

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ELIEZER SOARES FLORES

**Segmentação de Lesões Melanocíticas
Usando uma Abordagem Baseada no
Aprendizado de Dicionários**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Jacob Scharcanski

Porto Alegre
2015

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Flores, Eliezer Soares

Segmentação de Lesões Melanocíticas Usando uma Abordagem Baseada no Aprendizado de Dicionários / Eliezer Soares Flores. – Porto Alegre: PPGC da UFRGS, 2015.

85 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2015. Orientador: Jacob Scharcanski.

1. Segmentação. 2. Lesões melanocíticas. 3. Imagens macroscópicas. 4. Aprendizado de dicionários. I. Scharcanski, Jacob. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“If I have seen farther than others,
it is because I stood on the shoulders of giants.”*

— SIR ISAAC NEWTON

AGRADECIMENTOS

A Deus por tudo que eu vivi e que me possibilitou chegar até aqui, por ter me feito quem eu sou. Agradeço pelos bons momentos que me fizeram celebrar a vida e pelos momentos difíceis que me fizeram crescer. Agradeço por nunca ter me abandonado e por nunca ter deixado eu abandonar a minha fé. Por ter me dado saúde e força para encarar as dificuldades da vida e por ter colocado ao meu lado pessoas muito especiais sem as quais certamente eu não estaria aqui.

Ao meu orientador Prof. Dr. Jacob Scharcanski pela oportunidade única de desenvolver este trabalho, encorajando-me na busca de um futuro melhor. Por ter confiado no meu potencial e ter me colocado na direção correta para alcançar os meus objetivos.

Aos meus pais Mariza e Ildefonso pelo incentivo, amor e incontáveis conselhos que recebi. Agradeço por serem um exemplo de dignidade, humildade e perseverança e por serem hoje e sempre a minha principal fonte de inspiração.

A Lorena, amor da minha vida, por não permitir que a distância diminua em nada o amor que sentimos um pelo outro e por mesmo estando longe, estar sempre presente em todos os momentos da minha vida, ao infinito de além!

Ao meu irmão Filipe, meu melhor amigo, pelo companheirismo de hoje e sempre. Agradeço por ter aceitado sair de Bagé, vindo para Porto Alegre morar comigo, buscando também uma profissão digna e prazerosa.

Ao meu irmão Nei por ter me instigado a trabalhar com computadores desde cedo e pelo exemplo de coragem e superação.

Aos meus dindos Ana, Itamar, Mimi e Valdir (*in memoriam*), tios, primos e demais familiares por sempre terem acreditado e investido em mim.

Aos meus amigos, colegas e ex-colegas por dividirem seu tempo, experiências e momentos de alegria nessa jornada.

Ao meu orientador da graduação Prof. Fábio Ronei Rodrigues Padilha por toda a dedicação e paciência, por ter acreditado em mim até quando as coisas pareciam impossíveis e por ter me feito gostar de trabalhar com processamento de imagens, visão computacional e reconhecimento de padrões.

A todos os meus professores e ex-professores não somente por terem aberto para mim as portas do conhecimento, mas também por terem mostrado ao longo destes anos o

quanto o caráter e a afetividade são importantes no processo de formação profissional.

A todas as pessoas envolvidas com a Universidade Federal do Rio Grande do Sul (UFRGS), a Universidade Federal do Pampa (UNIPAMPA), a Companhia de Processamento de Dados do Estado do Rio Grande do Sul (PROCERGS), a Fundação Bradesco, a Igreja Anglicana e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelas inúmeras oportunidades a mim concedidas.

A todas as pessoas que cruzaram a minha vida e contribuíram direta ou indiretamente na minha jornada, muito obrigado!

RESUMO

Segmentação é uma etapa essencial para sistemas de pré-triagem de lesões melanocíticas. Neste trabalho, um novo método para segmentar lesões melanocíticas em imagens de câmera padrão (i.e., imagens macroscópicas) é apresentado. Inicialmente, para reduzir artefatos indesejáveis, os efeitos de sombra são atenuados na imagem macroscópica e uma pré-segmentação é obtida usando um esquema que combina a transformada wavelet com a transformada watershed. Em seguida, uma imagem de variação textural projetada para melhorar a discriminabilidade da lesão em relação ao fundo é obtida e a região pré-segmentada é usada para o aprendizado de um dicionário inicial e de uma representação inicial via um método de fatoração de matrizes não-negativas. Uma versão não-supervisionada e não-paramétrica do método de aprendizado de dicionário baseado em teoria da informação é proposta para otimizar esta representação, selecionando o subconjunto de átomos que maximiza a compactividade e a representatividade do dicionário aprendido. Por fim, a imagem da lesão de pele é representada usando o dicionário aprendido e segmentada com o método de corte normalizado em grafos. Nossos resultados experimentais baseados em uma base de imagens bastante utilizada sugerem que o método proposto tende a fornecer melhores resultados do que os métodos estado-da-arte analisados (em termos do erro XOR).

Palavras-chave: Segmentação. Lesões melanocíticas. Imagens macroscópicas. Aprendizado de dicionários.

Segmentation of Melanocytic Lesions Using a Dictionary Learning Based Approach

ABSTRACT

Segmentation is an essential step for the automated pre-screening of melanocytic lesions. In this work, a new method for segmenting melanocytic lesions in standard camera images (i.e., macroscopic images) is presented. Initially, to reduce unwanted artifacts, shading effects are attenuated in the macroscopic image and a pre-segmentation is obtained using a scheme that combines the wavelet transform and the watershed transform. Afterwards, a textural variation image designed to enhance the skin lesion against the background is obtained, and the pre-segmented skin lesion region is used to learn an initial dictionary and an initial representation via a nonnegative matrix factorization method. An unsupervised and non-parametric version of the information-theoretic dictionary learning method is proposed to optimize this representation by selecting the subset of atoms that maximizes the learned dictionary compactness and representation. Finally, the skin lesion image is represented using the learned dictionary and segmented with the normalized graph cuts method. Our experimental results based on a widely used image dataset suggest that the proposed method tends to provide more accurate skin lesion segmentations than comparable state-of-the-art methods (in terms of the XOR error).

Keywords: Segmentation. Melanocytic lesions. Macroscopic images. Dictionary learning.

LISTA DE FIGURAS

Figura 2.1 – Exemplo de imagem de lesão melanocítica.....	22
Figura 2.2 – Exemplos de histograma.....	23
Figura 2.3 – Representação de uma lesão melanocítica em RGB	24
Figura 2.4 – Ilustração do espaço de cores HSV	26
Figura 2.5 – Exemplo de filtragem pelo filtro da média.....	29
Figura 2.6 – Comparação entre o filtro da média e o filtro gaussiano.	30
Figura 2.7 – Representações em diferentes dicionários.....	43
Figura 2.8 – Comparação de resultados obtidos usando contornos ativos.	48
Figura 3.1 – Exemplo de atenuação de sombras.	61
Figura 3.2 – Exemplo de representação em três canais.....	64
Figura 4.1 – Visão geral do método proposto.	67
Figura 5.1 – Comparação visual entre algoritmos de segmentação.	79

LISTA DE TABELAS

Tabela 5.1 – Acurácia obtida variando os parâmetros w e \mathcal{K}	73
Tabela 5.2 – Especificidade obtida variando os parâmetros w e \mathcal{K}	75
Tabela 5.3 – Comparação entre algoritmos em termos de especificidade.	76
Tabela 5.4 – Erro XOR obtido variando os parâmetros w e \mathcal{K}	77
Tabela 5.5 – Comparação entre algoritmos em termos do erro XOR.....	78

LISTA DE ABREVIATURAS E SIGLAS

<i>RGB</i>	<i>Red, Green and Blue</i>
<i>HSV</i>	<i>Hue, Saturation and Value</i>
PCA	<i>Principal Component Analysis</i>
ICA	<i>Independent Component Analysis</i>
MOD	<i>Method of Optimal Directions</i>
SVD	<i>Singular Value Decomposition</i>
NMF	<i>Nonnegative Matrix Factorization</i>
ITDL	<i>Information-Theoretic Dictionary Learning</i>
UNPITDL	<i>Unsupervised and Non-Parametric Information-Theoretic Dictionary Learning</i>
GT	<i>Ground Truth</i>

LISTA DE SÍMBOLOS

\mathcal{J}	Imagem analógica
\mathbf{I}	Imagem digital
(i, j)	Par de índices em \mathbf{I}
$I_{(i,j)}$	Elemento de \mathbf{I}
m	Número de linhas em \mathbf{I}
n	Número de colunas em \mathbf{I}
ℓ	Número de bits por pixel em \mathbf{I}
\mathbf{R}	Primeiro canal do espaço de cores \mathbf{RGB}
\mathbf{G}	Segundo canal do espaço de cores \mathbf{RGB}
\mathbf{B}	Terceiro canal do espaço de cores \mathbf{RGB}
\mathbf{H}	Primeiro canal do espaço de cores \mathbf{HSV}
\mathbf{S}	Segundo canal do espaço de cores \mathbf{HSV}
\mathbf{V}	Terceiro canal do espaço de cores \mathbf{HSV}
Φ	Amostras de pixels de não-lesão
\mathbf{Z}	Modelo de iluminação de \mathbf{V}
\mathbf{L}	Canal de luminância
$\mathbf{I}^{(1)}$	Complemento do canal \mathbf{R}
$\mathbf{I}^{(2)}$	Canal de variação textural
$\mathbf{I}^{(3)}$	Canal de variação local da cor
w	Número de linhas (e colunas) em um bloco quadrado
\mathbf{Y}	Matriz dos blocos (representados como vetores)
M	Número de linhas em \mathbf{Y}
N	Número de colunas em \mathbf{Y}

$\{\mathbf{y}_i\}_{i=1}^N$ Colunas (blocos) de \mathbf{Y}

\mathbf{D} Dicionário

\mathcal{K} Número de colunas em \mathbf{D}

$\{\mathbf{d}_j\}_{j=1}^{\mathcal{K}}$ Colunas (átomos) de \mathbf{D}

\mathbf{X} Representação de \mathbf{Y} em \mathbf{D}

Γ Representação de \mathbf{X} como um grafo ponderado e não-direcionado

Λ Conjunto de vértices de Γ

Ξ Conjunto de arestas de Γ

Ω Matriz de similaridade de Γ

\mathbf{M} Imagem segmentada (binária)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivo	16
1.2	Contribuições deste Trabalho	16
1.3	Organização do Trabalho	20
2	REVISÃO BIBLIOGRÁFICA	21
2.1	Fundamentos de Processamento e Análise de Imagens	21
2.1.1	Definição de Imagem	21
2.1.2	Imagens Digitais	21
2.1.3	Histograma da Imagem	22
2.1.4	Representações de Cores	24
2.2	Conceitos de Pré-Processamento de Imagens Dermatológicas	27
2.2.1	Convolução	28
2.2.2	Filtros Passa-Baixas	28
2.2.3	Filtros Passa-Altas	30
2.3	Conceitos de Extração de Características de Imagens Dermatológicas	31
2.3.1	PCA	33
2.3.2	ICA	34
2.3.3	MOD e K-SVD	37
2.3.4	NMF	39
2.3.5	ITDL	40
2.4	Conceitos de Segmentação de Imagens Dermatológicas	42
2.4.1	Limiar de Otsu	44
2.4.2	Contornos Ativos e Fluxo do Vetor Gradiente	45
2.4.3	Segmentação Baseada nas Transformadas Wavelet e Watershed	47
2.4.4	<i>K</i> -médias	50
2.4.5	Corte Normalizado em Grafos	52

2.4.6	Morfologia Matemática	54
3	ESTADO-DA-ARTE	56
3.1	Limiar de Otsu na Imagem em Tons de Cinza	56
3.2	Segmentação Baseada no Fluxo do Vetor Gradiente.....	56
3.3	Atenuação de Sombras	57
3.4	Representação em Três Canais para Imagens de Lesões Melanocíticas	61
3.5	Segmentação Baseada em ICA.....	63
3.6	Segmentação Baseada em NMF	65
3.7	Outros Métodos.....	65
4	MÉTODO PROPOSTO	66
4.1	UNPITDL.....	68
4.2	Pós-Processamento	71
5	RESULTADOS EXPERIMENTAIS.....	72
6	CONCLUSÃO.....	80
6.1	Trabalhos Futuros	80
	Referências.....	81

1 INTRODUÇÃO

A lesão melanocítica maligna é um câncer de pele que ataca os melanócitos¹ e tem sido considerado o tipo mais perigoso de câncer de pele (MELANOMA RESEARCH FOUNDATION, 2015). Embora seja menos comum do que outros tipos de câncer de pele (i.e., representando apenas 4% do número total de casos), estas lesões causam aproximadamente 77% das mortes por câncer de pele (UNIVERSITY OF CALIFORNIA, SAN FRANCISCO SCHOOL OF MEDICINE, 2015). Os estágios de uma lesão melanocítica maligna são os seguintes: IA, IB, IIA, IIB, IIC, IIIA, IIIB, IIIC e IV. O estágio IA é o primeiro estágio da doença e se uma lesão melanocítica é detectada enquanto classificada no estágio IA, a taxa de sobrevivência esperada (para os próximos 5 anos) do paciente é de aproximadamente 97%. Porém, esta taxa diminui para 15 – 20% se a detecção ocorre enquanto a doença está no estágio IV (AMERICAN CANCER SOCIETY, 2015). Estes fatos destacam a importância de um diagnóstico rápido e eficiente para reduzir o número de vítimas de lesões melanocíticas malignas.

O diagnóstico de lesões melanocíticas malignas pode ser realizado por um dermatologista via um dermatoscópio digital (i.e., usando imagens dermatoscópicas). Porém, o tempo necessário para agendar/consultar com um dermatologista (especialmente em países em desenvolvimento) pode resultar em um atraso não desejável no diagnóstico. Além disso, uma pesquisa realizada com os membros da AMERICAN ACADEMY OF DERMATOLOGY (2015) mostrou que somente 48% dos dermatologistas usam dermatoscópios para diagnosticar lesões de pele (ENGASSER; WARSHAW, 2010).

Por outro lado, vários sistemas de visão computacional têm sido propostos para o diagnóstico de lesões melanocíticas malignas usando imagens macroscópicas² (WONG; SCHARCANSKI; FIEGUTH, 2011). Uma síntese do estado-da-arte em termos de técnicas de visão computacional para o diagnóstico de lesões melanocíticas malignas pode ser encontrado em (SCHARCANSKI; CELEBI, 2013). Dada uma fotografia de uma lesão melanocítica como entrada, o objetivo destes sistemas é detectar a lesão e identificar se ela é maligna ou benigna. Usualmente, estes sistemas executam os seguintes passos (CAVALCANTI; SCHARCANSKI; BARANOSKI, 2013): (a) pré-processamento, onde a imagem de entrada é processada para facilitar a segmentação da lesão de pele; (b) segmentação,

¹As células que produzem a melanina, responsável pela coloração da pele, dos cabelos e dos olhos.

²Fotografias coloridas da lesão de pele, obtidas sem iluminação especial ou equipamento dedicado.

onde o objetivo é delimitar a região da lesão de pele na imagem de entrada; (c) extração de características, onde a lesão é representada pelas suas características; (d) classificação, onde a lesão melanocítica é classificada como benigna ou maligna com base nas características extraídas anteriormente. Este trabalho tem como foco o passo (b), a segmentação de lesões melanocíticas.

1.1 Objetivo

O objetivo deste trabalho é apresentar um novo método para segmentar lesões melanocíticas em imagens macroscópicas usando uma abordagem baseada no aprendizado de dicionários.

1.2 Contribuições deste Trabalho

As principais contribuições deste trabalho são:

- um novo método para segmentar lesões melanocíticas em imagens macroscópicas;
- uma versão não-supervisionada e não-paramétrica do método de aprendizado de dicionário baseado em teoria da informação (UNPITDL).

Durante o desenvolvimento deste trabalho foram elaborados três artigos:

- um artigo de conferência aceito e publicado (Qualis B1);
- dois artigos submetidos para revistas internacionais (Qualis A1).

Estes artigos são apresentados e resumidos a seguir:

FLORES E. S.; SCHARCANSKI J. Segmentation of pigmented melanocytic skin lesions based on learned dictionaries and normalized graph cuts. In: CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES (SIBGRAPI).

RESUMO: A pré-triagem de lesões melanocíticas depende de uma segmentação adequada das regiões da imagem afetadas pela lesão de pele. Este artigo propõe um novo algoritmo para segmentar lesões melanocíticas em imagens macroscópicas. É assumido que somente uma lesão de pele está na imagem de entrada, localizada perto do centro da imagem. Assim, primeiramente, os efeitos de sombra são atenuados. Então, a imagem é convertida para uma nova representação em três canais que melhora a discriminabilidade entre regiões saudáveis e regiões não-saudáveis. Em seguida, um dicionário compacto e representativo é gerado para representar os blocos da imagem. O aprendizado do dicionário é realizado de forma não-supervisionada por uma versão modificada do método de aprendizado de dicionário baseado em teoria da informação, o qual foi originalmente proposto para o aprendizado supervisionado de dicionários. Por fim, o método de corte normalizado em grafos é usado para particionar em dois grupos o conjunto de blocos projetados no dicionário, resultando em uma máscara binária que rotula os pixels como saudáveis ou não-saudáveis. Nossos resultados experimentais obtidos em uma base de dados bastante utilizada são animadores e sugerem que o método proposto para segmentar lesões melanocíticas fornece, na média, um menor erro XOR do que outros métodos comparáveis propostos na literatura.

FLORES E. S.; SCHARCANSKI J. Segmentation of melanocytic skin lesions using feature learning and dictionaries. Submetido para: **Expert Systems with Applications**.

RESUMO: Sistemas de pré-triagem para o diagnóstico de lesões melano-cíticas dependem de uma segmentação adequada da região da imagem afetada pela lesão. Este artigo propõe um esquema de aprendizagem de características que encontra características relevantes para a segmentação da lesão de pele. O esquema proposto para a aprendizagem das características inicia com uma etapa de pré-processamento, onde os efeitos de sombra da imagem são atenuados. Então, uma imagem de variação textural é obtida para melhorar a discriminação entre os blocos da imagem. Em seguida, um método de fatoração de matrizes não-negativas é usado para gerar um dicionário inicial e uma representação inicial para os blocos da imagem de variação textural. Uma extensão não-supervisionada do método de aprendizado de dicionário baseado em teoria da informação é proposta para selecionar o subconjunto de átomos que maximiza a compactividade e a representatividade do dicionário. Por fim, as características aprendidas e o método de corte normalizado em grafos são empregados para segmentar a imagem de entrada. Os resultados experimentais sugerem que a abordagem proposta pode potencialmente fornecer resultados de segmentação mais precisos do que os demais métodos analisados.

FLORES E. S.; GOMES V.; SCHARCANSKI J. Segmentation of melanocytic skin lesions using unsupervised dictionary learning and robust watersheds. Submetido para: **Computer Vision and Image Understanding**.

RESUMO: Segmentação é uma etapa essencial para sistemas de pré-triagem de lesões melanocíticas. Neste trabalho, um novo método para segmentar lesões melanocíticas em imagens de câmera padrão (i.e., imagens macroscópicas) é apresentado. Inicialmente, para reduzir artefatos indesejáveis, os efeitos de sombra são atenuados na imagem macroscópica e uma pré-segmentação é obtida usando um esquema que combina a transformada wavelet com a transformada watershed. Em seguida, uma imagem de variação textural projetada para melhorar a discriminabilidade da lesão em relação ao fundo é obtida e a região pré-segmentada é usada para o aprendizado de um dicionário inicial e de uma representação inicial via um método de fatoração de matrizes não-negativas. Uma versão não-supervisionada e não-paramétrica do método de aprendizado de dicionário baseado em teoria da informação é proposta para otimizar esta representação, selecionando o subconjunto de átomos que maximiza a compactividade e a representatividade do dicionário aprendido. Por fim, a imagem da lesão de pele é representada usando o dicionário aprendido e segmentada com o método de corte normalizado em grafos. Nossos resultados experimentais baseados em uma base de imagens bastante utilizada sugerem que o método proposto tende a fornecer melhores resultados do que os métodos estado-da-arte analisados (em termos do erro XOR).

1.3 Organização do Trabalho

O restante deste trabalho está organizado da seguinte forma: O Capítulo 2 introduz alguns conceitos fundamentais de processamento de imagens e apresenta algoritmos fundamentais para o entendimento do método proposto e dos demais métodos comparáveis e representativos do estado-da-arte, os quais são apresentados no Capítulo 3. O Capítulo 4 descreve o método proposto para segmentar lesões melanocíticas em imagens macroscópicas. No Capítulo 5, são apresentados e discutidos os resultados obtidos pelo método proposto. Por fim, no Capítulo 6, são feitas as considerações finais a respeito deste trabalho.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo, primeiramente, serão apresentadas algumas definições fundamentais de processamento e análise de imagens. Em seguida, serão apresentados conceitos e algoritmos de pré-processamento, extração de características e segmentação de imagens dermatológicas.

2.1 Fundamentos de Processamento e Análise de Imagens

2.1.1 Definição de Imagem

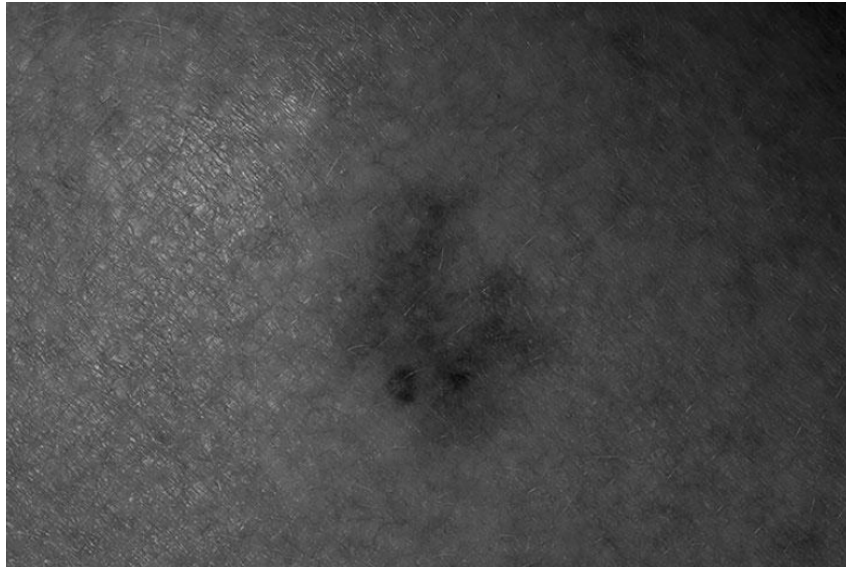
Uma imagem pode ser definida como uma função bidimensional \mathcal{J} das variáveis contínuas x e y . Assim, cada ponto (x, y) no domínio de \mathcal{J} é associado a um escalar $\mathcal{J}(x, y)$ (este escalar é usualmente denominado intensidade). Uma imagem conforme definida anteriormente é de natureza analógica, ou seja, é contínua tanto nos pontos (x, y) quanto nas intensidades $\mathcal{J}(x, y)$. Porém, visando o seu processamento computacional, esta deve ser digitalizada e representada por alguma estrutura de dado adequada (GONZALEZ; WOODS, 2008).

2.1.2 Imagens Digitais

Uma imagem ao ser digitalizada torna-se discreta nos pontos (onde o processo de discretização recebe o nome de amostragem) e nas intensidades (onde o processo de discretização recebe o nome de quantização). Basicamente, a digitalização (i.e., amostragem e quantização) converte uma imagem \mathcal{J} em uma matriz \mathbf{I} (i.e., uma imagem digital em tons de cinza) com m linhas e n colunas (SONKA; HLAVAC; BOYLE, 2014). O processo de digitalização envolve decisões a respeito dos valores de m , n e ℓ , variando muito de dispositivo para dispositivo (GONZALEZ; WOODS, 2008). Maiores valores de m e n implicam em imagens com maiores resoluções espaciais (i.e., mais pontos). Por outro lado, a quantização faz com que cada ponto na imagem assuma um valor inteiro contido em um intervalo de 0 a $2^\ell - 1$, onde $\ell \in \mathbb{N}^*$. O caso particular $\ell = 1$ define uma imagem binária. Daqui em diante, uma imagem digital em tons de cinza ou colorida (ver Subseção 2.1.4) será denotada por uma letra maiúscula em negrito e itálico (e.g., \mathbf{I}), um par de índices

(linha e coluna) da imagem será denotado (i, j) e um elemento de \mathbf{I} será denotado $I_{(i,j)}$. Além disso, um par ordenado $((i, j), I_{(i,j)})$ é a definição usual de pixel, sendo pixel uma aglutinação do termo em inglês *picture element* (elemento da imagem). A Figura 2.1 mostra um exemplo de imagem de lesão melanocítica apresentada em tons de cinza.

Figura 2.1: Exemplo de imagem de lesão melanocítica apresentada em tons de cinza.



2.1.3 Histograma da Imagem

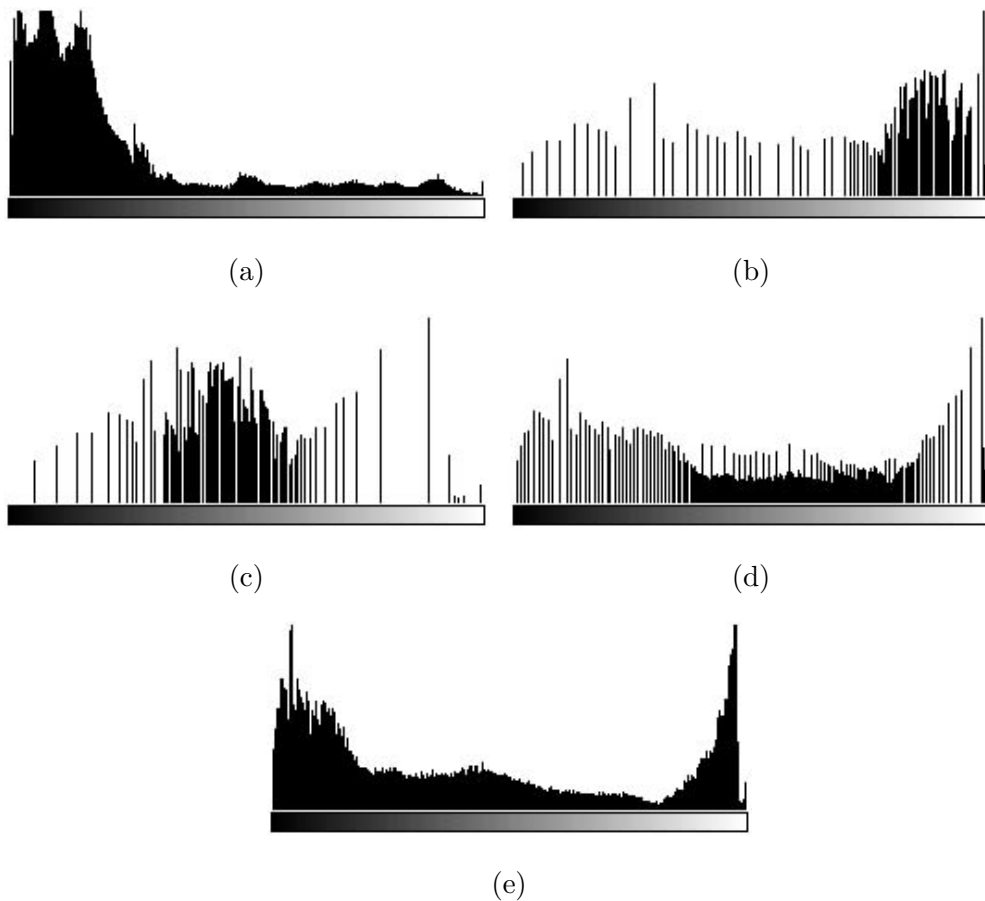
O histograma de uma imagem \mathbf{I} com m linhas e n colunas pode ser definido como um conjunto $\{p(I_{(i,j)} = 0), p(I_{(i,j)} = 1), \dots, p(I_{(i,j)} = N_\ell - 1)\}$, onde $N_\ell = 2^\ell$ denota o número de intensidades presentes na imagem e $p(I_{(i,j)} = \ell)$ denota a probabilidade de ocorrência da intensidade ℓ (MARQUES FILHO; VIEIRA NETO, 1999). A probabilidade $p(I_{(i,j)} = \ell)$ pode ser calculada por:

$$p(I_{(i,j)} = \ell) = \frac{\sum_{I_{(i,j)}=\ell} 1}{mn}. \quad (2.1)$$

Um histograma pode ser visto como uma função discreta e representado por um gráfico, de forma que o eixo das abcissas corresponde às intensidades presentes na imagem e o eixo das ordenadas corresponde às probabilidades de ocorrência destas intensidades (GONZALEZ; WOODS, 2008). A Figura 2.2 apresenta cinco exemplos de tipos de histogramas frequentemente encontrados em imagens dermatológicas. O histograma da

Figura 2.2a apresenta uma grande concentração nas intensidades mais baixas, caracterizando uma imagem predominantemente escura. O histograma da Figura 2.2b apresenta uma grande concentração nas intensidades mais altas, caracterizando uma imagem predominantemente clara. O histograma da Figura 2.2c apresenta uma grande concentração nas intensidades intermediárias, caracterizando uma imagem de brilho médio. Nas Figuras 2.2a até 2.2c os histogramas apresentam grandes concentrações em estreitas faixas de intensidades, caracterizando imagens de baixo contraste. Por outro lado, o histograma da Figura 2.2d está bem distribuído ao longo de quase todas as intensidades, caracterizando uma imagem de alto contraste. A Figura 2.2e mostra um histograma bimodal, isto é, um histograma que possui grandes concentrações em duas faixas de intensidade.

Figura 2.2: Exemplos de histograma.



Fonte: (MARQUES FILHO; VIEIRA NETO, 1999).

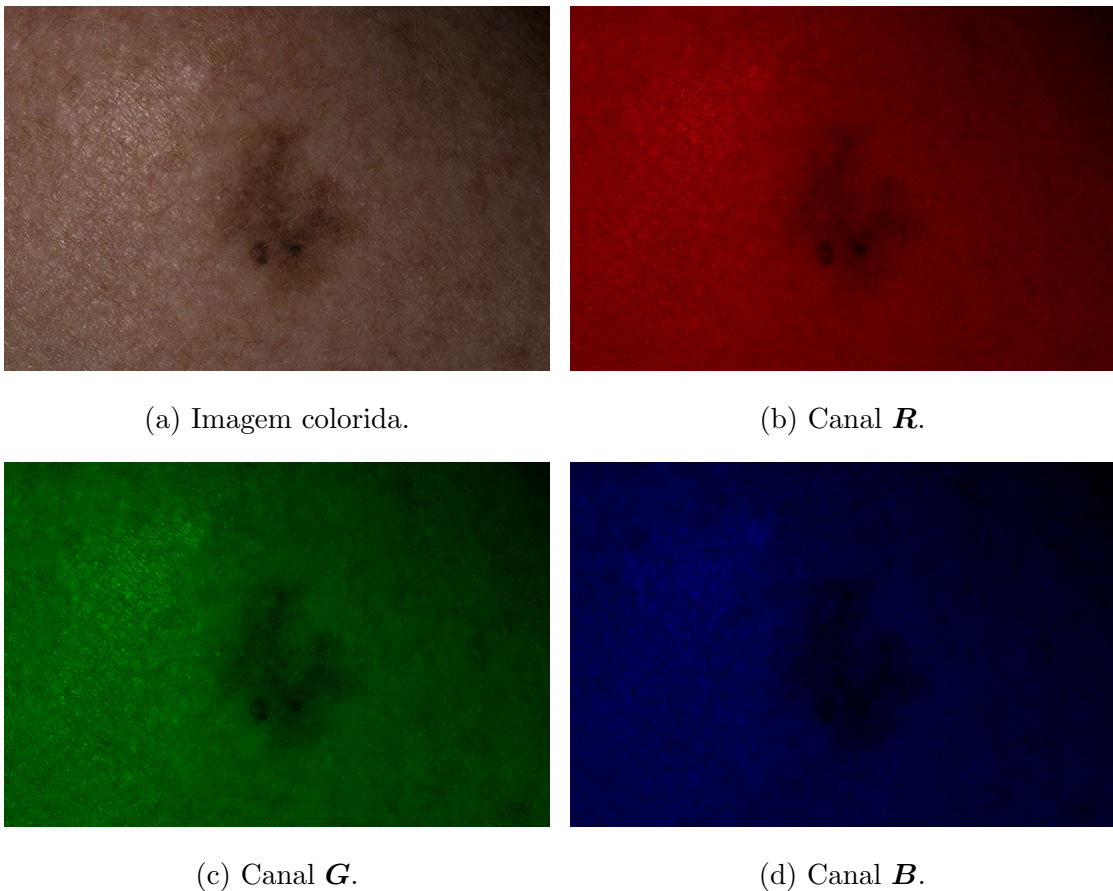
Note que o histograma não contém informação a respeito da posição onde ocorrem as intensidades. Assim, a representação de uma imagem através de um histograma possui duas principais limitações (MOESLUND, 2012):

1. duas imagens diferentes podem possuir o mesmo histograma;
2. uma imagem não pode ser reconstruída a partir do histograma.

2.1.4 Representações de Cores

Diversos modelos, conhecidos como espaços de cores, têm sido propostos para representar cores em imagens. Dentre estes, está o espaço de cores **RGB**. No espaço de cores **RGB**, uma imagem colorida é representada pela combinação de três imagens em tons de cinza, denominadas neste contexto canais da imagem colorida. Estes canais contêm as intensidades³de vermelho (i.e., **R**), verde (i.e., **G**) e azul (i.e., **B**) da imagem (SZELISKI, 2010). A Figura 2.3 mostra uma imagem de uma lesão melanocítica representada no espaço de cores **RGB**.

Figura 2.3: Representação de uma lesão melanocítica em **RGB**.



³Daqui em diante, considere as intensidades assumindo valores entre 0 e 1 (exceto quando especificado o contrário).

De acordo com RUSS (2011), o espaço de cores **RGB** é o modelo mais utilizado para a aquisição de imagens coloridas. Porém, este modelo não é adequado para a segmentação de imagens coloridas. Por outro lado, espaços de cores tais como o **HSV** são baseados na percepção humana e têm se mostrado uma melhor alternativa para a segmentação de imagens coloridas. O espaço de cores **HSV** pode ser visto como uma transformação não-linear do espaço de cores **RGB** e pode ser representado pelos seguintes canais (MOESLUND, 2012):

1. o canal de matiz (i.e., **H**) contém intensidades entre 0° e 360° , estas intensidades fornecem o ângulo (em graus) correspondente a uma determinada cor no círculo das cores (ver Figura 2.4). Este canal pode ser definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ em termos dos canais **R**, **G** e **B** por:

$$H_{(i,j)} = \begin{cases} 60^\circ \times \left(\frac{\Delta BG_{(i,j)}}{\text{MAX}_{(i,j)} - \text{MIN}_{(i,j)}} \bmod 6 \right), & \text{se } \text{MAX}_{(i,j)} = R_{(i,j)}, \\ 60^\circ \times \left(\frac{\Delta RB_{(i,j)}}{\text{MAX}_{(i,j)} - \text{MIN}_{(i,j)}} + 2 \right), & \text{se } \text{MAX}_{(i,j)} = G_{(i,j)}, \\ 60^\circ \times \left(\frac{\Delta GR_{(i,j)}}{\text{MAX}_{(i,j)} - \text{MIN}_{(i,j)}} + 4 \right), & \text{se } \text{MAX}_{(i,j)} = B_{(i,j)}, \end{cases} \quad (2.2)$$

onde $a \bmod b$ denota a operação módulo (resto da divisão de a por b), além disso:

$$\text{MAX}_{(i,j)} = \max\{R_{(i,j)}, G_{(i,j)}, B_{(i,j)}\}, \quad (2.3)$$

$$\text{MIN}_{(i,j)} = \min\{R_{(i,j)}, G_{(i,j)}, B_{(i,j)}\}, \quad (2.4)$$

$$\Delta BG_{(i,j)} = G_{(i,j)} - B_{(i,j)}, \quad (2.5)$$

$$\Delta GR_{(i,j)} = R_{(i,j)} - G_{(i,j)}, \quad (2.6)$$

$$\Delta RB_{(i,j)} = B_{(i,j)} - R_{(i,j)}; \quad (2.7)$$

2. o canal de saturação (i.e., \mathbf{S}) contém intensidades entre 0 e 1, estas intensidades fornecem informações sobre o quão pura (menos acinzentada) é uma determinada cor. Este canal pode ser definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

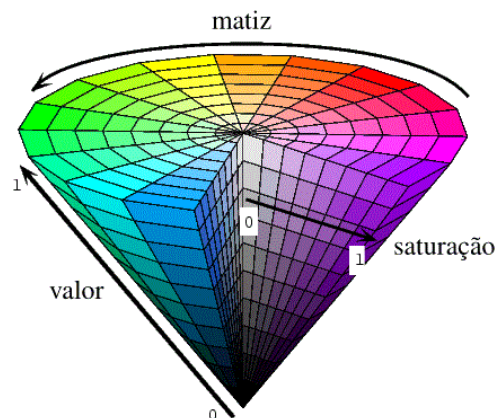
$$S_{(i,j)} = \frac{\text{MAX}_{(i,j)} - \text{MIN}_{(i,j)}}{\text{MAX}_{(i,j)}}; \quad (2.8)$$

3. o canal de valor (i.e., \mathbf{V}) contém intensidades entre 0 e 1, estas intensidades fornecem informações sobre o brilho de uma determinada cor. Este canal pode ser definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$V_{(i,j)} = \text{MAX}_{(i,j)}. \quad (2.9)$$

A Figura 2.4 ilustra o espaço de cores \mathbf{HSV} .

Figura 2.4: Ilustração do espaço de cores \mathbf{HSV} .



Fonte: (EARLEY, 2015).

Por fim, é importante mencionar que a transformação inversa, isto é, a transformação do espaço de cores \mathbf{HSV} para o espaço de cores \mathbf{RGB} é definida para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$[R_{(i,j)}, G_{(i,j)}, B_{(i,j)}]^T = \begin{cases} [\mathcal{C}_{(i,j)} + \Delta\mathcal{C}V_{(i,j)}, \mathcal{O}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{C}V_{(i,j)}]^T, & \text{se } 0^\circ < H_{(i,j)} \leq 60^\circ, \\ [\mathcal{O}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{C}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{C}V_{(i,j)}]^T, & \text{se } 60^\circ \leq H_{(i,j)} \leq 120^\circ, \\ [\mathcal{C}V_{(i,j)}, \mathcal{C}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{O}_{(i,j)} + \mathcal{C}V_{(i,j)}]^T, & \text{se } 120^\circ \leq H_{(i,j)} \leq 180^\circ, \\ [\mathcal{C}V_{(i,j)}, \mathcal{O}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{C}_{(i,j)} + \mathcal{C}V_{(i,j)}]^T, & \text{se } 180^\circ \leq H_{(i,j)} \leq 240^\circ, \\ [\mathcal{O}_{(i,j)} + \mathcal{C}V_{(i,j)}, \mathcal{C}V_{(i,j)}, \mathcal{C}_{(i,j)} + \mathcal{C}V_{(i,j)}]^T, & \text{se } 240^\circ \leq H_{(i,j)} \leq 300^\circ, \\ [\mathcal{C}_{(i,j)} + \Delta\mathcal{C}V_{(i,j)}, \Delta\mathcal{C}V_{(i,j)}, \mathcal{O}_{(i,j)} + \Delta\mathcal{C}V_{(i,j)}]^T, & \text{se } 300^\circ \leq H_{(i,j)} \leq 360^\circ, \end{cases} \quad (2.10)$$

onde:

$$\mathcal{C}_{(i,j)} = V_{(i,j)} \times S_{(i,j)}, \quad (2.11)$$

$$\Delta\mathcal{C}V_{(i,j)} = V_{(i,j)} - \mathcal{C}_{(i,j)}, \quad (2.12)$$

$$\mathcal{O}_{(i,j)} = \mathcal{C}_{(i,j)} \times (1 - |(H_{(i,j)}/60^\circ) \bmod 2 - 1|). \quad (2.13)$$

2.2 Conceitos de Pré-Processamento de Imagens Dermatológicas

O objetivo das técnicas de pré-processamento (ou filtragem) é processar uma imagem de modo que a imagem processada seja mais adequada do que a imagem original para uma aplicação específica (GONZALEZ; WOODS, 2008). Nesta seção, serão apresentados alguns conceitos de pré-processamento de imagens, fundamentais para métodos de segmentação de lesões melanocíticas.

2.2.1 Convolução

A filtragem é dita linear quando a intensidade em um ponto da imagem processada é resultado de uma combinação linear das intensidades da imagem de entrada (SZELISKI, 2010). Assim, a filtragem linear de uma imagem \mathbf{I} com m linhas e n colunas por uma máscara (ou filtro) \mathbf{G} com $2d_h + 1$ linhas e $2d_v + 1$ colunas é denominada convolução e pode ser definida para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por (JAIN, 1989):

$$(I \otimes F)_{(i,j)} = \begin{cases} \sum_{k=-d_h}^{d_h} \sum_{l=-d_v}^{d_v} I_{(i-k, j-l)} G_{(k,l)}, & \text{se } (i, j) \in \{d_h, \dots, m - d_h\} \times \{d_v, \dots, n - d_v\}, \\ 0, & \text{se } (i, j) \notin \{d_h, \dots, m - d_h\} \times \{d_v, \dots, n - d_v\}. \end{cases} \quad (2.14)$$

A escolha do filtro \mathbf{G} depende do objetivo da aplicação. De forma geral, filtros podem ser classificados como passa-baixas ou passa-altas (SONKA; HLAVAC; BOYLE, 2014).

2.2.2 Filtros Passa-Baixas

Filtros são denominados passa-baixas quando são projetados para atenuar ou eliminar as regiões de borda e/ou detalhes finos na imagem. O efeito da filtragem por um filtro passa-baixas é a suavização da imagem, provocando um borramento na mesma (GONZALEZ; WOODS, 2008).

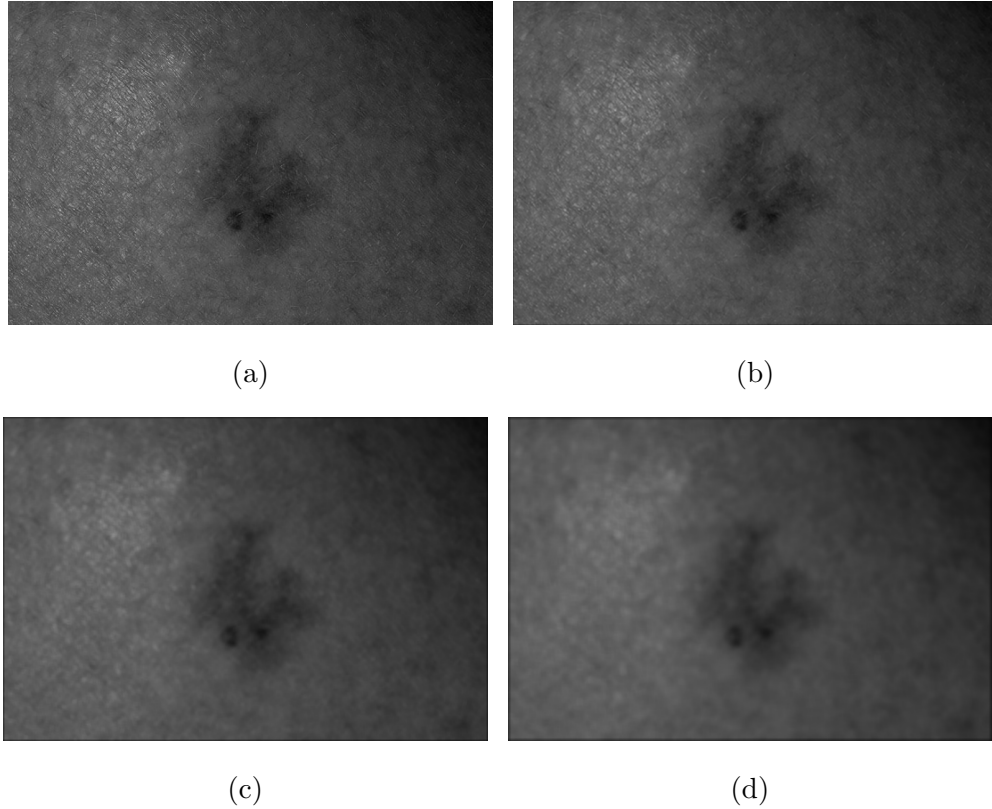
Os filtros passa-baixas mais utilizados são os filtros de média (MARQUES FILHO; VIEIRA NETO, 1999), definidos para todo $(k, l) \in \{-d_h, \dots, d_h\} \times \{-d_v, \dots, d_v\}$ por:

$$G_{(k,l)} = \frac{1}{(2d_h + 1)(2d_v + 1)}. \quad (2.15)$$

Na escolha do tamanho do filtro (i.e., na escolha de d_h e d_v) deve-se ter em mente que quanto maior o filtro, maior será o borramento na imagem resultante (MARQUES FILHO; VIEIRA NETO, 1999). Para exemplificar isso, considere a Figura 2.5.

Os filtros gaussianos, também são passa-baixas de particular interesse e podem ser definidos para um desvio padrão σ e $(k, l) \in \{-d_h, \dots, d_h\} \times \{-d_v, \dots, d_v\}$ por:

Figura 2.5: Exemplo de filtragem pelo filtro da média: (a) Imagem original; (b) Imagem filtrada pelo filtro da média 3×3 ; (c) Imagem filtrada pelo filtro da média 7×7 ; (d) Imagem filtrada pelo filtro da média 11×11 .

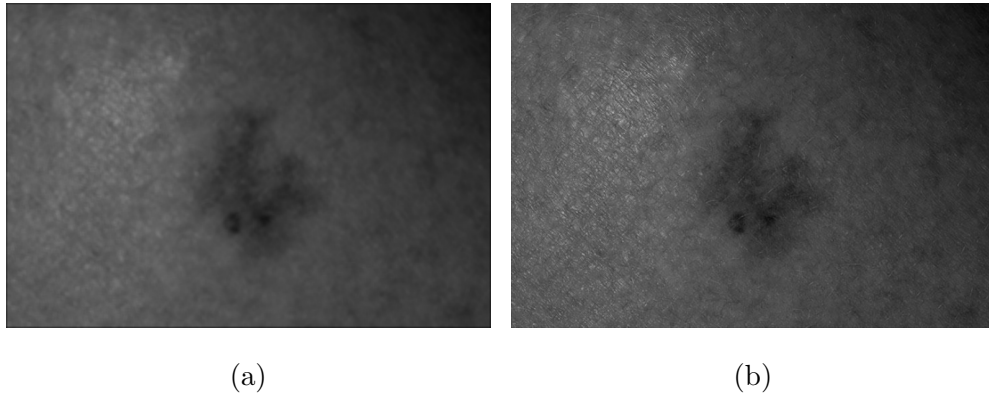


$$G_{\sigma(k,l)} = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{k^2 + l^2}{2\sigma^2} \right\}. \quad (2.16)$$

Filtros gaussianos são mais robustos do que os filtros de média no sentido de que tendem a borrar menos a imagem resultante (SZELISKI, 2010). Para exemplificar isso considere a Figura 2.6.

Conforme pode ser observado, a filtragem através de um filtro linear é completamente definida pelo filtro, pois é realizada através da convolução da imagem pelo filtro. Por outro lado, a filtragem através de um filtro não-linear é definida não só pelo filtro, mas também pela operação a ser realizada, uma vez que a filtragem não pode ser realizada através da operação de convolução. Filtros não-lineares projetados para produzir uma intensidade no ponto (i, j) dependente da ordenação das intensidades ao redor de (i, j) na imagem de entrada são denominados filtros não-lineares de estatística de ordem e são usualmente classificados como filtros passa-baixas. Estes filtros são usualmente mais robustos do que

Figura 2.6: Comparação entre o filtro da média e o filtro gaussiano: (a) Imagem filtrada pelo filtro da média 7×7 ; (b) Imagem filtrada pelo filtro gaussiano 7×7 ($\sigma = 0.5$).



filtros lineares, pois além de eliminarem detalhes finos (i.e., ruído impulsivo), tendem a preservar melhor as bordas da imagem. O filtro mais bem conhecido desta categoria é o filtro da mediana (GONZALEZ; WOODS, 2008). Uma imagem \mathbf{I}^G , resultado da filtragem da imagem \mathbf{I} pelo filtro da mediana, pode ser definida para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$I_{(i,j)}^G = \text{mediana}\{I_{(i-k,j-l)}, (k, l) \in \{-d_h, \dots, d_h\} \times \{-d_v, \dots, d_v\}\}. \quad (2.17)$$

2.2.3 Filtros Passa-Altas

Sabendo-se que o cálculo da média das intensidades em um trecho da imagem produz como efeito a remoção de bordas e detalhes finos da imagem e que o conceito de média é análogo à operação de integração, é razoável esperar que a diferenciação produza o efeito oposto e enfatize as bordas e as regiões de alto contraste da imagem. Baseado nisso, a forma mais comum de se realçar bordas e regiões de alto contraste em uma imagem é através da magnitude do gradiente (MARQUES FILHO; VIEIRA NETO, 1999). A magnitude do gradiente pode ser definida para uma imagem analógica \mathcal{J} (i.e., uma função das variáveis contínuas x e y) como:

$$|\nabla \mathcal{J}| = \sqrt{\left(\frac{\partial \mathcal{J}}{\partial x}\right)^2 + \left(\frac{\partial \mathcal{J}}{\partial y}\right)^2}. \quad (2.18)$$

A derivada $\frac{\partial \mathcal{J}}{\partial x}$ pode ser aproximada para uma imagem digital \mathbf{I} através da filtragem desta por um dos seguintes filtros, os quais são denominados filtro de Prewitt e filtro de Sobel, respectivamente (SONKA; HLAVAC; BOYLE, 2014):

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad (2.19)$$

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}. \quad (2.20)$$

Assim, a magnitude do gradiente $|\nabla \mathcal{J}|$ pode ser aproximada para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$|\nabla I_{(i,j)}| = \sqrt{(I \otimes G_x)_{(i,j)}^2 + (I \otimes G_y)_{(i,j)}^2}, \quad (2.21)$$

onde $\mathbf{G}_y = \mathbf{G}_x^\top$.

2.3 Conceitos de Extração de Características de Imagens Dermatológicas

O processo de digitalização permite representar uma imagem (sinal bidimensional) como combinação linear de impulsos unitários no espaço (JAIN, 1989):

$$\mathbf{I} = \sum_{k_1=1}^m \sum_{k_2=1}^n I_{(k_1, k_2)} \boldsymbol{\delta}_{(i-k_1, j-k_2)}, \quad (2.22)$$

onde $\boldsymbol{\delta}_{(i-k_1, j-k_2)}$ denota um impulso unitário no espaço, definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$\boldsymbol{\delta}_{(i-k_1, j-k_2)} = \begin{cases} 1, & \text{se } i = k_1 \text{ e } j = k_2, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.23)$$

De acordo com RUBINSTEIN; BRUCKSTEIN; ELAD (2010), esta representação é conveniente para exibir imagens, porém, é usualmente ineficiente para analisar imagens. Assim, técnicas de análise de imagens geralmente requerem representações mais robustas, as quais devem salientar as características mais relevantes da imagem. Por exemplo, para tarefas de classificação, a representação deve salientar as características mais discriminativas. Para tarefas de filtragem, a representação deve sinalizar eficientemente o ruído. Para tarefas de compressão, a representação deve capturar uma grande parte do sinal com apenas alguns poucos coeficientes.

A representação de um sinal $\mathbf{y} \in \mathbb{R}^M$ pode ser baseada em um dicionário $\mathbf{D} \in \mathbb{R}^{M \times \mathcal{K}}$, no qual as colunas $\{\mathbf{d}_j\}_{j=1}^{\mathcal{K}}$ são denominadas átomos do dicionário. Assim, a representação de \mathbf{y} em \mathbf{D} consiste basicamente em decompor o sinal \mathbf{y} como combinação linear dos átomos $\{\mathbf{d}_j\}_{j=1}^{\mathcal{K}}$. Se \mathbf{D} é ortogonal, o vetor $\mathbf{x} \in \mathbb{R}^{\mathcal{K}}$ com os coeficientes da combinação linear é calculado por produtos internos entre o sinal e os átomos do dicionário:

$$\mathbf{x} = \mathbf{D}^T \mathbf{y}. \quad (2.24)$$

Por outro lado, se \mathbf{D} não é ortogonal, a ideia é selecionar um vetor de coeficientes $\hat{\mathbf{x}}$ tal que (RUBINSTEIN; BRUCKSTEIN; ELAD, 2010):

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{E}(\mathbf{y}, \mathbf{D}\mathbf{x}), \quad (2.25)$$

onde \mathcal{E} é a função desejada para medir o erro da representação \mathbf{x} de \mathbf{y} em \mathbf{D} .

De forma geral, dicionários podem ser classificados como (WRIGHT et al., 2010):

- dicionários pré-definidos: são caracterizados por alguma formulação analítica. A transformada de Fourier e a transformada wavelet (MALLAT, 2008) são duas das ferramentas mais utilizadas para obter dicionários pré-definidos;
- dicionários aprendidos: são aprendidos especificamente para um determinado conjunto de sinais. Dicionários aprendidos têm proporcionado melhores resultados na grande maioria dos trabalhos estado-da-arte em restauração e classificação de imagens (MAIRAL; BACH; PONCE, 2012).

Os seguintes métodos podem ser empregados para o aprendizado de dicionários e servem de base para o entendimento do método proposto e para o entendimento de outros métodos de segmentação de lesões melanocíticas em imagens macroscópicas (apresentados no Capítulo 3):

- PCA: este método será apresentado na Subseção 2.3.1 e o nome PCA é um acrônimo para *Principal Component Analysis* - análise de componentes principais;
- ICA: este método será apresentado na Subseção 2.3.2 e o nome ICA é um acrônimo para *Independent Component Analysis* - análise de componentes independentes;
- MOD: este método será apresentado na Subseção 2.3.3 e o nome MOD é um acrônimo para *Method of Optimal Directions* - método de direções otimizadas;
- K-SVD: este método também será apresentado na Subseção 2.3.3. O nome K-SVD advém do fato de que este método pode ser visto como uma generalização do método de K -médias (apresentado na Subseção 2.4.4) e do fato deste método utilizar o método SVD, cujo nome é um acrônimo para *Singular Value Decomposition* - decomposição em valores singulares.
- NMF: este método será apresentado na Subseção 2.3.4 e o nome NMF é um acrônimo para *Nonnegative Matrix Factorization* - fatoração de matrizes não-negativas;
- ITDL: este método será apresentado na Subseção 2.3.5 e o nome ITDL é um acrônimo para *Information-Theoretic Dictionary Learning* - aprendizado de dicionário baseado em teoria da informação.

2.3.1 PCA

O método de PCA (JOLLIFFE, 2002) tem por objetivo representar um conjunto de sinais $\{\mathbf{y}_i\}_{i=1}^N$ em um conjunto de átomos $\{\hat{\mathbf{d}}_j\}_{j=1}^K$ de forma que este conjunto de átomos maximize a variância das representações $\{\mathbf{x}_i\}_{i=1}^N$. O primeiro passo do método é calcular o sinal médio:

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i. \quad (2.26)$$

Desta forma, um conjunto de sinais com média zero, $\{\bar{\mathbf{y}}_i\}_{i=1}^N$, pode ser obtido por:

$$\bar{\mathbf{y}}_i = \mathbf{y}_i - \bar{\mathbf{y}}. \quad (2.27)$$

Assim, o primeiro átomo, denominado componente principal, pode ser computado por:

$$\hat{\mathbf{d}}_1 = \arg \max_a \left\{ \frac{\mathbf{d}^\top \overline{\mathbf{Y}\mathbf{Y}^\top} \mathbf{d}}{\mathbf{d}^\top \mathbf{d}} \right\}, \quad (2.28)$$

onde $\overline{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_N]$. A Eq. (2.28) pode ser reconhecida como o quociente de Rayleigh, onde o quociente máximo é igual ao maior autovalor de $\overline{\mathbf{Y}\mathbf{Y}^\top}$ (GOLUB; VAN LOAN, 2013). Como resultado disso, o primeiro átomo $\hat{\mathbf{d}}_1$ é o autovetor correspondente ao maior autovalor de $\overline{\mathbf{Y}\mathbf{Y}^\top}$, o segundo átomo $\hat{\mathbf{d}}_2$ é o autovetor correspondente ao segundo maior autovalor de $\overline{\mathbf{Y}\mathbf{Y}^\top}$ e assim por diante. Por fim, um dicionário $\hat{\mathbf{D}} = [\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \dots, \hat{\mathbf{d}}_\mathcal{K}] \in \mathbb{R}^{M \times \mathcal{K}}$ com $\mathcal{K} < N$ átomos pode ser construído e a representação \mathbf{x}_i de cada sinal \mathbf{y}_i em $\hat{\mathbf{D}}$ pode ser obtida por:

$$\mathbf{x}_i = \hat{\mathbf{D}}^\top \bar{\mathbf{y}}_i. \quad (2.29)$$

2.3.2 ICA

O método de ICA (HYVÄRINEN; OJA, 2000) tem por objetivo representar um conjunto de sinais em um dicionário de forma que as representações dos sinais no dicionário sejam independentes e não-gaussianas. Para exemplificar uma aplicação de ICA, considere que dois microfones (e.g., microfone 1 e microfone 2) estão em diferentes localizações de uma sala onde duas pessoas (e.g., pessoa 1 e pessoa 2) estão falando simultaneamente. Considere que o microfone 1 gravou um sinal \mathbf{y}_1 e que o microfone 2 gravou um sinal \mathbf{y}_2 . Assim, os sinais \mathbf{y}_1 e \mathbf{y}_2 podem ser expressos, respectivamente, por:

$$\mathbf{y}_1 = d_{(1,1)} \mathbf{x}_1 + d_{(1,2)} \mathbf{x}_2, \quad (2.30)$$

$$\mathbf{y}_2 = d_{(2,1)}\mathbf{x}_1 + d_{(2,2)}\mathbf{x}_2, \quad (2.31)$$

onde \mathbf{x}_1 e \mathbf{x}_2 são os sinais fonte emitidos pela pessoa 1 e pela pessoa 2, respectivamente. Os coeficientes $d_{(1,1)}$, $d_{(1,2)}$, $d_{(2,1)}$ e $d_{(2,2)}$ são parâmetros dependentes da distância entre a pessoa e o microfone. Assim, o objetivo é encontrar os sinais originais (i.e., os sinais fonte) \mathbf{x}_1 e \mathbf{x}_2 a partir dos sinais observados \mathbf{y}_1 e \mathbf{y}_2 . Para resolver este problema, são assumidas duas premissas (HYVÄRINEN; OJA, 2000):

1. os sinais originais \mathbf{x}_1 e \mathbf{x}_2 são independentes;
2. os valores de \mathbf{x}_1 e de \mathbf{x}_2 possuem distribuições não-gaussianas.

Com base nestas premissas, diversos algoritmos tem sido propostos para ICA. Dentre estes, um dos mais populares é o FastICA (HYVÄRINEN, 1999). O algoritmo FastICA deve receber como entrada um conjunto de sinais observados $\{\mathbf{y}_i\}_{i=1}^N$ e o número desejado de sinais fonte $\mathcal{K} < N$. Assim como o algoritmo de PCA, este algoritmo também inicia pelo cálculo do sinal médio, $\bar{\mathbf{y}}$, através da Eq. (2.26) e um conjunto de sinais com média zero, $\{\bar{\mathbf{y}}_i\}_{i=1}^N$, pode ser obtido através da Eq. (2.27). Em seguida, o conjunto de sinais decorrelacionados e com média zero, $\{\hat{\mathbf{y}}_i\}_{i=1}^N$, pode ser obtido por:

$$\hat{\mathbf{Y}} = \mathbf{C}^{-1/2}\bar{\mathbf{Y}}, \quad (2.32)$$

onde:

$$\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_N], \quad (2.33)$$

$$\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_N], \quad (2.34)$$

$$\mathbf{C} = \frac{1}{N} \overline{\mathbf{Y}\mathbf{Y}^\top}. \quad (2.35)$$

Depois, o algoritmo FastICA executa as seguintes etapas:

1. $\mathbf{D} \leftarrow$ matriz pseudo-aleatória no $\mathbb{R}^{M \times \mathcal{K}}$;
2. enquanto $1 - \min(\text{abs}(\text{diag}(\mathbf{D}\mathbf{D}^{+\top}))) < \epsilon$:

$$(a) \mathbf{D}^{+\top} \leftarrow G'(\mathbf{D}^\top \widehat{\mathbf{Y}}) \widehat{\mathbf{Y}}^\top - \text{diag}[G''(\mathbf{D}^\top \widehat{\mathbf{Y}}) \mathbf{1}_N] \mathbf{D}^\top;$$

$$(b) \mathbf{D}^\top \leftarrow (\mathbf{D}^{+\top} \mathbf{D}^+)^{-1/2} \mathbf{D}^{+\top};$$

aqui, G' e G'' denotam as derivadas de primeira e segunda ordem de uma função não-linear e não-quadrática G (denominada função de contraste), $\mathbf{1}_N$ denota um vetor com dimensão $N \times 1$ de 1's. De acordo com HYVÄRINEN (1999), escolhas adequadas para $G(u)$ são:

$$G(u) = \frac{1}{\alpha} \log(\cosh(\alpha u)), \quad (2.36)$$

$$G(u) = -\frac{1}{\beta} \exp\left\{-\frac{\beta u^2}{2}\right\}. \quad (2.37)$$

Por fim, cada sinal \mathbf{x}_i pode ser obtido por:

$$\mathbf{x}_i = \mathbf{D}^\top \widehat{\mathbf{y}}_i. \quad (2.38)$$

Embora ICA tenha surgido como um método de separação cega de fontes, este método tem sido utilizado para diferentes aplicações e uma aplicação de particular interesse é a extração não-supervisionada de características (HYVÄRINEN; OJA, 2000).

2.3.3 MOD e K-SVD

Tanto MOD quanto K-SVD são algoritmos iterativos projetados para representar um conjunto de sinais $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{M \times N}$. Assim, o objetivo destes algoritmos é encontrar um dicionário $\widehat{\mathbf{D}} = [\widehat{\mathbf{d}}_1, \widehat{\mathbf{d}}_2, \dots, \widehat{\mathbf{d}}_{\mathcal{K}}] \in \mathbb{R}^{M \times \mathcal{K}}$ e um conjunto de vetores de coeficientes $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2, \dots, \widehat{\mathbf{x}}_N] \in \mathbb{R}^{\mathcal{K} \times N}$ que minimizem o erro de representação:

$$\{\widehat{\mathbf{D}}, \widehat{\mathbf{X}}\} = \arg \min_{\mathbf{D}, \mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ s.t } \forall i, \|\mathbf{x}_i\|_0 \leq t_0, \quad (2.39)$$

onde $\|\mathbf{x}_i\|_0$ denota o número de elementos não-nulos em \mathbf{x}_i (i.e., norma ℓ_0), t_0 denota o nível de esparsidade mínimo (parâmetro de entrada dos algoritmos) e $\|\cdot\|_F$ denota a norma de Frobenius, definida para uma matriz \mathbf{A} por:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{(i,j)} A_{(i,j)}^2}. \quad (2.40)$$

Os algoritmos MOD e K-SVD alternam entre os passos de codificação esparsa e de atualização do dicionário. Primeiramente, um dicionário \mathbf{D} é inicializado aleatoriamente e as colunas de \mathbf{D} são normalizadas pela norma ℓ_2 . Depois, a cada iteração, são executados os seguintes passos:

1. codificação esparsa: neste passo, \mathbf{D} é fixado e o seguinte problema de otimização é resolvido de forma aproximada usando algum algoritmo de representação esparsa tal como OMP (ELAD, 2010) para computar os vetores de coeficientes $\{\widehat{\mathbf{x}}_i\}_{i=1}^N$:

$$\widehat{\mathbf{x}}_i = \arg \min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \text{ s.t } \forall i, \|\mathbf{x}_i\|_0 \leq t_0. \quad (2.41)$$

2. atualização do dicionário: neste passo é que se diferenciam os algoritmos MOD e K-SVD. O algoritmo MOD atualiza todos os átomos simultaneamente para resolver o seguinte problema de otimização:

$$\widehat{\mathbf{D}} = \arg \min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\widehat{\mathbf{X}}\|_F^2, \quad (2.42)$$

cuja solução é dada por (PATEL; CHELLAPPA, 2013):

$$\widehat{\mathbf{D}} = \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^{-1}. \quad (2.43)$$

Apesar do algoritmo MOD usualmente convergir em poucas iterações, a inversão da matriz introduz uma alta complexidade no algoritmo. No caso do K-SVD, o dicionário é atualizado átomo por átomo, de uma maneira eficiente que não exige a inversão de matrizes (AHARON; ELAD; BRUCKSTEIN, 2006). Para isso, a Eq. (2.42) é reescrita como:

$$\widehat{\mathbf{D}} = \arg \min_{\mathbf{D}} \left\| \mathbf{Y} - \sum_{j \neq j_0} \mathbf{d}_j \mathbf{x}_j^\top - \mathbf{d}_{j_0} \mathbf{x}_{j_0}^\top \right\|_F^2. \quad (2.44)$$

Uma vez que o objetivo é atualizar \mathbf{d}_{j_0} e $\mathbf{x}_{j_0}^\top$, o primeiro termo da Eq. (2.44) pode ser pré-computado:

$$\mathbf{E}_{j_0} = \mathbf{Y} - \sum_{j \neq j_0} \mathbf{d}_j \mathbf{x}_j^\top. \quad (2.45)$$

Assim, os valores ótimos de \mathbf{d}_{j_0} e $\mathbf{x}_{j_0}^\top$ são encontrados através de uma decomposição em SVD (GOLUB; VAN LOAN, 2013). Em particular, enquanto fixada a cardinalidade de todas as representações, somente um subconjunto das colunas de \mathbf{E}_{j_0} é considerado. Desta forma, a atualização do dicionário usando K-SVD é mais rápida do que a atualização do algoritmo MOD (PATEL; CHELLAPPA, 2013).

Tanto MOD quanto K-SVD são métodos de aprendizagem de dicionário que fornecem uma representação esparsa, uma vez que somente alguns poucos coeficientes não-nulos são permitidos (i.e., o número máximo de coeficientes não-nulos é limitado por t_0). Representações esparsas têm sido consideradas um princípio fundamental tanto para a extração de características quanto para a classificação (BACH et al., 2012). Além disso, é também desejável que o dicionário seja compacto (QIU; PATEL; CHELLAPPA, 2014). Um dicionário compacto consiste de átomos incoerentes e encoraja sinais similares, os quais

pertencem provavelmente a uma mesma classe, a serem descritos por um conjunto similar de átomos com coeficientes similares (QIU; JIANG; CHELLAPPA, 2011). Porém, MOD e K-SVD são métodos para o aprendizado de dicionários redundantes, sendo usualmente adequados somente para tarefas de reconstrução (i.e., não são adequados para tarefas de classificação) (QIU; PATEL; CHELLAPPA, 2014). Para contornar este problema, considere o fato de que imagens são usualmente representadas por matrizes não-negativas⁴. Considere também a existência de uma relação entre não-negatividade e esparsidade, a qual pode ser explicada tanto pela teoria dos conjuntos ativos quanto pela teoria da inferência bayesiana (LI; NGOM, 2013). Assim, uma alternativa natural para gerar representações esparsas e ao mesmo tempo compactas são os métodos de NMF apresentados a seguir.

2.3.4 NMF

Dada uma matriz $\mathbf{Y} \in \mathbb{R}^{M \times N}$ não-negativa e um inteiro positivo $\mathcal{K} < \min\{M, N\}$, o problema da fatoração de matrizes não-negativas consiste em encontrar uma matriz não-negativa $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{\mathcal{K}}] \in \mathbb{R}^{M \times \mathcal{K}}$ (i.e, um dicionário) e uma matriz não-negativa $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{\mathcal{K} \times N}$ (i.e., uma representação de \mathbf{Y} em \mathbf{D}) que minimizem a seguinte função:

$$f = \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2. \quad (2.46)$$

Um desafio importante da minimização numérica da função definida pela Eq. (2.46) é a existência de mínimos locais devido à não-convexidade de f em relação a \mathbf{D} e \mathbf{X} (BERRY et al., 2007). Vários algoritmos têm sido propostos para resolver este problema, estes algoritmos podem ser classificados como: (1) algoritmos de atualização multiplicativa; (2) algoritmos de gradiente descendente; (3) algoritmos de mínimos quadrados alternados. Geralmente, os algoritmos de mínimos quadrados alternados convergem mais rápido, de forma mais consistente e são menos sensíveis a valores iniciais (MATHWORKS, 2015). Os algoritmos de mínimos quadrados alternados exploram o fato de que embora a função definida na Eq. (2.46) não seja simultaneamente convexa em \mathbf{D} e \mathbf{X} , ela é convexa em \mathbf{D}

⁴Uma matriz não-negativa tem todos os seus elementos maiores ou iguais a zero.

ou \mathbf{X} (BERRY et al., 2007). Estes algoritmos dependem das seguintes decomposições do objetivo para \mathbf{D} e \mathbf{X} :

$$\begin{aligned} \mathbf{Y} &= \mathbf{D}\mathbf{X} \\ \Leftrightarrow \mathbf{Y}\mathbf{X}^\top &= \mathbf{D}\mathbf{X}\mathbf{X}^\top \\ \Leftrightarrow \mathbf{D} &= \mathbf{Y}\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}, \end{aligned} \tag{2.47}$$

$$\begin{aligned} \mathbf{Y} &= \mathbf{D}\mathbf{X} \\ \Leftrightarrow \mathbf{D}^\top\mathbf{Y} &= \mathbf{D}^\top\mathbf{D}\mathbf{X} \\ \Leftrightarrow \mathbf{X} &= (\mathbf{D}^\top\mathbf{D})^{-1}\mathbf{D}^\top\mathbf{Y}. \end{aligned} \tag{2.48}$$

Uma implementação típica para a fatoração de matrizes não-negativas via mínimos quadrados alternados é mostrada no Algoritmo 1. Nesta implementação, em cada iteração, \mathbf{D} e \mathbf{X} são atualizados e este processo é repetido até satisfazer ao menos um dos critérios de convergência (i.e., $\Delta x \leq \text{tol}x$ ou $\Delta \text{fun} \leq \text{tol} \text{fun}$ ou $\text{iter} > \text{maxiter}$). Por fim, as colunas de \mathbf{D} e as linhas de \mathbf{X} são normalizadas pela norma ℓ_2 .

2.3.5 ITDL

QIU; PATEL; CHELLAPPA (2014) observaram que três critérios são críticos para dicionários em tarefas de classificação:

1. compactividade: um dicionário compacto consiste de átomos incoerentes e encoraja sinais similares, os quais provavelmente pertencem a uma mesma classe, a serem descritos por um conjunto similar de átomos com coeficientes similares (QIU; JIANG; CHELLAPPA, 2011);
2. discriminabilidade: um dicionário discriminativo encoraja sinais de diferentes classes a serem descritos por um conjunto diferente de átomos ou pelo mesmo conjunto de átomos com coeficientes diferentes (RODRIGUEZ; SAPIRO, 2008; MAIRAL et al., 2008);

Algoritmo 1 Algoritmo para computar a fatoração de uma matriz não-negativa \mathbf{Y} .

```

1: procedure NMF( $\mathbf{Y}, \mathcal{K}, maxiter, tol_x, tol_{fun}$ )
2:    $\mathbf{D} \leftarrow$  uma matriz pseudo-aleatória não-negativa do  $\mathbb{R}^{M \times \mathcal{K}}$ 
3:    $\mathbf{X} \leftarrow$  uma matriz pseudo-aleatória não-negativa do  $\mathbb{R}^{\mathcal{K} \times N}$ 
4:    $residual \leftarrow \frac{1}{\sqrt{MN}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F$ 
5:   for  $iter = 1$  to  $maxiter$  do
6:      $\mathbf{D}_{old} \leftarrow \mathbf{D}$ 
7:      $\mathbf{D} \leftarrow \mathbf{Y}\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}$ 
8:     setar todos os elementos negativos em  $\mathbf{D}$  para zero
9:      $\Delta\mathbf{D} \leftarrow \frac{\max(abs(\mathbf{D} - \mathbf{D}_{old}))}{\max(\mathbf{D}_{old}) + 10^{-9}}$ 
10:     $\mathbf{X}_{old} \leftarrow \mathbf{X}$ 
11:     $\mathbf{X} \leftarrow (\mathbf{D}^\top\mathbf{D})^{-1}\mathbf{D}^\top\mathbf{Y}$ 
12:    setar todos os elementos negativos em  $\mathbf{X}$  para zero
13:     $\Delta\mathbf{X} \leftarrow \frac{\max(abs(\mathbf{X} - \mathbf{X}_{old}))}{\max(\mathbf{X}_{old}) + 10^{-9}}$ 
14:     $\Delta x \leftarrow \max\{\Delta\mathbf{D}, \Delta\mathbf{X}\}$ 
15:     $residual_{old} \leftarrow residual$ 
16:     $residual \leftarrow \frac{1}{\sqrt{MN}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F$ 
17:     $\Delta fun \leftarrow |residual - residual_{old}|$ 
18:    if  $(\Delta x \leq tol_x)$  or  $(\Delta fun \leq tol_{fun})$  then
19:      break
20:    for  $i = 1$  to  $\mathcal{K}$  do
21:       $\alpha \leftarrow$  norma  $\ell_2$  da  $i$ -ésima linha de  $\mathbf{X}$ 
22:      if  $\alpha \neq 0$  then
23:        dividir todos os elementos da  $i$ -ésima linha de  $\mathbf{X}$  por  $\alpha$ 
24:        multiplicar todos os elementos da  $i$ -ésima coluna de  $\mathbf{D}$  por  $\alpha$ 
25:  return  $\mathbf{D}, \mathbf{X}$ 

```

3. representatividade: um dicionário representativo é aquele que fornece uma representação robusta para os sinais no sentido de minimizar o erro de representação dos sinais nos átomos do dicionário (YANG; ZHANG; FENG, 2011).

Com base nisso, QIU; PATEL; CHELLAPPA (2014) propuseram o método ITDL para aprender um dicionário que seja ao mesmo tempo compacto, representativo e discrimi-

nativo. Este método recebe como entrada um conjunto de sinais $\mathbf{Y} \in \mathbb{R}^{M \times N}$ e os seus respectivos rótulos $\mathbf{C} \in \mathbb{N}^N$, um dicionário inicial $\mathbf{D}^{(0)} \in \mathbb{R}^{M \times \mathcal{K}}$ aprendido via K-SVD, uma representação $\mathbf{X} \in \mathbb{R}^{\mathcal{K} \times N}$ de \mathbf{Y} em $\mathbf{D}^{(0)}$ e o parâmetro $k < \mathcal{K}$, correspondente ao número de átomos a serem selecionados. Assim, um subconjunto de k átomos de \mathbf{D} é selecionado:

$$\mathbf{D}_k = \arg \max_{\mathbf{D} \in \mathbb{R}^{M \times k}} \left\{ \lambda_1 \text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D}) + \lambda_2 \text{MI}(\mathbf{X}, \mathbf{C}) + \lambda_3 \text{MI}(\mathbf{Y}, \mathbf{D}) \right\}, \quad (2.49)$$

onde $\text{MI}(\mathbf{A}, \mathbf{B})$ denota a informação mútua entre as matrizes \mathbf{A} e \mathbf{B} , definida por:

$$\text{MI}(\mathbf{A}, \mathbf{B}) = \sum_{A_{(i,j)}} \sum_{B_{(i,j)}} p(A_{(i,j)}, B_{(i,j)}) \times \log \left(\frac{p(A_{(i,j)}, B_{(i,j)})}{p(A_{(i,j)})p(B_{(i,j)})} \right), \quad (2.50)$$

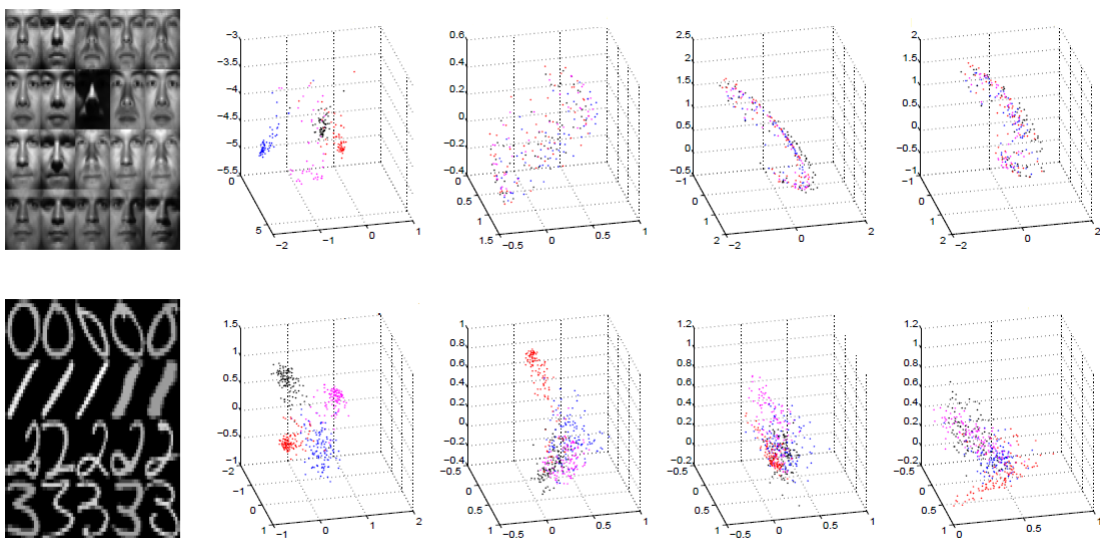
aqui, $p(A_{(i,j)}, B_{(i,j)})$ denota a probabilidade conjunta de \mathbf{A} e \mathbf{B} , $p(A_{(i,j)})$ e $p(B_{(i,j)})$ denotam as probabilidades marginais de \mathbf{A} e \mathbf{B} , respectivamente.

QIU; PATEL; CHELLAPPA (2014) mostraram que os termos de compactividade (i.e., $\text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D})$), discriminabilidade (i.e., $\text{MI}(\mathbf{X}_D, \mathbf{C})$) e representatividade (i.e., $\text{MI}(\mathbf{Y}, \mathbf{D})$) podem ser computados individualmente através de fórmulas fechadas. A Figura 2.7 compara representações obtidas pelo método ITDL com outros métodos estado-da-arte em aprendizado de dicionários para tarefas de classificação de imagens. Observe que o método ITDL não pode ser empregado em um esquema não-supervisionado de segmentação de lesões melanocíticas pois exige o conhecimento prévio da classe dos sinais (i.e., exige \mathbf{C}). Além disso, o parâmetro de entrada k não precisa ser fornecido como entrada uma vez que pode ser adicionado ao processo de maximização. Assim, uma versão não-supervisionada e não-paramétrica deste método é proposta na Seção 4.1.

2.4 Conceitos de Segmentação de Imagens Dermatológicas

O objetivo da segmentação é particionar uma imagem em conjuntos de pixels, os quais devem possuir uma forte correlação com os objetos ou áreas do mundo real que estão contidos na imagem (SONKA; HLAVAC; BOYLE, 2014). Na prática, a segmentação de uma imagem implica na divisão ou separação da mesma em regiões de características

Figura 2.7: Representações em diferentes dicionários, aprendidos respectivamente por ITDL (QIU; PATEL; CHELLAPPA, 2014), SOMP (TROPPEL; GILBERT; STRAUSS, 2006), MMI-1 (QIU; JIANG; CHELLAPPA, 2011) e MMI-2 (QIU; JIANG; CHELLAPPA, 2011): cada ponto corresponde a uma imagem, cada dicionário consiste de três átomos e cada cor identifica uma classe de imagens.



Fonte: (QIU; PATEL; CHELLAPPA, 2014).

semelhantes. A característica mais básica para segmentar imagens é a intensidade dos pixels (PRATT, 2007).

O processo mais simples para segmentar imagens é a limiarização, onde cada intensidade é comparada com um determinado limiar t e cada pixel da imagem é classificado como objeto ou fundo de acordo com a sua intensidade (SONKA; HLAVAC; BOYLE, 2014), ou seja:

$$M_{(i,j)} = \begin{cases} 1, & \text{se } I_{(i,j)} \geq t, \\ 0, & \text{se } I_{(i,j)} < t, \end{cases} \quad (2.51)$$

onde \mathbf{M} é uma imagem binária tal que $M_{(i,j)} = 1$ se o pixel (i, j) foi classificado como objeto ou $M_{(i,j)} = 0$ se o pixel (i, j) foi classificado como fundo.

Nesta seção, serão apresentados métodos que fornecem uma base para o entendimento do método proposto e/ou para o entendimento dos métodos tratados no Capítulo 3. A Subseção 2.4.1 apresenta método de Otsu, o qual permite selecionar automaticamente um limiar para métodos de segmentação baseados em limiarização. As Subseções 2.4.2 e 2.4.3 apresentam métodos de segmentação baseados em bordas. As Subseções 2.4.4 e 2.4.5 apresentam métodos de agrupamento, os quais podem ser empregados para segmentar lesões melanocíticas em imagens macroscópicas (mais detalhes na Seção 3.6 e no Capítulo 4). A morfologia matemática é uma ferramenta poderosa para a segmentação de imagens e pode ser particularmente útil para melhorar a qualidade da segmentação (WELFER; SCHARCANSKI; MARINHO, 2011). Assim, a Subseção 2.4.6 apresenta alguns conceitos fundamentais a respeito de morfologia matemática.

2.4.1 Limiar de Otsu

OTSU (1979) propôs um método para a seleção automática do limiar t em histogramas (ótimo para histogramas bimodais). Considere $p(I_{(i,j)} = 0), p(I_{(i,j)} = 1), \dots, p(I_{(i,j)} = N_I - 1)$ as probabilidades de ocorrência das diferentes intensidades presentes em uma dada imagem \mathbf{I} . Assumindo que o histograma de \mathbf{I} é bimodal, o problema consiste em determinar o limiar t que separa as duas classes (i.e., o objeto e o fundo). Observe que para cada limiar t podem ser determinadas as seguintes variâncias intra-classe:

$$\sigma_{\text{fundo}}^2(t) = \frac{\sum_{\ell=0}^t [\ell - \mu_{\text{fundo}}(t)]^2 \mathbb{P}(I_{(i,j)} = \ell)}{Q_{\text{fundo}}(t)}, \quad (2.52)$$

$$\sigma_{\text{objeto}}^2(t) = \frac{\sum_{\ell=t+1}^{N_t-1} [\ell - \mu_{\text{objeto}}(t)]^2 \mathbb{P}(I_{(i,j)} = \ell)}{Q_{\text{objeto}}(t)}, \quad (2.53)$$

onde:

$$\mu_{\text{fundo}} = \frac{\sum_{\ell=0}^t \ell \mathbb{P}(I_{(i,j)} = \ell)}{Q_{\text{fundo}}(t)}, \quad (2.54)$$

$$\mu_{\text{objeto}} = \frac{\sum_{\ell=t+1}^{N_t-1} \ell \mathbb{P}(I_{(i,j)} = \ell)}{Q_{\text{objeto}}(t)}, \quad (2.55)$$

e:

$$Q_{\text{fundo}}(t) = \sum_{\ell=0}^t \mathbb{P}(I_{(i,j)} = \ell), \quad (2.56)$$

$$Q_{\text{objeto}}(t) = \sum_{\ell=t+1}^{N_t-1} \mathbb{P}(I_{(i,j)} = \ell). \quad (2.57)$$

Assim, este método consiste em selecionar o limiar t que minimiza a seguinte combinação linear das variâncias intra-classe:

$$\sigma_{\text{intra}}^2(t) = Q_{\text{fundo}}(t) \sigma_{\text{fundo}}^2(t) + Q_{\text{objeto}}(t) \sigma_{\text{objeto}}^2(t). \quad (2.58)$$

2.4.2 Contornos Ativos e Fluxo do Vetor Gradiente

Considere \mathcal{J} uma imagem em tons de cinza vista como uma função de variáveis contínuas (x, y) . Assim, um contorno ativo pode ser definido como uma curva $\mathbf{x}(s) =$

$[x(s), y(s)]^\top$, $s \in [0, 1]$, projetada para minimizar a seguinte função de energia (KASS; WITKIN; TERZOPOULOS, 1988):

$$\xi = \int_0^1 \frac{1}{2} (\alpha |\mathbf{x}'(s)|^2 + \beta |\mathbf{x}''(s)|^2) + \xi_{\text{ext}}(\mathbf{x}(s)) ds, \quad (2.59)$$

onde α e β são os parâmetros que controlam a tensão e a rigidez do contorno ativo, respectivamente. $\mathbf{x}'(s)$ e $\mathbf{x}''(s)$ denotam as derivadas de primeira e segunda ordem de $\mathbf{x}(s)$ em relação a s . Por fim, ξ_{ext} é a função de energia externa, extraída diretamente a partir da imagem. Típicas funções de energia externa projetadas para levar um contorno ativo em direção as bordas são definidas por (KASS; WITKIN; TERZOPOULOS, 1988):

$$\xi_{\text{ext}} = -|\nabla \mathcal{J}|^2, \quad (2.60)$$

$$\xi_{\text{ext}} = -|\nabla(\mathcal{J} \otimes G_\sigma)|^2, \quad (2.61)$$

onde G_σ é uma função gaussiana bidimensional com desvio padrão σ e ∇ é o operador gradiente. Além disso, um contorno ativo que minimiza a Eq. (2.59) deve satisfazer (KASS; WITKIN; TERZOPOULOS, 1988):

$$\alpha \mathbf{x}''(s) - \beta \mathbf{x}''''(s) - \nabla \xi_{\text{ext}} = \mathbf{0}. \quad (2.62)$$

Para encontrar uma solução para a Eq. (2.62), o contorno ativo $\mathbf{x}(s)$ é tratado dinamicamente como $\mathbf{x}(s, t)$. Considere $\mathbf{x}_t(s, t)$ a derivada parcial de $\mathbf{x}(s, t)$ em relação a t de tal forma que:

$$\mathbf{x}_t(s, t) = \alpha \mathbf{x}''(s) - \beta \mathbf{x}''''(s) - \nabla \xi_{\text{ext}}. \quad (2.63)$$

Assim, quando $\mathbf{x}(s, t)$ estabiliza, o termo $\mathbf{x}_t(s, t)$ desaparece e uma solução para a Eq. (2.62) é encontrada. Se a Eq. (2.63) for discretizada e resolvida iterativamente, isso resulta

em um algoritmo de gradiente descendente, o qual pode ser utilizado para resolver a Eq. (2.59) (KASS; WITKIN; TERZOPOULOS, 1988). Observe que uma vez que o contorno é obtido de forma iterativa, um contorno inicial deve ser fornecido como entrada. Assim, o uso de contornos ativos para segmentar objetos em imagens possui duas limitações principais (XU; PRINCE, 1998):

1. os contornos ativos não alcançam bordas de objetos que estão distantes do contorno inicial;
2. os contornos ativos não alcançam concavidades ou recuos das bordas (e.g., o topo do carácter “U”).

Para evitar estas limitações, XU; PRINCE (1998) definiram o fluxo do vetor gradiente como um campo vetorial $\mathbf{v}(x, y) = [u(x, y), v(x, y)]^T$ para minimizar a seguinte função de energia:

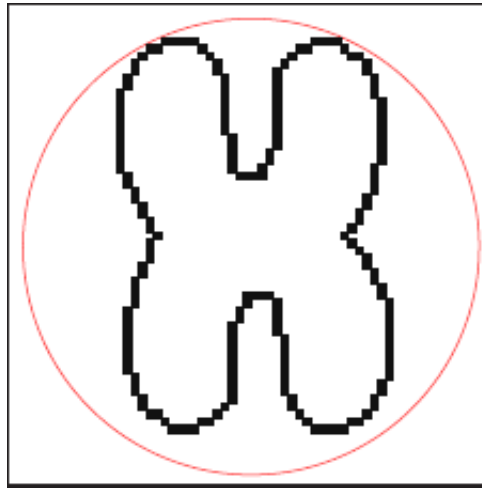
$$\xi_v = \int \int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla \mathbf{f}|^2 |\mathbf{v} - \nabla \mathbf{f}|^2 dx dy, \quad (2.64)$$

onde μ é um parâmetro de regularização que regula o *tradeoff* entre o primeiro e o segundo termo. Assim, substituir \mathbf{v} por $\nabla \xi_{\text{ext}}$ na Eq. (2.63) garante resultados mais adequados. A Figura 2.8 ilustra como o fluxo do vetor gradiente pode melhorar os resultados obtidos por contornos ativos.

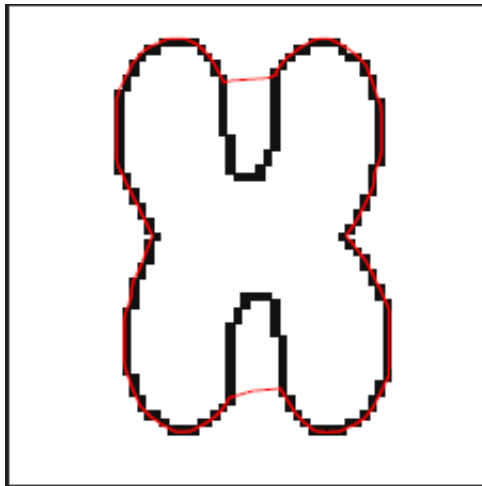
2.4.3 Segmentação Baseada nas Transformadas Wavelet e Watershed

O método proposto em (JUNG, 2007) combina as transformadas wavelet e watershed para segmentar uma imagem colorida. O primeiro passo deste método é a decomposição da imagem através da transformada wavelet. A decomposição de uma imagem \mathbf{I} (no espaço de cores **RGB**) em \mathcal{L} escalas através da transformada wavelet consiste em dois conjuntos de sub-imagens: $\mathcal{W}_{2^l}^{s,c}$ e $\mathcal{A}_{2^l}^c$. Assim, $c \in \{r, g, b\}$ é um índice que indica se uma sub-imagem corresponde a decomposição do canal **R** (i.e., $c = r$), **G** (i.e., $c = g$) ou **B** (i.e., $c = b$), $s \in \{h, v, d\}$ é um índice que indica se uma sub-imagem contém os detalhes horizontais (i.e., $s = h$), verticais (i.e., $s = v$) ou diagonais (i.e., $s = d$) e $l \in \{1, \dots, \mathcal{L}\}$ é um índice de escala de tal forma que $l = \mathcal{L}$ corresponde a resolução

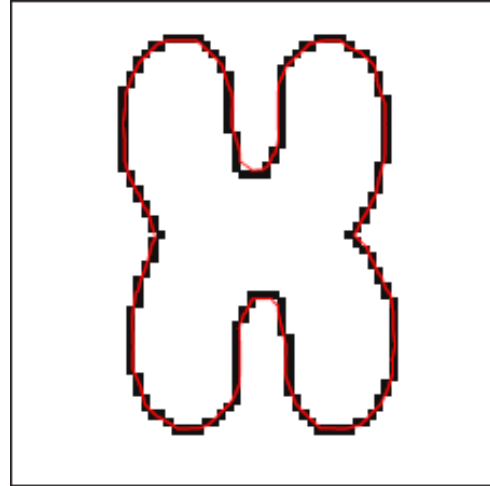
Figura 2.8: Ilustração de resultados obtidos por contornos ativos.



(a) Contorno inicial.



(b) Contorno obtido usando o ξ_{ext} da Eq. (2.60).



(c) Contorno obtido usando o fluxo do vetor gradiente.

Fonte: (AYALA, 2015).

mínima. As sub-imagens $\mathcal{W}_{2^i}^{s,c}$ e $\mathcal{A}_{2^i}^c$ são obtidas através de convoluções com filtros passa-baixas (associados com uma wavelet mãe) e decimações nos canais da imagem \mathbf{I} . Além disso, a transformada wavelet possui uma transformada inversa que permite a reconstrução da imagem original \mathbf{I} a partir das sub-imagens $\mathcal{W}_{2^i}^{s,c}$ e $\mathcal{A}_{2^i}^c$ (MALLAT, 2008). Devido à sua boa localização no espaço-frequência e baixa complexidade computacional, a wavelet de Haar foi escolhida no método proposto em (JUNG, 2007). Após decompor a imagem \mathbf{I} em \mathcal{L} escalas através da transformada wavelet de Haar, a magnitude do gradiente de cada sub-imagem $\mathcal{A}_{2^i}^c$ é computada (i.e., aproximada através o filtro de Prewitt) e denotada

$\mathcal{M}_{2^{\mathcal{L}}}^c$. A probabilidade de um dado pixel com magnitude do gradiente m corresponder a uma borda é computada por:

$$p(\text{borda} | \mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m) = \frac{(1 - w_{\text{ruído}})p_{\text{borda}}(\mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m)}{w_{\text{ruído}}p_{\text{ruído}}(\mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m) + (1 - w_{\text{ruído}})p_{\text{borda}}(\mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m)}, \quad (2.65)$$

onde $p_{\text{ruído}}$ é a distribuição das magnitudes associadas ao ruído, p_{borda} é a distribuição das magnitudes associadas às bordas e $w_{\text{ruído}}$ é a probabilidade *a priori* da distribuição $p_{\text{ruído}}$. As distribuições $p_{\text{ruído}}$ e p_{borda} podem ser estimadas por:

$$p_{\text{ruído}}(\mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m) = \frac{1}{8\sigma_{\text{ruído}}^6} m^5 \exp \left\{ -\frac{m^2}{2\sigma_{\text{ruído}}^2} \right\}, \quad (2.66)$$

$$p_{\text{borda}}(\mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m) = \frac{1}{8\sigma_{\text{borda}}^6} m^5 \exp \left\{ -\frac{m^2}{2\sigma_{\text{borda}}^2} \right\}, \quad (2.67)$$

onde $\sigma_{\text{ruído}}^2$ é a variância das magnitudes correspondentes ao ruído e σ_{borda}^2 é a variância das magnitudes correspondentes às bordas. Observe que os parâmetros da Eq. (2.65) (i.e., $\sigma_{\text{ruído}}$, σ_{borda} e $w_{\text{ruído}}$) podem ser estimados pelo método de máxima verossimilhança. Assim, os pixels com $p(\text{borda} | \mathcal{M}_{2^{\mathcal{L}}}^c(i,j) = m) < t_1$ são setados para “0” e a imagem $\widehat{\mathcal{M}}_{2^{\mathcal{L}}}^c$ ($\widehat{\mathcal{M}}_{2^{\mathcal{L}}}^c$ denota a imagem $\mathcal{M}_{2^{\mathcal{L}}}^c$ limiarizada) pode ser vista como um “relevo topográfico”. A transformada watershed é então usada para “inundadar” este relevo a partir de um mínimo local⁵ e resulta em um conjunto de regiões separadas por bordas. Considere $\mathcal{B}_{2^{\mathcal{L}}}^c$ uma representação de $\mathcal{A}_{2^{\mathcal{L}}}^c$ de tal forma que todos os pixels em uma região identificada pela transformada watershed são setados para a cor média desta região em $\mathcal{A}_{2^{\mathcal{L}}}^c$ e todos os pixels identificados como borda pela transformada watershed são setados para “0”. Assim, os coeficientes da escala $2^{\mathcal{L}}$ são atualizados para:

$$\widetilde{\mathcal{A}}_{2^{\mathcal{L}}}^c(i,j) = \begin{cases} \mathcal{A}_{2^{\mathcal{L}}}^c(i,j), & \text{se } (i,j) \text{ pertence às bordas,} \\ \mathcal{B}_{2^{\mathcal{L}}}^c(i,j), & \text{caso contrário,} \end{cases} \quad (2.68)$$

$$\widetilde{\mathcal{W}}_{2^{\mathcal{L}}}^{s,c}(i,j) = \begin{cases} \mathcal{W}_{2^{\mathcal{L}}}^{s,c}(i,j), & \text{se } (i,j) \text{ pertence às bordas,} \\ 0, & \text{caso contrário.} \end{cases} \quad (2.69)$$

⁵A transformada watershed baseia-se em operações morfológicas (Subseção 2.4.6).

Em seguida, a transformada wavelet inversa é aplicada para $\tilde{\mathcal{A}}_{2^z}^c$ e $\tilde{\mathcal{W}}_{2^z}^{s,c}$, gerando um novo conjunto de imagens $\mathcal{S}_{2^{z-1}}^c$. Observe que cada platô em $\mathcal{B}_{2^z}^c$ é mapeado para um platô em $\mathcal{S}_{2^{z-1}}^c$. Porém, os coeficientes de detalhe atualizados geram flutuações (i.e., pixels difusos) entre platôs adjacentes. Considere um pixel difuso (i, j) e os seus 8 pixels vizinhos. A dissimilaridade de cor ρ_κ entre um pixel difuso e um pixel vizinho pode ser computada através da distância euclideana. Então, (i, j) é atribuído para a região homogênea do vizinho com menor ρ_κ (desconsidere os vizinhos difusos). Depois, uma representação $\mathcal{B}_{2^{z-1}}^c$ é obtida e os coeficientes $\tilde{\mathcal{A}}_{2^{L-1}}^c$ e $\tilde{\mathcal{W}}_{2^{L-1}}^{s,c}$ são obtidos usando as Eqs. (2.68) e (2.69). Este processo é repetido até a resolução máxima ser computada. Por fim, a distância euclidiana no espaço de cores **Lab** (GONZALEZ; WOODS, 2008) é computada entre as cores médias de cada par de regiões adjacentes e os pares de regiões adjacentes com diferenças menores do que um limiar t_2 (enquanto houverem) são fundidos (a cor média da região fundida deve ser recomputada).

2.4.4 K -médias

O algoritmo de K -médias (KANUNGO et al., 2002) tem por objetivo particionar um conjunto com N pontos no \mathbb{R}^M em $K \leq N$ agrupamentos. Considere o problema de particionar a matriz $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$, na qual cada coluna \mathbf{x}_i corresponde a um ponto no \mathbb{R}^M . Além disso, considere a matriz $\mathcal{R} \in \mathbb{R}^{N \times K}$, construída de maneira que $\mathcal{R}_{(i,\kappa)} = 1$ se \mathbf{x}_i é atribuído para o agrupamento κ e $\mathcal{R}_{(i,\kappa')} = 0$ para $\kappa' \neq \kappa$. Assim, a seguinte função objetivo pode ser definida:

$$J = \sum_{i=1}^N \sum_{\kappa=1}^K \mathcal{R}_{(i,\kappa)} \|\mathbf{x}_i - \boldsymbol{\mu}_\kappa\|^2, \quad (2.70)$$

onde $\boldsymbol{\mu}_\kappa$ denota a média dos pontos do agrupamento κ . Esta função representa a soma dos quadrados das distâncias de cada ponto \mathbf{x}_i para a média $\boldsymbol{\mu}_\kappa$ do agrupamento ao qual ele foi atribuído. O objetivo é encontrar a matriz \mathcal{R} e o conjunto de médias $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$ que minimizam J . Isso pode ser feito através de um processo iterativo, o qual executa as seguintes etapas (BISHOP, 2006):

1. selecionar os valores iniciais para as médias $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$;

2. (a) minimizar J em relação a \mathcal{R} , mantendo o conjunto $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$ fixo;
- (b) minimizar J em relação a $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$, mantendo a matriz \mathcal{R} fixa;
3. se pelo menos um ponto mudou de agrupamento e o número máximo de iterações não foi excedido, executar novamente a etapa 2.

Uma vez que a função J é linear em relação a $\mathcal{R}_{(i,\kappa)}$, a obtenção dos valores $\mathcal{R}_{(i,\kappa)}$ que compõem \mathcal{R} e minimizam J para $\boldsymbol{\mu}_\kappa$ fixo possui solução fechada. Assim, a matriz \mathcal{R} é construída de tal maneira que cada ponto \mathbf{x}_i seja atribuído para o agrupamento com a média $\boldsymbol{\mu}_\kappa$ mais próxima de \mathbf{x}_i :

$$\mathcal{R}_{(i,\kappa)} = \begin{cases} 1, & \text{se } \kappa = \arg \min_{\kappa'} \|\mathbf{x}_i - \boldsymbol{\mu}_{\kappa'}\|, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.71)$$

Uma vez que a função J é quadrática em relação a $\boldsymbol{\mu}_\kappa$, a sua minimização para $\mathcal{R}_{(i,\kappa)}$ fixo pode ser obtida igualando a derivada de J (em relação a $\boldsymbol{\mu}_\kappa$) a zero:

$$2 \sum_{i=1}^N \mathcal{R}_{(i,\kappa)} (\mathbf{x}_i - \boldsymbol{\mu}_\kappa) = 0, \quad (2.72)$$

resultando em:

$$\boldsymbol{\mu}_\kappa = \frac{\sum_{i=1}^N \mathcal{R}_{(i,\kappa)} \mathbf{x}_i}{\sum_{i=1}^N \mathcal{R}_{(i,\kappa)}}. \quad (2.73)$$

Uma questão importante é a seleção dos valores iniciais para as médias $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$. Para isso, ARTHUR; VASSILVITSKII (2007) propuseram inicializar as médias $\{\boldsymbol{\mu}_\kappa\}_{\kappa=1}^K$ pelo algoritmo K -médias++, o qual diminui o tempo de execução e melhora a qualidade do algoritmo de K -médias (ARTHUR; VASSILVITSKII, 2007), executando as seguintes etapas:

1. selecionar um ponto aleatório em \mathbf{X} para ser a média $\boldsymbol{\mu}_1$, considerando \mathbf{X} uma variável aleatória uniformemente distribuída;
2. computar as distâncias entre cada ponto \mathbf{x}_i e $\boldsymbol{\mu}_1$:

$$d_i = \|\mathbf{x}_i - \boldsymbol{\mu}_1\|; \quad (2.74)$$

3. selecionar um ponto aleatório em \mathbf{X} para ser a média $\boldsymbol{\mu}_2$, considerando \mathbf{X} uma variável aleatória com a seguinte função densidade de probabilidade:

$$p(\mathbf{x}_i) = \frac{d_i^2}{\sum_{i=1}^N d_i^2}; \quad (2.75)$$

4. se $K > 2$, executar para $\kappa = 2, \dots, K$:

- (a) computar as distâncias entre cada ponto \mathbf{x}_i e a média do agrupamento mais próximo:

$$d_i = \min\{\|\mathbf{x}_i - \boldsymbol{\mu}_1\|, \dots, \|\mathbf{x}_i - \boldsymbol{\mu}_\kappa\|\}; \quad (2.76)$$

- (b) selecionar um ponto aleatório em \mathbf{X} para ser a média $\boldsymbol{\mu}_{\kappa+1}$, considerando \mathbf{X} uma variável aleatória com a seguinte função densidade de probabilidade:

$$p(\mathbf{x}_i) = \frac{d_i^2}{\sum_{i=1}^N d_i^2}. \quad (2.77)$$

A principal limitação do algoritmo de K -médias é o fato de que este algoritmo não pode separar agrupamentos que não são linearmente separáveis no espaço de entrada (TIRAJU; MEHTA, 2008). Uma alternativa bem conhecida para superar este problema é modelar a classificação não-supervisionada (i.e., o agrupamento) como um problema de particionamento em grafo.

2.4.5 Corte Normalizado em Grafos

Considere novamente o problema de particionar $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]$. Considere também a representação de \mathbf{X} por um grafo ponderado e não-direcionado $\Gamma = \{\Lambda, \Xi, \Omega\}$. Neste grafo Γ , $\Lambda = \{\Lambda_1, \Lambda_2, \dots, \Lambda_N\}$ é um conjunto de vértices (ou nodos), onde Λ_i representa \mathbf{x}_i . Adicionalmente, Ξ é o conjunto de arestas, onde uma aresta $(i, j) \in \Xi$ se e somente se Λ_i é adjacente a Λ_j e $\Lambda_i, \Lambda_j \in \Lambda$. A matriz de similaridade (ou afinidade) Ω pondera cada aresta de tal forma que $\Omega_{(i,j)}$ representa a similaridade entre Λ_i e Λ_j (e.g., $\Omega_{(i,j)}$ pode ser a distância euclidiana entre \mathbf{x}_i e \mathbf{x}_j). Este grafo pode ser particionado em dois sub-grafos \mathbf{A} e $\mathbf{B} = \Lambda - \mathbf{A}$, esta partição de grafo é obtida pela remoção (i.e., pelo corte) do conjunto de arestas que conecta \mathbf{A} e \mathbf{B} . Então, o peso de um corte de grafo é dado por:

$$\text{cut}(\mathbf{A}, \mathbf{B}) = \sum_{\Lambda_i \in \mathbf{A}, \Lambda_j \in \mathbf{B}} \Omega_{(i,j)}. \quad (2.78)$$

Um desafio importante é encontrar o corte mínimo de um grafo e ao mesmo tempo evitar o agrupamento de pequenos conjuntos isolados de nodos (SHI; MALIK, 2000). O corte normalizado apresentado a seguir fornece uma alternativa adequada para o corte de grafos:

$$\text{Ncut}(\mathbf{A}, \mathbf{B}) = \frac{\text{cut}(\mathbf{A}, \mathbf{B})}{\text{assoc}(\mathbf{A}, \Lambda)} + \frac{\text{cut}(\mathbf{A}, \mathbf{B})}{\text{assoc}(\mathbf{B}, \Lambda)}, \quad (2.79)$$

onde:

$$\text{assoc}(\mathbf{A}, \Lambda) = \sum_{\Lambda_i \in \mathbf{A}, \Lambda_h \in \Lambda} \Omega_{(i,h)}, \quad (2.80)$$

$$\text{assoc}(\mathbf{B}, \Lambda) = \sum_{\Lambda_j \in \mathbf{B}, \Lambda_h \in \Lambda} \Omega_{(j,h)}. \quad (2.81)$$

SHI; MALIK (2000) mostraram que:

$$\arg \min_{\mathbf{A}, \mathbf{B}} \{\text{Ncut}(\mathbf{A}, \mathbf{B})\} = \arg \min_{\mathbf{U}} \left\{ \frac{\mathbf{U}^\top (\mathbf{D} - \Omega) \mathbf{U}}{\mathbf{U}^\top \mathbf{D} \mathbf{U}} \right\}, \quad (2.82)$$

onde $\mathbf{D} \in \mathbb{R}^{N \times N}$ é a matriz diagonal definida por:

$$\mathcal{D}_{(i,i)} = \sum_j \Omega_{(i,j)}, \quad (2.83)$$

$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]^\top \in \{0, 1\}^{N \times 2}$ é a matriz de rótulos definida por:

$$\mathbf{u}_i = \begin{cases} [1, 0]^\top, & \text{se } \Lambda_i \in \mathbf{A}, \\ [0, 1]^\top, & \text{se } \Lambda_i \in \mathbf{B}. \end{cases} \quad (2.84)$$

O corte normalizado na Eq. (2.82) é o quociente de Rayleigh (GOLUB; VAN LOAN, 2013) e uma vez que \mathbf{U} assume valores reais o problema pode ser resolvido através do seguinte sistema de autovalores generalizado:

$$(\mathcal{D} - \Omega)U = \lambda \mathcal{D}U, \quad (2.85)$$

onde λ denota um autovalor associado a um autovetor de $\mathcal{D}^{\frac{1}{2}}U$ e o autovetor associado com o segundo menor autovalor fornece uma solução aproximadamente ótima para o problema (LIU; ZENG, 2012).

2.4.6 Morfologia Matemática

O princípio básico da morfologia matemática consiste em extrair as informações relativas à geometria e a topologia de um conjunto desconhecido (uma imagem) pela transformação através de outro conjunto completamente definido, denominado elemento estruturante (GONZALEZ; WOODS, 2008). Portanto, a base da morfologia matemática é a teoria de conjuntos. Por exemplo, o conjunto de todos os pixels pretos em uma imagem binária descreve completamente a imagem uma vez que os demais pixels só podem ser brancos (MARQUES FILHO; VIEIRA NETO, 1999). Sendo assim, considere os conjuntos $\mathbf{A}, \mathbf{B} \subseteq \{1, \dots, m\} \times \{1, \dots, n\}$, cujos elementos são $\mathbf{a} = (a_1, a_2)$ e $\mathbf{b} = (b_1, b_2)$, respectivamente. Assim, a translação de \mathbf{A} por $\mathbf{p} = (p_1, p_2)$ é definida como:

$$(\mathbf{A})_{\mathbf{p}} = \{\mathbf{q} | \mathbf{q} = \mathbf{a} + \mathbf{p}\}, \quad (2.86)$$

por sua vez, a reflexão de \mathbf{A} é definida como:

$$\widehat{\mathbf{A}} = \{\mathbf{p} | \mathbf{p} = -\mathbf{a}\}. \quad (2.87)$$

As definições nas Eqs. (2.86) e (2.87) permitem definir a primeira operação morfológica básica, a dilatação, como:

$$\mathbf{A} \oplus \mathbf{B} = \{\mathbf{p} | [(\widehat{\mathbf{B}})_{\mathbf{p}} \cap \mathbf{A}] \subseteq \mathbf{A}\}. \quad (2.88)$$

Portanto, o processo de dilatação consiste em obter a reflexão de \mathbf{B} e depois deslocar por \mathbf{p} , resultando no conjunto de todos os deslocamentos \mathbf{p} para os quais a intersecção de $(\widehat{\mathbf{B}})_{\mathbf{p}}$ e \mathbf{A} inclui pelo menos um elemento. Neste contexto, o conjunto \mathbf{B} é usualmente denominado como elemento estruturante. A segunda operação morfológica básica, a erosão, é definida como:

$$\mathbf{A} \ominus \mathbf{B} = \{\mathbf{p} | (\mathbf{B})_{\mathbf{p}} \subseteq \mathbf{A}\}. \quad (2.89)$$

Portanto, a erosão de \mathbf{A} por \mathbf{B} resulta em um conjunto de pontos \mathbf{p} tais que \mathbf{B} transladado de \mathbf{p} está contido em \mathbf{A} . A operação de dilatação usualmente é aplicada para expandir uma imagem binária enquanto a operação de erosão usualmente é aplicada para encolher uma imagem binária (MARQUES FILHO; VIEIRA NETO, 1999).

Um algoritmo baseado em operações morfológicas de particular interesse é o algoritmo de preenchimento de regiões. Considere um conjunto \mathbf{A} representando uma imagem binária que contém um contorno fechado de 1's. Partindo de um ponto \mathbf{p} situado dentro do contorno, o objetivo deste algoritmo é preencher o interior do contorno (i.e., a região) com 1's. Para isso, o algoritmo resolve iterativamente o seguinte passo (enquanto $\mathfrak{C}^{(t)} \neq \mathfrak{C}^{(t-1)}$) (SOILLE, 2010):

$$\mathfrak{C}^{(t)} = (\mathfrak{C}^{(t-1)} \oplus \mathbf{B}) \cap \overline{\mathbf{A}}, \quad (2.90)$$

onde $\mathfrak{C}^{(t)} = \mathfrak{C}^{(t-1)}$ é a imagem binária com as regiões preenchidas, $\mathfrak{C}_0 = \mathbf{p}$ e \mathbf{B} é o elemento estruturante definido por:

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2.91)$$

3 ESTADO-DA-ARTE

Neste capítulo, serão apresentados métodos de segmentação de lesões melanocíticas em imagens macroscópicas representativos do estado-da-arte.

3.1 Limiar de Otsu na Imagem em Tons de Cinza

Baseado no princípio de que lesões melanocíticas são depigmentações da pele e a fim de reduzir o custo computacional, RUIZ et al. (2008) propuseram um algoritmo de segmentação baseado na imagem em tons de cinza. Assim, o primeiro passo deste algoritmo é converter a imagem de entrada para tons de cinza. Depois, o limiar de Otsu é utilizado. Por fim, operações morfológicas são aplicadas para melhorar a qualidade da segmentação.

3.2 Segmentação Baseada no Fluxo do Vetor Gradiente

TANG (2009) propôs segmentar lesões melanocíticas em imagens macroscópicas usando contornos ativos e fluxo do vetor gradiente. Para isso, TANG (2009) propôs um novo mapa de bordas \mathbf{f} para evitar que o fluxo do vetor gradiente convirja para bordas espúrias da imagem. Assim, baseado em uma segmentação inicial mais “grosseira”, é computado o centro da lesão (i_0, j_0) e também o vetor $\widehat{\mathbf{d}}_{(i,j)} = [\widehat{d}_i, \widehat{d}_j]^\top$, definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$\widehat{d}_i = \frac{i_0 - i}{\sqrt{(i_0 - i)^2 + (j_0 - j)^2}}, \quad (3.1)$$

$$\widehat{d}_j = \frac{j_0 - j}{\sqrt{(i_0 - i)^2 + (j_0 - j)^2}}. \quad (3.2)$$

Em seguida, é selecionado o vetor $\mathbf{d}_{(i,j)}$ mais próximo (de acordo com a distância cosseno) de $\widehat{\mathbf{d}}_{(i,j)}$, para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$, onde $\mathbf{d}_{(i,j)} = [d_i, d_j]^\top$ é um dos seguintes vetores: $[-1, -1]^\top$, $[-1, 0]^\top$, $[-1, 1]^\top$, $[0, -1]^\top$, $[0, 0]^\top$, $[0, 1]^\top$, $[1, -1]^\top$, $[1, 0]^\top$, $[1, 1]^\top$. Assim, o mapa de bordas inicial $\widehat{\mathbf{f}}$ é computado para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$\widehat{f}_{(i,j)} = I_{(i+d_i,j+d_j)} - I_{(i,j)}. \quad (3.3)$$

Por fim, o mapa de bordas \mathbf{f} é computado para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$f_{(i,j)} = \begin{cases} \widehat{f}_{(i,j)}, & \text{se } \widehat{f}_{(i,j)} < 0, \\ 0, & \text{se } \widehat{f}_{(i,j)} \geq 0. \end{cases} \quad (3.4)$$

TANG (2009) sugeriu uma segmentação inicial gerada através de uma suavização da imagem de entrada por difusão anisotrópica (PERONA; MALIK, 1990), seguida por uma limiarização multi-estágio adaptativa (YAN; ZHANG; KUBE, 2005).

3.3 Atenuação de Sombras

Imagens macroscópicas de lesões melanocíticas podem conter áreas sombreadas, as quais podem ser confundidas com lesões de pele pelo algoritmo de segmentação. Para evitar isso, CAVALCANTI; SCHARCANSKI; LOPES (2010) propuseram um algoritmo de atenuação de sombras, o qual tem sido utilizado em diversos métodos estado-da-arte em segmentação de lesões melanocíticas em imagens macroscópicas (CAVALCANTI; YARI; SCHARCANSKI, 2010; CAVALCANTI et al., 2011; CAVALCANTI; SCHARCANSKI, 2011, 2013; CAVALCANTI et al., 2014). O algoritmo de atenuação de sombras proposto em (CAVALCANTI; YARI; SCHARCANSKI, 2010) será apresentado a seguir.

Primeiramente, é importante mencionar que o canal \mathbf{V} definido pela Eq. (2.9) possui um papel fundamental neste algoritmo uma vez que os efeitos de sombra podem ser melhor identificados neste canal. Sendo assim, considere os seguintes conjuntos de índices em \mathbf{V} :

$$\begin{aligned} \Phi_{ul} &= \{(i, j) | 1 \leq i \leq \kappa, 1 \leq j \leq \kappa\}, \\ \Phi_{ur} &= \{(i, j) | 1 \leq i \leq \kappa, n - \kappa + 1 \leq j \leq n\}, \\ \Phi_{l} &= \{(i, j) | m - \kappa + 1 \leq i \leq m, 1 \leq j \leq \kappa\}, \\ \Phi_{lr} &= \{(i, j) | m - \kappa + 1 \leq i \leq m, n - \kappa + 1 \leq j \leq n\}. \end{aligned} \quad (3.5)$$

Estes conjuntos contém os índices da região de dimensão $\kappa \times \kappa$ mais acima e mais à esquerda em $\mathbf{V}(\Phi_{ul})$, os índices da região de dimensão $\kappa \times \kappa$ mais acima e mais à direita em $\mathbf{V}(\Phi_{ur})$, os índices da região de dimensão $\kappa \times \kappa$ mais abaixo e mais à esquerda em $\mathbf{V}(\Phi_{ll})$ e os índices da região de dimensão $\kappa \times \kappa$ mais abaixo e mais à direita em $\mathbf{V}(\Phi_{lr})$. É assumido que pixels de lesão não estão nas regiões mencionadas anteriormente. Assim, estas regiões correspondem às amostras de pixels de não-lesão e são usadas para ajustar o seguinte modelo de iluminação da imagem (CAVALCANTI; SCHARCANSKI; LOPES, 2010):

$$Z_{(i,j)} = P_1 i^2 + P_2 j^2 + P_3 ij + P_4 i + P_5 j + P_6. \quad (3.6)$$

A soma dos quadrados dos resíduos pode ser definida por:

$$\varepsilon = \sum_{(i,j) \in \Phi} (Z_{(i,j)} - V_{(i,j)})^2, \quad (3.7)$$

onde $\Phi = \Phi_{ul} \cup \Phi_{ur} \cup \Phi_{ll} \cup \Phi_{lr}$. O objetivo aqui é encontrar $\{P_\kappa\}_{\kappa=1}^{\kappa=6}$ tais que:

$$\frac{\partial \varepsilon}{\partial P_\kappa} = 0. \quad (3.8)$$

Resolvendo a Eq. (3.8) para $\kappa = 1, 2, \dots, 6$, as seguintes equações lineares são obtidas:

$$\sum_{(i,j) \in \Phi} i^4 P_1 + \sum_{(i,j) \in \Phi} i^2 j^2 P_2 + \sum_{(i,j) \in \Phi} i^3 j P_3 + \sum_{(i,j) \in \Phi} i^3 P_4 + \sum_{(i,j) \in \Phi} i^2 j P_5 + \sum_{(i,j) \in \Phi} i^2 P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)} i^2 \quad (\text{para } \kappa = 1), \quad (3.9)$$

$$\sum_{(i,j) \in \Phi} i^2 j^2 P_1 + \sum_{(i,j) \in \Phi} j^4 P_2 + \sum_{(i,j) \in \Phi} ij^3 P_3 + \sum_{(i,j) \in \Phi} ij^2 P_4 + \sum_{(i,j) \in \Phi} j^3 P_5 + \sum_{(i,j) \in \Phi} j^2 P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)} j^2 \quad (\text{para } \kappa = 2), \quad (3.10)$$

$$\sum_{(i,j) \in \Phi} i^3 j P_1 + \sum_{(i,j) \in \Phi} i j^3 P_2 + \sum_{(i,j) \in \Phi} i^2 j^2 P_3 + \sum_{(i,j) \in \Phi} i^2 j P_4 + \sum_{(i,j) \in \Phi} i j^2 P_5 + \sum_{(i,j) \in \Phi} i j P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)} i j$$

(para $\kappa = 3$), (3.11)

$$\sum_{(i,j) \in \Phi} i^3 P_1 + \sum_{(i,j) \in \Phi} i j^2 P_2 + \sum_{(i,j) \in \Phi} i^2 j P_3 + \sum_{(i,j) \in \Phi} i^2 P_4 + \sum_{(i,j) \in \Phi} i j P_5 + \sum_{(i,j) \in \Phi} i P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)} i$$

(para $\kappa = 4$), (3.12)

$$\sum_{(i,j) \in \Phi} i^2 j P_1 + \sum_{(i,j) \in \Phi} j^3 P_2 + \sum_{(i,j) \in \Phi} i j^2 P_3 + \sum_{(i,j) \in \Phi} i j P_4 + \sum_{(i,j) \in \Phi} j^2 P_5 + \sum_{(i,j) \in \Phi} j P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)} j$$

(para $\kappa = 5$), (3.13)

$$\sum_{(i,j) \in \Phi} i^2 P_1 + \sum_{(i,j) \in \Phi} j^2 P_2 + \sum_{(i,j) \in \Phi} i j P_3 + \sum_{(i,j) \in \Phi} i P_4 + \sum_{(i,j) \in \Phi} j P_5 + \sum_{(i,j) \in \Phi} P_6 = \sum_{(i,j) \in \Phi} V_{(i,j)}$$

(para $\kappa = 6$), (3.14)

As Eqs. (3.9) até (3.14) podem ser rearranjadas como o seguinte sistema de equações lineares na forma matricial:

$$\Theta P = \Psi, \quad (3.15)$$

onde:

$$\Theta = \begin{bmatrix} \sum_{(i,j) \in \Psi} i^4 & \sum_{(i,j) \in \Phi} i^2 j^2 & \sum_{(i,j) \in \Phi} i^3 j & \sum_{(i,j) \in \Phi} i^3 & \sum_{(i,j) \in \Phi} i^2 j & \sum_{(i,j) \in \Phi} i^2 \\ \sum_{(i,j) \in \Phi} i^2 j^2 & \sum_{(i,j) \in \Phi} j^4 & \sum_{(i,j) \in \Phi} i j^3 & \sum_{(i,j) \in \Phi} i j^2 & \sum_{(i,j) \in \Phi} j^3 & \sum_{(i,j) \in \Phi} j^2 \\ \sum_{(i,j) \in \Phi} i^3 j & \sum_{(i,j) \in \Phi} i j^3 & \sum_{(i,j) \in \Phi} i^2 j^2 & \sum_{(i,j) \in \Phi} i^2 j & \sum_{(i,j) \in \Phi} i j^2 & \sum_{(i,j) \in \Phi} i j \\ \sum_{(i,j) \in \Phi} i^3 & \sum_{(i,j) \in \Phi} i j^2 & \sum_{(i,j) \in \Phi} i^2 j & \sum_{(i,j) \in \Phi} i^2 & \sum_{(i,j) \in \Phi} i j & \sum_{(i,j) \in \Phi} i \\ \sum_{(i,j) \in \Phi} i^2 j & \sum_{(i,j) \in \Phi} j^3 & \sum_{(i,j) \in \Phi} i j^2 & \sum_{(i,j) \in \Phi} i j & \sum_{(i,j) \in \Phi} j^2 & \sum_{(i,j) \in \Phi} j \\ \sum_{(i,j) \in \Phi} i^2 & \sum_{(i,j) \in \Phi} j^2 & \sum_{(i,j) \in \Phi} i j & \sum_{(i,j) \in \Phi} i & \sum_{(i,j) \in \Phi} j & \sum_{(i,j) \in \Phi} 1 \end{bmatrix}, \quad (3.16)$$

$$\mathbf{\Psi} = \begin{bmatrix} \sum_{(i,j) \in \Phi} V_{(i,j)} i^2 \\ \sum_{(i,j) \in \Phi} V_{(i,j)} j^2 \\ \sum_{(i,j) \in \Phi} V_{(i,j)} ij \\ \sum_{(i,j) \in \Phi} V_{(i,j)} i \\ \sum_{(i,j) \in \Phi} V_{(i,j)} j \\ \sum_{(i,j) \in \Phi} V_{(i,j)} \end{bmatrix}, \quad (3.17)$$

e $\mathbf{P} = [P_1, \dots, P_6]^\top$. A solução deste sistema linear fornece os valores P_1, \dots, P_6 e é dada por:

$$\mathbf{P} = (\mathbf{\Theta}^\top \mathbf{\Theta})^{-1} \mathbf{\Theta}^\top \mathbf{\Psi}. \quad (3.18)$$

Por fim, o canal \mathbf{V}^* onde os efeitos de sombra são atenuados é dado por:

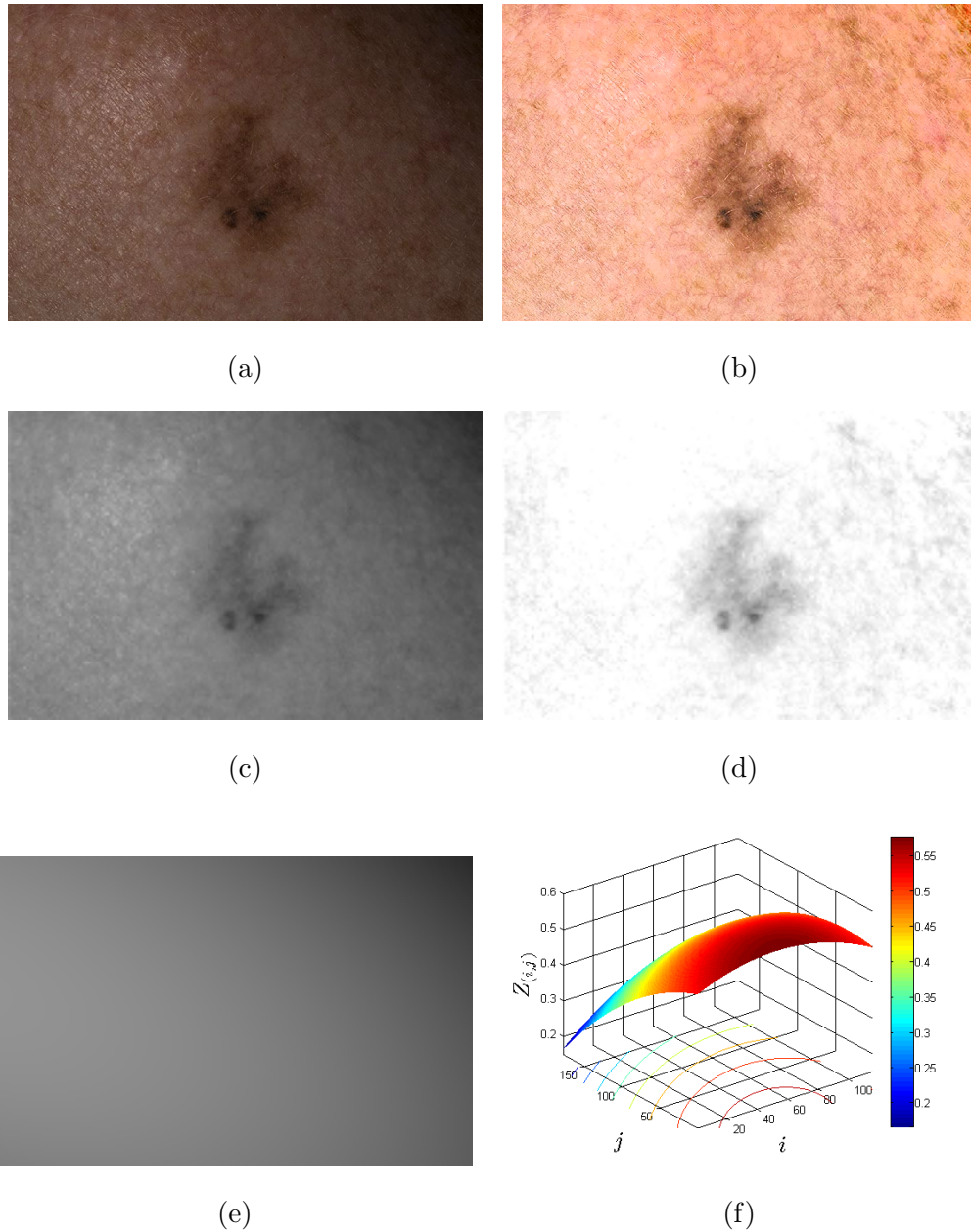
$$V_{(i,j)}^* = \frac{V_{(i,j)}}{Z_{(i,j)}}. \quad (3.19)$$

O algoritmo de atenuação de sombras proposto em (CAVALCANTI; SCHARCANSKI; LOPES, 2010) pode ser resumido como segue:

1. converter a imagem do espaço de cores **RGB** para o espaço de cores **HSV**;
2. computar \mathbf{V}^* e substituir \mathbf{V} por \mathbf{V}^* ;
3. converter de volta a imagem com os efeitos de sombra atenuados do espaço de cores **HSV*** para o espaço de cores **RGB**.

É importante mencionar que todos os algoritmos apresentados daqui em diante utilizam este algoritmo de atenuação de sombras como uma etapa de pré-segmentação.

Figura 3.1: Exemplo de resultado obtido pelo algoritmo de atenuação de sombras: (a) imagem antes da atenuação; (b) imagem depois da atenuação; (c) canal V antes da atenuação; (d) canal V depois da atenuação (i.e., V^*); (e) visualização bidimensional do modelo de iluminação (i.e., Z); (f) visualização tridimensional do modelo de iluminação.



3.4 Representação em Três Canais para Imagens de Lesões Melanocíticas

Imagens macroscópicas de lesões melanocíticas podem conter artefatos tais como pêlos ou sardas, os quais podem ser confundidos com lesões e afetar negativamente o processo de segmentação. Para evitar isso, CAVALCANTI; SCHARCANSKI (2011) propuseram

uma representação em três canais para imagens de lesões melanocíticas. Esta representação ajuda a maximizar a discriminação entre regiões de lesão e regiões de não-lesão e é composta por três canais, os quais serão apresentados a seguir:

1. o primeiro canal, $\mathbf{I}^{(1)}$, é o complemento do canal \mathbf{R} (i.e., o primeiro canal do espaço de cores \mathbf{RGB}) e é definido para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$I_{(i,j)}^{(1)} = 1 - R_{(i,j)}. \quad (3.20)$$

Este canal contém informações úteis porque as regiões de não-lesão usualmente contém intensidades mais elevadas no canal \mathbf{R} do que as regiões de lesão;

2. o segundo canal, $\mathbf{I}^{(2)}$, é uma representação da variação textural. A fim de extrair informações sobre a variação textural, a imagem de luminância, $\mathbf{L} \in \mathbb{R}^{m \times n}$, é computada para todo $(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\}$ por:

$$L_{(i,j)} = \frac{R_{(i,j)} + G_{(i,j)} + B_{(i,j)}}{3}. \quad (3.21)$$

Além disso, η filtros gaussianos $\{\mathbf{S}^{(s)}\}_{s=1}^{\eta}$ de dimensões $2\tau - 1 \times 2\tau - 1$ são computados:

$$\mathcal{S}_{(i,j)}^{(s)} = \frac{1}{2\pi(1 + \frac{\tau s}{2\tau-1})^2} \exp\left(-\frac{(i - \tau)^2 + (j - \tau)^2}{2(1 + \frac{\tau s}{2\tau-1})^2}\right), \quad (3.22)$$

onde $(i, j) \in \{1, \dots, 2\tau - 1\} \times \{1, \dots, 2\tau - 1\}$. Assim, o canal de variação textural $\mathcal{T} \in \mathbb{R}^{m \times n}$ é computado por:

$$\mathcal{T}_{(i,j)} = \max_{s \in \{1, \dots, \eta\}} \left\{ L_{(i,j)} \cdot \left(\frac{1 - (L \otimes S^{(s)})_{(i,j)}}{(L \otimes S^{(s)})_{(i,j)}} \right) \right\}, \quad (3.23)$$

onde $(L \otimes S^{(s)})_{(i,j)}$ denota o pixel indexado por (i, j) em $\mathbf{L} \otimes \mathbf{S}^{(s)}$ (“ \otimes ” denota a operação de convolução). O canal de variação textural é normalizado por:

$$T_{(i,j)} = \frac{\mathcal{T}_{(i,j)} - \min \{\mathcal{T}\}}{\max \{\mathcal{T}\} - \min \{\mathcal{T}\}}, \quad (3.24)$$

onde $\min \{\mathcal{T}\}$ e $\max \{\mathcal{T}\}$ denotam a menor e a maior intensidade em \mathcal{T} . Por fim, considere:

$$\zeta_{(i,j)} = (L \otimes S^{(n)})_{(i,j)}. \quad (3.25)$$

Se uma região da imagem tem predominantemente baixas intensidades, como frequentemente ocorre nas áreas de lesão, então $(1 - \zeta_{(i,j)}) > \zeta_{(i,j)}$ e a razão $(1 - \zeta_{(i,j)})/\zeta_{(i,j)}$ tende a aumentar, enfatizando a região de lesão no canal de variação textural. Se a região tem predominantemente altas intensidades, como frequentemente ocorre nas áreas de não-lesão, então $(1 - \zeta_{(i,j)}) < \zeta_{(i,j)}$ e a razão $(1 - \zeta_{(i,j)})/\zeta_{(i,j)}$ tende a diminuir, desenfazando a região de não-lesão no canal de variação textural (CAVALCANTI; SCHARCANSKI, 2011);

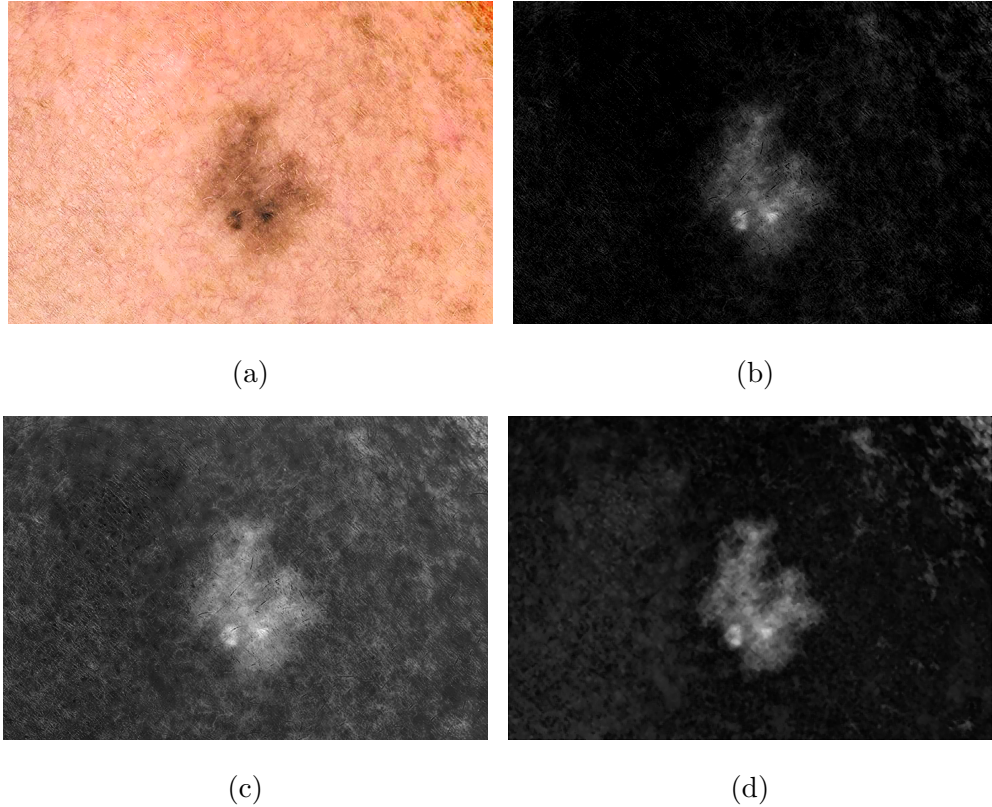
- o terceiro canal, $\mathbf{I}^{(3)}$, descreve a variação local da cor e é dado pela normalização da representação da imagem em **RGB** no seu subespaço de maior variância, o qual é obtido por PCA. Por fim, este canal é filtrado pelo filtro da mediana 5×5 para reduzir o ruído impulsivo do canal. O terceiro canal é útil porque é assumido que regiões de lesão e não-lesão têm diferentes distribuições de cor e esta informação torna-se mais evidente no canal $\mathbf{I}^{(3)}$.

A Figura 3.2 mostra um exemplo desta representação em três canais para uma imagem de lesão melanocítica.

3.5 Segmentação Baseada em ICA

CAVALCANTI et al. (2011) propuseram usar o algoritmo FastICA (com $\mathcal{K} = 3$) para gerar um contorno inicial adequado e em seguida refinar este contorno pelo método de Chan-Vese de contornos ativos (CHAN; SANDBERG; VESE, 2000). Para isso, primeiramente, os canais **R**, **G** e **B** da imagem de entrada **I**, com m linhas e n colunas, são

Figura 3.2: Exemplo de representação em três canais: (a) imagem de uma lesão melanocítica, a qual pode ser representada pelos canais $\mathbf{I}^{(1)}$, $\mathbf{I}^{(2)}$ e $\mathbf{I}^{(3)}$, mostrados respectivamente em (b), (c) e (d).



representados como vetores de dimensões $1 \times mn$ e a matriz $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{3 \times mn}$ é gerada através da concatenação destes vetores. Depois, o conjunto de sinais $\{\mathbf{y}_i\}_{i=1}^N$ é fornecido como entrada para o método FastICA que retorna três sinais fonte $\{\mathbf{d}_j\}_{j=1}^3$ e um conjunto de sinais $\{\mathbf{x}_i\}_{i=1}^N$. CAVALCANTI et al. (2011) assumem que o método FastICA gera um sinal fonte para as regiões de lesão, um sinal fonte para as regiões da pele não-afetadas pela lesão e um sinal fonte para os artefatos indesejáveis da imagem. Para identificar os canais fonte, considere a matriz $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{3 \times N}$. Considere também $\{\mathbf{x}_j\}_{j=1}^3$ denotando as linhas da matriz \mathbf{X} . Assim, a seguinte função pode ser computada pra os vetores $\{\mathbf{x}_j\}_{j=1}^3$:

$$J(\mathbf{x}_j) = H(\mathbf{x}_j) - H(\mathbf{G}_\sigma), \quad (3.26)$$

onde H denota a entropia e \mathbf{G}_σ é uma gaussiana com a variância σ igual a variância de

\mathcal{X}_j . Por fim, o vetor \mathcal{X}_j que maximiza J implica na identificação de \mathbf{d}_j como sinal fonte das regiões de lesão.

3.6 Segmentação Baseada em NMF

CAVALCANTI et al. (2014) propuseram um algoritmo de segmentação baseado em NMF e K -médias. O primeiro passo deste algoritmo é representar os blocos⁶ $w \times w$ dos canais \mathbf{R} , \mathbf{G} e \mathbf{B} da imagem de entrada \mathbf{I} , com m linhas e n colunas, como vetores de dimensões $w^2 \times 1$ e concatenar estes vetores para gerar a matriz $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{3w^2 \times mn}$. Em seguida, o algoritmo de NMF apresentado na Subseção 2.3.4 é usado para obter um dicionário \mathbf{D} e uma representação \mathbf{X} de \mathbf{Y} em \mathbf{D} (i.e., $\mathbf{Y} \approx \mathbf{D}\mathbf{X}$). Depois, as colunas $\{\mathbf{x}_i\}_{i=1}^N$ são fornecidas como entradas para o algoritmo de K -médias (com $K = 2$), que por sua vez, particiona estas entradas de forma que um agrupamento corresponda aos blocos de lesão e o outro agrupamento corresponda aos blocos de não-lesão, resultando assim, em uma segmentação da lesão.

3.7 Outros Métodos

CAVALCANTI; YARI; SCHARCANSKI (2010) propuseram aplicar o limiar de Otsu no canal \mathbf{R} do espaço de cores \mathbf{RGB} , tirando proveito do fato de que regiões de não-lesão tendem a ser mais avermelhadas. CAVALCANTI; SCHARCANSKI (2011) propuseram aplicar o limiar de Otsu nos três canais por eles propostos e classificar cada pixel de acordo com a classificação de pelo menos dois destes canais. CAVALCANTI; SCHARCANSKI (2013) propuseram usar o limiar de Otsu no canal de variação textural proposto em CAVALCANTI; SCHARCANSKI (2011) para gerar um contorno inicial e em seguida refinar este contorno pelo método de Chan-Vese de contornos ativos (CHAN; SANDBERG; VESE, 2000).

⁶O termo em inglês *patch* será designado ao longo do texto por bloco.

4 MÉTODO PROPOSTO

Imagens macroscópicas de lesões melanocíticas podem conter áreas sombreadas, as quais podem ser confundidas com lesões de pele pelo algoritmo de segmentação. Para evitar isso, o método proposto inicia com a aplicação do algoritmo de atenuação de sombras proposto em (CAVALCANTI; SCHARCANSKI; LOPES, 2010) e discutido na Seção 3.3, com o parâmetro $\kappa = 20$.

Para delimitar a região de interesse (i.e., a região da imagem que contém a lesão) e também para reduzir o custo computacional do método, a imagem com os efeitos de sombra atenuados é pré-segmentada pelo algoritmo proposto em (JUNG, 2007) e discutido na Subseção 2.4.3, com os parâmetros $\mathfrak{L} = 5$, $t_1 = \frac{1}{4}$ e $t_2 = \frac{10}{255}$. Em seguida, é selecionado somente o segmento maior e mais central (com base na Eq. (4.13), discutida na Seção 4.2). Por fim, para garantir que a pré-segmentação contém toda a lesão, é aplicada uma dilatação morfológica e selecionado o menor retângulo que envolve a região dilatada. Um disco de raio 50 foi utilizado como elemento estruturante para a dilatação.

O retângulo pré-segmentado é usado para delimitar a região de interesse no canal de variação textural, o qual fornece uma maior discriminabilidade entre regiões afetadas pela lesão e regiões não-afetadas pela lesão. Este canal é gerado conforme proposto em (CAVALCANTI; SCHARCANSKI; LOPES, 2010) e discutido na Seção 3.4, com os parâmetros $\eta = 9$ e $\tau = 4$. A fim de reduzir ainda mais o custo computacional do método, o canal de variação textural é dividido em blocos não-sobrepostos de tamanho $w \times w$, os resultados do método foram avaliados para blocos de tamanho 5×5 , 6×6 , 7×7 , 8×8 e 9×9 .

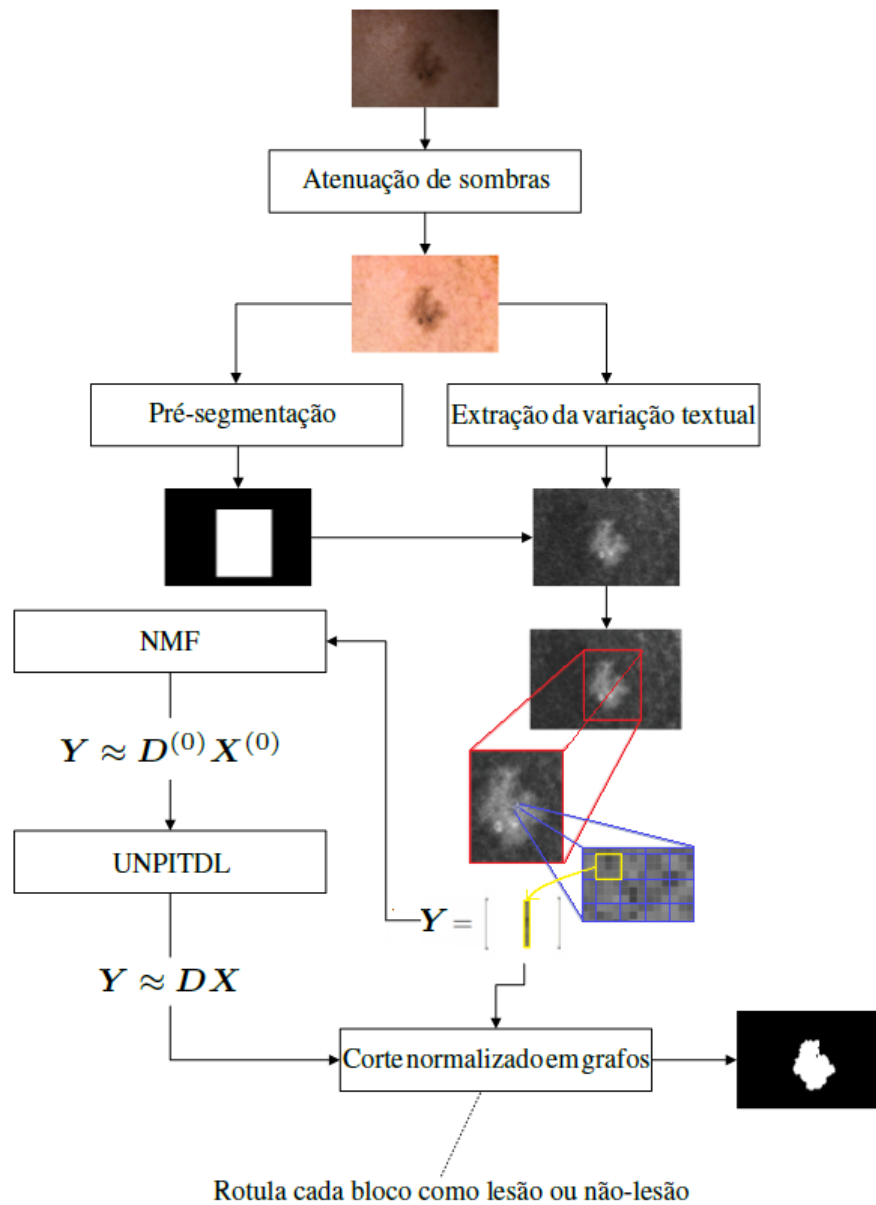
Os blocos são usados para construir um dicionário inicial e uma representação inicial via NMF, usando o algoritmo de mínimos quadrados alternados apresentado na Subseção 2.3.4 (Algoritmo 1), os resultados do método foram avaliados para dicionários iniciais de tamanho $\mathcal{K} = 2$, $\mathcal{K} = 3$, $\mathcal{K} = 4$ e $\mathcal{K} = 5$. Em seguida, o algoritmo UNPITDL proposto na Seção 4.1 é empregado para selecionar o subconjunto de átomos que maximiza simultaneamente a compactividade e a representatividade do dicionário.

Os blocos representados no dicionário compacto e representativo são então representados por um grafo totalmente conectado, não-direcionado e ponderado pela distância euclidiana entre os blocos. Em seguida, o método de corte normalizado em grafos discu-

tido na Subseção 2.4.5 é usado para realizar o particionamento do grafo, resultando em uma segmentação da imagem.

Por fim, são aplicadas algumas operações de pós-processamento na segmentação binária para melhorar a qualidade da segmentação. As etapas de pós-processamento são tratadas na Seção 4.2. A Figura 4.1 apresenta uma visão geral do método proposto para segmentar lesões melanocíticas em imagens macroscópicas.

Figura 4.1: Visão geral do método proposto para segmentar lesões melanocíticas.



4.1 UNPITDL

Dada a matriz dos blocos $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$, onde $\{\mathbf{y}_i\}_{i=1}^N$ são representações dos blocos como vetores colunas. Dado também um dicionário inicial $\mathbf{D}^{(0)}$ e uma representação inicial $\mathbf{X}^{(0)}$ (ambos obtidos a partir de \mathbf{Y} via NMF). Considere o problema de selecionar o subconjunto dos átomos de $\mathbf{D}^{(0)}$ que maximiza simultaneamente a compactividade e a representatividade do dicionário formado pelos seus átomos. Considere também o fato de não ser conhecida *a priori* a classe de nenhum dos blocos nem o número de átomos a ser selecionado (i.e., o aprendizado do dicionário é não-supervisionado e não-paramétrico). Este problema pode ser modelado de tal forma que o objetivo é encontrar um dicionário \mathbf{D} para maximizar a seguinte função (QIU; PATEL; CHELLAPPA, 2014):

$$g = \lambda_1 \text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D}) + \lambda_2 \text{MI}(\mathbf{Y}, \mathbf{D}), \quad (4.1)$$

onde λ_1 e λ_2 são os coeficientes para balancear as contribuições da compactividade $\text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D})$ e da representatividade $\text{MI}(\mathbf{Y}, \mathbf{D})$, os quais são obtidos com base em \mathbf{Y} e $\mathbf{D}^{(0)}$ como segue:

$$\begin{aligned} \lambda_1 &= 1, \\ \lambda_2 &= \max_{j \in \{1, \dots, \mathcal{K}\}} \left\{ \frac{\text{MI}(\mathbf{Y}, \mathbf{d}_j)}{\text{MI}(\mathbf{d}_j, \mathbf{D}^{(0)} - \mathbf{d}_j)} \right\}, \end{aligned} \quad (4.2)$$

aqui, \mathbf{d}_j denota a j -ésima coluna de $\mathbf{D}^{(0)}$ (i.e., o j -ésimo átomo do dicionário) e \mathcal{K} é o tamanho do dicionário inicial $\mathbf{D}^{(0)}$ (i.e., o número de átomos do dicionário). Conforme sugerido em (QIU; PATEL; CHELLAPPA, 2014), a maximização da informação mútua para o aprendizado do dicionário recai na seguinte função objetivo sub-modular⁷:

$$\begin{aligned} \hat{g} &= \lambda_1 [\text{MI}(\mathbf{D} \cup \mathbf{d}_j, \mathbf{D}^{(0)} - (\mathbf{D} \cup \mathbf{d}_j)) - \text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D})] \\ &\quad + \lambda_2 [\text{MI}(\mathbf{Y}, \mathbf{D} \cup \mathbf{d}_j) - \text{MI}(\mathbf{Y}, \mathbf{D})]. \end{aligned} \quad (4.3)$$

⁷Consequentemente, uma abordagem gulosa permite uma solução sub-ótima.

Assim, o algoritmo proposto inicia com $\mathbf{D} = \emptyset$ e para $j = 1, 2, \dots, \mathcal{K}$, seleciona o átomo $\mathbf{d}_j \in \mathbf{D}^{(0)} - \mathbf{D}$ que fornece o maior incremento na função \hat{g} expressa pela Eq. (4.3) (i.e., $\mathbf{D} \leftarrow \mathbf{D} \cup \mathbf{d}_j$). Observe que no fim deste processo, o dicionário \mathbf{D} resulta em uma reordenação de $\mathbf{D}^{(0)}$ de tal forma que as colunas mais a esquerda de \mathbf{D} correspondem aos átomos mais significativos e as colunas mais a direita correspondem aos átomos menos significativos (em termos de compactividade e representatividade). De fato, os dois termos associados a λ_1 e λ_2 na Eq. (4.3) podem ser considerados separadamente como segue:

1. encontrar um átomo \mathbf{d}_j que maximiza $\text{MI}(\mathbf{D} \cup \mathbf{d}_j, \mathbf{D}^{(0)} - (\mathbf{D} \cup \mathbf{d}_j)) - \text{MI}(\mathbf{D}, \mathbf{D}^{(0)} - \mathbf{D})$ equivale a encontrar um átomo \mathbf{d}_j tal que (QIU; JIANG; CHELLAPPA, 2011):

$$\mathbf{d}_j = \arg \max_{\mathbf{d}_{j'} \in \mathbf{D}^{(0)} - \mathbf{D}} \left\{ \frac{[\Sigma_{\mathbf{D}}]_{(j'j')} - \boldsymbol{\sigma}_{\mathbf{D}_{j'}}^T \Sigma_{\mathbf{D}}^{-1} \boldsymbol{\sigma}_{\mathbf{D}_{j'}}}{[\Sigma_{\overline{\mathbf{D}}}]_{(j'j')} - \boldsymbol{\sigma}_{\overline{\mathbf{D}}_{j'}}^T \Sigma_{\overline{\mathbf{D}}}^{-1} \boldsymbol{\sigma}_{\overline{\mathbf{D}}_{j'}}} \right\}, \quad (4.4)$$

onde $\overline{\mathbf{D}}$ denota $\mathbf{D}^{(0)} - (\mathbf{D} \cup \mathbf{d}_{j'})$, $\Sigma_{\mathbf{D}} = [\boldsymbol{\sigma}_{\mathbf{D}_1}, \boldsymbol{\sigma}_{\mathbf{D}_2}, \dots, \boldsymbol{\sigma}_{\mathbf{D}_{|\mathbf{D}|}}]$ denota a matriz de covariância de \mathbf{D} e $\Sigma_{\overline{\mathbf{D}}} = [\boldsymbol{\sigma}_{\overline{\mathbf{D}}_1}, \boldsymbol{\sigma}_{\overline{\mathbf{D}}_2}, \dots, \boldsymbol{\sigma}_{\overline{\mathbf{D}}_{|\overline{\mathbf{D}}|}}]$ denota a matriz de covariância de $\overline{\mathbf{D}}$;

2. assumindo que os blocos são independentes, encontrar um átomo \mathbf{d}_j para maximizar $\text{MI}(\mathbf{Y}, \mathbf{D} \cup \mathbf{d}_j) - \text{MI}(\mathbf{Y}, \mathbf{D})$ equivale a encontrar um átomo \mathbf{d}_j tal que (QIU; PATEL; CHELLAPPA, 2014):

$$\mathbf{d}_j = \arg \max_{\mathbf{d}_{j'} \in \mathbf{D}^{(0)} - \mathbf{D}} \left\{ \sum_{i=1}^N \text{H}(\mathbf{y}_i | \mathbf{D}) - \sum_{i=1}^N \text{H}(\mathbf{y}_i | \mathbf{D} \cup \mathbf{d}_{j'}) \right\}, \quad (4.5)$$

onde $\text{H}(\mathbf{y}_i | \mathbf{D})$ denota a entropia de $p(\mathbf{y}_i | \mathbf{D})$, a qual é definida como:

$$\text{H}(\mathbf{y}_i | \mathbf{D}) = \sum_{i=1}^N p(\mathbf{y}_i | \mathbf{D}) \log \left(\frac{1}{p(\mathbf{y}_i | \mathbf{D})} \right), \quad (4.6)$$

aqui, $p(\mathbf{y}_i | \mathbf{D})$ pode ser estimado por:

$$p(\mathbf{y}_i | \mathbf{D}) = \exp \left(-\frac{\|\mathbf{r}_i\|^2}{2\sigma^2(\mathbf{r}_i)} \right), \quad (4.7)$$

onde $\mathbf{r}_i = \mathbf{y}_i - \mathbf{D}\mathbf{x}_i$ é a diferença entre um bloco \mathbf{y}_i e a sua reconstrução no dicionário $\mathbf{D}\mathbf{x}_i$, \mathbf{x}_i é a representação de \mathbf{y}_i baseada no dicionário \mathbf{D} e $\sigma^2(\mathbf{r}_i)$ é a variância de \mathbf{r}_i . Deve ser observado que a entropia $H(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_j)$ pode ser expressa por:

$$H(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_{j'}) = \sum_{i=1}^N p(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_{j'}) \log \left(\frac{1}{p(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_{j'})} \right), \quad (4.8)$$

aqui, $p(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_i)$ pode ser estimado por:

$$p(\mathbf{y}_i|\mathbf{D} \cup \mathbf{d}_{j'}) = \exp \left(-\frac{\|\mathbf{r}'_i\|^2}{2\sigma^2(\mathbf{r}'_i)} \right), \quad (4.9)$$

onde $\mathbf{r}'_i = \mathbf{y}_i - (\mathbf{D} \cup \mathbf{d}_{j'})\mathbf{x}'_i$ é a diferença entre um bloco \mathbf{y}_i e a sua reconstrução $(\mathbf{D} \cup \mathbf{d}_{j'})\mathbf{x}'_i$, \mathbf{x}'_i é a representação de \mathbf{y}_i baseada no dicionário $(\mathbf{D} \cup \mathbf{d}_{j'})$ e $\sigma^2(\mathbf{r}'_i)$ é a variância de \mathbf{r}'_i .

Assim, a função definida na Eq. (4.3) é calculada iterativamente para cada átomo ainda não incluído em \mathbf{D} e em cada iteração é selecionado para a inclusão em \mathbf{D} o átomo que fornece o maior incremento nesta função. Este átomo é selecionado resolvendo os termos da função separadamente através das Eqs. (4.4) e (4.5). Desta forma, \mathbf{D} torna-se uma reordenação do dicionário inicial $\mathbf{D}^{(0)}$ de tal forma que as colunas mais a esquerda correspondem aos átomos mais significativos e as colunas mais a direita correspondem aos átomos menos significativos.

Por fim, o dicionário \mathbf{D}_k é construído com os k átomos mais significativos em \mathbf{D} (i.e., as k colunas mais a esquerda de \mathbf{D}). Uma representação \mathbf{X} dos blocos \mathbf{Y} é obtida usando o dicionário aprendido \mathbf{D}_k por:

$$\mathbf{X} = \mathbf{D}_k^\dagger \mathbf{Y}, \quad (4.10)$$

onde \mathbf{D}_k^\dagger denota a pseudo-inversa do dicionário \mathbf{D}_k , definida por:

$$\mathbf{D}_k^\dagger = (\mathbf{D}_k^\top \mathbf{D}_k)^{-1} \mathbf{D}_k^\top. \quad (4.11)$$

4.2 Pós-Processamento

Os dois conjuntos resultantes do particionamento de \mathbf{X} , \mathbf{A} e \mathbf{B} , podem ser usados para gerar uma segmentação binária temporária, $\mathbf{M}^{(0)} \in \{0, 1\}^{m \times n}$, de tal forma que $M_{(i,j)}^{(0)} = 0$ se o pixel indexado por (i, j) foi agrupado na partição \mathbf{A} ou $M_{(i,j)} = 1$ se o pixel indexado por (i, j) foi agrupado na partição \mathbf{B} . Para detectar qual destas partições contém os blocos classificados como lesão e qual destas partições contém os blocos classificados como não-lesão, é assumido que a maior partição (i.e., a maioria dos blocos) corresponde a pixels de não-lesão (e.g., $|\mathbf{A}| > |\mathbf{B}|$). Assim, a segmentação binária \mathbf{M} seta os pixels de lesão para “1” e é dada por:

$$M_{(i,j)} = \begin{cases} M_{(i,j)}^{(0)}, & \text{se } |\mathbf{A}| \geq |\mathbf{B}| \text{ (}\mathbf{B} \text{ contém os pixels de lesão)}, \\ 1 - M_{(i,j)}^{(0)}, & \text{se } |\mathbf{A}| < |\mathbf{B}| \text{ (}\mathbf{A} \text{ contém os pixels de lesão)}. \end{cases} \quad (4.12)$$

É também assumido que somente uma lesão está presente na imagem de entrada e que esta lesão está posicionada perto do centro da imagem (i.e., esta lesão não toca nas bordas exteriores na imagem de entrada). Sendo assim, somente a região previamente identificada como lesão e que maximiza a seguinte medida de tamanho e de centralidade é mantida na imagem binária e retida para o futuro processamento:

$$Q(\psi_h) = \sum_{(i,j) \in \psi_h} \exp \left[-\frac{1}{\rho} \cdot \left(\frac{\Delta i^2}{m} + \frac{\Delta j^2}{n} \right) \right], \quad (4.13)$$

onde $\psi_h \in \{\psi_h\}_{h=1}^{h=H}$ denota um conjunto de pixels em uma região previamente identificada como lesão, $\Delta i = i - i_0$, $\Delta j = j - j_0$ e $(i_0, j_0) = (\lfloor m/2 \rfloor, \lfloor n/2 \rfloor)$ denota o pixel central da imagem. Resultados adequados são obtidos usando $\rho = 5$. Por fim, é aplicada uma dilatação morfológica e em seguida o algoritmo de preenchimento de regiões. Um disco com 5 pixels de raio foi utilizado como elemento estruturante para a dilatação.

5 RESULTADOS EXPERIMENTAIS

O desempenho do método foi avaliado na base de imagens proposta em (ALCÓN et al., 2009). Esta base tem sido amplamente utilizada (CAVALCANTI; YARI; SCHARCANSKI, 2010; ALCÓN et al., 2009; CAVALCANTI et al., 2011; CAVALCANTI; SCHARCANSKI, 2011, 2013; CAVALCANTI et al., 2014) e consiste em 152 imagens macroscópicas⁸ de lesões melanocíticas (i.e., 45 nevos atípicos e 107 melanomas malignos) retiradas do DERMNET SKIN DISEASE ATLAS (2015). Esta base é de particular interesse pois dispõe dos 152 *ground truths*, os quais foram criados manualmente e validados por um especialista (CAVALCANTI; SCHARCANSKI, 2013).

Para cada imagem na base, o método proposto gerou uma imagem binária \mathbf{M} (i.e., uma segmentação), esta imagem foi então comparada com o *ground truth* \mathbf{GT} correspondente e as seguintes medidas foram computadas:

- TP : corresponde ao número de pixels de lesão em \mathbf{M} , de acordo com \mathbf{GT} , corretamente indentificados como lesão;
- FP : corresponde ao número de pixels de não-lesão em \mathbf{M} , de acordo com \mathbf{GT} , incorretamente indentificados como lesão;
- TN : corresponde ao número de pixels de não-lesão em \mathbf{M} , de acordo com \mathbf{GT} , corretamente indentificados como não-lesão;
- FN : corresponde ao número de pixels de lesão em \mathbf{M} , de acordo com \mathbf{GT} , incorretamente indentificados como não-lesão.

As medidas definidas anteriormente permitem computar a acurácia de uma dada segmentação. A acurácia de uma segmentação \mathbf{M} indica a probabilidade do método proposto classificar corretamente um pixel de \mathbf{M} de acordo com \mathbf{GT} e é definida por:

$$\text{acurácia} = \frac{TP + TN}{TP + FP + TN + FN}. \quad (5.1)$$

A Tabela 5.1 apresenta a média, a mediana e o desvio padrão da acurácia com base nas 152 segmentações obtidas pelo método proposto variando os parâmetros w (i.e., o tamanho

⁸Estas imagens possuem, na média, 500×710 pixels.

de um bloco quadrado) e \mathcal{K} (i.e., o número de átomos no dicionário inicial aprendido via NMF). Conforme pode ser observado na Tabela 5.1, o método proposto tende a obter uma acurácia elevada, apresentando uma média de aproximadamente 0.97, uma mediana de aproximadamente 0.98 e um desvio padrão de aproximadamente 0.03 para as diferentes escolhas dos parâmetros w e \mathcal{K} observados.

Tabela 5.1: Média, mediana e desvio padrão da acurácia obtida variando os parâmetros w e \mathcal{K} .

Média da acurácia (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	97.27%	97.16%	97.27%	97.08%
$w = 6$	97.23%	97.27%	97.16%	97.20%
$w = 7$	97.26%	97.29%	97.03%	97.17%
$w = 8$	97.26%	97.18%	97.11%	97.17%
$w = 9$	97.28%	97.19%	97.12%	97.10%
Mediana da acurácia (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	98.45%	98.47%	98.41%	98.38%
$w = 6$	98.39%	98.38%	98.45%	98.31%
$w = 7$	98.41%	98.39%	98.40%	98.33%
$w = 8$	98.40%	98.39%	98.35%	98.32%
$w = 9$	98.39%	98.40%	98.42%	98.30%
Desvio padrão da acurácia (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	3.41%	3.65%	3.40%	3.51%
$w = 6$	3.35%	3.42%	3.52%	3.37%
$w = 7$	3.37%	3.39%	3.67%	3.35%
$w = 8$	3.30%	3.43%	3.46%	3.43%
$w = 9$	3.29%	3.41%	3.48%	3.42%

Uma alta acurácia nos resultados de segmentação é desejável. Porém, uma baixa taxa FP é ainda mais desejável uma vez que o contrário implica em pixels de não-lesão sendo identificados como lesão e impossibilitando a classificação da lesão. A especificidade, definida a seguir, aumenta a medida que FP diminui e é usualmente mais adequada para

avaliar o resultado de uma segmentação do que a acurácia:

$$\text{especificidade} = \frac{TN}{FP + TN}. \quad (5.2)$$

A Tabela 5.2 mostra a média, a mediana e o desvio padrão da especificidade com base nas 152 segmentações obtidas pelo método proposto variando os parâmetros w e \mathcal{K} . Conforme pode ser observado na Tabela 5.2, o método proposto tende a obter uma especificidade elevada, apresentando uma média e uma mediana de aproximadamente 0.99, com um desvio padrão de aproximadamente 0.01 para as diferentes escolhas dos parâmetros w e \mathcal{K} observados.

A Tabela 5.3 compara o método proposto e outros métodos representativos do estado-da-arte (usando a mesma base de imagens) em termos de especificidade.

Considere o erro XOR, definido por:

$$\text{erro XOR} = \frac{FP + FN}{TP + FN}. \quad (5.3)$$

Observe que se o número de pixels incorretamente segmentados (i.e., $FP + FN$) é maior do que o número de pixels de lesão de acordo com **GT** (i.e., $TP + FP$), o erro XOR é maior do que 100%. Assim, o erro XOR tende a penalizar de forma mais rigorosa segmentações indesejáveis do que a acurácia e do que a especificidade. O erro XOR tem sido preferido para avaliar os resultados em diversos métodos de segmentação de lesões melanocíticas em imagens macroscópicas (SCHARCANSKI; CELEBI, 2013; CAVALCANTI; YARI; SCHARCANSKI, 2010; CAVALCANTI et al., 2011; CAVALCANTI; SCHARCANSKI, 2011; CAVALCANTI et al., 2014).

A Tabela 5.4 mostra a média, a mediana e o desvio padrão do erro XOR com base nos 152 resultados de segmentação obtidos com o método proposto variando os parâmetros w e \mathcal{K} .

A Tabela 5.5 compara o método proposto com outros métodos representativos do estado-da-arte (usando a mesma base de dados) em termos do erro XOR. Conforme pode ser observado, o método proposto com $w = 7$ e $\mathcal{K} = 3$ obteve um erro XOR menor do que 0.15 na média, menor do que 0.12 na mediana e um desvio padrão menor do que 0.11. É importante mencionar que estes resultados seriam maiores do que 0.16 na

Tabela 5.2: Média, mediana e desvio padrão da especificidade obtida variando os parâmetros w e \mathcal{K} .

Média da especificidade (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	99.05%	99.04%	98.98%	99.16%
$w = 6$	98.95%	99.10%	99.14%	98.95%
$w = 7$	98.93%	99.12%	99.13%	99.03%
$w = 8$	98.94%	99.07%	99.06%	99.04%
$w = 9$	98.98%	99.12%	99.22%	99.07%
Mediana da especificidade (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	99.25%	99.36%	99.25%	99.49%
$w = 6$	99.23%	99.36%	99.32%	99.29%
$w = 7$	99.18%	99.33%	99.39%	99.23%
$w = 8$	99.22%	99.31%	99.33%	99.25%
$w = 9$	99.15%	99.36%	99.44%	99.29%
Desvio padrão da especificidade (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	1.30%	1.80%	1.61%	1.38%
$w = 6$	1.35%	1.62%	1.38%	1.39%
$w = 7$	1.52%	1.41%	1.22%	1.11%
$w = 8$	1.20%	1.35%	1.45%	1.25%
$w = 9$	1.07%	1.28%	1.18%	1.29%

média, maiores do que 0.12 na mediana e com um desvio padrão maior do que 0.12 se o esquema UNPITDL não tivesse sido aplicado. Com base na Tabela 5.5, pode-se concluir que o método proposto com $w = 7$ e $\mathcal{K} = 3$ e o método proposto em (CAVALCANTI; SCHARCANSKI, 2013) tendem a produzir menos resultados indesejáveis do que os demais métodos analisados.

Por fim, a Figura 5.1 mostra uma comparação visual entre os resultados de segmentação obtidos pelo método proposto e os resultados de segmentação obtidos por outros dois métodos representativos do estado-da-arte. Estes resultados ilustram situações onde

Tabela 5.3: Comparação entre o método proposto e outros métodos representativos do estado-da-arte em termos de especificidade.

Método	Especificidade (%)		
	Média	Mediana	Desvio Padrão
Limiar de Otsu na imagem em tons de cinza (RUIZ et al., 2008)	97.95	99.20	5.33
Limiar de Otsu no canal R (CAVALCANTI; YARI; SCHARCANSKI, 2010)	98.49	99.47	4.70
Segmentação baseada no Fluxo do Vetor Gradiente (TANG, 2009)	99.62	99.89	0.76
Segmentação baseada em ICA (CAVALCANTI et al., 2011)	98.34	99.59	8.38
Limiar de Otsu na representação em três canais (CAVALCANTI; SCHARCANSKI, 2011)	99.72	99.98	2.08
Limiar de Otsu no canal de variação textural (CAVALCANTI; SCHARCANSKI, 2013)	99.11	99.41	2.16
Segmentação baseada em NMF (CAVALCANTI et al., 2014)	99.51	99.80	1.79
Método proposto ($w = 7, \mathcal{K} = 3$)	99.12	99.33	1.41

os outros dois métodos não segmentam a lesão melanocítica corretamente (e.g., identificando pelo como borda da lesão) e o método proposto apresenta resultados satisfatórios visualmente e em termos das métricas de acurácia, especificidade e do erro XOR.

Uma avaliação preliminar do tempo de execução⁹ mostrou que o algoritmo proposto com $w = 7$ e $\mathcal{K} = 3$ segmentou cada imagem, na média, em 2.4 segundos. Por outro lado, o método proposto em (CAVALCANTI; SCHARCANSKI, 2013) segmentou cada imagem, na média, em 79.0 segundos.

⁹Implementações em MATLAB em um computador com uma CPU Intel Core i7-4770 (3.4GHz, 8-GB RAM).

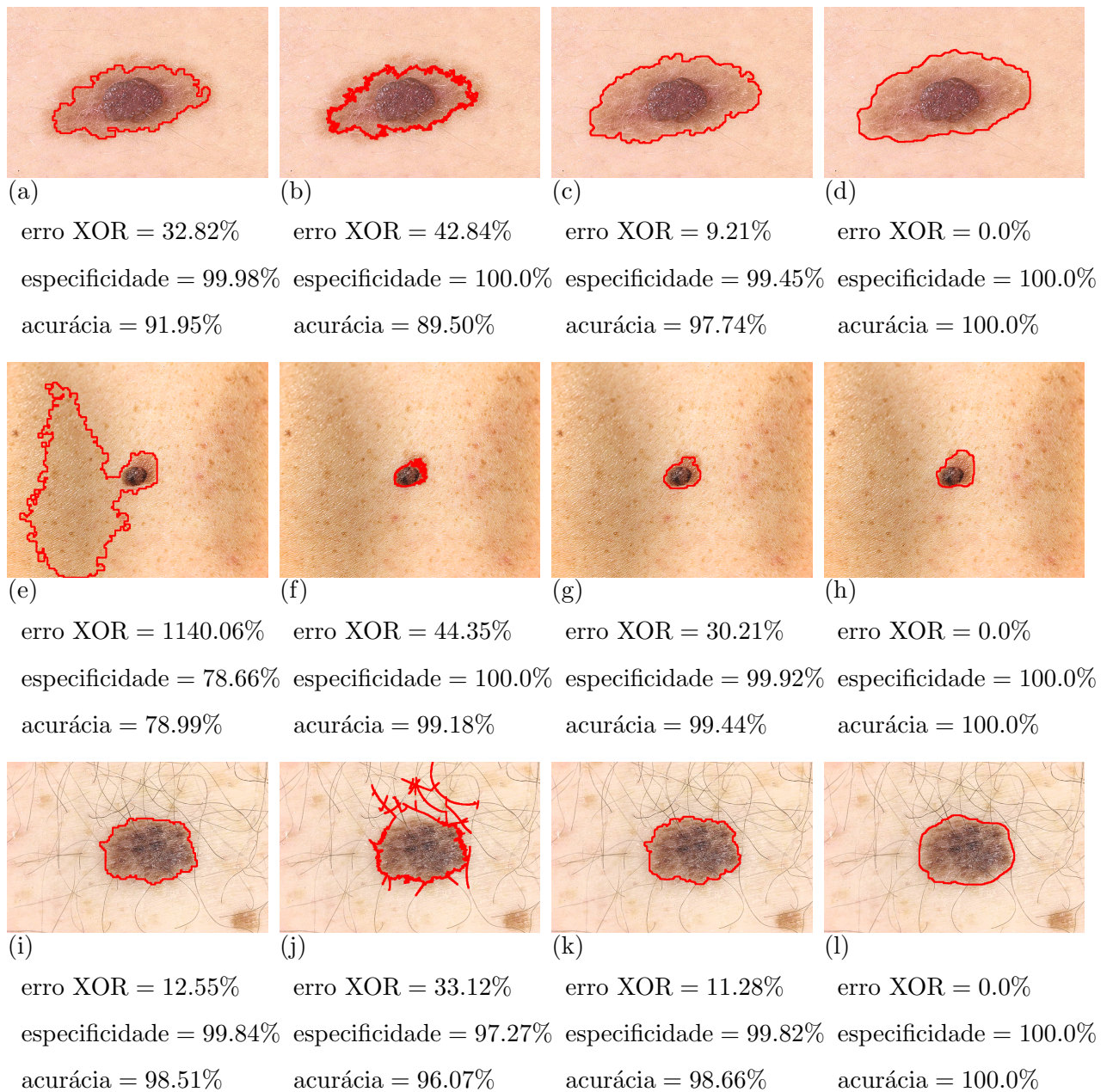
Tabela 5.4: Média, mediana e desvio padrão do erro XOR obtido variando os parâmetros w e \mathcal{K} .

Média do erro XOR (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	15.05%	15.40%	14.82%	15.88%
$w = 6$	15.51%	14.94%	15.65%	15.37%
$w = 7$	15.45%	14.87%	16.23%	15.68%
$w = 8$	15.53%	15.31%	15.88%	15.68%
$w = 9$	15.06%	15.45%	16.13%	16.21%
Mediana do erro XOR (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	11.96%	11.05%	11.96%	12.01%
$w = 6$	12.07%	11.42%	11.40%	11.59%
$w = 7$	11.41%	11.86%	12.26%	12.32%
$w = 8$	11.75%	11.94%	11.90%	12.32%
$w = 9$	11.84%	11.95%	12.36%	12.55%
Desvio padrão do erro XOR (%)				
	$\mathcal{K} = 2$	$\mathcal{K} = 3$	$\mathcal{K} = 4$	$\mathcal{K} = 5$
$w = 5$	11.82%	12.69%	10.68%	11.12%
$w = 6$	11.82%	11.34%	12.42%	10.56%
$w = 7$	12.18%	10.58%	13.02%	11.96%
$w = 8$	11.68%	10.37%	12.22%	11.68%
$w = 9$	9.89%	10.93%	12.67%	11.95%

Tabela 5.5: Comparação entre o método proposto e outros métodos representativos do estado-da-arte em termos do erro XOR.

Método	Erro XOR (%)		
	Média	Mediana	Desvio Padrão
Limiar de Otsu na imagem em tons de cinza (RUIZ et al., 2008)	42.33	13.13	192.15
Limiar de Otsu no canal R (CAVALCANTI; YARI; SCHARCANSKI, 2010)	38.63	15.01	160.49
Esquema baseado no Fluxo do Vetor Gradiente (TANG, 2009)	59.60	50.32	40.96
Segmentação baseada em ICA (CAVALCANTI et al., 2011)	28.33	12.88	108.00
Limiar de Otsu na representação em três canais (CAVALCANTI; SCHARCANSKI, 2011)	34.83	21.90	107.37
Limiar de Otsu no canal de variação textural (CAVALCANTI; SCHARCANSKI, 2013)	15.60	12.33	11.93
Segmentação baseada em NMF (CAVALCANTI et al., 2014)	25.99	14.97	91.69
Método proposto ($w = 7, \mathcal{K} = 3$)	14.87	11.86	10.58

Figura 5.1: Comparação visual entre algoritmos de segmentação. A primeira coluna apresenta três resultados de segmentação obtidos pelo método proposto em (CAVALCANTI et al., 2014). A segunda coluna mostra três resultados de segmentação obtidos (para as mesmas imagens) pelo método proposto em (CAVALCANTI; SCHARCANSKI, 2013). A terceira coluna mostra os resultados obtidos pelo método proposto (com $w = 7$ e $\mathcal{K} = 3$). A quarta coluna mostra os *ground truths*.



6 CONCLUSÃO

Neste trabalho, um novo método para a segmentação de lesões melanocíticas em imagens macroscópicas é proposto.

O método proposto inicia com a aplicação de um algoritmo de atenuação de sombras na imagem de entrada para reduzir a influência dos efeitos de sombra, os quais são frequentemente encontrados em imagens de camera padrão de lesões melanocíticas. Em seguida, um algoritmo de segmentação baseado na transformada wavelet e na transformada watershed foi utilizado para gerar uma pré-segmentação da lesão melanocítica. Depois, um canal de variação textural projetado para destacar a lesão perante o fundo (i.e., para maximizar a discriminabilidade entre regiões de lesão e não-lesão) e para reduzir a influência de artefatos indesejáveis (e.g., pêlos e sardas) é obtido e a região deste canal previamente segmentada é particionada em blocos. A segmentação final é então obtida através de um esquema que combina a aprendizagem de dicionário com o algoritmo de corte normalizado em grafos.

Para aprender um dicionário adequado para representar lesões melanocíticas, foi proposto o método UNPITDL, o qual tem por objetivo maximizar simultaneamente a compactividade e a representatividade do dicionário. Este método é inicializado por um dicionário aprendido via NMF para representar os blocos da região pré-segmentada. Por fim, o algoritmo de corte normalizado em grafos e a representação no dicionário aprendido são aplicados para particionar os blocos, gerando uma segmentação final para a lesão melanocítica.

Uma série de experimentos realizados em uma base de imagens popular sugere que o método proposto neste trabalho para segmentar lesões melanocíticas em imagens macroscópicas pode ser potencialmente mais robusto do que outros métodos comparáveis e representativos do estado-da-arte (em termos do erro XOR).

6.1 Trabalhos Futuros

Na próxima etapa deste trabalho, pretende-se realizar uma extensa avaliação clínica do método proposto e integrá-lo a um sistema completo de pré-triagem de lesões melanocíticas com base em imagens macroscópicas.

REFERÊNCIAS

- AHRON, M; ELAD, M; BRUCKSTEIN, A.. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. **IEEE Transactions on Signal Processing**. Piscataway, p. 4311-4322. 2006.
- ALCÓN, J. F. et al.. Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. **IEEE Journal of Selected Topics in Signal Processing**. Piscataway, p. 14-25. 2009.
- AMERICAN CANCER SOCIETY. **Melanoma skin cancer: what are the survival rates for melanoma skin cancer by stage?**. Disponível em: <<http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-survival-rates>>. Acesso em: 13 Feb. 2015.
- ARTHUR, D.; VASSILVITSKII, S.. K-means++: the advantages of careful seeding. In: ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, 18., 2007, New Orleans. **Proceedings...** . New York: ACM, 2007. p. 1027 - 1035.
- AYALA, A. C.. **Gradient vector flow snakes**. Disponível em: <http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV1011/cartas.pdf>. Acesso em: 13 Feb. 2015.
- BACH, F. et al. Optimization with sparsity-inducing penalties. **Foundations and Trends in Machine Learning**. Hanover, p. 1-106. 2012.
- BERRY, M. W. et al. Algorithms and applications for approximate nonnegative matrix factorization. **Computational Statistics and Data Analysis**. Amsterdam, p. 155-173. 2007.
- BISHOP, C. M.. **Pattern Recognition and Machine Learning**. Heidelberg: Springer, 2006.
- CAVALCANTI, P. G. et al. An ICA-based method for the segmentation of pigmented skin lesions in macroscopic images. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 33., 2011, Boston. **Proceedings...** . Piscataway: IEEE, 2011. p. 5993 - 5996.
- CAVALCANTI, P. G. et al. Segmentation of pigmented skin lesions using nonnegative matrix factorization. In: IEEE INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 29., 2014, Montevideo. **Proceedings...** . Piscataway: IEEE, 2014. p. 72-75.
- CAVALCANTI, P. G.; SCHARCANSKI, J.. Automated prescreening of pigmented skin lesions using standard cameras. **Computerized Medical Imaging and Graphics**. Oxford, U. K., p. 481-491. 2011.
- CAVALCANTI, P. G.; SCHARCANSKI, J.. A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images. **Computer Methods and Programs in Biomedicine**. Maryland Heights, p. 684-693. 2013.
- CAVALCANTI, P. G.; SCHARCANSKI, J.; BARANOSKI, G. V.. A two-stage approach for discriminating melanocytic skin lesions using standard cameras. **Expert Systems with Applications**. Oxford, U. K., p. 4054-4064. 2013.

- CAVALCANTI, P. G.; SCHARCANSKI, J.; LOPES, C. B. O.. Shading attenuation in human skin color images. In: INTERNATIONAL SYMPOSIUM ON VISUAL COMPUTING, 6., 2010, Las Vegas. **Proceedings...** . Heidelberg: Springer, 2010. p. 190 - 198.
- CAVALCANTI, P. G.; YARI, Y.; SCHARCANSKI, J.. Pigmented skin lesion segmentation on macroscopic images. In: INTERNATIONAL CONFERENCE OF IMAGE AND VISION COMPUTING NEW ZELAND, 25., 2010, Queenstown. **Proceedings...** . Piscataway: IEEE, 2010. p. 1 - 7.
- CHAN, T. F.; SANDBERG, B. Y.; VESE, L. A.. Active contours without edges for vector-valued images. **Journal of Visual Communication and Image Representation**. Maryland Heights, p. 130-141. 2000.
- DERMNET SKIN DISEASE ATLAS. **Melanoma (skin cancer), nevi, and moles**. Disponível em: <<http://www.dermnet.com/images/Melanoma-Skin-Cancer-Nevi-and-Moles>>. Acesso em: 13 Feb. 2015.
- EARLEY, J.. **Color theory basics**: Understanding color & avoid some common mistakes. Disponível em: <<http://jonteaches.com/101/color-theory-basics>>. Acesso em: 13 Feb. 2015.
- ELAD, M.. **Sparse and Redundant Representations**: from theory to applications in signal and image processing. Heidelberg: Springer, 2010.
- ENGASSER, H. C.; WARSHAW, E. M.. Dermatoscopy use by US dermatologists: a cross-sectional survey. **Journal of the American Academy of Dermatology**. Schaumburg, p. 412-419. 2010.
- GOLUB, G. H.; VAN LOAN, C. F.. **Matrix Computations**. 4. ed. Baltimore: JHU Press, 2013.
- GONZALEZ, R. C.; WOODS, R. E.. **Digital Image Processing**. 3. ed. Upper Saddle River: Prentice Hall, 2008.
- HYVÄRINEN, A.. Fast and robust fixed-point algorithms for independent component analysis. **IEEE Transactions on Neural Networks**. Piscataway, p. 626-634. 1999.
- HYVÄRINEN, A.; OJA, E.. Independent component analysis. **Neural Networks**. Oxford, U. K., p. 411-430. 2000.
- JAIN, A. K.. **Fundamentals of Digital Image Processing**. Englewood Cliffs: Prentice Hall, 1989.
- JOLLIFFE, I.. **Principal Component Analysis**. New York: John Wiley & Sons, 2002.
- JUNG, C. R.. Unsupervised multiscale segmentation of color images. **Pattern Recognition Letters**. Oxford, U. K., p. 523-533. 2007.
- KANUNGO, T. et al. An efficient K-means clustering algorithm: analysis and implementation. **IEEE Transactions on Pattern Recognition and Machine Intelligence**. Piscataway, p. 881-892. 2002.
- KASS, M.; WITKIN, A.; TERZOPOULOS, D.. Snakes: active contour models. **International Journal of Computer Vision**. Amsterdam, p. 321-331. 1988.

LI, Y.; NGOM, A.. Sparse representation approaches for the classification of high-dimensional biological data. **BMC Systems Biology**. London, p. S6. 2013.

LIU, L.; ZENG, F.. Digital mammogram segmentation based on normalized cuts. In: IEEE GLOBAL HIGH TECH CONGRESS ON ELETRONICS, 1., 2012, Shenzhen. **Proceedings...** . Piscataway: IEEE, 2012. p. 127 - 130.

MAIRAL, J.; BACH, F.; PONCE, J.. Task-driven dictionary learning. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. Piscataway, p. 791-804. 2012.

MAIRAL, J. et al. Discriminative learned dictionaries for local image analysis. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 26., 2008, Anchorage. **Proceedings...** . Piscataway: IEEE, 2008. p. 1 - 8.

MALLAT, S.. **A Wavelet Tour of Signal Processing: the sparse way**. 3. ed. Burlington: Academic Press, 2008.

MARQUES FILHO, O.; VIEIRA NETO, H.. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999.

MATHWORKS. **Nonnegative matrix factorization**. Disponível em: <<http://www.mathworks.com/help/stats/nonnegative-matrix-factorization-1.html>>. Acesso em: 13 Feb. 2015.

MELANOMA RESEARCH FOUNDATION. **What is melanoma?** Disponível em: <<http://www.melanoma.org/understand-melanoma/what-is-melanoma>>. Acesso em: 13 Feb. 2015.

MOESLUND, T. B.. **Introduction to Video and Image Processing: build real systems and applications**. Heidelberg: Springer, 2012.

OTSU, N. A.. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man and Cybernetics**. Piscataway, p. 629-639. 1979.

PATEL, V. M.; CHELLAPPA, R.. **Sparse Representation and Compressive Sensing**. Heidelberg: Springer, 2013.

PERONA, P.; MALIK, J.. Scale-space and edge detection using anisotropic diffusion. **IEEE Transactions on Pattern Recognition and Machine Intelligence**. Piscataway, p. 629-639. 1990.

PRATT, W. K.. **Digital Image Processing**. New York: John Wiley & Sons, 2007.

QIU, Q.; JIANG, Z.; CHELLAPPA, R.. Sparse dictionary-based representation and recognition of action attributes. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 15., 2011, Barcelona. **Proceedings...** . Piscataway: IEEE, 2011. p. 707 - 714.

QIU, Q.; PATEL, V.; CHELLAPPA, R.. Information-theoretic dictionary learning. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. Piscataway, p. 2173-2184. 2014.

RODRIGUEZ, F.; SAPIRO, G.. **Sparse representations for image classification: learning discriminative and reconstructive non-parametric dictionaries**. Minneapolis: DTIC Document, 2008.

RUBISTEIN, R.; BRUCKSTEIN, A. M.; ELAD, M.. Dictionaries for sparse representation modeling. **Proceedings of the IEEE**. Piscataway, p. 1045-1057. 2010.

RUIZ, D. et al. A cooperative approach for the diagnosis of the melanoma. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 30., 2008, Vancouver. **Proceedings...** . Piscataway: IEEE, 2008. p. 5144 - 5147.

RUSS, J. C.. **The Image Processing Handbook**. 6. ed. Boca Raton: CRC Press, 2011.

SCHARCANSKI, J.; CELEBI, M. E.. **Computer Vision Techniques for the Diagnosis of Skin Cancer**. Heidelberg: Springer, 2013.

SHI, J.; MALIK, J.. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. Piscataway, p. 888-905. 2000.

SOILLE, P.. **Morphological Image Analysis: principles and applications**. 2. ed. Heidelberg: Springer, 2010.

SONKA, M.; HLAVAC, V.; BOYLE, R.. **Image Processing, Analysis, and Machine Vision**. 4. ed. Stamford: Cengage Learning, 2014.

SZELISKI, R.. **Computer Vision: algorithms and applications**. Heidelberg: Springer, 2010.

TANG, J.. Multi-direction GVF snake for the segmentation of skin cancer images. **Pattern Recognition**. Oxford, U. K., p. 1172-1179. 2009.

TATIRAJU, S.; MEHTA, A.. **Image segmentation using K-means clustering, EM and Normalized Cuts**. Irvine: Department of EECS, 2008.

TROPP, J. A.; GILBERT, A. C.. Algorithms for simultaneous sparse approximation. Part I: greedy pursuit. **Signal Processing**. Amsterdam, p. 572-588. 2006.

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO SCHOOL OF MEDICINE. **Melanoma**. Disponível em: <<http://dermatology.ucsf.edu/skincancer/general/types/melanoma.aspx>>. Acesso em: 13 Feb. 2015.

WELFER, D.; SCHARCANSKI, J.; MARINHO, D.. Fovea center detection based on the retina anatomy and mathematical morphology. **Computer Methods and Programs in Biomedicine**. Maryland Heights, p. 397-409. 2011.

WONG, A.; SCHARCANSKI, J.; FIEGUTH, P.. Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Transactions on Information Technology in Biomedicine*. Piscataway, p. 929-936. 2011.

WRIGHT, J. et al. Sparse representation for computer vision and pattern recognition. **Proceedings of the IEEE**. Piscataway, p. 1031-1044. 2010.

XU, C.; PRINCE, J. L.. Snakes, shapes, and gradient vector flow. **IEEE Transactions on Image Processing**. Piscataway, p. 359-369. 1998.

YAN, F.; ZHANG, H.; KUBE, C. R.. A multistage adaptive thresholding method. **Pattern Recognition Letters**. Oxford, U. K., p. 1183-1191. 2005.

YANG, M.; ZHANG, D.; FENG, X.. Fisher discrimination dictionary learning for sparse representation. In: *IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION*, 15., 2011, Barcelona. **Proceedings...** . Piscataway: IEEE, 2011. p. 543 - 550.