral para a classificação de *pixels* usando ambas fontes ao mesmo tempo. Além disso, Shashi et al. [18] empregaram um classificador Bayesiano com diferentes modelos para capturar o contexto espacial, e assim fornecer comparações entre autoregressão espacial e dois diferentes modelos Markovianos. Outro trabalho relacionado foi proposto por Zhang et al. [19], os quais utilizaram SVM com um modelo MRF adaptativo (a-MRF) para evitar correção excessiva na fronteira das classes. Abordagens interessantes para lidar com classificação contextual de imagens de sensoriamento remoto usando descritores de características hierárquicas e de multiresolução são descritas em [20, 21]. Já Cai et al. [22] aplicaram MRF em conjunto com redes neurais artificiais para classificação de imagens.

No contexto de classificação da cobertura do solo usando informação contextual, podem ser encontrados alguns trabalhos interessantes. Laha et al. [23], por exemplo, aplicaram a teoria da evidência para incorporar informação contextual para posterior classificação de imagens usando regras nebulosas (*fuzzy*). Ghimire et al. [24] aplicaram florestas aleatórias para classificação contextual de cobertura do solo, e Stuckens et al. [25] empregaram um algoritmo de agrupamento baseado em conectividade para o mesmo propósito. Sarkar et al. [26] também apresentaram uma abordagem híbrida entre MRF e a teoria de Dempster-Shafer para classificação da cobertura do solo. Também é importante mencionar o trabalho seminal de Wharton [27], o qual apresentou uma classificação contextual

da cobertura do solo baseado na distribuição local de rótulos próximos. Por último, mas não menos importante, Mahmoudi et al. [28] evidenciaram que as relações contextuais podem superar alguns desafios em problemas de reconhecimento de áreas urbanas baseadas em imagens de satélite.

Além disso, diferentes modelos Markovianos têm sido utilizados em aplicações com diversas finalidades, sendo que o modelo mais utilizado é o modelo de Potts. Por exemplo, Martins et al. [29] empregaram uma técnica baseada em máxima probabilidade *a posteriori* que considera a distribuição contínua dos *pixels* vizinhos chamada *Generalized Isotropic Multi-Level Logistic* - GIMLL, na qual a contribuição é dada pela norma L1. Já Chellappa et al. [30] empregaram GMRF para classificação de texturas de uma imagem. Manjunath et al. [31] também utilizaram o método GMRF para segmentação não-supervisionada de texturas, dentre outros trabalhos.

1.2 Objetivos e Contribuições

Embora existam ainda muitas outras abordagens que levam em consideração o contexto no processo de reconhecimento de padrões, ou seja, são utilizadas informações a respeito da distribuição dos rótulos das classes das instâncias vizinhas obtidas a partir de um conhecimento *a priori*, ainda não existe nenhuma versão do classificador OPF nesse sentido. Assim sendo, um dos principais objetivos desta tese é propor o OPF-MRF, o qual faria uso então de informações contextuais para auxiliar na tarefa de classificação de dados. Contudo, uma vez que os modelos Markovianos são dependentes de parâmetros e, muitas vezes, não é possível encontrar o valor ótimo em tempo viável (poucos segundos/horas), pois podem assumir infinitas soluções, no presente trabalho também é apresentada outra abordagem propondo modelar o processo de encontrar tais parâmetros como sendo um problema de otimização, em que a função de aptidão a ser maximizada é a acurácia do OPF sobre um conjunto rotulado. Além disso, visando diminuir os artefatos remanescentes na imagem final, também foi proposto um esquema de classificação contextual seguida por uma etapa de pós-processamento, onde os resultados parciais do OPF-MRF são filtrados usando o filtro da moda aplicado sob algumas condições para evitar a correção excessiva em regiões de alta frequência.

Outra contribuição desta tese é o de propor um esquema de classificação contextual baseado em blocos, que tenta explorar os seguintes fatos: (i) embora os computadores tenham evoluído ao longo dos anos, muitas vezes não são capazes de lidar com imagens de alta resolução, (ii) mesmo obtendo resultados satisfatórios utilizando o esquema de otimização para definir o parâmetro dos modelos Markovianos, ainda podem ser considerados inviáveis quando comparados com as versões não-contextuais e (iii) tal parâmetro é definido utilizando toda a imagem, não levando em consideração as particularidades de

cada região da imagem. Finalmente, foi proposto um modelo Markoviano baseado na teoria de informação de Fisher, o qual tende a ser mais discriminante do que seus respectivos modelos Markovianos.

Assumindo que as instâncias vizinhas tendem a pertencer à mesma classe, é possível supor que essas instâncias tendem a ter probabilidades condicionais locais similares e, portanto, podem ser boas características a serem exploradas com o intuito de melhorar o resultado de classificação. Contudo, tal abordagem de classificação contextual poderia apresentar o efeito Hughes [32], o qual indica que a acurácia tende a diminuir se o número de características for aumentado. Assim, a hipótese mais provável é que o resultado irá depender de uma boa estimativa de parâmetros e do modelo Markoviano a ser empregado. Também é esperado que a contribuição da informação contextual seja limitada, pois a contribuição fornecida pelos atributos das instância são mantidos ao longo das iterações, enquanto que a informação contextual tende a chegar a um equilíbrio.

1.3 Organização do Trabalho

A presente tese está organizada como segue. No Capítulo 2 é apresentada a revisão bibliográfica a respeito da abordagem de reconhecimento de padrões tradicional e baseado em contexto. Nos Capítulos 3 e 4 são apresentadas a técnica de classificação baseada em grafos OPF, assim como o método de integração de informação contextual através de MRF, respectivamente. Já no Capítulo 5 é discutida a metodologia de classificação contextual utilizando OPF-MRF, assim como os métodos de otimização empregados para modelar o problema de encontrar valores adequados para os modelos Markovianos como sendo um problema de otimização. Um método de classificação contextual com uma etapa de pós-processamento também é apresentado, bem como o esquema de classificação baseado em blocos e o modelo Markoviano baseado na teoria da informação de Fisher. No Capítulo 6 são apresentados os resultados experimentais obtidos sobre imagens de ressonância magnética, bem como os resultados sobre imagens de sensoriamento remoto. O Capítulo 7 apresenta as conclusões e considerações finais, bem como as perspectivas futuras para a continuidade das pesquisas relacionadas ao projeto. Finalmente, o Capítulo 8 apresenta os trabalhos publicados.

Capítulo 2

Revisão Bibliográfica

2.1 Reconhecimento de Padrões

O interesse na área de reconhecimento de padrões tem-se renovado devido ao aumento da disponibilidade de grandes bases de dados e à demanda por aplicações cada vez mais complexas, sendo não apenas desafiadoras mas também necessitam de maior poder computacional como, por exemplo, mineração de dados, classificação de documentos, recuperação de dados multimídia, biometria, entre outras [33, 34].

Apesar dos avanços, os melhores reconhecedores de padrões na maioria dos casos ainda são humanos, e como os humanos reconhecem padrões ainda não é perfeitamente conhecido. Embora seja uma tarefa enfadonha para os seres humanos, é uma tarefa extremamente difícil para computadores, pois o ser humano é capaz de associar informações de diferentes fontes para determinar um padrão. Contudo, os computadores são capazes de analisar grandes volumes de dados e distinguir pequenas variações, que muitas vezes são imperceptíveis ao ser humano.

Segundo Jain [33], o reconhecimento de padrões pode ser definido como o estudo de como as máquinas podem observar o ambiente, aprender a distinguir padrões de interesse e tomar decisões sobre as categorias de padrões. Watanabe [35] define padrão como o oposto do caos, uma entidade vagamente definida que poderia ser dado um nome. Já Raittz [34] define reconhecimento de padrões como a aplicação de técnicas de raciocínio indutivo sobre um conjunto de dados conhecidos com o objetivo de construir um modelo para que novos dados não conhecidos possam ser classificados de forma conveniente.

Apesar da indução ser o recurso mais utilizado pelo cérebro humano para derivar conhecimento novo, ela deve ser utilizada com cautela, pois se o número de instâncias for insuficiente, ou se as instâncias não forem bem escolhidas, as hipóteses obtidas podem ser de pouco valor. Portanto, as hipóteses geradas através de inferência indutiva podem ou não preservar a verdade [9].

Dado um padrão, seu reconhecimento/classificação consiste de uma das seguintes tarefas [35]: (i) classificação supervisionada, cujo padrão de entrada é identificado como um membro de uma classe predefinida; e (ii) classificação não-supervisionada, cujo padrão é atribuído à uma classe definida com base na similaridade de padrões.

O projeto de um sistema de reconhecimento de padrões essencialmente envolve três aspectos [33]: (1) aquisição dos dados e pré-processamento, (2) representação dos dados e (3) tomada de decisão. Tais aspectos estão relacionados ao processo de aprendizado, o qual mostra-se essencial para que uma máquina possa exibir um comportamento inteligente [36]. Além disso, o domínio do problema influencia a escolha dos sensores, das técnicas de pré-processamento, dos esquemas de representações e dos modelos de tomada de decisão. É comumente aceito que um problema de reconhecimento bem definido e suficientemente detalhado (pequena variação intraclasse e grande variação interclasse) irá conduzir a uma representação compacta de padrões e uma estratégia de tomada de decisão simplificada.

As quatro abordagens mais conhecidas de reconhecimento de padrões são: correspondência por padrão (*template matching*), classificação estatística, sintática ou estrutural e redes neurais. Tais abordagens não são necessariamente independentes, e algumas vezes o mesmo método de reconhecimento de padrões existe com interpretações diferentes [33].

Em uma abordagem estatística, cada padrão é representado em termos de características, ou atributos, e pode ser visto como um ponto no espaço de características. Frequentemente, os padrões são caracterizados por uma grande quantidade de atributos. Contudo, o número de características levadas em consideração no processo de aprendizado pode comprometer o desempenho das técnicas de classificação, uma vez que a maioria dos algoritmos computacionalmente viáveis de Aprendizado de Máquina não trabalham bem na presença de um grande número de características [36]. Para garantir não só melhor precisão na classificação, mas também melhorar a eficiência do processo de treinamento, é importante escolher características relevantes que possam descrever um padrão da melhor forma possível, pois se as características não expressarem as relações inerentes ao processo de classificação, as generalizações produzidas por qualquer algoritmo de aprendizado terão baixo desempenho na classificação de novos casos [37].

2.2 Métodos de Classificação Tradicionais

Métodos de classificação tradicionais, comumente chamados de métodos pontuais, assumem que as instâncias são independentes e identicamente distribuídas, isto é, apenas os dados observáveis da instância são utilizados para determinar sua classificação. Assim, cada instância pode ser classificada sem a influência ou interferência do restante da amostra, ou seja, a informação a respeito das correlações entre os membros da amostra não

é utilizada na classificação. Contudo, tal situação pode conduzir a erros de classificação.

Inúmeras técnicas de classificação supervisionadas foram propostas. Dentre as mais conhecidas, podemos citar redes neurais artificiais, as quais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. Para o seu treinamento, o algoritmo de retropropagação é o mais utilizado. O principal objetivo deste algoritmo é desenvolver uma regra de treinamento com o intuito de minimizar o erro quadrático total entre as saídas desejadas e as saídas reais dos nós em uma camada de saída.

Outro método amplamente utilizado é o SVM, o qual propõe resolver o problema de classificação de padrões assumindo ser possível separar as classes em um espaço de mais alta dimensão. Suponha uma situação na qual os dados não são linearmente separáveis. Tais amostras podem ser dicotomizadas usando curvas ou círculos como superfícies de decisão, porém encontrar tais limiares é uma tarefa custosa. A principal idéia da técnica SVM é pré-processar os dados de tal forma que o problema de encontrar uma função discriminante não linear seja transformado em um problema de encontrar um hiperplano, ou seja, mapear os dados que estão em uma dimensão qualquer para outra maior, tornando os mesmos linearmente separáveis. Isto é realizado definindo um mapeamento que transforma o vetor de entrada em um outro (usualmente maior) vetor. Espera-se que, escolhendo um mapeamento adequado, o novo conjunto de treinamento seja linearmente separável.

Outro método bastante utilizado é o conhecido k -vizinhos mais próximos (k -NN - *k-Nearest Neighbors*) [38] o qual, basicamente, classifica as amostras com base nos rótulos das amostras mais próximas do conjunto de treinamento. Na fase de treinamento, o espaço de atributos é particionado em regiões de acordo com os rótulos das amostras de treinamento. Uma amostra é associada à classe mais freqüente de suas k -vizinhas mais próximas na etapa de classificação. Apesar de ser uma abordagem simples de ser implementada, a acurácia do algoritmo k -NN pode ser severamente degradada pela presença de ruído (*outliers*) ou dados irrelevantes. O custo computacional aumenta com o valor de k , sendo sua escolha um grande problema. Desta forma, a melhor escolha de k depende intrinsecamente do seu conjunto de dados: altos valores de k tendem a amenizar o efeito do ruído, porém criam superfícies de separação que tornam as classes menos distintas. Geralmente, o parâmetro k é escolhido através do método de otimização por validação cruzada. Um caso particular do k -NN é quando a classe de uma amostra do conjunto de teste é dada pelo rótulo de seu vizinho mais próximo. Neste caso, temos o algoritmo de vizinho mais próximo (NN - *Nearest Neighbor*) [39].

2.3 Métodos de classificação Contextual

Particularmente, uma imagem a ser classificada é formada por regiões cujos tamanhos geralmente são maiores do que o tamanho de um *pixel*. Supondo que as instâncias espacialmente próximas possuem uma alta probabilidade de pertencerem à mesma classe, a informação adquirida a partir da vizinhança de uma instância poderia aumentar a capacidade discriminante entre as instâncias da amostra. Com base nessa premissa, diversas abordagens foram propostas com o intuito de melhorar o resultado de classificação utilizando a informação adquirida a respeito das correlações da instância a ser classificada com as instâncias vizinhas. Em Kittler et al. [17], por exemplo, propuseram uma abordagem composta por um classificador Bayesiano e MRF Gaussiano que combina a informação espacial e espectral para a classificação de *pixels* usando ambas fontes ao mesmo tempo. Em outras palavras, são adicionados novos componentes ao vetor de características original obtidos por meio do modelo de campos aleatórios Gaussiano, que extrai as informações espaciais a partir do resultado da classificação tradicional inicial. Posteriormente, a classificação contextual é efetuada utilizando esse novo vetor de características.

Um dos métodos de classificação contextual mais conhecidos é o SVM-MRF proposto por Tarabalka et al. [11], o qual apresentaram uma abordagem híbrida composta por Máquinas de Vetores de Suporte e Campos Aleatórios Markovianos para classificação de imagens de sensoriamento remoto chamada SVM-MRF. A ideia proposta foi efetuar uma classificação não contextual utilizando SVM convencional no primeiro estágio e extrair informação espectral e espacial em um segundo estágio a partir do mapa de rótulos para refinar os resultados de classificação. Neste segundo passo, uma regularização por MRF é executada no mapa de rótulos através da regra de decisão máximo *a posteriori*. Tal regularização é feita executando os seguintes passos: (i) primeiramente escolhe-se aleatoriamente uma possível classe; (ii) posteriormente são calculadas as energias espectral, que é calculada através de uma classificação binária, e espacial, que é obtida pelo cálculo da probabilidade condicional local em um sistema de vizinhança e ponderada pelo parâmetro de dependência espacial; (iii) finalmente a regularização é efetuada se a energia da classe escolhida for menor do que a atual classe da instância, caso contrário a classe da instância ainda pode ser alterada com uma dada probabilidade, que é definida com base nas diferenças entre as energias e no parâmetro de controle que varia ao longo das iterações. Além disso, foram definidas regras nebulosas aplicadas na função de energia a partir da imagem de gradiente para preservar a borda entre as regiões enquanto executa a regularização espacial. Contudo, o operador de detecção de borda pode ser sensível a ruído nas variações locais. Seguindo uma ideia similar, Zhang et al. [19] propuseram uma abordagem utilizando SVM com um modelo MRF adaptativo (a-MRF) para evitar correção excessiva na fronteira das classes, cujo coeficiente de dependência espacial é determinado por um

índice de homogeneidade relativa.

Em Moser e Serpico [12], também foi proposta uma abordagem semelhante para classificação contextual apresentando uma única formulação para ambos SVM e MRF, que executa classificação contextual em um simples passo. Tal abordagem define um relacionamento analítico entre a regra de decisão de minimização de energia dos modelos Markovianos, e a aplicação de um SVM onde o espaço de características é adequadamente transformado pela influência de um *kernel* baseado em MRF. Em Wu et al. [13], também foi proposta uma versão diferente do classificador SVM-MRF para segmentação de imagens do cérebro. Portanto, a ideia principal por trás desses classificadores contextuais consiste em modelar algum conhecimento *a priori* como um MRF localmente dependente, objetivando manter a suposição de suavidade espacial dos rótulos dos *pixels* [14, 15, 16].

Em contraste aos Campos Aleatórios Markovianos, Campos Aleatórios Condicionais (*Conditional Random Fields* - CRF) [40] são modelos discriminativos que essencialmente fornecem uma forma de combinar as vantagens de classificação e modelagem gráfica [41], sendo uma ferramenta efetiva para uma variedade de diferentes tarefas de segmentação e classificação de dados [42]. Também podem ser entendidos como uma extensão do classificador de regressão logística para estruturas gráficas arbitrárias ou como um modelo discriminativo análogo ao generativo de dados estruturados, tais como modelos Markovianos [41].

A habilidade de classificar múltiplas variáveis que dependem uns dos outros é fundamental para muitas aplicações [41] como, por exemplo, classificação de regiões de uma imagem [43], segmentação de cadeias de DNA [44], e extração de sintaxe de texto em linguagem natural [45].

Modelos baseados em CRF têm se destacado pela habilidade de segmentar/rotular diretamente uma dada imagem observada, além de incorporar facilmente funções arbitrárias de características observadas no processo de treinamento [42]. Assim, CRF têm provado serem uma abordagem flexível para classificação contextual em uma ampla variedade de tarefas de classificação. Por exemplo, Albert et al. [46] investigaram a classificação do uso do solo usando CRF para verificação de base de dados geoespaciais. Zheng et al. [47] empregaram CRF como Redes Neurais Recorrentes para segmentação semântica de imagens. Schmidt et al. [48] investigaram a relevância de diferentes características para classificação de pontos do sensor Lidar em áreas costeiras, bem como a interação entre pontos vizinhos, empregando Florestas Aleatórias (*Random Forest* - RF) com CRF. Em Niemeyer et al. [49], foi empregado CRF com um modelo de iteração capaz de modelar estruturas espaciais locais de dados para classificação contextual de dados do sensor Lidar e detecção de construções em áreas urbanas. He et al. [50] propuseram uma abordagem CRF multiescala, onde componentes são extraídos a partir de um campo aleatório utilizando sistemas de vizinhança de diferentes ordens, e as distribuições de probabilidade

condicional dos componentes são combinados. Já Kuan et al. [51] propuseram uma abordagem baseada no conceito do CRF chamada de *Discriminative Random Fields* (DRF) para classificação binária de regiões da imagem, que tem o intuito de aperfeiçoar o CRF propondo o uso de modelos discriminativos locais para capturar as associações das classes nas instâncias individuais, bem como as interações com as instâncias vizinhas. Além disso, o modelo DRF proposto permite interações em ambos dados observáveis e rótulos. Para tal, ao invés de modelar as funções de características como sendo combinações lineares de um conjunto de características predefinidas a partir de uma sequência de texto, como é feito no CRF, a função de características, chamada de função potencial de associação, é modelada como sendo um termo de decisão local que decide a associação de uma dada instância a uma certa classe ignorando seus vizinhos, e a outra função chamada de função potencial de interação é vista como sendo uma função de suavização de dados dependentes. Tal abordagem foi avaliada comparando os resultados de detecção nas imagens de teste utilizando três diferentes métodos aplicando (i) classificador logístico com inferência por máxima probabilidade *a posteriori* (*Maximum a posteriori* - MAP), e (ii) MRF com inferência pelo método ICM e (iii) DRF com inferência pelo método ICM.

2.4 Métodos de Otimização

Embora seja possível extrair informações a partir de modelos probabilísticos baseados em Campos Aleatórios Markovianos, tais modelos requerem que seja definido o parâmetro de dependência espacial, o qual controla a quantidade de informação local no processo de aprendizado. Uma vez que existem poucas expressões fechadas para determinar tal parâmetro, uma alternativa para resolver o problema de encontrar valores de parâmetros adequados seria a utilização de métodos de otimização, que são métodos capazes de encontrar valores pseudo-ótimos dentro do espaço de busca que maximizam a função de densidade de probabilidade em tempo viável.

A busca de solução para resolver problemas de nível elevado de complexidade computacional tem sido um desafio constante para pesquisadores das mais diversas áreas [52]. Em geral, a tarefa é otimizar certas propriedades de um sistema escolhendo os parâmetros pertinentes do mesmo, sendo que o objetivo da otimização de um problema é melhorar a sua eficiência e diminuir seu custo. Portanto, a otimização busca determinar qual é a melhor solução, sem que seja necessário computar todas as possíveis alternativas [53]. Assim, a escolha do método de resolução a ser utilizado depende principalmente da razão entre a qualidade da solução gerada pelo método e o tempo gasto para encontrar essa solução [54].

Quando o custo da função é não-linear e não-diferenciável, abordagens de busca direta podem ser a escolha mais adequada. A ideia central de todo método de busca direta é

a geração de variações do vetor de parâmetros. Uma vez que a variação é gerada, uma decisão deve ser tomada se os novos parâmetros devem ser aceitos ou não, sendo utilizada uma estratégia gulosa para tomar esta decisão na maioria dos métodos de busca direta. Sob a estratégia gulosa, um novo parâmetro é aceito se, e somente se, o custo da função for reduzido. Um dos melhores métodos conhecidos é o método Nelder-Mead, que visa encontrar pontos estacionários em espaços livres multidimensionais. Por exemplo, Luersen et al. [55] aplicaram o método Nelder-Mead na maximização da carga de flambagem.

Outro método proposto por Storn et al. [53], chamado *Differential Evolution*, é um método paralelo de busca direta que utiliza vetores de parâmetros N -dimensionais, cujo vetor inicial é escolhido aleatoriamente e deve cobrir o espaço de parâmetros, e novos vetores de parâmetros são gerados adicionando a diferença ponderada entre dois vetores de parâmetros, operação esta chamada de mutação. Em seguida, tal vetor é mesclado com outro vetor predefinido, operação chamada de cruzamento ou *crossover*. Posteriormente, é executada uma operação chamada Seleção, que seleciona o vetor de parâmetros que produz o menor valor de custo da função.

Porém, segundo Koide [56], tais métodos podem tender para um ponto de ótimo local, dependendo do ponto inicial. Além disso, se o ponto inicial estiver em uma região inviável, tais métodos podem convergir a um ótimo local inviável. Para tal problema, métodos de otimização estocásticos têm se destacado por serem menos sensíveis ao ponto inicial, evitando permanecerem em um ótimo local. Assim, métodos de otimização não determinísticos têm sido propostos, como algoritmos heurísticos, que são algoritmos exploratórios não-exatos, e meta-heurísticos, que buscam solucionar problemas genéricos onde não se tem um algoritmo eficiente.

Os métodos de otimização meta-heurísticos são métodos inteligentes e flexíveis, pois possuem uma estrutura com componentes genéricos que são adaptados ao problema que se quer resolver. Tais métodos possuem uma certa facilidade em incorporar novas situações e exploram o espaço de busca permitindo a escolha estratégica de soluções piores do que as já encontradas na tentativa de superar a otimalidade local [54].

Por exemplo, a técnica de otimização global conhecida como Algoritmo Genético é um algoritmo de busca estocástico proposto por Holland [57], o qual combina o sobrevivente mais apto com operadores genéticos abstraídos a partir de uma seleção natural para formar um mecanismo robusto que é adequado a uma variedade de problemas de otimização. Por exemplo, Ishibuchi et al. [58] utilizaram algoritmos genéticos mono-objetivo e multi-objetivos para a seleção de regras linguísticas para problemas de reconhecimento de padrões. Liu et al. [59] desenvolveram um algoritmo genético com novas características permutativas com o intuito de obterem a sequência de empilhamento de lâminas para resolver o problema da carga máxima de flambagem.

Otimização por Enxame de Partículas (*Particle Swarm Optimization* - PSO) é uma

técnica evolutiva proposta por Kennedy et al. [60] que simula o comportamento de indivíduos em um grupo para maximizar a sobrevivência da espécie, onde cada partícula se move em uma direção baseando-se na sua experiência e de todo o grupo [61, 62]. PSO é similar aos Algoritmos Genéticos no sentido de que o sistema é inicializado com uma população de soluções escolhidas aleatoriamente, mas diferente no fato de que também é atribuída uma velocidade aleatória à cada solução potencial, e então as potenciais soluções, chamadas partículas, se movem pelo espaço de busca [63]. Assim, o PSO tem se tornado um candidato para muitas aplicações devido à sua alta performance e flexibilidade. Por exemplo, Omran et al. [62] propuseram uma abordagem de classificação não-supervisionada de imagens usando PSO. Melgani et al. [64] propuseram a otimização dos parâmetros do classificador SVM utilizando PSO para melhorar a classificação de sinais de eletrocardiograma, e Van der Merwe et al. [65] propuseram abordagens de agrupamento de dados empregando PSO.

Outra técnica de otimização meta-heurística bastante conhecida é chamada de Recozimento Simulado (*Simulated Annealing* - SA), que foi proposta por [66] e simula o processo de recozimento de metais [67, 67]. Neste processo, os metais são aquecidos acima de seu ponto de fusão seguido por um resfriamento lento obtido pela diminuição da temperatura do ambiente em passos, onde a temperatura é mantida constante por um período de tempo a cada passo até que o metal alcance o equilíbrio térmico [68]. SA explora uma vizinhança a cada iteração de acordo com suas estratégias, e escolhe apenas um elemento dessa vizinhança a cada passo. Esse tipo de varredura do espaço de busca gera uma trajetória de soluções obtida pela transição de uma solução para outra de acordo com os movimentos permitidos segundo uma certa probabilidade, evitando que o sistema fique preso em regiões de atração de um mínimo local [54]. Por exemplo, Selim et al. [68] empregaram SA para problemas de agrupamentos de dados, e Akbulut et al. [69] investigaram a aplicação da técnica SA para minimização da espessura de laminados.

Capítulo 3

Classificação por Floresta de Caminhos Ótimos

O classificador OPF modela o problema de reconhecimento de padrões como sendo uma tarefa de particionamento em um grafo induzido pelo conjunto de dados. Cada instância da base de dados, a qual é representada pelo seu vetor de características, é tratada como sendo o nó de um grafo completo, sendo que as arestas entre elas são ponderadas pela distância entre seus respectivos vetores de características. O conjunto de dados é dividido em amostras para realizar o treinamento e teste. Na etapa de treinamento, cada nó do grafo é conectado com as demais amostras através de arestas ponderadas (Figura 3.1a). Em seguida, algumas instâncias (protótipos) de cada classe são escolhidas. A escolha dos protótipos leva em consideração a distância entre os nós do grafo, os quais são escolhidos como protótipos os nós mais próximos de classes diferentes. Em seguida, os custos associados à cada nó do grafo são inicializados com valor de custo mínimo para os nós escolhidos como protótipos e valor de custo máximo para os demais nós (Figura 3.1a). É importante destacar que, como consequência dos custos iniciais associados aos protótipos, um protótipo não conseguirá conquistar outro, pois já possuem valores de custo mínimo e não existem arcos ponderados com valores negativos. Finalmente, uma competição entre os nós é iniciada a partir dos protótipos com o intuito de conquistar os demais nós do conjunto de dados de treinamento, oferecendo a eles valores de caminho de custo ótimo. Após esse processo de competição, cada protótipo será a raiz de uma árvore de caminhos ótimos, a qual contém as instâncias mais fortemente conexas a ele, ou seja, as instâncias irão pertencer à mesma classe do protótipo as quais estão conectadas. A coleção dessas árvores nos remete a uma floresta de caminhos ótimos, que dá o nome ao referido classificador.

Seja Z uma base de dados rotulada e Z_1 e Z_2 os conjuntos de treinamento e teste, respectivamente, com $|Z_1|$ e $|Z_2|$ instâncias, tal que $Z = Z_1 \cup Z_2$ e $Z_1 \cap Z_2 = \emptyset$. Seja $\lambda(s)$

uma função que associa o rótulo correto i , $i = 1, 2, \dots, C$ da classe i a qualquer instância $s \in Z_1 \cup Z_2$. Seja $S \in Z_1$ um conjunto de protótipos de todas as classes (isto é, instâncias que melhor representam as classes), e \vec{d} um algoritmo que extrai n atributos de qualquer instância $s \in Z_1 \cup Z_2$, e retorna um vetor de atributos $\vec{d}(s) \in \mathbb{R}^m$. A distância $d(s, t)$ entre duas instâncias, s e t , é dada pela distância entre seus vetores de atributos $\vec{d}(s)$ e $\vec{d}(t)$, respectivamente. Nosso problema consiste então em usar S , (\vec{d}, d) e Z_1 para projetar um classificador ótimo, o qual pode prever o rótulo correto $\lambda(s)$ de qualquer instância $s \in Z_2$.

Seja (Z_1, A) um grafo completo cujos nós são as instâncias em Z_1 , onde qualquer par de instâncias define um arco em A (isto é, $A = Z_1 \times Z_1$). Note que os arcos não precisam ser armazenados e o grafo não precisa ser explicitamente representado. Temos, também, que um caminho em um grafo é uma sequência de k instâncias $\pi_{s_k} = \langle s_1, s_2, \dots, s_k \rangle$, onde $(s_i, s_{i+1}) \in A$ para $1 \leq i \leq k - 1$, com término em s_k . Um caminho é dito ser trivial se $\pi_s = \langle s \rangle$. Associa-se à cada caminho π_s um valor dado por uma função de valor de caminho f , denotada por $f(\pi_s)$. Diz-se que um caminho π_s é ótimo se $f(\pi_s) \leq f(\tau_s)$ para qualquer caminho τ_s , onde π_s e τ_s terminam na mesma instância s , independente de sua origem. Também denota-se $\pi_s \cdot (s, t)$ a concatenação do caminho π_s com término em s e o arco (s, t) . O algoritmo OPF pode ser utilizado com qualquer função de valor de caminho suave [70], sendo que uma função de valor de caminho f é suave quando, para qualquer instância t , existe um caminho ótimo π_t o qual é trivial ou possui a forma $\pi_s \cdot (s, t)$, onde:

- $f(\pi_s) \leq f(\pi_t)$;
- π_s é ótimo, e
- para qualquer caminho ótimo τ_s , $f(\tau_s \cdot \langle s, t \rangle) = f(\pi_t)$.

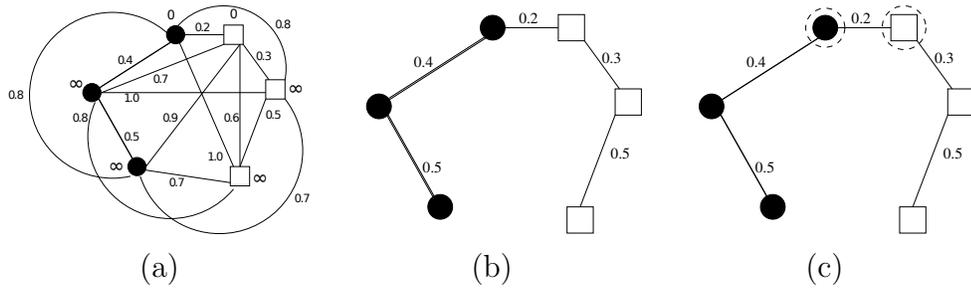


Figura 3.1: (a) Grafo completo ponderado nas arestas para um determinado conjunto de treinamento e inicialização dos custos dos nós do grafo. (b) MST do grafo completo. (c) Protótipos escolhidos como sendo os elementos adjacentes de classes diferentes na MST (nós circutados).

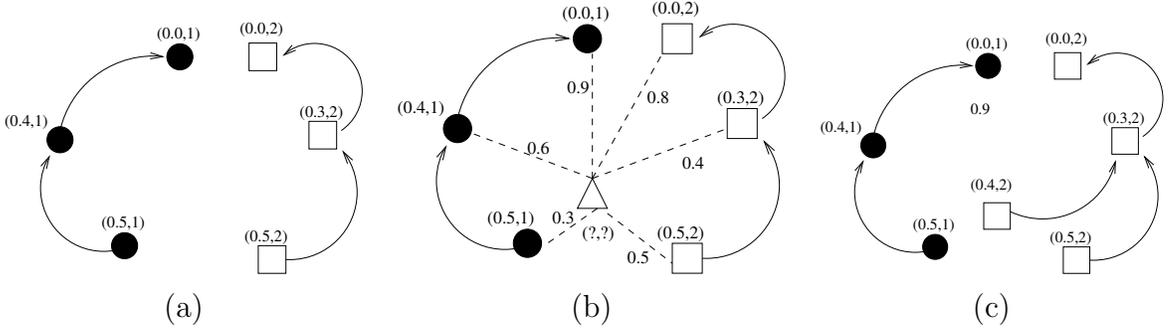


Figura 3.2: (a) Floresta de caminhos ótimos resultante para a Figura 3.1a utilizando a função de valor de caminho f_{max} e dois protótipos. Os identificadores (x, y) acima dos nós são, respectivamente, o custo e o rótulo dos mesmos. A seta indica o nó predecessor no caminho ótimo. (b) Uma instância de teste (triângulo) da classe 2 e suas conexões (linhas pontilhadas) com os nós do conjunto de treinamento. (c) O caminho ótimo do protótipo mais fortemente conexo, seu rótulo 2 e o custo de classificação 0.4 são associados a instância de teste. Note que, mesmo a instância de teste estando mais próxima de um nó da classe 1, ela foi classificada como sendo da classe 2.

Na versão OPF com grafo completo, a função de custo abordada foi a f_{max} , a qual é definida como:

$$f_{max}(\langle s \rangle) = \begin{cases} 0 & \text{se } s \in S, \\ +\infty & \text{caso contrário} \end{cases}$$

$$f_{max}(\pi_s \cdot (s, t)) = \max\{f_{max}(\pi_s), d(s, t)\}, \quad (3.1)$$

onde $f_{max}(\pi_s)$ computa a distância máxima entre instâncias adjacentes em π_s , quando π_s não é um caminho trivial.

O algoritmo baseado em OPF associa um caminho ótimo $P^*(s)$ de S a toda instância $s \in Z_1$, formando uma floresta de caminhos ótimos P (uma função sem ciclos, a qual associa a todo $s \in Z_1$ seu predecessor $P(s)$ em $P^*(s)$, ou uma marca *nil* quando $s \in S$, como mostrado na Figura 3.2a). Seja $R(s) \in S$ a raiz de $P^*(s)$ a qual pode ser alcançada por $P(s)$. O algoritmo computa, para cada $s \in Z_1$, o custo $V(s)$ de $P^*(s)$, o rótulo $L(s) = \lambda(R(s))$ e o seu predecessor $P(s)$, como segue no Algoritmo 1 na página seguinte.

As Linhas 1–5 do Algoritmo 1 inicializam os mapas e inserem protótipos em Q (Figura 3.1a). O laço principal calcula um caminho ótimo de S para cada instância $s \in Z_1$ em uma ordem não decrescente de custos (Linhas 6–14), utilizando como base de cálculo a função de custo mostrada na Equação 3.2. A cada iteração um caminho de custo ótimo $V(s)$ é obtido em P (Linha 9). Empates são resolvidos em Q utilizando a política FIFO (*first-in-first-out*), ou seja, quando dois caminhos atingem uma determinada instância s

com o mesmo custo mínimo, s é associado ao primeiro caminho que o atingiu. O restante das linhas avalia se o caminho que atinge uma instância adjacente t através de s é mais barato que o caminho que termina em t . Em caso de positivo, atualiza Q , $P(t)$, $L(t)$ e $V(t)$. No final do algoritmo, V armazena o valor do custo do caminho ótimo de S a cada instância $s \in Z_1$ de acordo com f_{max} (Figura 3.2a).

Algoritmo 1: Classificador baseado em Floresta de Caminhos Ótimos

Entrada: Um conjunto de treinamento Z_1 λ -rotulado, protótipos $S \subset Z_1$ e o par (v, d) para vetor de atributos e cálculo das distâncias.

Saída: Floresta de caminhos ótimos P , mapa de valores de custo de caminhos V e mapa de rótulos L .

Estruturas auxiliares: Fila de prioridades Q , e variável cst .

```

1 para todo  $s \in Z_1$ , faça
2    $P(s) \leftarrow nil$  e  $V(s) \leftarrow +\infty$ .
3 para todo  $s \in S$ , faça
4    $V(s) \leftarrow 0$ ,  $P(s) \leftarrow nil$ ,  $L(s) = \lambda(s)$ 
5   Insira  $s$  em  $Q$ .
6 enquanto  $Q$  é não vazia, faça
7   Remova de  $Q$  uma instância  $s$  tal que  $V(s)$  é mínimo.
8   para cada  $t \in Z_1$  tal que  $s \neq t$  e  $V(t) > V(s)$  faça
9     Calcule  $cst \leftarrow \max\{V(s), d(s, t)\}$ .
10    se  $cst < V(t)$ , então
11      se  $V(t) \neq +\infty$ , então
12        remova  $t$  de  $Q$ .
13         $P(t) \leftarrow s$ ,  $L(t) \leftarrow L(s)$  e  $V(t) \leftarrow cst$ .
14        Insira  $t$  em  $Q$ .

```

3.1 Treinamento

A fase de treinamento do classificador baseado em floresta de caminhos ótimos usando o grafo completo consiste, basicamente, em encontrar o conjunto S de protótipos, ou seja, os elementos mais representativos de cada classe, e iniciar o processo de competição entre eles no conjunto de treinamento. Várias heurísticas poderiam ser adotadas como, por exemplo, uma escolha aleatória de protótipos. Entretanto, tal escolha pode prejudicar o desempenho do classificador, tornando-o instável e com um alto grau de sensibilidade

com relação aos protótipos escolhidos. Desejamos, assim, estimar protótipos nas regiões de sobreposição de instâncias e nas fronteiras entre as classes, visto que são regiões muito susceptíveis a erros de classificação.

Computando uma MST no grafo completo (Z_1, A) , obtém-se um grafo conexo acíclico cujos nós são todas as instâncias em Z_1 , e os arcos são não direcionados e ponderados (Figura 3.1b). Seus pesos são dados pela distância d entre os vetores de atributos de instâncias adjacentes. Esta árvore de espalhamento é ótima no sentido em que a soma dos pesos de seus arcos é mínima se comparada a outras árvores de espalhamento no grafo completo. Os protótipos a serem escolhidos são os elementos conectados na MST com diferentes rótulos em Z_1 , isto é, elementos mais próximos de classes diferentes (Figura 3.1c). Removendo-se os arcos entre classes diferentes, tais instâncias adjacentes tornam-se protótipos em S e o Algoritmo 1 pode computar uma floresta de caminhos ótimos em Z_1 (Figura 3.2a). Note que uma dada classe pode ser representada por múltiplos protótipos (isto é, árvores de caminhos ótimos) e deve existir pelo menos um protótipo por classe.

3.2 Classificação

Para qualquer instância $t \in Z_2$, consideram-se todos os arcos conectando t com instâncias $s \in Z_1$, tornando t como se fosse parte do grafo original (ver Figura 3.2b, onde a instância t é representada pelo triângulo no grafo). Considerando todos os possíveis caminhos entre S e t , deseja-se encontrar o caminho ótimo $P^*(t)$ de S até t com a classe $\lambda(R(t))$ de seu protótipo $R(t) \in S$ mais fortemente conexo. Este caminho pode ser identificado incrementalmente, avaliando o valor do custo ótimo $V(t)$ como segue:

$$V^*(t) = \min\{V(t), \max\{V(s), d(s, t)\}\}. \quad \forall s \in Z_1. \quad (3.2)$$

Seja $s^* \in Z_1$ o nó que satisfaz a equação acima (isto é, o predecessor $P(t)$ no caminho ótimo $P^*(t)$), ou seja, se o custo ótimo oferecido por s^* ($V^*(t)$) for menor do que o custo atual de t ($V(t)$), então a classificação simplesmente associa $L(s^*)$ como sendo a classe de t (Figura 3.2c). Um erro ocorre quando $L(s^*) \neq \lambda(R(t))$, ou seja, a instância t é ligada a um subgrafo enraizado a um protótipo $R(s)$ diferente de sua classe verdadeira.

Capítulo 4

Campos Aleatórios de Markov

Campos Aleatórios de Markov são modelos probabilísticos que podem ser utilizados para integrar informação espacial contextual em problemas de classificação de imagens [43, 11]. Tais modelos são baseados na premissa de que um *pixel* pertencente a uma determinada classe possui uma probabilidade maior de ter vizinhos também dessa classe do que qualquer outra. Os MRFs 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 vizinhança não paramétricos, é uma extensão direta dos modelos ocultos de Markov (*Hidden Markov Models* - HMM). Nessa abordagem, é preciso impor um ordenamento artificial para compensar a falta da noção natural de causalidade dos dados espaciais em relação a séries temporais, o que frequentemente ocasiona no aparecimento de artefatos direcionais nos campos aleatórios [71]. Entretanto, uma vantagem dessa abordagem consiste na grande quantidade de ferramentas matemáticas disponíveis envolvendo cadeias de Markov unidimensionais [72, 73]. Na outra abordagem, o ponto de partida foram as ideias e ferramentas utilizadas na mecânica estatística para caracterizar a energia de sistemas físicos de partículas orgânicas em um reticulado bidimensional [74]. 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.

A base teórica para trabalhos que empregam MRFs em processamento de imagens surgiram, principalmente, a partir da década de 1970 [75, 14]. A partir daí, várias outras abordagens foram propostas, principalmente em segmentação e restauração de imagens [76, 15]. Mais recentemente, alguns trabalhos têm utilizado campos aleatórios no contexto de reconhecimento de padrões com o intuito de aumentar a eficácia dessas abordagens [77, 78, 11]. Wu e Ouyang [78], por exemplo, propuseram um modelo híbrido para

classificação de imagens utilizando SVM e campos aleatórios de Markov. Tal abordagem, chamada SVM-MRF, consiste de duas etapas: (i) classificação dos dados utilizando SVM e (ii) posterior pós-processamento da imagem utilizando MRF. A ideia consiste, basicamente, em modelar a imagem classificada como um campo aleatório, e o rótulo de cada *pixel* atribuído pelo classificador é analisado em relação à sua vizinhança com o intuito de reclassificá-lo caso atenda à alguns critérios estabelecidos pela abordagem proposta.

Como pode ser notado, a utilização de MRF no contexto de análise de imagens e reconhecimento de padrões tem crescido ultimamente, principalmente pelo fato da complexidade inerente de determinadas imagens, como cenas naturais, por exemplo. Nesses casos, uma classificação baseada em *pixels*, muitas vezes, pode levar-nos a resultados não muito satisfatórios, podendo gerar o que é chamado de “efeito sal e pimenta” (ruído ou erros de classificação) sobre as imagens resultantes, quando não leva-se em consideração a informação contextual espacial [79, 18]. Portanto, explorar as correlações entre a instância e sua vizinhança pode melhorar o desempenho das técnicas de classificação baseadas em *pixel*, uma vez que os *pixels* adjacentes provavelmente pertencem à mesma classe. A próxima seção apresenta uma fundamentação teórica acerca de campos aleatórios de Markov.

4.1 Fundamentação Teórica

Tem-se que um campo aleatório consiste em um conjunto de variáveis aleatórias $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\}$, sendo que y_s é dito ser uma realização de Y_s e pode assumir valores dentro de um alfabeto Σ_{Y_s} , tal que esse campo pode ser caracterizado por uma distribuição de probabilidade conjunta P_Y [15, 80, 81].

Assim sendo, tem-se que uma imagem rotulada pode ser modelada por um campo aleatório $\mathcal{Y} = \{Y_s | s \in \Omega\}$, onde $\Omega = \{(i, j) | 1 \leq i \leq l, 1 \leq j \leq c\}$ denota o conjunto de índices de locais de um reticulado bidimensional contendo $l \times c$ pontos ou *pixels*. De modo geral, em imagens rotuladas, assume-se que o alfabeto $\Sigma_{Y_s} = \{1, 2, \dots, C\}$ é discreto, representando os rótulos para qualquer $s \in \Omega$.

Dado um campo aleatório, pode-se ainda definir um sistema de vizinhança sobre ele. Assim sendo, um sistema de vizinhança η com relação a um local s , isto é, η_s , pode ser definido como um subconjunto de Ω tal que $\eta_s = \{t | d(s, t) \leq \phi\}$, onde ϕ denota um limiar de distância, e $d, s \in \mathbb{R}\}$, onde cada local $t \in \eta_s$ é denominado vizinho de s . Com base nessa ideia, uma ordem pode ser dada pelo seu número de vizinhos como, por exemplo, quatro vizinhos na primeira ordem e oito vizinhos na segunda. As três primeiras ordens são ilustradas na Figura 4.1, já que são as mais utilizadas. Embora a escolha da ordem seja bastante importante, pois interfere na solução final e no custo computacional, tal problema não será abordado. Na próxima seção serão abordados alguns dos modelos de

MRF mais utilizados na literatura.

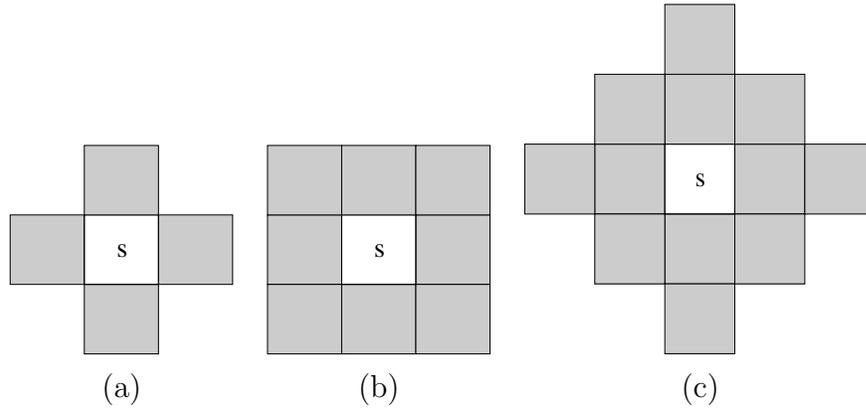


Figura 4.1: Ordem de sistemas de vizinhança de um *pixel* s : (a) primeira ordem com 4 vizinhos, (b) segunda ordem com 8 vizinhos e (c) terceira ordem com 12 vizinhos.

4.2 Modelos de Campos Aleatórios Markovianos

Diversos modelos Markovianos que consideram a informação contextual foram propostos na literatura, sendo que tais modelos são adequados, principalmente, para representar o conhecimento *a priori* por meio de distribuições de probabilidade.

Um dos modelos mais utilizados é o chamado Modelo de Potts (Potts-Strauss) [15], o qual é um modelo Markoviano discreto. Tal modelo é aplicado como MRF *a priori* na maioria dos problemas de problemas de processamento de imagens que envolvem mais de duas classes. Esse modelo pode ser definido por meio das probabilidades condicionais locais (*Local Conditional Density Functions* - LCDFs), conforme segue:

$$p(y_s = m | \eta_{ij}, \beta) = \frac{\exp\{\beta \mathcal{U}_{ij}(m)\}}{\sum_{k=1}^C \exp\{\beta \mathcal{U}_{ij}(k)\}}, \quad (4.1)$$

onde $m \in \Sigma_{y_s}$ representa um rótulo dentre C possíveis, \mathcal{U}_{ij} denota uma função de energia, que pode ser, por exemplo, o número de vizinhos que possuem o mesmo valor de m , e β é um parâmetro que representa o peso dado à função de energia e que deverá ser estimado. Cabe destacar que, quanto maior o valor de β , maior será a dependência espacial entre os elementos do campo, pois maior a probabilidade do elemento central pertencer à classe dominante. Note, porém, que quando $\beta = 0$, ou seja, não há dependência espacial, a função distribuição do modelo se degenera para a função de probabilidade de uma variável

discreta com distribuição uniforme (todos os valores são equiprováveis), e a informação contextual é completamente desprezada, pois $\exp(0) = 1$ e assim $p(y_s = m | \eta_{ij}, \beta) = \frac{1}{C}$.

A função de energia usada no modelo de Potts pode ser definida como:

$$\mathcal{U}_{ij}(m) = \sum_{y_t, t \in \eta_s} \delta(m, y_t), \quad (4.2)$$

onde $\delta(a, b)$ é função delta de Kronecker, versão discreta da função delta de Dirac, dada por:

$$\delta(a, b) = \begin{cases} 1 & \text{se } a = b, \\ 0 & \text{se } a \neq b. \end{cases}$$

Uma extensão do Modelo de Potts proposta por Li et al. [43] que considera uma distribuição contínua dos *pixels* vizinhos é chamada de *Generalized Isotropic Multi-Level Logistic* - GIMLL, a qual pode ser obtida modificando apenas a definição da função de energia \mathcal{U}_{ij} para:

$$\mathcal{U}_{ij}(m) = \sum_{y_t, t \in \eta_{ij}} [1 - 2e^{-(m-y_t)^2}]. \quad (4.3)$$

Caso observemos essa função de energia, vemos que a contribuição é dada por uma norma L2. Assim, é comum chamarmos esse modelo de GIMLL-L2. Podemos, ainda, definir uma outra versão do GIMLL considerando agora uma norma L1, a qual é denominada GIMLL-L1:

$$\mathcal{U}_{ij}(m) = \sum_{y_t, t \in \eta_{ij}} [1 - 2e^{-|m-y_t|}]. \quad (4.4)$$

Alguns autores argumentam que tais modelos são capazes de modelar as interações entre os *pixels* vizinhos de forma mais suave do que o modelo de Potts, sendo adequados para resolver problemas de aumento de resolução espacial e também para restauração de imagens, uma vez que apenas os *pixels* vizinhos que pertencem à mesma classe contribuem para a probabilidade *a priori*. Embora Martins et al. [82] tenham relatado que os resultados obtidos foram promissores, tais funções tiveram um alto custo computacional devido ao número de imagens e à dimensão da imagem.

Finalmente, um outro modelo bastante utilizado é o MRF Gaussiano (*Gaussian MRF* - GMRF), cujas probabilidades condicionais locais são dadas por [43]:

$$p(y_s = m | \{y_t, t \in \eta_s\}, \beta, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} [m - \mu - \sum_{y_t, t \in \eta_s} \beta(y_t - \mu)]^2}, \quad (4.5)$$

onde μ e σ^2 correspondem à média e variância globais estimadas sobre as amostras, respectivamente. Para estimar o parâmetro β no modelo GMRF, existe uma fórmula fechada para o estimador de Máxima Pseudo-Verossimilhança (*Maximum Pseudo-Likelihood*) proposta por Besag [14], que é definida por:

$$\hat{\beta}_{MPL} = \frac{\sum_s [(y_s - \mu) \sum_{y_t} (y_t - \mu)]}{\sum_s [\sum_{y_t, t \in \eta_s} (y_t - \mu)^2]}, \quad (4.6)$$

onde μ corresponde agora à média global da imagem observada rotulada.

4.3 Modo Condicional Iterativo (ICM)

Um dos grandes problemas para estimar uma imagem classificada a partir de um MRF é encontrar o mínimo global de uma função de energia, especialmente se tal função possuir vários mínimos locais. Embora Li et al. [43] tenham relatado que não existe um algoritmo eficiente que encontre o mínimo global da função com garantia de sucesso, uma boa alternativa que vem sendo utilizada é o método *Iterated Conditional Modes* (ICM) que foi proposto por Besag [76], o qual modela o problema da maximização da probabilidade conjunta de um MRF sequencialmente. Tal algoritmo usa a estratégia gulosa em uma abordagem iterativa para encontrar uma energia mínima local, sendo que sua convergência para o máximo global da função não é garantida. Porém, com uma boa estimativa inicial, são garantidos bons resultados em poucas iterações. Assim, os resultados são extremamente sensíveis à estimativa inicial, especialmente em espaços de alta dimensão com energias não convexas devido ao grande número de mínimos locais. Além disso, o ICM incorpora conhecimento adquirido de uma instância a partir da escolha de um sistema de vizinhança, uma função de energia e um parâmetro de dependência espacial. De maneira geral, este algoritmo tenta explorar a tendência de que *pixels* adjacentes provavelmente possuem a mesma cor.

Em suma, a ideia por trás do método ICM para classificação de imagens é atualizar cada *pixel* maximizando iterativamente $p(y_s = m | \eta_{ij}, \beta)$, que é obtida pela maximização da função de energia $\mathcal{U}_{ij}(m)$ dada pela Equação 4.2, caso contrário pode ser obtida pela minimização da função de energia se forem utilizadas as Equações 4.3 e 4.4, assumindo as hipóteses de que as instâncias são independentes entre si e que um *pixel* depende apenas dos valores dos *pixels* em uma vizinhança local.

4.4 Teoria da Informação de Fisher

A Informação de Fisher é uma medida para a quantidade de informação que é esperado de uma amostra sobre o parâmetro de interesse β [83]. É definida como a variância da função-objetivo, ou seja, a derivada do logaritmo da função de verossimilhança com relação ao parâmetro, sob condições adicionais de regularidade.

Em uma distribuição cujas instâncias são independentes e identicamente distribuídas, o cálculo da Informação de Fisher é trivial, pois basta multiplicar a contribuição de uma única instância pelo tamanho da amostra. Entretanto, ao lidar com variáveis espacialmente dependentes, como no caso de modelos Markovianos, diferentes padrões de configuração espaciais definidos em termos de um sistema de vizinhança fornecem diferentes contribuições à Informação de Fisher global [83].

Porém, em alguns casos, não é possível calcular a Informação de Fisher na prática, devido à inexistência de expressões fechadas para os valores esperados. Nessas situações, pode-se adotar uma alternativa denominada de Informação de Fisher observada.

Seja $G = (V, E)$ um grafo rotulado com $|V|$ vértices e $|E|$ arestas, onde y_i representa o rótulo do vértice $v_i \in V$. Suponha ainda que o grafo rotulado possa ser modelado como um MRF definido pelo conjunto de LCDFs $p(y_i | \eta_i, \beta, \vec{\theta})$, $i = 1, 2, \dots, |V|$, com $\vec{\theta}$ denotando o vetor de parâmetros do modelo (como apresentado na Seção 4.2, por exemplo) e β denotando o parâmetro de dependência espacial entre os rótulos do grafo. Seja, ainda, $\varphi_\beta(y_i)$ definido como:

$$\varphi_\beta(y_i) = \left[\frac{\delta}{\delta\beta} \log p(y_i | \eta_i, \beta, \vec{\theta}) \right]^2, \quad (4.7)$$

que denota a Informação de Fisher Observada Local em relação a β calculada a partir da primeira derivada para o vértice v_i .

O princípio do mínimo da Informação de Fisher declara que é necessário escolher a distribuição de probabilidade que minimiza a Informação de Fisher sujeita às restrições conhecidas sobre o sistema [84]. Note que a distribuição de probabilidade que minimiza a Informação de Fisher também irá maximizar o erro médio quadrático.

A minimização de $\varphi_\beta(y_i)$ pode ser um critério bastante interessante na solução de problemas envolvendo modelos Markovianos, pois ao escolher o rótulo y_i^* que minimiza $\varphi_\beta(y_i)$, estaremos minimizando o quadrado da taxa de variação da função densidade condicional local em relação a β . Quando essa taxa é próxima de zero, estamos perto de um ponto com alta verossimilhança. Assim, a ideia da minimização da Informação de Fisher está diretamente relacionada com a maximização da verossimilhança local.

Dado o vértice $v_i \in V$, o cálculo do nível de informação presente em cada vértice da Equação 4.1 é dado por [85, 86]:

$$\varphi_\beta(y_i) = \frac{\sum_{l=1}^C \sum_{k=1}^C \{[U_i(m) - U_i(l)][U_i(m) - U_i(k)]e^{\beta[U_i(l)+U_i(k)]}\}}{\sum_{l=1}^C \sum_{k=1}^C e^{\beta[U_i(l)+U_i(k)]}}. \quad (4.8)$$

onde φ_i é a informação do vértice i , C é a quantidade de classes, U_i é um vetor de C posições, onde cada posição representa a quantidade de vértices ligantes de determinada classe ao vértice i , m é a posição no vetor U_i da classe do vértice i .

Embora possa ser aplicado como um modelo Markoviano, ou seja, $\varphi_\beta(y_i) = p(y_i|\eta_i, \beta)$, tal equação não garante as seguintes condições:

$$\sum_{i=1}^q \mathcal{P}(y_i = m|\eta_i, \beta) = 1 \quad e \quad (4.9)$$

$$\mathcal{P}_i(y_i = m|\eta_i, \beta) \geq 0, \quad \text{para } m = 1, \dots, C \quad (4.10)$$

Tais condições podem ser satisfeitas alterando a Equação 4.8 como a seguir:

$$\mathcal{P}_i(y_i = m|\eta_i, \beta) = \frac{\sum_{l=1}^C \sum_{k=1}^C \{[U_i(m) - U_i(l)][U_i(m) - U_i(k)]e^{\beta[U_i(l)+U_i(k)]}\}}{\sum_{p=1}^C \sum_{l=1}^C \sum_{k=1}^C \{[U_i(p) - U_i(l)][U_i(p) - U_i(k)]e^{\beta[U_i(l)+U_i(k)]}\}}, \quad (4.11)$$

onde $m = 1, \dots, C$. Note que $[U_i(m) - U_i(l)][U_i(m) - U_i(k)]$ agora assume apenas valores positivos e que $\mathcal{P}_i(y_i = m|\eta_i, \beta)$ é normalizado.

De forma similar, tal modelo também pode ser aplicado utilizando os modelos GIMLL-L1 e GIMLL-L2 (Equações 4.4 e 4.3, respectivamente), uma vez que são extensões do modelo de Potts para valores contínuos. Porém, a maximização de $\mathcal{P}_i(y_i = m|\eta_i, \beta)$ para tais modelos é dada pela maximização da Informação de Fisher.

Capítulo 5

Classificação Contextual com Floresta de Caminhos Ótimos

Esta seção apresenta as abordagens propostas pela presente tese de doutorado.

5.1 OPF-MRF: Uma Abordagem Contextual Baseada em Floresta de Caminhos Ótimos

Nesta seção é apresentado um método de classificação contextual por meio do classificador OPF com Campos Aleatórios Markovianos, o qual considera a informação contextual dentro de uma vizinhança obtida por meio de modelos de campos aleatórios. A abordagem OPF-MRF foi apresentada em [87, 88, 89]

Segundo Kittler et al. [17], em uma abordagem de classificação baseada em contexto, pode-se integrar a informação contextual em três diferentes etapas do processo de classificação: (i) pós-processamento, (ii) pré-processamento, e (iii) extração de características. Uma técnica amplamente utilizada é editar os *pixels* classificados de acordo com algumas regras. Este pós-processamento é guiado por regras inferidas a partir de um conhecimento contextual. Exemplos mais simples de pós-processamento podem ser feitos através de filtragem da imagem classificada com qualquer tipo de filtro, convolução ou filtros não lineares [90, 91], ou mesmo através de regras sintáticas [92, 93]. Dentro desta mesma categoria, porém mais sofisticada, pode-se citar a abordagem proposta por Wharton [27].

Outra abordagem é a integração de informação contextual na etapa de pré-processamento por meio de um algoritmo escolhido apropriadamente, onde a imagem é dividida em regiões homogêneas que são então classificadas como objetos. Embora a classificação desses objetos normalmente seja mais confiável do que *pixels* classificados individualmente, ainda temos que enfrentar o problema de segmentação. Isto é difícil tanto quanto o seu processo de classificação, portanto as suas vantagens são limitadas.

Uma terceira abordagem compreende métodos que adicionam novos componentes aos vetores de características, sendo a classificação realizada usando este vetor estendido. A forma mais simples consiste em adicionar os valores dos *pixels* vizinhos ou uma função apropriada deles (como, por exemplo, média e variância) ao vetor original [94]. Alternativamente, novos componentes podem ser derivados a partir de descritores de textura como, por exemplo, coeficientes de Fourier ou matriz de co-ocorrência [95, 96, 97]. Problemas comuns nesses métodos são a excessiva dimensionalidade dos vetores estendidos, alto custo computacional e baixo desempenho de classificação na fronteira dos objetos.

A quarta abordagem combina informação espectral (valores radiométricos associados aos *pixels* nas diferentes bandas espectrais) e espacial, e classifica os *pixels* usando ambas fontes ao mesmo tempo, sendo similar à família de métodos do parágrafo anterior. Esta abordagem permite um maior uso da informação disponível e, se necessário, podem ser executadas operações de pré ou pós-processamento. Ou seja, depois da classificação da imagem, é definido um novo vetor de características a partir do rótulo dos *pixels* em uma dada vizinhança, e durante um segundo passo esses vetores são reclassificados. A informação contextual é então utilizada no segundo processo de classificação. Neste capítulo, é apresentada uma abordagem de classificação contextual usando uma ideia similar combinando a técnica OPF com MRF.

É importante salientar que tais abordagens não alteram a base das técnicas de classificação tradicionais, podendo assim ser facilmente implementadas para executar classificação contextual. Contudo, a desvantagem comum a todos esses métodos é o fato de que tentam recuperar a informação perdida, ou seja, o relacionamento de um *pixel* com seus *pixels* vizinhos, e tais tentativas podem ser bem sucedidas até um certo ponto, pois a qualidade da informação recuperada depende de uma boa estimativa inicial. Além disso, tais abordagens são limitadas apenas para problemas de classificação de *pixels* de uma imagem, uma vez que é utilizado o modelo de vizinhança baseado na proximidade espacial entre os *pixels*.

A ideia principal por trás do OPF-MRF é explorar a informação contextual por meio de um modelo de vizinhança baseado em um reticulado bidimensional, ou seja, uma estrutura em forma de uma grade regular (*grid*) estabelecida sobre a área estudada. Basicamente, o algoritmo de aprendizado baseado em contexto proposto para o classificador OPF pode ser dividido em três fases distintas: (i) uma classificação inicial dos *pixels* é executada pelo OPF padrão para gerar o mapa de rótulos inicial (imagem classificada); (ii) para cada instância, a probabilidade condicional local $\mathcal{P}(y|n_{ij}, \beta)$, $y = \{1, 2, \dots, C - 1, C\}$, dada pela Equação 4.1 é calculada para cada classe sobre o mapa de rótulos fornecido no passo anterior, e então novos componentes são adicionados ao vetor de características, ou seja, dado o vetor de características $\vec{o}_s = \{o_1, o_2, \dots, o_n\}$ de uma dada instância $s \in \mathcal{Z}$ com n características, o resultado do cálculo das probabilidades condicionais locais de cada

classe é adicionado como característica e, portanto, o vetor de características \vec{o}_s passa a ser composto por $n + C$ características; em seguida (iii) é realizado um novo treinamento e uma nova classificação usando o OPF padrão com o intuito de gerar um mapa de rótulos atualizado. O processo é iterativo do passo (ii) ao passo (iii), até que o critério de convergência seja alcançado, sendo que tal processo iterativo implementa o método ICM. Basicamente, a ideia do OPF-MRF é maximizar iterativamente uma função de densidade de probabilidade local seguida pela minimização da função de custo de caminho f_{max} (Equação 4.9) para todas instâncias do conjunto de dados. A Figura 5.1 esquematiza o fluxograma do OPF-MRF.

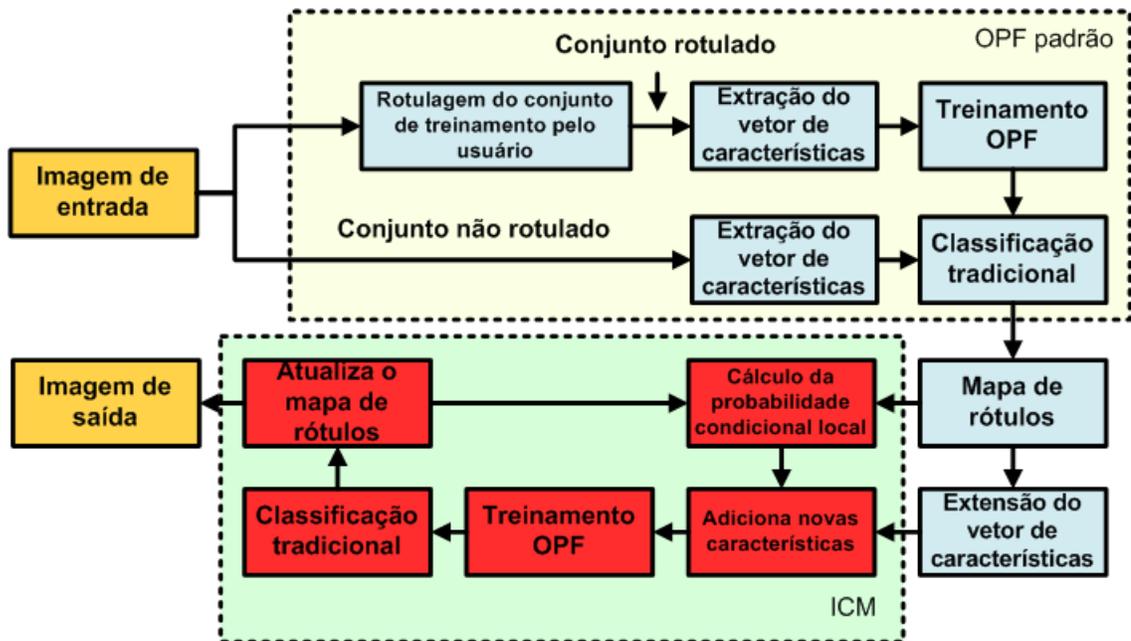


Figura 5.1: Fluxograma do OPF-MRF.

Pode-se perceber que o primeiro passo do OPF-MRF segue o procedimento de classificação tradicional, isto é, têm-se os conjuntos rotulado e não-rotulado, sendo o último usado para gerar o mapa de rótulos (imagem classificada) que será utilizado como entrada para a classificação contextual. Enquanto o método ICM maximiza a densidade de probabilidade local da imagem classificada, o OPF emprega tais informações para uma nova etapa de classificação. Portanto, a sinergia entre o OPF e ICM permite que o OPF-MRF refine a classificação baseada em *pixel*. O Algoritmo 2 implementa esta ideia usando o algoritmo ICM, que maximiza o MRF iterativamente [76]. Note que o algoritmo emprega o modelo de Potts. Além disso, nesta abordagem, o conjunto de dados foi dividido em 2 subconjuntos $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$ originando os conjuntos rotulado e não rotulado, respectivamente.

Algoritmo 2: OPF-MRF

Entrada: Um conjunto rotulado \mathcal{Z}_1 , conjunto não rotulado \mathcal{Z}_2 , número de iterações do OPF-MRF e o parâmetro de dependência espacial β .

Saída: Um mapa de rótulos \mathcal{L}

- 1 Defina um modelo de Campos Aleatórios Markovianos, \mathcal{M} (Exemplo. modelo de Potts).
- 2 Treine o OPF usando \mathcal{Z}_1 (Algoritmo 1).
- 3 Classifique \mathcal{Z}_2 gerando uma estimativa inicial \mathcal{L} .
- 4 Crie novas instâncias de \mathcal{Z}'_1 e \mathcal{Z}'_2 estendendo \mathcal{Z}_1 e \mathcal{Z}_2
- 5 com a frequência absoluta dos rótulos dos vizinhos dados por \mathcal{L} .
- 6 $i \leftarrow 1, \Delta \leftarrow +\infty$.
- 7 **enquanto** $i \leq T$ e $\Delta > 0.01\%$, **faça**
- 8 $[\mathcal{Z}^*_1, \mathcal{Z}^*_2] \leftarrow \mathcal{M}(\mathcal{Z}'_1, \mathcal{Z}'_2, \beta, \mathcal{L})$.
- 9 Treine OPF usando \mathcal{Z}^*_1 .
- 10 Classifique \mathcal{Z}^*_2 gerando mapa de rótulos atualizada \mathcal{L}' .
- 11 Calcula Δ usando \mathcal{L} e \mathcal{L}' .
- 12 $\mathcal{L} \leftarrow \mathcal{L}', \mathcal{Z}'_1 \leftarrow \mathcal{Z}^*_1, \mathcal{Z}'_2 \leftarrow \mathcal{Z}^*_2$ e $i \leftarrow i + 1$.

Na Linha 1, define-se um modelo MRF (diferentes modelos podem ser vistos na Seção 4.2). Por exemplo, aplicando o modelo de Potts, basicamente é considerado a distribuição de frequência de cada rótulo em uma dada vizinhança de um *pixel* (i, j) , definida por n_{ij} . Nas Linhas 2 – 3, o classificador OPF é treinado sobre o conjunto rotulado \mathcal{Z}_1 e o conjunto não rotulado \mathcal{Z}_2 é classificado, gerando uma estimativa inicial (para posterior extração da informação contextual a partir do mapa de rótulos) \mathcal{L} . As Linhas 4 – 5 criam novos conjuntos rotulado e não rotulado estendendo o vetor atual de características de cada instância a partir de \mathcal{Z}_1 e \mathcal{Z}_2 , nos quais são acrescentadas C características fornecidas pelos modelos Markovianos, onde C é o número de possíveis classes. Além disso, o laço nas Linhas 7 – 12 (módulos “vermelhos” na Figura 5.1) implementa o algoritmo ICM, sendo responsável por refinar a solução. Para tal, são calculadas as probabilidades condicionais locais de cada instância e adicionadas como sendo novas características aos seus respectivos vetores de características estendidos \mathcal{Z}'_1 e \mathcal{Z}'_2 (Linha 8) fornecendo os conjuntos \mathcal{Z}^*_1 e \mathcal{Z}^*_2 , respectivamente. É possível observar que um dado modelo Markoviano M necessita de 4 entradas: conjuntos rotulado \mathcal{Z}'_1 e não rotulado \mathcal{Z}'_2 estendidos, o parâmetro de dependência espacial β e o mapa de rótulos gerado pelo OPF padrão. Depois disso, as saídas são os novos conjuntos rotulado \mathcal{Z}^*_1 e não rotulado \mathcal{Z}^*_2 , nos quais os vetores de características dos conjuntos são atualizados. É importante notar que na linha 7 existem dois critérios de parada: o número máximo de iterações T e o número

de instâncias do conjunto não rotulado nas quais os rótulos são diferentes entre a iteração atual e anterior. Assim, a variável Δ pode ser entendida como a quantidade de mudanças que ocorreram no mapa de rótulos entre as iterações. Embora Besag [76] tenha relatado que, em termos de taxa de convergência, o máximo local pode ser alcançado com 6 a 8 iterações do ICM, foi utilizado em nossos experimentos $T = 10$ (escolhido empiricamente).

Em seguida, o classificador OPF é aplicado sobre \mathcal{Z}_1^* (treinamento) e \mathcal{Z}_2^* (etapa de classificação) nas Linhas 9 – 10. A Linha 10 gera um mapa de rótulos atualizado (imagem classificada) \mathcal{L}' , que é então utilizado na próxima iteração do ICM. Finalmente, a Linha 12 atualiza as variáveis para a próxima iteração do OPF-MRF. Adicionalmente, é possível observar que o OPF-MRF possui dois parâmetros: número de iterações T e o parâmetro de dependência espacial β . Assim, uma das contribuições da próxima seção é apresentar uma abordagem de otimização por meta-heurísticas para encontrar β automaticamente.

5.2 Estimando a Dependência Espacial Utilizando Algoritmos Evolutivos

Nesta seção, é apresentada uma abordagem baseada em otimização por meta-heurísticas para estimar os parâmetros utilizados para classificação contextual do OPF-MRF apresentado na Seção 5.1. Essa abordagem foi proposta em [98, 88].

Conforme descrito na Equação 4.1, o Modelo de Potts é uma abordagem discreta para estimar funções de densidade de probabilidade locais, sendo que a variável β define o peso que a função de energia irá representar no cálculo desse modelo. Quando $\beta = 0$, tem-se que a relação de vizinhança é desprezada e o modelo torna-se o tradicional, ou seja, aquele que não leva em consideração o contexto. Embora Levada et al. [99, 100, 77] tenham desenvolvido equações para a estimativa desse parâmetro para vizinhanças de segunda ($k = 8$) e terceira ordens ($k = 12$), em modelos de vizinhança mais robustos sua escolha ainda é empírica, o que torna-os bastante sensíveis à experiência do usuário. Nesse contexto, esta seção propõe a estimativa do parâmetro β utilizando algoritmos de otimização evolutivos.

A estimativa do parâmetro de dependência espacial foi feita por meio de um conjunto de avaliação¹, ou seja, a técnica de otimização objetiva encontrar o valor de β que maximiza a taxa de acerto do classificador OPF nesse conjunto. Ao inicializar-se uma população de possíveis indivíduos-solução, a cada um deles é atribuído um valor

¹Conforme foi mencionado na seção anterior, na prática, o usuário que deseja classificar uma imagem, por exemplo, deverá selecionar (marcar) alguns *pixels* de cada classe, os quais são utilizados como conjunto de treinamento. Na abordagem proposta nesta seção, o conjunto de treinamento rotulado foi particionado pelo usuário em um novo conjunto de treinamento e um conjunto de validação, que será então utilizado para estimar o valor de β .

de $\beta \in [0, \beta_{max}]$, sendo que β_{max} corresponde ao maior valor de β desejado e pode ser expresso por $\beta_{max} = \ln(1 + \sqrt{C})$ [101], sendo C o número máximo de classes possíveis. Embora o número de soluções seja limitado pelo número de classes, podem-se ter infinitas soluções dentro deste intervalo. Para valores acima de β_{max} , o sistema sempre se comporta de maneira organizada, com prevalência de alinhamento entre elementos vizinhos e praticamente não revelando diferença alguma. É importante salientar que este é o valor crítico de β para o Modelo de Potts, não sendo válido para os demais modelos. Contudo, para limitar o problema de encontrar o valor de β , tal valor crítico também foi utilizado para os demais modelos. Assim, após o aprendizado do classificador OPF em um conjunto rotulado, o conjunto de classificação já rotulado pelo classificador seria modelado como sendo um MRF e re-classificado novamente utilizando o Modelo de Potts e o valor de β encontrado pela abordagem de otimização, diga-se β^* . Assim sendo, em uma nova geração de indivíduos, a técnica de otimização seria guiada pelos melhores valores de β . A Figura 5.2 mostra a metodologia empregada neste trabalho.

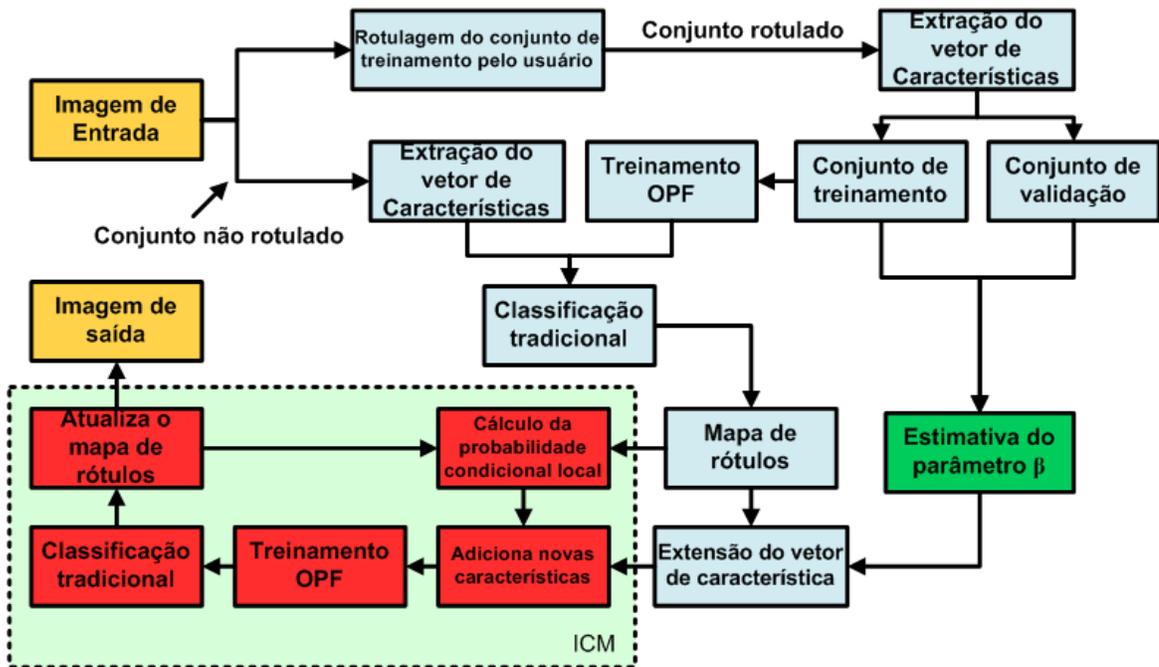


Figura 5.2: Metodologia proposta para encontrar β^* .

O primeiro passo consiste em extrair as características de imagens para construir o conjunto de dados, o qual é composto por instâncias baseadas em *pixels*. Em seguida, o conjunto de dados \mathcal{Z} pode ser agora particionado em $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_v \cup \mathcal{Z}_2$, como sendo os con-

juntos rotulados de treinamento e de validação, e não rotulado de teste, respectivamente. A ideia é usar ambos \mathcal{Z}_1 e \mathcal{Z}_v para guiar os algoritmos de otimização para maximizar a taxa de reconhecimento sobre \mathcal{Z}_v , onde cada agente tem uma posição no espaço de busca definido por $\beta \in [0, \beta_{max}]$, o qual é inicializado aleatoriamente. Em seguida, o Algoritmo 2 é executado toda vez que um agente muda sua posição até que o critério de convergência seja satisfeito. Tal procedimento é codificado no módulo “verde” da Figura 5.2. Além disso, depois que um valor adequado para β é encontrado, isto é, o valor que maximiza a taxa de reconhecimento sobre \mathcal{Z}_v , o mesmo é utilizado para orientar o algoritmo ICM a maximizar a taxa de reconhecimento sobre o mapa de rótulos na etapa de refinamento, etapa esta executada pelos módulos “vermelhos” (Figura 5.2).

Embora qualquer outra técnica de otimização pudesse ser empregada para encontrar β , neste trabalho foi proposta uma abordagem por meta-heurísticas para tal propósito, uma vez que tais tipos de técnicas apresentam um esquema elegante e amigável para aplicações orientadas à otimização. Foram utilizadas duas técnicas de otimização: Otimização por Enxame de Partículas (*Particle Swarm Optimization* - PSO) [60] e Busca Harmônica (*Harmony Search* - HS) [102]. Enquanto a primeira abordagem é baseada na dinâmica social de um grupo de indivíduos e a interação entre eles com o objetivo de encontrar uma fonte de comida, HS modela o problema de otimização de variáveis como sendo uma tarefa de encontrar uma música com a melhor harmonia possível. Neste caso, cada variável de decisão é modelada como um instrumento musical, o qual pode se originar da memória harmônica (memória do artista) ou até mesmo do processo de improvisação do artista. Portanto, enquanto o PSO tem sido amplamente usado em diversas aplicações, HS tem sido uma das técnicas de otimização mais rápidas, tornando-se, assim, bastante interessante no contexto deste trabalho. Uma breve explicação será dada sobre a implementação com otimização de parâmetros (módulo verde na Figura 5.2) usando Otimização por Enxame de Partículas e Busca Harmônica a seguir.

5.2.1 Otimização por Enxame de Partículas

A Otimização por Enxame de Partículas [60] é um método iterativo para otimização global, que trabalha atualizando um enxame de partículas que se movem em torno do espaço de busca influenciadas pelo aperfeiçoamento descoberto por outras partículas. O PSO foi empregado com o intuito de encontrar o valor de β que maximiza as dependências espaciais/temporais, tentando fornecer melhor acurácia do que uma escolha aleatória. O Algoritmo 3 implementa este procedimento.

Primeiramente, um conjunto de possíveis valores é escolhido aleatoriamente e associados à uma partícula $b_k \in \mathcal{B}$ (Linha 1), onde \mathcal{B} corresponde ao conjunto de partículas, respeitando o intervalo de possíveis valores no espaço de busca. Além disso, para cada

Algoritmo 3: Procedimento PSO para encontrar β^*

Entrada: Conjuntos de treinamento λ -rotulado \mathcal{Z}_1 e validação \mathcal{Z}_v , um conjunto de partículas \mathcal{B} , modelo MRF \mathcal{M} , número de iterações T , número de partículas U e número de iterações do ICM n .

Saída: Valor de β^* que maximiza a acurácia do OPF sobre o conjunto validação.

```

1 Defina um conjunto de partículas  $\mathcal{B}$ .
2 para cada iteração  $i$  ( $i = 1, \dots, T$ ), faça
3   para cada partícula  $k$  ( $k = 1, \dots, U$ ), faça
4     Treine OPF usando  $\mathcal{Z}_1$ .
5     Classifica  $\mathcal{Z}_v$  gerando uma estimativa inicial  $\mathcal{L}$ .
6     Crie novas instâncias  $\mathcal{Z}'_1$  e  $\mathcal{Z}'_v$  estendendo  $\mathcal{Z}_1$  e  $\mathcal{Z}_v$ 
7     com a frequência dos rótulos dos vizinhos dados por  $\mathcal{L}$ .
8     para cada iteração  $j$  ( $j = 1, \dots, n$ ), faça
9        $[\mathcal{Z}^*_1, \mathcal{Z}^*_v] \leftarrow \mathcal{M}(\mathcal{Z}'_1, \mathcal{Z}'_v, b_k, \mathcal{L})$ .
10      Treine OPF usando  $\mathcal{Z}^*_1$ .
11      Classifique  $\mathcal{Z}^*_v$  gerando um mapa de rótulos atualizado  $\mathcal{L}'$ ,
12      e armazene a acurácia em  $A$ .
13       $\mathcal{L} \leftarrow \mathcal{L}'$ ,  $\mathcal{Z}'_1 \leftarrow \mathcal{Z}^*_1$  e  $\mathcal{Z}'_v \leftarrow \mathcal{Z}^*_v$ .
14     $f_k \leftarrow A$ .
15  $\beta^* \leftarrow \max(\vec{f})$ .

```

iteração do PSO (repetição nas Linhas 2 – 14), executa-se um procedimento similar dado pelo Algoritmo 2, isto é, cada partícula k armazena seu próprio valor de β em b_k , que é então utilizado como entrada para o cálculo da probabilidade condicional local de um dado modelo Markoviano \mathcal{M} junto com o mapa de rótulos \mathcal{L} , e os conjuntos de treinamento e validação (Linha 9). Depois da execução do algoritmo ICM (repetição nas Linhas 8 – 13), tem-se a acurácia da última iteração sobre o conjunto de validação, que é armazenada em f_k . O procedimento acima é executado mais $T - 1$ iterações, e o valor final de β é o que maximiza a acurácia armazenada em \vec{f} (Linha 15).

5.2.2 Busca Harmônica

A Busca Harmônica [103] é um algoritmo meta-heurístico de otimização, que foi inspirado pela forma como os músicos criam músicas. No processo de improvisação musical, um músico normalmente segue três regras: (a) toca um trecho exato de uma música famosa da memória dele/dela; (b) toca algo similar a um trecho conhecido, mas com leve afinamento; ou (c) compõe uma nova ou aleatória harmonia. Estas regras são implementadas usando os parâmetros “taxa de consideração da memória harmônica” (*Harmony Memory Consideration Rate* - HMCR), a qual define a probabilidade de utilizar um valor da memória harmônica ou compor a nova harmonia, “taxa de ajuste de som” (*Pitch Adjustment Rate* - PAR) e “Largura de Banda” (*Bandwidth* - BW), as quais controlam a quantidade de ruído adicionado a uma possível solução para evitar armadilhas de ótimos locais. Baseado nas ideias acima descritos, o Algoritmo 4 apresenta a abordagem proposta para encontrar β .

Primeiramente, a Memória Harmônica é inicializada com valores de β dentro do intervalo na Linha 1. Nas Linhas 2 – 3, o classificador OPF tradicional é treinado sobre \mathcal{Z}_1 e sua eficácia é avaliada sobre \mathcal{Z}_v . A repetição nas Linhas 4 – 24 executa o algoritmo HS, e a Linha 6 gera um número aleatório dentro do intervalo $[0, 1]$. As Linhas 7 – 14 são responsáveis por criar uma nova harmonia, etapa que pode ser realizada usando valores a partir da Memória Harmônica (Linhas 8 – 12) ou então podem ser empregados apenas valores aleatórios (Linha 14).

As Linhas 15 – 16 criam versões estendidas de \mathcal{Z}_1 e \mathcal{Z}_v , empregando-as no algoritmo ICM nas Linhas 17 – 22, como mencionado no Algoritmo 4. Além disso, as Linhas 23 – 24 são responsáveis por substituir a pior harmonia na Memória Harmônica (MH) pela nova harmonia, caso esta possua uma melhor acurácia. O procedimento acima é executado sobre $T - 1$ iterações, sendo o valor de β final que maximiza a acurácia armazenado em \vec{f} (Linha 25). É importante notar que tais algoritmos de otimização utilizam os subconjuntos \mathcal{Z}_1 e \mathcal{Z}_v referentes aos conjuntos treinamento e validação, respectivamente, para encontrar o melhor valor de β a ser usado posteriormente no Algoritmo 2 com os subconjuntos \mathcal{Z}_1

e \mathcal{Z}_2 referentes aos conjuntos treinamento e teste, respectivamente.

O *Improved Harmony Search* (IHS) [104] é um método variante do HS e difere do HS tradicional por atualizar os valores dos parâmetros $PAR \in [PAR_{min}, PAR_{max}]$ e $BW \in [BW_{min}, BW_{max}]$ dinamicamente. Já o método *Global-best Harmony Search* (GHS) [105] emprega a mesma modificação proposta por IHS com relação aos valores do PAR escolhidos dinamicamente. No entanto, não é empregado o conceito de largura de banda, sendo os valores da nova harmonia baseados a partir da nova harmonia.

5.3 Estimando a Dependência Espacial com Métodos de Otimização Numéricos

Utilizando a mesma ideia aplicada na Seção 5.2, também foi proposta a utilização de métodos numéricos para estimar o parâmetro de dependência espacial. Embora o resultado de classificação contextual seja bastante imprevisível quando o valor do parâmetro é variado, é esperado que os métodos numéricos apresentem uma boa aproximação para o parâmetro β que maximiza a taxa de acerto do OPF-MRF, especialmente se o resultado de classificação apresentar um comportamento suave quando o parâmetro é ajustado. Para tal tarefa, foram utilizados os métodos Nelder-Mead (NM) e Gradiente Descendente (GD) apresentados a seguir.

5.3.1 Método Nelder-Mead

O método Nelder-Mead é um método heurístico iterativo de busca direta proposto para encontrar pontos estacionários em problemas livres multidimensionais [106]. Tal método é utilizado quando não é possível encontrar o valor máximo ou mínimo da função ou quando necessita de alto custo computacional para determinar o valor que maximiza ou minimiza a função.

Sendo β^0 a estimativa inicial do parâmetro β , o método Nelder-Mead cria um simplex $\mathcal{S}^0 = \{\beta_0, \beta_1, \dots, \beta_n\} \in \mathbb{R}^n$ em torno de β^0 com $n + 1$ pontos. Existem diferentes formas para gerar um simplex inicial e o seu tamanho pode influenciar a solução a ser obtida. O método proposto emprega a abordagem clássica para gerar \mathcal{S}^0 conforme a Equação 5.1:

$$\begin{aligned} \beta_{\mathbf{0}} &= \beta^0 \quad e \\ \beta_{\mathbf{k}} &= \beta_{\mathbf{0}} + s e_k \quad k \in \{1, 2, \dots, n\}, \end{aligned} \tag{5.1}$$

onde s é o tamanho do passo que determina o tamanho do simplex e $e = \{1, 1, \dots, 1\} \in \mathbb{R}^n$ é o vetor diagonal com tamanho \sqrt{n} . Assim, o simplex inicial \mathcal{S}^0 tem todas as bordas com o mesmo tamanho s .

Algoritmo 4: Procedimento HS para encontrar β

Entrada: Um conjunto de treinamento λ -rotulado \mathcal{Z}_1 , um conjunto validação \mathcal{Z}_v , um modelo MRF \mathcal{M} , Taxa de Consideração da Memória Harmônica (HMCR), Taxa de Ajuste de Som (PAR), Largura de banda (BW), número de iterações da Busca Harmônica T , número de harmonias U e número de iterações do ICM n .

Saída: Valor de β que maximiza a acurácia do OPF sobre o conjunto validação.

- 1 Inicializa a MH com U harmonias.
- 2 Treina OPF usando \mathcal{Z}_1 .
- 3 Classifica \mathcal{Z}_v gerando uma estimativa inicial \mathcal{L} .
- 4 **para cada iteração i ($i = 1, \dots, T$) faça**
 - 5 $min \leftarrow$ encontre a harmonia com a pior acurácia na MH.
 - 6 $r \leftarrow$ Random[0,1].
 - 7 **se (HMCR > r) então**
 - 8 $h \leftarrow$ Random[1, U].
 - 9 $X_{new} \leftarrow$ MH[h].
 - 10 $r \leftarrow$ Random[0,1].
 - 11 **se (PAR > r) então**
 - 12 $X_{new} \leftarrow$ afina(X_{new} , BW).
 - 13 **senão**
 - 14 $X_{new} \leftarrow$ Cria uma nova harmonia aleatoriamente.
 - 15 Cria novas instâncias \mathcal{Z}'_1 e \mathcal{Z}'_v estendendo \mathcal{Z}_1 e \mathcal{Z}_v com a frequência dos rótulos dos vizinhos dado por \mathcal{L} .
 - 16 **para cada iteração j ($j = 1, \dots, n$), faça**
 - 17 $[\mathcal{Z}'_1, \mathcal{Z}'_v] \leftarrow \mathcal{M}(\mathcal{Z}'_1, \mathcal{Z}'_v, X_{new}, \mathcal{L})$.
 - 18 Treina OPF usando \mathcal{Z}'_1 .
 - 19 Classifica \mathcal{Z}'_v gerando o mapa de rótulos atualizada \mathcal{L}' , e armazena a acurácia em f_i .
 - 20 $\mathcal{L} \leftarrow \mathcal{L}'$, $\mathcal{Z}'_1 \leftarrow \mathcal{Z}'_1$ e $\mathcal{Z}'_v \leftarrow \mathcal{Z}'_v$.
 - 21 **se ($f_i > \text{MH}[min]$) então**
 - 22 $\text{MH}[min] \leftarrow X_{new}$.
- 23 $\beta \leftarrow \max(\vec{f})$.

Depois da construção do simplex \mathcal{S}^i , o método inicia o processo iterativo para determinar um ponto estacionário β^* . O primeiro passo é calcular o valor de $p_k(y_{ij} = m|\eta_{ij}, \beta_k)$ $\forall 0 \leq k \leq n$ de cada instância e $p_k(y_{ij} = m|\eta_{ij}, \beta_k)$ é adicionado ao vetor de característica da instância. A seguir, é executada a classificação contextual com OPF-MRF para determinar os índices w , v e b , que representam os índices do pior, segundo pior e do melhor resultado de classificação, respectivamente, de acordo com a taxa de acerto Acc obtida pela classificação contextual. Finalmente, é calculado o centróide $c = \frac{1}{n} \sum_{k \neq w} \beta_j$ de todos os pontos amostrais, exceto o pior resultado.

Além disso, o ponto de reflexão $\beta_r = \mathbf{c} + \vartheta(\mathbf{c} - \beta_w)$ é calculado como segue: se $Acc_v < Acc_r \leq Acc_b$, então a instância do simplex β_w é substituída por β_r e a iteração termina. Caso contrário, se $Acc_r > Acc_b$, o ponto de expansão $\beta_e = c + \varsigma(\beta_r - c)$ e seu valor $Acc_e = Acc(\beta_e)$ são calculados. Se $Acc_e > Acc_r$, então β_e é selecionado e β_w é descartado; caso contrário, β_r é aceito e β_w é descartado. Agora, se $Acc_r \leq Acc_v$, o ponto de contração β_c (Equação 5.2) é calculado usando o melhor valor entre β_r e β_v :

$$\begin{aligned} \beta_c &= \mathbf{c} + \gamma(\beta_r - \mathbf{c}) && \text{se}(Acc_w \leq Acc_r \leq Acc_v) \\ \beta_c &= \mathbf{c} + \gamma(\beta_w - \mathbf{c}) && \text{se}(Acc_r < Acc_w). \end{aligned} \quad (5.2)$$

onde $\beta_{\mathbf{brv}}$ denota o ponto com o valor mais alto entre β_r e β_v , isto é, $\beta_{\mathbf{brv}} = \beta_r$ se $Acc_r \geq Acc_w$; e $\beta_{\mathbf{brv}} = \beta_w$ caso contrário. Se $Acc_e \geq Acc_{brv}$, β_c é aceito; caso contrário, é necessário criar um novo simplex, que pode ser calculado atualizando os vértices como a seguir:

$$\beta_j = \beta_j + \rho(\beta_j - \beta_b), \quad \forall j \neq b, \quad (5.3)$$

onde $j = 1, 2, \dots, n$. O processo iterativo é repetido até que o número máximo de iterações for alcançado ou se o critério de convergência for satisfeito. Note que o método Nelder-Mead possui os seguintes parâmetros: ϑ , γ , ρ e ς .

5.3.2 Método do Gradiente Descendente

O método do Gradiente Descendente é um método de otimização para minimizar/maximizar a função-objetivo a partir de um estado inicial, o qual o gradiente da função indica a direção do maior crescimento da função. Embora tal método seja bastante simples, sua taxa de convergência é fortemente dependente do problema [107], uma vez que exige que o espaço de busca seja convexo [108, 109].

Seja $\beta^{k=0}$ o ponto inicial, a estimativa do próximo ponto é calculada como a seguir:

$$\beta_{k+1} = \beta_k + \alpha \nabla_{\beta}, \quad (5.4)$$

onde α é o passo de atualização e ∇_{β} o operador gradiente. O processo se repete até que não haja mais ganho significativo. O Algoritmo 5 apresenta o método para maximizar a taxa de acerto Acc obtida pela classificação contextual com β_k , onde $\|\cdot\|_2$ representa a norma euclidiana, ϵ é o erro médio quadrático e α é o passo de atualização. As recursões nas Linhas (7 – 13) e (17 – 23) implementam o método ICM. Note que Acc_k armazena a taxa de acerto obtida pela classificação de \mathcal{Z}_v^* da última iteração do método ICM. As Linhas 8 e 18 calculam $p(y_i = m|\beta_k)$, $m = 1, \dots, C$, da instância i e são adicionados como C novas características, onde C é o número de classes. A linha 15 determina o próximo ponto β_k que maximiza a taxa de acerto do OPF-MRF.

5.4 Aplicando uma etapa de pós-processamento

Embora os modelos Markovianos utilizados tenham apresentado resultados melhores do que os obtidos pela versão não-contextual, alguns artefatos ainda podem ser observados em áreas compostas por pequenas regiões rodeadas por amostras de diferentes classes. Uma alternativa para tentar eliminar tal efeito seria aplicar um filtro da moda como uma etapa de pós-processamento. Contudo, tal filtragem tende a perder alguns detalhes nas regiões de alta frequência e poderia degradar o resultado final, principalmente se o mesmo for aplicado consecutivamente. Assim, para evitar tal problema e também auxiliar na classificação contextual, foi proposta uma filtragem da moda, mas que somente é aplicada quando o valor da moda atinge um valor mínimo, sendo apresentada em [110]. Ou seja, dada uma vizinhança de um dado *pixel*, é definida a classe mais frequente l através do cálculo da frequência absoluta de cada classe. A filtragem só será aplicada se $n > M$, onde n é o número de *pixels* classificados como sendo da classe l , e M o valor mínimo para que a filtragem seja aplicada. Tal procedimento é aplicado sobre o mapa de rótulos obtido a cada iteração para que as correções obtidas com a filtragem sejam propagadas ao longo da classificação contextual. A Figura 5.3 ilustra esse procedimento.

5.5 Classificação contextual com estimativa de parâmetros em uma imagem dividida em blocos

Imagens de alta resolução frequentemente contém uma diversidade de canais espectrais, aumentando a quantidade de informação a ser processada e, portanto, necessitando de mais memória e técnicas de processamento eficientes. Em classificação de imagens de grande escala, principalmente na etapa de treinamento, o processo de aprendizado deve lidar com uma enorme quantidade de dados, tornando-os computacionalmente inviáveis. Uma vez que encontrar parâmetros baseados em heurísticas frequentemente necessitam

Algoritmo 5: Procedimento GD para encontrar β

Entrada: Um conjunto de treinamento λ -rotulado \mathcal{Z}_1 , um conjunto validação \mathcal{Z}_v , um modelo MRF \mathcal{M} , passo de atualização α e número de iterações do ICM g .

Saída: Valor de β que maximiza a acurácia do OPF sobre o conjunto validação.

- 1 Treina OPF usando \mathcal{Z}_1 .
 - 2 Classifica \mathcal{Z}_v gerando uma estimativa inicial \mathcal{L} .
 - 3 $k \leftarrow 0$.
 - 4 Escolha do ponto inicial β_k .
 - 5 Cria novas instâncias \mathcal{Z}'_1 e \mathcal{Z}'_v estendendo \mathcal{Z}_1 e \mathcal{Z}_v .
 - 6 $q \leftarrow 0$.
 - 7 **enquanto** $q < g$ **faça**
 - 8 $[\mathcal{Z}_1^*, \mathcal{Z}_v^*] \leftarrow \mathcal{M}(\mathcal{Z}'_1, \mathcal{Z}'_v, \beta_k, \mathcal{L})$.
 - 9 Treina OPF usando \mathcal{Z}_1^* .
 - 10 Classifica \mathcal{Z}_v^* gerando o mapa de rótulos atualizada \mathcal{L}' ,
 - 11 e armazena a acurácia em Acc_k .
 - 12 $\mathcal{L} \leftarrow \mathcal{L}'$, $\mathcal{Z}'_1 \leftarrow \mathcal{Z}_1^*$ e $\mathcal{Z}'_v \leftarrow \mathcal{Z}_v^*$.
 - 13 $q \leftarrow q + 1$.
 - 14 **enquanto** $\|\nabla_{Acc_k}\|_2 > \epsilon$ **faça**
 - 15 $\beta_{k+1} \leftarrow \beta_k + \alpha \nabla_k$.
 - 16 $q \leftarrow 0$.
 - 17 **enquanto** $q < g$ **faça**
 - 18 $[\mathcal{Z}_1^*, \mathcal{Z}_v^*] \leftarrow \mathcal{M}(\mathcal{Z}'_1, \mathcal{Z}'_v, \beta_k, \mathcal{L})$.
 - 19 Treina OPF usando \mathcal{Z}_1^* .
 - 20 Classifica \mathcal{Z}_v^* gerando o mapa de rótulos atualizada \mathcal{L}'
 - 21 e armazena a acurácia em Acc_k .
 - 22 $\mathcal{L} \leftarrow \mathcal{L}'$, $\mathcal{Z}'_1 \leftarrow \mathcal{Z}_1^*$ e $\mathcal{Z}'_v \leftarrow \mathcal{Z}_v^*$.
 - 23 $q \leftarrow q + 1$.
 - 24 $k \leftarrow k + 1$.
 - 25 $\beta \leftarrow \max(\vec{Acc})$.
-

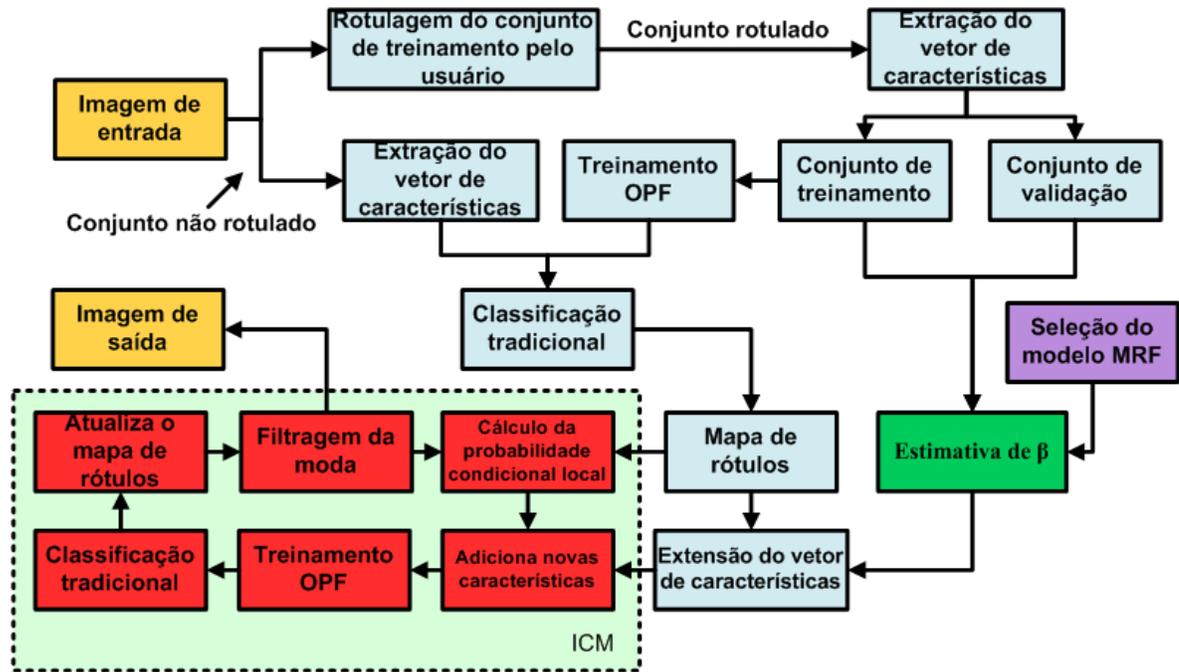


Figura 5.3: Metodologia proposta para classificação contextual com pós-processamento.

de diversas execuções da fase de treinamento, o processo de aprendizado poderia consumir uma considerável quantidade de tempo. Para classificação contextual, em particular, os resultados tendem a ser promissores, especialmente quando os parâmetros são corretamente configurados. Porém, para determinar valores de parâmetros que maximizam a taxa de acerto através de técnicas de otimização, necessitam do cálculo da função “fitness” diversas vezes, sendo necessário um alto custo computacional.

Embora os computadores tenham evoluído ao longo dos anos, ainda exigem grande quantidade de memória para o processamento de imagens de alta resolução. Uma forma de superar tal limitação é utilizar a estratégia “dividir e conquistar”, onde a imagem é dividida em blocos menores para posterior classificação de cada bloco individualmente. Contudo, tal processo de classificação pode falhar ao tentar manter a hipótese de suavidade, especialmente nas regiões de fronteira entre os blocos, quando a imagem é reconstruída.

Tal ideia poderia auxiliar na definição do parâmetro pelos métodos de otimização. Assim, é esperado que a abordagem seja mais rápida e tenha maior taxa de acerto em relação à classificação contextual com a imagem toda, pois o valor do parâmetro β pode variar de acordo com o bloco da imagem. Contudo, o resultado pode não ser satisfatório dependendo do número de divisões da imagem, uma vez que a amostragem de treinamento

também irá diminuir. Assim, foi proposta uma abordagem que encontra valores de β para diferentes regiões das imagens. Tal abordagem é parte integrante do artigo *A Block-Based Markov Random Field Model Estimation for Contextual Classification Using Optimum-Path Forest* que foi aceito para publicação no *IEEE International Symposium on Circuits and Systems*.

Seja \mathcal{D} o conjunto de dados compostos de vetores de características extraídos dos *pixels* da imagem \mathcal{I} . Assuma que \mathcal{I} possa ser dividida em r blocos, tal que \mathcal{I}_s e \mathcal{D}_s é um bloco da imagem e seu respectivo conjunto de dados, respectivamente, $s = 1, 2, \dots, r$. Além disso, assumamos que \mathcal{D}_s possa ser particionado em $\mathcal{D}_s = \mathcal{D}_s^{tr} \cup \mathcal{D}_s^v \cup \mathcal{D}_s^{ts}$, correspondendo aos conjuntos de treinamento, de validação e de teste, respectivamente. A ideia é considerar diferentes MRFs para cada bloco e, assim, resultando em um β_s para cada bloco da imagem. Assim, a abordagem estima β_s como uma tarefa de otimização, os quais \mathcal{D}_s^{tr} e \mathcal{D}_s^v são utilizados para guiar os algoritmos de otimização, ou seja, o método OPF-MRF é treinado com \mathcal{D}_s^{tr} e sua acurácia sobre \mathcal{D}_s^v é usado como a função “*fitness*”. A Figura 5.4 apresenta a abordagem proposta para estimar o parâmetro β , classificar cada bloco da imagem separadamente e, finalmente, é efetuada a junção dos mapas de rótulos resultantes de cada bloco da imagem para gerar a imagem rotulada (“módulo laranja” da Figura 5.4).

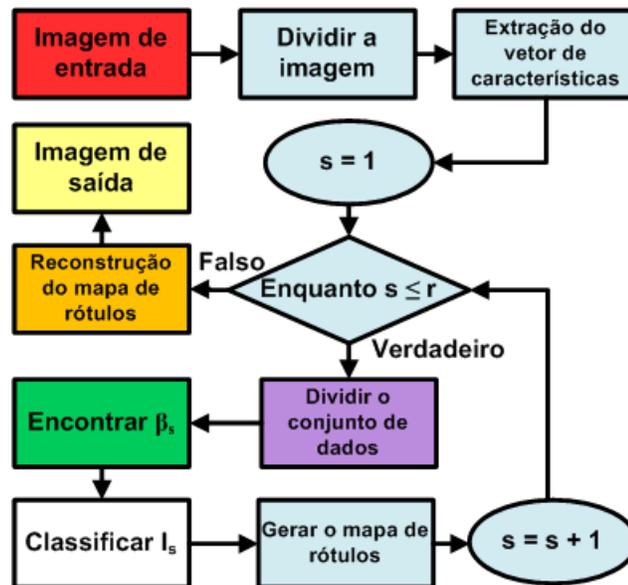


Figura 5.4: Metodologia para classificação contextual por divisão da imagem em blocos.

No capítulo seguinte, são apresentados os resultados experimentais das abordagens de classificação contextual propostas para a técnica OPF utilizando Campos Aleatórios

5.5. *Classificação contextual com estimativa de parâmetros em uma imagem dividida em blocos*

Markovianos para integração da informação contextual discutidas até aqui.

Capítulo 6

Simulações e Resultados

Com o intuito de validar as abordagens propostas para classificação contextual utilizando OPF-MRF, os experimentos foram conduzidos empregando OPF-MRF sobre imagens de ressonância magnética do cérebro humano (*Magnetic Resonance Image - MRI*) do conjunto de dados IBSR¹, bem como imagens de sensoriamento remoto (Figuras 6.1 e 6.3a-d) e imagem de radar (Figuras 6.2 e 6.3e).

A Tabela 6.1 apresenta uma descrição detalhada a respeito do número de classes das imagens de sensoriamento remoto para cada cobertura de área, isto é, Itatinga (Figuras 6.3a e 6.3b) e Duque de Caxias (Figuras 6.3c and 6.3d). A Tabela 6.2 fornece as cores associadas à cada classe nas imagens de sensoriamento remoto e de radar. Além disso, as Tabelas 6.3 e 6.4 mostram o número de amostras por classe para as coberturas de área de Itatinga e Duque de Caxias, respectivamente.

Com relação à taxa de reconhecimento, foi empregada a medida de acurácia proposta por Papa et al. [6], a qual é similar ao índice Kappa [112], porém sendo mais restritivo. Se existem duas classes, por exemplo, com grande diferença de tamanho e um classificador sempre atribui o rótulo da maior classe, sua acurácia irá cair drasticamente devido à alta taxa de erro na menor classe. A acurácia é calculada levando em consideração que as classes poderiam ter diferentes tamanhos em \mathcal{Z}_v (definição similar é aplicada para \mathcal{Z}_2), sendo definida como:

$$e_{i,1} = \frac{FP_i}{|\mathcal{Z}_v| - |\mathcal{Z}_v^i|} \quad (6.1)$$

e

$$e_{i,2} = \frac{FN_i}{|\mathcal{Z}_v^i|}, i = 1, 2, \dots, C, \quad (6.2)$$

onde $|\mathcal{Z}_v^i|$ é o número de instâncias em \mathcal{Z}_v que são da classe i , FP_i e FN_i são os falsos

¹A descrição de dados foi fornecida pelo “*Center for Morphometric Analysis*” do Hospital Geral de Massachusetts e disponível em <http://www.cma.mgh.harvard.edu/ibsr/>.

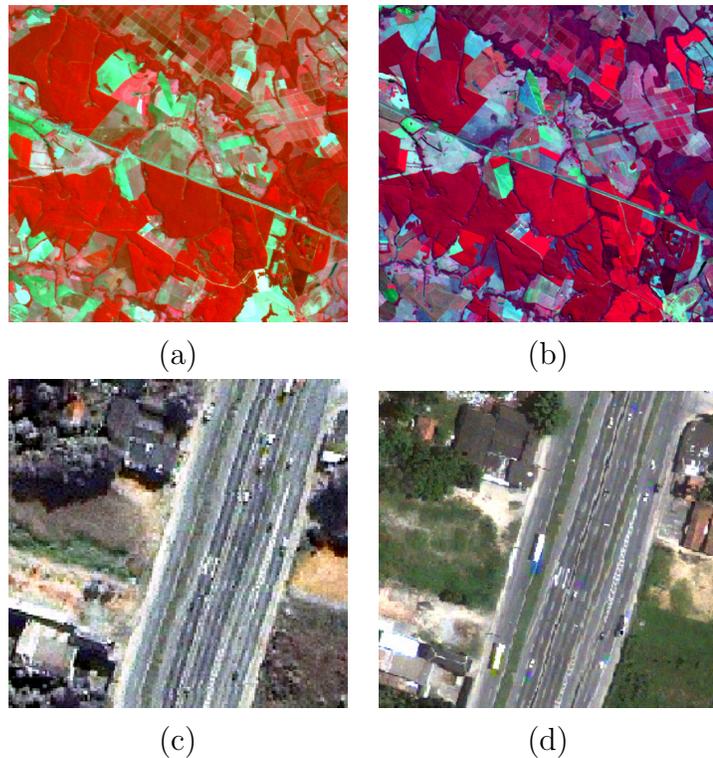


Figura 6.1: Imagens de satélite utilizadas nos experimentos: cobrindo a área de Itatinga-SP, Brasil por (a) CBERS-2B CCD (20m) sensor (R2G3B4) e (b) Landsat 5 TM (30m) sensor (R4G3B5), e cobrindo a área de Duque de Caxias-RJ, Brasil por (c) Ikonos-2 MS sensor (R4G3B2) e (d) Geoeye sensor (R5G4B3). As imagens CBERS-2B e Landsat 5 TM possuem 526×492 *pixels*, e as imagens Ikonos-2 MS e Geoeye possuem 258×250 e 268×250 *pixels*, respectivamente. Note que as imagens Ikonos-2 MS e Geoeye foram obtidas através de um processo de fusão entre imagens correspondentes dos sensores MS (4m) e PAN (1m) usando o método *pan-sharpening*. A imagem final tem uma resolução espacial de 1m.

positivos e falsos negativos para a classe i , respectivamente. Isto é, FP_i corresponde ao número de instâncias de outras classes que foram classificadas como sendo de outras classes em \mathcal{Z}_v . Os termos $e_{i,1}$ e $e_{i,2}$ são então utilizados para definir o total de erros da classe i :

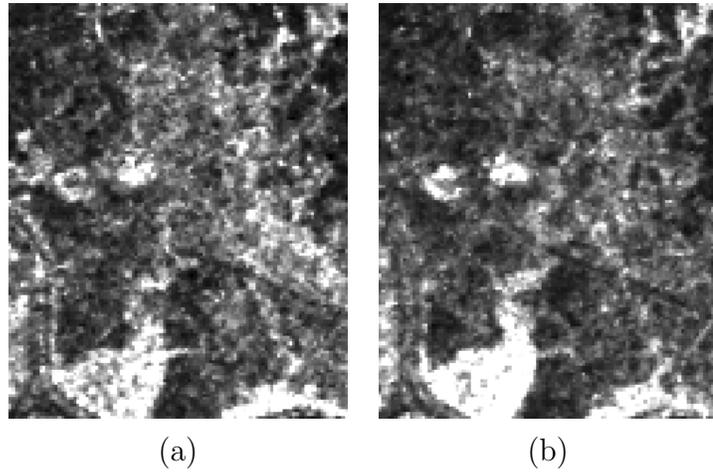


Figura 6.2: Imagem do radar ALOS-PALSAR utilizada nos experimentos com polarização (a) HH e (b) HV. Tal sensor possui capacidade de imageamento com múltipla polarização (HH, VV, HV e VH; H = polarização horizontal, V = polarização vertical, 1^a letra = radiação emitida, 2^a letra = radiação recebida) [111].

$$E_i = e_{i,1} + e_{i,2}. \quad (6.3)$$

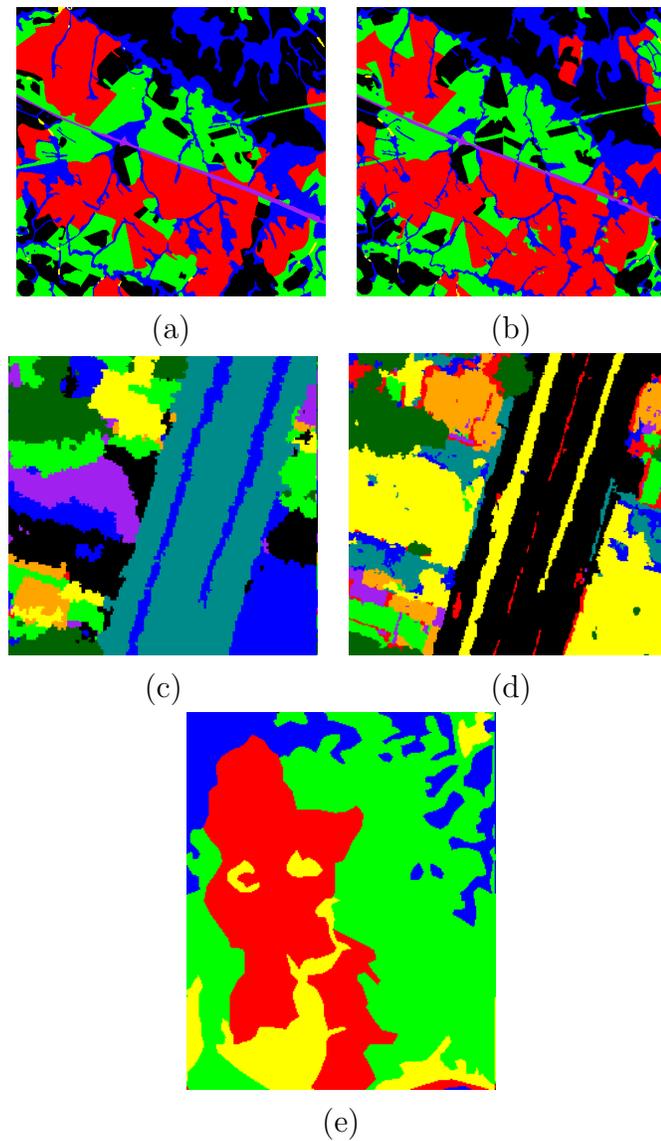


Figura 6.3: Imagens rotuladas utilizadas nos experimentos: (a), (b), (c) e (d) referem-se às imagens apresentadas nas Figuras 6.1a, 6.1b, 6.1c e 6.1d, respectivamente, e (e) refere-se às imagens apresentadas nas Figuras 6.2a e 6.2b.

Finalmente, a acurácia Acc então é definida como a seguir:

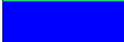
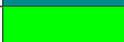
Cobertura de área	Sensor	#	Classes de uso do solo
Itatinga	CBERS-2B	6	pastagens, reflorestamento, culturas, estradas, lagos e mata nativa
Itatinga	Landsat 5 TM	6	pastagens, reflorestamento, culturas, estradas, lagos e mata nativa
Duque de Caxias	Ikonos-2 MS	8	estradas, pastagens, solo exposto úmido, cobertura de árvores, cobertura de tons claros, cobetura de tons escuros, solo exposto com tons claros e cobertura de sombras
Duque de Caxias	Geoeye	9	estradas, pastagens, solo exposto úmido, cobertura de árvores, cobertura de tons claros, cobertura de tonalidade média, cobetura de tons escuros, solo exposto com tons claros e cobertura de sombras

Tabela 6.1: Descrição dos rótulos para cada cobertura de área.

$$Acc = 1 - \frac{\sum_{i=1}^C E_i}{2C}. \quad (6.4)$$

É importante destacar que todos os experimentos foram conduzidos em um computador equipado com um processador Intel CPU i3[®] 540 3.07GHZ, 4GB RAM e sistema operacional Ubuntu 10.04 LTS.

Tabela 6.2: Cores associadas com cada classe de cobertura do solo.

Sensor	Cores	Classes de cobertura do solo
CBERS-2B e Landsat 5		Pastagens
		Mata nativa
		Culturas
		Reflorestamento
		Lagos
		Estradas
Ikonos-2 MS		Cobertura de sombras
		Pastagens
		Solo exposto com tons claros
		Cobertura de tonalidade média
		Cobertura de tons escuros
		Solo exposto úmido
		Cobertura de árvores
		Cobertura de tons claros
		Estradas
Geoeye		Cobertura de tonalidade média
		Solo exposto úmido
		Estradas
		Cobertura de sombras
		Pastagens
		Cobertura de tons claros
		Cobertura de árvores
		Cobertura de tons escuros
		Solo exposto claro
ALOS-PALSAR		Áreas planas com solo exposto e áreas secas
		Pastagens com baixo nível de umidade
		Pastagens com nível de umidade médio
		Pântano e vegetação com altos níveis de umidade

Classe	CBERS-2B	LANDSAT-5 TM
culturas	18407	13356
mata nativa	13779	12387
lagos	142	123
pastagens	14794	17567
reflorestamento	18825	22581
estradas	803	736

Tabela 6.3: Número de amostras para a cobertura de área de Itatinga.

Classe	Ikonos-2 MS	Geoeye
cobertura de árvores	5914	6132
cobertura de sombras	6481	2822
pastagens	12054	19370
cobertura de tons escuros	3578	5073
estradas	22871	22924
solo exposto úmido	4417	2380
solo exposto com tons claros	7400	4490
cobertura de tons claros	1785	1026
cobertura de tons médios	-	2783

Tabela 6.4: Número de amostras para a cobertura de área de Duque de Caxias.

6.1 Classificação Contextual com OPF

6.1.1 Classificação contextual de imagens de ressonância magnética utilizando o modelo de Potts

Nesta seção, são apresentados os resultados experimentais obtidos pelo classificador OPF para classificação contextual utilizando Campos Aleatórios Markovianos (OPF-MRF). Esta seção corresponde ao artigo *OPF-MRF: Optimum-Path Forest and Markov Random Fields for Contextual-based Image Classification* [87], o qual foi apresentado no “15th International Conference on Computer Analysis of Images and Patterns” (CAIP/ 2013), e propôs o classificador OPF-MRF.

Para avaliar a técnica de classificação OPF baseada em contexto, foram conduzidos experimentos sobre imagens de ressonância magnética (Magnetic Resonance Image - MRI). A seguir, é descrita a metodologia aplicada para avaliação do método proposto. Com relação à extração de características, cada *voxel* é representado pelo brilho do *voxel* de seus 26-vizinhos, o qual é referido aqui como o método tradicional ou *baseline* (isto é, in-

formação não contextual é empregada neste caso), gerando um total de 26 características para cada instância. Com o intuito de avaliar a robustez do OPF-MRF, foram executados experimentos em 4 imagens do IBSR (a Figura 6.4a apresenta um exemplo dessa base, bem como sua versão rotulada na Figura 6.4b). Para tal, cada *voxel* é uma instância usada para compor o conjunto de dados de sua imagem correspondente, que é particionada em conjunto de treinamento e teste. Foi utilizado o procedimento de validação cruzada com 10 rodadas com proporções de 0.05% e 99.95% para os conjuntos de treinamento e teste, respectivamente. Além disso, o campo aleatório foi modelado empregando o modelo de Potts e a probabilidade condicional local maximizada com o método ICM utilizando possíveis valores de β , os quais foram definidos empiricamente.

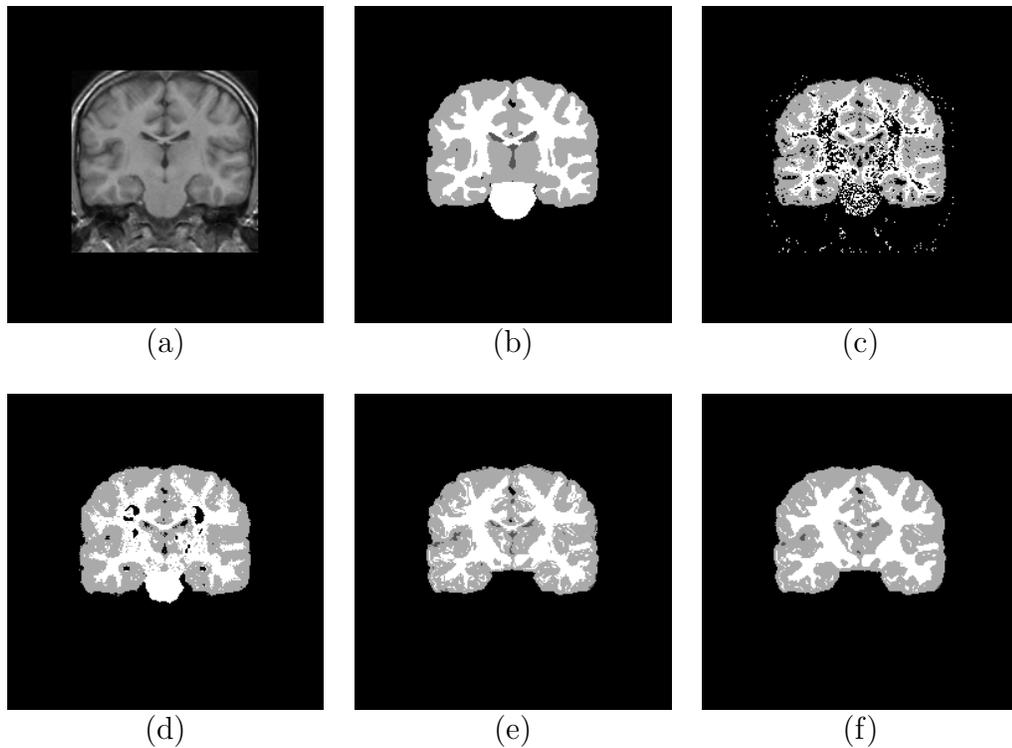


Figura 6.4: (a) IBSR-01 fatia 58, (b) dados de referência, (c) *baseline* da classificação (OPF tradicional), (d) classificação com OPF-MRF, (e) Abordagem proposta em [113] e (f) OPF-MRF aplicado usando técnicas de pré-processamento proposta em [113].

As Figuras 6.5 e 6.6 mostram os resultados do κ de Cohen [112] para cada imagem usando três diferentes valores para o β (Equação 4.1), o qual controla a quantidade de informação contextual utilizada no processo de aprendizado. Para cada imagem, foram executadas 10 rodadas para compor o procedimento de validação cruzada. Portanto, o valor #1 no eixo horizontal para um dado β , por exemplo, corresponde ao valor médio

do OPF-MRF sobre 10 rodadas de validação cruzada durante a iteração #1 do algoritmo proposto (foi empregado $T = 10$ iterações). É possível observar que o OPF-MRF (Figura 6.4d) tem fornecido melhores resultados do que a abordagem tradicional (*baseline*), isto é, OPF sem informação contextual (Figura 6.4c). É importante notar que, removendo o conteúdo intra-craniano e o tronco cerebral em uma imagem MRI da cabeça (processo referido como *skull stripping*), e aplicando técnicas para correção de homogeneidade, a técnica proposta obteve melhores resultados (ver Figura 6.4 (e) e (f) - com κ igual a 0.8812 e 0.8995, respectivamente).

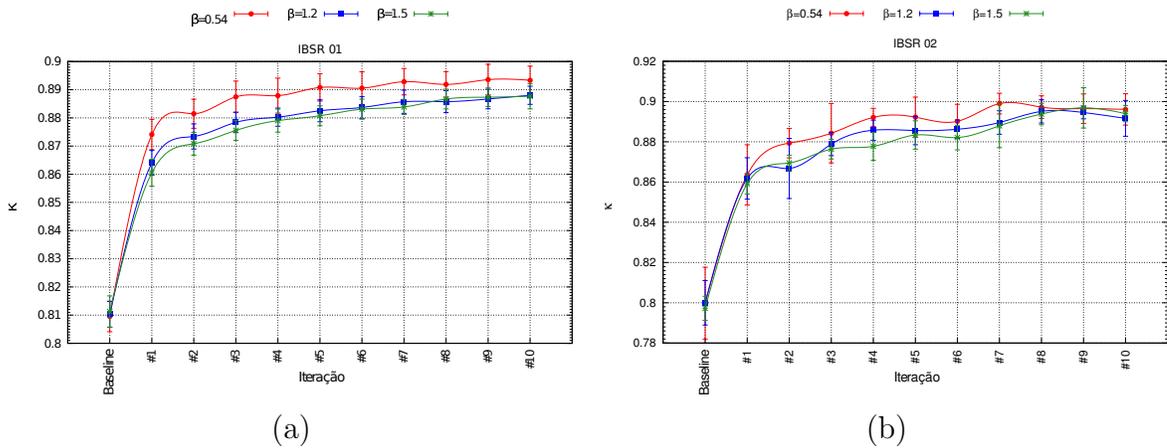


Figura 6.5: Valores globais de κ do Cohen obtida através de validação cruzada para os conjuntos de dados IBSR-01 e IBSR-02.

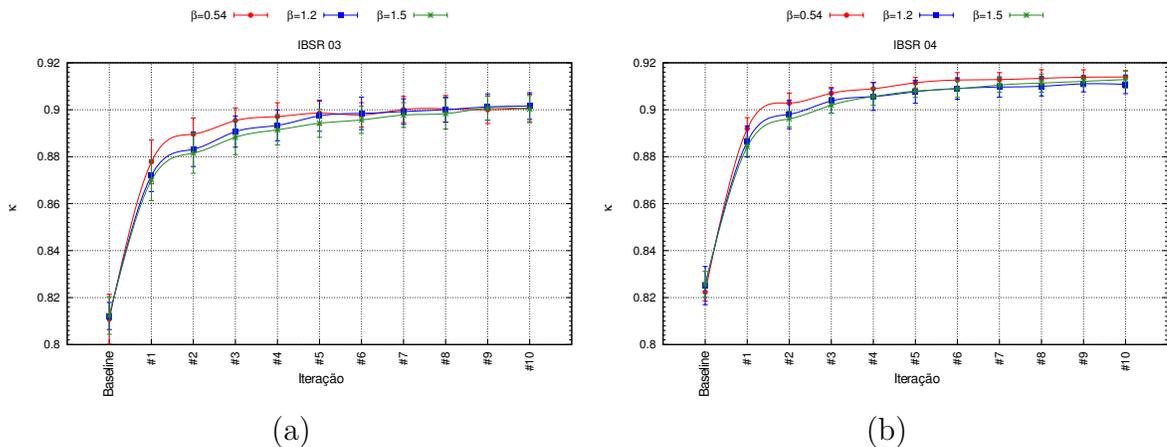


Figura 6.6: Valores globais de κ do Cohen obtida através de validação cruzada para os conjuntos de dados IBSR-03 e IBSR-04.

A abordagem proposta pôde fornecer uma solução mais homogênea para o mapa de classificação. Também deve ser observado que o método é altamente dependente da

solução inicial. O melhor resultado foi obtido com $\beta = 0.54$, o que poderia indicar que este valor é o que melhor representa a relação custo/benefício entre a observação e um conhecimento *a priori*.

Na seção seguinte, são avaliados outros modelos Markovianos sobre imagens de sensoriamento remoto em classificação contextual com OPF-MRF.

6.1.2 Sob a Influência de Modelos Markovianos para Classificação usando Floresta de Caminhos Ótimos Baseado em Contexto

Com o intuito de comparar as diferentes funções para modelagem de Campos Aleatórios Markovianos no contexto do classificador OPF-MRF, esta seção descreve a metodologia aplicada para classificação contextual, bem como os resultados experimentais obtidos utilizando diferentes modelos Markovianos. Esta seção diz respeito ao artigo *On the Influence of Markovian Models for Contextual-based Optimum-Path Forest Classification* [114], que foi aceito para publicação ao 19th *Iberoamerican Congress on Pattern Recognition* (CIARP/2014). O trabalho foi validado no contexto de classificação do uso do solo em imagens de sensoriamento remoto obtidas pelos satélites CBERS-2B cobrindo a área de Itatinga-SP, Brasil (Figura 6.1a), e Ikonos-2 MS cobrindo a área de Duque de Caxias-RJ, Brasil (Figura 6.1b), bem como suas versões rotuladas nas Figuras 6.3a e 6.3b, respectivamente.

Metodologia

Os experimentos foram conduzidos como a seguir: para cada imagem de satélite, foi empregada uma amostra de 5% (*pixels*) para compor o conjunto de treinamento e os 95% remanescentes para compor o conjunto de teste. Note que tais valores foram empiricamente escolhidos. Além disso, os vetores de características são compostos pelos valores RGB dos 8 *pixels* vizinhos e a posição do *pixel* central (x, y), resultando em um total de 28 características. Com relação à implementação do OPF, foi utilizada a biblioteca LibOPF [115]. Também foi estimado o parâmetro β (Equação 4.1) usando uma busca pseudo-exaustiva (*Brute Force* - BF) dentro da faixa $[0, \beta_{max}]$ com passos de 0.1. Esta abordagem foi empregada para simular uma referência “ótima” para a taxa de reconhecimento usando diferentes valores de β . Assim, o valor de β que maximiza a taxa de reconhecimento sobre o conjunto de teste é então utilizado para calcular as acurácias médias². Esta estratégia foi aplicada para os modelos Markovianos de Potts, GIMLL-L1, GIMLL-L2 e GMRF (Equações 4.1, 4.4, 4.3 e 4.5, respectivamente). Finalmente, foram

²Não foi utilizado o conjunto de validação para determinar o valor de β .

utilizadas 10 iterações para o método ICM, sendo calculada a acurácia média para cada iteração sobre um procedimento de validação cruzada com 5 rodadas usando o valor de β encontrado por BF.

Resultados Experimentais

Esta seção apresenta os resultados experimentais usando a metodologia descrita na Seção 6.1.2. Com intuito de avaliar o comportamento do OPF-MRF sob diferentes modelos Markovianos, foram comparados quatro modelos no contexto de classificação do uso do solo utilizando duas imagens de satélite com diferentes resoluções espaciais (Figuras 6.1a e 6.1b). A Figura 6.7 apresenta as imagens classificadas usando as técnicas OPF e OPF-MRF.

É possível observar que OPF-MRF tem maior acurácia do que o OPF padrão para ambas imagens, uma vez que os resultados do OPF apresentaram o efeito de sal-e-pimenta em algumas regiões das imagens (regiões central e inferior esquerda na imagem CBERS-2B (Figura 6.7a), e superior esquerda e inferior esquerda na imagem Ikonos-2 (Figura 6.7f)). Tais regiões foram suavizadas por todos modelos Markovianos considerando o OPF-MRF, uma vez que a informação contextual foi utilizada. No entanto, é visualmente difícil notar diferenças entre os quatro modelos Markovianos. Com intuito de responder a esta questão, a Figura 6.8 mostra a curva de acurácia média ao longo das iterações para o OPF-MRF. Conforme mencionado acima, foi utilizado o melhor resultado obtido a partir do valor do β que maximizou a taxa de reconhecimento do OPF-MRF para os modelos Potts, GIMLL-L1, GIMLL-L2 e GMRF. Note que a acurácia especificada na iteração #0 é a acurácia obtida a partir da classificação usando OPF padrão, isto é, onde nenhuma informação contextual é considerada.

Na Figura 6.8 é possível observar três importantes informações: (i) o OPF-MRF usando todos os modelos claramente aperfeiçoa o OPF padrão para ambas imagens (as taxas de reconhecimento na iteração #1 são melhores do que as taxas de reconhecimento na iteração #0), (ii) o OPF-MRF tem convergido ao longo das iterações para todos os modelos e (iii) o modelo de Potts superou os modelos GIMLL-L1, GIMLL-L2 e GMRF em ambas imagens. As primeiras duas observações também foram constatadas por Nakamura et al. [87] e Osaku et al. [98], sendo que a principal contribuição deste trabalho é avaliar se existe um aperfeiçoamento significativo considerando outros modelos Markovianos para classificação com o OPF baseado em contexto.

Uma possível razão por trás do melhor resultado obtido pelo modelo de Potts é devido ao fato de que tal modelo foi proposto para modelar situações quando as partículas de um dado campo aleatório assumem valores discretos. Na formulação do OPF-MRF, cada elemento (instância) assume um valor discreto representado pelo seu rótulo (um número

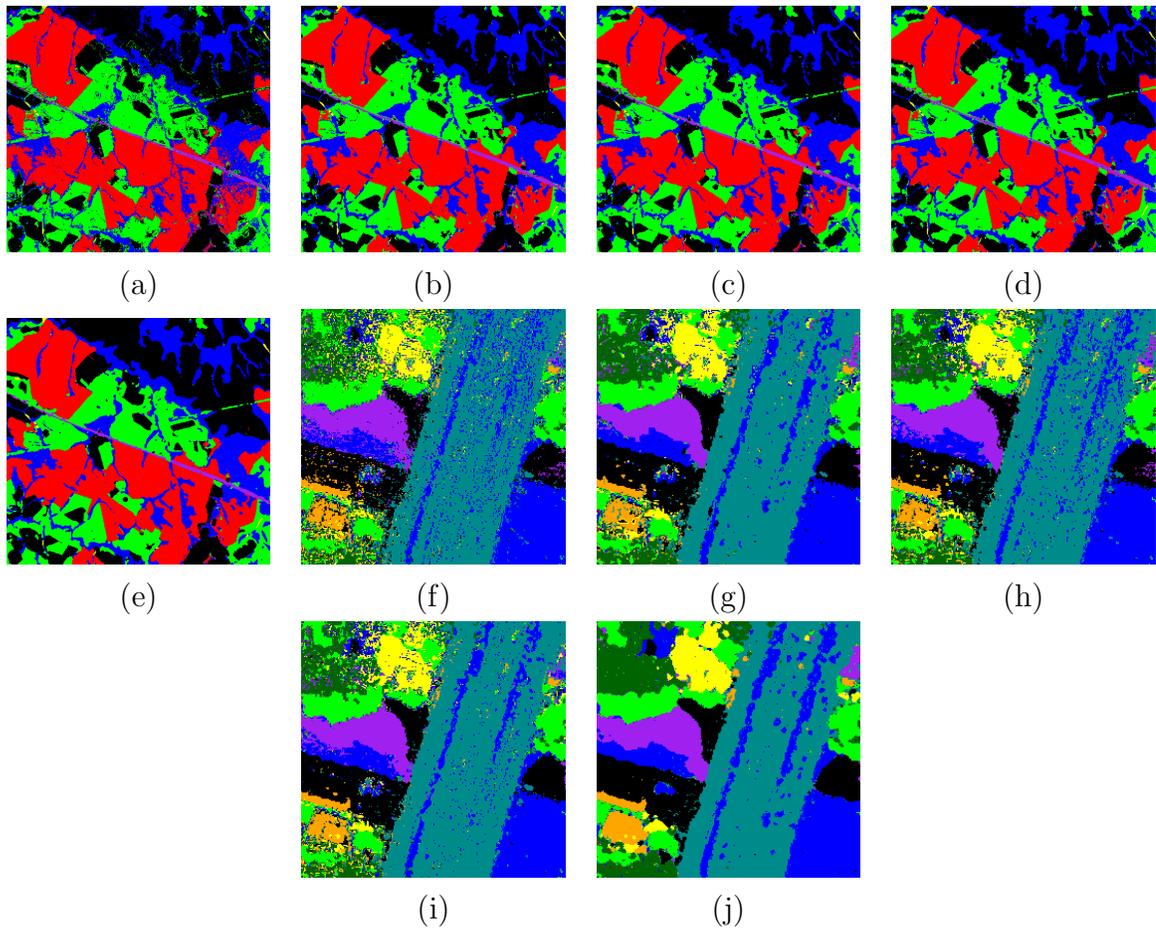


Figura 6.7: Imagem CBERS-2B classificada considerando o OPF tradicional em (a), e OPF-MRF com os modelos GIMLL-L1 (b), GIMLL-L2 (c), GMRF (d) e de Potts (e). Imagem Ikonos-2 classificada considerando o OPF tradicional OPF em (f), e OPF-MRF com os modelos GIMLL-L1 (g), GIMLL-L2 (h), GMRF (i) e de Potts (j).

natural entre $[1, 2, \dots, C]$, onde C é o número total de classes) e, portanto, favorecendo este modelo. Os modelos restantes, isto é, GIMLL-L1, GIMLL-L2 e GMRF, são mais adequados para lidar com situações em que as variáveis assumem valores contínuos.

Além dos modelos avaliados, foi avaliado o modelo baseado em teoria da informação proposto, cujos resultados serão apresentados na seção seguinte.

6.1.3 Um modelo Markoviano baseado em teoria da informação

Nesta seção são apresentados os resultados experimentais efetuando classificação contextual com OPF-MRF utilizando um modelo baseado na teoria da informação, cuja probabilidade condicional é dada pelo cálculo da Informação de Fisher Observada Local para o

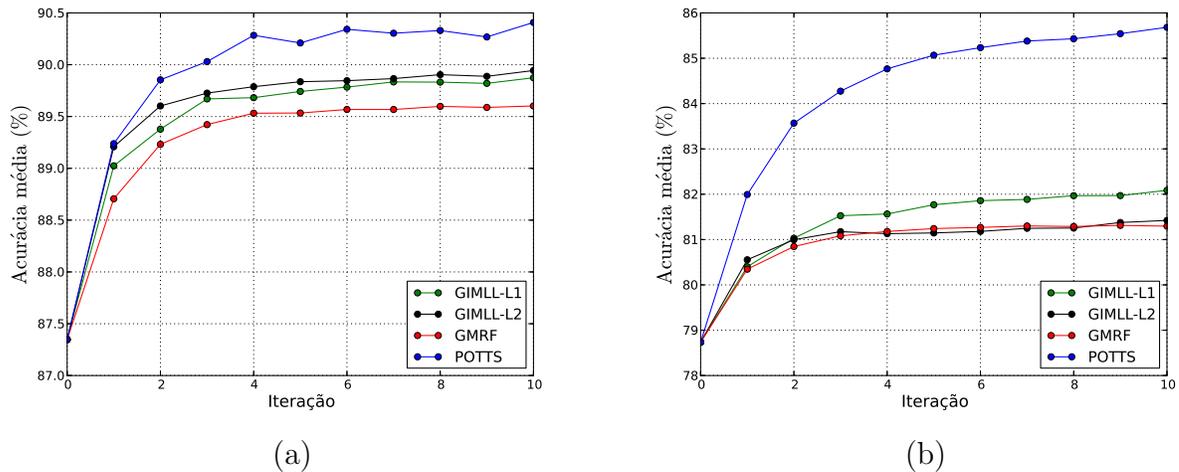


Figura 6.8: Curva de acurácia média obtida ao longo das iterações através de validação cruzada para as imagens (a) CBERS-2B e (b) Ikonos-2.

modelo de Potts e suas variantes para valores contínuos GIMLL-L1 e GIMLL-L2. Esses resultados foram submetidos para publicação no congresso *SIBGRAPI 2016 - Conference on Graphics, Patterns and Images*.

Para avaliar o modelo proposto, foram usadas as imagens de sensoriamento remoto CBERS-2B, Ikonos-2 MS e Geoeye (Figuras 6.1a, 6.1c e 6.1d, respectivamente). Para cada imagem, foram aplicadas uma amostragem de 5% (*pixels*) para o conjunto de treinamento, 15% para o conjunto de validação e 80% para o conjunto de teste. Além disso, foram avaliados dois vetores de características distintos, um composto pelos valores RGB dos 8 *pixels* vizinhos e a posição do *pixel* (i, j), dando um total de 26 características e outro composto pelos valores RGB do *pixel* e a posição do *pixel* central (i, j), dando um total de 5 características. Para classificação tradicional e ponto de partida, foi utilizada a técnica OPF. Para classificação contextual, foi utilizada a técnica OPF-MRF, cujo parâmetro β foi estimado executando 50 iterações com o método de busca pseudo-exaustiva sobre os conjuntos de treinamento e validação. Também foi utilizado um sistema de vizinhança 3×3 para o cálculo da função de energia e 10 iterações do método ICM.

Com o intuito de avaliar os modelos propostos, foi executada a classificação contextual com os modelos Markovianos e seus respectivos modelos baseados na teoria da informação sobre as imagens de sensoriamento remoto. Assim, as Tabelas 6.5 e 6.6 apresentam as acurácias médias, bem como os respectivos desvios padrão, utilizando vetores de características compostos por 5 e 26 características, respectivamente.

É possível verificar que a equação baseado na Informação de Fisher Observada Local (Equação 4.11) fornece informação contextual mais discriminante do que a Equação 4.1,

Tabela 6.5: Taxa de acurácia média usando diferentes modelos Markovianos com vetor de 5 características. A melhor acurácia de cada imagem está em negrito.

Técnica	CBERS2-B	Ikonos-2 MS	Geoeye
OPF (baseline)	87.35% \pm 0.47	85.82% \pm 0.36	88.03% \pm 0.30
OPF-MRF/Potts	90.93% \pm 0.24	88.64% \pm 0.67	90.18% \pm 0.39
OPF-MRF/GIMLL-L1	90.38% \pm 0.39	87.86% \pm 0.52	89.46% \pm 0.15
OPF-MRF/GIMLL-L2	90.33% \pm 0.24	87.87% \pm 0.52	89.47% \pm 0.19
OPF-MRF/Potts/Fisher	91.13% \pm 0.80	88.49% \pm 0.45	90.17% \pm 0.09
OPF-MRF/GIMLL-L1/Fisher	91.19% \pm 0.66	88.81% \pm 0.52	90.46% \pm 0.17
OPF-MRF/GIMLL-L2/Fisher	91.08% \pm 0.54	89.03% \pm 1.02	90.32% \pm 0.22

Tabela 6.6: Taxa de acurácia média usando diferentes modelos Markovianos com vetor de 26 características. A melhor acurácia de cada imagem está em negrito.

Técnica	CBERS2-B	Ikonos-2 MS	Geoeye
OPF (baseline)	87.17% \pm 0.44	78.74% \pm 0.15	83.16% \pm 0.15
OPF-MRF/Potts	90.15% \pm 0.70	85.40% \pm 0.42	88.30% \pm 0.34
OPF-MRF/GIMLL-L1	89.63% \pm 0.42	81.72% \pm 0.38	84.98% \pm 0.14
OPF-MRF/GIMLL-L2	89.80% \pm 0.45	81.10% \pm 0.16	85.23% \pm 0.25
OPF-MRF/Potts/Fisher	90.21% \pm 0.80	85.36% \pm 0.53	88.41% \pm 0.23
OPF-MRF/GIMLL-L1/Fisher	90.59% \pm 0.68	85.52% \pm 0.39	88.45% \pm 0.31
OPF-MRF/GIMLL-L2/Fisher	90.43% \pm 0.51	85.45% \pm 0.41	88.42% \pm 0.26

sendo as acurácias médias obtidas ligeiramente melhores para os modelos propostos. Particularmente, os modelos GIMLL-L1 e GIMLL-L2 apresentaram os melhores resultados, sendo que em alguns casos aumentaram cerca de 1% na acurácia média para vetor de 5 características e 4% para o vetor de 26 características.

Com base nos obtidos pela classificação com OPF, mostrados na Figura 6.9, é possível observar, nas Figuras 6.10 e 6.11, que a classificação contextual com OPF-MRF utilizando modelos baseados na Informação de Fisher Observada Local podem ser uma boa alternativa para refinar os resultados de classificação tradicional.

As seções seguintes apresentam os resultados obtidos pela classificação contextual com OPF-MRF, cujo problema de encontrar valores de β que maximizam a taxa de reconhecimento foi tratado como um problema de otimização.

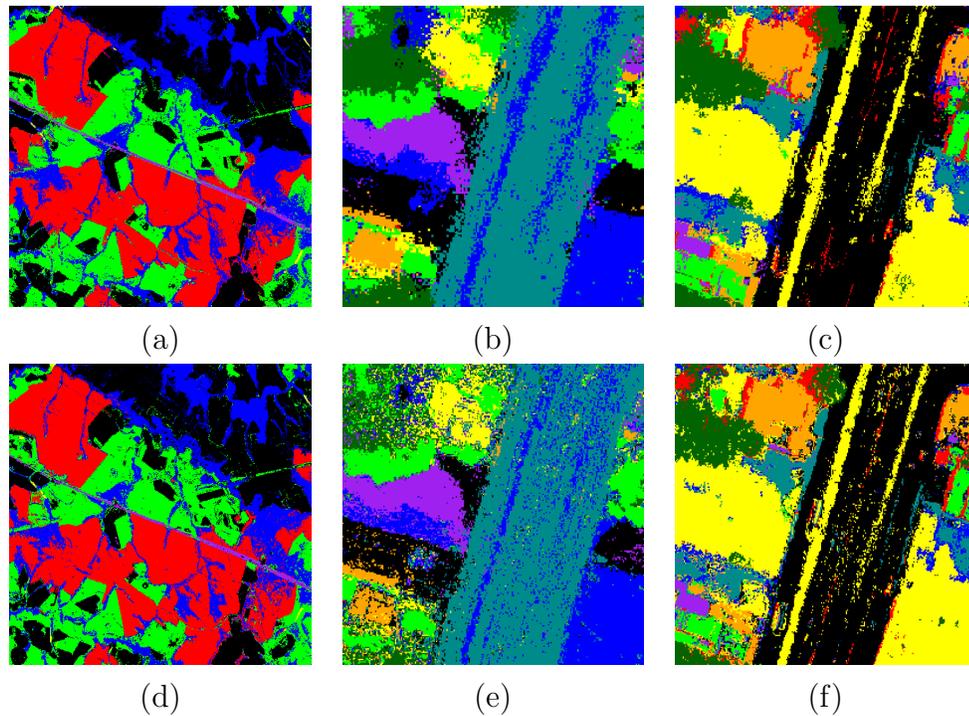


Figura 6.9: Imagens CBERS-2B, Ikonos-2 MS e Geoeye classificadas por OPF (*baseline*) utilizando: vetor de 5 características em (a), (b) e (c); e vetor de 26 características em (d), (e) e (f).

6.2 Estimando a Dependência Espacial Utilizando Algoritmos Evolutivos

Esta seção diz respeito ao artigo *Optimizing Contextual-based Optimum-Forest Classification through Swarm Intelligence* [98] que foi apresentado no *Advanced Concepts for Intelligent Vision Systems (ACIVS/2013)*. Com o intuito de avaliar as abordagens de otimização evolutiva empregadas para determinar o parâmetro β , foram conduzidos experimentos para comparar as técnicas de otimização evolutivas HS e PSO utilizando imagens de ressonância magnética do conjunto de dados IBSR.

A Tabela 6.7 apresenta os valores para cada parâmetro dos métodos de otimização PSO e HS empregados neste experimento. Note que uma atenção especial foi dada para o parâmetro BW . Ao invés de usar um valor fixo para tal, o qual poderia causar uma lenta convergência, foi proposta a seguinte regra:

$$BW = e^{-x}, \quad (6.5)$$

tal que

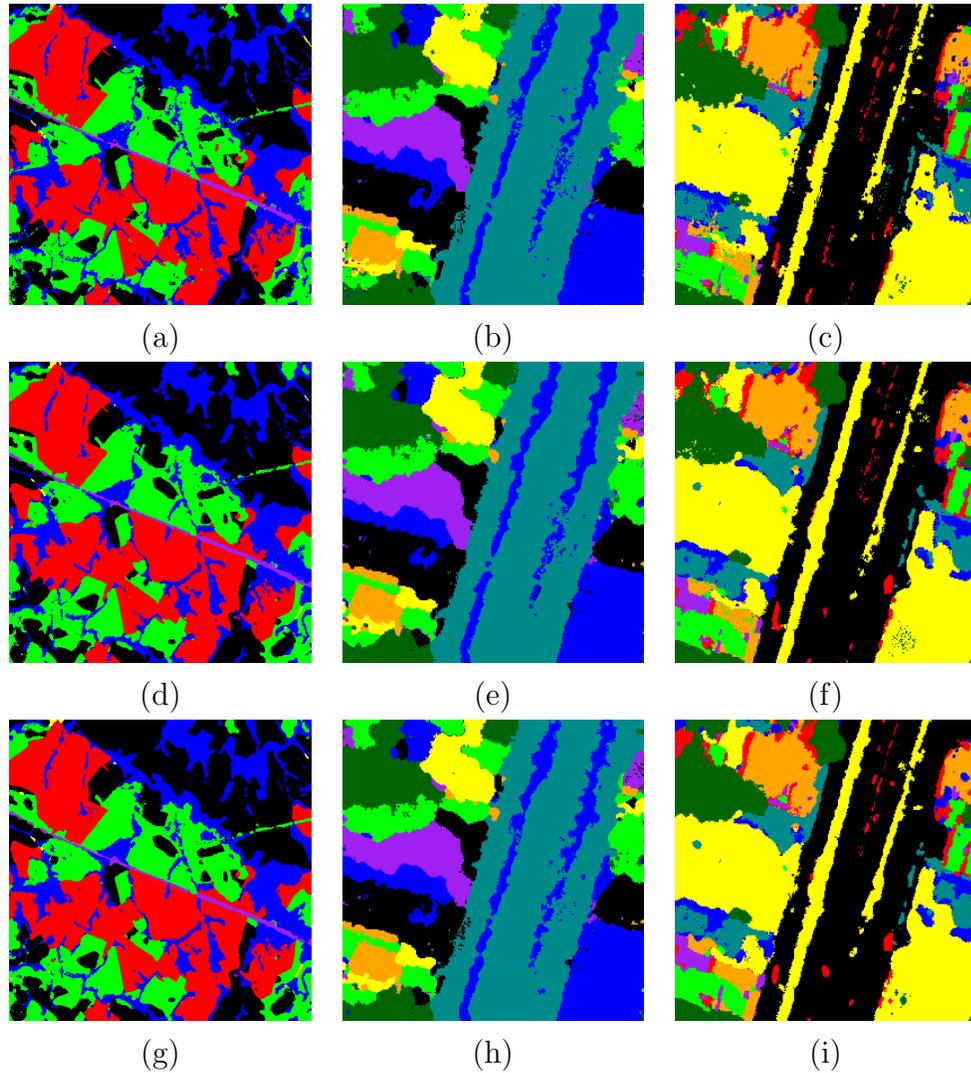


Figura 6.10: Imagens CBERS-2B, Ikonos-2 MS e Geosyde classificadas por OPF-MRF utilizando vetor de 5 características com os modelos: Potts/Fisher em (a), (b) e (c), respectivamente; GIMLL-L1/Fisher em (d), (e) e (f); e GIMLL-L2 em (g), (h) e (i).

$$x = \begin{cases} 2.3 + (t - 1) * 0.092 & \text{se } 1 \leq t \leq 25 \\ 4.6 & \text{caso contrário,} \end{cases} \quad (6.6)$$

onde t é o número de iterações, tal que $1 \leq t \leq T_{opt} = 50$ (laço principal nas Linhas 4 – 24 do Algoritmo 4).

Utilizando as Equações 6.5 e 6.6, pode-se iniciar com valores altos para BW ($BW = e^{-2.3} = 0.1$ para a iteração #1) até o algoritmo alcançar a metade do seu “tempo de vida”,

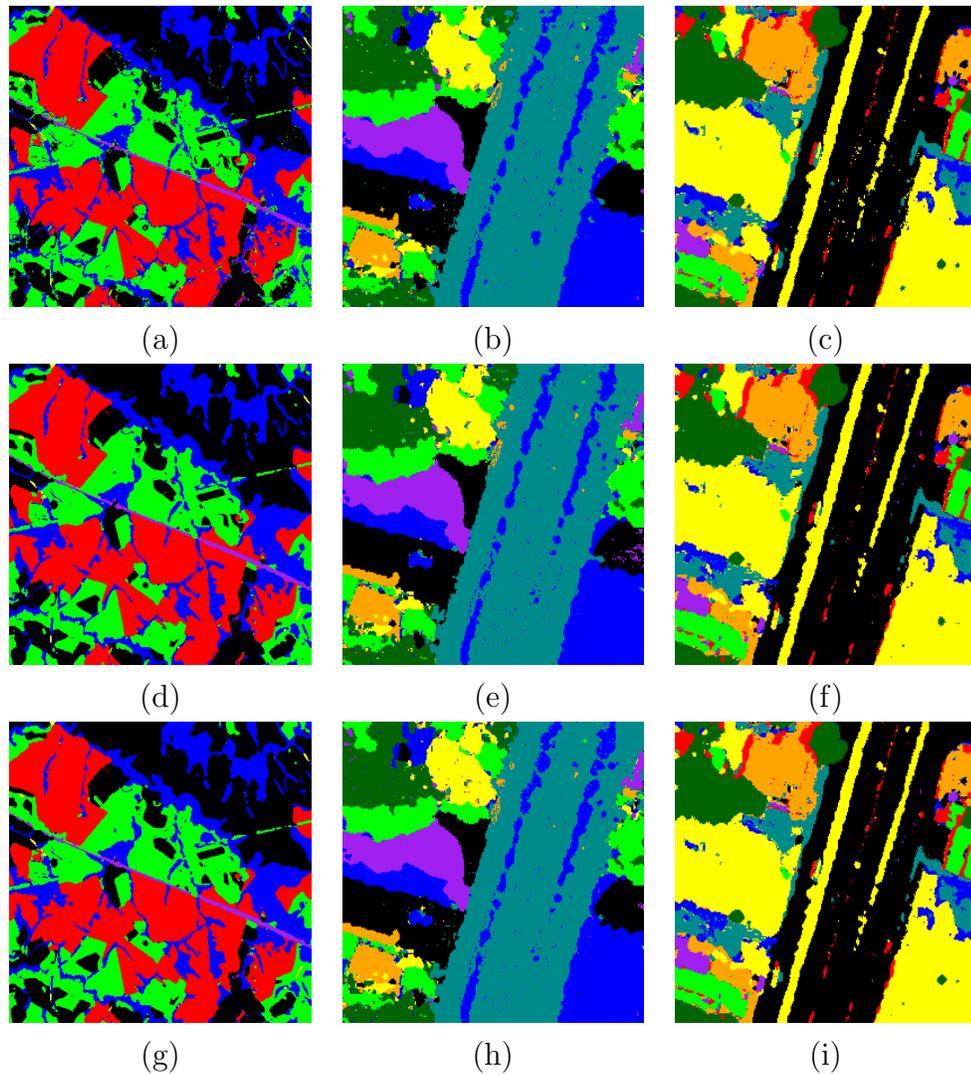


Figura 6.11: Imagens CBERS-2B, Ikonos-2 MS e Geoeye classificadas por OPF-MRF utilizando vetor de 26 características com os modelos: Potts/Fisher em (a), (b) e (c), respectivamente; GIMLL-L1/Fisher em (d), (e) e (f); e GIMLL-L2 em (g), (h) e (i).

Técnica	Parâmetros
HS	HMCR= 0.9, PAR= 0.5
PSO	$c_1 = c_2 = 2, w = 0.5$

Tabela 6.7: Parâmetros utilizados para o PSO e HS, sendo tais valores empiricamente escolhidos.

isto é, $t = 25$ para um total de 50 iterações. Portanto, enquanto x aumenta em passos de 0.092, e^{-x} diminui até $t = 25$. Depois disso, isto é, para $26 \leq t \leq 50$, BW é limitado

como sendo um parâmetro estático com valor $BW = e^{-4.6} = 0.01$. Assim, a principal ideia é permitir uma rápida convergência no início da execução do HS e então diminuir tal comportamento com o intuito de refinar a solução e, com isso, evitar armadilhas de ótimos locais em torno da solução global.

Utilizando imagens MRI, foram comparados os resultados obtidos das otimizações dos métodos PSO e métodos baseados em HS em conjunto com a abordagem BF, isto é, uma busca pseudo-exaustiva para valores de β dentro do intervalo $]0, \beta_{max}]$, onde o valor crítico de β é dado por $\beta_{max} = \ln(1 + \sqrt{C})$. Para tal propósito, foram utilizadas 5 iterações para o algoritmo ICM. As porcentagens dos conjuntos de treinamento, validação e teste utilizadas foram 0.03%, 19.97% e 80.00%, respectivamente, sendo escolhidas empiricamente. Com relação à descrição da imagem, foram extraídas 27 características de cada *voxel*, ou seja, uma janela 3×3 para um sistema de vizinhança 3-D. As características são os valores de cinza dos *voxels* dividido pelo *voxel* com maior brilho na imagem.

Como mencionado anteriormente, a abordagem proposta primeiramente itera sobre os conjuntos treinamento/validação com o intuito de encontrar o melhor valor de β fornecido pelo algoritmo de otimização usando 5 partículas e 10 iterações para o PSO, e 5 harmonias e 50 iterações para o HS (foram utilizadas mais iterações para HS porque ele tem sido consistentemente mais rápido do que o PSO). Posteriormente, foi empregado o algoritmo ICM sobre os conjuntos de treinamento/teste usando o valor de β que foi obtido pelo PSO/HS. Com o intuito de fornecer resultados mais conclusivos, foram executadas quatro validações cruzadas sobre os conjuntos de treinamento, validação e teste. Com relação ao método BF, foram usados valores de β com passos de 0.1. A Figura 6.12 mostra os resultados médios para o método BF sobre o conjunto de teste para as imagens 1–6. Pode ser constatada a variação da acurácia média sobre diferentes valores de β , principalmente para a imagem 06, onde a acurácia média caiu de 89.06% para 84.59% com $\beta = 0.8$ e $\beta = 0.9$, respectivamente. É importante lembrar que o procedimento BF foi realizado sobre os conjuntos de treinamento e teste, sendo o conjunto de validação empregado apenas para encontrar o valor do β pelo PSO e HS.

A Tabela 6.8 apresenta as acurácias médias obtidas para cada imagem empregando as abordagens BF, PSO e HS. Com relação aos resultados de BF, foi empregado o valor com a melhor acurácia de cada imagem, conforme apresentado na Figura 6.12. Pode-se observar que BF tem superado os resultados de PSO e HS em 3 das 6 imagens, mas é perceptível que todas abordagens têm obtido resultados próximos, que foi o resultado esperado. As Figuras 6.13d, 6.13e e 6.13f mostram os resultados para as abordagens PSO, HS e BF para uma imagem de exemplo (Figura 6.13a), respectivamente.

No entanto, o ponto principal está em torno dos tempos de execução médios, que são apresentados na Tabela 6.9 no seguinte formato $x:y:z$, onde x , y e z são horas, minutos e segundos, respectivamente. O tempo de execução considera todo o algoritmo de apren-

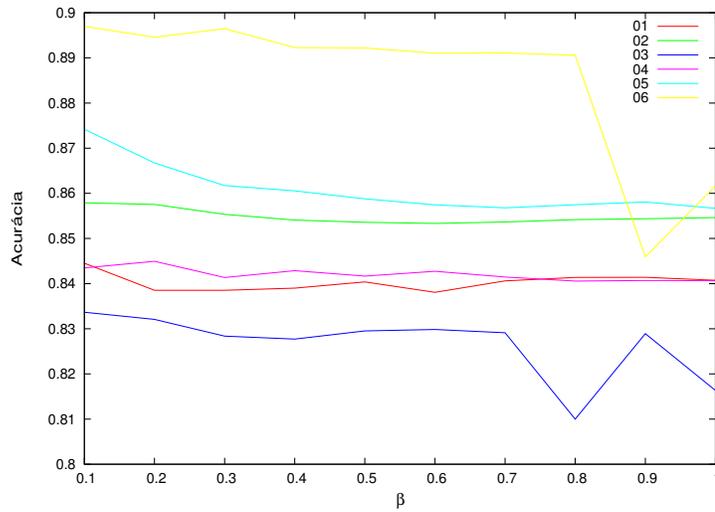


Figura 6.12: Acurácia média sobre o conjunto de teste para diferentes valores de β usando BF.

Tabela 6.8: Acurácia média (\bar{x}) e desvio padrão (σ) sobre as imagens empregadas. As melhores acurácias estão em negrito.

		01	02	03	04	05	06
HS	\bar{x}	84.38%	85.84%	83.07%	84.78%	87.32%	89.81%
	σ	1.952	1.673	0.841	1.026	2.112	0.976
PSO	\bar{x}	84.61%	85.81%	83.95%	84.77%	87.29%	89.83%
	σ	2.059	1.490	0.558	1.496	2.043	0.811
BF	\bar{x}	84.50%	86.05%	83.36%	84.67%	87.43%	89.91%
	σ	1.852	1.563	0.799	1.159	2.011	0.990

dizagem para encontrar β no caso de PSO e HS (Algoritmos 3 e 4). Claramente, pode-se observar que os métodos de otimização foram mais rápidos do que BF. Por exemplo, se for considerada a imagem 03, PSO foi 71.56% mais rápido do que BF, sendo que o PSO ainda tem maior acurácia do que BF para a mesma imagem (Tabela 6.8). Um comportamento similar pode ser visto para as imagens restantes.

Tabela 6.9: Tempo de execução médio sobre as imagens. As técnicas mais rápidas estão em negrito.

	01	02	03	04	05	06
HS	3:21:40	3:32:00	3:04:46	2:55:11	3:01:45	3:44:31
PSO	3:39:33	3:31:29	2:18:58	3:13:16	3:04:15	4:02:06
BF	4:24:36	4:32:00	3:58:25	4:11:12	3:53:43	4:52:56

6.2.1 Melhorando a Classificação da Cobertura do Solo Utilizando OPF-MRF

Esta seção diz respeito ao artigo *Improving Land Cover Classification Through Contextual-based Optimum-Path Forest* [88], publicado no periódico *Information Sciences*.

Os experimentos foram conduzidos com o intuito de realizar uma análise comparativa da abordagem de otimização proposta com os classificadores baseados em contexto OPF-MRF e SVM-MRF. Assim, a metodologia empregada é descrita nesta seção para validar o OPF-MRF proposto contra as abordagens OPF e SVM tradicionais, bem como a implementação do classificador SVM-MRF utilizando uma ideia similar ao que foi aplicada na abordagem OPF-MRF, o qual foi inspirado no trabalho de Kittler et al. [17]. Para tal, foi utilizado o modelo de Potts para implementação das versões contextuais, o qual foi projetado para lidar com modelagem de problemas em que as variáveis assumem valores discretos. Os resultados experimentais foram obtidos utilizando imagens de sensoriamento remoto adquiridas dos sensores Landsat 5 e CBERS-2B cobrindo a área de Itatinga-SP, Brasil, e imagens adquiridas pelos sensores Geoeye e Ikonos-2 MS cobrindo a área de Duque de Caxias-RJ, Brasil. A Figura 6.1 mostra estas imagens, sendo suas respectivas versões de dados de referência ilustradas na Figura 6.3.

Com relação às implementações do OPF e SVM, foram empregados os pacotes LibOPF [115] e LibSVM [116], respectivamente. Como mencionado anteriormente, o SVM padrão foi modificado para o SVM-MRF operar de maneira semelhante ao trabalho de Tarabalka et al. [11], mas a base do algoritmo ainda “permanece” na LibSVM. Além disso, foi utilizado o *kernel* “função de base radial” (*Radial Basis Function* - RBF) para ambos classificadores SVM e SVM-MRF. Note que os parâmetros do *kernel* RBF e os classificadores baseados em SVM foram otimizados através de um procedimento de validação cruzada utilizando os valores de parâmetros $C = 2^k$ e $\gamma = 2^m$, onde $k = -5, \dots, 15$ e $m = -15, \dots, 3$, ambos com passos de 1.

Assim, para cada imagem, foram executadas cinco rodadas de ambas técnicas SVM-

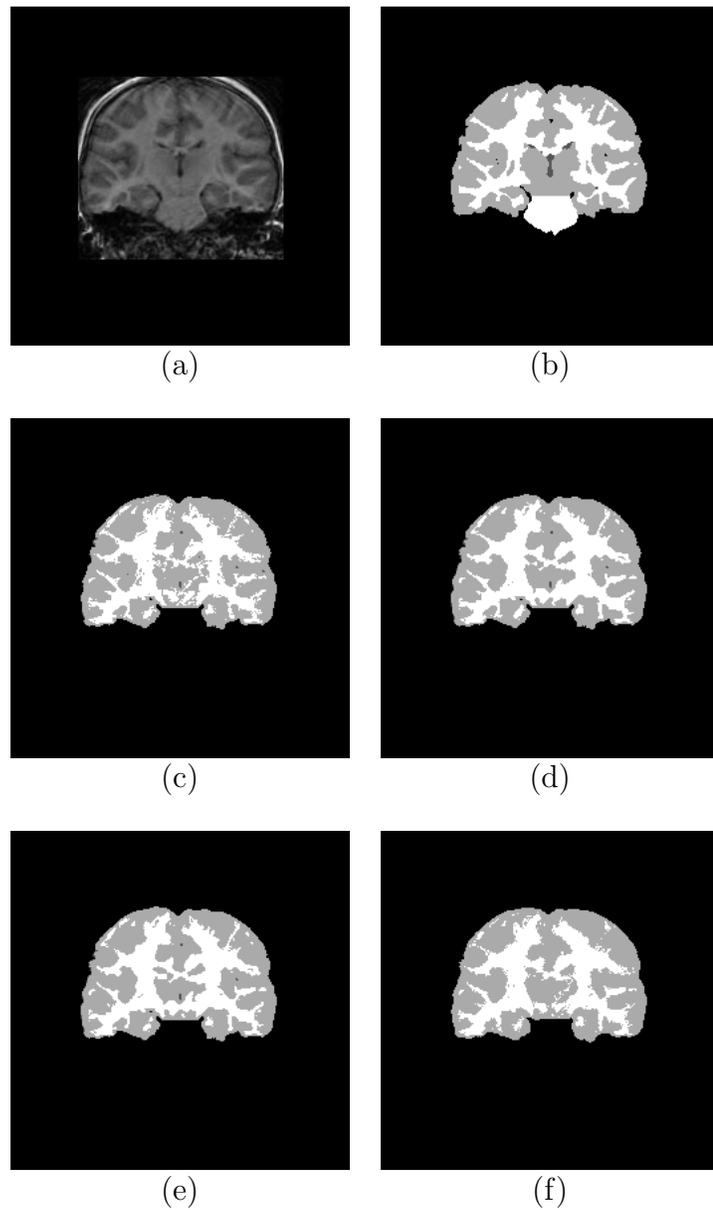


Figura 6.13: Classificação de imagens MRI-T1: (a) IBSR-04 fatia 58, (b) dados de referência, (c) linha de base, (d), (e) e (f) são os resultados obtidos usando os métodos PSO, HS e BF, respectivamente.

MRF e OPF-MRF com 5%, 15% e 80% para os conjuntos de treinamento, validação e teste, respectivamente. Além disso, cada *pixel* da imagem é uma instância do problema, sendo o seu vetor de características composto pela intensidade de brilho de seus vizinhos nas diferentes bandas no sistema de cor RGB e a localização espacial do *pixel*. Para a

extração de informação contextual, foi utilizado um sistema de vizinhança 3×3 , sendo a classificação contextual realizada utilizando um número máximo de 10 iterações para o método ICM, onde os critérios de parada são o número máximo de iterações ou se o número de mudanças entre uma iteração e outra for menor do que 0.1% da quantidade de instâncias (*pixels*) de uma imagem. Finalmente, o parâmetro β foi estimado utilizando os conjuntos de treinamento e validação para os métodos PSO, HS, IHS e GHS, para os quais foram empregadas 10 iterações e 5 partículas para o PSO, e 50 iterações e 5 harmonias para o HS e suas variantes. Tais métodos possuem parâmetros que foram configurados conforme a Tabela 6.10. Além disso, também empregamos o método BF com valores dentro da faixa $[0, \beta_{max}]$ com passos de β_{max}/n , com $n = 50$. Esta última abordagem foi empregada para simular uma referência “ótima” (linha de base) para a taxa de reconhecimento usando diferentes valores de β . Com relação à taxa de reconhecimento, foi utilizada a mesma medida de acurácia proposta por Papa et al. [6] e citada anteriormente.

Técnica	Parâmetros
HS	HMCR= 0.9, PAR= 0.5
IHS	HMCR= 0.9, PAR _{min} = 0.1, PAR _{max} = 0.9 BW _{min} = 0.1, BW _{max} = 0.9
GHS	HMCR= 0.9, PAR= 0.5
PSO	$c_1 = c_2 = 2, w = 0.5$

Tabela 6.10: Parâmetros utilizados para o PSO, HS, IHS e GHS: Tais parâmetros foram escolhidos empiricamente.

Com isso, foram calculadas as acurácias médias sobre uma validação cruzada com 5 rodadas em cada imagem, usando a acurácia encontrada na última iteração dos métodos de classificação contextual como sendo o seu resultado final. Com o intuito de executar uma análise estatística, foram comparadas todas as combinações de classificadores usando o teste de McNemar [117], que é descrito pela Equação 6.7:

$$\chi^2 = \frac{(|N_{01} - N_{10}| - 1)^2}{N_{01} + N_{10}}, \quad (6.7)$$

onde N_{01} representa o número de vezes que o primeiro classificador classificou erroneamente e o segundo classificou corretamente, e N_{10} o número de vezes que o primeiro classificador classificou corretamente e o segundo classificou erroneamente. Assume-se a hipótese nula quando ambos classificadores tem desempenhos semelhantes, ou seja, se o valor de χ^2 for maior do que 10.83 ($p = 0.001$), pode-se rejeitar a hipótese nula e concluir que os classificadores tem desempenhos distintos.

Resultados Experimentais

A seguir, são apresentados os resultados experimentais obtidos utilizando os classificadores de padrões OPF e SVM, bem como suas versões contextuais OPF-MRF e SVM-MRF, respectivamente. Com o intuito de deixar os resultados de maneira mais clara, os experimentos foram divididos de acordo com cada imagem empregada.

Imagem CBERS-2B

Nesta seção apresentamos os resultados obtidos sobre a imagem CBERS-2B (Figura 6.1a). A Figura 6.14 mostra a curva de acurácia média ao longo das iterações para o OPF-MRF com β estimado usando PSO, HS, IHS, GHS e BF. A acurácia mostrada na iteração #0 é o resultado da classificação utilizando OPF padrão, isto é, classificação sem considerar a informação contextual.

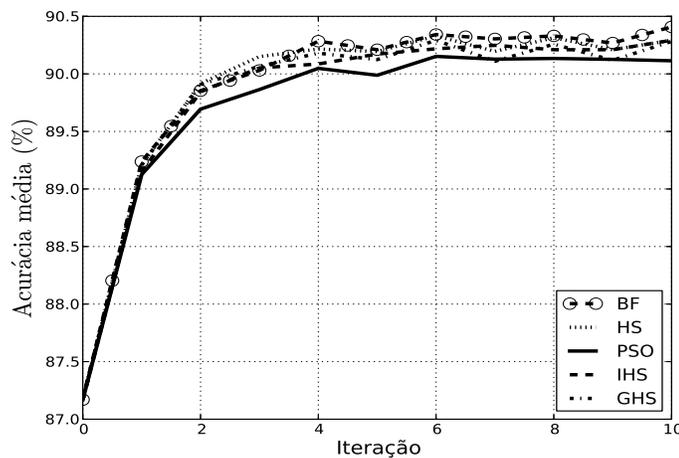


Figura 6.14: Curva de acurácia para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem CBERS-2B.

É possível extrair cinco informações importantes da Figura 6.14: (i) OPF-MRF claramente aperfeiçoa a classificação do OPF padrão (iteração #0), (ii) a acurácia do OPF-MRF ao longo das iterações aumenta até a sua convergência (iteração #10), (iii) todas as abordagens foram capazes de encontrar valores similares aos valores (pseudo-ótimo) obtidos pelo BF, (iv) a acurácia média obtida pelo HS é ligeiramente melhor do que a acurácia obtida pelo PSO (se for considerada apenas uma inspeção visual na Figura 6.14) e, finalmente, (v) o HS com esquema proposto para calcular o valor de BW dinamicamente (Equações 6.5 e 6.6) alcançou resultado ligeiramente melhor do que GHS. Adicionalmente, a Figura 6.15 mostra as acurácias médias obtidas pelo SVM-MRF ao longo das iterações.

Novamente, os valores encontrados pelas abordagens empregadas foram bastante similares aos obtidos pelo método BF. Desta figura, três importantes informações pode ser destacadas: (i) os resultados do OPF-MRF são ligeiramente melhores do que o SVM-MRF, (ii) SVM-MRF ligeiramente aperfeiçoa o resultado do SVM padrão³, e (iii) PSO obteve melhores resultados do que HS, diferentemente dos resultados obtidos pelo OPF-MRF. Portanto, é possível extrair dois aspectos importantes: (i) a versão contextual do OPF melhorou a sua versão padrão, (ii) bem como o esquema proposto de otimização do parâmetro de dependência espacial funcionou para ambos classificadores OPF e SVM, evidenciando a segunda contribuição deste experimento.

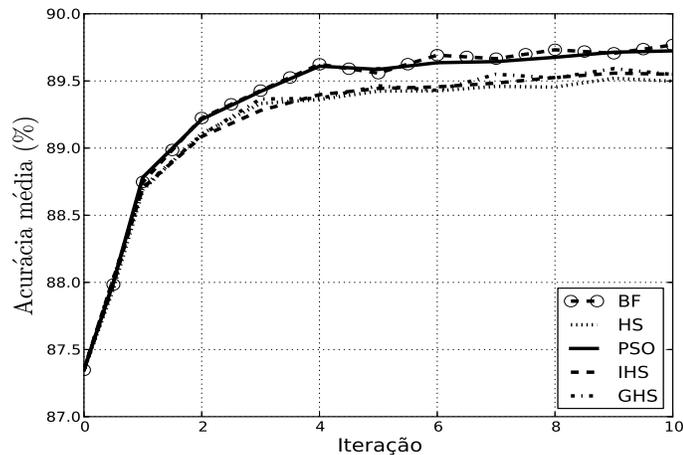


Figura 6.15: Curva de acurácia para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem CBERS-2B.

A Tabela 6.11 apresenta os tempos médios de execução obtidos sobre as imagens no formato $x:y:z$, onde x , y e z denotam as horas, minutos e segundos, respectivamente. Aparentemente, para todas as imagens, as abordagens PSO, HS e suas variantes foram mais rápidas do que BF levando-se em consideração o resultado obtido por OPF-MRF, tornando-as abordagens interessantes para evitar a sobrecarga de avaliações empíricas de β . Embora as técnicas baseadas no método HS tenham menor complexidade do que o PSO, são necessárias mais iterações para a convergência, uma vez que apenas um agente (harmonia) é avaliado a cada iteração, enquanto que no PSO todos agentes (partículas) são avaliados. Além disso, é possível notar que o método SVM-MRF foi mais lento do que o OPF-MRF, uma vez que está sendo considerada a carga computacional para otimização dos parâmetros do SVM.

³Note que o Algoritmo 2 também pode ser aplicado para construir o classificador SVM-MRF: a ideia é apenas mudar o OPF por SVM nas linhas 2 – 3 e 9 – 10.

Tabela 6.11: Tempo de execução média para estimar β .

Técnica		CBERS-2B	Landsat 5	Geoeye	Ikonos-2	ALOS-PALSAR
OPF-MRF	HS	5:18:19	4:48:30	0:17:30	0:15:57	0:21:35
	PSO	6:36:06	5:49:47	0:26:21	0:22:46	0:29:54
	BF	7:42:56	8:07:27	0:31:45	0:29:35	0:30:43
	IHS	4:50:15	4:37:56	0:20:24	0:19:00	0:22:52
	GHS	2:01:38	1:52:04	0:07:14	0:07:23	0:09:15
SVM-MRF	HS	12:19:53	10:52:49	0:44:43	0:51:50	0:11:29
	PSO	13:09:12	11:41:29	0:51:49	0:54:00	0:13:37
	BF	12:45:35	11:15:26	0:53:02	1:01:40	0:15:46
	IHS	12:22:13	11:10:33	0:45:23	0:52:50	0:11:26
	GHS	10:46:13	9:43:24	0:37:24	0:42:52	0:04:37

As Figuras 6.16 e 6.17 apresentam a imagem CBERS-2B classificada utilizando as técnicas OPF e SVM, respectivamente, bem como as suas versões contextuais (OPF-MRF e SVM-MRF) com β estimado pelo PSO e técnicas baseadas no HS. É possível observar boas taxas de reconhecimento para as classes “pastagens” e “reflorestamento”. No entanto, nota-se uma confusão entre as classes “mata nativa” e “culturas” com as classes “pastagens”, “mata nativa” e “estradas”. Com relação à imagem classificada pelo SVM, é possível verificar alguns pequenos erros: uma confusão entre “culturas” com “estradas” e “mata nativa”. As classes “reflorestamento” e “pastagens” mostraram melhores taxas de reconhecimento. Resultados similares foram notados para o OPF-MRF com β estimado pelo HS, porém com uma confusão entre “cultura” com outras classes. SVM-MRF mostrou uma mistura entre “culturas” e “reflorestamento” com “mata nativa”, “pastagens” e “estradas”.

As Tabelas 6.12 e 6.13 apresentam os resultados do teste estatístico de McNemar sobre a imagem CBERS-2B classificada considerando as técnicas baseadas em OPF e SVM, respectivamente. Na Tabela 6.12, é possível observar duas principais informações: (i) o OPF-MRF aperfeiçoou o OPF padrão usando as abordagens PSO, HS, IHS, GHS e BF (valores de χ^2 maiores do que 10.83), e (ii) os pares de técnicas que apresentaram desempenho similares foram (BF,HS) e (HS,GHS) (negrito).

Já com relação ao resultado obtido pelas técnicas baseadas em SVM, é possível observar na Tabela 6.13 que o SVM-MRF superou o resultado do SVM padrão, e também pode-se notar que ambas técnicas HS e PSO foram capazes de encontrar bons valores para β , uma vez que tais técnicas foram consideradas estatisticamente similares ao BF.

Tabela 6.12: Análise estatística considerando os classificadores baseados em OPF sobre a imagem CBERS-2B. Técnicas similares estão em negrito.

χ^2	OPF	HS	IHS	GHS	PSO	BF
OPF	-	5055.2609	5018.7463	5079.8849	4901.2751	5014.5422
HS	5055.2609	-	47.9359	0.0396	83.3416	10.49462
IHS	5018.7463	47.9359	-	50.8094	11.2752	19.1954
GHS	5079.8849	0.0396	50.8094	-	88.1695	11.7650
PSO	4901.2751	83.3416	11.27524	88.1695	-	47.5805
BF	5014.5422	10.4946	19.1954	11.7650	47.5805	-

Tabela 6.13: Análise estatística considerando os classificadores baseados em SVM sobre a imagem CBERS-2B. Técnicas similares estão em negrito.

χ^2	SVM	HS	IHS	GHS	PSO	BF
SVM	-	2426.4230	2541.7600	2541.7600	2336.4964	2381.6485
HS	2426.4230	-	3.7832	3.7832	4.0742	1.4819
IHS	2541.7600	3.7832	-	-	13.0907	7.9324
GHS	2541.7600	3.7832	-	-	13.0907	7.9324
PSO	2336.4964	4.0742	13.0907	13.0907	-	0.9674
BF	2381.6485	1.4819	7.9324	7.9324	0.9674	-

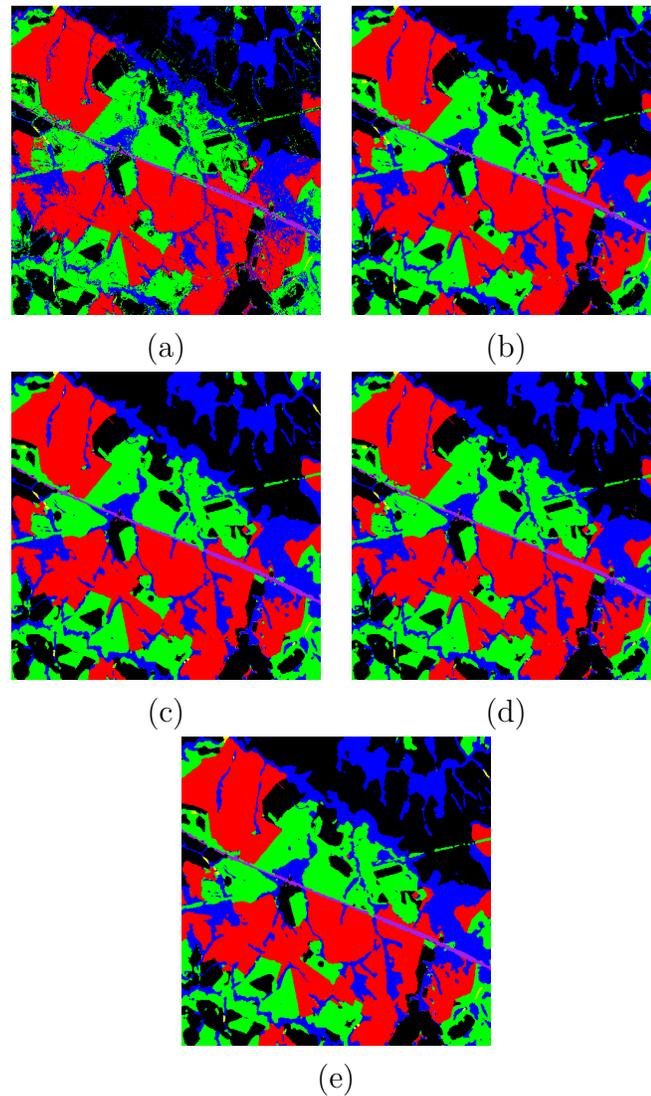


Figura 6.16: Imagem CBERS-2B classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.

Imagem Landsat 5

Esta seção apresenta os resultados obtidos sobre a imagem Landsat 5 TM (Figura 6.1b). A Figura 6.18 mostra a curva de acurácia média ao longo das iterações para o OPF-MRF com β estimado usando PSO, HS, IHS, GHS e BF. Claramente, é possível notar que o BF superou as técnicas de otimização em 1%, sendo PSO ligeiramente melhor do que técnicas baseadas em HS. No entanto, tal melhoria pode não ser adequada, uma vez que PSO tem

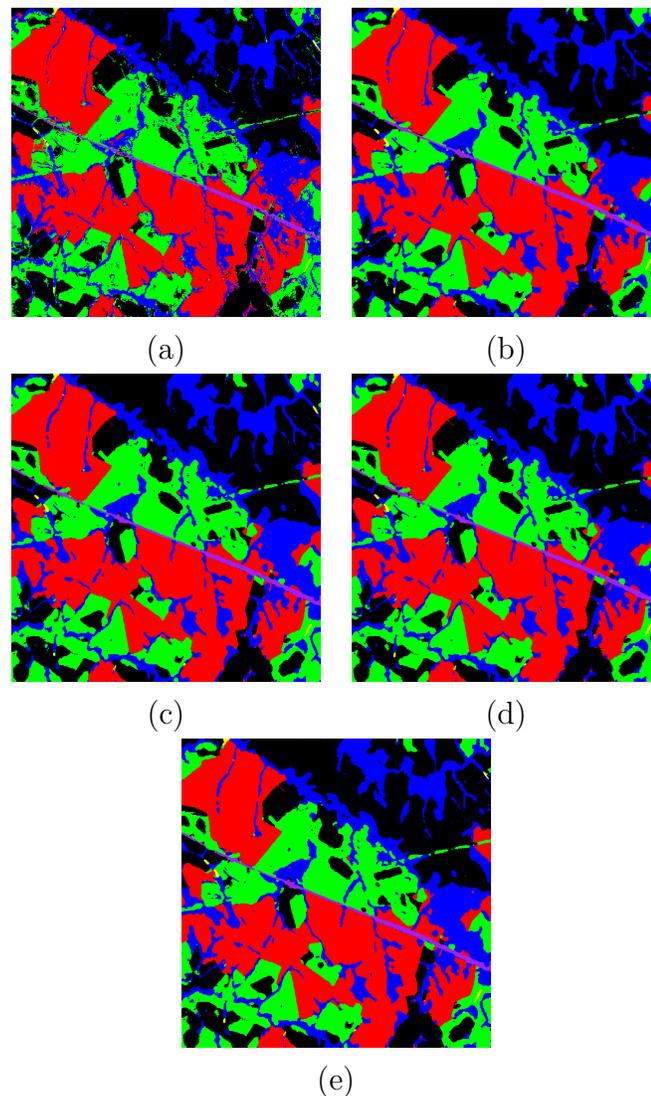


Figura 6.17: Imagem CBERS-2B classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.

uma alta carga computacional, como apresentado na Tabela 6.11. Note que a acurácia observada na iteração #0 corresponde à classificação do OPF padrão, isto é, classificação sem considerar a informação contextual.

A Figura 6.19 apresenta os resultados obtidos pelo SVM-MRF sobre a imagem Landsat 5 TM. Basicamente, os melhores resultados são limitados pelo BF, o qual obteve melhores taxas de reconhecimento, e os piores resultados são limitados pelo PSO, que alcançou as menores taxas de reconhecimento. No entanto, deve-se notar que os resultados não

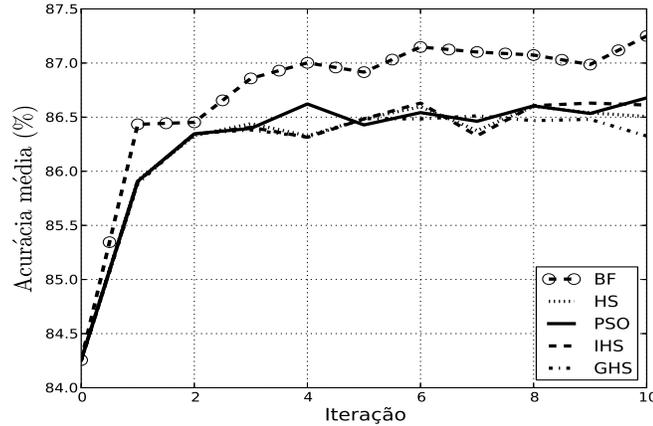


Figura 6.18: Curva de acurácia para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem Landsat 5 TM.

diferem muito uns dos outros.

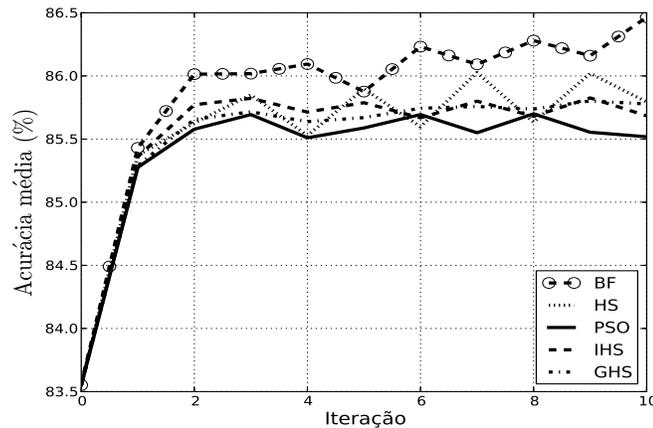


Figura 6.19: Curva de acurácia para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF sobre a imagem Landsat 5 TM.

As Figuras 6.20 e 6.21 mostram a imagem Landsat 5 TM classificada usando os classificadores OPF e SVM, respectivamente, bem como suas versões contextuais (OPF-MRF e SVM-MRF) com β estimado por técnicas meta-heurísticas e BF. Novamente, as versões contextuais obtiveram melhores resultados do que suas versões tradicionais, bem como a imagem classificada com β estimado por GHS apresentou melhor qualidade visual, principalmente em relação à porção superior direita da imagem.

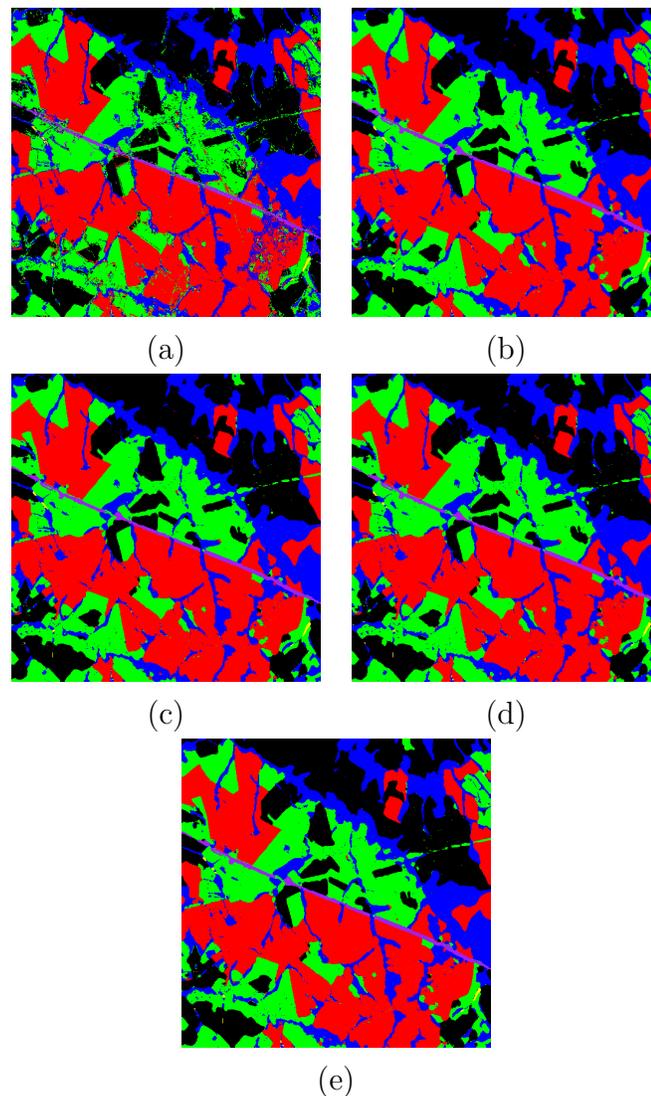


Figura 6.20: Imagem Landsat 5 TM classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.

Novamente, os resultados obtidos pelo OPF-MRF superaram os obtidos pelo OPF tradicional, sendo todas as técnicas meta-heurísticas consideradas similares entre si. Considerando os resultados obtidos pelo SVM-MRF, pode-se observar uma similaridade entre os pares (HS,BF), (HS,IHS), (IHS,GHS), (GHS,PSO), (BF,HS) e (BF,IHS). Portanto, BF, HS e IHS podem ser eleitas as abordagens com maiores acurácias. Assim sendo, HS e IHS podem ser uma boa escolha se for considerada a relação custo/benefício entre eficácia e eficiência (Tabela 6.11).

Tabela 6.14: Análise estatística com relação aos classificadores baseados em OPF sobre a imagem Landsat 5 TM. Técnicas similares estão em negrito.

χ^2	OPF	HS	IHS	GHS	PSO	BF
OPF	-	3392.9098	3406.9396	3355.1024	3418.9078	3576.0901
HS	3392.9098	-	0.6946	1.0466	4.085106	0.0780
IHS	3406.9396	0.6946	-	3.1515	0.1307	0.6277
GHS	3355.1024	1.0466	3.1515	-	2.0459	0.1898
PSO	3418.9078	4.0851	0.1307	2.0459	-	0.3243
BF	3576.0901	0.0780	0.6277	0.1898	0.3243	-

Tabela 6.15: Análise estatística com relação aos classificadores baseados em SVM sobre a imagem Landsat 5 TM. Técnicas similares estão em negrito.

χ^2	SVM	HS	IHS	GHS	PSO	BF
SVM	-	1842.3732	1774.2948	1675.0094	1634.8654	1828.2435
HS	1842.3732	-	3.5124	18.6501	25.0703	0.0274
IHS	1774.2948	3.5124	-	9.2691	13.9426	2.9302
GHS	1675.0094	18.6501	9.2691	-	1.1287	17.4896
PSO	1634.8654	25.0703	13.9426	1.1287	-	23.6006
BF	1828.2435	0.0274	2.9302	17.4896	23.6006	-

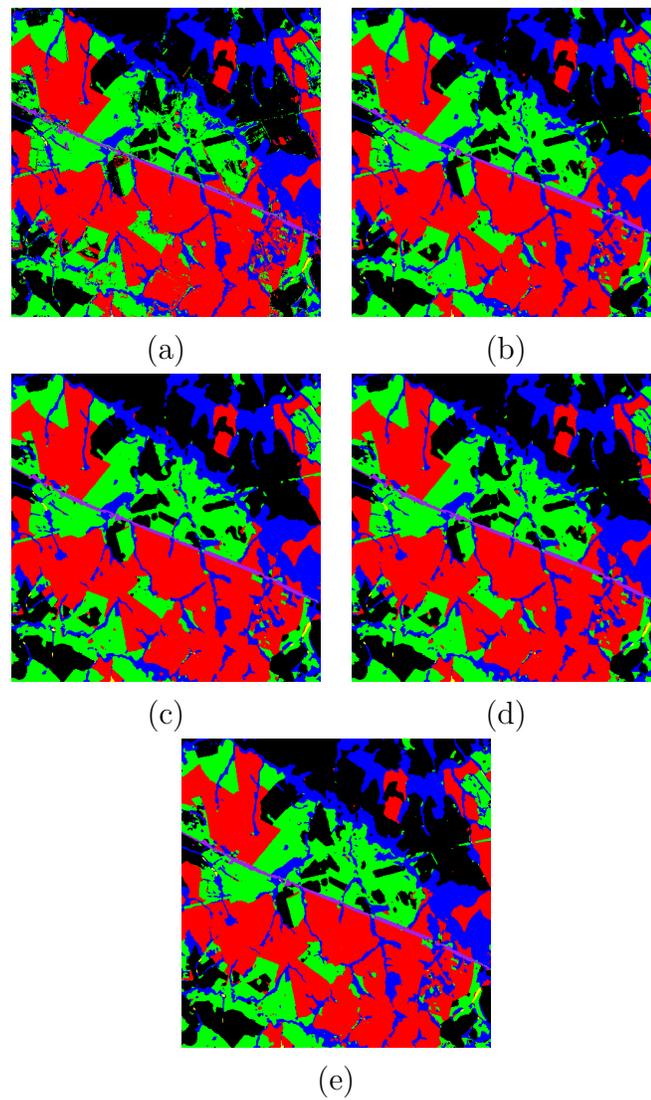


Figura 6.21: Imagem Landsat 5 TM classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.

Imagem Ikonos-2 MS

Esta seção apresenta os resultados de classificação obtidos sobre a imagem Ikonos-2 MS (Figura 6.1c). A Figura 6.22 mostra a curva de acurácia média ao longo das iterações do OPF-MRF usando β estimado por PSO, HS, IHS, GHS e BF. Um comportamento semelhante pode ser observado para todas as técnicas meta-heurísticas, sendo os resultados obtidos por BF ligeiramente melhores. O ganho na acurácia foi mais alto do que nos experimentos anteriores, evidenciando que o OPF-MRF é adequado para o aprendizado com informação contextual. Novamente, o esquema proposto baseado em métodos meta-heurísticos foi capaz de encontrar valores semelhantes de β considerando o método BF. Além disso, pode-se observar na Tabela 6.11 que todas técnicas meta-heurísticas foram mais rápidas do que BF em ambos OPF-MRF e SVM-MRF.

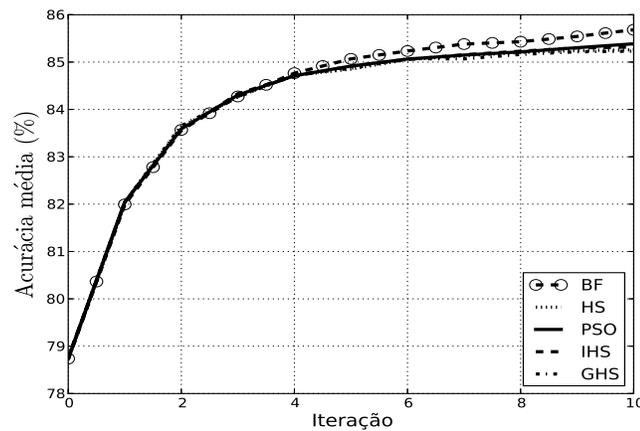


Figura 6.22: Curva de acurácia média para o OPF-MRF usando β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Ikonos-2 MS.

A Figura 6.23 apresenta os resultados obtidos pelo SVM-MRF sobre a imagem Ikonos-2 MS. Neste caso, é possível notar que SVM-MRF melhorou o resultado do SVM padrão, sendo a taxa de reconhecimento dos métodos PSO, HS e suas variantes abaixo do resultado obtido pelo BF. Outro ponto interessante a ser observado é que, aparentemente, o OPF adaptou-se melhor para a abordagem contextual do que o SVM, uma vez que a acurácia na iteração #0 foi próxima de 79%, enquanto que a acurácia do SVM na mesma iteração foi próxima de 82%. No entanto, a acurácia do OPF na última iteração foi ligeiramente maior do que a taxa de reconhecimento do SVM.

As Figuras 6.24 e 6.25 apresentam a imagem Ikonos-2 MS classificada usando as técnicas OPF e SVM, respectivamente, bem como suas versões contextuais (OPF-MRF e SVM-MRF) com β estimado por PSO e técnicas baseadas em HS. Uma confusão na maior

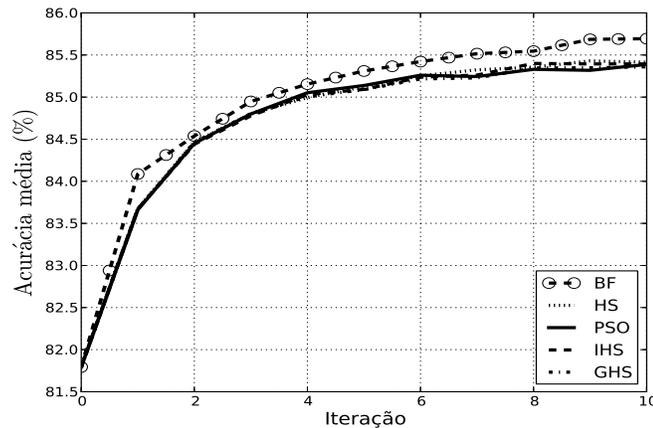


Figura 6.23: Curva de acurácia média para o SVM-MRF usando β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Ikonos-2 MS.

parte das classes pode ser observada, com exceção para “cobertura de árvores” e “pastagens” com OPF padrão. O OPF-MRF com β estimado pelo PSO mostrou boas taxas de reconhecimento para a maioria das classes, com exceção para “estradas”. Também foi observada uma confusão de “estradas” com outras classes e de “solo exposto úmido” e “pastagens” com “estradas”, com resultado similar para o OPF-MRF com β estimado por HS. Com relação ao classificador SVM, foi observada uma mistura de “cobertura de tons escuros” e “cobertura de árvores” com “solo exposto com tons claros” e “cobertura de tons claros”. Resultados similares para o SVM-MRF com β estimado pelas abordagens de otimização também podem ser observados.

As Tabelas 6.16 e 6.17 apresentam o teste estatístico da imagem Ikonos-2 MS considerando as técnicas baseadas em OPF e SVM, respectivamente. Na Tabela 6.16, pode-se observar que as técnicas baseadas em HS são estatisticamente semelhantes, sendo as demais consideradas estatisticamente distintas entre si. Na Tabela 6.17, pode-se observar que todas as técnicas baseadas em HS e o PSO podem ser consideradas semelhantes entre si.

Imagem Geoeye

Esta seção apresenta os resultados obtidos sobre a imagem Geoeye (Figura 6.1d). A Figura 6.26 apresenta a curva de acurácia média ao longo das iterações para OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF. Considerando esse experimento, são observados comportamentos similares entre todas as técnicas, os quais podem ser notados na validação estatística da Tabela 6.18 (todos os pares de técnicas meta-heurísticas similares

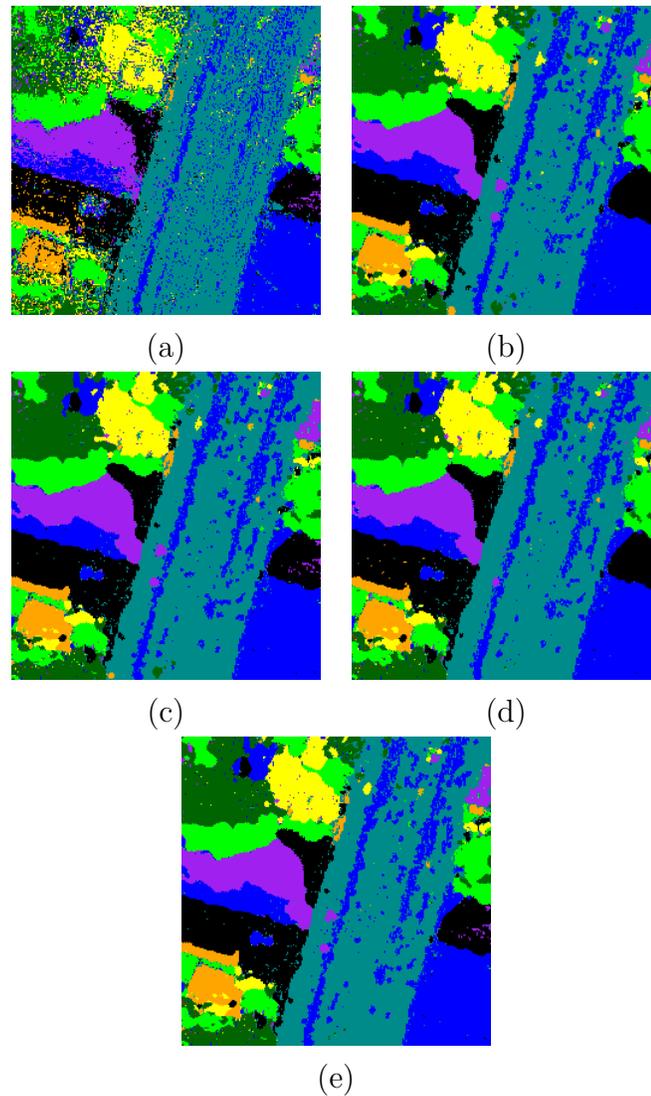


Figura 6.24: Imagem Ikonos-2 MS classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.

entre si estão em negrito).

A Figura 6.27 apresenta a curva de acurácia média obtida pelo SVM-MRF com β otimizado por PSO, HS, IHS, GHS e BF. É possível observar um comportamento semelhante ao da Figura 6.26, porém com resultados mais próximos em todas as técnicas e, assim, sendo consideradas similares entre si (Tabela 6.19). As Figuras 6.28 e 6.29 apresentam as imagens classificadas com OPF-MRF e SVM-MRF, respectivamente.

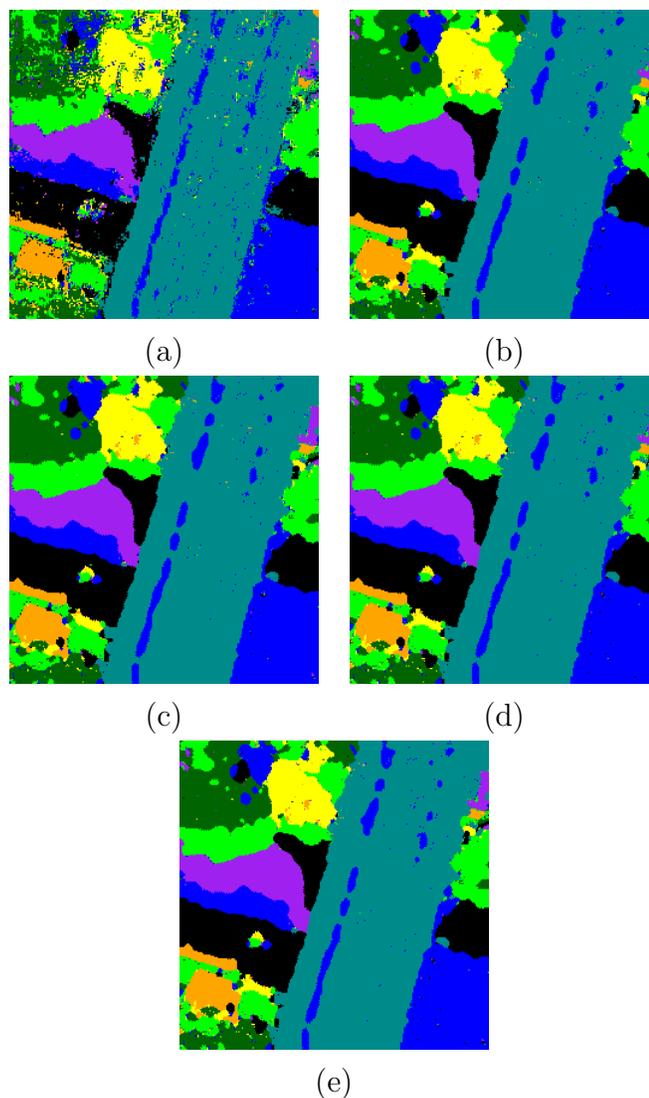


Figura 6.25: Imagem Ikonos-2 MS classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.

Imagem ALOS-PALSAR

Finalmente, esta seção apresenta os resultados sobre a imagem de radar ALOS-PALSAR. A Figura 6.30 apresenta a curva de acurácia média obtida pela classificação contextual com OPF-MRF usando β estimado através de um processo de otimização. Um comportamento semelhante pode ser observado entre todas as técnicas, exceto para BF, que aparentemente oscila da iteração #4 até #10. A análise estatística apresentada na Tabela 6.20 eviden-

Tabela 6.16: Análise estatística com relação aos classificadores baseados em OPF sobre a imagem Ikonos-2 MS. Técnicas similares estão em negrito.

χ^2	OPF	HS	IHS	GHS	PSO	BF
OPF	-	2899.7008	3273.1485	2900.9332	2727.6816	4025.6730
HS	2899.7008	-	34.3607	0.2874	11.5361	262.8288
IHS	3273.1485	34.3607	-	30.5201	73.0151	147.5630
GHS	2900.9332	0.2874	30.5201	-	13.7118	260.9796
PSO	2727.6816	11.5361	73.0151	13.7118	-	321.392031
BF	4025.6730	262.828804	147.5630	260.9796	321.3920	-

Tabela 6.17: Análise estatística com relação aos classificadores baseados em SVM sobre a imagem Ikonos-2 MS. Técnicas similares estão em negrito.

χ^2	SVM	HS	IHS	GHS	PSO	BF
SVM	-	1251.1060	1288.0363	1288.0363	1319.2329	1396.8101
HS	1251.1060	-	2.1485	2.1485	4.8520	22.9165
IHS	1288.0363	2.1485	-	-	0.6677	13.7734
GHS	1288.0363	2.1485	-	-	0.6677	13.7734
PSO	1319.2329	4.8520	0.6677	0.6677	-	10.3854
BF	1396.8101	22.9165	13.7734	13.7734	10.3854	-

Tabela 6.18: Análise estatística em relação aos classificadores baseados em OPF sobre a imagem Geoeye. Técnicas similares estão em negrito.

χ^2	OPF	HS	IHS	GHS	PSO	BF
OPF	-	1751.3133	1780.2405	1798.6286	1825.1182	1788.4409
HS	1751.3133	-	0.0216	0.0079	0.7689	0.0005
IHS	1780.2405	0.0216	-	0.0964	1.3823	0.0152
GHS	1798.6286	0.0079	0.0964	-	0.8444	0.0257
PSO	1825.1182	0.7689	1.3823	0.8444	-	2.4362
BF	1788.4409	0.0005	0.0152	0.0257	2.4362	-

ciou que o OPF-MRF utilizando diferentes métodos de otimização para a estimativa do parâmetro β pode ser considerado estatisticamente distinto em relação ao OPF, sendo o

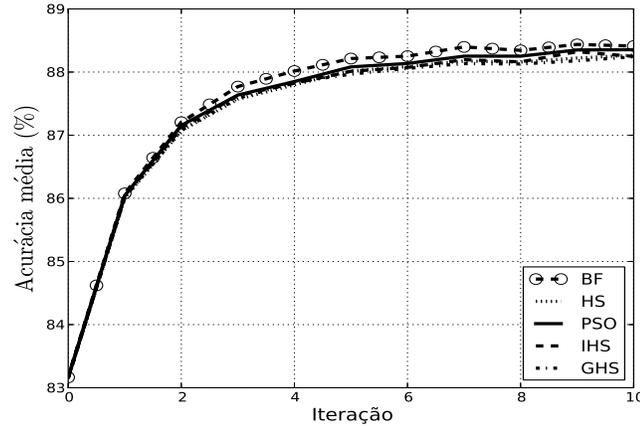


Figura 6.26: Curva de acurácia média para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Geoeeye.

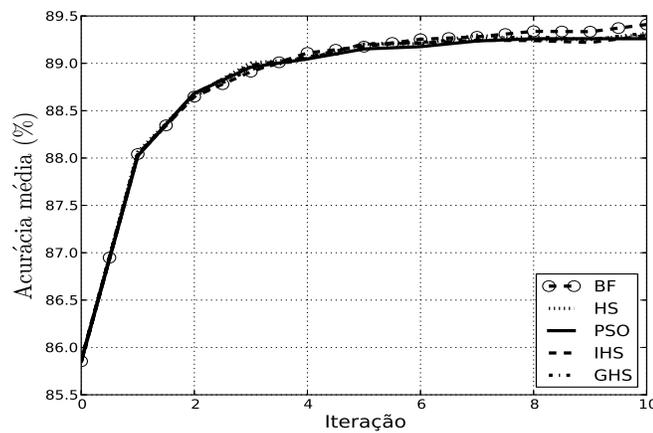


Figura 6.27: Curva de acurácia média para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem Geoeeye.

método HS semelhante à todas técnicas comparadas.

A Figura 6.31 apresenta a curva de acurácia obtida pela classificação contextual com SVM-MRF. Novamente, a abordagem proposta superou a abordagem de classificação tradicional (iteração #0), mas com um comportamento instável para todas as técnicas. Neste caso, os resultados do método BF foram melhores do que os resultados obtidos pelo método HS, levando-se a concluir que a estimativa do parâmetro β através de métodos de otimização para o SVM-MRF é um problema mais complexo do que para o OPF-MRF. A análise estatística apresentada na Tabela 6.21 evidenciou que os métodos HS e BF

Tabela 6.19: Análise estatística em relação aos classificadores baseados em SVM sobre a imagem Geoeye. Técnicas similares estão em negrito.

χ^2	SVM	HS	IHS	GHS	PSO	BF
SVM	-	886.6530	858.3420	858.3420	882.1317	872.5408
HS	886.6530	-	2.5739	2.5739	0.0671	0.4419
IHS	858.3420	2.5739	-	-	1.9760	0.3081
GHS	858.3420	2.5739	-	-	1.9760	0.3081
PSO	882.1317	0.0671	1.9760	1.9760	-	0.2853
BF	872.5408	0.4419	0.3081	0.3081	0.2853	-

Tabela 6.20: Análise estatística em relação aos classificadores baseados em OPF sobre a imagem ALOS-PALSAR. Técnicas similares estão em negrito.

χ^2	OPF	HS	IHS	GHS	PSO	BF
OPF	-	2724.9776	2641.7124	2734.4859	2592.8110	2853.7568
HS	2724.9775	-	2.6495	0.0033	6.7906	6.9341
IHS	2641.7124	2.6495	-	2.6675	1.3819	16.1200
GHS	2734.4859	0.0033	2.6675	-	6.5583	6.6103
PSO	2592.8110	6.7906	1.3819	6.5583	-	23.2619
BF	2853.7568	6.9341	16.1200	6.6103	23.2619	-

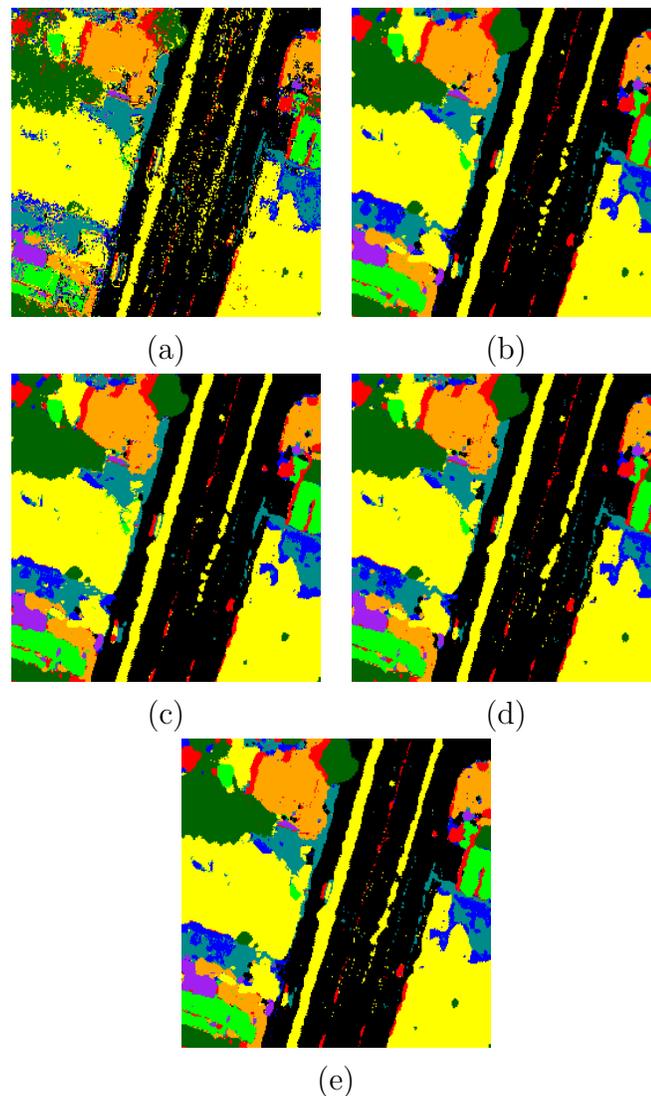


Figura 6.28: Imagem Geocye classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.

obtiveram taxas de reconhecimento semelhantes, e tal resultado pode ser devido ao valor alto de desvio padrão obtido pelo HS durante o processo de aprendizado. Com relação às cargas computacionais apresentadas na Tabela 6.9, GHS foi a técnica mais rápida novamente, com o resultado obtido pelo método PSO próximo ao resultado obtido pelo método BF.

As Figuras 6.32 e 6.33 apresentam a imagem ALOS-PALSAR classificada por OPF-MRF e SVM-MRF, respectivamente. Embora a imagem classificada por SVM aparente-

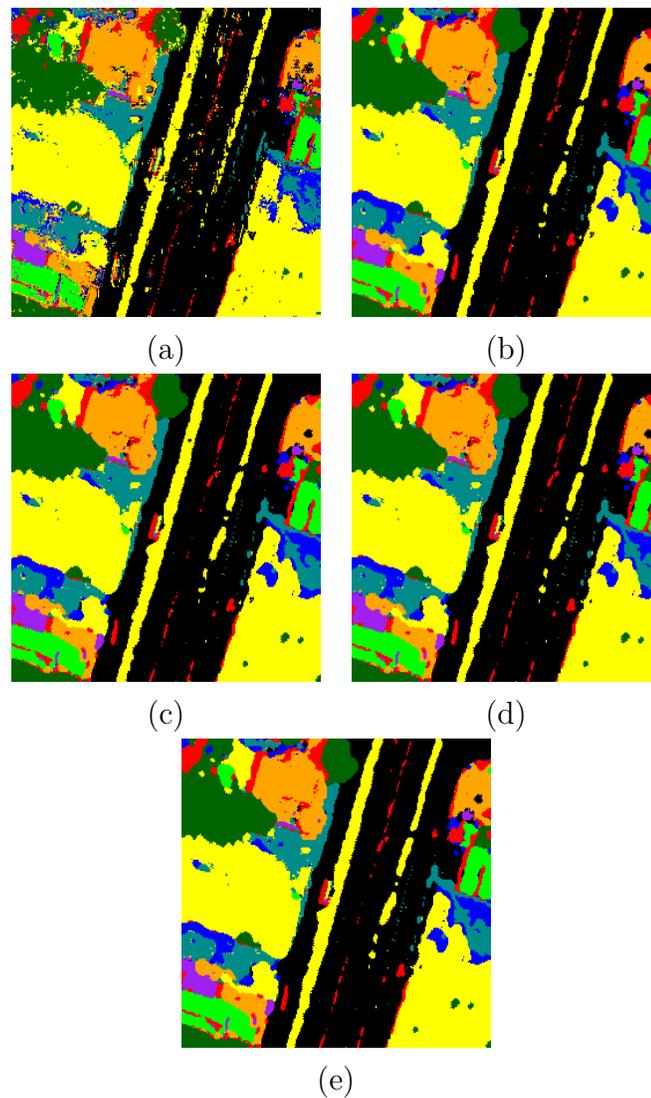


Figura 6.29: Imagem Geoeye classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.

mente obteve maior acurácia do que os resultados obtidos por OPF, tal dado foi obtido a partir de apenas uma rodada, e não reflete o resultado apresentado nas Figuras 6.30 e 6.31.

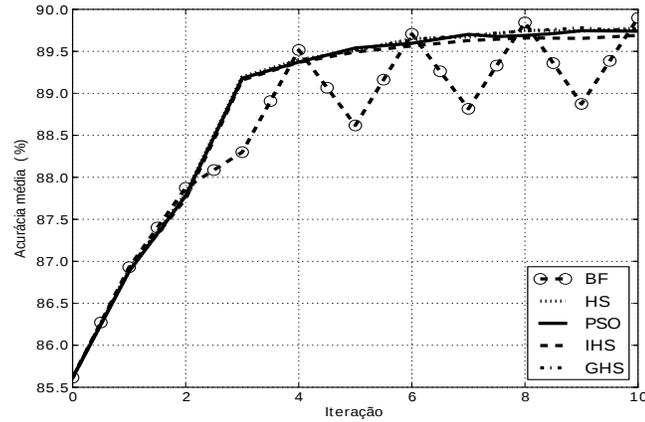


Figura 6.30: Curva de acurácia média para o OPF-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem ALOS-PALSAR.

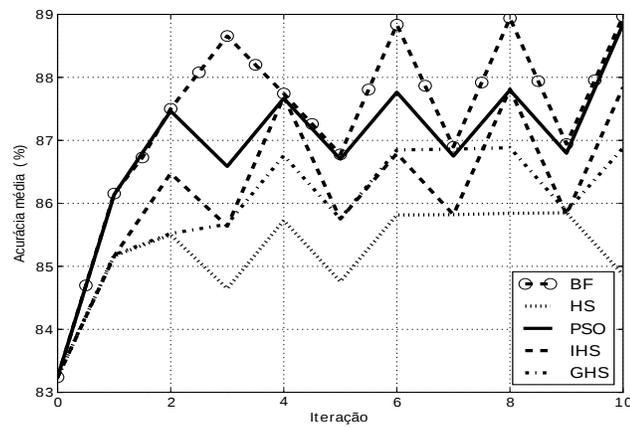


Figura 6.31: Curva de acurácia média para o SVM-MRF com β estimado por PSO, HS, IHS, GHS e BF aplicados sobre a imagem ALOS-PALSAR.

A seção seguinte complementa os resultados obtidos nesta seção aplicando diferentes modelos Markovianos, avaliando diferentes métodos de extração de características e propondo uma etapa adicional de pós-processamento.

Tabela 6.21: Análise estatística em relação aos classificadores baseados em SVM sobre a imagem ALOS-PALSAR. Técnicas similares estão em negrito.

χ^2	SVM	HS	IHS	GHS	PSO	BF
SVM	-	1062.3209	1062.3209	994.7229	648.8286	1048.5504
HS	1062.3209	-	0.0000	6.9963	138.7408	0.9298
IHS	1062.3209	0.0000	-	6.9963	138.7408	0.9298
GHS	994.7229	6.9963	6.9963	-	104.1773	7.3518
PSO	648.8286	138.7408	138.7408	104.1773	-	191.3284
BF	1048.5504	0.9298	0.9298	7.3518	191.3284	-

6.2.2 Avaliando a Classificação da Cobertura do Solo através de Floresta de Caminhos Ótimos Baseada em Contexto sob a Influência de Modelos Markovianos Otimizados

Nesta seção são avaliados os resultados de classificação contextual sobre imagens de sensoriamento remoto utilizando diferentes modelos Markovianos com o parâmetro de dependência espacial estimado através de métodos de otimização evolutivos e numéricos. Também foram avaliados dois vetores de características diferentes com as seguintes representações: (I) um vetor de característica composto pelos valores RGB dos *pixels* vizinhos e pela localização do *pixel* central (x, y) , e (II) outro composto apenas pelos valores RGB do *pixel* central e pela localização do mesmo, gerando um total de 26 e 5 características, respectivamente. A ideia principal é avaliar se é realmente necessário um grande número de características com o intuito de evitar o efeito Hughes [32], o qual indica que a acurácia tende a diminuir se o número de características for aumentado. Além disso, outra contribuição apresentada nesta seção foi avaliar a classificação contextual propondo uma etapa de pós-processamento, cujo intuito foi refinar o mapa de rótulos entre as iterações do OPF-MRF. Esses experimentos fazem parte do artigo intitulado “*Fine-tuning Contextual-based Optimum Path Forest for Land Cover Classification*” [110], o qual foi publicado no periódico *IEEE Geoscience and Remote Sensing Letters*.

Para avaliar os métodos de extração de características, bem como a classificação contextual com pós-processamento, foram utilizadas duas imagens obtidas pelos satélites CBERS-2B e Ikonos-2 MS. Assim, os experimentos foram conduzidos da seguinte forma: para cada imagem de satélite, foram empregadas uma amostragem de 5% dos *pixels* para compor o conjunto de treinamento, 15% para o conjunto de validação e os 80% remanescentes para compor o conjunto de teste, que foram escolhidos empiricamente. Assim, foi empregado um sistema de vizinhança 3×3 escolhido empiricamente. Para estimar

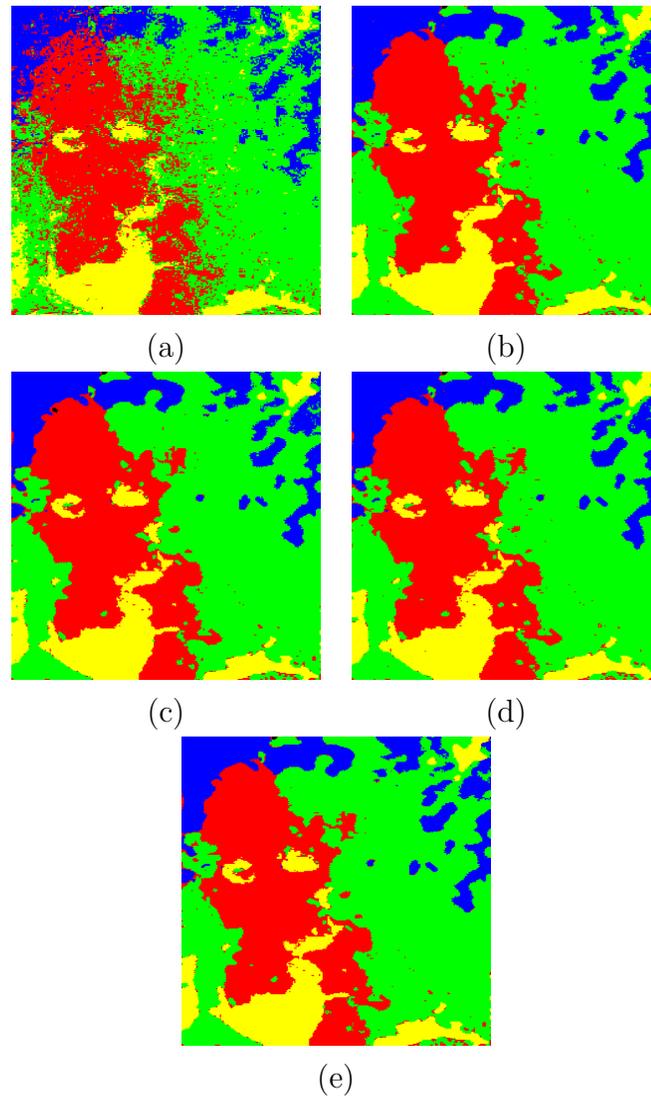


Figura 6.32: Imagem ALOS-PALSAR classificada por: (a) OPF tradicional, (b) OPF-MRF com β estimado por PSO, (c) OPF-MRF com β estimado por HS, (d) OPF-MRF com β estimado por IHS e (e) OPF-MRF com β estimado por GHS.

valores adequados para o parâmetro β , foram empregadas as técnicas de otimização por meta-heurísticas HS e PSO e comparadas com o método BF. Para tal, foram utilizadas 10 iterações e 5 partículas para o PSO, e 50 iterações e 5 harmonias para o HS, cujos parâmetros foram escolhidos empiricamente. Também foram utilizadas 10 iterações para o OPF-MRF nas etapas de otimização e teste. Finalmente, foram calculadas as acurácias médias sobre uma validação cruzada com 10 rodadas. Os demais parâmetros são apresentados na Tabela 6.7.

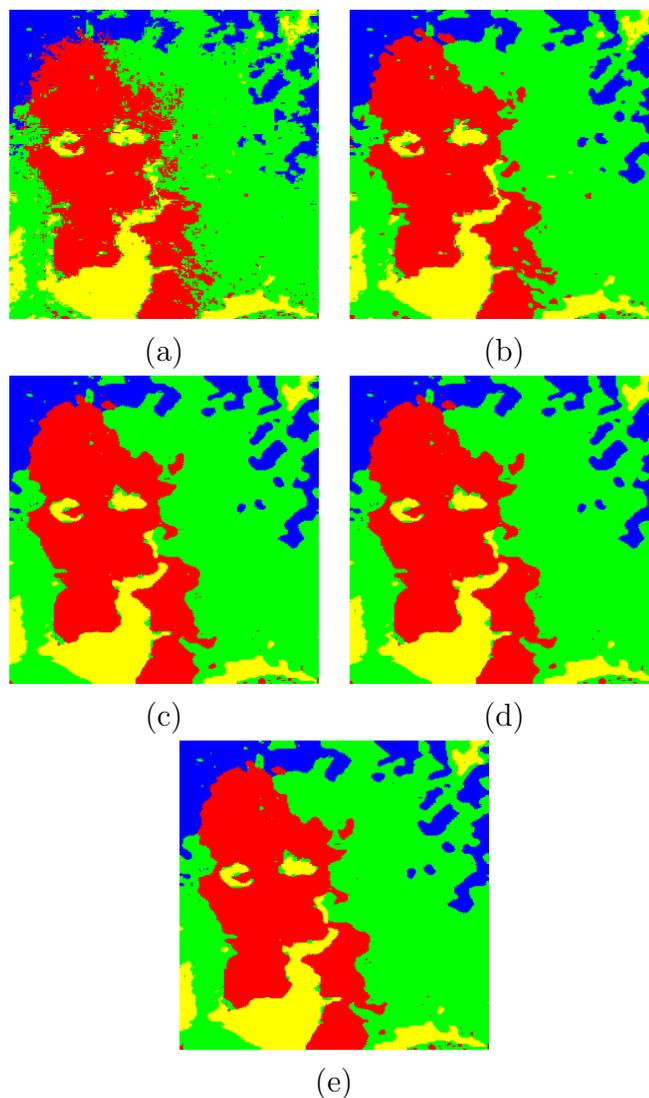


Figura 6.33: Imagem ALOS-PALSAR classificada por: (a) SVM tradicional, (b) SVM-MRF com β estimado por PSO, (c) SVM-MRF com β estimado por HS, (d) SVM-MRF com β estimado por IHS e (e) SVM-MRF com β estimado por GHS.

Embora a abordagem OPF-MRF tenha superado o OPF padrão no contexto de classificação da cobertura do solo, a imagem classificada final ainda poderia conter alguns *pixels* classificados erroneamente em áreas compostas por pequenas regiões cercadas por instâncias de diferentes classes. Para lidar com tal problema, foi proposta uma simples mas eficaz abordagem: basicamente, foi aplicado um filtro da moda como uma etapa de pós-processamento na imagem classificada, mas apenas quando um critério predefinido for satisfeito com o intuito de evitar correções excessivas sobre regiões de alta frequência,

as quais poderiam degradar os resultados anteriores. O critério é definido calculando a frequência absoluta de cada rótulo dentro de um sistema de vizinhança 3×3 centrado em um dado *pixel*, e então selecionado o rótulo mais frequente l . O valor da moda de M significa que o filtro da moda é aplicado se $N > M$, onde N denota o número de *pixels* classificados com o rótulo l . Assim, foi utilizado $M \in \{3, 4, 5\}$ nos experimentos⁴. É importante observar que, se não houver ruído no mapa de rótulos, tal abordagem não terá o efeito desejado. Também é válido destacar que, além de reduzir os artefatos gerados por erros de classificação, evitamos que o mapa de rótulos seja degradado pelo filtro ao longo das iterações, especialmente nas fronteiras das classes.

Resultados Experimentais

Os experimentos apresentados a seguir foram obtidos utilizando a metodologia descrita acima. No primeiro experimento, são comparados os resultados obtidos utilizando duas técnicas de extração de características, isto é, considerando um vetor de 26 características (I) e outro de composto por 5 características (II). As Tabelas 6.22 e 6.23 apresentam os resultados sobre as imagens CBERS-2B e Ikonos-2 MS, respectivamente⁵.

A ideia principal é comparar os métodos de classificação contextual OPF-MRF e SVM-MRF com o OPF e SVM padrão, respectivamente, utilizando diferentes modelos Markovianos, cujo parâmetro β é definido através dos métodos de otimização. Também foram definidas duas composições diferentes do vetor de características para serem avaliadas. Outra contribuição é comparar os resultados obtidos pelos métodos PSO, HS, NM e GD com o método BF.

Utilizando parâmetros selecionados apropriadamente, é possível notar que os resultados de alguns modelos Markovianos foram semelhantes entre si. Outra interessante conclusão com relação aos métodos de seleção de características é que, na média, com o vetor de características (II), os resultados obtiveram taxas de reconhecimento ligeiramente melhores, mas ainda similares aos obtidos com o vetor de características (I) para ambas técnicas SVM-MRF e OPF-MRF. Tais resultados são muito interessantes, uma vez que pode ser um indicativo de que não é necessário analisar toda a vizinhança de uma dada instância. Embora a imagem Ikonos-2 MS tenha menor resolução do que a imagem CBERS-2B, parece ser mais complicado efetuar sua classificação, uma vez que possui maior número de classes. É possível notar pequenas regiões de diferentes classes nas regiões inferior esquerda e superior direita da Figura 6.3c.

O próximo experimento tem o intuito de verificar a viabilidade de modelar o problema de ajustar o parâmetro β como sendo um problema baseado em otimização evolutiva ou

⁴Procedimento executado sobre o mapa de rótulos, o qual foi gerado em cada iteração pelo OPF-MRF.

⁵Foi utilizado o método BF para a otimização do parâmetro neste primeiro experimento, uma vez que a ideia era apenas avaliar a importância dos diferentes algoritmos de extração de características.

Tabela 6.22: Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II).

Técnica	(I)	(II)
OPF (<i>baseline</i>)	87.17% ± 0.44	87.42% ± 0.47
OPF-MRF/Potts	90.15%±0.70	90.90%±0.40
OPF-MRF/GIMLL-L1	89.63%±0.42	90.38%±0.39
OPF-MRF/GIMLL-L2	89.80%±0.45	90.33%±0.25
OPF-MRF/GMRF	89.54%±0.54	89.64%±0.31
SVM (<i>baseline</i>)	87.35%±0.32	87.93% ± 0.26
SVM-MRF/Potts	89.73%±0.35	90.36%±0.23
SVM-MRF/GIMLL-L1	89.61%±0.21	90.28%±0.47
SVM-MRF/GIMLL-L2	89.67%±0.24	90.06%±0.35
SVM-MRF/GMRF	89.00%±0.28	89.34%±0.20

Tabela 6.23: Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II).

Técnica	(I)	(II)
OPF (<i>baseline</i>)	78.74% ± 0.15	85.83% ± 0.36
OPF-MRF/Potts	85.40%±0.42	88.65%±0.67
OPF-MRF/GIMLL-L1	81.72% ± 0.38	87.87%±0.53
OPF-MRF/GIMLL-L2	81.10% ± 0.16	87.88%±0.53
OPF-MRF/GMRF	81.35% ± 0.11	87.40%±0.49
SVM (<i>baseline</i>)	81.79% ± 0.15	86.92%±0.36
SVM-MRF/Potts	85.50%±0.46	87.88%±0.60
SVM-MRF/GIMLL-L1	84.31% ± 0.33	87.37%±0.47
SVM-MRF/GIMLL-L2	84.64%±0.50	87.63%±0.57
SVM-MRF/GMRF	83.58% ± 0.13	87.05%±0.58

baseado em otimização numérica. Assim, as Tabelas 6.24 e 6.26 apresentam as acurácias médias e os desvios padrões sobre as imagens CBERS-2B e Ikonos-2 MS, respectivamente, obtidas a partir de uma validação cruzada com 10 rodadas com OPF-MRF para todos os modelos Markovianos com o parâmetro otimizado por HS, PSO, NM e GD. Resultados obtidos pelo SVM-MRF podem ser vistos nas Tabelas 6.25 e 6.27, para as imagens CBERS-2B e Ikonos-2 MS, respectivamente. Nessas Tabelas, é possível chegarmos a algumas conclusões: (i) primeiramente, os resultados estendem o trabalho da seção 6.1.2, o qual foi observado que o modelo de Potts alcançou melhor desempenho do que os modelos GIMLL-L1, GIMLL-L2 e GMRF, mas nesse experimento é possível notar que todos modelos obtiveram resultados similares entre si quando são empregadas uma abordagem confiável de ajuste fino de parâmetro e um método adequado de extração de características; (ii) em geral, os métodos de otimização obtiveram resultados similares aos obtidos pelo método BF (ver Tabelas 6.22 e 6.23), mas no que diz respeito ao tempo de execução médio, é possível concluir que tais abordagens foram mais eficientes analisando os resultados obtidos nas Tabelas 6.32 e 6.34 para o OPF MRF e nas Tabelas 6.33 e 6.35 para o SVM-MRF.

Tabela 6.24: Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.

	Técnica	HS	PSO	NM	GD
(I)	OPF (<i>baseline</i>)	87.17% ± 0.44			
	OPF-MRF/Potts	90.29%±0.37	90.11%±0.75	90.18%±0.66	90.13%±0.60
	OPF-MRF/GIMLL-L1	89.53%±0.49	89.49%±0.50	89.67%±0.38	89.67%±0.46
	OPF-MRF/GIMLL-L2	89.77%±0.50	89.76%±0.59	89.79%±0.36	89.86%±0.33
	OPF-MRF/GMRF	89.03%±0.55	89.24%±0.46	89.36%±0.35	89.34%±0.39
(II)	OPF (<i>baseline</i>)	87.42% ± 0.47			
	OPF-MRF/Potts	90.82%±0.22	90.85%±0.33	90.74%±0.61	90.72%±0.56
	OPF-MRF/GIMLL-L1	90.33%±0.45	90.32%±0.49	90.36%±0.42	90.21%±0.55
	OPF-MRF/GIMLL-L2	90.33%±0.38	90.35%±0.43	90.34%±0.31	90.19%±0.42
	OPF-MRF/GMRF	89.71%±0.32	89.69%±0.33	89.61%±0.52	89.48%±0.58

A seguir, foi avaliado um esquema de classificação contextual híbrida, ou seja, uma classificação contextual com OPF-MRF seguida por um pós-processamento. Assim, as Tabelas 6.28 e 6.29 apresentam as acurácias médias sobre as imagens CBERS-2B e Ikonos-2 MS, respectivamente, obtidas a partir de uma validação cruzada com 10 rodadas com OPF-MRF para todos os modelos Markovianos com o parâmetro otimizado por BF, HS,

Tabela 6.25: Taxa de acurácia média usando diferente modelos Markovianos sobre a imagem CBERS-2B com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.

	Técnica	HS	PSO	NM	GD
(I)	SVM (<i>baseline</i>)	87.35%±0.32			
	SVM-MRF/Potts	89.64%±0.35	89.57%±0.31	89.68%±0.33	89.38%±0.68
	SVM-MRF/GIMLL-L1	89.65%±0.21	89.62%±0.19	89.58%±0.19	89.36%±0.34
	SVM-MRF/GIMLL-L2	89.67%±0.23	89.60%±0.24	89.64%±0.25	89.54%±0.29
	SVM-MRF/GMRF	88.91%±0.28	89.00%±0.24	88.91%±0.30	88.38%±0.31
(II)	SVM (<i>baseline</i>)	87.93% ± 0.26			
	SVM-MRF/Potts	90.44%±0.33	90.48%±0.22	90.35%±0.36	90.07%±0.74
	SVM-MRF/GIMLL-L1	90.14%±0.48	90.14%±0.38	90.38%±0.23	89.69%±1.17
	SVM-MRF/GIMLL-L2	90.24%±0.31	90.05%±0.26	90.19%±0.20	89.28%±0.65
	SVM-MRF/GMRF	89.28%±0.71	89.32%±0.29	88.81%±0.97	88.15%±1.09

Tabela 6.26: Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.

	Técnica	HS	PSO	NM	GD
(I)	OPF (<i>baseline</i>)	78.74% ± 0.15			
	OPF-MRF/Potts	85.26%±0.42	85.39%±0.36	84.96%±0.31	85.20%±0.65
	OPF-MRF/GIMLL-L1	81.58%±0.32	81.84%±0.36	81.70%±0.29	81.74%±0.36
	OPF-MRF/GIMLL-L2	81.03%±0.20	81.00%±0.28	81.01%±0.31	81.13%±0.30
	OPF-MRF/GMRF	80.82%±0.15	80.93%±0.06	80.91%±0.06	80.43%±0.68
(II)	OPF (<i>baseline</i>)	85.83% ± 0.36			
	OPF-MRF/Potts	88.51%±0.61	88.51%±0.67	88.34%±0.58	88.84%±0.63
	OPF-MRF/GIMLL-L1	87.63%±0.40	87.66%±0.47	87.77%±0.47	87.70%±0.62
	OPF-MRF/GIMLL-L2	87.93%±0.50	87.90%±0.47	88.11%±0.64	87.98%±0.61
	OPF-MRF/GMRF	87.02%±0.97	87.07%±0.97	86.63%±1.18	86.13%±1.27

Tabela 6.27: Taxa de acurácia média usando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com algoritmos de extração de características (I) e (II). A(s) melhor(es) acurácia(s) de cada par (modelo Markoviano, algoritmo de extração de características) está(ão) em negrito.

	Técnica	HS	PSO	NM	GD
(I)	SVM (baseline)	81.79% ± 0.15			
	SVM-MRF/Potts	85.40%±0.50	85.41%±0.43	85.37%±0.53	85.20%±0.12
	SVM-MRF/GIMLL-L1	84.20%±0.28	84.31%±0.27	84.22%±0.29	84.24%±0.26
	SVM-MRF/GIMLL-L2	84.61%±0.29	84.61%±0.43	84.58%±0.40	84.38%±0.42
	SVM-MRF/GMRF	83.53%±0.13	83.51%±0.16	83.38%±0.44	82.7%±0.78
(II)	SVM (baseline)	86.92%±0.36			
	SVM-MRF/Potts	87.63%±0.54	87.68%±0.66	87.74%±0.62	87.48%±0.55
	SVM-MRF/GIMLL-L1	87.41%±0.52	87.43%±0.30	87.61%±0.58	87.29%±0.41
	SVM-MRF/GIMLL-L2	87.14%±0.68	87.41%±0.43	87.61%±0.54	87.33%±0.50
	SVM-MRF/GMRF	87.10%±0.58	86.75%±0.84	86.82%±0.69	86.65%±0.25

PSO, NM e GD⁶. Resultados obtidos pelo SVM-MRF podem ser vistos nas Tabelas 6.30 e 6.31 para as imagens CBERS-2B e Ikonos-2 MS, respectivamente. É possível notar que a abordagem de pós-processamento proposta para evitar a correção excessiva em regiões de alta frequência foi eficaz para ambos modelos discretos e contínuos, porém sendo mais adequada para modelos contínuos. Uma vez que o modelo de Potts foi especialmente projetado para trabalhar sobre variáveis discretas, aparentemente os resultados iniciais já foram muito próximos dos melhores resultados, isto é, não houve ganho significativo com a técnica de pós-processamento, enquanto que os resultados dos modelos contínuos (GIMLL-L1, GIMLL-L2 e GMRF) foram melhorados, especialmente na imagem Ikonos-2 MS, denotando que os resultados de classificação contextual usando o modelo de Potts produzem menos artefatos no mapa de rótulos que os demais modelos.

As Figuras 6.34 e 6.35 apresentam alguns resultados de classificação sobre as imagens CBERS-2B e Ikonos-2 MS, respectivamente. Primeiramente, é possível notar que o OPF-MRF reduziu o efeito de sal e pimenta em relação ao resultado do OPF padrão. Além disso, a abordagem proposta de pós-processamento reduziu a correção excessiva em regiões compostas por instâncias próximas de classes diferentes (por exemplo, na região superior central das Figuras 6.35e e 6.35f).

Embora o tempo de execução médio tenha diminuído com os métodos de otimização, efetuar uma classificação contextual em tempo viável ainda é um desafio. Além disso, o parâmetro β para classificação contextual é estimado para ser usado para toda ima-

⁶O símbolo ‘-’ denota que a abordagem de pós-processamento não foi aplicada.

Tabela 6.28: Acurácias médias obtidas pelo OPF-MRF considerando diferentes modelos Markovianos sobre a imagem CBERS-2B com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.

Técnica	M	BF	HS	PSO	NM	GD
OPF		87.42% \pm 0.37				
OPF-MRF/ Potts	-	90.90% \pm 0.40	90.82% \pm 0.22	90.85% \pm 0.33	90.74% \pm 0.61	90.72% \pm 0.56
	3	88.03% \pm 0.45	88.03% \pm 0.44	88.05% \pm 0.43	87.88% \pm 0.56	87.88% \pm 0.56
	4	91.08%\pm0.35	91.00%\pm0.35	90.90% \pm 0.42	90.99% \pm 0.57	90.98%\pm0.91
	5	91.03% \pm 0.22	90.96% \pm 0.39	91.05%\pm0.41	91.05%\pm0.58	90.75% \pm 0.78
OPF-MRF/ GIMLL-L1	-	90.38% \pm 0.39	90.33% \pm 0.45	90.32% \pm 0.49	90.36% \pm 0.42	90.21% \pm 0.55
	3	87.79% \pm 0.40	87.77% \pm 0.41	87.78% \pm 0.39	87.71% \pm 0.43	87.74% \pm 0.43
	4	90.79%\pm0.42	90.75%\pm0.45	90.77%\pm0.44	90.79% \pm 0.46	90.77%\pm0.51
	5	90.71% \pm 0.43	90.74% \pm 0.45	90.71% \pm 0.42	90.81%\pm0.31	90.73% \pm 0.33
OPF-MRF/ GIMLL-L2	-	90.33% \pm 0.25	90.33% \pm 0.38	90.35% \pm 0.43	90.34% \pm 0.31	90.19% \pm 0.42
	3	87.81% \pm 0.47	87.82% \pm 0.38	87.78% \pm 0.40	87.69% \pm 0.45	87.74% \pm 0.51
	4	90.82%\pm0.4	90.80%\pm0.50	90.80%\pm0.46	90.81%\pm0.52	90.81%\pm0.41
	5	90.70% \pm 0.42	90.77% \pm 0.36	90.73% \pm 0.36	90.77% \pm 0.39	90.76% \pm 0.34
OPF-MRF/ GMRF	-	89.64% \pm 0.31	89.71% \pm 0.32	89.69% \pm 0.33	89.61% \pm 0.52	89.48% \pm 0.58
	3	87.71% \pm 0.46	87.68% \pm 0.45	87.70% \pm 0.43	87.66% \pm 0.51	87.60% \pm 0.52
	4	90.26%\pm0.44	90.30%\pm0.43	90.31%\pm0.43	90.37%\pm0.59	90.14%\pm0.57
	5	90.15% \pm 0.31	90.08% \pm 0.29	90.17% \pm 0.26	90.21% \pm 0.36	90.11% \pm 0.44

Tabela 6.29: Acurácias médias obtidas pelo OPF-MRF considerando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.

Técnica	M	BF	HS	PSO	NM	GD
OPF		85.87% \pm 0.42				
OPF-MRF/ Potts	-	88.65% \pm 0.67	88.51% \pm 0.61	88.51% \pm 0.67	88.34% \pm 0.58	88.84% \pm 0.63
	3	88.00% \pm 0.40	88.13% \pm 0.36	88.06% \pm 0.40	87.88% \pm 0.56	87.99% \pm 0.38
	4	88.88%\pm0.61	89.00%\pm0.63	88.96%\pm0.67	88.73% \pm 0.40	88.89% \pm 0.61
	5	88.85% \pm 0.79	88.80% \pm 0.69	88.88% \pm 0.72	88.87%\pm0.70	89.00%\pm0.68
OPF-MRF/ GIMLL-L1	-	87.87% \pm 0.53	87.63% \pm 0.40	87.66% \pm 0.47	87.77% \pm 0.47	87.70% \pm 0.62
	3	88.01% \pm 0.48	88.02% \pm 0.41	88.04% \pm 0.43	87.71% \pm 0.43	88.02% \pm 0.37
	4	88.77%\pm0.39	88.88%\pm0.55	88.87%\pm0.44	88.82%\pm0.12	88.79%\pm0.41
	5	88.15% \pm 0.55	88.10% \pm 0.54	88.13% \pm 0.53	88.30% \pm 0.53	88.11% \pm 0.76
OPF-MRF/ GIMLL-L2	-	87.88% \pm 0.53	87.93% \pm 0.50	87.90% \pm 0.47	88.11% \pm 0.64	87.98% \pm 0.61
	3	88.03% \pm 0.49	88.02% \pm 0.50	88.03% \pm 0.49	87.69% \pm 0.45	87.99% \pm 0.38
	4	89.09%\pm0.58	89.00%\pm0.69	89.02%\pm0.51	88.98%\pm0.47	89.02%\pm0.50
	5	88.43% \pm 0.72	88.33% \pm 0.66	88.46% \pm 0.64	88.50% \pm 0.75	88.46% \pm 0.70
OPF-MRF/ GMRF	-	87.40% \pm 0.49	87.02% \pm 0.97	87.07% \pm 0.97	86.63% \pm 1.18	86.13% \pm 1.27
	3	88.16% \pm 0.41	88.10% \pm 0.38	88.14% \pm 0.38	87.66% \pm 0.51	87.90% \pm 0.40
	4	88.86%\pm0.57	88.66%\pm0.56	88.72%\pm0.54	88.56%\pm0.36	88.22%\pm0.42
	5	87.87% \pm 0.53	87.73% \pm 0.67	87.81% \pm 0.59	87.42% \pm 0.73	87.37% \pm 0.65

Tabela 6.30: Acurácias médias obtidas pelo SVM-MRF considerando diferentes modelos Markovianos sobre a imagem CBERS-2 B com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.

Técnica	M	BF	HS	PSO	NM	GD
SVM		87.93% \pm 0.22				
SVM-MRF/ Potts	-	90.36% \pm 0.23	90.44% \pm 0.32	90.48% \pm 0.22	90.35% \pm 0.35	90.07% \pm 0.74
	3	87.09% \pm 0.52	87.07% \pm 0.53	87.09% \pm 0.54	87.09% \pm 0.52	87.06% \pm 0.62
	4	90.08% \pm 0.42	90.05% \pm 0.34	90.10% \pm 0.39	90.13% \pm 0.46	89.92% \pm 0.53
	5	90.53%\pm0.34	90.47%\pm0.32	91.05%\pm0.41	90.70%\pm0.32	90.41%\pm0.54
SVM-MRF/ GIMLL-L1	-	90.28% \pm 0.47	90.14% \pm 0.48	90.14% \pm 0.38	90.38% \pm 0.23	89.69% \pm 1.17
	3	87.02% \pm 0.47	87.02% \pm 0.55	87.02% \pm 0.46	87.00% \pm 0.48	86.98% \pm 0.47
	4	89.98% \pm 0.32	89.87% \pm 0.39	89.92% \pm 0.31	89.94% \pm 0.38	89.77% \pm 0.38
	5	90.51%\pm0.35	90.33%\pm0.22	90.71%\pm0.42	90.58%\pm0.12	89.96%\pm0.87
SVM-MRF/ GIMLL-L2	-	90.06% \pm 0.35	90.24% \pm 0.31	90.05% \pm 0.26	90.18% \pm 0.20	89.28% \pm 0.65
	3	87.00% \pm 0.48	87.00% \pm 0.47	86.98% \pm 0.46	86.98% \pm 0.46	86.94% \pm 0.47
	4	89.91% \pm 0.31	89.84% \pm 0.35	89.95% \pm 0.35	89.96% \pm 0.31	89.56% \pm 0.32
	5	90.26%\pm0.21	90.49%\pm0.27	90.73%\pm0.36	90.33%\pm0.25	89.94%\pm0.63
SVM-MRF/ GMRF	-	89.34% \pm 0.20	89.28% \pm 0.71	89.32% \pm 0.29	88.81% \pm 0.97	88.15% \pm 1.09
	3	86.89% \pm 0.51	86.89% \pm 0.52	86.90% \pm 0.53	86.83% \pm 0.48	86.73% \pm 0.37
	4	89.37% \pm 0.26	89.43%\pm0.27	89.39% \pm 0.24	89.14% \pm 0.55	89.07%\pm0.32
	5	89.40%\pm0.36	89.40% \pm 0.60	90.17%\pm0.26	89.15%\pm0.88	88.91% \pm 0.56

Tabela 6.31: Acurácias médias obtidas pelo SVM-MRF considerando diferentes modelos Markovianos sobre a imagem Ikonos-2 MS com β otimizado por BF, HS, PSO, NM e GD. A melhor acurácia de cada par (modelo Markoviano, técnica de otimização) está em negrito.

Técnica	M	BF	HS	PSO	NM	GD
SVM		86.92% \pm 0.36				
SVM-MRF/ Potts	-	87.88% \pm 0.60	87.63% \pm 0.54	87.68% \pm 0.67	87.74% \pm 0.62	87.48% \pm 0.55
	3	87.46% \pm 0.37	87.48% \pm 0.42	87.43% \pm 0.40	87.46% \pm 0.37	87.50% \pm 0.34
	4	88.30%\pm0.56	88.25%\pm0.61	88.15%\pm0.51	88.21%\pm0.52	88.12% \pm 0.41
	5	88.05% \pm 0.52	88.01% \pm 0.71	88.06% \pm 0.59	88.05% \pm 0.48	88.15%\pm0.55
SVM-MRF/ GIMLL-L1	-	87.37% \pm 0.47	87.41% \pm 0.52	87.43% \pm 0.30	87.61% \pm 0.57	87.29% \pm 0.41
	3	87.45% \pm 0.40	87.44% \pm 0.39	87.46% \pm 0.41	87.41% \pm 0.35	87.44% \pm 0.40
	4	88.18%\pm0.52	88.24%\pm0.51	88.11%\pm0.44	88.22%\pm0.52	88.00% \pm 0.53
	5	87.76% \pm 0.63	87.96% \pm 0.64	87.89% \pm 0.53	87.93% \pm 0.53	88.02%\pm0.63
SVM-MRF/ GIMLL-L2	-	87.63% \pm 0.57	87.14% \pm 0.68	87.41% \pm 0.43	87.61% \pm 0.54	87.33% \pm 0.50
	3	87.46% \pm 0.40	87.47% \pm 0.40	87.47% \pm 0.40	87.46% \pm 0.41	87.44% \pm 0.35
	4	88.24%\pm0.52	88.20%\pm0.48	88.26%\pm0.50	88.22%\pm0.53	88.08%\pm0.49
	5	88.09% \pm 0.57	87.86% \pm 0.33	87.81% \pm 0.31	88.02% \pm 0.54	87.98% \pm 0.65
SVM-MRF/ GMRF	-	87.05% \pm 0.58	87.10% \pm 0.58	86.75% \pm 0.84	86.82% \pm 0.69	86.65% \pm 0.25
	3	87.45% \pm 0.44	87.46% \pm 0.44	87.44% \pm 0.42	87.42% \pm 0.41	87.34% \pm 0.39
	4	88.03%\pm0.62	88.02%\pm0.67	88.03%\pm0.65	87.84%\pm0.63	87.72%\pm0.47
	5	87.60% \pm 0.73	87.71% \pm 0.64	87.58% \pm 0.59	87.46% \pm 0.70	87.40% \pm 0.26

Tabela 6.32: Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com OPF-MRF sobre a imagem CBERS-2B. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.

		HS	PSO	BF	NM	GD
(I)	Potts	3:36:52	6:36:07	6:52:55	2:04:36	1:38:03
	GIMLL-L1	5:15:44	6:44:24	7:20:26	1:50:33	1:35:58
	GIMLL-L2	4:32:08	6:55:45	7:06:33	2:11:27	1:33:59
	GMRF	4:50:47	6:07:10	10:22:34	2:08:01	1:46:10
(II)	Potts	03:05:21	04:14:26	05:13:41	1:28:56	1:00:24
	GIMLL-L1	03:10:37	04:01:57	05:17:43	1:31:32	1:02:07
	GIMLL-L2	03:00:48	04:18:10	05:13:37	1:35:17	1:15:50
	GMRF	03:40:30	03:45:57	06:14:07	1:42:09	1:02:56

Tabela 6.33: Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com SVM-MRF sobre a imagem CBERS-2B. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.

		HS	PSO	BF	NM	GD
(I)	Potts	12:12:03	12:55:51	14:03:33	10:54:57	10:13:50
	GIMLL-L1	12:18:56	13:29:13	14:14:24	11:01:56	10:17:53
	GIMLL-L2	12:21:17	13:14:41	14:11:47	11:01:25	10:19:53
	GMRF	12:45:37	13:28:07	14:59:27	11:06:57	10:18:24
(II)	Potts	10:14:46	10:52:16	11:36:45	8:23:14	7:31:28
	GIMLL-L1	9:28:49	10:54:40	11:37:05	8:19:19	7:29:42
	GIMLL-L2	9:42:22	10:29:09	11:28:35	8:29:03	7:37:57
	GMRF	10:38:17	11:38:15	13:06:04	8:30:41	7:36:12

Tabela 6.34: Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com OPF-MRF sobre a imagem Ikonos-2 MS. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) estão em negrito.

		HS	PSO	BF	NM	GD
(I)	Potts	0:15:57	0:22:46	0:27:34	0:04:27	0:06:36
	GIMLL-L1	0:20:50	0:28:16	0:28:58	0:08:23	0:05:27
	GIMLL-L2	0:19:42	0:26:44	0:28:55	0:08:23	0:06:22
	GMRF	0:19:55	0:26:25	0:42:37	0:08:43	0:08:12
(II)	Potts	00:11:51	00:14:17	00:18:26	0:04:37	0:03:42
	GIMLL-L1	00:13:03	00:16:17	00:20:05	0:05:42	0:04:09
	GIMLL-L2	00:12:47	00:16:49	00:19:48	0:05:38	0:04:40
	GMRF	00:15:07	00:15:27	00:25:14	0:06:50	0:06:36

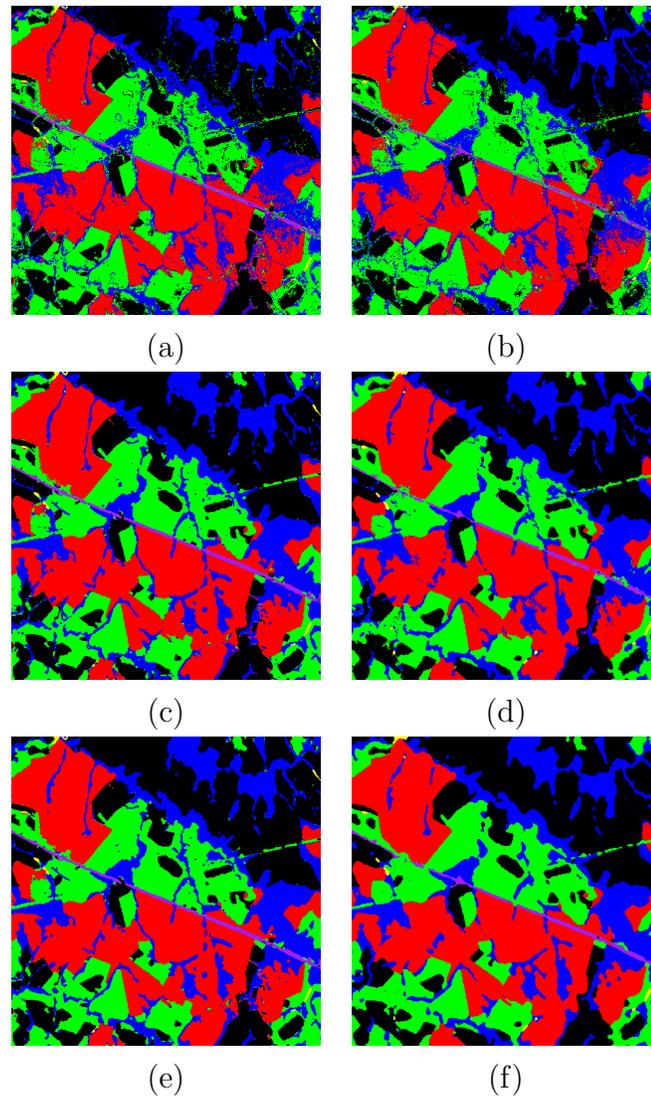


Figura 6.34: Resultados de Classificação sobre a imagem CBERS2-B obtidos pelo OPF padrão usando as abordagens de extração de características (I) e (II) em (a) e (b), respectivamente; resultados de classificação obtidos pelo OPF-MRF sem a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) em (c) e (d), respectivamente, e resultados de classificação obtidos pelo OPF-MRF com a etapa de pós-processamento usando as abordagens de extração de características (I) (valor da moda igual a 5) e (II) (valor da moda igual a 4) em (e) e (f), respectivamente. Os resultados de classificação considerando ambas abordagens de extração de características para o OPF-MRF foram obtidos através do modelo de Potts otimizado por meio do método BF.

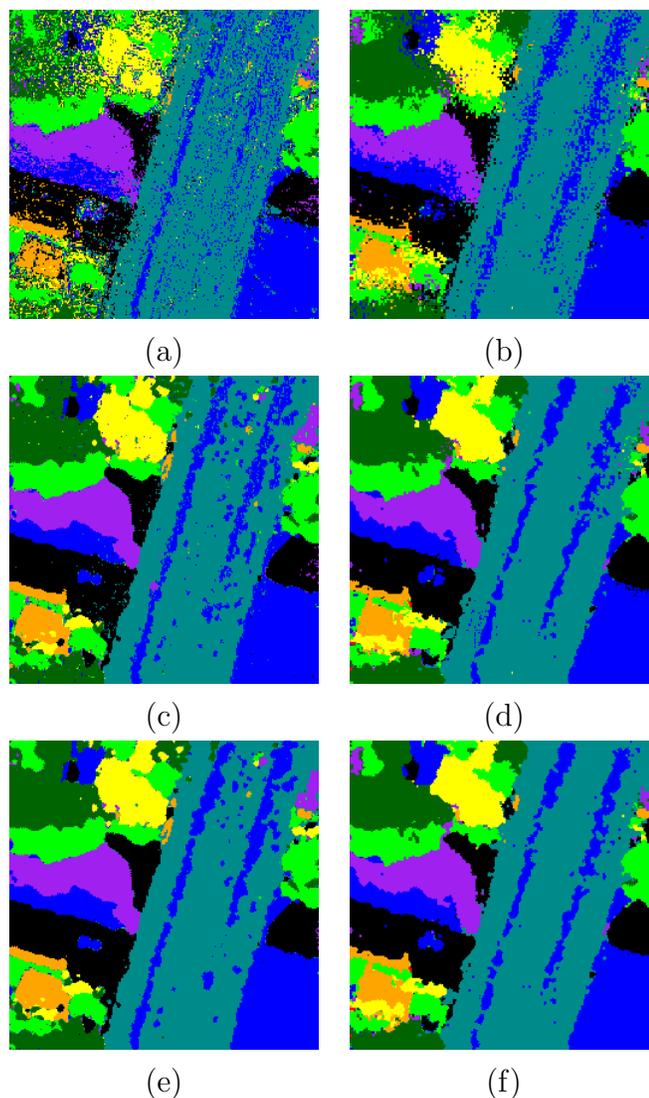


Figura 6.35: Resultados de Classificação sobre a imagem Ikonos-2 MS obtidos pelo OPF padrão usando as abordagens de extração de características (I) e (II) em (a) e (b), respectivamente; resultados de classificação obtidos pelo OPF-MRF sem a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) em (c) e (d), respectivamente, e resultados de classificação obtidos pelo OPF-MRF com a etapa de pós-processamento usando as abordagens de extração de características (I) e (II) (valor da moda igual a 5) em (e) e (f), respectivamente. Os resultados de classificação considerando ambas abordagens de extração de características para o OPF-MRF foram obtidos através do modelo de Potts otimizado por meio do método BF.

gem, não levando em consideração as particularidades de uma região. A seção seguinte apresenta os resultados obtidos pelo esquema proposto para enfrentar estas questões.

Tabela 6.35: Tempo de execução médio para estimar o parâmetro β e efetuar classificação contextual com SVM-MRF sobre a imagem Ikonos-2 MS. As técnicas mais rápidas de cada par (modelo Markoviano, técnica de extração de características) está em negrito.

		HS	PSO	BF	NM	GD
(I)	Potts	0:51:31	0:54:09	0:59:07	0:42:51	0:40:15
	GIMLL-L1	0:52:36	0:58:55	1:01:52	0:45:18	0:40:34
	GIMLL-L2	0:53:57	0:59:38	1:01:44	0:44:44	0:40:36
	GMRF	0:53:35	0:57:19	1:03:35	0:44:51	0:40:38
(II)	Potts	0:33:51	0:35:16	0:38:57	0:28:13	0:26:33
	GIMLL-L1	0:33:18	0:35:17	0:39:19	0:28:31	0:26:22
	GIMLL-L2	0:34:02	0:35:56	0:39:11	0:28:33	0:26:10
	GMRF	0:34:31	0:34:49	0:41:04	0:27:55	0:26:33

6.2.3 Classificação contextual em uma imagem dividida em blocos

Esta seção diz respeito ao artigo *A Block-based Markov Random Field Model Estimation for Contextual Classification Using Optimum-Path Forest* que foi aceito para publicação no *IEEE International Symposium on Circuits and Systems*, onde são apresentados os resultados obtidos pela classificação contextual utilizando OPF-MRF com modelo de Potts sobre imagens de sensoriamento remoto divididas em blocos. A ideia principal é dividir a imagem em blocos e classificá-los separadamente, associando um campo aleatório a cada bloco, bem como o parâmetro de dependência espacial, visando evitar a influência das regiões vizinhas na estimativa do parâmetro e também a carga computacional necessária para o seu processamento. Assim, os experimentos foram conduzidos sobre quatro imagens de sensoriamento remoto como a seguir: (i) cada imagem \mathcal{I} foi particionada em $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \dots \cup \mathcal{I}_r$ blocos, onde $r \in \{4, 8, 16\}$; (ii) cada *pixel* (i, j) de \mathcal{I}_s é uma instância representada por 5 características fornecidas pelos seus valores RGB e sua localização espacial; (iii) para cada bloco, as instâncias são divididas em conjuntos de treinamento (5%), validação (15%) e de teste (80%); O conjunto de validação é então utilizado para otimizar o parâmetro β_s para cada \mathcal{I}_s usando os métodos BF, bem como HS e PSO; (v) finalmente, a acurácia é calculada executando a junção dos resultados de classificação (mapa de rótulos) de cada bloco.

A Tabela 6.36 apresenta a acurácia média e o desvio padrão para as imagens particionadas em 4, 8 e 16 blocos, bem como os resultados obtidos por OPF-MRF e OPF sem particionamento da imagem. É possível observar que a abordagem proposta obteve os

melhores resultados para todas imagens usando 4 blocos, exceto para a imagem Landsat-5 TM, cujo melhor resultado foi obtido usando 8 blocos. Aparentemente, existe uma relação custo-benefício entre o número de blocos e a acurácia final, uma vez que menores conjuntos de treinamento conduzem à menores taxas de reconhecimento. Tais hipóteses são mantidas em situações onde pequenos conjuntos de treinamento não são suficientemente representativos para um bom processo de aprendizado.

Tabela 6.36: Acurácia média sobre as imagens CBERS-2B (I), Geoeye (II), Ikonos-2 MS (III) e Landsat-5 TM (IV) particionadas em 4, 8 e 16 blocos com o parâmetro β otimizado por BF, HS e PSO. Note que o símbolo ‘-’ significa que não houve particionamento da imagem, isto é, naïve OPF-MRF. O melhor resultado para cada imagem está em negrito.

		Técnica			
		OPF	OPF-MRF/ BF	OPF-MRF/ HS	OPF-MRF/ PSO
I	-	87.25%±0.24	90.73%±0.19	90.72%±0.42	90.86%±0.23
	4	87.29%±0.25	91.50%±0.25	91.53%±0.28	91.54%±0.23
	8	87.71%±0.27	91.40%±0.32	91.12%±0.26	91.05%±0.42
	16	87.55%±0.36	91.21%±0.56	91.32%±0.47	91.05%±0.86
II	-	88.07%±0.40	90.23%±0.90	90.08%±0.59	90.22%±0.67
	4	88.08%±0.42	90.69%±0.57	90.65%±0.66	90.58%±0.50
	8	87.45%±0.39	89.28%±0.49	89.33%±0.53	89.29%±0.52
	16	88.10%±0.45	90.26%±0.44	90.22%±0.36	90.25%±0.44
III	-	86.61%±0.74	89.47%±1.00	89.51%±1.05	89.42%±1.11
	4	86.60%±0.74	90.17%±0.89	90.11%±0.89	90.29%±0.65
	8	85.77%±0.57	89.08%±0.56	89.29%±0.46	89.06%±0.21
	16	85.84%±0.70	89.34%±0.65	89.41%±0.47	89.46%±0.47
IV	-	83.69%±0.44	86.82%±0.71	86.91%±1.02	86.69%±0.81
	4	83.84%±0.46	87.89%±0.65	88.06%±0.57	87.59%±0.99
	8	84.01%±0.44	88.28%±1.00	88.09%±0.83	88.43%±1.04
	16	84.33%±0.68	88.22%±0.80	87.85%±0.85	87.77%±0.69

Embora alguns resultados tenham sido similares ao OPF-MRF padrão, sua aplicação é justificada quando é levada em consideração a redução da carga computacional obtida, como podem ser observados na Tabela 6.37, cujos tempos totais de execução foram calculadas através do somatório da carga computacional para cada bloco. Assim, ainda é possível ser beneficiado com uma implementação paralela do OPF-MRF baseado em blocos. Considerando tal possibilidade, a eficiência ainda pode ser aumentada por um fator

de n , onde n é o número de blocos usados. Mesmo considerando uma execução sequencial de cada bloco, a carga computacional pode ser dramaticamente reduzida.

Tabela 6.37: Tempo de execução médio usando 4, 8 e 16 blocos com o parâmetro β otimizado por BF, HS e PSO. Note que o símbolo ‘-’ significa que não houve particionamento da imagem, isto é, naïve OPF-MRF. Os resultados são mostrados no seguinte formato $h:m:s$, onde h , m e s denotam horas, minutos e segundos, respectivamente.

		CBERS-2B	Geoeye	Ikonos-2 MS	Landsat-5 TM
BF	-	5:10:08	0:21:01	0:19:47	5:07:32
	4	1:16:08	0:04:36	0:04:18	1:13:33
	8	0:33:29	0:02:08	0:01:58	0:32:23
	16	0:15:49	0:00:50	0:00:46	0:14:52
HS	-	3:22:41	0:14:03	0:13:25	3:06:41
	4	0:51:14	0:03:06	0:02:52	0:49:20
	8	0:22:18	0:01:26	0:01:15	0:22:04
	16	0:10:27	0:00:26	0:00:26	0:9:52
PSO	-	4:11:14	0:18:26	0:16:27	3:43:38
	4	1:02:03	0:03:58	0:03:33	1:00:46
	8	0:26:17	0:01:41	0:01:33	0:27:28
	16	0:12:07	0:00:32	0:00:35	0:12:10

As Figuras 6.36, 6.37, 6.38 e 6.39 apresentam os resultados de classificação sobre as imagens CBERS-2B, Landsat-5 TM, Ikonos-2 MS e Geoeye, respectivamente. Para todas imagens, particionando a imagem em 4 e 8 blocos, os resultados de classificação com contêm regiões mais heterogêneas do que com OPF e naïve OPF-MRF (regiões destacadas). Contudo, alguns erros de classificação associados ao particionamento das imagens podem ser notados, especificamente na fronteira entre blocos.

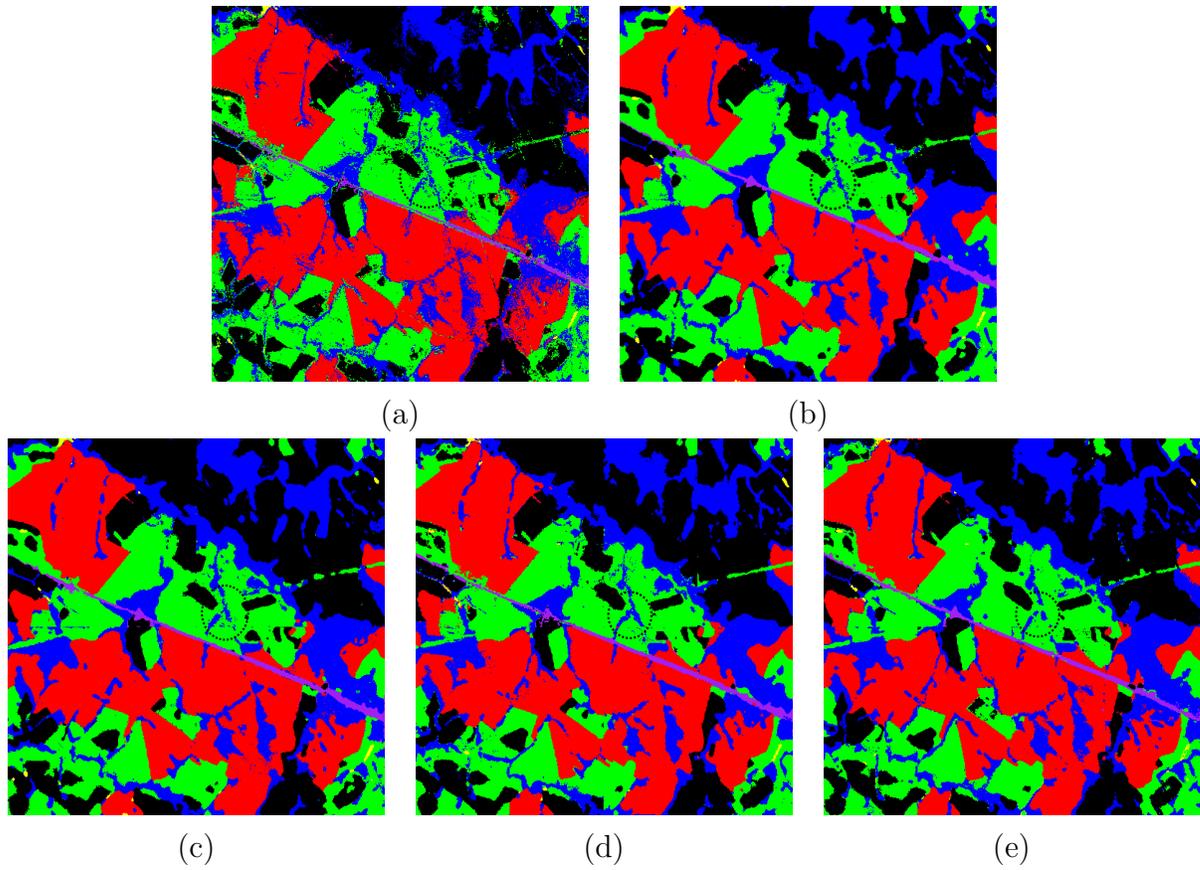


Figura 6.36: Imagem CBERS-2B classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos.

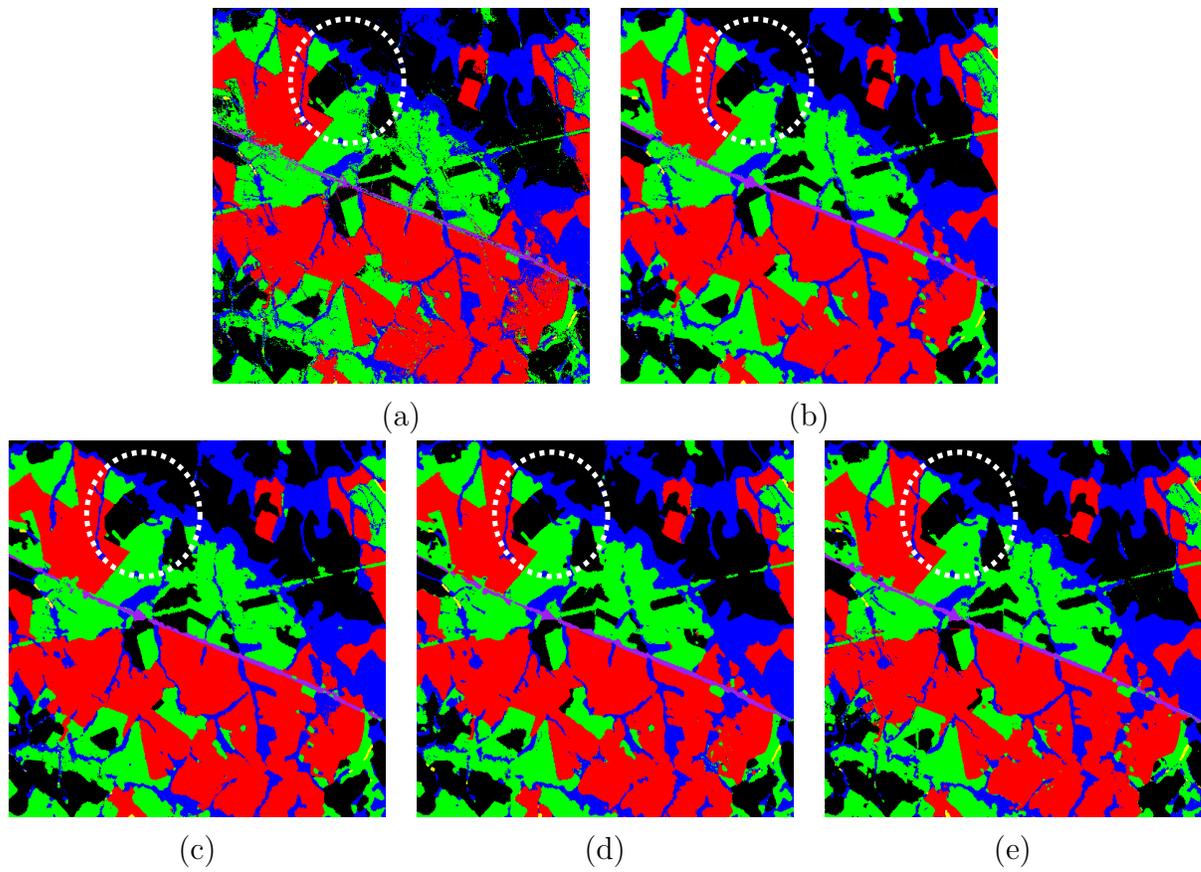


Figura 6.37: Imagem Landsat-5 TM classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos.

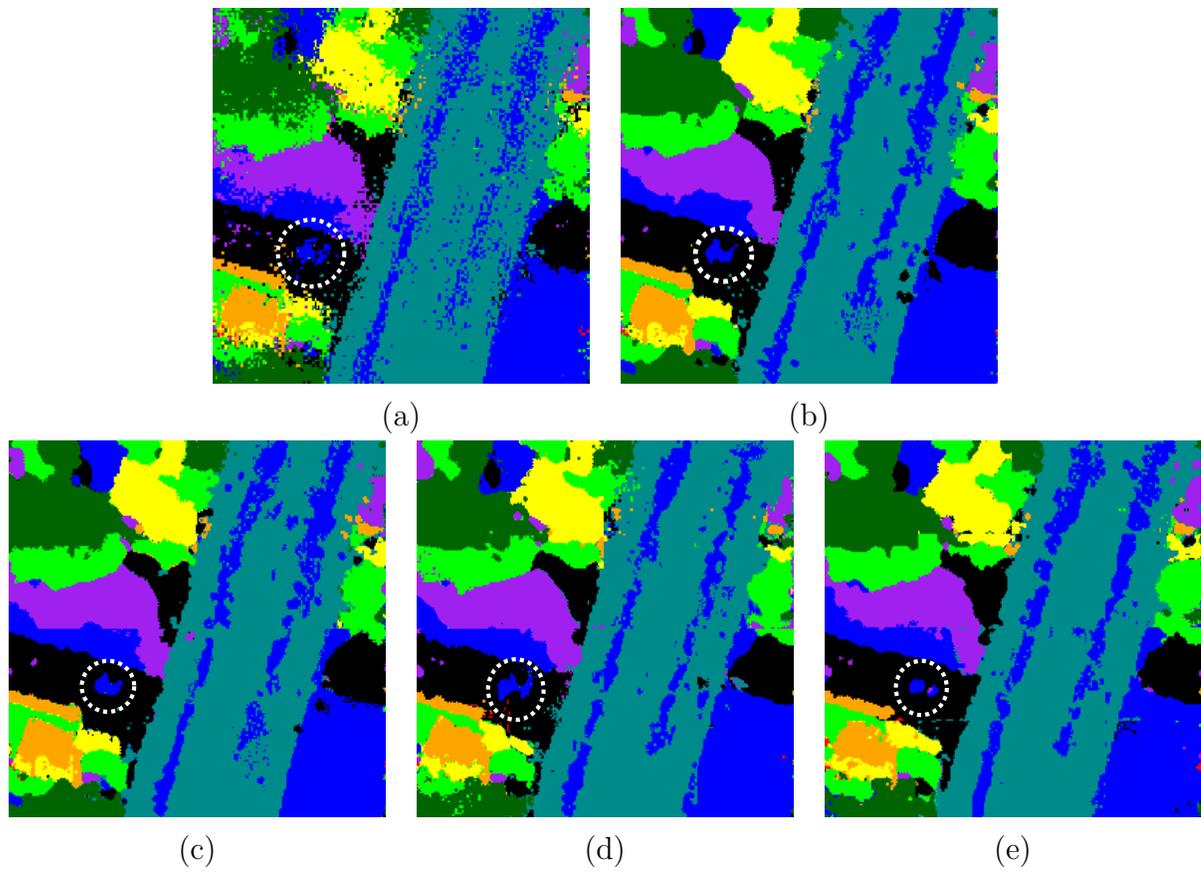


Figura 6.38: Imagem IKONOS-2 MS classificada por: (a) OPF, (b) naïve OPF-MRF, e OPF-MRF baseado em blocos otimizado por BF com (c) 4, (d) 8 and (e) 16 blocos.

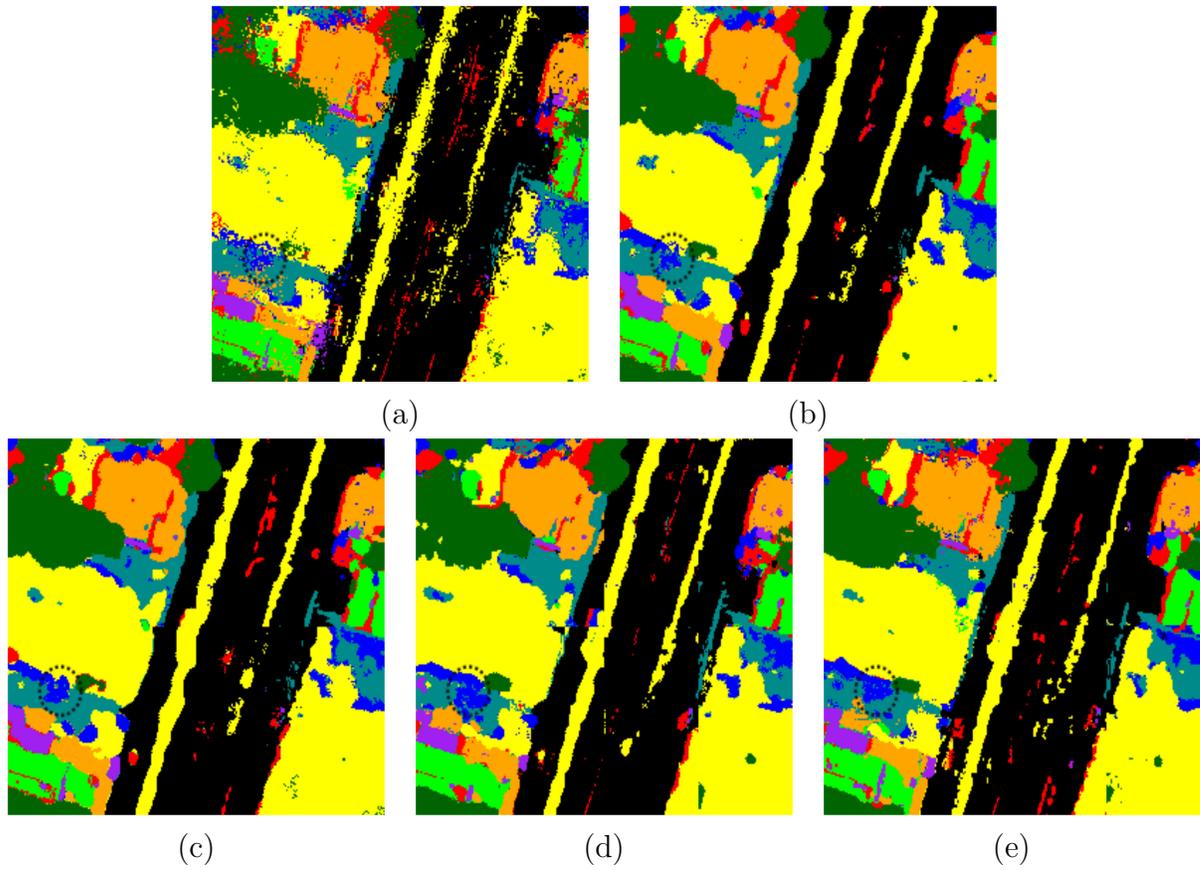


Figura 6.39: Imagem Geoeye classificada por: (a) OPF, (b) naïve OPF-MRF e OPF-MRF baseado em blocos otimizados por BF com (c) 4, (d) 8 e (e) 16 blocos.

Capítulo 7

Conclusões

7.1 Considerações finais

Técnicas de classificação de padrões baseadas em *pixels* têm sido empregadas por um longo tempo. No entanto, muitos trabalhos têm chamado a atenção para técnicas de classificação “baseadas em contexto”, as quais também consideram a vizinhança das instâncias no processo de decisão.

Uma das formas de modelar a informação contextual é empregar Campos Aleatórios Markovianos em conjunto com uma função de energia, a qual é maximizada através de um processo de aprendizagem. Assim, foi proposta uma nova abordagem para explorar informação contextual por meio do classificador OPF chamada OPF-MRF, onde a informação contextual é tratada em termos da maximização de uma probabilidade condicional dada por um modelo Markoviano em conjunto com a minimização de custo de caminho fornecido pelo classificador OPF padrão. Tal modelo permite o uso de informação adicional fornecida pela vizinhança de uma dada instância. Embora tenham sido utilizadas 10 iterações para o OPF-MRF, é possível notar que o máximo local pode ser alcançado com 6 a 8 iterações, pois o ganho a partir dessas iterações não é significativo, fato que já havia sido relatado por Besag [76] e percebido em todos os experimentos.

Com o intuito de apresentar uma aplicação das abordagens propostas, foi abordada a classificação de tecidos sobre o conjunto de dados IBSR, que contém imagens MRI de cérebros humanos. Diferentes valores de β (restrição espacial) escolhidos empiricamente foram avaliados com o intuito de encontrar um valor que forneça a melhor taxa de reconhecimento para o OPF-MRF. Com os resultados obtidos nos experimentos, ficou evidenciado que OPF-MRF obteve melhores resultados do que OPF padrão em todos os casos, o que parece ser uma abordagem adequada para classificação contextual. Contudo, também é possível notar que tal abordagem de classificação é fortemente dependente da solução inicial, ou seja, um resultado não satisfatório obtida pela classificação tradicional

poderia afetar o desempenho da classificação contextual, visto que a precisão no cálculo da densidade condicional local está intimamente ligada à qualidade do mapa de rótulos fornecido pela classificação tradicional.

Uma vez que o intervalo do parâmetro β de restrição espacial poderia conter infinitos valores, sendo a escolha empírica uma tarefa que poderia tornar-se de difícil escolha, foi proposto um esquema de otimização por meta-heurísticas para encontrar valores adequados do parâmetro β para o modelo de Potts, complementando o trabalho anterior de Nakamura et al. [87], que propuseram o OPF-MRF, uma abordagem baseada em contexto que combina o classificador Floresta de Caminhos Ótimos com Campos Aleatórios Markovianos. É importante destacar que a abordagem proposta pode ser usada com qualquer algoritmo de otimização, sendo que foram empregadas duas abordagens: Otimização por Enxame de Partículas e Busca Harmônica. Os experimentos compararam PSO e HS contra a busca por Força Bruta (pseudo-exaustiva) sobre o intervalo de valores possíveis para β em 6 imagens MRI do cérebro. Outra contribuição deste trabalho é a metodologia utilizada para encontrar o β , onde foi empregado um conjunto de treinamento e validação para tal propósito, sendo a imagem refinada com o algoritmo ICM. Conclui-se que o PSO e HS podem alcançar resultados similares aos obtidos por BF, mas podem ser muito mais rápidos do que a busca pseudo-exaustiva (algumas vezes acima de 71%). Este comportamento foi observado em todas imagens.

Posteriormente, também foi avaliado o OPF-MRF proposto no contexto de classificação da cobertura do solo em quatro imagens de satélite fornecidas pelos sensores CBERS-2B, Landsat 5 TM, Ikonos-2 MS e Geoeye, sendo seus resultados melhores do que os obtidos pelo OPF padrão. Além disso, foi avaliado o esquema de otimização por meta-heurísticas para encontrar valores adequados do parâmetro β . Para tal propósito, foram empregadas as técnicas de otimização PSO, HS, IHS e GHS, as quais foram comparadas com uma busca pseudo-exaustiva sobre uma faixa de possíveis valores de β . Pode-se concluir que a abordagem proposta pôde encontrar valores similares aos valores “ótimos” encontrados pela busca pseudo-exaustiva. A abordagem proposta também foi avaliada considerando o classificador SVM, mostrando que pode ser aplicado à qualquer técnica de reconhecimento de padrões.

Além disso, foi avaliada a influência de diferentes modelos Markovianos para classificação baseada em contexto por meio do classificador Floresta de Caminhos Ótimos na tarefa de classificação do uso do solo em duas imagens de satélite. Foi observado que o modelo de Potts apresentou melhor resultado para ambas imagens, uma vez que foi projetado para lidar com situações em que as variáveis assumem apenas valores discretos, ao contrário dos modelos GIMLL-L1, GIMLL-L2 e GMRF, os quais assumem uma distribuição contínua das variáveis.

Também foi proposto um esquema de otimização utilizando métodos evolutivos e méto-

dos numéricos para os modelos Markovianos não-discretos, bem como uma abordagem de classificação contextual seguida por um pós-processamento para refinar o mapa de rótulos com o intuito de diminuir o efeito de sal e pimenta remanescentes, porém evitando a correção excessiva em regiões de alta frequência. Os resultados obtidos demonstraram que a abordagem proposta foi capaz de produzir melhores resultados do que o OPF-MRF original, bem como ambos modelos discreto e não-discretos obtiveram resultados semelhantes entre si quando é empregado um ajuste adequado do parâmetro de dependência espacial, sendo o método Nelder-Mead, o esquema de otimização que apresentou o melhor custo-benefício em termos de acurácia média e tempo de execução. Embora o método de otimização GD tenha utilizado menor carga computacional, o resultado em termos de acurácia média foi pouco conclusivo no que diz respeito à convexidade do espaço de busca, uma vez que a quantidade de mínimos locais está associada ao problema (imagem) e, portanto, não sendo recomendado para resolução deste tipo de problema.

Com relação à abordagem de classificação contextual em imagens divididas em blocos, é possível observar melhores taxas de reconhecimento aplicando OPF-MRF em uma imagem dividida em blocos. Outro ponto positivo é a carga computacional, que é menor do que o OPF-MRF tradicional. Contudo, é necessário analisar o custo benefício entre o número de blocos e a taxa de reconhecimento final, uma vez que quanto menor o tamanho do conjunto de treinamento, menor será a acurácia do classificador.

Finalmente, foi proposta uma nova modelagem da informação contextual baseando-se na teoria da informação, onde foi utilizado um modelo baseado na Informação de Fisher Observada Local para ser utilizado como um modelo Markoviano. Os resultados obtidos mostraram que a utilização da Informação de Fisher como medida da quantidade de informação local pode ser uma boa forma de extrair informações contextuais que auxiliem na etapa de classificação, muito embora os resultados obtidos tenham sido similares aos comparados para algumas imagens.

7.2 Contribuições

A principal contribuição desta tese é uma abordagem de classificação contextual de imagens por meio do Classificador Floresta de Caminhos Ótimos com Campos Aleatórios Markovianos. Embora existam diversas versões contextuais para o classificador SVM, ainda não existia nenhuma versão contextual para o classificador OPF na literatura.

No que diz respeito à abordagem para a estimativa do parâmetro de dependência espacial, a principal contribuição foi a elaboração de uma metodologia para encontrar valores de parâmetros pseudo-ótimos que maximizam a função de densidade de probabilidade local como sendo uma tarefa de otimização em um tempo viável. Com relação à abordagem de classificação contextual de imagens divididas em blocos, as principais contribuições

foram o aumento da taxa de acerto do classificador contextual OPF-MRF e a redução da carga computacional para classificação contextual.

Outra contribuição desta tese foi modelagem da informação contextual empregando Informação de Fisher Observada Local, cujos resultados foram promissores, abrindo novas possibilidades de aplicação em outras etapas de classificação.

7.3 Trabalhos Futuros

A abordagem de classificação contextual de imagens apresentada leva em consideração a informação contextual no processo de classificação estendendo o vetor de características e adicionando novos componentes obtidos por meio de Campos Aleatórios Markovianos. Embora os resultados apresentados aparentemente foram promissores, é possível supor que refinamento da solução atinge até um certo ponto, pois a contribuição fornecida pelos atributos de uma instância para sua classificação permanece a mesma durante todo o processo de classificação. Assim, propõe-se uma abordagem de classificação contextual, que considera a informação contextual numa etapa de pós-processamento, a qual é feita uma regularização por MRF utilizando máxima probabilidade *a posteriori* e o algoritmo Metropolis para estimar o mapa de rótulos verdadeiro dado que um mapa de classificação inicial é fornecido. Também pretende-se propor uma versão composta pelo OPF-MRF seguida por uma regularização por MRF.

Com isso, tal abordagem abre a possibilidade de utilizar o conceito de classificação contextual no espaço de características e não apenas em dados espacialmente dependentes como imagens. Para tal, baseado na similaridade entre as instâncias, é possível determinar os vizinhos mais próximos no espaço de características e utilizá-los para o cálculo da probabilidade condicional local, assumindo a hipótese de que as instâncias próximas tendem a pertencer à mesma classe. Dessa forma, pretende-se investigar a efetividade de tal hipótese.

Capítulo 8

Trabalhos Publicados

Conferências

- Rodrigo Yuji Mizobe Nakamura, Daniel Osaku, Alexandre Luís Magalhães Levada, João Paulo Papa, Fábio Augusto Menocci Cappabianco e Alexandre Xavier Falcão. **OPF-MRF: Optimum-Path Forest and Markov Random Fields for Contextual-based Image Classification.** *Proceedings of the 15th International Conference on Computer Analysis of Images and Patterns*, York - Reino Unido, 2013.
- Daniel Osaku, Rodrigo Yuji Mizobe Nakamura, João Paulo Papa, Alexandre Luís Magalhães Levada, Fábio Augusto Menocci Cappabianco e Alexandre Xavier Falcão. **Optimizing Contextual-Based Optimum-Forest Classification through Swarm Intelligence.** *Advanced Concepts for Intelligent Vision Systems*, pp. 203–214, Poznan - Polônia, 2013.
- Daniel Osaku, Alexandre Luís Magalhães Levada, João Paulo Papa. **On the Influence of Markovian Models for Contextual-Based Optimum-Path Forest Classification.** *19th Iberoamerican Congress on Pattern Recognition*, Puerto Vallarta - México, 2014.
- Daniel Osaku, Alexandre Luís Magalhães Levada, João Paulo Papa. **A Block-Based Markov Random Field Model Estimation for Contextual Classification Using Optimum-Path Forest.** *IEEE International Symposium on Circuits & Systems*, Montreal, 2016. (aceito para publicação).
- Daniel Osaku, Alexandre Luís Magalhães Levada e João Paulo Papa. **Learning Contextual Classification with Fisher Information.** *SIBGRAPI 2016 - Conference on Graphics, Patterns and Images*, 2016, Brasil. (submetido).

Periódicos

- Daniel Osaku, Rodrigo Yuji Mizobe Nakamura, Luís Augusto Martins Pereira, Rodrigo José Pisani, Alexandre Luís Magalhães Levada, Fábio Augusto Menocci Cappabianco, Alexandre Xavier Falcão e João Paulo Papa. **Improving land cover classification through contextual-based optimum-path forest.** *Information Sciences*, 2015.
- Daniel Osaku, Danilo R. Pereira, Alexandre Luís Magalhães Levada, João Paulo Papa. **Fine-Tuning Contextual-based Optimum Path Forest for Land Cover Classification.** *IEEE Geoscience and Remote Sensing Letters*, 2016.

Referências Bibliográficas

- [1] S. Haykin. *Neural Networks: A comprehensive foundation*. Prentice-Hall, 1998.
- [2] C. Cortes and V. Vapnik. Support Vector Networks. *Machine Learning*, 20:273–297, 1995.
- [3] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2 edition, 2000.
- [4] T. Joachims. Transductive inference for text classification using Support Vector Machines. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 200–209, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.
- [5] C. T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transaction on Computers*, 20(1):68–86, 1971.
- [6] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki. Supervised pattern classification based on Optimum-Path Forest. *International Journal of Imaging Systems and Technology*, 19(2):120–131, 2009.
- [7] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares. Efficient supervised Optimum-Path Forest classification for large datasets. *Pattern Recognition*, 45(1):512–520, 2012.
- [8] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.
- [9] M. Monard, Carolina and J. A. Baranauskas. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, 1:1, 2003.
- [10] P. M. Atkinson and P. Lewis. Geostatistical classification for remote sensing: an introduction. *Computers & Geosciences*, 26(4):361–371, 2000.

- [11] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson. SVM and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 7(4):736–740, 2010.
- [12] G. Moser and S. B. Serpico. Combining Support Vector Machines and Markov Random Fields in an integrated framework for contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–19, 2012.
- [13] T. Wu, M.H. Bae, M. Zhang, R. Pan, and A. Badae. A prior feature SVM-MRF based method for mouse brain segmentation. *NeuroImage*, 59(3):2298–2306, 2012.
- [14] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.
- [15] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [16] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, 51(2):271–279, 1989.
- [17] J. Kittler and J. Föglein. Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2(1):13 – 29, 1984.
- [18] S. Shekhar, P.R. Schrater, R. R. Vatsavai, W. Weili, and S. Chawla. Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transactions on Multimedia*, 4(2):174–188, 2002.
- [19] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng. Adaptive Markov Random Field approach for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(5):973–977, 2011.
- [20] L. Bruzzone and L. Carlin. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2587–2600, 2006.
- [21] E. Binaghi, I. Gallo, and M. Pepe. A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2906–2922, 2003.

- [22] N. Cai, J. Yang, K. Hu, and H. Xiong. MRF-MBNN: a novel Neural Network architecture for image processing. In *Proceedings of the Second international conference on Advances in neural networks - Volume Part II*, pages 673–678, Berlin, Heidelberg, 2005. Springer-Verlag.
- [23] A. Laha, N. R. Pal, and J. Das. Land cover classification using Fuzzy rules and aggregation of contextual information through evidence theory. *IEEE Transactions on Geoscience and Remote Sensing*, 6(4):1633–1642, 2009.
- [24] B. Ghimire, J. Rogan, and J. Miller. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using Random Forests and the getis statistic. *Remote Sensing Letters*, 1(1):45–54, 2010.
- [25] J. Stuckens, P.R. Coppin, and M.E. Bauer. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment*, 71(3):282–296, 2000.
- [26] A. Sarkar, A. Banerjee, N. Banerjee, S. Brahma, B. Kartikeyan, M. Chakraborty, and K. L. Majumder. Landcover classification in MRF context using Dempster-Shafer fusion for multisensor imagery. *IEEE Transactions on Image Processing*, 14:634–645, 2005.
- [27] S. W. Wharton. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. *Pattern Recognition*, 15(4):317–324, 1982.
- [28] F. T. Mahmoudi, F. Samadzadegan, and P. Reinartz. Object oriented image analysis based on multi-agent recognition system. *Computers & Geosciences*, 54(0):219–230, 2013.
- [29] AL Martins, A Levada, Murillo RP Homem, and N Mascarenhas. MAP-MRF super-resolution image reconstruction using maximum pseudo-likelihood parameter estimation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1165–1168. IEEE, 2009.
- [30] R. Chellappa and S. Chatterjee. Classification of textures using Gaussian Markov Random Fields. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(4):959–963, Aug 1985.
- [31] B.S. Manjunath and R. Chellappa. Unsupervised texture segmentation using Markov Random Field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):478–482, 1991.

- [32] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, Jan 1968.
- [33] A. K Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.
- [34] R. T. Raittz. *Fan 2002: Um modelo neuro-Fuzzy para reconhecimento de padrões*. PhD thesis, Universidade Federal de Santa Catarina, 2002.
- [35] S. Watanabe. *Pattern Recognition: Human and Mechanical*. John Wiley & Sons, Inc., New York, NY, USA, 1985.
- [36] H. D. Lee. *Seleção e construção de features relevantes para o aprendizado de máquina*. PhD thesis, Universidade de São Paulo, 2000.
- [37] M. D. C. Nicoletti. *Ampliando os limites do aprendizado indutivo de máquina através das abordagens construtiva e relacional*. PhD thesis, Universidade de São Paulo, 1994.
- [38] K. Fukunaga and P.M. Narendra. A branch and bound algorithms for computing k-nearest neighbors. *IEEE Transactions on Computers*, 24(7):750–753, 1975.
- [39] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.
- [40] J. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. 2001.
- [41] C. Sutton and A. McCallum. An introduction to Conditional Random Fields. *Machine Learning*, 4(4):267–373, 2011.
- [42] J. Verbeek and W. Triggs. Scene segmentation with CRFs learned from partially labeled images. In *NIPS 2007-Advances in Neural Information Processing Systems*, volume 20, pages 1553–1560. MIT Press, 2008.
- [43] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition, 2009.
- [44] A. Bernal, K. Crammer, A. Hatzigeorgiou, and F. Pereira. Global discriminative learning for higher-accuracy computational gene prediction. *PLoS Comput Biol*, 3(3):e54, 2007.
- [45] B. Taskar, D. Klein, M. Collins, D. Koller, and C. D. Manning. Max-margin parsing. In *EMNLP*, volume 1, page 3. Citeseer, 2004.

- [46] L. Albert, F. Rottensteiner, and C. Heipke. Land use classification using Conditional Random Fields for the verification of geospatial databases. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(4):1, 2014.
- [47] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr. Conditional Random Fields as recurrent Neural Networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.
- [48] A. Schmidt, J. Niemeyer, F. Rottensteiner, and U. Soergel. Contextual classification of full waveform lidar data in the wadden sea. *IEEE Geoscience and Remote Sensing Letters*, 11(9):1614–1618, Sept 2014.
- [49] J. Niemeyer, F. Rottensteiner, and U. Soergel. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing*, 87:152–165, 2014.
- [50] X. He, R. S. Zemel, and M. A. Carreira-Perpinan. Multiscale Conditional Random Fields for image labeling. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–695–II–702 Vol.2, June 2004.
- [51] Sanjiv Kumar and Martial Hebert. Discriminative Random Fields: A discriminative framework for contextual interaction in classification. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1150–1157. IEEE, 2003.
- [52] A. A. Bona. *Algoritmo de otimização combinatorial*. Florianópolis, SC, 2005.
- [53] R. Storn and K. Price. Differential Evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.
- [54] J. E. C. Arroyo. Heurísticas e metaheurísticas para otimização combinatória multi-objetivo. 2002.
- [55] M. A. Luersen and R. Le Riche. Globalized nelder–mead method for engineering optimization. *Computers & structures*, 82(23):2251–2260, 2004.
- [56] R. M. KOIDE. Algoritmo de colônia de formigas aplicado à otimização de materiais compostos laminados. 2010.

- [57] J. H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press, 1975.
- [58] H. Ishibuchi, T. Murata, and I. B. Türksen. Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets and Systems*, 89(2):135 – 150, 1997.
- [59] B. Liu, R. T. Haftka, M. A. Akgün, and A. Todoroki. Permutation genetic algorithm for stacking sequence design of composite laminates. *Computer methods in applied mechanics and engineering*, 186(2):357–372, 2000.
- [60] J. Kennedy and R. Eberhart. Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks*, volume 4, pages 1942 –1948 vol.4, nov/dec 1995.
- [61] A. Lazinica. *Particle Swarm Optimization*. InTech Kirchengasse, 2009.
- [62] M. Omran, A. Salman, and A. P. Engelbrecht. Image classification using Particle Swarm Optimization. In *Proceedings of the 4th Asia-Pacific conference on simulated evolution and learning*, volume 1, pages 18–22. Singapore, 2002.
- [63] R. C. Eberhart and Y. Shi. Particle Swarm Optimization: developments, applications and resources. In *evolutionary computation, 2001. Proceedings of the 2001 Congress on*, volume 1, pages 81–86. IEEE, 2001.
- [64] F. Melgani and Y. Bazi. Classification of electrocardiogram signals with Support Vector Machines and Particle Swarm Optimization. *IEEE Transactions on Information Technology in Biomedicine*, 12(5):667–677, Sept 2008.
- [65] D. W. van der Merwe and A. P. Engelbrecht. Data clustering using Particle Swarm Optimization. In *Evolutionary Computation, 2003. CEC '03. The 2003 Congress on*, volume 1, pages 215–220 Vol.1, Dec 2003.
- [66] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
- [67] R.W. Eglese. Simulated Annealing: A tool for operational research. *European Journal of Operational Research*, 46(3):271 – 281, 1990.
- [68] S. Z. Selim and K. Alsultan. A Simulated Annealing algorithm for the clustering problem. *Pattern Recognition*, 24(10):1003 – 1008, 1991.

- [69] M. Akbulut and F. O. Sonmez. Optimum design of composite laminates for minimum thickness. *Computers & Structures*, 86(21):1974–1982, 2008.
- [70] A. X. Falcão, J. Stolfi, and R. A. Lotufo. The Image Foresting Transform: Theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29, 2004.
- [71] C. S. Won and R. M. Gray. *Stochastic Image Processing*. Kluwer Academic/Plenum Publishers, New York, 2004.
- [72] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37(6):1554–563, 1966.
- [73] L. R. Rabiner. A tutorial in hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [74] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A: Hadrons and Nuclei*, 31(1):253–258, 1925.
- [75] J. H.ammersley and P. Clifford. *Random Fields on Finite Graphs and Lattices*. Não Publicado, 1971.
- [76] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.
- [77] A. L. M. Levada, , N. D. A. Mascarenhas, and A. Tannús. A novel MAP-MRF approach for multispectral image contextual classification using combination of sub-optimal iterative algorithms. *Pattern Recognition Letters*, 31(13):1795–1808, 2010.
- [78] Z. Wu and Q. Ouyang. SVM- and MRF-based method for contextual classification of polarimetric SAR images. In *Proceedings of the 2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, pages 818–821, 2011.
- [79] H. Bischof, W. Schneider, and A.J. Pinz. Multispectral classification of landsat-images using Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–490, 1992.
- [80] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, London, UK, UK, 1995.

- [81] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov Random Fields. In *Computer Vision–ECCV 2006*, pages 16–29. Springer, 2006.
- [82] A. L. D. Martins, N. D. A. Mascarenhas, and C. A. T. Suazo. Spatio-temporal resolution enhancement of vocal tract MRI sequences based on image registration. *Integrated Computer-Aided Engineering*, 18(2):143–155, 2011.
- [83] A. L. M. Levada and N. D. A. Mascarenhas. Filtragem adaptativa de ruído Gaussiano em imagens através da minimização da Informação de Fisher Observada.
- [84] M. Reginatto. Derivation of the equations of nonrelativistic quantum mechanics using the principle of minimum Fisher Information. *Physical Review A*, 58(3):1775, 1998.
- [85] A. K. Hiraga and A.L.M. Levada. Informação de Fisher no modelo de Potts definido em grafos para classificação supervisionada.
- [86] N. D. A. Mascarenhas A. L. M. Levada and A. Tannus. Statistical inference on Markov Random Fields: Parameter estimation, asymptotic evaluation and contextual classification of nmr multispectral images. In *Pattern Recognition*. InTech, 2009.
- [87] R. Nakamura, D. Osaku, A. Levada, J.P. Papa, F. Cappabianco, and A. Falcão. OPF-MRF: Optimum-Path Forest and Markov Random Fields for contextual-based image classification. In *Proceedings of the 15th International Conference on Computer Analysis of Images and Patterns*, 2013.
- [88] D. Osaku, R.Y.M. Nakamura, L.A.M. Pereira, R.J. Pisani, A.L.M. Levada, F.A.M. Cappabianco, A.X. Falcão, and J.P. Papa. Improving land cover classification through contextual-based Optimum-Path Forest. *Information Sciences*, 324:60 – 87, 2015.
- [89] D. Osaku, A.L.M. Levada, and J.P. Papa. On the influence of Markovian models for contextual-based Optimum-Path Forest classification. In E. Bayro-Corrochano and E. Hancock, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8827 of *Lecture Notes in Computer Science*, pages 462–469. Springer International Publishing, 2014.
- [90] A. Rosenfeld and A.C. Kak. *Digital picture processing*, volume 1. Elsevier, 2014.

- [91] K.I. Itten and F. Fasler. *Thematic adaptive spatial filtering of Landsat landuse classification results. A*: [Arbeiten aus dem Geographischen Institut. Geogr. Inst. Univ. Zürich, 1980.
- [92] J.M. Brayer, P.H. Swain, and K.S. Fu. Modeling of earth resources satellite data. In KingSun Fu, editor, *Syntactic Pattern Recognition, Applications*, volume 14 of *Communication and Cybernetics*, pages 215–242. Springer Berlin Heidelberg, 1977.
- [93] K.S. Fu. *Syntactic pattern recognition and applications*. Prentice-Hall advanced reference series: Computer science. Prentice-Hall, 1982.
- [94] N. Ahuja, A. Rosenfeld, and R.M. Haralick. Neighbor gray levels as features in pixel classification. *Pattern Recognition*, 12(4):251 – 260, 1980.
- [95] J. Desachy. Texture features in remote sensing imagery. In J.C. Simon and R.M. Haralick, editors, *Digital Image Processing*, volume 77 of *NATO Advanced Study Institutes Series*, pages 203–210. Springer Netherlands, 1981.
- [96] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, Nov 1973.
- [97] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6(4):269–285, April 1976.
- [98] D. Osaku, R. Y. M. Nakamura, J. P. Papa, A. L. M. Levada, F. A. M. Cappabianco, and A. X. Falcão. Optimizing contextual-based optimum-forest classification through swarm intelligence. In *Advanced Concepts for Intelligent Vision Systems*, pages 203–214, 2013.
- [99] A. Levada, N. D. A. Mascarenhas, and A. Tannús. Pseudolikelihood equations for Potts MRF model parameter estimation on higher order neighborhood systems. *IEEE Geoscience and Remote Sensing Letters*, 5(3):522–526, 2008.
- [100] A. Levada, N. D. A. Mascarenhas, and A. Tannús. 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*, 23(2):120–140, 2009.
- [101] F.Y. Wu. The Potts model. *Review of Modern Physics*, 54:235–268, 1982.

- [102] Z. W. Geem. *Music-Inspired Harmony Search Algorithm: Theory and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2009.
- [103] Z. W. Geem, J. H. Kim, and G. V. Loganathan. A new heuristic optimization algorithm: Harmony Search. *SIMULATION*, 76(2):60–68, 2001.
- [104] M. Mahdavi, M. Fesanghary, and E. Damangir. An improved Harmony Search algorithm for solving optimization problems. *Applied Mathematics and Computation*, 188(2):1567 – 1579, 2007.
- [105] G.H.O. Mahamed and M. Mehrdad. Global-best Harmony Search. *Applied Mathematics and Computation*, 198(2):643 – 656, 2008.
- [106] J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [107] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [108] J. Settin. Um algoritmo de aprendizado por reforço para Redes Neurais utilizando metaotimização estatística. 2010.
- [109] A. R. Gonçalves. *Otimização em ambientes dinâmicos com variáveis contínuas empregando algoritmos de estimação de distribuição*. PhD thesis, Universidade Estadual de campinas, 2011.
- [110] D. Osaku, D. R. Pereira, A. L. M. Levada, and J. P. Papa. Fine-tuning Contextual-Based Optimum-Path Forest for land-cover classification. *IEEE Geoscience and Remote Sensing Letters*, 13(5):735–739, May 2016.
- [111] Elaine Marra dos Santos. Análise de imagens do satélite alos palsar (banda l, modo polarimétrico) para discriminação de coberturas de terras no distrito federal. 2009.
- [112] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [113] F.A.M. Cappabianco, A.X. Falcão, C.L. Yasuda, and J.K. Udupa. Brain tissue MR-image segmentation via Optimum-Path Forest clustering. *Computer Vision and Image Understanding*, 116(10):1047–1059, October 2012.
- [114] D. Osaku, A. L. M. Levada, and J. P. Papa. *On the Influence of Markovian Models for Contextual-Based Optimum-Path Forest Classification*, pages 462–469. Springer International Publishing, Cham, 2014.

- [115] J. P. Papa, C. T. N. S., and A. X. Falcão. *LibOPF: A library for the design of Optimum-Path Forest classifiers*, 2009. Software version 2.0 available at <http://www.ic.unicamp.br/~afalcao/LibOPF>.
- [116] C. C. Chang and C. J. Lin. *LIBSVM: A Library for Support Vector Machines*, 2001. Software available at url <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [117] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.