



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA DE
TELEINFORMÁTICA

GERALDO LUIS BEZERRA RAMALHO

**ANÁLISE DE IMAGENS POR MEIO DA MATRIZ DE
INTERDEPENDÊNCIA E DA TRANSFORMAÇÃO ESTRUTURAL
MULTIESCALA**

FORTALEZA
2013

GERALDO LUIS BEZERRA RAMALHO

ANÁLISE DE IMAGENS POR MEIO DA MATRIZ DE
INTERDEPENDÊNCIA E DA TRANSFORMAÇÃO ESTRUTURAL
MULTIESCALA

Tese apresentada à Coordenação do Curso de Pós-Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará como requisito parcial para a obtenção do grau de **Doutor em Engenharia de Teleinformática**. Área de concentração: Sinais e Sistemas.

Orientadora: Profa. Dra. Fátima N. Sombra de Medeiros.

FORTALEZA

2013

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Pós-Graduação em Engenharia - BPGE

-
- R135a Ramalho, Geraldo Luis Bezerra.
Análise de imagens por meio da matriz de interdependência e da transformação estrutural multiescala / Geraldo Luis Bezerra Ramalho. – 2013
163 f. : il. color., enc. ; 30 cm.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Departamento de Engenharia de Teleinformática, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2013.
Concentração: Sinais e Sistemas.
Orientação: Profa. Dra. Fátima N. Sombra de Medeiros.
1. Teleinformática. 2. Processamento de imagens. I. Título.

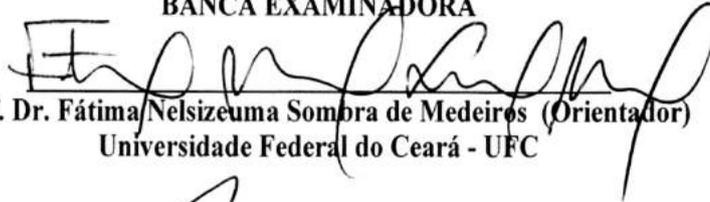
GERALDO LUÍS BEZERRA RAMALHO

**ANÁLISE DE IMAGENS POR MEIO DA MATRIZ DE INTERDEPENDÊNCIA E DA
TRANSFORMAÇÃO ESTRUTURAL MULTIESCALA**

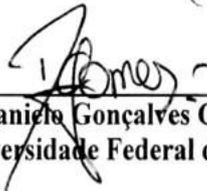
Tese submetida à Coordenação do Curso de Pós-Graduação em Engenharia de Teleinformática, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Doutor em Engenharia de Teleinformática, área de concentração Sinais e Sistemas.

Aprovada em 02/12/2013.

BANCA EXAMINADORA



Prof. Dr. Fátima Nelsizeuma Sombra de Medeiros (Orientador)
Universidade Federal do Ceará - UFC



Prof. Dr. Daniela Gonçalves Gomes (PPGETI/UFC)
Universidade Federal do Ceará - UFC



Prof. Dr. Marco Antônio Botelho Soares
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE



Prof. Dr. Francisco Nivando Bezerra
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE



Profa. Dra. Andrea Gomes Campos Bianchi
Universidade Federal de Ouro Preto - UFOP



Profa. Dra. Nina Sumiko Tomita Hirata
Universidade de São Paulo - USP

À minha mãe e guerreira Helena.

À minha esposa e companheira Raquel.

Ao meu abençoado filho Guilherme.

AGRADECIMENTOS

A todos os professores, colegas, alunos e amigos que contribuíram para que este trabalho pudesse ser concluído.

Aos colegas do departamento de indústria do *Campus* Maracanaú do IFCE, Professores Samuel Dias, Luiz Daniel, Celso Schmidlin, Nélio Costa, Pedro Pedrosa, Adriano Holanda, Ciro dos Santos, José Daniel, Venceslau Xavier, Frederico Matos, Francisco Oliveira e Rodrigo Freitas pelas contribuições, apoio e paciência.

Ao colega Prof. Daniel Ferreira, pela grande colaboração com suas críticas e sugestões e à colega Profa. Janaina Cruz, pelo grande apoio.

Ao Prof. Dr. Hélio Cordeiro, coordenador do ENGESOLDA, por ter disponibilizado o banco de imagens que deu origem a este trabalho e ao Engenheiro do LACAM, Dr. Flávio Herculano pela prestimosa contribuição.

Ao aluno e colega Victor Moita, pela prestimosa ajuda nas atividades de laboratório.

Aos colegas do LABVIS/GPI pela valorosa colaboração.

À Professora Dra. Fátima N. Sombra de Medeiros, pela sua dedicação e por acreditar neste trabalho.

Resumo

A análise de imagens é uma tarefa fundamental em visão computacional. Ela influencia o desenvolvimento de algoritmos de processamento digital de imagens e as abordagens de avaliação dos resultados produzidos por estes algoritmos. Esta tese introduz uma metodologia para a análise estrutural de imagens baseada no uso combinado de uma transformação estrutural multiescala e da extração de características estruturais por meio da matriz de interdependência espacial. A transformação estrutural multiescala é um algoritmo baseado no arcabouço da morfologia matemática que mapeia os níveis de cinza da imagem de entrada para um espaço no qual esses níveis de cinza estão reagrupados em diferentes escalas de estruturas que formam os objetos. A transformação pode ser aplicada no realce de imagens em níveis de cinza e na decomposição de imagens binárias em estruturas elementares. A matriz de interdependência espacial é um algoritmo baseado na estatística de coocorrência que produz uma representação global das coincidências das estruturas de duas imagens de entrada. Essa matriz provê quatro atributos, a saber, correlação, momento de diferença inverso, coeficiente chi-quadrado e entropia, os quais podem ser utilizados como descritores globais das estruturas da imagem. A metodologia proposta é validada com os resultados obtidos para diferentes aplicações: a detecção de corrosão atmosférica em fotografias de superfícies metálicas, a detecção de doenças pulmonares em imagens de tomografia computadorizada, a avaliação referenciada da qualidade da imagens, a segmentação dos vasos da retina em retinografias e a avaliação da qualidade de algoritmos de segmentação de vasos de retina.

Palavras-chave: análise estrutural, matriz de interdependência espacial, extração de características, transformação estrutural multiescala, realce, decomposição.

Abstract

Image analysis is a fundamental task in computer vision. It influences the development of algorithms for digital image processing and approaches for evaluating the results produced by these algorithms. This thesis introduces a methodology for the structural analysis of images based on the combined use of a multiscale structural transformation and extraction of structural features through spatial interdependence matrix. The multiscale structural transformation is an algorithm based on mathematical morphology framework that maps the gray levels of the input image into a space in which these gray levels are grouped into different scales of structures that form objects. The transformation can be applied in enhancement of gray level images and decomposition of binary images into elementary structures. The spatial interdependence matrix is an algorithm based on cooccurrence statistics that produces a global representation of the structural coincidences of two images. This matrix provides four attributes, namely, correlation, inverse difference moment, chi-square coefficient and entropy, which can be used as global descriptors of the image structures. The proposed methodology is validated with the results obtained for different applications: the detection of atmospheric corrosion of metal surfaces in photographs, the detection of lung disease in computerized tomography images, the referenced evaluation of image quality, the segmentation of retinal vessels in retinography and the quality assessment of retinal vessels segmentation algorithms.

Keywords: structural analysis, spatial interdependence matrix, feature extraction, multiscale structural transform, enhancement, decomposition.

Sumário

Lista de Figuras	viii
Lista de Tabelas	xiv
Lista de Símbolos	xv
Lista de Siglas	xviii
1 Introdução	1
1.1 Motivação	3
1.2 Objetivos	6
1.3 Organização da Tese	7
2 Fundamentos Teóricos	9
2.1 Introdução	9
2.1.1 Características estruturais da imagem	9
2.1.2 Processamento de imagens	11
2.2 Medidas clássicas de avaliação de imagens	14
2.3 Estatística de coocorrência	15
2.3.1 Matriz de coocorrência de níveis de cinza	15
2.3.2 Atributos da GLCM	17
2.3.3 Complexidade computacional da GLCM	18
2.4 Morfologia matemática	19
2.4.1 Elemento estruturante	19
2.4.2 Dilatação e erosão	19
2.4.3 Abertura e fechamento	20
2.4.4 <i>Top hat</i>	20
2.4.5 Granulometria	21
2.4.6 Decomposição por limiar	21
2.4.7 Decomposição multiescala por <i>top hat</i>	23
2.4.8 Complexidade computacional da MM	24
2.5 Conclusão	24

3	Estado da Arte da Análise Estrutural	26
3.1	Introdução	26
3.2	A análise estrutural não referenciada	27
3.3	A análise estrutural referenciada	31
3.3.1	Avaliação referenciada da qualidade	32
3.3.2	Avaliação referenciada da segmentação	36
3.4	Análise por decomposição ou transformação	38
3.4.1	Realce morfológico	38
3.4.2	Avaliação da qualidade por decomposição em valores singulares	40
3.5	Conclusão	41
4	Matriz de Interdependência Espacial	43
4.1	Introdução	43
4.2	Matriz de interdependência espacial	44
4.2.1	Informação extraída da MIDE	48
4.2.2	Interpretação dos atributos da MIDE	50
4.2.3	Interpretação visual da MIDE	51
4.3	Extração de características	55
4.3.1	Descritores estruturais	55
4.4	Semelhança estrutural entre imagens	56
4.5	Avaliação da qualidade	59
4.5.1	Índice de qualidade estrutural	60
4.5.2	Mapa local de qualidade	61
4.6	Complexidade computacional da MIDE	62
4.7	Conclusão	62
5	Transformação Estrutural Multiescala	64
5.1	Introdução	64
5.2	Transformação estrutural	65
5.2.1	Transformação estrutural de imagens binárias	65
5.2.2	Transformação estrutural de imagens em níveis de cinza	70
5.2.3	Parâmetros da transformação	71
5.2.4	Decomposição de imagens estruturais por limiar	76
5.3	Realce estrutural de imagens em níveis de cinza	77
5.4	Mapeamento de imagens binárias para níveis de cinza	79
5.5	Complexidade computacional da TEM	81
5.6	Conclusão	81
6	Análise Estrutural Multiescala por meio da MIDE e TEM	83
6.1	Introdução	83
6.1.1	Base de imagens	84
6.2	Caso I: detecção de corrosão atmosférica	89
6.3	Caso II: detecção de doenças em TC de pulmões	94
6.4	Caso III: avaliação da qualidade de imagens	99
6.5	Caso IV: segmentação de vasos de retina	107
6.6	Caso V: avaliação da qualidade da segmentação de vasos de retina	113

6.7	Conclusão	118
7	Conclusões	120
7.1	Produção científica e tecnológica	121
7.2	Perspectivas de trabalhos futuros	121
	Apêndice A Ferramentas de Avaliação	123
	Apêndice B Aplicações Suplementares	135
	Referências Bibliográficas	153

Lista de Figuras

1.1	Diferentes métodos de análise estrutural por decomposição de formas. (a) Decomposição em discos (Adaptado de Xu (2007)). (b) Decomposição por particionamento (Adaptado de Kim, Yun e Lee (2002)). (c) Decomposição em retângulos (Adaptado de Suk e Flusser (2010)).	3
1.2	Resultados de métodos de segmentação da rede de vasos de uma mesma retinografia usando (a) morfologia matemática (ZANA; KLEIN, 2001), (b) mistura gaussiana (SOARES <i>et al.</i> , 2006) e (c) momentos invariantes (MARÍN <i>et al.</i> , 2011).	4
1.3	(a) Imagem “Lena” e (b-d) diferentes versões degradadas, ambas com EQM = 225. Adaptado de Wang e Bovik (2009).	4
1.4	(a) Tubulação industrial submetida à corrosão atmosférica. (b,c) Recorte da imagem (a) ilustrando regiões rugosas sem corrosão e com corrosão, respectivamente.	5
1.5	Uso da metodologia proposta em um sistema de visão computacional.	7
2.1	(a) Imagem estruturada. (b) Histograma da imagem (a). (c) Versão embaralhada da imagem (a). (d) Histograma da imagem (c).	11
2.2	Exemplo de degradação estrutural causada pela contaminação por ruído gaussiano. (a) Região em destaque. (b) Sem ruído. (c) $\mu = 0$, $\sigma = 0,002$, PSNR = 30,3, EQM = 60,8. (d) $\mu = 0$, $\sigma = 0,02$, PSNR = 19,9, EQM = 655,4.	12
2.3	(a) Uma imagem contendo $N_g = 4$ níveis de cinza. (b) Exemplo de GLCM para a vizinhança de um pixel à direita. O elemento em destaque registra o número de transições do par de intensidades (2,1).	16
2.4	Elementos estruturantes simétricos (a-e) disco e (f-j) quadrado de diferentes escalas.	20
2.5	Exemplo de operações morfológicas sobre uma (a) imagem binária. Resultado das operações de (b) abertura e (c) <i>top hat</i>	21
2.6	Modelo generalizado de uma decomposição por limiar de uma imagem em níveis de cinza.	22
2.7	Modelo esquemático de uma decomposição multiescala por <i>top hat</i>	23

3.1	Exemplo da análise de textura com GLCM. (a) Imagem original. (b) Mapa de textura com o atributo de contraste. (c) Mapa de textura com o atributo de homogeneidade.	28
3.2	Exemplo da análise de qualidade com GLCM. (a) Imagem de referência. (b) Imagem contaminada com ruído gaussiano com $\mu = 0$ e $\sigma \in [0, 0,1]$. (c) Avaliação do nível de degradação com os atributos da GLCM.	29
3.3	GLCM 16 níveis da imagem Lena 512 x 512 após diferentes degradações.	30
3.4	Exemplo de análise com a SSIM. (a) Imagem de referência. (b) Imagem contaminada com ruído sal e pimenta. (c) Mapa SSIM com MS-SSIM = 0,81.	34
3.5	Exemplo da análise com MSVD. (a) Imagem de referência. (b) Imagem contaminada com ruído sal e pimenta. (c) Mapa MSVD com valor médio 59,72.	41
4.1	Modelo esquemático da comparação de intensidades das estruturas de duas imagens	44
4.2	Modelo esquemático para obtenção da interdependência espacial . . .	45
4.3	Exemplo de matriz de interdependência entre as imagens I e J , sem alterações estruturais.	46
4.4	Exemplo de matriz de interdependência entre as imagens I e J , com alterações estruturais em J . A região que sofreu mudança estrutural aparece em destaque.	47
4.5	Quadrantes da MIDE.	48
4.6	(a) Imagem original. (b-f) Ruído gaussiano com $\mu = 0$ para diferentes valores de σ . (g) Comparação entre PSNRn, COR, MDI, ENTn e CHI.	51
4.7	Imagem em nível cinza e diferentes níveis de variação de nível DC. . .	52
4.8	MIDE imagem nível cinza e diferentes níveis de degradação por ruído Poisson.	53
4.9	MIDE imagem binária. Sensibilidade da MIDE ao ruído S&P.	54
4.10	Exemplo de texturas representadas pelo vetor de atributos extraído da MIDE. (a) Texturas separadas em classes por linha. (b) Espaço de descritores identificando agrupamentos de texturas similares.	56
4.11	Sensibilidade da MIDE a imagens binárias semelhantes. (a,b) Sem ruído. A imagem 5 é a referência. (c,d) Com ruído.	57
4.12	Sensibilidade da MIDE a imagens em níveis de cinza semelhantes. (a,b) Sem ruído. A imagem 1 é a referência. (c,d) Com ruído.	58
4.13	(a) Imagem original. (b-f) Nível DC modificado. (g) Comparação entre $r \times \text{COR}$ e $ent \times \text{ENT}$. As medidas de entropia são apresentadas normalizadas.	59
4.14	Detalhe da imagem “Lena” com diferentes degradações. Adaptado de (WANG; BOVIK, 2002a).	60
4.15	MIDE 16 níveis. Imagem “Lena” com diferentes ruídos.	61
4.16	Mapa local de qualidade obtido com o atributo CHI.	62
5.1	Modelo esquemático da TEM com elemento estruturante quadrado. .	66

5.2	Exemplo de realce multiescala de uma imagem binária com elemento estruturante quadrado. (a) Imagem binária original formada por pixels de valor 1. (b) Imagem transformada contendo pixels de valores 1, 2 e 3, conforme a escala das estruturas.	67
5.3	Resultado da transformação para diferentes funções de ponderação.	69
5.4	Exemplos de transformação estrutural multiescala de uma imagem em níveis de cinza. (a) Imagem original com 4 níveis de cinza. (b) A imagem estrutural resultante da TEM apresenta pixels com valores diferentes, os quais correspondem ao nível de escala das saliências destacadas pela TEM.	70
5.5	Comparação entre imagem original e resultado da TEM. (a) Imagem original em 8 níveis de cinza e (b) imagem realçada. (c) Histograma dos níveis de cinza da imagem original. (d) Histograma dos níveis de cinza da imagem transformada, com destaque para a linha que separa o histograma em duas classes de pixels.	71
5.6	Exemplos de transformação estrutural multiescala de uma imagem binária representando diferentes interpretações. (a) Transformação em 2 níveis utilizando um elemento estruturante circular. (b) Transformação em 3 níveis utilizando um elemento estruturante retangular.	72
5.7	(a-c) Funções de ponderação de escala e (d-f) os resultados da TEM para a imagem da Figura 5.2 (a).	75
5.8	(a) Transformação de estruturas com peso inversamente proporcional às escalas. (b) Evolução do tamanho do elemento estruturante e seus respectivos pesos.	76
5.9	(a) Transformação de estruturas com peso diretamente proporcional às escalas. (b) Evolução do tamanho do elemento estruturante e seus respectivos pesos.	76
5.10	Cortes da transformação de “dog-1” da base MPEG-7, usando a TEM.	77
5.11	Visualização tridimensional da imagem “cameraman”. (a) Original e (b) realçada. As áreas mais claras (em tons de amarelo e vermelho) representam intensidades maiores.	78
5.12	Preservação de bordas na transformação TEM. (a) Imagem original “cameraman”. (c,d,e) Imagens realçadas para $N = 2$, $N = 4$ e $N = 8$. (b,f,g,h) Detecção de bordas pelo método de Canny (1986).	79
5.13	Preservação da topologia da TEM. (a) Imagem original “shark” MPEG-7. (b-c) Imagem realçada com 4 e 8 níveis. (d-e) detecção de bordas pelo método de Canny (1986).	80
5.14	Visualização tridimensional da imagem MPEG-7 “shark”. (a) original e (b) realçada, na qual somente as saliências possuem valores maiores do que zero.	81
6.1	Modelo genérico de aplicação da MIDE e TEM na análise estrutural de imagens.	85
6.2	(a) Exemplo de imagem não degradada da base LIVE e (b-f) diferentes tipos de degradação.	86

6.3	Exemplos de retinografias da base (a-b) DRIVE e (c-d) STARE.	87
6.4	(a) Tubulação exposta à corrosão atmosférica destacando três regiões monitoradas. (b) Mosaico dos recortes de superfícies de equipamentos contendo corrosão atmosférica (três primeiras colunas) e livre de corrosão (demais colunas à direita).	88
6.5	Exemplos da base de imagens de TC de pulmões.	89
6.6	Modelo da detecção da corrosão atmosférica.	89
6.7	Mapa de descritores da (a,c) MIDE e (b,d) GLCM para os atributos de corrosão. As cores identificam as imagens sem corrosão (laranja) e com corrosão (azul).	91
6.8	(a,c) Matrizes de confusão e (b,d) curvas ROC ilustrativos do desempenho da detecção de corrosão atmosférica (Classe 2) utilizando atributos extraídos da MIDE e da GLCM.	92
6.9	Exemplo de imagens de corrosão classificadas incorretamente usando o conjunto de descritores da (a-d) MIDE e (e-g) GLCM.	93
6.10	Modelo da classificação de doenças pulmonares.	94
6.11	(a-c) Matriz-U da MIDE, GLCM e VIF para classificação de TC de pulmão. (d-f) Espaço de descritores em que as cores identificam as classes PS (vermelho), FC (amarelo) e DPOC (azul).	96
6.12	Exemplos de pulmões presentes na fronteira das classes e suas respectivas matrizes MIDE.	97
6.13	Modelo para avaliação da qualidade de imagens. A linha pontilhada indica um procedimento opcional.	99
6.14	Exemplo da análise de qualidade com MIDE, GLCM (GADKARI, 2004) e MSSIM (WANG <i>et al.</i> , 2004), para efeitos combinados de 4 diferentes tipos de degradação. (a) $GLCM \times MIDE$. (b) $MSSIM \times MIDE$. A medida de referência PSNR está normalizada como $PSNR/\max(PSNR)$	101
6.15	Regressão logística das medidas de qualidade de imagem.	105
6.16	Tempo de processamento por imagem para a base LIVE. (a) Métodos mais rápidos (b) Métodos mais lentos.	107
6.17	Representação 3D da detecção da rede de vasos em retinografias.	108
6.18	Segmentação da rede de vasos de retina baseada na TEM.	109
6.19	Comparação entre retinografias. A região destacada contém vasos de pequeno calibre. (a) Canal verde da retinografia original. (b) Retinografia regularizada com base na TEM.	110
6.20	(a) Detecção da rede de vasos em retinografia da base DRIVE, sem FOV (sem máscara). (b) Comparação com o 1º observador da base DRIVE. Os pixels em vermelho e em azul representam os erros por ausência e excesso, respectivamente.	110
6.21	Imagens utilizadas na avaliação supervisionada da segmentação de vasos de retina. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Superposição das distorções mais comuns: SUB + DES.	114

6.22	Detalhe das imagens utilizadas na avaliação supervisionada da segmentação de vasos de retina. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Espelhamento. (h) Rotação. (i) Superposição das distorções mais comuns: SUB + DES.	115
6.23	Modelo do algoritmo para cálculo do índice de qualidade da segmentação.	115
6.24	Imagens após TEM. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Espelhamento. (h) Rotação. (i) Superposição das distorções mais comuns: SUB + DES.	116
6.25	Sensibilidade das medidas de avaliação de segmentação de vasos quanto ao tipo de distorção. (a) Sobresegmentação. (b) Subsegmentação. (c) Ausência de vasos finos. (d) Descontinuidades. (e) Linha central.	117
A.1	(a) Modelo de uma rede ELM com 3 neurônios na camada escondida e 2 neurônios na camada de saída. As setas representam os pesos sinápticos da camada escondida w_{in} e da camada de saída β_{jm} . (b) Modelo de uma rede SOM com 12 neurônios em um mapa 4×3 . As linhas representam a relação de vizinhança entre os neurônios.	124
A.2	(a) Gráfico de espalhamento dos dados simulados com 3 dimensões. (b) Mapa bidimensional (Matriz-U). (c) Mapa rotulado.	127
A.3	Contribuição das variáveis x , y e z para o mapa da Figura A.2 (b).	128
A.4	Elementos de um <i>boxplot</i>	133
B.1	Mapa de descritores da MIDE para falhas de curto circuito. As cores identificam as classes A (vermelho), B (laranja), C (amarelo) e D (azul).	137
B.2	Espaço dos descritores da MIDE para falhas de curto circuito. As cores identificam as classes A (círculo vermelho), B (retângulo laranja), C (triângulo amarelo) e D (estrela azul).	137
B.3	Gráfico da função de ponderação utilizada na detecção de saliências.	138
B.4	Deteção de saliências baseada na TEM. (a) Imagem original “Lena”. (b) Imagem estrutural com 5 níveis. (c) Regiões relevantes na imagem.	138
B.5	Algoritmo de deteção de pontos de interesse.	139
B.6	Deteção de saliências com TEM e MIDE. (a) Imagem original e o recorte de 34×29 pixels utilizado na busca. (b) Regiões de busca resultantes da deteção de pontos de interesse para variações geométricas e de luminosidade.	139
B.7	Deteção de cantos com a TEM. (a) Imagem original “shark” da base MPEG-7. (b) Imagem realçada com 4 níveis. (c) Cantos detetados pelo método de Harris e Stephens (1988). (d) Cantos detetados com $\tau > 1$	141
B.8	Algoritmo para deteção de cantos.	141

B.9	Deteção de cantos com a TEM para formas da base MPEG-7 CE Shape-1 da classe “Device”. (a-c) Em verde os cantos que coincidem com o padrão outro. (d) Em vermelho os cantos não detetados ou não coincidentes com a verdade terrestre.	142
B.10	Base MIAS. Densa, Fatty, Densa.	143
B.11	Função de ponderação para realce de microcalcificações com a TEM.	144
B.12	Realce de estruturas que contêm microcalcificações em uma mamografia 1024 x 1024, 200 μ m. Imagem 253 da base MIAS. (a) Imagem original com relevo suave. (b) Imagem realçada R . (c) Detalhe da imagem original em uma região de microcalcificação. (d) Detalhe da imagem realçada em uma região de microcalcificação.	145
B.13	Detalhe do relevo de uma região de mamografia contendo um aglomerado de microcalcificações. Imagem 253 da base MIAS (a) antes e (b) após realce estrutural.	146
B.14	Realce de precipitados em micrografias. (a,d) Ditch. (b,e) Dual. (c,f) Step.	148
B.15	Algoritmo para segmentação de precipitados intergranulares.	148
B.16	Classificação das amostras de acordo com a área normalizada dos precipitados (ordenados por tipo de aço 304 e 316). Os marcadores escuros indicam as amostras classificadas corretamente.	149
B.17	Classificação de microestruturas com relação ao grau de susceptibilidade à IGC. (a) Matriz de confusão para as classes <i>Step</i> (1), <i>Dual</i> (2) e <i>Ditch</i> (3). (b) Curva ROC.	150
B.18	Função de ponderação para realce de microcalcificações com a TEM.	151
B.19	(a) Imagem original (gentilmente cedida pelo Prof. Dr. Marco Botelho, RENORBIO). (b) Detalhe de uma nanopartícula. O contraste foi alterado para melhor exibição. (c) Limiarização pelo método de Otsu (1979). (d) Rótulos da segmentação pelo método proposto. (e) Histograma do diâmetro em pixels das nanopartículas da imagem (c). (f) Histograma do diâmetro em pixels das nanopartículas maiores da imagem (d). (g) Histograma do diâmetro em pixels das nanopartículas menores da imagem (d).	152

Lista de Tabelas

3.1	Trabalhos de análise de imagens com base na estatística de coocorrência	27
3.2	Trabalhos de avaliação da qualidade imagens com base na estatística de coocorrência	31
3.3	Trabalhos de análise por transformação.	39
5.1	Classificação das funções de ponderação de escala	74
6.1	Subconjuntos da base LIVE, adaptado de (SHEIKH; SABIR; BOVIK, 2006)	86
6.2	Matriz de confusão para os atributos da MIDE	98
6.3	Matriz de confusão para os atributos da GLCM	98
6.4	Comparação entre medidas de avaliação da qualidade de imagem sobre a base LIVE	102
6.5	Análise comparativa do desempenho de iqe e iqe_m com MSSIM, MS-SSIM e VIF para os subconjuntos da base LIVE	103
6.6	Medidas de avaliação da qualidade de imagem das quais o iqe mais se aproxima considerando os tipos de degradação da base LIVE	104
6.7	Medidas de avaliação da qualidade de imagem das quais o iqe_{ms} mais se aproxima considerando os tipos de degradação da base LIVE	104
6.8	Tempo de processamento das medidas de avaliação da qualidade de imagens	106
6.9	Base de retinografias DRIVE	112
6.10	Base de retinografias STARE	112
A.1	Índice κ	132
B.1	Matriz de confusão média de 100 testes de classificação	136

Lista de Símbolos

Conjuntos

\mathbb{Z}	conjunto dos números inteiros
\mathbb{R}	conjunto dos números reais
\mathbb{Z}^n	espaço discreto n -dimensional
\mathbb{R}^n	espaço euclidiano n -dimensional
$\#\{\cdot\}$	cardinalidade de $\{\cdot\}$
$\{x : \star\}$	conjunto de todos os elementos x que satisfazem a propriedade \star

Imagens

I	imagem
E	imagem estrutural
R	imagem realçada
\mathcal{D}	domínio de uma imagem
M_i	número colunas
M_j	número de linhas
N_g	número de níveis de cinza
N	número de níveis de decomposição
p	posição de um pixel p no domínio da imagem
$\mathbf{I}(p)$	intensidade de cinza no pixel p no domínio da imagem I
Ψ	operação de transformação de uma imagem

Morfologia Matemática

S	elemento estruturante
d	distância geodésica
δ	dilatação

ε	erosão
γ	abertura
ϕ	fechamento
τ	operador <i>top hat</i>
\flat	operador <i>bottom hat</i>
ξ	operador de esqueletização
ψ	operador de incremento
Φ	operador de corte
φ	transformação morfológica multiescala
f_p	função de ponderação de escala

Estatística

r	coeficiente de correlação linear (Pearson)
R^2	coeficiente de determinação
ρ	coeficiente de correlação de Spearman
τ	coeficiente de correlação de Kendall
Mcc	coeficiente de correlação de Matthews
\mathbf{d}	vetor de distância
χ^2	distribuição chi-quadrado
\mathfrak{C}	matriz de coocorrência de níveis de cinza
\mathfrak{c}_{ij}	frequência de ocorrência do par (i,j)
\mathfrak{c}	elemento da matriz de coocorrência
Cor	correlação da matriz de coocorrência de níveis de cinza
Hom	momento de diferença inverso da matriz de coocorrência de níveis de cinza
Ent	entropia da matriz de coocorrência de níveis de cinza
Con	contraste da matriz de coocorrência de níveis de cinza

Análise Estrutural

\mathfrak{M}	matriz de interdependência espacial
$\mathfrak{m}_{i,j}$	frequência de ocorrência do par (i,j) de \mathfrak{M}
COR	correlação de \mathfrak{M}
MDI	momento de diferença inverso de \mathfrak{M}
ENT	entropia de \mathfrak{M}
CHI	chi-quadrado de \mathfrak{M}

iqe	índice de qualidade estrutural
iqe _m	índice de qualidade estrutural multiescala
iqs	índice de qualidade da segmentação

Lista de Siglas

BG	Borramento Gaussiano
DMM	Decomposição Morfológica Multiescala
DMOS	Diferença do Escore Médio do Observador (<i>Difference Mean Observer Score</i>)
DPOC	Doença Pulmonar Obstrutiva Crônica
DRIVE	Base de imagens de retina
EQM	Erro Quadrático Médio
FF	<i>fast fading</i>
GLCM	Matriz de Coocorrência de Níveis de Cinza (<i>Gray Level Cooccurrence Matrix</i>)
HVS	Sistema Visual Humano (<i>Human Visual System</i>)
IQM	Medida de Qualidade de Imagem (<i>Image Quality Measure</i>)
LIVE	Base de imagens naturais degradadas
LM	Algoritmo de otimização de Levenberg-Marquardt
MC	Microcalcificação
MIAS	Base de imagens mamográficas
MIDE	Matriz de Interdependência Espacial
MM	Morfologia Matemática
MSSIM	Média da Similaridade Estrutural (<i>Mean Structural Similarity</i>)
MS-SSIM	Similaridade Estrutural Multiescala (<i>Multiscale Structural Similarity</i>)
M-SVD	Medida de qualidade baseada em Decomposição de Valores Singulares (<i>Singular Value Decomposition</i>)
PDI	Processamento Digital de Imagens
PDF	Função Densidade de Probabilidade (<i>Probability Density Function</i>)
PO	Padrão Ouro

PSNR	Pico da Relação Sinal-Ruído (<i>Peak Signal-to-Noise Ratio</i>)
RB	Ruído Branco
RNA	Redes Neurais Artificiais
RP	Reconhecimento de Padrões
SFIM	Matriz de Interdependência Espacial de Atributos (<i>Spatial Feature Interdependence Matrix</i>)
SOM	Mapa auto-organizável (<i>Self-Organizing Map</i>)
SSIM	Similaridade Estrutural (<i>Structural Similarity</i>)
STARE	Base de imagens de retina
SVC	Sistema de Visão Computacional
SVD	Decomposição de Valor Singular (<i>Singular Value Decomposition</i>)
TEM	Transformação Estrutural Multiescala
TC	Tomografia Computadorizada
VIF	Fidelidade de Informação Visual <i>Visual Information Fidelity</i>

Introdução

A visão humana é um processo complexo que resulta em uma descrição útil e ordenada de informações relevantes a partir de imagens do mundo externo sendo, portanto, uma tarefa de processamento de informações (MARR, 1982). Uma das principais tarefas desse processo é a análise de imagens, a qual consiste em extrair informações de uma cena. Essa tarefa pode ser tão sofisticada quanto a identificação e descrição de objetos presentes em uma cena para que seja possível interpretá-la.

Pela sua complexidade, a visão é um dos mais estimulantes campos de pesquisa, tanto sob a perspectiva científica quanto tecnológica (COSTA; CESAR, 2009). Nesse contexto, a visão computacional pode ser entendida como um conjunto de técnicas de Processamento Digital de Imagens (PDI) e de Reconhecimento de Padrões (RP) que busca extrair significado de uma cena registrada digitalmente. Uma imagem digital é uma representação espacial de um objeto em uma cena bidimensional. A representação mais elementar de uma imagem é chamada pixel, o qual corresponde a uma intensidade ou nível de cinza, para o caso de imagens monocromáticas, a um valor lógico para imagens binárias ou a uma classe para imagens com múltiplas regiões segmentadas (HARALICK; SHAPIRO, 1992). Na análise de imagens, saliências visuais são as regiões de uma imagem que naturalmente chamam a atenção do observador, porque seu conteúdo se destaca da sua vizinhança (ITTI; KOCH, 2001; XU; CHEN, 2004; HARDING; ROBERTSON, 2013). A relação espacial entre as intensidade dos pixels vizinhos formam estruturas que possuem correspondência com as saliências¹ da imagem. Cantos, bordas e textura são exemplos de estruturas

¹Nesta tese, os termos saliência, marco estatístico, ponto de interesse, alvo e canto são considerados equivalentes no que se refere à descrição de estruturas de uma imagem (binária ou em níveis de cinza) que se destacam da sua vizinhança.

que, dependendo da forma como estão arranjadas espacialmente, podem destacar objetos ou partes deles. Há um entendimento de que a interpretação do arranjo dessas estruturas é a forma como o sistema visual humano (HVS, do inglês *human visual system*) reconhece objetos em uma cena (WANG; BOVIK, 2002b).

Koenderink (1984) descreve as imagens em termos estruturais sendo que Koenderink e Doorn (1982) definem que o sistema visual é extensivo² o suficiente para conseguir representar a imagem da retina em todos os níveis de resolução simultaneamente. Entretanto, a relevância dos detalhes de uma imagem muitas vezes reside em uma faixa restrita de escala. De modo geral, a saliência é uma estrutura que chama a atenção do observador. Na observação de uma forma binária, por exemplo uma estrela, seus cantos e vértices se destacam. Em uma imagem em níveis de cinza, por exemplo um retrato, as saliências são as estruturas correspondentes a olhos, nariz e boca. De fato, estas estruturas e a relação espacial entre elas, são características importantes em um processo de reconhecimento de objetos, como na tarefa de detecção de faces (ROTSHEIN *et al.*, 2008).

Uma abordagem para o reconhecimento de objetos consiste na decomposição em suas partes constituintes, a qual tem um apelo intuitivo (ULLMAN, 1996). A decomposição em partes assume que os objetos podem ser formados por um conjunto pequeno de componentes genéricos e que todos os objetos podem ser descritos como diferentes combinações dos mesmos componentes.

Esses conceitos sugerem que a análise de imagens deve incorporar observações sobre suas partes constituintes e estruturas para tentar aproximar a percepção do HVS. Portanto, é importante que o desenvolvimento de Sistemas de Visão Computacional (SVCs) contemple a análise estrutural de imagens que, nesta tese, pode ser expressa por três princípios:

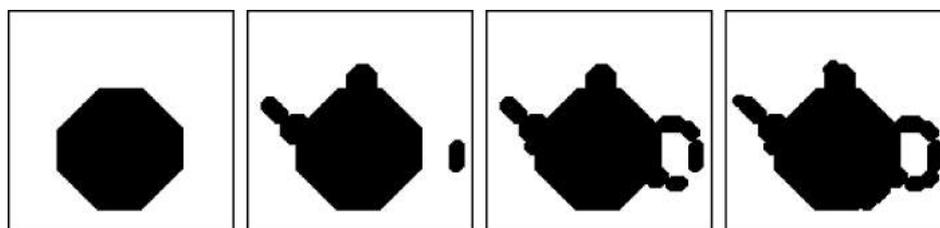
- ▶ A forma como um objeto aparece na imagem tem relação direta com a configuração espacial dos valores dos pixels e tem pouca relação com um valor específico (HARALICK; SHAPIRO, 1992).
- ▶ Perceber uma imagem é transformá-la (SERRA, 1982).
- ▶ Não existe uma estratégia única que seja apropriada para todos os problemas que envolvem a análise de imagens (ULLMAN, 1996).

²que pode ser aplicado em casos análogos

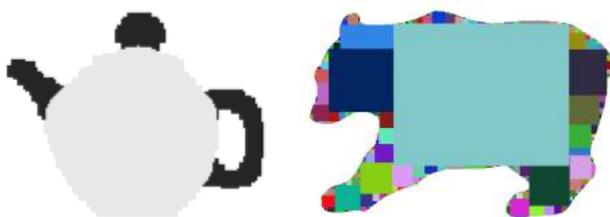
1.1 Motivação

Algoritmos de análise estrutural de imagens têm se mostrado eficazes em aproximar a percepção do HVS, conforme observado nos resultados de diversos trabalhos (ZHANG *et al.*, 2013; MASCHAL *et al.*, 2013; LEITE, 2012; GEGÚNDEZ-ARIAS *et al.*, 2012; FRAZ *et al.*, 2012; SUK; HOSCHL; FLUSSER, 2012; WANG *et al.*, 2004; SOARES *et al.*, 2006; SHEIKH; SABIR; BOVIK, 2006). Esses trabalhos reportam diferentes abordagens de análise estrutural que incorporam algum nível de subjetividade, se aproximam da percepção do HVS e possibilitam sua aplicação na análise de imagens em níveis de cinza e binárias.

Braga-Neto e Goutsias (2005) introduziram um método de análise multiescala de imagens com base na conectividade das estruturas. Os experimentos apresentados incluem aplicações em granulometria, segmentação e classificação de estruturas. Além disso, alguns trabalhos relatam resultados da análise de imagens por meio da decomposição em elementos geométricos simples (SUK; HOSCHL; FLUSSER, 2012; SUK; FLUSSER, 2010; XU, 2007) ou por meio do particionamento do objeto em componentes (KIM; YUN; LEE, 2002). A Figura 1.1 ilustra exemplos de diferentes abordagens da análise estrutural de formas binárias.



(a) Discos sobrepostos



(b) Particionamento

(c) Retângulos

Figura 1.1: Diferentes métodos de análise estrutural por decomposição de formas. (a) Decomposição em discos (Adaptado de Xu (2007)). (b) Decomposição por particionamento (Adaptado de Kim, Yun e Lee (2002)). (c) Decomposição em retângulos (Adaptado de Suk e Flusser (2010)).

A análise estrutural possibilita a identificação de estruturas de interesse,

como ocorre em problemas de segmentação de estruturas delgadas (LEITE, 2012; SOARES *et al.*, 2006; ZANA; KLEIN, 2001; MENDONCA; CAMPILHO, 2006; VILLALOBOS-CASTALDI; FELIPE-RIVERÓN; SÁNCHEZ-FERNÁNDEZ, 2010; FRAZ *et al.*, 2012). Exemplos típicos do resultado de métodos de segmentação da rede de vasos a partir de retinografias estão ilustrados na Figura 1.2.

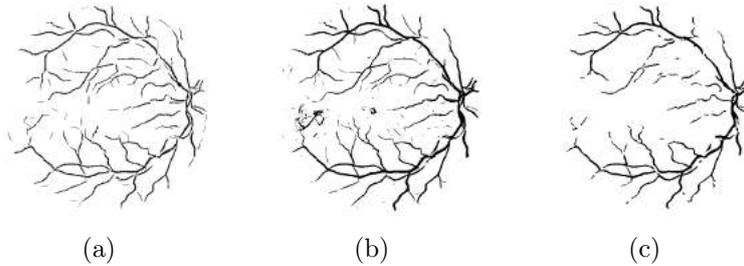


Figura 1.2: Resultados de métodos de segmentação da rede de vasos de uma mesma retinografia usando (a) morfologia matemática (ZANA; KLEIN, 2001), (b) mistura gaussiana (SOARES *et al.*, 2006) e (c) momentos invariantes (MARÍN *et al.*, 2011).

Adicionalmente, a análise estrutural está presente na avaliação da qualidade de imagens. Métodos de avaliação que não levam em consideração as estruturas da imagem, como o erro quadrático médio (EQM), que se baseia no valor médio das diferenças pixel a pixel, tendem a falhar nessa tarefa (WANG; BOVIK, 2002a). A Figura 1.3 ilustra exemplos em que a adoção do EQM como medida de avaliação não é capaz de descrever a forma como o HVS interpreta perdas de informação em uma imagem degradada.



Figura 1.3: (a) Imagem “Lena” e (b-d) diferentes versões degradadas, ambas com EQM = 225. Adaptado de Wang e Bovik (2009).

Diversas propostas buscam resolver problemas específicos (SUK; HOSCHL; FLUSSER, 2012; GEGÚNDEZ-ARIAS *et al.*, 2012), embora existam exceções, como os métodos baseados na extração de características e na avaliação da qualidade

de imagens (WANG; BOVIK, 2009; LEITE, 2012; BRAGA-NETO; GOUTSIAS, 2005; CHOI; KIM, 2005). A Figura 1.4 ilustra um problema de detecção de corrosão abordada pela extração de atributos de textura (CHOI; KIM, 2005). Contudo, o mesmo problema poderia ser abordado utilizando a avaliação da qualidade de imagens (SHEIKH; BOVIK; VECIANA, 2005) como método de extração de informação estrutural.

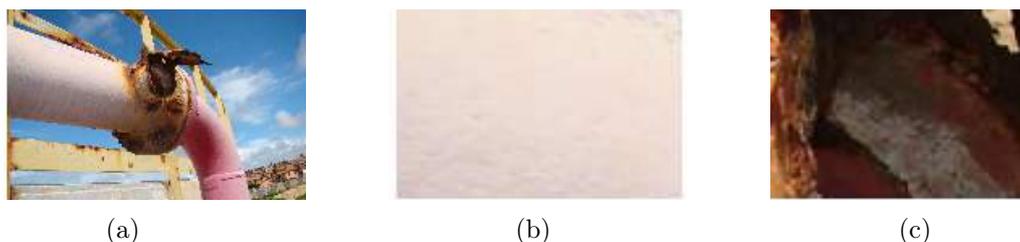


Figura 1.4: (a) Tubulação industrial submetida à corrosão atmosférica. (b,c) Recorte da imagem (a) ilustrando regiões rugosas sem corrosão e com corrosão, respectivamente.

Da literatura se observa que alguns métodos de avaliação estrutural têm seu emprego restrito a imagens binárias ou a imagens em níveis de cinza (XU, 2007; SHEIKH; BOVIK; VECIANA, 2005). Outros métodos possuem elevado custo computacional (ZANA; KLEIN, 2001; SHEIKH; BOVIK, 2006) e existem ainda aqueles que dependem de uma etapa de treinamento (SOARES *et al.*, 2006; MARÍN *et al.*, 2011; CHOI; LEE, 2011) para extrair a informação desejada, sendo frequentemente dependentes do banco de dados utilizado para ajuste de parâmetros. Essas características podem inviabilizar a aplicação desses métodos em sistemas que exigem o processamento em tempo real ou na solução de problemas com restrições para a obtenção de padrões conhecidos, necessários ao treinamento.

Além disso, alguns trabalhos de segmentação reportam resultados avaliados por métodos que não consideram as características estruturais dos objetos contidos na cena (SOARES *et al.*, 2006; VILLALOBOS-CASTALDI; FELIPE-RIVERÓN; SÁNCHEZ-FERNÁNDEZ, 2010). Na avaliação da segmentação da rede de vasos de retina, por exemplo, seria útil uma transformação capaz de realizar sua decomposição em estruturas com diferentes graus importância. Dessa forma seria possível comparar o resultado de diferentes algoritmos de segmentação obedecendo requisitos da aplicação.

1.2 Objetivos

Esta tese apresenta uma metodologia de análise estrutural que pode ser aplicada tanto a imagens digitais binárias quanto em níveis de cinza. Para alcançar esse objetivo geral, foram traçados os seguintes objetivos específicos:

- ▶ Propor uma metodologia de avaliação referenciada de imagens binárias ou em níveis de cinza, baseado em estatística de coocorrência (interdependência espacial), por meio de mapeamento para um espaço diferente daquele da imagem de entrada;
- ▶ Propor uma transformação de imagens em níveis de cinza ou binárias para um espaço no qual os pixels sejam reagrupados com base na escala das estruturas formadas por eles (transformação estrutural multiescala);
- ▶ Validar os métodos propostos por meio de experimentos em diferentes aplicações de visão computacional.

A proposta deste trabalho se enquadra na classe de métodos de avaliação objetiva direta de imagens, com base em uma referência. O diagrama apresentado na Figura 1.5 ilustra as possibilidades de uso da metodologia em um sistema de visão computacional. A transformação estrutural multiescala é aplicada na segmentação de imagens e no realce de estruturas, saliências ou alvos, conforme a aplicação. A estatística de interdependência espacial pode ser aplicada na avaliação de imagens, por meio da obtenção de medidas de similaridade e de distorção estrutural. Em conjunto com a transformação, a interdependência espacial pode ser utilizada na extração de atributos para a interpretação de objetos da imagem.

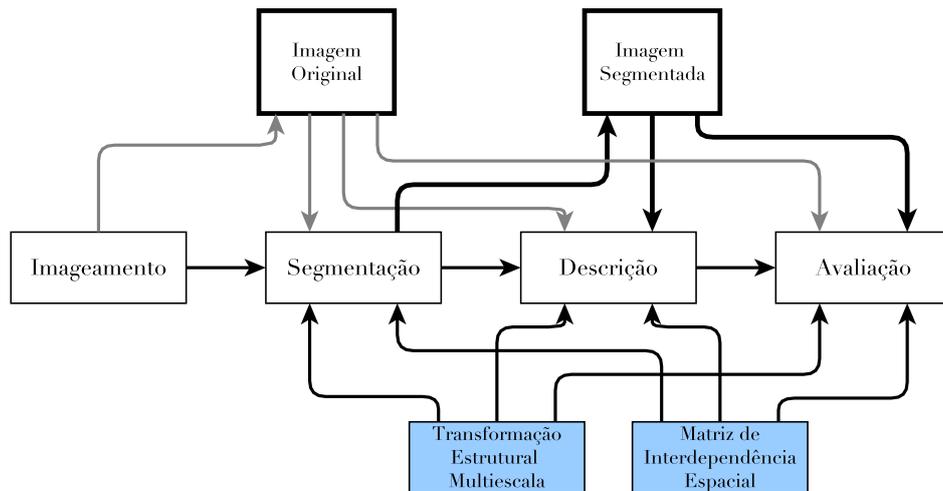


Figura 1.5: Uso da metodologia proposta em um sistema de visão computacional.

1.3 Organização da Tese

Este trabalho foi organizado de modo que sejam introduzidas as ferramentas e conceitos básicos da análise de imagens no contexto estrutural, com o objetivo de dar destaque às contribuições da pesquisa.

Conceitos teóricos utilizados no desenvolvimento desta tese são apresentados no Capítulo 2 e estão relacionados à Morfologia Matemática e à Estatística de Coocorrência. Este capítulo também apresenta as medidas de desempenho do resultado dos algoritmos avaliados em experimentos de testes no Capítulo 6.

O Capítulo 3 apresenta os métodos mais recentes no contexto da análise estrutural de imagens, incluindo métodos de decomposição de formas e avaliação da qualidade de imagens.

O Capítulo 4 introduz o conceito de Matriz de Interdependência Espacial (MIDE) e a interpretação da informação extraída dessa matriz. A MIDE, concebida como uma ferramenta de análise de imagens em níveis de cinza ou binárias, é proposta com o intuito de fornecer uma representação visual e numérica das distorções estruturais entre uma imagem de teste e uma imagem de referência.

O Capítulo 5 introduz um método de transformação de imagens por meio da superposição multiescala baseado em operações *top hat*. Esta transformação, nomeada Transformação Estrutural Multiescala (TEM), é uma ferramenta de mapeamento da imagem de entrada para um espaço que representa suas estruturas em um contexto multiescala.

O Capítulo 6 destaca as principais aplicações dos métodos propostos e apresenta uma discussão sobre os resultados obtidos em análise estrutural de imagens por meio da MIDE e da TEM. É apresentada uma comparação do desempenho com métodos clássicos e do estado-da-arte.

No Capítulo 7 estão as conclusões finais desta tese e sugestões de investigações futuras no tema estudado.

As ferramentas de reconhecimento de padrões e estatísticas, empregadas nos experimentos e na avaliação de resultados, são apresentadas no Apêndice A.

No Apêndice B são apresentadas aplicações suplementares que ilustram o emprego da metodologia em análise de sinais unidimensionais, detecção de saliências e realce de alvos.

Capítulo 2

Fundamentos Teóricos

Este capítulo apresenta uma revisão de técnicas de análise no domínio espacial comumente empregadas na filtragem, decomposição e extração de características de imagens. As técnicas pertencem a duas áreas de estudo, a saber, a Morfologia Matemática e a Estatística de Coocorrência, constituindo o arcabouço teórico para a compreensão dos métodos propostos e para a avaliação dos resultados obtidos.

2.1 Introdução

A análise de imagens é um processo que precede a interpretação dos objetos contidos em uma cena. Nesta tese, analisar uma imagem tem um significado amplo que vai desde a avaliação da sua qualidade, passando pelo processamento e interpretação dos objetos presentes na cena. O processamento envolve a realização de transformações na imagem a fim de filtrar, realçar ou segmentar objetos de interesse. Em alguns casos, é importante definir diferentes tipos de descritores que possam ser utilizados como auxílio na segmentação ou mesmo fornecer meios para a interpretação dos objetos ou da imagem como um todo. Neste contexto, duas técnicas são amplamente empregadas na avaliação da qualidade de imagens, extração de características ou segmentação de uma cena e na decomposição ou realce de imagens. Estas técnicas compreendem, respectivamente, os descritores de textura baseados na estatística de coocorrência de níveis de cinza (HARALICK; DINSTEIN, 1973) e os operadores morfológicos (SOILLE, 2002; SERRA, 1982).

2.1.1 Características estruturais da imagem

Existem quatro fatores responsáveis pela intensidade do pixel em uma imagem: a geometria e a refletância da superfície visível dos objetos, a iluminação da cena

e o ponto de vista do observador. Todos esses fatores estão misturados em uma imagem bidimensional. Dentre os diferentes aspectos da informação visual, a forma dos objetos é uma das mais importantes (COSTA; CESAR, 2009). Baseando-se nas propriedades físicas do processo de obtenção de imagens, Marr (1982) fez algumas suposições sobre como o observador percebe os objetos da cena por meio da análise das estruturas que os formam:

- ▶ Existência de superfícies: o mundo físico pode ser entendido como sendo uma composição de superfícies suaves que possuem função de refletância de tal modo que as estruturas espaciais podem ser elaboradas.
- ▶ Organização hierárquica: As estruturas estão organizadas espacialmente de acordo com as propriedades dos objetos em diferentes escalas; a organização espacial da função de refletância de uma superfície é frequentemente produzida por uma quantidade de diferentes processos, cada um operando em uma escala diferente.
- ▶ Similaridade: os itens gerados pelo processo de geração de refletância em uma dada superfície agindo sobre uma dada escala tendem a ser mais parecidos com outros de mesmo tamanho, contraste local, cor e organização espacial.
- ▶ Continuidade espacial: em geral, os locais de descontinuidades em profundidade ou em orientação da superfície são suaves.
- ▶ Continuidade de fluxo: se uma direção de movimento é descontínua em mais de um ponto – ao longo de uma linha por exemplo – então um contorno de objeto está presente.

De modo geral, a interpretação de uma cena pode ser obtida a partir da análise das estruturas presentes na imagem, pertencentes aos objetos e às suas partes constituintes. Este processo envolve, especialmente, a detecção de mudanças de intensidade, a representação e análise de estruturas geométricas locais, além da detecção de efeitos como fontes de luz, saliência e transparência (MARR, 1982). É importante que o desenvolvimento de métodos de análise de imagens tenha inspiração nessas propriedades, com o objetivo de se aproximar da função do sistema visual humano (HVS, do inglês *human visual system*) na interpretação de cenas, seja por meio da extração de informação diretamente da imagem ou de uma transformação da mesma.

Seguindo o raciocínio de Marr (1982), a propriedade da similaridade, por exemplo, permite atribuir um significado a estruturas que compartilham o mesmo nível de intensidade luminosa ou de escalas semelhantes. Este é o princípio básico por trás da técnica de avaliação estatística de textura proposta por Haralick (1979), na qual os conceitos de homogeneidade e de contraste são explorados. A observação da continuidade espacial fornece uma importante pista para a análise de textura e também do nível de degradação de uma imagem. Por fim, a continuidade de fluxo é um critério para a detecção e identificação de objetos, conceito que foi empregado no detetor de bordas de Canny (1986).

Conclui-se que as características estruturais de uma imagem correspondem a um arranjo entre as partes constituintes de um objeto. Este arranjo pode ser descrito pela relação entre os valores de intensidade dos pixels em uma dada vizinhança, ou seja, o agrupamento desses pixels em estruturas. Nas Figuras 2.1 (a) e (c) são apresentadas duas imagens que possuem a mesma distribuição de intensidades, conforme ilustram os respectivos histogramas de intensidades nas Figuras 2.1 (b) e (d). Contudo, a primeira imagem apresenta uma alta organização estrutural, determinada pelo arranjo espacial das intensidades (ou níveis de cinza) dos pixels, formando objetos que são facilmente reconhecidos pelo HVS. De fato, a imagem da Figura 2.1 (c) contém os mesmos pixels da imagem da Figura 2.1 (a), entretanto a distribuição espacial aleatória dos seus pixels torna impossível a identificação dos objetos daquela imagem.

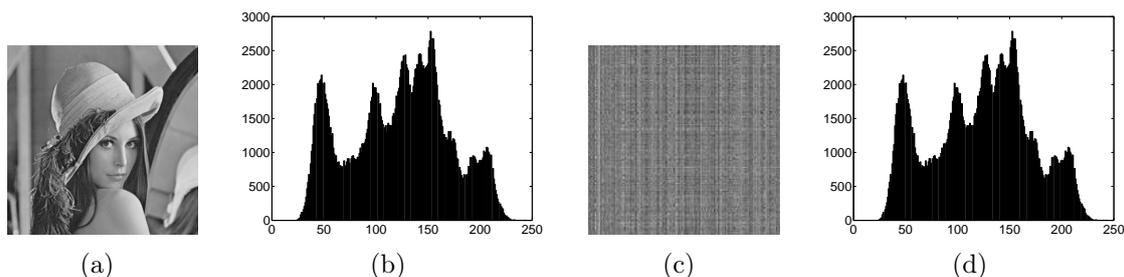


Figura 2.1: (a) Imagem estruturada. (b) Histograma da imagem (a). (c) Versão embaralhada da imagem (a). (d) Histograma da imagem (c).

2.1.2 Processamento de imagens

Wang e Bovik (2002b) reportam que a principal função do olho humano é extrair informação estrutural e as imagens degradadas comprometem a capacidade do HVS interpretar corretamente a cena registrada.

Em geral, as imagens digitais estão sujeitas a degradações durante sua aquisição, processamento, armazenamento, transmissão e reprodução. Como pode ser observado no detalhe da Figura 2.2 (b), a proximidade entre os valores de intensidade de pixels vizinhos permite a percepção de estruturas circulares que definem os componentes elementares de um olho humano, das extremidades para o centro: globo ocular, íris e pupila. As imagens das Figuras 2.2 (c) e (d) ilustram como a degradação por ruído gaussiano de média μ e desvio padrão σ é capaz de modificar as estruturas de uma imagem, neste caso, podendo torná-la incompreensível. Percebe-se que o ruído altera a relação entre as intensidades das estruturas vizinhas, que antes formavam um arranjo estrutural homogêneo. A compreensão do objeto fica bastante prejudicada conforme aumenta o grau de degradação, ilustrado na Figura 2.2 (d). Neste exemplo, os valores de EQM e pico da relação sinal-ruído (PSNR, do inglês *peak signal-to-noise ratio*) foram utilizados para fornecer uma medida da degradação.

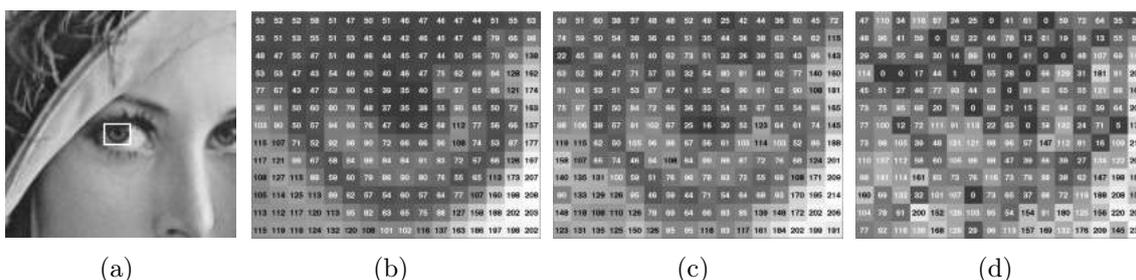


Figura 2.2: Exemplo de degradação estrutural causada pela contaminação por ruído gaussiano. (a) Região em destaque. (b) Sem ruído. (c) $\mu = 0$, $\sigma = 0,002$, PSNR = 30,3, EQM = 60,8. (d) $\mu = 0$, $\sigma = 0,02$, PSNR = 19,9, EQM = 655,4.

A classificação da informação observada em “ruído” ou detalhe “relevante” depende da interpretação do observador. As técnicas de filtragem e análise de imagens devem, em algum nível, corresponder a essa subjetividade (SETHIAN, 1999). O desenvolvimento ou escolha da técnica de filtragem mais adequada pode ser positivamente influenciada por uma análise prévia da imagem, principalmente quando há disponível algum método computacional de avaliação capaz de aproximar-se da avaliação subjetiva da visão.

O reconhecimento de objetos é outro importante aspecto da percepção visual (ULLMAN, 1996). Segundo Haralick e Shapiro (1992), reconhecer um objeto em uma imagem significa conhecê-lo quando é visto de uma orientação particular. Uma

das etapas da tarefa de reconhecer objetos consiste em conseguir interpretá-los no todo ou em partes. Por isso, a segmentação automática, também possui fundamental importância para a análise de uma imagem por meio de visão computacional. Zhang (1996) afirma que muitos autores concordam que o desempenho do processo de segmentação possui influência direta no resultado de um sistema de visão.

O reconhecimento de objetos e a interpretação de cenas por meio da visão computacional apresenta inúmeros desafios. Fatores adversos da aquisição de imagens digitais influenciam negativamente no desempenho dos sistemas de visão computacional. Dependendo da aplicação, é possível obter excelentes resultados incorporando algum nível de conhecimento prévio do problema à tarefa de segmentação. A dependência desse conhecimento prévio, contudo, torna o projeto dos sistemas de visão computacional susceptíveis a diferentes interpretações e os resultados podem variar significativamente dentro de um mesmo escopo (SETHIAN, 1999). Isto sugere que, uma técnica de análise de imagens desenvolvida com vistas em uma aplicação específica pode não ser capaz de generalização, ou seja, ser empregada em problemas diversos.

Duas técnicas de processamento de imagens merecem destaque pela sua capacidade de generalização para diferentes problemas. Dentre os métodos espaciais de análise de imagens, vale destacar a análise de textura por meio da estatística de coocorrência e a análise multiescala, ou granulometria por meio de operações morfológicas.

A matriz de coocorrência de níveis de cinza (GLCM, do inglês *gray level cooccurrence matrix*) baseia-se na estatística de coocorrência e é amplamente empregado na extração de descritores de textura de uma imagem, por exemplo para fins de classificação como sugerem Bevk e Kononenko (2002) e Cai-xiang e Shu-bo (2009). A GLCM, apresentado originalmente por Haralick (1979), utiliza o conceito de dependência espacial dos níveis de cinza de uma imagem para produzir uma matriz que incorpora informação de vizinhança entre pixels. Haralick e Shapiro (1992) e Haralick (1979) descrevem diversas abordagens baseadas na análise estrutural. As informações extraídas dessa matriz são utilizadas para descrever uma imagem em função da textura de seus componentes. Contudo, a utilização da estatística de coocorrência é ainda mais abrangente, dado que os atributos extraídos da GLCM podem ser utilizados para criar imagens descritivas e também com o objetivo de segmentar imagens para identificar objetos presentes na cena.

A análise espacial multiescala, também conhecida como filtragem morfológica

ou granulometria, encontra suas bases na teoria da Morfologia Matemática (MM) (SOILLE, 2002; SERRA, 1982). Essa teoria matemática apresenta ferramentas para a decomposição de formas, filtragem e realce. A MM é um arcabouço poderoso na análise de imagens, mas que pode ser compreendido por duas operações elementares: erosão e dilatação. Diversos operadores morfológicos são derivados dessas operações, destacando-se aqueles que aparecem em métodos de análise encontrados na literatura, a saber, as operações de abertura, fechamento e *top hat*. Esses operadores morfológicos utilizam estruturas elementares para detetar ou remover estruturas similares na imagem, destacando assim os objetos de interesse (ZANA; KLEIN, 2001; HU; TIAN, 2006; MENDONCA; CAMPILHO, 2006). Hu e Tian (2006) desenvolveram uma metodologia baseada em MM para extração de contorno de objetos. Essa metodologia adota a superposição de efeitos com elementos estruturantes direcionais nebulosos para realçar e detetar o contorno de formas. Esse tipo de segmentação é importante para a análise de imagens, porque incorpora um conjunto de conhecimentos em um algoritmo computacional. Os autores basearam-se na suposição de que os elementos estruturantes lineares não detetam estruturas pontuais e, por isso, o método proposto é robusto à presença de ruído.

2.2 Medidas clássicas de avaliação de imagens

Considere uma imagem em escala de cinza (*grayscale*) \mathbf{I} definida como um mapeamento de um subconjunto retangular \mathcal{D}_I do plano \mathbb{Z}^2 em um conjunto discreto $\{0, 1, 2, \dots, N_g - 1\}$, em que N_g é o número de níveis de cinza possíveis. Seja \mathbf{J} uma versão degradada dessa imagem, por exemplo, contaminada por algum tipo de ruído. Duas medidas clássicas são frequentemente utilizadas para fornecer uma indicação do grau de degradação de \mathbf{J} , admitindo que essa degradação é proporcional às diferenças entre as duas imagens. Vale observar que, quando se utilizam 8 bits de dados para armazenar as intensidades dos pixels, o valor de $N_g = 256$. As imagens binárias constituem um caso particular para o qual $N_g = 2$. Considerando,

$$\mathbf{I} : \mathcal{D}_I \subset \mathbb{Z}^2 \rightarrow \{0, \dots, N_g - 1\}, \quad (2.1)$$

para todo pixel p de coordenadas (i, j) no subconjunto \mathcal{D}_I de uma imagem, a sua intensidade (valor de luminância ou nível de cinza) é representada por $\mathbf{I}(p)$.

O EQM é um valor escalar que mede a diferença pixel a pixel entre as duas imagens \mathbf{I} e \mathbf{J} , sendo definido por

$$\text{EQM} = \frac{1}{M_i M_j} \sum_p (\mathbf{I}(p) - \mathbf{J}(p))^2 \quad \forall p \in \mathcal{D}_I \quad (2.2)$$

em que M_i e M_j representam o número de linhas e de colunas das imagens, respectivamente. A variável p representa as posições dos pixels nas imagens \mathbf{I} e \mathbf{J} . O problema com o uso dessa medida na avaliação da qualidade é que diferentes imagens submetidas ao mesmo tipo de degradação podem apresentar valores EQM similares. Isto não corresponde objetivamente à percepção do HVS, já que este avalia de forma bastante diferente fotografias desfocadas de uma paisagem e de um rosto.

Uma medida mais eficaz é obtida com o valor de PSNR entre as duas imagens. Essa medida é definida pela expressão

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}_I}{\text{EQM}} \right), \quad (2.3)$$

em que MAX_I representa o maior valor de intensidade da imagem \mathbf{I} . Trata-se de uma medida objetiva com baixo custo computacional. Por isso, o PSNR é utilizado, em geral, na avaliação da qualidade de vídeo. Além disso, esta medida aproxima melhor a avaliação subjetiva da qualidade de imagens do que o EQM para alguns tipos de degradação (SHEIKH; SABIR; BOVIK, 2006).

2.3 Estatística de coocorrência

A estatística de coocorrência é um método que tem bases na teoria da informação e busca descrever a relação entre subconjuntos, classes de elementos ou suas características. A forma mais elementar de coocorrência foi apresentado por Jaccard (1912). O coeficiente de Jaccard é uma estatística utilizada para comparar a similaridade ou diversidade entre dois conjuntos amostrais A e B , por meio da expressão

$$J(A, B) = \frac{\#\{A \cap B\}}{\#\{A \cup B\}}, \quad (2.4)$$

em que $\#\{\cdot\}$ é o operador de cardinalidade que determina a frequência de ocorrência dos elementos.

2.3.1 Matriz de coocorrência de níveis de cinza

Quando coeficientes de Jaccard são obtidos a partir de níveis de cinza de uma imagem, segundo algum critério de relação de vizinhança entre pares de pixels, obtém-se uma GLCM. A GLCM foi criada, inicialmente, para extração de

atributos visando descrever texturas em imagens (HARALICK; DINSTEIN, 1973). Entretanto, a matriz de coocorrência (*cooccurrence matrix*) pode ser explorada de diferentes formas. Trabalhos recentes têm empregado a matriz de coocorrência em aplicações como ferramenta de busca de limiares de segmentação (SAHOO; ARORA, 2004), mineração de dados (LEYDESDORFF; VAUGHAN, 2006), matriz de coocorrência generalizada para recuperação de imagens (PIANPIAN *et al.*, 2008) e matriz de coocorrência binária em classificação de tecidos (ROOMI; S.SARANYA, 2012).

A matriz de coocorrência \mathfrak{C} é uma matriz quadrada de dimensões $N_g \times N_g$ que mapeia o número de transições de intensidades entre pixels de um imagem \mathbf{I} , segundo um critério de vizinhança entre eles. Sejam $i, j = \{0, 1, \dots, N_g - 1\}$ as intensidades possíveis, os elementos da matriz \mathfrak{C} são definidos por

$$c_{ij} = \#\{(i, j) : q \in V_p(d, \theta), i = \mathbf{I}(p), j = \mathbf{I}(p + q)\} \quad \forall p \in \mathcal{D}_I, \quad (2.5)$$

em que V_p é uma função de vizinhança do pixel p , dado um deslocamento $d \in \mathbb{Z}^+$ e um ângulo θ em uma das quatro direções possíveis, ou seja, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. $\mathbf{I}(p + q)$ é uma translação de $\mathbf{I}(p)$ por q . O operador $\#\{\cdot\}$ denota a cardinalidade, ou seja, o número de elementos do conjunto.

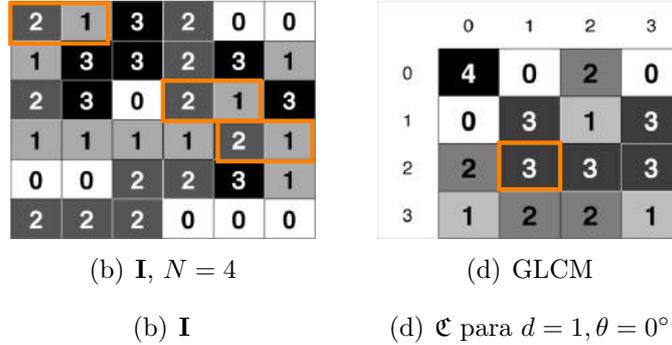


Figura 2.3: (a) Uma imagem contendo $N_g = 4$ níveis de cinza. (b) Exemplo de GLCM para a vizinhança de um pixel à direita. O elemento em destaque registra o número de transições do par de intensidades (2,1).

Uma matriz de coocorrência simétrica é obtida calculando-se as frequências de ocorrência de níveis de cinza de pixels separados tanto por d quanto por $-d$ para a mesma direção θ . Uma forma otimizada para obter matriz de coocorrência simétrica \mathfrak{C}^s é expressá-la como sendo

$$\mathbf{c}^s = \mathbf{c} + \mathbf{c}^T, \quad (2.6)$$

em que \mathbf{c}^T é a matriz transposta de \mathbf{c} .

A matriz de coocorrência simétrica \mathbf{c}^s é convenientemente normalizada de forma a aproximar densidades de probabilidade conjunta discreta dos valores de coocorrência de níveis de cinza. Uma normalização pode ser obtida apropriadamente por meio da divisão de cada elemento da matriz pelo total de ocorrências pareadas. Sendo assim, a matriz normalizada \mathbf{c}^n é obtida pela expressão

$$\mathbf{c}^n = \frac{\mathbf{c}^s}{\sum_{i,j=0}^{N_g-1} \mathbf{c}_{ij}^s}, \quad (2.7)$$

em que \mathbf{c}_{ij}^s representa o elemento da matriz simétrica \mathbf{c}^s .

2.3.2 Atributos da GLCM

A GLCM é uma forma compacta de representação da dependência espacial entre níveis de cinza de uma imagem segundo um critério de vizinhança que inclui um parâmetro de distância e outro de direção (GONZALEZ; WOODS, 2001). Haralick e Dinstein (1973) propuseram um conjunto de 28 atributos de textura extraídos da GLCM para descrever texturas em imagens. Muitos desses atributos são correlacionados entre si e acrescentam pouca capacidade de discriminação. Um subconjunto desses atributos, especificado como segue, é o mais frequentemente utilizado em aplicações práticas:

- ▶ Energia (*Ene*): energia, uniformidade ou segundo momento angular, é uma medida do nível total de intensidade. Seja \mathbf{c} a matriz normalizada e \mathbf{c}_{ij} seus elementos, a expressão da energia é definida como

$$Ene = \sum_{i,j=0}^{N_g-1} (\mathbf{c}_{ij})^2. \quad (2.8)$$

- ▶ Contraste (*Con*): mede o quanto a intensidade do pixel é similar à de sua vizinhança. Valores da GLCM concentrados na diagonal mostram que a imagem não possui contraste. O contraste aumenta à medida que os maiores pesos são encontrados fora da diagonal. O descritor de contraste é definido por

$$Con = \sum_{i,j=0}^{N_g-1} |i-j|^k (\mathbf{c}_{ij})^l, \quad (2.9)$$

em que $k \geq 1$ define o quão rápido a medida deve variar com a distância da diagonal. Para $k = 1$ e $l = 1$ essa medida se torna uma medida de dissimilaridade de primeira ordem.

- Homogeneidade (*Hom* ou *Mdi*): mede o quanto a distribuição dos elementos da GLCM está próxima da diagonal. Esta medida também é chamada de momento de diferença inverso (*Mdi*) e a expressão que a define é

$$Mdi = \sum_{i,j=0}^{N_g-1} \frac{\mathbf{c}_{ij}}{1 + |i-j|}, \quad (2.10)$$

em que $|\cdot|$ expressa o valor absoluto.

- Entropia (*Ent*): mede a aleatoriedade dos níveis de cinza e é expressa por

$$Ent = - \sum_{i,j=0}^{N_g-1} \mathbf{c}_{ij} \ln(\mathbf{c}_{ij}). \quad (2.11)$$

- Correlação (*Cor*): mede a probabilidade conjunta da ocorrência de pares de pixels da GLCM. Um valor absoluto elevado significa que existe uma forte dependência linear entre os pares de tons de cinza em uma determinada direção. Este atributo é descrito por

$$Cor = \sum_{i,j=0}^{N_g-1} \frac{ij\mathbf{c}_{ij} - \mu_i\mu_j}{\sigma_i\sigma_j}, \quad (2.12)$$

em que μ e σ representam, respectivamente, a média e o desvio padrão em uma linha i ou coluna j .

2.3.3 Complexidade computacional da GLCM

A complexidade computacional para o cálculo da GLCM é definida por $\mathcal{O}(M_i M_j N_g^2)$ em que M_i e M_j são as dimensões da imagem, N_g é o número de níveis de cinza considerados e $\mathcal{O}(\cdot)$ é o operador que define a complexidade de um algoritmo computacional em termos de quantidade de operações. Em geral, o

tamanho da imagem é muito maior do que o número de níveis de cinza. Portanto, a complexidade converge assintoticamente para $\mathcal{O}(M_i M_j)$.

2.4 Morfologia matemática

A MM é um arcabouço matemático que se baseia em um conjunto de transformações espaciais dentro de um mesmo domínio. A MM é uma teoria aplicada na análise de estruturas espaciais, como a forma de objetos, em imagens digitais (SERRA, 1982) (SOILLE, 2002). Frequentemente autores se referem às operações da MM como operações de filtragem morfológica. O uso da MM tem sido amplamente adotada em diversas aplicações desde inspeção industrial (CHOI; KIM, 2005) a diagnósticos médicos (COLOMÉ-SERRA *et al.*, 1992; ZANA; KLEIN, 2001; HANBURY, 2002).

As operações morfológicas básicas são definidas sob o escopo de um conjunto de operações geodésicas sobre uma imagem binária $\mathbf{I}(p) \in \{0,1\}$ ou imagem em níveis de cinza (*grayscale*) $\mathbf{I}(p) \in \{0,1,2,\dots,N_g - 1\}$ (VINCENT, 1993). Para simplificar, daqui por diante será utilizado apenas \mathbf{I} para representar uma imagem binária ou em níveis de cinza. Quando necessário, a diferenciação será referenciada no texto.

2.4.1 Elemento estruturante

As operações morfológicas atuam na análise local de estruturas conectadas segundo um critério de vizinhança. Esse critério é definido por um padrão de pixels dispostos espacialmente em um arranjo bidimensional chamado elemento estruturante planar (*flat*) $S_n : \mathcal{D}_S \subset \mathbb{Z}^2 \in \{0,1\}$, em que n representa sua escala. As dimensões do elemento estruturante, assim como a distribuição espacial da vizinhança, dependem da sua escala e forma, conforme ilustrado na Figura 2.4.

2.4.2 Dilatação e erosão

Dado um elemento estruturante planar (SOILLE, 2002) e simétrico S , a operação de dilatação morfológica consiste em encontrar a intensidade máxima dos pixels da imagem \mathbf{I} que coincidem com S centrado em $p \in \mathbf{I}$. Uma definição simplificada para a dilatação pode ser expressa por

$$\delta_S(\mathbf{I}) = \max\{\mathbf{I}(p + q) : q \in S\}, \quad (2.13)$$

em que $\mathbf{I}(p + q)$ é uma translação de \mathbf{I} por q .

Por definição, a erosão é uma operação dual da dilatação. Portanto, a operação

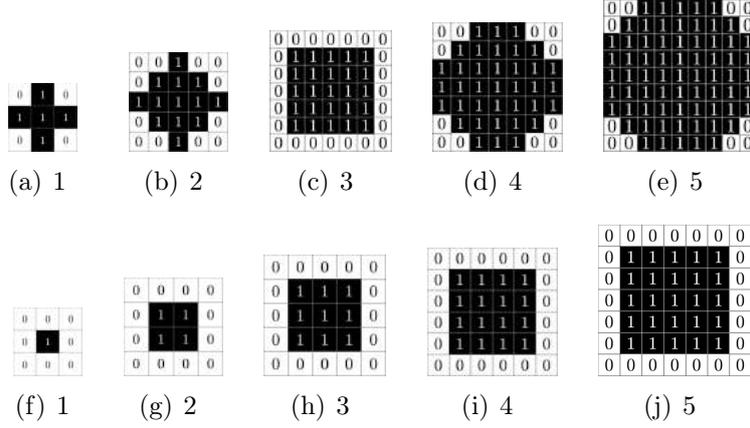


Figura 2.4: Elementos estruturantes simétricos (a-e) disco e (f-j) quadrado de diferentes escalas.

de erosão consiste em encontrar a intensidade mínima da região que coincide com o elemento estruturante S . A erosão pode ser definida pela expressão

$$\varepsilon_S(\mathbf{I}) = \min\{\mathbf{I}(p + q) : q \in S\}. \quad (2.14)$$

2.4.3 Abertura e fechamento

As demais operações morfológicas empregadas no contexto desta tese são derivadas de equações 2.13 e 2.14. A abertura $\gamma_S(\mathbf{I})$ e o fechamento $\phi_S(\mathbf{I})$ são operações morfológicas definidas, respectivamente, na equação 2.15 e equação 2.16. A abertura e o fechamento de uma imagem em níveis de cinza modificam seu contraste e morfologia por meio do preenchimento de pixels claros ou escuros, de acordo com a forma e tamanho do elemento estruturante e dos níveis de cinza da imagem.

$$\gamma_S(\mathbf{I}) = \delta_S(\varepsilon_S(\mathbf{I})) \quad (2.15)$$

$$\phi_S(\mathbf{I}) = \varepsilon_S(\delta_S(\mathbf{I})) \quad (2.16)$$

2.4.4 Top hat

O operador morfológico *top hat* é uma transformação que atua como detetor de picos ou vales. Esse operador é definido como

$$\tau_S(\mathbf{I}) = \mathbf{I} - \gamma_S(\mathbf{I}), \quad (2.17)$$

e seu dual, chamado *bottom hat*, é determinado pela expressão

$$b_S(\mathbf{I}) = \phi_S(\mathbf{I}) - \mathbf{I}. \quad (2.18)$$

Importante destacar o aspecto multiescala das operações de abertura e *top hat* no contexto da decomposição ou granulometria. A imagem da Figura 2.5 (b) ilustra uma operação de abertura aplicada sobre a imagem da Figura 2.5 (a), utilizando o elemento estruturante da Figura 2.4 (a). Observa-se que a abertura elimina os pixels que combinam com a escala do elemento estruturante, preservando apenas as estruturas de escalas maiores. Já a operação *top hat*, ilustrada na Figura 2.5 (c), produz um resultado complementar. Em outras palavras, a operação *top hat* preserva apenas as estruturas cuja escala é menor ou igual à do elemento estruturante.

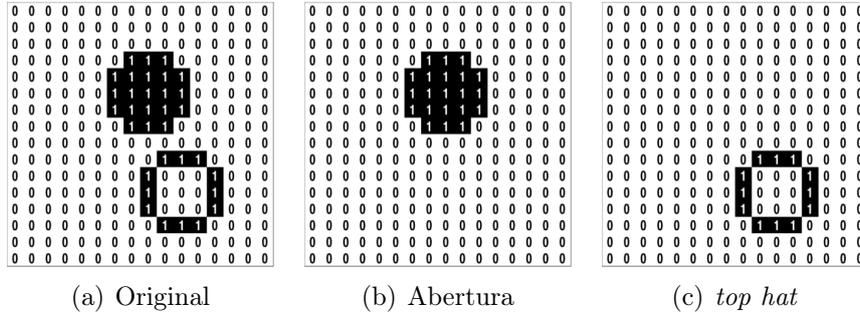


Figura 2.5: Exemplo de operações morfológicas sobre uma (a) imagem binária. Resultado das operações de (b) abertura e (c) *top hat*.

2.4.5 Granulometria

Também chamada de filtragem morfológica multiescala (SOILLE, 2002), a granulometria é uma forma de detetar estruturas e também de decompor formas binárias. Existem várias combinações possíveis para um processo de granulometria, as quais dependem do tipo, forma e escala de estrutura que se deseja analisar. A granulometria empregada nos exemplos desta tese baseia-se na expressão

$$\vartheta(\mathbf{I}) = \max \gamma_{S_n}(\mathbf{I}), \quad (2.19)$$

em que, S_n é um elemento estruturante de escala n .

2.4.6 Decomposição por limiar

A decomposição por limiar é um método clássico de decomposição de imagens em níveis de cinza, também chamada de corte no nível $0 < n < N_g - 1$, em que N_g é o número de níveis de cinza da imagem. Considere a imagem \mathbf{I} como um sólido cuja

altura da superfície é definida pelas intensidades dos pixels. Um corte de \mathbf{I} pode ser entendido como a interseção desse sólido com um plano paralelo em um nível n .

Para entender melhor a decomposição por limiar, considere $\mathbf{I}(p) \in \{0,1,2,3\}$ uma imagem em níveis de cinza em que $N_g = 4$. A operação de corte $\Phi(\mathbf{I})$ corresponde a um mapeamento do sólido limitado pelas intensidades de \mathbf{I} para uma superfície binária no domínio $\mathcal{D}_I \in \{0,1\}$. Diferentes cortes podem ser obtidos, conforme ilustrado na Figura 2.6. Os valores dos pixels da imagem \mathbf{I} estão representados na figura por cores que vão do azul escuro (menor intensidade) ao vermelho (maior intensidade). O operador corte

$$\Phi_n(\mathbf{I}) = \{p \in \mathbf{I} : \mathbf{I}(p) \geq n\}, \quad (2.20)$$

retorna uma imagem que contém apenas os pixels $p \in \mathbf{I}$ que satisfazem a condição $\mathbf{I}(p) \geq n$.

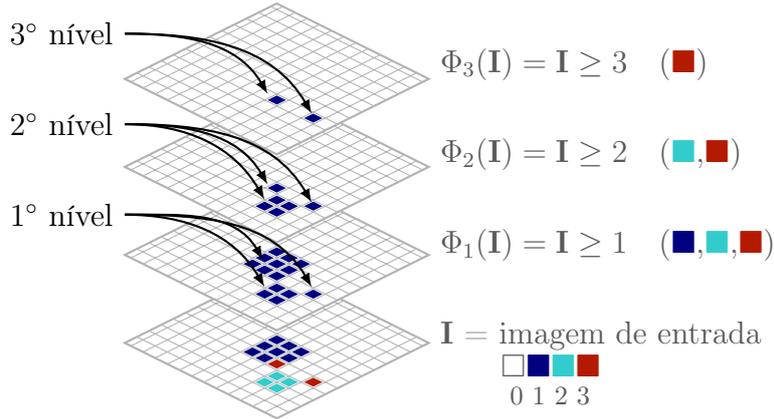


Figura 2.6: Modelo generalizado de uma decomposição por limiar de uma imagem em níveis de cinza.

A operação de cortes sucessivos resulta em uma pilha de imagens na qual cada nível contém apenas os pixels correspondentes que satisfazem a desigualdade. Este conceito é amplamente utilizado na MM para definir as operações elementares no domínio das imagens em níveis de cinza.

A partir desse conceito, duas conclusões podem ser obtidas. A primeira diz respeito à redefinição de uma imagem \mathbf{I} como sendo o somatório dos seus cortes

$$\mathbf{I} = \sum_{n=0}^{N_g-1} \Phi_n(\mathbf{I}). \quad (2.21)$$

A segunda conclusão é que qualquer imagem em níveis de cinza pode ser

representada topologicamente por uma decomposição multiescala, ou pilha de cortes (GONZALEZ; WOODS, 2001; VINCENT, 1993; MARAGOS, 1989).

2.4.7 Decomposição multiescala por *top hat*

A operação morfológica *top hat* pode ser utilizada para destacar estruturas como picos ou saliências presentes em uma cena. Escolhidas adequadamente, a forma e tamanho do elemento estruturante possibilitam controlar a escala de estrutura que se deseja destacar. A conectividade entre pixels vizinhos tem influência no resultado e, portanto, na escolha do formato do elemento estruturante.

Uma sequência de operações *top hat* com elementos estruturantes isotrópicos de escalas progressivas pode ser utilizada para decompor uma imagem em componentes de tamanhos específicos, que são componentes dos objetos presentes na cena. O uso de um elemento estruturante isotrópico apresenta a vantagem de combinar estruturas da imagem distribuídas em qualquer direção.

A operação de abertura, definida na equação 2.15, remove as partes menores do que S . A operação *top hat* da equação 2.17 isola essas partes removidas de tal forma que sua união com o resultado da abertura resulta na imagem original.

Uma decomposição multiescala por *top hat* está ilustrada no modelo da Figura 2.7. Importante destacar que, a partir de um determinado tamanho de S , as operações *top hat* passam a ser idempotentes se a escala da maior estrutura da imagem for menor que o S .

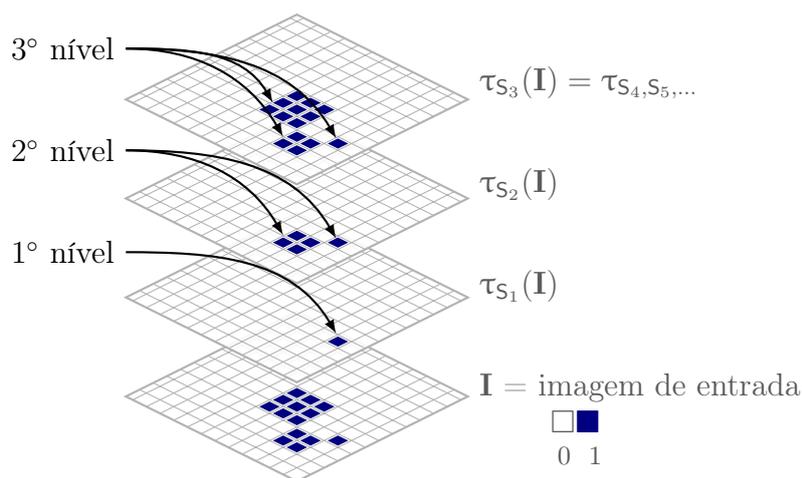


Figura 2.7: Modelo esquemático de uma decomposição multiescala por *top hat*.

Em uma operação de decomposição por *top hat*, o tamanho do elemento estruturante deve ser modificado para combinar com o tamanho das estruturas.

Em um processo como o da granulometria, o tamanho do elemento estruturante assume diversos valores dentro de uma faixa de tamanhos pré-estabelecida. A forma do elemento estruturante, por sua vez, exige uma escolha mais crítica. Em geral, deve-se escolher um elemento estruturante de acordo com a forma canônica¹ das estruturas presentes na imagem.

Um critério de parada ou de estabilização é necessário para definir quando interromper as sucessivas operações *top hat*. O critério mais simples consiste em monitorar as sucessivas operações e verificar quando não existe mais variação entre as imagem resultante de níveis adjacentes da decomposição.

2.4.8 Complexidade computacional da MM

Para análises realizadas com um elemento estruturante de tamanho fixo, em geral de dimensão bem menor do que a da imagem, as operações fundamentais da MM, erosão ou dilatação, possuem complexidade assintótica $\mathcal{O}(M_1M_2)$, em que M_1 e M_2 são as dimensões da imagem.

Contudo, as operações de decomposição utilizam elementos estruturantes que mudam de tamanho a cada iteração. Portanto, é conveniente apresentar a complexidade assintótica dessas operações por $\mathcal{O}(M_1M_2\lambda_1\lambda_2)$, em que λ_1 e λ_2 são as dimensões do maior elemento estruturante.

Além disso, a computação paralela baseada em vetores de processadores baseados em unidade de processamento gráfico (GPU, do inglês *Graphic Processing Unit*) possibilita a redução significativa do tempo de processamento (THURLEY; DANELL, 2012).

2.5 Conclusão

Neste capítulo foram apresentados os conceitos básicos da estatística de coocorrência, por meio da qual se estima uma matriz de probabilidades de coocorrência. A partir dessa matriz é possível extrair informações relevantes sobre uma imagem e os objetos presentes na mesma. Em particular, a matriz de coocorrência de níveis de cinza foi apresentada como forma de extrair atributos relativos à textura da imagem. Dependendo da forma como é construída a matriz de coocorrência, diferentes tipos de informação podem ser extraídas, seja para uso como descritores globais de textura ou como parâmetros para limiares de segmentação local.

¹com base em conhecimento prévio.

Também foi apresentada a teoria básica da Morfologia Matemática, com destaque para a decomposição de imagens por limiar e para as operações de incremento ou de superposição de efeitos.

No próximo capítulo serão apresentados métodos considerados o estado da arte em avaliação estrutural de imagens, especialmente envolvendo as tarefas de avaliação da qualidade de imagens e de decomposição de formas.

Estado da Arte da Análise Estrutural

Neste capítulo é apresentada uma revisão bibliográfica de métodos de análise estrutural de imagens recentes da literatura. As abordagens apresentadas compreendem métodos não referenciados, métodos referenciados e métodos baseados em transformação. No grupo de métodos não referenciados, é abordada a extração de atributos visando a caracterização de imagens ou regiões dela. O segundo grupo engloba a avaliação da qualidade de imagens em níveis de cinza e a avaliação do resultados de algoritmos de segmentação. Para o terceiro grupo, é apresentada uma metodologia comum ao realce e segmentação de estruturas pequenas ou delgadas baseada em morfologia matemática e no conceito da superposição de efeitos. Também é apresentada uma metodologia de análise por meio de decomposição.

3.1 Introdução

A análise estrutural de imagens compreende a interpretação de uma cena ou dos objetos nela contidos, a partir das estruturas que os compõem. O ponto de partida de alguns métodos de análise envolve transformações na imagem de entrada a fim de filtrar, realçar ou segmentar objetos de interesse. Em alguns casos, é possível definir descritores que são utilizados na segmentação dos objetos ou mesmo fornecer meios para a interpretação desses objetos ou da imagem como um todo. Descritores podem ser obtidos diretamente da intensidade dos pixels no mesmo domínio da imagem original, ou a partir de uma transformação dessa imagem para um domínio no qual certas estruturas e as relações entre elas são destacadas.

Embora existam diferentes abordagens de análise estrutural, nesta tese são destacadas aquelas baseadas em estatística de coocorrência e Morfologia Matemática (MM) por possuir relação direta com a metodologia desenvolvida. Para fins de

comparação, também foram incluídos no estudo métodos que disponibilizam meios para avaliação objetiva de resultados, seja por meio de um banco de dados ou por meio do código fonte da sua implementação computacional.

3.2 A análise estrutural não referenciada

A análise não referenciada de imagens é um processo que extrai da imagem de entrada a informação desejada, dispensando uso de imagens adicionais como referência. Diversos métodos que se baseiam na estatística de coocorrência, como os relacionados na Tabela 3.1, se enquadram nessa classe. Estes trabalhos introduzem métodos de extração de atributos ou abordagens baseadas em estatística de coocorrência para cálculo de similaridade entre descritores de objetos e são empregados em uma gama de aplicações que compreendem a segmentação, a classificação de cenas e a caracterização de objetos.

Tabela 3.1: Trabalhos de análise de imagens com base na estatística de coocorrência

Trabalho	Método e aplicação
Gadkari (2004)	Extração de atributos da matriz de coocorrência de níveis de cinza para avaliação da qualidade de imagens
Sahoo e Arora (2004)	Estimação de limiar através de histograma bidimensional para a segmentação de imagens
Rotshtein <i>et al.</i> (2008)	Extração de características por meio da relação espacial de segunda ordem para reconhecimento de faces
Nguyen-Duc <i>et al.</i> (2010)	Recuperação de imagens por meio de descritor baseado em coocorrência e transformada <i>contourlet</i> (decomposição piramidal)
Eleyan e Demirel (2011)	Atributos estatísticos da matriz de coocorrência para reconhecimento de faces
Yao e Yu (2011) e Yao e Yu (2013)	Extração de atributos por meio da interdependência espacial de características para reconhecimento de faces
Mittal, Moorthy e Bovik (2012)	Avaliação não referenciada da qualidade de imagens
Mitrea <i>et al.</i> (2012)	Atributos de ordem superior da matriz de coocorrência de níveis de cinza para caracterização de tumores no fígado
Shi <i>et al.</i> (2013)	Análise de textura por meio de atributos estatísticos e estruturais, com base na matriz de coocorrência de níveis de cinza, para caracterização de tumores na bexiga

A matriz de coocorrência de níveis de cinza (GLCM, do inglês *gray level*

cooccurrence matrix), introduzida no Capítulo 2, é uma ferramenta de avaliação não referenciada que tem sido utilizada para diferentes fins (GADKARI, 2004; MASCARO *et al.*, 2009; ELEYAN; DEMIREL, 2011; ASHA; BHAJANTRI; NAGABHUSHAN, 2011; MITREA *et al.*, 2012). Na Figura 3.1 é ilustrado o uso clássico da GLCM como ferramenta de análise de textura. Os mapas de textura ilustrados nas Figuras 3.1 (b) e (c) são representações da textura local em uma janela 3×3 . Por meio desses mapas é possível segmentar regiões de interesse como, por exemplo, áreas heterogêneas que contêm informação para a interpretação da cena.

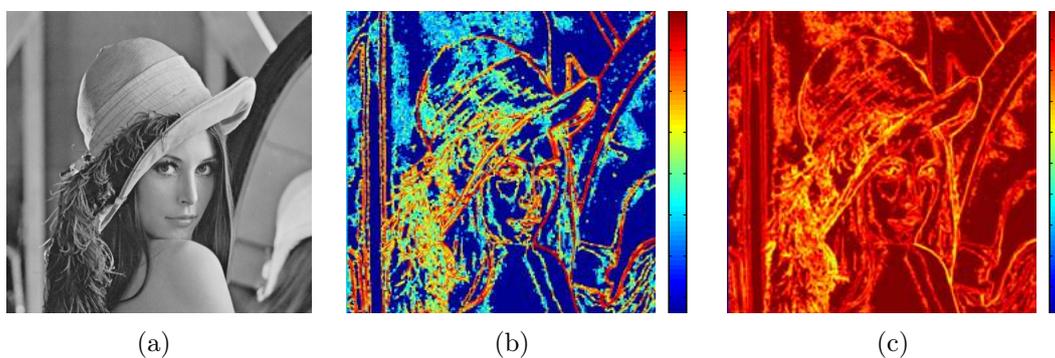


Figura 3.1: Exemplo da análise de textura com GLCM. (a) Imagem original. (b) Mapa de textura com o atributo de contraste. (c) Mapa de textura com o atributo de homogeneidade.

A detecção de manchas de óleo no mar, utilizando imagens de Radar de Abertura Sintética (SAR, do inglês *Synthetic Aperture Radar*), envolve a segmentação e posterior identificação por meio das características geométricas (FRATE *et al.*, 2000; SOLBERG *et al.*, 2004), ou por meio da análise de características de textura (ASSILZADEH; MANSOR, 2001; MERCIER; GIRARD-ARDHUIN, 2005; BREKKE; SOLBERG, 2005; SOLBERG *et al.*, 2004; RAMALHO; MEDEIROS, 2006). O conceito por trás dessa abordagem é a extração de características da imagem que contém um determinado tipo de mancha para sua posterior classificação utilizando Redes Neurais Artificiais (RNAs). Um conjunto de descritores de forma e textura é utilizado para caracterizar as regiões da imagem que podem ou não corresponder a um derramamento de óleo.

Choi e Kim (2005) introduziram um método para classificação de danos utilizando atributos morfológicos, a saber: textura, cor e forma. O método emprega um algoritmo de agrupamento para classificar danos causados por diferentes mecanismos de corrosão, dentre os quais: corrosão transgranular e intergranular,

corrosão localizada (*pitting*), corrosão por desgaste e corrosão uniforme.

Outras aplicações da GLCM compreendem a caracterização de tumores (MITREA *et al.*, 2012), o reconhecimento de faces (YAO; YU, 2011; YAO; YU, 2013; ROTSHTEIN *et al.*, 2008) e a análise da qualidade de imagens (GADKARI, 2004). O método robusto de reconhecimento de faces de Yao e Yu (2011) merece destaque porque apresenta um certo grau de similaridade com o método de avaliação estrutural proposto nesta tese. Yao e Yu (2011) propuseram o matriz de interdependência espacial de atributos (SFIM, do inglês *spatial feature interdependence matrix*) como forma de avaliar a similaridade pela distância χ^2 entre atributos extraídos de imagens contendo faces. Esses atributos, vale notar, não possuem relação com a avaliação estrutural proposta nesta tese, porque esta medida expressa a similaridade entre mapas de atributos bidimensionais extraídos da imagem de entrada e não a similaridade entre as estruturas dessas imagens.

O exemplo ilustrado na Figura 3.2 demonstra o uso da GLCM como ferramenta de análise da qualidade de imagens, de maneira similar ao proposto por Gadkari (2004). O resultado do experimento ilustrado pelas curvas da Figura 3.2 (c) não exhibe um padrão consistente com os diferentes níveis de degradação da imagem. De fato, o perfil parabólico das curvas mostra que, para intensidades diferentes de degradação determinado por σ , os atributos apresentam valores similares.

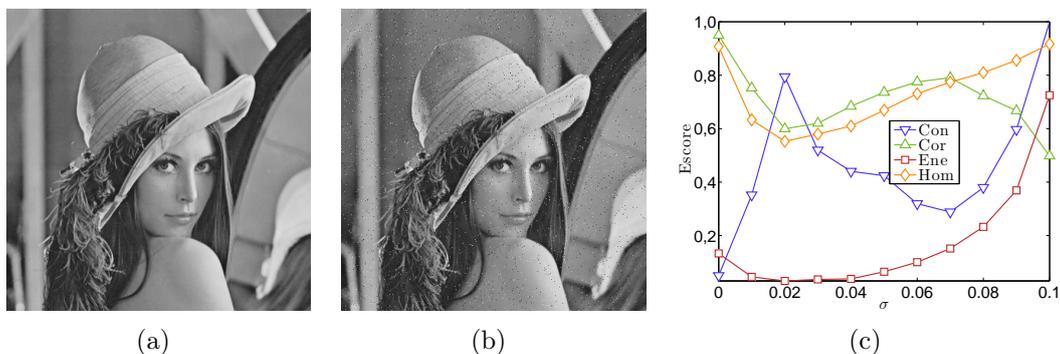


Figura 3.2: Exemplo da análise de qualidade com GLCM. (a) Imagem de referência. (b) Imagem contaminada com ruído gaussiano com $\mu = 0$ e $\sigma \in [0, 0,1]$. (c) Avaliação do nível de degradação com os atributos da GLCM.

Entretanto, por meio de uma análise visual das matrizes da Figura 3.3, percebe-se que a GLCM é capaz de expressar visualmente as diferenças estruturais em imagens que sofreram algum tipo de degradação. Notadamente, essas diferenças não aparecem refletidas significativamente nos valores dos atributos extraídos da GLCM. Importante notar que os atributos *Cor* e *Hom*, por exemplo, apresentam

inconsistências à medida que aumenta o nível de degradação. De fato, a primeira derivada de todos os atributos deste exemplo apresenta mudança de sinal, indicando que estes atributos produzem valores ambíguos.

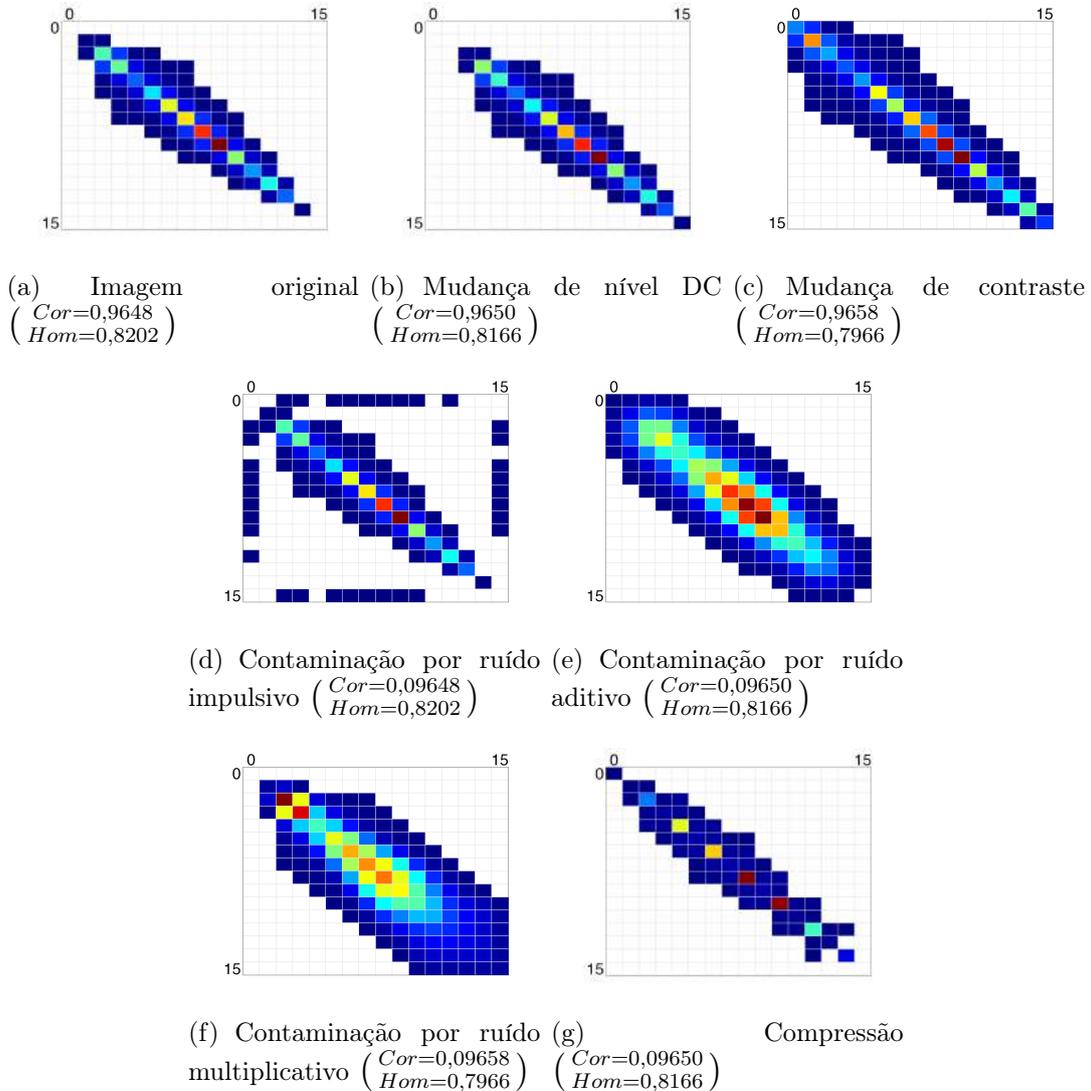


Figura 3.3: GLCM 16 níveis da imagem Lena 512 x 512 após diferentes degradações.

O método de avaliação não referenciada da qualidade de imagens proposto por Mittal, Moorthy e Bovik (2012), procura fornecer uma interpretação da qualidade de uma imagem avaliando somente a imagem degradada. Essa abordagem tem grande importância já que, na prática, as imagens que serão avaliadas sofreram algum tipo de degradação sem que uma cópia original tenha sido preservada. Entretanto, métodos não referenciados como o proposto por Mittal, Moorthy e Bovik (2012) requerem uma etapa de aprendizagem e, portanto, ficam restritos à disponibilidade

de um número razoável de imagens de treinamento e ao poder de generalização do algoritmo de aprendizagem utilizado.

3.3 A análise estrutural referenciada

No âmbito da avaliação referenciada, na qual uma das imagens de entrada é necessariamente uma referência conhecida, destacam-se os métodos de avaliação da qualidade por meio de análise estrutural. Diversas abordagens, resumidas na Tabela 3.2, baseiam-se no uso de modelos do sistema visual humano (HVS, do inglês *human visual system*), estatística de coocorrência e decomposição multiescala. Esses trabalhos apresentam métodos de análise recentes relacionados ao escopo desta tese. Uma parte representativa dos trabalhos da Tabela 3.2 foi incluída na pesquisa como forma de referência para validação da metodologia proposta.

Tabela 3.2: Trabalhos de avaliação da qualidade imagens com base na estatística de coocorrência

Trabalho	Aplicação
Wang e Bovik (2002a)	Índice de avaliação da qualidade de imagens por meio de parâmetros estatísticos locais
Wang, Simoncelli e Bovik (2003)	Avaliação de imagens por meio da similaridade estrutural multiescala
Wang, Luc e Bovik (2004)	Avaliação da qualidade de vídeo com base na similaridade estrutural
Sheikh, Bovik e Veciana (2005)	Modelo do HVS para avaliação da qualidade de imagens por meio de informação visual
Sampat <i>et al.</i> (2009)	Medida de qualidade de imagem com base em similaridade estrutural de decomposições por ondaletas complexas
Gao <i>et al.</i> (2009)	Análise geométrica de decomposições multiescala para avaliação da qualidade de imagens
Li e Bovik (2010)	Qualidade de vídeo por similaridade estrutural com base em conteúdo ponderado
Chetouania, Beghdadia e Deriche (2012)	Sistema híbrido supervisionado, baseado na escolha de algoritmos adaptados a determinadas degradações na avaliação da qualidade

A avaliação objetiva da qualidade de imagens não é uma tarefa trivial. Wang *et al.* (2004) reportam que essa tarefa vem sendo desenvolvida há décadas com base na medição de erro entre uma imagem de referência e sua versão degradada.

Porém, os mesmos autores relatam que as medidas convencionais não correspondem à percepção humana de qualidade visual de uma imagem e que uma análise da degradação de suas estruturas é mais efetiva, conforme mostram alguns estudos (WANG; BOVIK, 2002a; WANG; SIMONCELLI; BOVIK, 2003; SHEIKH; BOVIK; VECIANA, 2005; SHEIKH; BOVIK, 2006; ROUSE; HEMAMI, 2008; GAO *et al.*, 2009; WANG; BOVIK, 2009). Chetouania, Beghdadia e Deriche (2012) concluem que a eficiência dos algoritmos de medida de qualidade de imagem (IQM, do inglês *image quality measure*) dependem do tipo de degradação da imagem e que alguns deles estão mais adaptados ao julgamento subjetivo do que outros.

Sheikh, Sabir e Bovik (2006) realizaram um estudo estatístico de 10 algoritmos de avaliação referenciada que fornecem uma IQM, algumas delas estatisticamente indistinguíveis da medida de pico da relação sinal-ruído (PSNR, do inglês *peak signal-to-noise ratio*). Alguns algoritmos avaliados são baseados em propriedades psicofísicas do HVS, a saber, a sensibilidade ao contraste e a percepção do contraste. Os autores observaram que o PSNR é uma excelente medida não referenciada de qualidade para imagens degradadas por ruído branco e que uma parte dos algoritmos de IQM apenas conseguem aproximar seu desempenho global.

Nesta tese, são abordados os métodos referenciados que possuem relação com a análise estrutural, a saber, a similaridade estrutural (WANG *et al.*, 2004), sua variante multiescala (WANG; SIMONCELLI; BOVIK, 2003; ROUSE; HEMAMI, 2008) e a modelagem matemática do HVS (SHEIKH; BOVIK; VECIANA, 2005). Estes métodos foram escolhidos porque apresentam excelentes resultados globais, quando avaliados para a mesma base de imagens contendo diferentes tipos de degradação (SHEIKH; BOVIK, 2006).

3.3.1 Avaliação referenciada da qualidade

Na avaliação referenciada da qualidade de imagens, uma imagem de referência livre de distorções é comparada com sua versão distorcida. Portanto se enquadram na categoria de avaliação referenciada (FR, do inglês *full reference*). A avaliação referenciada da qualidade de imagens requer a definição de uma medida de similaridade bidimensional. Uma medida de distância frequentemente utilizada é a diferença média entre os níveis de intensidade de pixels correspondentes (WANG *et al.*, 2004).

Essa abordagem é útil quando se deseja avaliar o desempenho de algoritmos de compressão e filtragem por meio de um índice de comparação. Em aplicações

práticas, entretanto, nem sempre a imagem original está disponível, o que é a principal desvantagem dos métodos referenciados de avaliação da qualidade de imagens.

Similaridade estrutural

O método de avaliação de imagens proposto por Wang *et al.* (2004) foi desenvolvido para avaliar a qualidade de imagens monocromáticas, com isso se aproxima ao HVS em aplicações de avaliação da qualidade de imagens. Este método adota como modelo as observações de Marr (1982) sobre a formação das estruturas, conforme comentado no Capítulo 1. Outras versões foram propostas (WANG; SIMONCELLI; BOVIK, 2003; SHEIKH; BOVIK, 2006; PÉREZ *et al.*, 2012; GINESU; MASSIDDA; GIUSTO, 2006), tendo obtido resultados próximos ao HVS para diferentes aplicações.

No trabalho de Wang e Bovik (2002b), os autores buscam um modo de avaliar objetivamente a qualidade de imagens em substituição ao EQM. Em outro estudo, Wang *et al.* (2004) propõem uma métrica de avaliação da qualidade de imagens em níveis de cinza.

O índice de similaridade estrutural (SSIM, do inglês *structural similarity*), proposto por Wang *et al.* (2004), mede a similaridade entre duas imagens. Trata-se, portanto, de uma medida referenciada e que quantifica a distorção por compressão ou ruído, quando comparada à sua versão original não degradada. Segundo os autores, o algoritmo SSIM foi desenvolvido para melhorar o resultado da avaliação objetiva da qualidade de imagens tradicionalmente efetuada com o PSNR e EQM. Ainda segundo Wang *et al.* (2004), a diferença da medida fornecida pela SSIM para as medidas EQM e PSNR está no fato de que estas são medidas de erro, enquanto a SSIM considera a degradação da imagem como uma mudança estrutural.

A informação estrutural pressupõe a idéia de que os valores dos pixels possuem uma forte dependência entre si, especialmente quando eles estão espacialmente próximos. Esta dependência fornece uma informação significativa a respeito da estrutura dos objetos da cena.

A medida SSIM é calculada localmente em um janela quadrada de dimensões $M \times M$ centrada nos pixels $x \in \mathbf{I}$ e $y \in \mathbf{J}$. Wang *et al.* (2004) afirmam que $M = 8$ é uma boa escolha e que a medida SSIM é pouco influenciada pelo valor de M . A janela pode ser deslocada pixel a pixel ao longo da imagem. Entretanto, os autores propuseram o uso de um conjunto reduzido de possibilidades para reduzir a

complexidade computacional dos cálculos. A expressão que define a SSIM é

$$\text{SSIM}(x,y) = [l(x,y)]^\alpha [c(x,y)]^\beta [s(x,y)]^\gamma, \quad (3.1)$$

em que $l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}$ é a componente de luminância, $c(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}$ a componente de contraste e $s(x,y) = \frac{2\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$ a componente estrutural (correlação cruzada). $C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$ e $C_3 = C_2/2$ são variáveis para estabilização da divisão com um denominador fraco, L é a faixa dinâmica de intensidade dos pixels (ex. 255), $k_1 = 0,01$ e $k_2 = 0,03$ são valores padrões. Os valores μ_x , μ_y , σ_x e σ_y são, respectivamente, as médias e desvios padrões dos valores dos pixels das janelas $M \times M$ centradas em x e em y . O desvio padrão σ_{xy} é uma medida de variabilidade dos valores dos pixels entre as duas janelas.

O resultado obtido com a SSIM é um mapa de mesmas dimensões das imagens de entrada que apresenta localmente o grau de similaridade entre as mesmas. Quanto mais próximo de zero, maior a degradação estrutural presente em uma determinada região da imagem. Isto pode ser observado na Figura 3.4 (c). A escala de cores vai do azul escuro para valores de similaridade próximos de zero, até o vermelho escuro para valores de similaridade iguais a 1.



Figura 3.4: Exemplo de análise com a SSIM. (a) Imagem de referência. (b) Imagem contaminada com ruído sal e pimenta. (c) Mapa SSIM com $\text{MS-SSIM} = 0,81$.

Para obter um valor numérico da qualidade da imagem, utiliza-se o valor médio das janelas SSIM definido por

$$\text{MSSIM} = \sum_{x,y} \text{SSIM}(x,y)/(M_iM_j), \quad (3.2)$$

em que M_i e M_j representam o número de linhas e colunas das imagens, respectivamente. O resultado da equação 3.2 representa o nível de degradação da

imagem, como ilustrado na Figura 3.4.

Similaridade estrutural multiescala

Uma versão multiescala da SSIM, a similaridade estrutural multiescala (MS-SSIM, do inglês *multiscale structural similarity*), foi proposta por Wang, Simoncelli e Bovik (2003). A imagem de entrada é submetida a um filtro passa-baixa e sucessivas subamostragens por um fator de 2. Esse algoritmo produz um resultado com $N - 1$ iterações, em que N é a escala máxima. O valor final é o resultado do produto das componentes SSIM para cada escala, conforme a expressão

$$\text{MS-SSIM}(x,y) = [l_N(x,y)]^{\alpha_M} \prod_{j=1}^N [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}. \quad (3.3)$$

Importante destacar que o valor de luminância (ou intensidade) é calculado somente na escala N . Na prática, a equação 3.3 diminui a importância da degradação na componente de luminância com relação às demais componentes.

O valor MS-SSIM aproxima melhor a avaliação da qualidade, porque incorpora informação local em diferentes escalas. Por exemplo, o valor MS-SSIM para a imagem da Figura 3.4 (b) é 0,91, bem maior do que o obtido pela SSIM. Este valor aproxima melhor a percepção de que a degradação causada pelo ruído sal e pimenta, na intensidade do exemplo, não afeta substancialmente as estruturas da imagem.

Rouse e Hemami (2008) propuseram uma versão modificada do MS-SSIM, na qual as constantes $C_1 = C_2 = 0$, porém mantendo inalterados os demais parâmetros. Os autores realizaram experimentos mostrando que a sua versão modificada, MS-SSIM*, obteve um desempenho melhor na tarefa de estabelecer um limiar para o reconhecimento estrutural de imagens naturais. Vale ressaltar, entretanto, que a proposta dos autores resume-se a um caso particular da MS-SSIM.

Fidelidade da informação visual

Sheikh, Bovik e Veciana (2005) propuseram um modelo que aproxima funcionalmente a informação extraída pelo HVS de uma cena. Esse modelo tem por base a estatística de cenas naturais, por meio da mistura gaussiana de escalas $\mathcal{C} = \{\vec{C}_i : i \in I\}$, sendo I um conjunto de índices espaciais para um campo aleatório gaussiano. A expressão pode ser decomposta como $\mathcal{C} = \{S_i \vec{U}_i : i \in I\}$, em que S_i são escalares e \vec{U}_i são vetores gaussianos m -dimensionais. Considere $\vec{C}^N = \{\vec{C}_1, \vec{C}_2, \dots, \vec{C}_N\}$ e S^N descritores dos N elementos de \mathcal{C} . Sabendo que $\mathcal{E} = \mathcal{C} + \mathcal{N}$ e $\mathcal{F} = \mathcal{D} + \mathcal{N}'$ definem, respectivamente, a imagem de referência e

a imagem de teste para degradações \mathcal{N} e \mathcal{N}' , é possível quantificar a informação mútua (MACKAY, 2005) $I(\vec{C}^N; \vec{E}^N : S^N = s^N)$ e $I(\vec{C}^N; \vec{F}^N : S^N = s^N)$, sendo que s^N denota uma realização de S^N . Similarmente, E^N e F^N são elementos de \mathcal{E} e \mathcal{F} , respectivamente. A função $I(\cdot)$ quantifica a informação que pode ser extraída da saída do HVS pelo cérebro, quando uma imagem é observada. Importante destacar que as transformações e decomposições presentes neste modelo visam representar a forma como o HVS percebe as estruturas de uma imagem. Maiores detalhes do modelo podem ser encontrados no trabalho de Sheikh, Bovik e Veciana (2005).

Com base nesse modelo, Sheikh e Bovik (2006) propuseram a Informação de Fidelidade Visual (VIF, do inglês *Visual Information Fidelity*). A VIF fornece uma medida da qualidade de imagens calculada para uma coleção de coeficientes das decomposições por ondaletas (*wavelets*) (DAUBECHIES, 1992) em escalas que formam as subbandas da imagem, por meio da expressão

$$\text{VIF} = \frac{\sum_{j \in \text{subbandas}} I(\vec{C}^{N,j}; \vec{F}^{N,j} : s^{N,j})}{\sum_{j \in \text{subbandas}} I(\vec{C}^{N,j}; \vec{E}^{N,j} : s^{N,j})}. \quad (3.4)$$

Uma subbanda pode representar tanto a imagem completa quanto uma região espacialmente localizada dos coeficientes das decomposições por ondaletas. Portanto, a equação 3.4 pode ser empregada em uma janela da imagem ou na sua totalidade. No primeiro caso é obtido um mapa de qualidade da imagem. No segundo caso, obtém-se uma medida que quantifica a fidelidade de informação e que pode ser empregada como estimativa da qualidade de uma imagem degradada.

3.3.2 Avaliação referenciada da segmentação

A avaliação do resultados de algoritmos de segmentação está em pauta permanente porque, em geral, os métodos de análise que buscam extrair informação dos objetos de uma cena dependem de um bom pré-processamento e de uma boa segmentação. Diversas abordagens de avaliação, referenciada ou não, de algoritmos de segmentação vêm sendo estudados (ZHANG, 1996; ZHANG; FRITTS; GOLDMAN, 2008). Dentre esses métodos estão aqueles que visam resolver problemas específicos, como a avaliação referenciada da segmentação de retina (GEGÚNDEZ-ARIAS *et al.*, 2012). Alguns métodos são mais gerais, como o índice de Delves *et al.* (1992), desenvolvido para avaliar algoritmos de segmentação de imagens SAR e que, contudo, também pode ser utilizado para avaliar a segmentação

de pulmões em imagens de tomografia computadorizada (TC) (REBOUÇAS FILHO, 2013).

No contexto da avaliação estrutural, é importante destacar o método proposto por Gegúndez-Arias *et al.* (2012) projetado para avaliar algoritmos de segmentação de vasos de retina. Este método emprega o conhecimento prévio sobre a estrutura de rede de vasos, o que o torna um método especializado para o que se propõe. O método tem a propriedade de penalizar a descontinuidade dos vasos segmentados, diferença de diâmetro e tamanho da rede. Tais aspectos tornam este método mais objetivo na avaliação do que medidas clássicas que se baseiam na estatística de coocorrência.

O método de Gegúndez-Arias *et al.* (2012) é composto de três medidas: conectividade (C), área (A) e comprimento (L) da rede de vasos. A medida de conectividade dos vasos baseia-se na proporção entre componentes conexos das imagens segmentadas, conforme define a expressão

$$C(S, S_G) = 1 - \min \left(1, \frac{\#_C(S_G) - \#_C(S)}{\#(S_G)} \right), \quad (3.5)$$

em que $c(\cdot)$ é a operação que determina o número de componentes conexos, S é a imagem segmentada de teste e S_G é a imagem segmentada de referência ou padrão ouro (PO). $\#(\cdot)$ representa a cardinalidade do conjunto e $\min(a, b)$ retorna o menor valor entre a e b .

Para obter uma medida de avaliação da área total da rede de vasos segmentada, adota-se a expressão

$$A(S, S_G) = \frac{\#[(\delta_{S_\alpha}(S) \cap S_G) \cup (S \cap \delta_{S_\alpha}(S_G))]}{\#(S \cup S_G)}, \quad (3.6)$$

em que δ_{S_α} é a operação morfológica de dilatação com o elemento estruturante S_α . O comprimento total da rede é medido por meio da expressão

$$L(S, S_G) = \frac{\#[(\xi(S) \cap \delta_{S_\beta}(S_G)) \cup (\delta_{S_\beta}(S) \cap \xi(S_G))]}{\#[\xi(S) \cup \xi(S_G)]} \quad (3.7)$$

em que $\xi(\cdot)$ é uma operação morfológica de esqueletização homotópica (SERRA, 1982) e δ_{S_β} é a operação morfológica de dilatação com o elemento estruturante S_β , conforme definido na equação 2.13. Os valores de α e β nas operações de dilatação das equações 3.6 e 3.7, respectivamente, determinam o grau de tolerância a falhas na rede de vasos.

O índice de qualidade da segmentação de Gegúndez-Arias *et al.* (2012), CAL , é definido pelo produto das equações 3.5, 3.6 e 3.7, ou seja

$$CAL = C \cdot A \cdot L. \quad (3.8)$$

3.4 Análise por decomposição ou transformação

Trabalhos relevantes que abordam a análise de imagens por meio de decomposição ou transformação estão resumidos na Tabela 3.3. Estes trabalhos ilustram a busca pela descrição de objetos por meio de decomposição de formas em imagens binárias ou baseadas em transformações, comumente destinadas ao realce de estruturas em imagens em níveis de cinza. Embora essa lista inclua diferentes abordagens, nesta tese é dado destaque para aquelas que utilizam o arcabouço da MM visando o realce de saliências e a decomposição de imagens em níveis de cinza por meio de algoritmos não supervisionados.

3.4.1 Realce morfológico

Por realce entende-se a tarefa de filtrar ou destacar objetos de interesse. O realce tem relação direta com a transformação das intensidades dos pixels de imagens em níveis de cinza. Mukhopadhyay e Chanda (2000) apresentaram uma abordagem multiescala para realce de contraste baseado em MM. O método consiste em realizar a superposição de mudanças locais de contraste em diferentes escalas. A alteração de contraste local é obtida por meio de sucessivas operações *top hat* da equação 2.17, combinadas com operações *bottom hat* da equação 2.18.

No contexto da análise estrutural por transformações, dois métodos de realce por meio de operações morfológicas merecem destaque. O primeiro emprega uma operação morfológica que utiliza elementos estruturantes de tamanho pré-determinado, escolhidos para serem compatíveis com as estruturas que se deseja realçar na imagem. Esse método, às vezes chamado de granulometria, possui aplicação tanto na filtragem quanto no destaque de estruturas. O conceito básico por trás da granulometria está nas operações morfológicas erosão e dilatação, ou uma combinação entre elas, e no tamanho e forma dos elementos estruturantes (SOILLE, 2002).

Zana e Klein (2001) propuseram um método de segmentação de vasos de retina baseado em MM, utilizando soma de sucessivas filtragens por operações *top hat* para realçar retinografias. Este processo prepara a imagem para uma fase anterior à

Tabela 3.3: Trabalhos de análise por transformação.

Trabalho	Aplicação
Zana e Klein (2001)	Realce da rede de vasos de retina por meio de operações morfológicas
Soares <i>et al.</i> (2006)	Realce da rede de vasos de retina por meio de ondaletas Gabor
Hu e Tian (2006)	Realce de bordas em imagens binárias por meio de operadores morfológicos nebulosos
Shnayderman e Gusev (2006)	Avaliação da qualidade de imagens por meio da decomposição em valores singulares
Vizireanu (2007)	Decomposição de formas binárias por meio de operações morfológicas generalizadas
Stojić e Reljin (2010)	Realce de microcalcificações em mamogramas utilizando teoria multifractal e morfologia matemática
Welfer <i>et al.</i> (2010)	Segmentação do disco ótico em imagens de retina utilizando realce por operações <i>top hat</i> e filtragem gaussiana passa-baixa para eliminação de ruído
Pedrosa e Barcelos (2010)	Deteção de cantos por meio de transformada por ondaletas
Pedrosa, Batista e Barcelos (2013)	Recuperação de imagens com base em descritores baseados na deteção de saliências
Gibson <i>et al.</i> (2013)	Visão aumentada através do realce do contorno de objetos

segmentação, que se baseia na análise de curvatura da rede de vasos. A soma de *top hats* proposta por Zana e Klein (2001) utiliza elementos estruturantes direcionais cujo objetivo é buscar correspondências com a rede de vasos e, ao mesmo tempo, realizar uma superposição de efeitos. O tamanho do elemento estruturante incorpora o conhecimento prévio do diâmetro esperado da rede de vasos. O resultado é uma imagem realçada que preserva os aspectos morfológicos dos vasos e elimina artefatos que não coincidem com os elementos direcionais.

Zana e Klein (2001) e Mendonca e Campilho (2006) exploraram o conceito de superposição de efeitos por meio de operações *top hat* em algoritmos de segmentação de vasos sanguíneos em retinografias. Foram utilizados elementos estruturantes lineares direcionais com tamanho fixo para realçar a rede de vasos em uma etapa anterior à segmentação. Uma abordagem semelhante foi utilizada por Hu e Tian (2006) em um método de realce de bordas e filtragem de ruído simultâneos em uma

aplicação de segmentação de caracteres chineses. Entretanto, os autores empregaram uma operação de dilatação baseada na MM nebulosa (SHINHA; DOUGHERTY, 1992).

3.4.2 Avaliação da qualidade por decomposição em valores singulares

Considere a imagem \mathbf{I} como uma matriz genérica de dimensões $M_i \times M_j$ em que os elementos representam a intensidade dos pixels. A decomposição de valores singulares (SVD, do inglês *singular value decomposition*) (HORN; JOHNSON, 1991) dessa matriz consiste em satisfazer a expressão

$$\mathbf{I} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$$

em que \mathbf{U} é uma matriz unitária quadrada $M_i \times M_i$, \mathbf{S} é uma matriz retangular $M_j \times M_i$ de valores reais não negativos na diagonal e \mathbf{V} é uma matriz unitária quadrada $M_j \times M_j$, sendo \mathbf{V}^\top a sua transposta. Os elementos dispostos na diagonal de \mathbf{S} são chamados valores singulares de \mathbf{I} . As colunas de \mathbf{U} e de \mathbf{V} são vetores singulares à esquerda e à direita de \mathbf{I} , respectivamente.

Com base na SVD, Shnayderman e Gusev (2006) propuseram uma medida que representa o valor médio da avaliação local da imagem em blocos de dimensões $M \times M$, definida por

$$\text{MSVD} = \frac{\sum_{i=1}^{(n/k) \times (n/k)} |D_i - D_{mid}|}{\binom{n}{k} \times \binom{n}{k}} \quad (3.9)$$

em que D_{mid} representa o ponto médio do vetor D_i ordenado, $k = MM$ é o tamanho do bloco e $n = M_i M_j$ é o tamanho da imagem. O valor D_i é um caso particular da métrica de Minkowski $(\sum_k |s_k - \hat{s}_k|)^{1/\beta}$, para $\beta = 2$, ou seja

$$D_i = \sqrt{\left[\sum_{i=1}^k (s_i - \hat{s}_i)^2 \right]}, \quad (3.10)$$

em que, s_i e \hat{s}_i são as componentes S da SVD obtidas com a imagem de referência e com a imagem de teste, respectivamente.

O algoritmo MSVD é semelhante ao da SSIM, tanto na representação gráfica quanto nos parâmetros utilizados. Entretanto, a complexidade assintótica do MSVD é de ordem $\mathcal{O}(n^3)$, em que n é o número de pixels da imagem, enquanto a complexidade da SSIM é $\mathcal{O}(n^2)$. Embora seja baseado em decomposição singular,

ao contrário da abordagem estrutural da SSIM, o resultado também é bastante semelhante, entretanto com um custo computacional elevado. Shnayderman e Gusev (2006) afirmam que o emprego de blocos 8×8 , similarmente utilizado na SSIM, reduz significativamente o tempo de processamento. Apesar disso, aproximadamente dois terços do tempo de processamento de uma imagem é utilizado apenas no cálculo das decomposições dos blocos.

A medida MSVD apresenta um mapa de erro nas mesmas dimensões da imagem de entrada. Esse mapa de erro fornece uma informação similar àquela disponibilizada pelo mapa da SSIM, conforme ilustrado na Figura 3.5. Contudo, o valor numérico obtido com a equação 3.9 é não limitado dentro de uma faixa regular, o que dificulta a sua interpretação. Para este exemplo, os valores da escala de cores vão de 0 (sem degradação) a 59,72 (degradação mais elevada) e podem variar de uma imagem para outra assim como entre imagens idênticas que foram submetidas a degradações distintas.

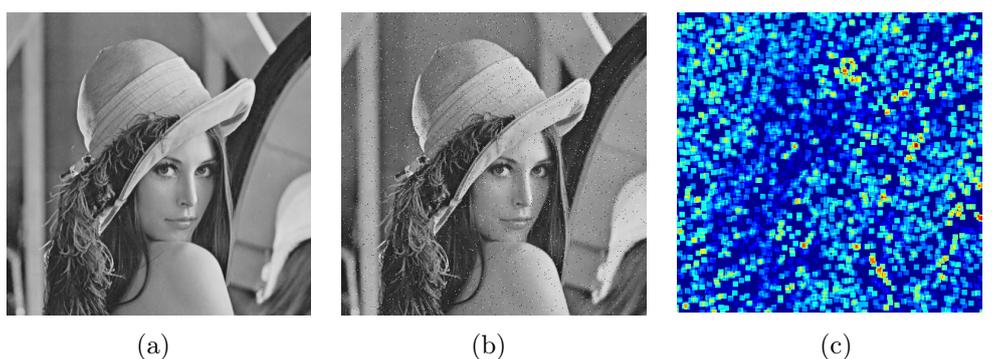


Figura 3.5: Exemplo da análise com MSVD. (a) Imagem de referência. (b) Imagem contaminada com ruído sal e pimenta. (c) Mapa MSVD com valor médio 59,72.

3.5 Conclusão

Neste capítulo foram apresentados os conceitos básicos sobre métodos empregados em análise estrutural de imagens no domínio espacial.

Os métodos de análise apresentados tentam aproximar uma característica do sistema visual humano (HVS, do inglês *human visual system*), que é a capacidade de extrair informação estrutural relevante das cenas. O conceito de relevância depende do contexto mas, em geral, está relacionado tanto com o realce de estruturas pequenas e finas quanto com a filtragem de estruturas pequenas causadas por algum tipo de degradação.

Em alguns processos de análise apresentados, é necessário realizar transformações na imagem para obter uma imagem realçada ou segmentada, facilitando sua compreensão. Esse tipo de transformação realça saliências visuais, como o contorno de objetos e elimina estruturas indesejáveis, como o ruído.

A análise estrutural também é realizada a partir da extração direta ou indireta de características presentes na imagem. Algumas dessas características, ao contrário da forma e escala de objetos, podem não ser facilmente mensuráveis como a textura e as distorções estruturais causadas por degradação.

No próximo capítulo é apresentada uma nova proposta de análise estrutural utilizando a estatística de coocorrência.

Matriz de Interdependência Espacial

Este capítulo introduz um método para análise de imagens construído sobre conceitos básicos da estatística de coocorrência. O método de análise de imagens por meio da Matriz de Interdependência Espacial (MIDE) baseia-se na estatística de coocorrência entre os níveis de cinza de duas imagens. Trata-se, portanto, de um método de análise referenciado.

Os atributos extraídos da MIDE, a saber, correlação, momento de diferença inverso, entropia e coeficiente chi-quadrado, permitem que a análise seja realizada em um domínio diferente daquele das imagens de entrada, o que ajuda a abstrair a interpretação contextual das imagens e dar ênfase às alterações estruturais entre elas. Além disso, fornece uma interpretação visual das alterações estruturais.

4.1 Introdução

Considere uma imagem de referência **I** e uma imagem a ser analisada **J**. A análise estrutural de **J** por meio da MIDE consiste em verificar se as estruturas presentes na imagem de referência são encontradas na imagem analisada ou de teste.

O princípio básico do método é apresentado na Figura 4.1. Para cada pixel da imagem de referência, é verificado se existe pixel de intensidade correspondente na imagem de teste. As coocorrências são contabilizadas para cada nível de cinza entre as duas imagens e armazenadas em uma matriz de frequências $N_g \times N_g$, em que N_g corresponde ao número de níveis de cinza considerados. Admitindo que as estruturas de uma imagem são formadas pela relação espacial de intensidades de uma vizinhança de pixels, essa matriz de frequências representa as coincidências e as diferenças estruturais entre **I** e **J**.

O método de avaliação apresentado neste capítulo foi desenvolvido com o

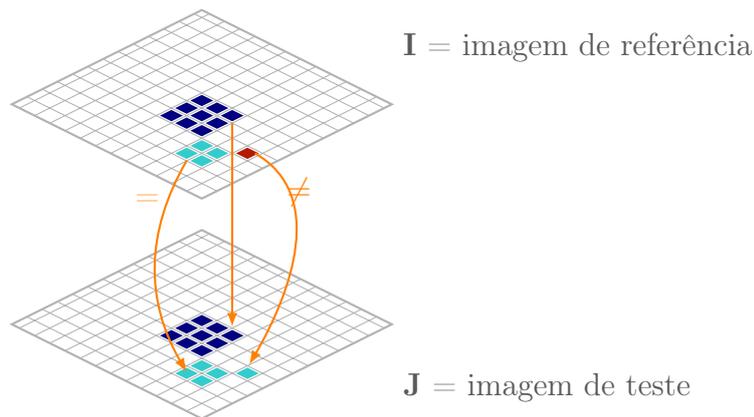


Figura 4.1: Modelo esquemático da comparação de intensidades das estruturas de duas imagens

propósito de:

- i. permitir uma análise em um espaço diferente daquele da imagem, utilizando uma representação visual na forma de matriz de dependência espacial das estruturas;
- ii. extrair informações da matriz de frequências que representem as alterações estruturais na imagem analisada;
- iii. possibilitar um ajuste de sensibilidade às alterações estruturais, realizado por meio do controle de quantização dos níveis de cinza.

4.2 Matriz de interdependência espacial

A dependência espacial é a base teórica utilizada por Haralick e Dinstein (1973) para a análise de textura em imagens, por meio da matriz de coocorrência de níveis de cinza (GLCM, do inglês *gray level cooccurrence matrix*). A dependência espacial baseia-se na coocorrência de níveis de cinza de acordo com a vizinhança dos pixels dentro de uma imagem. A interdependência espacial consiste em uma adaptação desse conceito para a análise estrutural entre duas imagens. A forma básica da MIDE não depende de direção e distância como a GLCM, mas necessita de duas imagens de entrada.

A matriz de interdependência de níveis de cinza (MIDE), portanto, é uma matriz de frequências construída a partir de duas imagens por meio da contagem da coocorrência dos pares de intensidade de pixel entre elas. O modelo apresentado na Figura 4.2 exemplifica o cálculo da MIDE, sendo \mathbf{I} a imagem de referência e

\mathbf{J} a imagem de teste. A matriz representada na Figura 4.2 acumula a frequência de ocorrência das intensidades dos pares de pixels espacialmente correspondentes nas imagens de entrada. As frequências relativas às estruturas similares entre \mathbf{I} e \mathbf{J} ficam registradas na diagonal da matriz ou próximas a ela. Para imagens idênticas, a MIDE resulta em uma matriz diagonal. As diferenças estruturais são registradas fora da diagonal da MIDE.

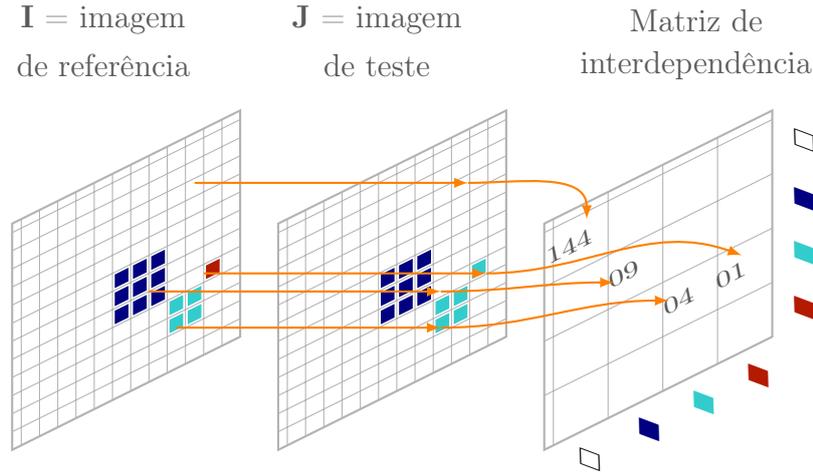


Figura 4.2: Modelo esquemático para obtenção da interdependência espacial

Considere \mathbf{I} e \mathbf{J} as imagens de referência e de teste, respectivamente, definidas no domínio \mathcal{D}_I . Sejam $i, j \in \{0, 1, \dots, N_g - 1\}$ as intensidades dos pixels, os elementos \mathbf{m}_{ij} de \mathfrak{M} são determinados por

$$\mathbf{m}_{ij} = \#\{(i, j) : i = \mathbf{I}(p), j = \mathbf{J}(p)\} \quad \forall p \in \mathcal{D}_I, \quad (4.1)$$

em que $\#\{\cdot\}$ é a cardinalidade do conjunto. Essa forma básica da MIDE é um caso particular da relação espacial entre p e q em que ambos correspondem a uma mesma posição no domínio \mathcal{D}_I das imagens.

Importante notar que, se o número total de níveis de cinza considerados no cálculo de \mathfrak{M} for menor do que N_g , obtém-se uma quantização das intensidades, ou seja, um mapeamento de intensidades $\{0, 1, \dots, N_g - 1\}$ para $\{0, 1, \dots, N\}$. Já que a MIDE baseia-se no princípio de que uma estrutura é formada por intensidades similares em uma vizinhança de p , essa quantização não altera a interpretação da MIDE e ainda reduz o custo computacional do seu cálculo.

Quando utilizada como medida de similaridade, é conveniente tornar a MIDE uma matriz simétrica por meio da expressão

$$\mathfrak{M}^s = \mathfrak{M} + \mathfrak{M}^\top, \quad (4.2)$$

em que \mathfrak{M}^\top representa a transposição de \mathfrak{M} . Neste caso, $\mathfrak{M}^s(\mathbf{I}, \mathbf{J}) = \mathfrak{M}^s(\mathbf{J}, \mathbf{I})$.

Uma aproximação das probabilidades de coocorrência entre intensidades é obtida com a normalização de \mathfrak{M}^s por meio da expressão

$$\mathfrak{M}^n = \frac{\mathfrak{M}^s}{\sum_{i,j=0}^{N_g-1} m_{ij}^s}, \quad (4.3)$$

em que m_{ij}^s são os elementos de \mathfrak{M}^s . Os elementos de \mathfrak{M}^n são representados por m^n .

Para simplificar a notação, doravante a matriz \mathfrak{M}^n será representada simplesmente por \mathfrak{M} . Entretanto, computacionalmente pode ser conveniente trabalhar na forma da matriz da equação 4.1 ou da equação 4.2.

Os elementos m_{ij} da matriz \mathfrak{M} normalizada estão no intervalo $[0,1]$. Portanto, cada elemento fornece uma estimativa da probabilidade de ocorrência de níveis de cinza entre pixels correspondentes a uma mesma posição espacial nas imagens de entrada. Este conceito é similar ao apresentado por Sahoo e Arora (2004) para obter um histograma bidimensional, utilizando a coocorrência entre a imagem original \mathbf{I} e sua média $\bar{\mathbf{I}}$.

Exemplos de análise por meio da MIDE são apresentados nas Figuras 4.3 e 4.4. A Figura 4.3 ilustra um caso em que duas imagens idênticas são comparadas. Observa-se que a MIDE resultante tem todas as frequências de coocorrência de níveis de cinza concentradas na sua diagonal, conforme ilustrado na Figura 4.3 (c). Este exemplo demonstra um casamento perfeito entre a estruturas das duas imagens.

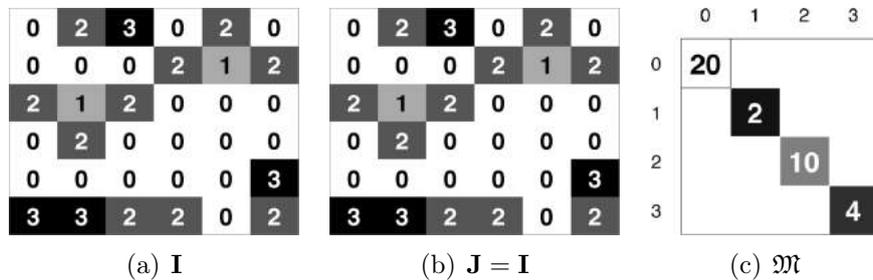


Figura 4.3: Exemplo de matriz de interdependência entre as imagens \mathbf{I} e \mathbf{J} , sem alterações estruturais.

Quando existem diferenças estruturais, como ilustrado no destaque da Figura

4.4, a MIDE correspondente apresenta uma dispersão dos valores de coocorrência que se afastam da diagonal. Quanto maior a dispersão em torno da diagonal, maiores as diferenças estruturais entre as imagens de entrada.

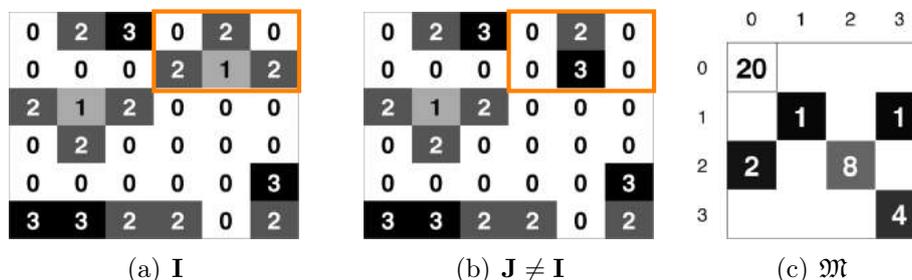


Figura 4.4: Exemplo de matriz de interdependência entre as imagens \mathbf{I} e \mathbf{J} , com alterações estruturais em \mathbf{J} . A região que sofreu mudança estrutural aparece em destaque.

Importante observar que a degradação estrutural presente na imagem de teste \mathbf{J} é perceptível na MIDE e independe do contexto da cena. A análise por meio da MIDE, portanto, elimina influências subjetivas às quais um observador humano não está imune.

A MIDE concentra a informação estrutural conforme ilustrado no gráfico da Figura 4.5. No quadrante I estão as frequências de coocorrência entre estruturas escuras da imagem. Em geral, estas estruturas não são objeto de interesse imediato e podem, eventualmente, ser descartadas da análise. De fato, a frequência das transições de nível de cinza correspondentes ao fundo da imagem são elevadas, como pode ser observado no valor do elemento \mathfrak{m}_{00} da matriz na Figura 4.4 (c). Portanto, em alguns casos, podem influenciar negativamente no resultado numérico da análise. No quadrante III ficam registradas as frequências das estruturas claras, as quais geralmente representam as saliências, ou seja, os objetos de interesse na imagem. Nos quadrantes II e IV estão as frequências relativas a regiões de contraste entre estruturas, ou seja, registram as diferenças entre as duas imagens. Para uma MIDE simétrica, os quadrantes II e IV possuem o mesmo significado. Entretanto, o significado pode variar quando a MIDE assimétrica é analisada. Importante notar que os quadrantes não possuem necessariamente as mesmas dimensões, porque a concentração das frequências na matriz depende do nível médio de intensidade das imagens de entrada.



Figura 4.5: Quadrantes da MIDE.

4.2.1 Informação extraída da MIDE

As medidas extraídas da MIDE, úteis para avaliação da relação estrutural entre duas imagens são definidas como segue.

Correlação

A correlação é um atributo da MIDE que mede a dependência linear entre os tons de cinza dos pixels da imagem analisada, que são espacialmente correspondentes com os pixels da imagem de referência. Considerando que a organização espacial desses pixels formam estruturas, admite-se que este atributo é uma medida da similaridade estrutural entre duas imagens. Pequenas variações de intensidade na vizinhança de pixels, pertencentes a uma mesma estrutura, afetam pouco a correlação.

$$\text{COR} = \sum_{i,j=0}^{N_g-1} \mathfrak{m}_{ij} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right] \in [-1,1]. \quad (4.4)$$

em que μ e σ são, respectivamente, a média e o desvio padrão dos valores de \mathfrak{M} para cada linha i e coluna j .

Uma correlação próxima de 1 é obtida quando as maiores frequências de coocorrência estão concentradas na diagonal principal da MIDE. À medida que as frequências mais elevadas se afastam da diagonal principal, a correlação tende a diminuir. Esta medida é sensível a variações de intensidade dos pixels, o que inclui mudanças no valor médio ou no contraste da imagem.

Momento de diferença inverso

O Momento de Diferença Inverso (MDI) é um momento de primeira ordem que mede o quanto os valores da matriz estão concentrados em torno da diagonal. A dispersão dos valores da matriz está relacionada a variações de intensidade dos pixels em uma vizinhança. Os pesos \mathbf{m}_{ij} decaem quando se afastam da diagonal principal e isto acontece quando a imagem analisada não possui estruturas espacialmente correspondentes àquelas da imagem de referência. O MDI mede o quão afastados estão os pesos da diagonal da matriz.

$$\text{MDI} = \sum_{i,j=0}^{N_g-1} \frac{\mathbf{m}_{ij}}{1 + (|i - j|)^{1/k}} \in [0,1], \quad (4.5)$$

em que k é um número inteiro diferente de zero. O valor de k determina o grau de importância dos pesos mais afastados da diagonal da matriz. Para valores positivos $k > 1$, o valor de MDI decai exponencialmente. O uso de $k = 1$ é suficiente para estimar o nível de distorção, sem elevar a complexidade computacional.

O valor da medida é mais próximo de zero quando a imagem de teste possui variações espaciais bruscas, como no caso da presença de ruído ou quando essa imagem é proveniente de uma cena diferente. Isto corresponde a mudanças estruturais na imagem de teste com relação à imagem de referência. De fato, as discordâncias entre as estruturas ficam registradas nos pesos mais afastados da diagonal.

Entropia

A entropia está associada à variabilidade dos pesos da MIDE simétrica. Se as estruturas variam espacialmente e em intensidade entre as imagens analisadas, os pesos da MIDE estarão bastante dispersos. Em outras palavras, a entropia definida por

$$\text{ENT} = \sum_{i,j=0}^{N_g-1} \mathbf{m}_{ij} [-\log(\mathbf{m}_{ij})], \quad (4.6)$$

mede o grau de desordem entre as estruturas das duas imagens.

Chi-quadrado

O atributo chi-quadrado mede o desvio entre as frequências de coocorrência pertencentes à diagonal da MIDE e o valor esperado dessas frequências quando

as imagens de entrada são idênticas. Por motivo de padronização, nesta tese essa medida é representada por CHI, embora formalmente sua representação seja χ^2 . Considere os conjuntos que contêm as frequências observadas e frequências esperadas da MIDE **assimétrica**, respectivamente definidos por $O_i = \{\mathbf{m}_{ij} : i = j\}$ e $E_i = \sum_{j=0}^{N_g-1} \{\mathbf{m}_{ij} : j \geq i\}$, sendo $i = \{0, \dots, N_g - 1\}$. As frequências observadas em O são sempre menores ou iguais às frequências esperadas em E . Quando as imagens de entrada são idênticas, o valor de CHI é zero. Portanto, o atributo CHI pode ser entendido como uma medida da independência estrutural entre as imagens. A expressão para $\text{CHI} \in [0,1]$ é definida por

$$\text{CHI} = \sum_{i=0}^{N_g-1} \frac{(O_i - E_i)^2}{E_i}. \quad (4.7)$$

4.2.2 Interpretação dos atributos da MIDE

A imagem da Figura 4.6 (a) corresponde a uma região de interesse que contém diferentes informações estruturais e que formam objetos distintos, porém bem definidos. As Figuras 4.6 (b-f) foram obtidas a partir da aplicação do ruído gaussiano de média $\mu = 0$ para diversos valores de desvio padrão σ . Os valores PSNRn e ENTn são as medidas PSNR e ENT normalizados para a faixa $[0,1]$ fazendo-as coincidir com a mesma escala dos demais atributos. O comportamento dos atributos COR, MDI, ENT e CHI, para este exemplo em particular, é ilustrado nas curvas Figura 4.6 (g). Quando comparada com a medida clássica pico da relação sinal-ruído (PSNR, do inglês *peak signal-to-noise ratio*), os atributos da MIDE apresentam um comportamento similar, ou seja, variam bruscamente e depois tendem a um valor limite conforme ocorre o aumento da degradação da imagem. Contudo, diferente do PSNR que apresenta pouca variação a partir de um certo nível de degradação, ambas as medidas extraídas da MIDE sofrem uma modificação gradual e significativa com o aumento da intensidade do ruído.

Algumas observações adicionais podem ser obtidas a partir da análise da Figura 4.6. Os atributos apresentam uma variação brusca à medida em que aumenta a degradação e depois convergem para um valor assintótico. A curva COR sugere seu uso como medida de similaridade, porque depende menos da intensidade da degradação do que as outras medidas. A medida MDI, entretanto, é mais sensível e pode ser utilizada para estimar o nível de degradação da imagem. A curva ENT condiz com o aumento da degradação, como esperado, podendo representar perdas

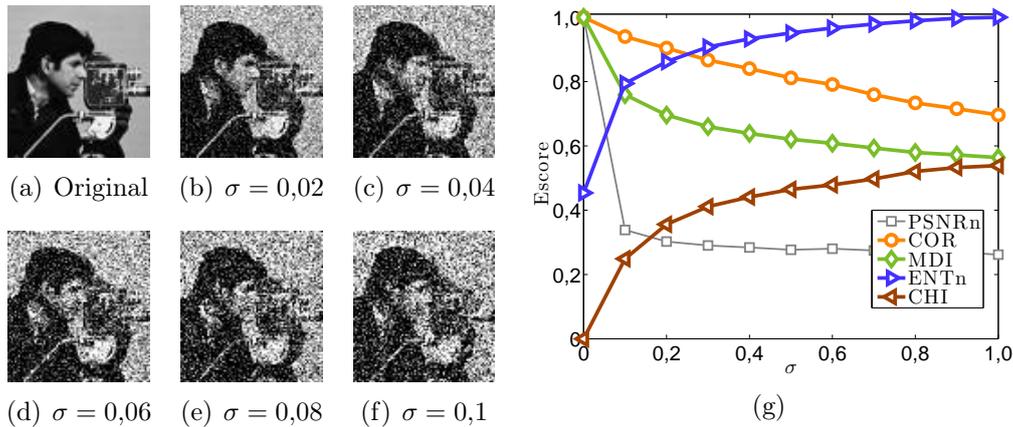


Figura 4.6: (a) Imagem original. (b-f) Ruído gaussiano com $\mu = 0$ para diferentes valores de σ . (g) Comparação entre PSNRn, COR, MDI, ENTn e CHI.

estruturais. A curva do atributo CHI, por sua vez, sugere uma distância estrutural entre as imagens de referência e de teste. A independência estrutural entre as imagens cresce com o aumento do nível de degradação.

4.2.3 Interpretação visual da MIDE

Além das medidas extraídas da MIDE, a própria matriz proporciona uma forma de interpretação visual das diferenças estruturais entre as imagens de entrada. Duas imagens sintéticas, uma em níveis de cinza e outra binária, são utilizadas para ilustrar o significado visual da MIDE. Ambas imagens sofreram sucessivas distorções com diferentes graus de intensidade e as medidas extraídas da MIDE foram registradas.

O exemplo da Figura 4.7 ilustra a variação de COR e MDI para uma imagem em nível de cinza após sucessivos incrementos no seu nível DC. O aspecto das imagens degradadas aparece na Figura 4.7 (b), numeradas de acordo com o incremento no seu nível DC. A imagem 1 é idêntica à imagem original, ou seja, não contém distorções. Na Figura 4.7 (c) são exibidas as respectivas MIDEs. O aumento de nível DC causa deslocamento das frequências para fora da diagonal da matriz. De fato, as estruturas da imagem da Figura 4.7 (a) não estão mais presentes nas imagens da Figura 4.7 (b), numeradas de 2 a 9, para os mesmos níveis de cinza.

Um observador atento poderá não perceber diferenças entre as imagens 1, 2 e 3, ou entre as imagens 8 e 9 da Figura 4.7 (b). Contudo, o registro da MIDE na Figura 4.7 (c) não deixa dúvidas de que existe algum nível de distorção do nível DC nas imagens 2 e 3, assim como permite perceber a diferença de degradação entre as

imagens 8 e 9. Esta interpretação é reforçada quando se observa as curvas da Figura 4.7 (d).

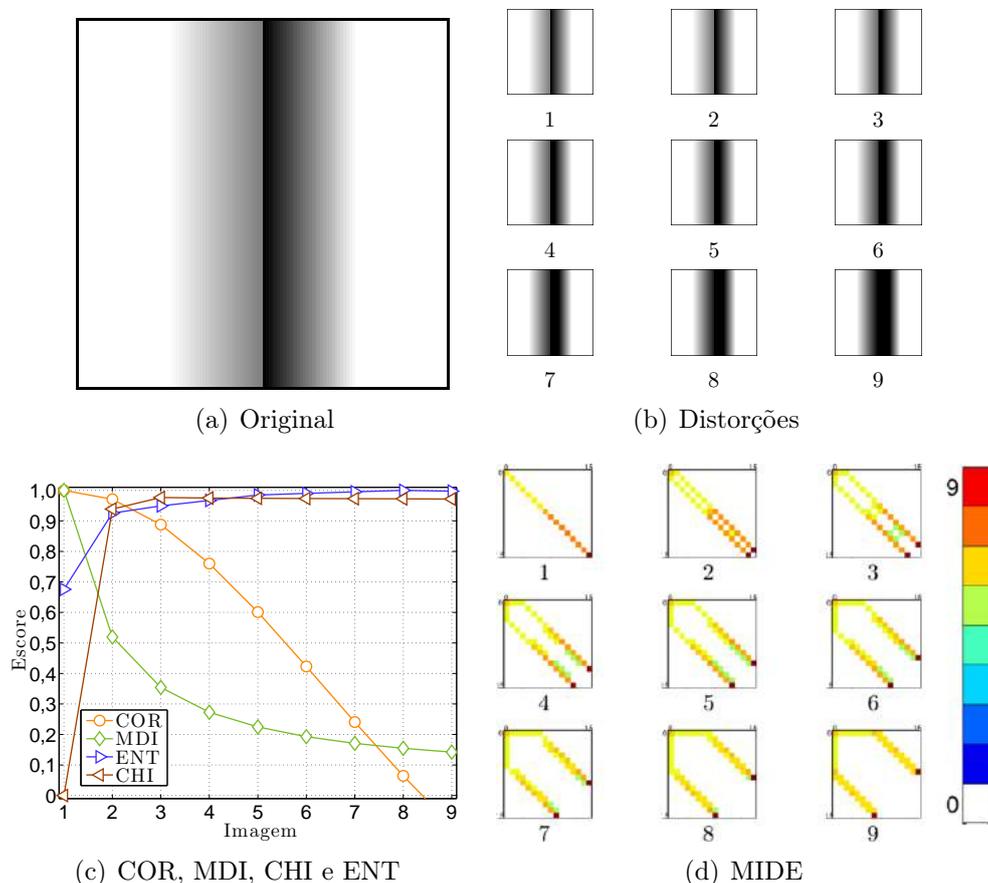


Figura 4.7: Imagem em nível cinza e diferentes níveis de variação de nível DC.

Observa-se, ainda, que a medida COR decai significativamente somente a partir do nível 3 de degradação. Isto coincide com a percepção visual de que as imagens 4 a 9 apresentam diferenças estruturais com relação à imagem original. Por exemplo, a região escura das imagens 4 a 9 possui largura significativamente maior do que aquela da imagem 1 da Figura 4.7 (b).

A percepção da distorção varia para observadores diferentes (WANG; BOVIK, 2002b). Por exemplo, caso o observador se concentre na parte clara das imagens, poderá discordar da informação fornecida pela COR. Neste caso, nota-se que as regiões claras permanecem sem grande variação em suas dimensões em todas as imagens. Por outro lado, ao observar as imagens da Figura 4.7 (c), não restam dúvidas de que existe uma diferença significativa entre as imagens. Mais que isso, apenas observando o aspecto da MIDE é possível inferir o grau de degradação das

imagens. Os valores COR e MDI do gráfico da Figura 4.7 (d) aproximam as duas interpretações. A progressiva dissimilaridade entre as imagens 2 a 9 comparadas com a imagem 1 é definida pelo decaimento aproximadamente linear de COR.

O exemplo da Figura 4.8 ilustra a variação de COR e MDI para uma imagem em níveis de cinza depois de sofrer contaminação por ruído Poisson. Neste exemplo, a contaminação por ruído é adicionada cumulativamente a cada novo nível de degradação.

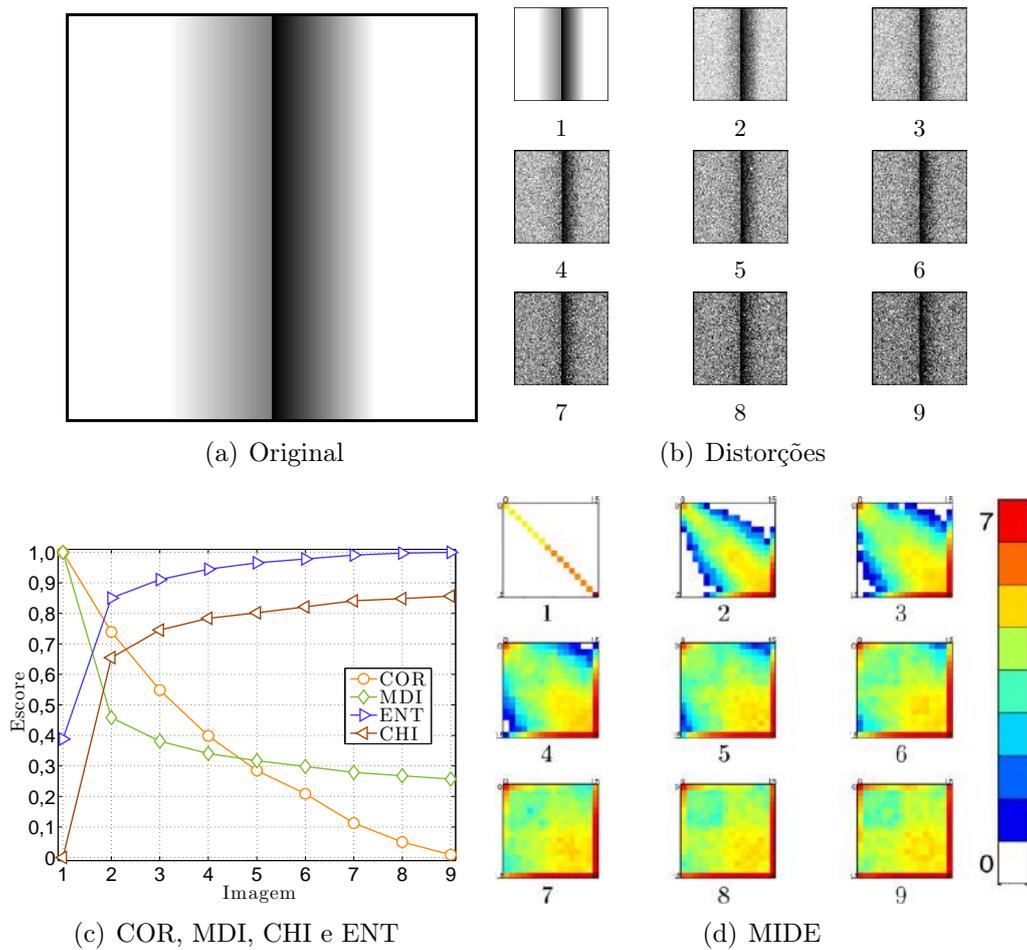


Figura 4.8: MIDE imagem nível cinza e diferentes níveis de degradação por ruído Poisson.

Imagens binárias

As imagens binárias podem ser consideradas como um caso particular das imagens em níveis de cinza. Entretanto, as ferramentas de análise empregadas para imagens em níveis de cinza nem sempre podem ser aplicadas em imagens binárias. No caso do uso da MIDE em imagens binárias, os resultados numéricos produzem interpretações diferentes daquelas obtidas com imagens em níveis de cinza. Além

disso, a interpretação visual da MIDE pode não ser única quando avaliada por observadores distintos. O principal motivo é que a MIDE resultante da análise de imagens binárias possui apenas dois níveis e, portanto, uma interpretação visual fica prejudicada quanto aos detalhes.

O exemplo da Figura 4.9 ilustra a variação de COR e MDI para uma imagem binária após sofrer contaminação por ruído sal e pimenta (S&P). Para imagens binárias, a medida COR apresenta boa sensibilidade às distorções, perceptível na variação de valores. Por outro lado a medida MDI é bem menos sensível do que no caso das imagens em níveis de cinza. Adicionalmente, percebe-se que, para imagens binárias a MIDE não possibilita uma interpretação visual objetiva sem a observação explícita dos valores dos atributos extraídos da matriz.

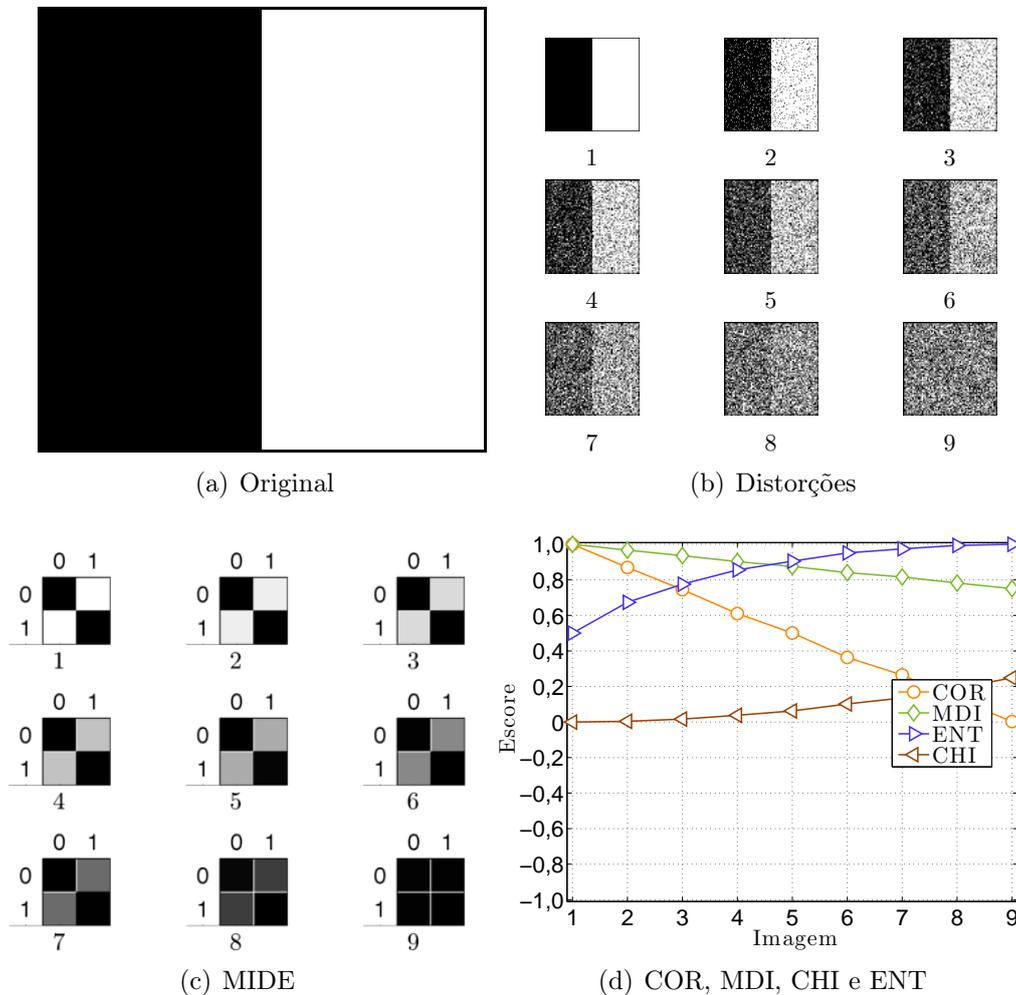


Figura 4.9: MIDE imagem binária. Sensibilidade da MIDE ao ruído S&P.

Embora as medidas COR, MDI, ENT e CHI apresentem correlação entre si,

quando combinadas podem fornecer novos significados. Por exemplo, a combinação $COR \times MDI$ ou $COR + MDI$ resulta em diferentes interpretações de acordo com o tipo de degradação analisada.

4.3 Extração de características

A MIDE pode ser aplicada como extrator de características quando as imagens de entrada são a imagem original (ou imagem de referência) e uma cópia dessa imagem degradada artificialmente. Em geral, uma suavização é suficiente para obter um conjunto de características da imagem por meio dos atributos da MIDE. Assume-se, portanto, que as saliências da imagem são estruturas de interesse que, degradadas por uma operação de suavização, podem ser detetadas pelos atributos da MIDE.

Textura ou rugosidade representam saliências comumente avaliadas em análise de imagens médicas (WU *et al.*, 2013), detecção de superfícies atacadas por corrosão (CHOI; KIM, 2005), detecção de manchas de óleo (RAMALHO; MEDEIROS, 2007) e detecção de falhas em tecidos (ROOMI; S.SARANYA, 2012; ASHA; BHAJANTRI; NAGABHUSHAN, 2011). Os exemplos desta seção ilustram o uso da MIDE nesse tipo de aplicação.

4.3.1 Descritores estruturais

Por ser um método de análise referenciado, a MIDE exige duas imagens de entrada. Contudo, é possível realizar a análise de uma imagem sem necessariamente ter disponível uma referência.

A suavização de uma imagem elimina detalhes que, em geral, estão relacionadas a altas frequências ou alto contraste. Portanto, o processo de suavização atenua essas saliências e resulta em uma degradação estrutural da imagem que é proporcional ao nível de suavização e à quantidade e escala das estruturas. Essa degradação pode ser medida pela MIDE comparando-se a imagem de teste com a sua versão suavizada.

Seja uma imagem \mathbf{I} e sua versão suavizada $\mathbf{J} = \mathcal{S}(\mathbf{I})$, sendo $\mathcal{S}(\cdot)$ uma função de suavização qualquer. A matriz \mathfrak{M} é capaz de representar a degradação estrutural das saliências de \mathbf{J} através do vetor de atributos $\mathbf{A} = \{COR, MDI, CHI, ENT\}$. Estes atributos são sensíveis à degradação de \mathbf{J} e formam um conjunto de descritores que podem ser utilizados para caracterizar uma imagem ou uma região da mesma.

O exemplo ilustrado na Figura 4.10 foi construído com o conjunto de atributos da MIDE sobre um subconjunto da base de texturas de Lazebnik, Schmid e Ponce (2005), dividido em três classes, a saber: madeira (classe A), água (classe B) e

granito (classe C). Foram utilizados os quatro descritores da MIDE com $N = 16$ para a imagem de entrada e sua versão borrada com o filtro da média em uma janela 3×3 . Como pode ser observado na projeção bidimensional do espaço de descritores representada pela matriz-U (Apêndice A) da Figura 4.10 (b), é possível discriminar as três classes de textura sem dificuldade. Essa matriz-U ilustra quão próximos entre si estão os elementos de uma mesma classe e também mostra regiões vazias que separam classes vizinhas.

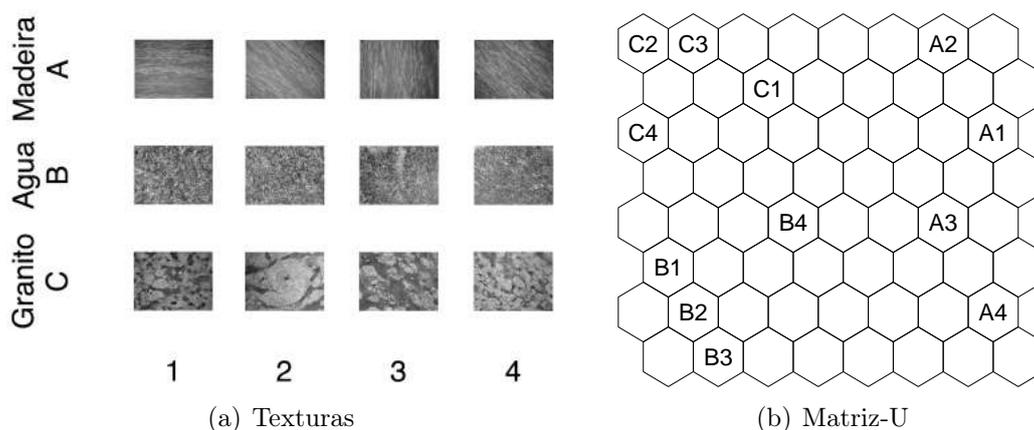


Figura 4.10: Exemplo de texturas representadas pelo vetor de atributos extraído da MIDE. (a) Texturas separadas em classes por linha. (b) Espaço de descritores identificando agrupamentos de texturas similares.

Vale destacar que a função de suavização $\mathcal{S}(\cdot)$ deve ser escolhida de tal forma que apenas as estruturas de interesse sejam degradadas. De modo geral, as saliências são representadas por estruturas pequenas, com relação aos objetos presentes na cena. Portanto, uma função de suavização pode ser obtida com a média da vizinhança dos pixels ou um borramento gaussiano de média zero e variância unitária, ambos em janelas $M \times M$. O tamanho M da janela depende somente do tamanho das estruturas. Como a MIDE é bastante sensível a degradações estruturais, $M = 3$ é uma escolha adequada, tendo em vista obter o menor custo computacional possível com esta operação.

4.4 Semelhança estrutural entre imagens

A correlação da MIDE fornece um índice de semelhança estrutural entre duas imagens **I** e **J**. Este índice mede o nível de dependência entre níveis de cinza que formam as estruturas da imagem.

Para imagens idênticas, dois atributos da MIDE assumem valores unitários (COR

e MDI) e um terceiro assume valor nulo (CHI). À medida que aparecem diferenças estruturais, ambos atributos apresentam variações significativas. A sensibilidade, entretanto, é diferente e depende da imagem de entrada.

O exemplo da Figura 4.11 ilustra a variação dos atributos para uma imagem binária depois de sofrer distorções geométricas. Como pode ser observado na Figura 4.11, a COR apresenta robustez à presença de ruído e outros tipos de degradação de intensidade. Entretanto, o tipo de degradação, assim como a sua intensidade, podem fazer com que a COR, eventualmente, retorne valores muito próximos para imagens distintas.

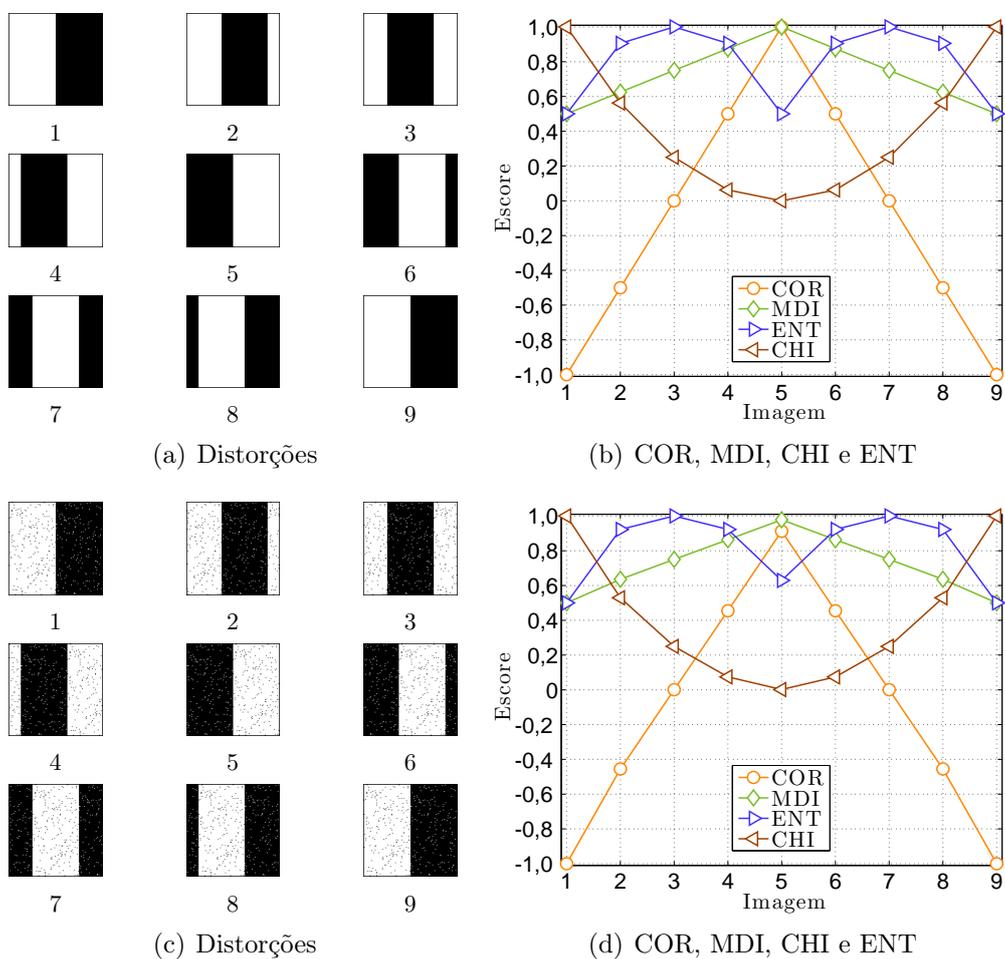


Figura 4.11: Sensibilidade da MIDE a imagens binárias semelhantes. (a,b) Sem ruído. A imagem 5 é a referência. (c,d) Com ruído.

O valor COR difere da correlação linear de Pearson r , definida pela equação A.14, porque esta última é uma estatística de primeira ordem. Em outras palavras, r não leva em consideração a distribuição espacial dos valores dos pixels. Por outro

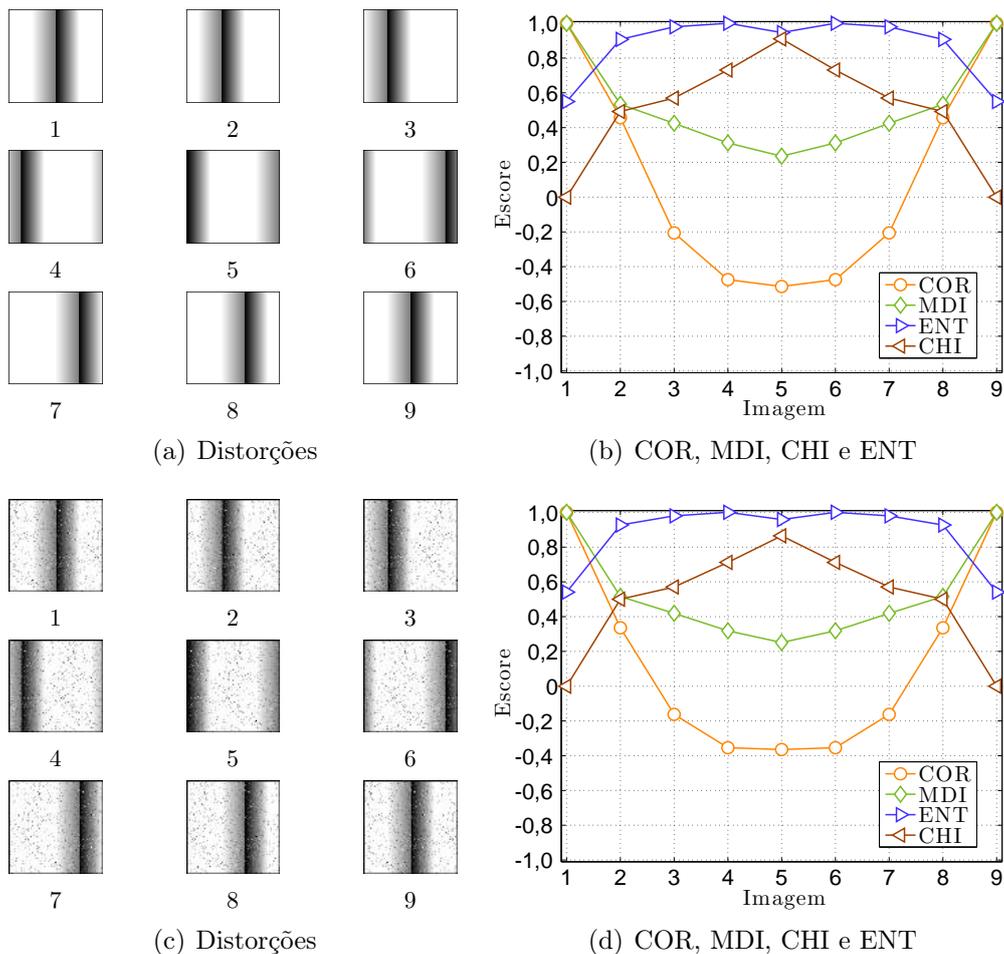


Figura 4.12: Sensibilidade da MIDE a imagens em níveis de cinza semelhantes. (a,b) Sem ruído. A imagem 1 é a referência. (c,d) Com ruído.

lado, ao contrário da correlação estatística, a COR é sensível a pequenas variações do valor médio (ou nível DC) e contraste.

Considere a imagem em níveis de cinza \mathbf{I} da Figura 4.13 (a) e uma versão degradada \mathbf{J} formada a partir da redução pela metade do valor de uma quantidade determinada de pixels de \mathbf{I} . A intensidade da degradação varia de 0% a 100% da quantidade total de pixels da imagem. Como a correlação estatística não leva em consideração a distribuição espacial dos valores dos pixels, o valor de r não é capaz de expressar adequadamente as diferenças estruturais presentes nas versões degradadas ilustradas nas Figuras 4.13 (b-f). Isto pode ser observado no comportamento de r , registrado no gráfico da Figura 4.13 (g). Importante destacar que a medida de entropia das diferenças entre as imagens, definida como $ent(|\mathbf{I}-\mathbf{J}|) = -\sum(h \log(h))$, em que h é um vetor normalizado das frequências de ocorrência dos níveis de cinza

do argumento de ent , também difere da medida ENT da MIDE, como apresentado no gráfico da Figura 4.13 (g). Em ambos os casos, as medidas r e ent apresentam informação ambígua para níveis de degradação bastante distintos, ao contrário dos atributos da MIDE.

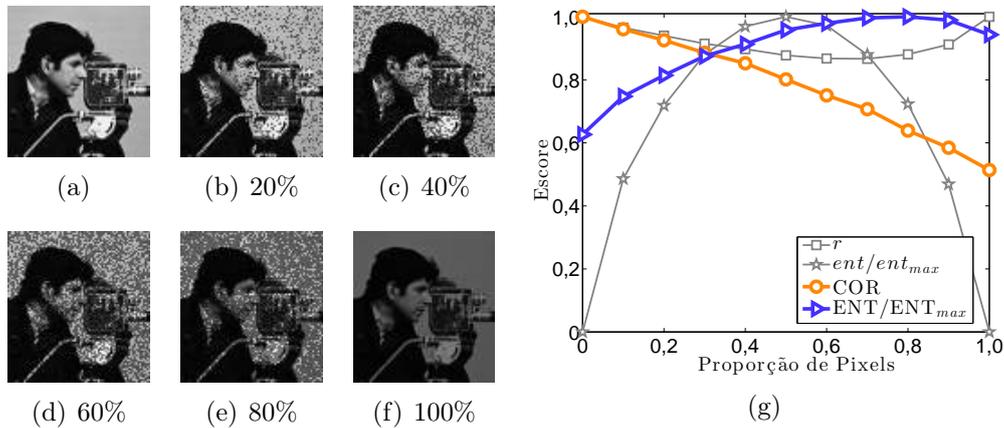


Figura 4.13: (a) Imagem original. (b-f) Nível DC modificado. (g) Comparação entre $r \times$ COR e $ent \times ENT$. As medidas de entropia são apresentadas normalizadas.

4.5 Avaliação da qualidade

Muitas aplicações que envolvem captura e transmissão de imagens requerem o uso de algum tipo de avaliação da qualidade. Em geral, são utilizados o EQM ou o PSNR como medidas de avaliação objetiva do grau de degradação da imagem, quando comparada com sua versão original. Entretanto, é comum que imagens submetidas a tipos diferentes de degradação possuam o mesmo valor de EQM, conforme ilustrado na Figura 4.14 (WANG; BOVIK, 2002a).

A MIDE é uma ferramenta que permite obter indicadores do grau de degradação dessas imagens. Na Figura 4.15 são apresentadas as matrizes \mathfrak{M} e alguns de seus atributos que representam cada degradação da Figura 4.14. Percebe-se que, numericamente os atributos COR e MDI expressam a intensidade da degradação. Além disso, o aspecto visual da MIDE apresenta diferença entre os tipos de degradação. Por exemplo, as Figuras 4.15 (e) e (f) apresentam uma dispersão em torno da diagonal causada pela contaminação por ruído aditivo e multiplicativo, respectivamente. Numericamente a diferença não é tão significativa. Contudo, a percepção visual da matriz deixa claro que o ruído speckle afeta com maior intensidade as regiões mais claras da imagem, enquanto o ruído gaussiano afeta a imagem como um todo. Os valores COR e MDI da imagem degradada com ruído

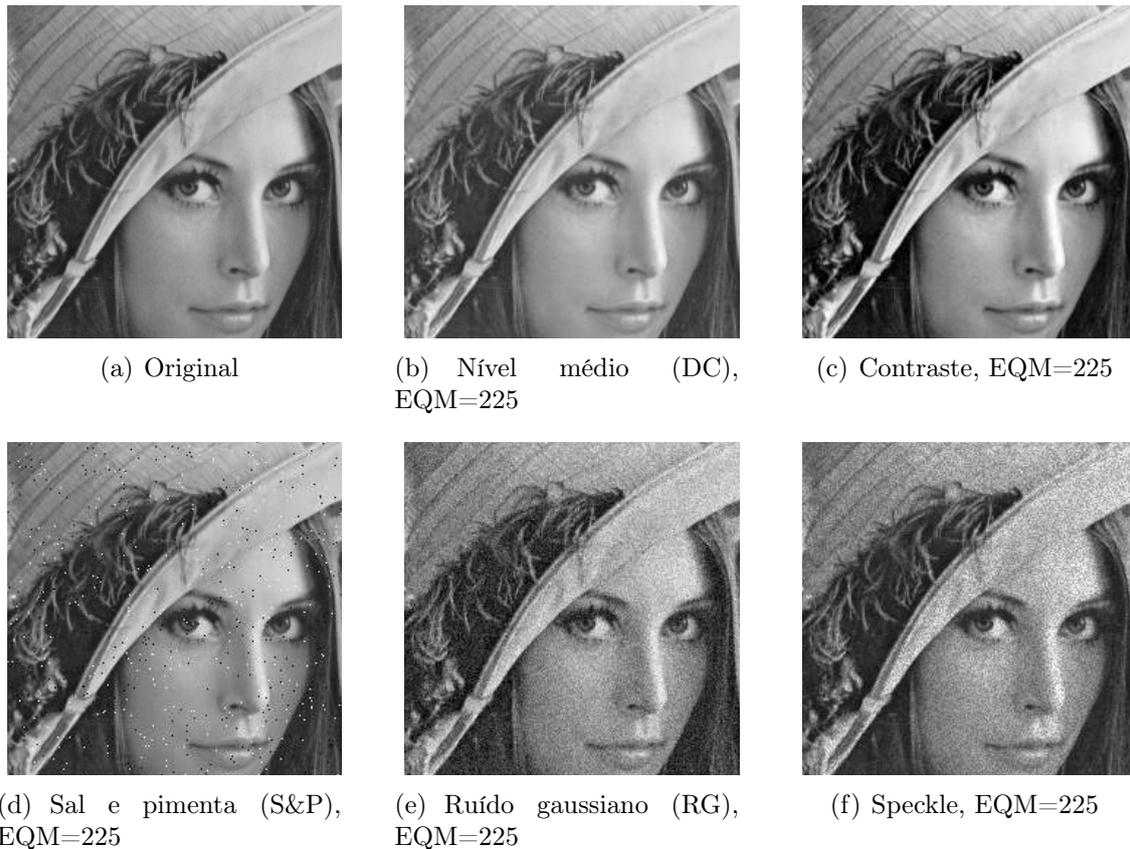


Figura 4.14: Detalhe da imagem “Lena” com diferentes degradações. Adaptado de (WANG; BOVIK, 2002a).

Speckle são superiores aos respectivos valores da sua versão contaminada com ruído gaussiano.

4.5.1 Índice de qualidade estrutural

O índice de qualidade estrutural iqe mede a diferença estrutural entre uma imagem e sua versão degradada. É um índice derivado das medidas COR e MDI, obtido por meio de uma soma ponderada entre o atributo correlação e o atributo homogeneidade da MIDE, ou seja,

$$iqe = \lambda COR + (1 - \lambda)MDI, \quad (4.8)$$

em que $\lambda \in [0,1]$ é uma ponderação da prevalência de COR sobre MDI. Em aplicações onde se deseja obter controle sobre o resultado, o valor de λ pode ser atribuído empiricamente.

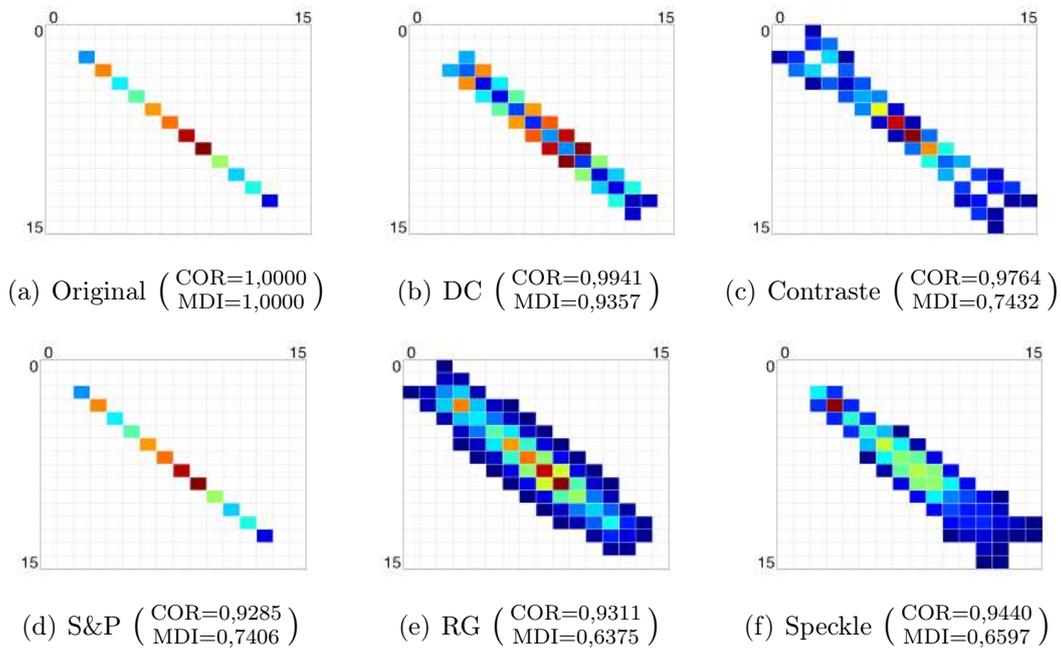


Figura 4.15: MIDE 16 níveis. Imagem “Lena” com diferentes ruídos.

4.5.2 Mapa local de qualidade

Assim como o mapa local da similaridade estrutural (SSIM, do inglês *structural similarity*), ilustrado na Figura 3.4, também é possível obter uma mapa local de qualidade utilizando o atributo CHI da MIDE. Um exemplo desse mapa local de qualidade é ilustrado na Figura 4.16. O mapa apresenta uma informação do nível de degradação em cada ponto da imagem, sendo 0 a menor degradação e 1 a maior. Contudo, o mapa local obtido com a MIDE calculada em janelas 3×3 requer um maior número de operações.

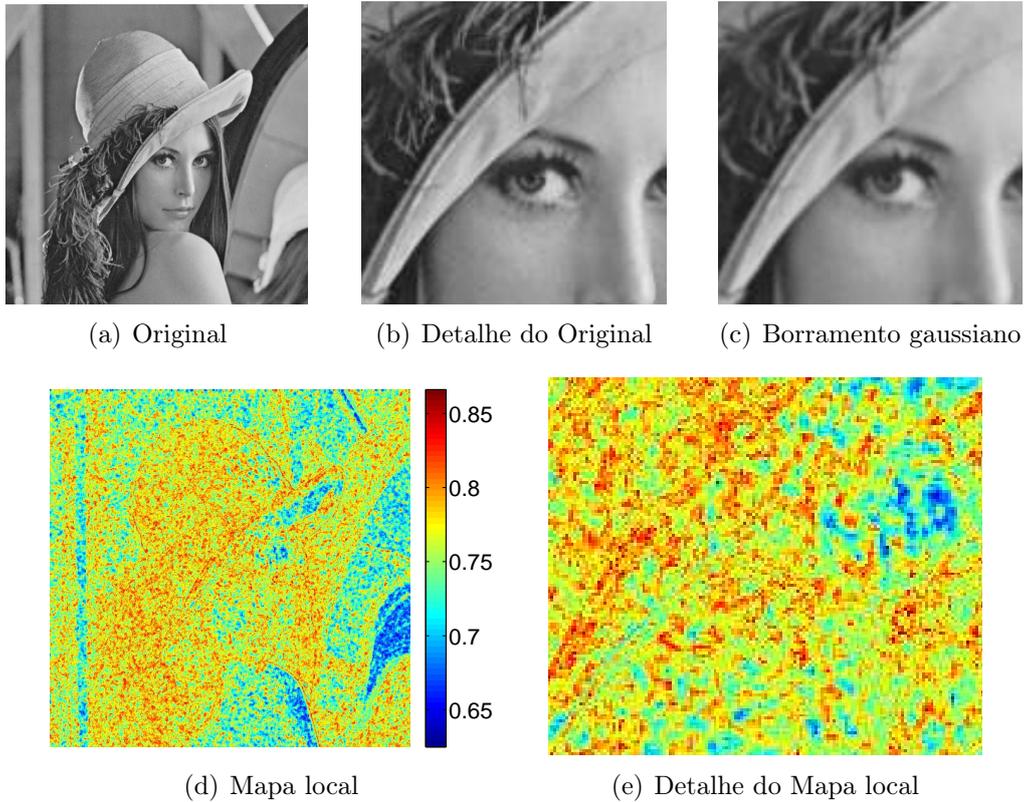


Figura 4.16: Mapa local de qualidade obtido com o atributo CHI.

4.6 Complexidade computacional da MIDE

A complexidade computacional assintótica da MIDE é a mesma do algoritmo SSIM, ou seja, $\mathcal{O}(M_1M_2)$, em que M_1 e M_2 são as dimensões da imagem de entrada. Entretanto, o algoritmo SSIM extrai informação local por meio de um janelamento, o que torna o processo mais demorado. Por outro lado, para obter um mapa local de qualidade, a complexidade assintótica é $\mathcal{O}(M_1M_2N_g^2)$, porque o número de níveis N_g da MIDE se torna relevante.

4.7 Conclusão

Neste capítulo, foi apresentado um método de avaliação referenciada de imagens que pode ser utilizado para medir o grau de similaridade estrutural entre imagens binárias ou em níveis de cinza. Os testes realizados mostram que o método pode ser empregado como medida da qualidade de imagens em níveis de cinza submetidas a algum tipo de degradação.

A Matriz de Interdependência Espacial (MIDE) apresenta uma estimativa essencialmente global do quanto uma imagem sofreu degradação estrutural. Embora

a MIDE possa ser utilizada em avaliação local, assim como a similaridade estrutural (SSIM, do inglês *structural similarity*), o seu custo computacional é mais elevado.

Baseada na estatística de coocorrência, a MIDE é calculada para um par de imagens, visando mapear as estruturas de uma imagem de referência em uma imagem de teste, ou vice-versa. O aspecto visual da MIDE permite uma interpretação direta e objetiva do nível de degradação e do grau de similaridade estrutural entre duas imagens, em um domínio independente de contexto.

Neste capítulo foram propostas medidas extraídas da MIDE e exemplos de uso em análise de imagens. As medidas COR e MDI são representações numéricas do aspecto visual da MIDE. O atributo COR sugere uma medida de similaridade estrutural, enquanto o atributo MDI sugere uma medida do grau de degradação estrutural, independente de contexto. Um terceiro atributo, a ENT, fornece uma informação adicional e descorrelacionada das demais. O atributo CHI mede o quanto a imagem analisada preserva a informação estrutural da imagem de referência.

Embora, numericamente, a MIDE possa ser empregada na análise de imagens binárias, a interpretação visual fica prejudicada, porque não é possível distinguir os padrões da matriz 2×2 formada por apenas dois níveis de cinza.

O próximo capítulo apresenta uma metodologia de transformação de imagens que viabiliza o emprego da MIDE na análise de imagens binárias.

Transformação Estrutural Multiescala

Neste capítulo é apresentado o método de Transformação Estrutural Multiescala (TEM) de imagens que se baseia em conceitos da Morfologia Matemática (MM), por meio da operação *top hat* e da superposição de efeitos introduzido no Capítulo 3.

A TEM mapeia imagens para um novo domínio em um contexto estrutural e multiescala. O objetivo principal deste mapeamento consiste em destacar saliências dos objetos presentes na imagem. A escala das estruturas salientes é definida por uma função de ponderação de escala que depende do contexto no qual a TEM é empregada. A TEM transforma imagens binárias em imagens em níveis de cinza de modo que os valores dos pixels representam as escalas das estruturas. Isto possibilita a identificação de diferenças estruturais entre imagens binárias usando métodos originalmente desenvolvidos para imagens em níveis de cinza.

Quando aplicada a imagens em níveis de cinza, a TEM atua como uma ferramenta de realce de saliências numa escala determinada pela função de ponderação. Isto facilita a detecção de pontos de interesse e auxilia em tarefas de segmentação.

5.1 Introdução

Uma forma de dar destaque às estruturas salientes dos objetos de uma imagem é decompor os mesmos em suas partes constituintes, preservando a conectividade entre elas. Utilizando os conceitos básicos da MM, é possível desenvolver técnicas de decomposição de imagens, seja por operações morfológicas ou por limiar. Em aplicações nas quais a imagem é binária, é interessante destacar uma determinada

escala de estruturas ao mesmo tempo que atenua as demais, porém preservando a conectividade original. Assim, é possível detectar estruturas em escalas escolhidas arbitrariamente, ou seja, saliências de pequena escala (cantos) ou as partes constituintes de um objeto.

O método de transformação apresentado neste capítulo foi desenvolvido para:

- i. destacar saliências em um contexto multiescala;
- ii. preservar a relação espacial e de escala das estruturas com sua vizinhança;
- iii. possibilitar o emprego em imagens binárias de ferramentas desenvolvidas para imagens em níveis de cinza.

5.2 Transformação estrutural

A transformação proposta atua sobre imagens binárias de forma inversa à decomposição por limiar de imagens em níveis de cinza. Esta transformação deve ocorrer para escalas específicas com o objetivo de destacar determinadas estruturas.

5.2.1 Transformação estrutural de imagens binárias

Tomando por base o conceito de superposição da MM, uma sequência de operações τ com elementos estruturantes isotrópicos pode ser utilizada para construir uma imagem em níveis de cinza a partir da imagem binária. Um modelo para a transformação multiescala com base em sucessivas operações τ_S , sendo S um elemento estruturante isotrópico, é apresentado na Figura 5.1. O conceito por trás deste método de transformação advém do fato de que a operação τ_S filtra apenas objetos de tamanho menor ou igual a S . O somatório das operações τ_S é incremental, ou seja, para cada tamanho de elemento estruturante S , tanto os objetos que combinam com o tamanho de S quanto os de escala menores são preservados.

Considere $\mathbf{I} : \mathcal{D}_I \rightarrow \{0,1\}$ uma imagem binária composta de objetos de diferentes escalas. Uma transformação da imagem binária \mathbf{I} para uma imagem em níveis de cinza \mathbf{R} é determinada pela operação de superposição

$$\mathbf{R} = \sum_{n=1}^N \tau_{S_n}(\mathbf{I}), \quad (5.1)$$

em que S_n é um elemento estruturante invariante à rotação e de tamanho n . O tamanho máximo de escala é determinado por N .

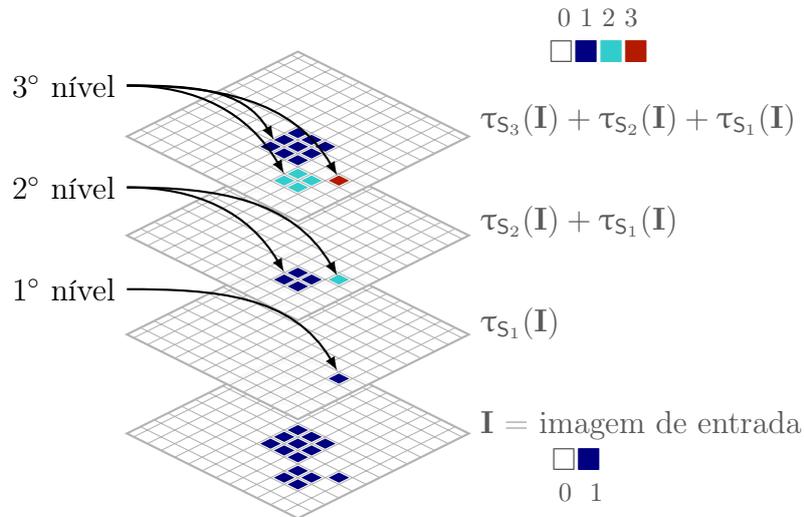


Figura 5.1: Modelo esquemático da TEM com elemento estruturante quadrado.

Analisando a Figura 5.1, observa-se que o nível de cinza dos pixels da imagem resultante da transformação é determinado pelo tamanho do elemento estruturante S . Para imagens binárias, portanto, a transformação proporciona um mapeamento de $\{0,1\}$ para níveis de cinza $\{0, 1, 2, \dots, N - 1\}$.

Na Figura 5.2 é apresentado um exemplo dessa transformação usando blocos empilhados. Os blocos brancos representam os pixels de valor 1 na imagem. O empilhamento dos blocos é proporcional ao valor do pixel, sendo que os valores mais altos são representados em tons mais escuros. Na Figura 5.2 (b), as estruturas destacadas pela operação τ_S nos primeiros níveis correspondem a saliências de menor escala da imagem original. As escalas de S definem agrupamentos de pixels em diferentes estruturas. As estruturas de escala mais alta (por exemplo, do nível 3) se conectam às estruturas de escalas menores. Este é um aspecto interessante da TEM para imagens binárias, porque nenhuma característica original é descartada após a transformação.

A TEM é, portanto, uma transformação que agrupa as estruturas da imagem em diferentes escalas, as quais são determinadas pelo tamanho e forma do elemento estruturante. De acordo com a equação 5.1, observa-se que quanto menor o tamanho do objeto destacado, maior é o valor atribuído aos pixels pertencentes a ele na escala de níveis de cinza. A imagem transformada conterá tantos níveis quantos sejam os detalhes de tamanhos diferentes recuperados pelas sucessivas operações τ_S . No exemplo da Figura 5.2, a imagem resultante contém 3 níveis de cinza, porque é composta de objetos detectados por elementos estruturantes de três tamanhos

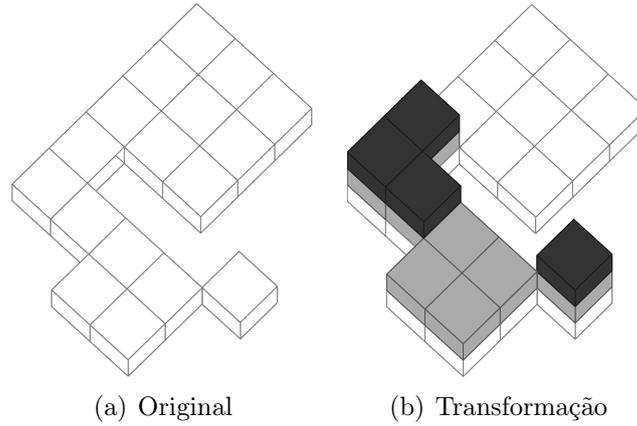


Figura 5.2: Exemplo de realce multiescala de uma imagem binária com elemento estruturante quadrado. (a) Imagem binária original formada por pixels de valor 1. (b) Imagem transformada contendo pixels de valores 1, 2 e 3, conforme a escala das estruturas.

diferentes.

Portanto, a TEM aplicada a uma imagem binária \mathbf{I} é definida como uma superposição das operações morfológicas $\tau_{\mathbf{S}}$ sobre essa imagem para diferentes tamanhos de \mathbf{S} , ou seja,

$$\varphi_N(\mathbf{I}) = \sum_{n=1}^N \tau_{\mathbf{S}_n}(\mathbf{I}), \quad (5.2)$$

em que φ_N é a transformação de superposição de N escalas.

À medida que o processo de superposição das operações $\tau_{\mathbf{S}}$ evolui, com escalas diferentes, as estruturas menores ganham destaque. As estruturas maiores que combinam com o tamanho do elemento estruturante \mathbf{S} são adicionadas nos níveis inferiores.

A intensidade máxima da imagem estrutural resultante depende do tamanho máximo do elemento estruturante N . Isto apenas altera o valor médio de intensidade da imagem obtida pela transformação sem, contudo, alterar a relação entre níveis de cinza nas diferentes escalas. Portanto, a partir de um determinado tamanho do elemento estruturante \mathbf{S} , as operações $\tau_{\mathbf{S}}$ são idempotentes e o processo pode ser interrompido.

Um critério de escolha do número de níveis N é determinado pela escala máxima dos detalhes na imagem que se deseja destacar. Quanto maior o objeto de interesse, maior deve ser o tamanho de \mathbf{S} e, conseqüentemente, maior o número de níveis

N . Essa é uma informação que, em geral, exige um conhecimento prévio das características da imagem que será analisada.

Na sua forma mais geral, a transformação estrutural é realizada por meio de uma ponderação, ou seja

$$\varphi_N(\mathbf{I}) = \sum_{n=1}^N f_p(n) \cdot \tau_{S_n}(\mathbf{I}), \quad (5.3)$$

em que $f_p(n)$ é uma função de ponderação de escala definida para cada valor de n . A função f_p modifica o contraste entre as escalas das estruturas e pode ser escolhida de acordo com a aplicação desejada.

No contexto da TEM, o contraste se refere à transformação controlada for f_p , capaz de diferenciar estruturas de escalas adjacentes. Um exemplo de contraste de estruturas corresponde a dar importância maior para uma escala n , por meio da atribuição de um nível de cinza mais elevado, enquanto as escalas adjacentes menos importantes, $n - 1$ e $n + 1$ devem receber um nível de cinza mais baixo.

A transformação $\mathbf{R} = \varphi(\mathbf{I})$ por meio da equação 5.3 resulta em uma imagem \mathbf{R} cujas valores : $\mathcal{D}_I \in \mathbb{R}^2$, no caso em que $f_p \in \mathbb{R}$. Nos casos em que é conveniente que a imagem estrutural \mathbf{R} esteja na mesma faixa de níveis de cinza da imagem original \mathbf{I} , pode-se empregar um procedimento de normalização

$$\mathbf{R}_N = \frac{\mathbf{R} - \min(\mathbf{R})}{\max(\mathbf{R}) - \min(\mathbf{R})} \in [0,1], \quad (5.4)$$

e o posterior mapeamento de \mathbf{R}_N para a faixa de níveis de cinza da imagem de entrada \mathbf{I} , por meio da expressão

$$\mathbf{R}_I = \lfloor \mathbf{R}_N(\max(\mathbf{I}) - \min(\mathbf{I})) + \min(\mathbf{I}) \rfloor, \quad (5.5)$$

em que os operadores \max e \min retornam os valores máximo e mínimo da imagem \mathbf{I} , respectivamente. O operador $\lfloor \cdot \rfloor$ retorna o menor valor inteiro mais próximo. \mathbf{R}_I representa a imagem reconstruída mapeada no domínio da imagem \mathbf{I} . Vale notar que este procedimento resulta em uma quantização dos valores de \mathbf{R} , o que pode alterar o contraste entre escalas. Em outras palavras, escalas de \mathbf{R} com valores próximos podem ser agrupadas em um mesmo nível em \mathbf{R}_I , embora as estruturas pertencessem originalmente a escalas diferentes.

Uma vantagem do uso da transformação generalizada da equação 5.3 sobre a transformação da equação 5.2 é a possibilidade de modificar de forma controlada as

escalas durante a operação de transformação. Em determinadas aplicações, algumas escalas não são desejadas como, por exemplo, as escalas que contêm ruído. A função de ponderação de escala pode, então, ser ajustada para atenuar ou remover os objetos de determinadas escalas. Essa estratégia pode ser aplicada na filtragem de ruído ou granulometria. Por exemplo, para atenuar objetos ou partes deles em escalas de tamanho menores do que 3 e realçar os demais, adota-se $f_p = \{0, 0, 1, 1, \dots\}$. Neste caso, as estruturas das escalas $n \leq 2$ possuem peso 0 e serão eliminadas, enquanto as demais escalas serão preservadas e realçadas. Com isso, a função de transformação f_p da TEM generalizada da equação 5.3 atribui valores diferentes para determinadas escalas de detalhes, resultando em uma imagem com maior ou menor contraste entre os níveis de escala.

Um exemplo do uso da função de ponderação pode ser observado na Figura 5.3. Sabe-se que as estruturas menores também são capturadas pela operação *top hat* nas escalas maiores. Por isso uma função de ponderação uniforme resulta em uma imagem na qual os valores dos pixels são mais altos para as escalas menores, como ilustrado na Figura 5.3 (a). Na Figura 5.3 (b) foi utilizada uma função de ponderação em que o peso é zero apenas na escala 2. O resultado é uma imagem transformada que destaca apenas as estruturas de escala 1. Um resultado semelhante para a escala 2 é obtido com a função de ponderação da Figura 5.3 (b). Se o objetivo é eliminar a escala maior, então o peso respectivo dessa escala deve ser zerado, como ilustrado na Figura 5.3 (c).

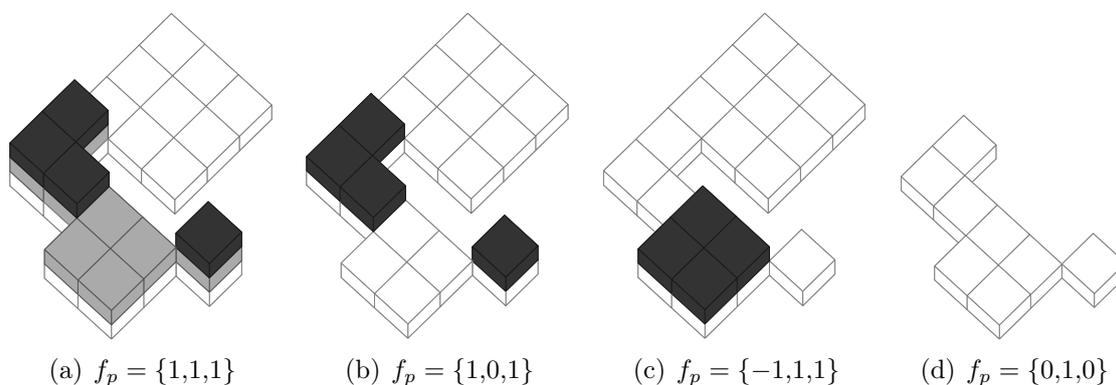


Figura 5.3: Resultado da transformação para diferentes funções de ponderação.

A observação das Figuras 5.3 e 5.2 permite concluir que a TEM para imagens

binárias tem duas aplicações diretas. A primeira é decompor a imagem em estruturas de diferentes escalas espacialmente vizinhas de escalas adjacentes. A segunda é produzir uma imagem estrutural, na qual as estruturas são representadas por intensidades diferentes. Para esta imagem estrutural podem ser empregadas ferramentas desenvolvidas para imagens em níveis de cinza, dentre elas a Matriz de Interdependência Espacial (MIDE) e as operações de corte por limiar, por exemplo.

5.2.2 Transformação estrutural de imagens em níveis de cinza

Assim como as operações morfológicas, a TEM pode ser estendida para imagens em níveis de cinza. Numa imagem em níveis de cinza, a TEM atua destacando estruturas em escalas correspondentes a S . As imagens da Figura 5.4 ilustram o resultado da transformação de uma imagem de $N_g = 4$ níveis de cinza. Para simplificar a exibição das imagens, os pixels correspondentes ao nível 0 não são exibidos. A imagem estrutural resultante da TEM, ilustrada na Figura 5.4 (b), apresenta pixels com valores que correspondem ao nível de escala das saliências destacadas. Pode-se observar, também, que o contraste entre os níveis de cinza foi modificado, aumentando o destaque entre as estruturas das escalas menores e maiores.

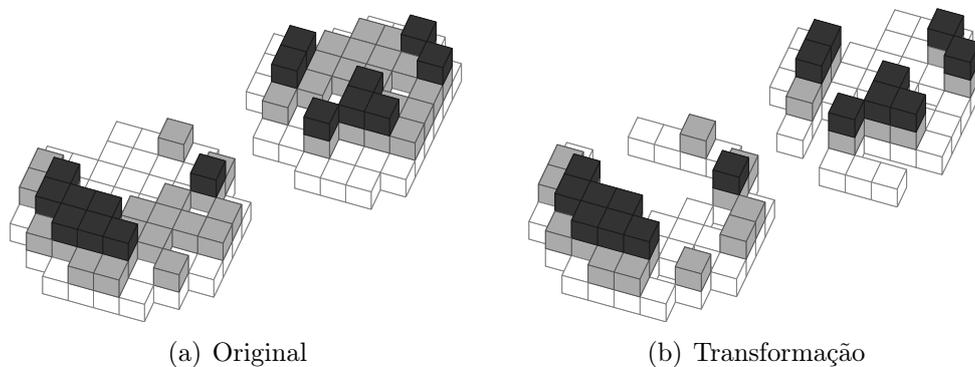


Figura 5.4: Exemplos de transformação estrutural multiescala de uma imagem em níveis de cinza. (a) Imagem original com 4 níveis de cinza. (b) A imagem estrutural resultante da TEM apresenta pixels com valores diferentes, os quais correspondem ao nível de escala das saliências destacadas pela TEM.

No exemplo da Figura 5.5, é possível notar que a TEM aumenta o contraste entre os picos e vales. Este efeito é comprovado quando se compara os histogramas da imagem original e da imagem transformada, como ilustram os gráficos da Figura 5.5 (c) e (d), respectivamente. Observa-se que duas modas são perceptíveis no histograma da imagem transformada. Este resultado é importante para auxiliar

na tarefa de segmentação automática, porque facilita a escolha de um limiar. O aumento do contraste entre picos e vales é uma etapa intermediária importante para a detecção de saliências.

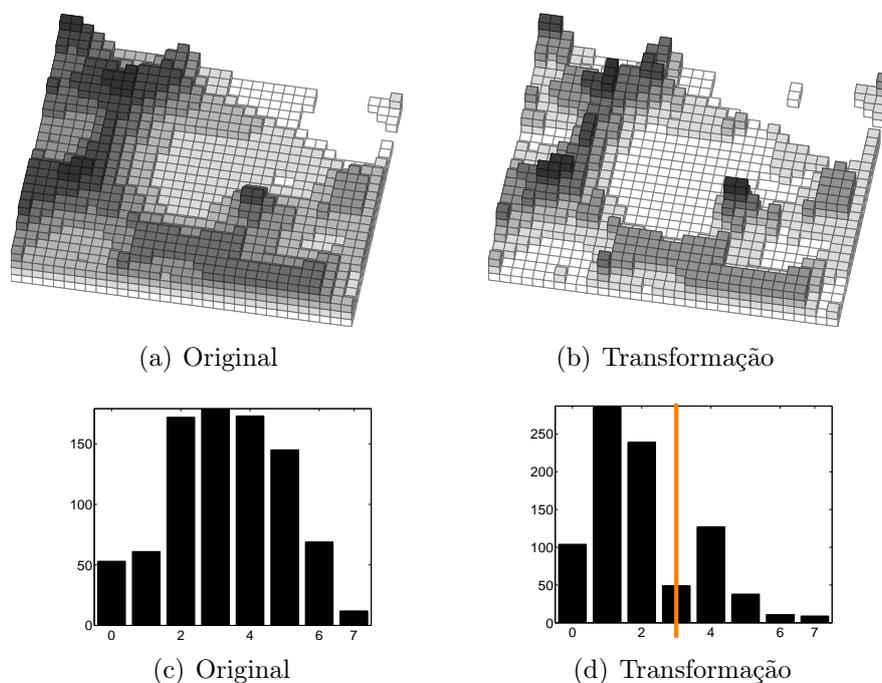


Figura 5.5: Comparação entre imagem original e resultado da TEM. (a) Imagem original em 8 níveis de cinza e (b) imagem realçada. (c) Histograma dos níveis de cinza da imagem original. (d) Histograma dos níveis de cinza da imagem transformada, com destaque para a linha que separa o histograma em duas classes de pixels.

5.2.3 Parâmetros da transformação

A escolha dos parâmetros da transformação pode ser realizada de acordo com o problema abordado. Os parâmetros dizem respeito à forma e tamanho do elemento estruturante, um critério para a interrupção da iteração da transformação e a definição de uma função de ponderação de escala.

Forma e tamanho do elemento estruturante S

A adoção de um elemento estruturante isotrópico, como os ilustrados na Figura 2.4, permite analisar estruturas independente da sua orientação na imagem. De modo geral, quando se deseja utilizar a TEM para promover uma separação de escalas de objetos retangulares em níveis de intensidade, um elemento estruturante de forma

retangular é a escolha apropriada. Por outro lado, um elemento estruturante circular combina melhor com uma variedade de estruturas componentes dos objetos.

A escolha da forma do elemento estruturante deve levar em consideração um critério de custo computacional envolvido no processamento. Quanto maior o elemento estruturante, maior é o custo computacional. A complexidade do elemento estruturante também é importante, porque sua forma incide sobre a sua decomposição em componentes mais simples. Um quadrado 4×4 , como o da Figura 2.4 (i), pode ser decomposto em quatro quadrados menores 2×2 (VINCENT, 1991; MARAGOS, 1989). Por outro lado, o disco da Figura 2.4 (d) é decomposto em um número maior de formas, o que aumenta a complexidade das operações.

A forma e tamanho da estrutura analisada na imagem dependem da relação de intensidade dos pixels com a sua vizinhança e, portanto, o elemento estruturante deve combinar com as estruturas que se deseja salientar. O resultado desejado depende do formato do elemento estruturante e, portanto, a imagem estrutural \mathbf{R} será diferente caso a forma adotada seja um disco ou um quadrado. Um exemplo disto pode ser observado na imagem da Figura 5.6. A escolha do elemento estruturante modifica o resultado da transformação e possibilita uma interpretação diferente. Na Figura 5.6 (b), o uso de elemento estruturante circular permite que pixels de estruturas salientes pequenas, tenham seu valor elevado, destacando cantos da forma. Quando se utiliza um elemento estruturante retangular, por outro lado, as estruturas salientadas são componentes constituintes da forma original, como pode ser observado na Figura 5.6 (c). Em ambos os casos, a escala dos elementos estruturantes define o nível de cinza das estruturas após a transformação.

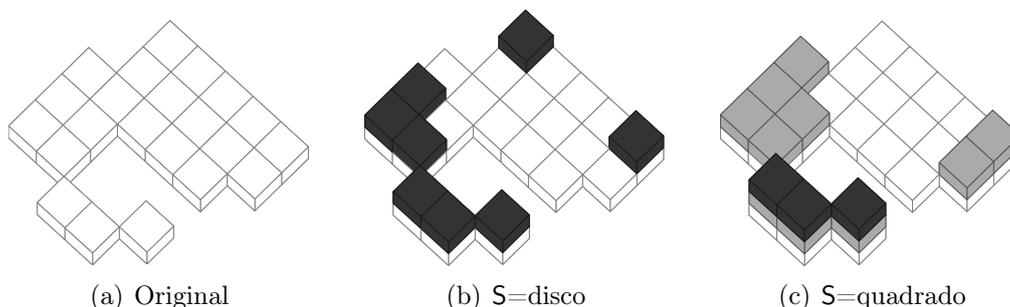


Figura 5.6: Exemplos de transformação estrutural multiescala de uma imagem binária representando diferentes interpretações. (a) Transformação em 2 níveis utilizando um elemento estruturante circular. (b) Transformação em 3 níveis utilizando um elemento estruturante retangular.

Critério de convergência

A convergência da transformação é alcançada no nível máximo N quando não existe mudança significativa entre o corte em um nível $n = N$ e o corte em um nível $n - 1$. Isso acontece porque a operação *top hat* se torna idempotente a partir de um tamanho $n \geq N$ do elemento estruturante. Em outras palavras, quando o elemento estruturante é suficientemente grande para conter todas as escalas de estruturas da imagem, o resultado da operação *top hat* não apresenta mudança significativa do nível anterior.

Portanto, um critério suficiente de convergência consiste em monitorar o número de pixels que mudam de nível entre dois cortes adjacentes. Isto significa encontrar o argumento que minimiza a diferença $\varphi_n - \varphi_{n-1}$, ou seja

$$N = \arg \min_n (\varphi_n - \varphi_{n-1}), \quad (5.6)$$

sabendo que N_{max} pode ser estimado como sendo a menor dimensão da imagem.

Outra possibilidade consiste em utilizar um dos atributos da MIDE calculado para níveis adjacentes da transformação. Neste caso, o tamanho máximo do elemento estruturante é determinado pelo argumento que maximiza o valor do atributo A extraído de $\mathfrak{M}(\varphi_n, \varphi_{n-1})$, ou seja,

$$N = \arg \max_n (A_{\mathfrak{M}(\varphi_n, \varphi_{n-1})}). \quad (5.7)$$

Como o atributo $CHI = 0$ para imagens idênticas, é conveniente utilizá-lo na equação 5.7, contudo substituindo o operador \max por \min . Dessa forma, o critério baseado na MIDE se assemelha à equação 5.6, já que a CHI é uma medida de desvio.

Escolha da função de ponderação

A função f_p é escolhida de acordo com a necessidade de atribuir maior ou menor grau de importância a uma determinada escala de objeto. Em uma aplicação na qual as estruturas pequenas e de baixo contraste em níveis de cinza merecem um destaque maior do que o restante da imagem, pode-se utilizar uma função rampa ou exponencial. O número de níveis N da imagem resultante é escolhido conforme a necessidade da aplicação a que se destina a imagem realçada.

Para várias aplicações, a escolha da função de ponderação estará atrelada ao conhecimento prévio do problema abordado (Apêndice B). Possíveis funções de ponderação de escala f_p e suas aplicações são listadas na Tabela 5.1. A f_p pode

ser qualquer tipo de mapeamento linear ou não-linear, isto é, pode ser obtida tanto a partir uma regressão linear dos valores dos pixels da imagem de entrada, quanto a partir da saída de uma Rede Neural Artificial (RNA) (HAYKIN, 2001) ou lógica nebulosa (*Fuzzy*) (ZIMERMANN, 1992).

Tabela 5.1: Classificação das funções de ponderação de escala

Classe	Função	Aplicação
I	$f_p(n) = k, \forall n \geq 1$	ponderação por uma constante k : realça estruturas pequenas (modo padrão)
II	$f_p(n) = L(n)$	modelo linear diretamente ou inversamente proporcional à escala de objetos
III	$f_p(n) = K(n)$	modelo não linear da escala de objetos: conjunto de pesos, rede neural, conjunto de regras
IV	$f_p(n) = P(n)$	modelo probabilístico ou paramétrico da escala de objetos

A Figura 5.7 ilustra alguns exemplos de função de ponderação e o resultado obtido tem a Figura 5.2 (a) como imagem de entrada, a qual é formada por três escalas de estruturas, a saber, quadrados de 1, 2 e 3 pixels de lado. Neste exemplo, a TEM utiliza elementos estruturantes quadrados com tamanhos de 1 a 4. Por isso, todas as imagens resultantes possuem nível de cinza máximo $N = 4$, como é possível observar pela altura da pilha de blocos. Para a função de modo padrão, $f_p(n) = 1$, apenas as escalas menores que 3 são destacadas. As escalas 1 e 2 receberam mesmo valor por causa do arredondamento numérico na quantização em 4 níveis. A função $f_p(n) = N - n$ produz um maior contraste entre as escalas, destacando as de menor tamanho ao atribuir pesos maiores para escalas menores. A função $f_p = N \exp(-(n - 2)^2)$ produz um resultado em que apenas a escala $n = 2$ é destacada. Também neste caso, o arredondamento fez com que as estruturas de escalas 1 e 2 recebessem o mesmo valor. Importante notar que a Figura 5.2 (c) difere da Figura 5.2 (a) no valor da estrutura de escala 3.

Quando existe algum conhecimento *a priori* sobre a relação espacial entre os pixels da imagem, é possível usá-lo para modelar a função de ponderação de escala. Suponha que a transformação de escala de objetos desejada segue um modelo normal de distribuição de tamanho dos mesmos na imagem. Uma escolha adequada para a função de ponderação pode ser a função de densidade de probabilidade

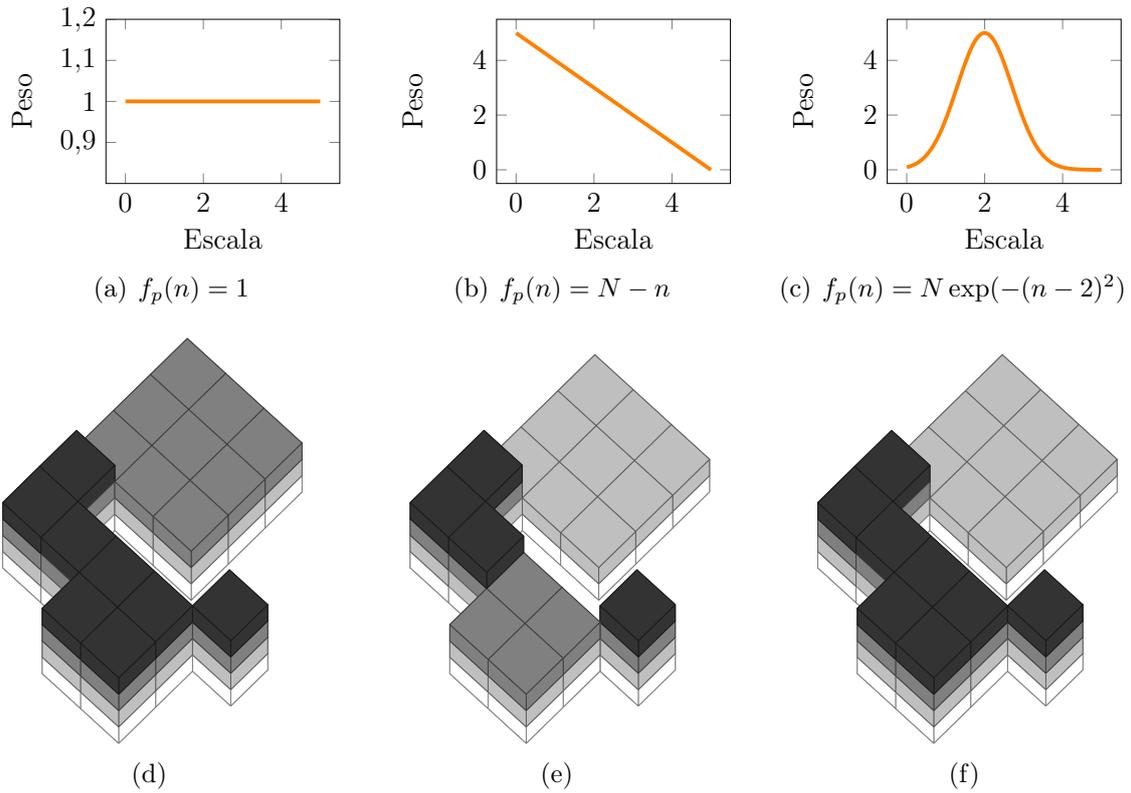


Figura 5.7: (a-c) Funções de ponderação de escala e (d-f) os resultados da TEM para a imagem da Figura 5.2 (a).

$f_p(n) = P(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{1}{2} \frac{(n-\mu_n)^2}{\sigma_n^2}}$, dados uma média μ_n e uma variância σ_n^2 das escalas das estruturas. Os parâmetros μ_n e σ_n^2 permitem definir qual a amplitude e faixa de escalas n é mais importante na ponderação atuando, assim, como um filtro passa-faixa.

As imagens da Figura 5.8 e Figura 5.9 ilustram casos particulares da transformação de uma imagem binária para uma função de ponderação que prioriza escalas pequenas $f_p(n) = 40 - 2n$ e outra que prioriza as escalas maiores $f_p(n) = \frac{5}{4}n - 50$, respectivamente. Os tons de vermelho representam as escalas com maior ponderação.

A finalidade da função de ponderação é modificar o contraste entre as escalas da imagem estrutural ou da imagem realçada. Neste sentido, a função não está atrelada a um conhecimento prévio sobre a formação estrutural dos objetos da imagem de origem. Por isso, no modo padrão $f_p(n) = k, \forall n \geq 1$ para $k = 1$, a TEM resulta em uma imagem estrutural cujos pixels são mapeados para níveis de cinza correspondentes às escalas das estruturas presentes na imagem. Entretanto, quando

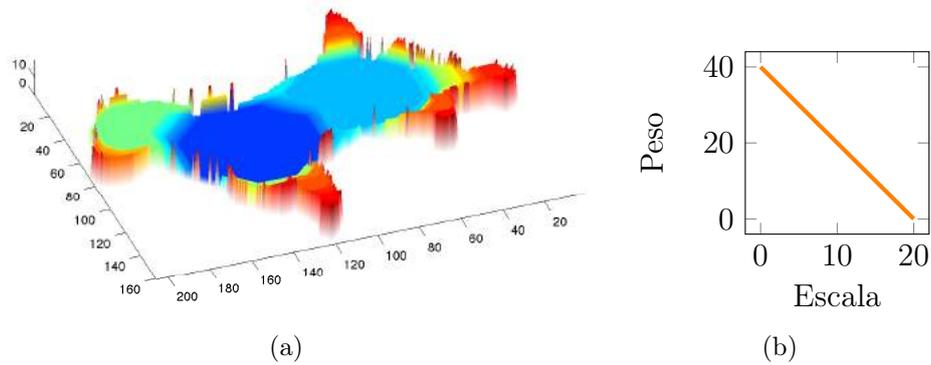


Figura 5.8: (a) Transformação de estruturas com peso inversamente proporcional às escalas. (b) Evolução do tamanho do elemento estruturante e seus respectivos pesos.

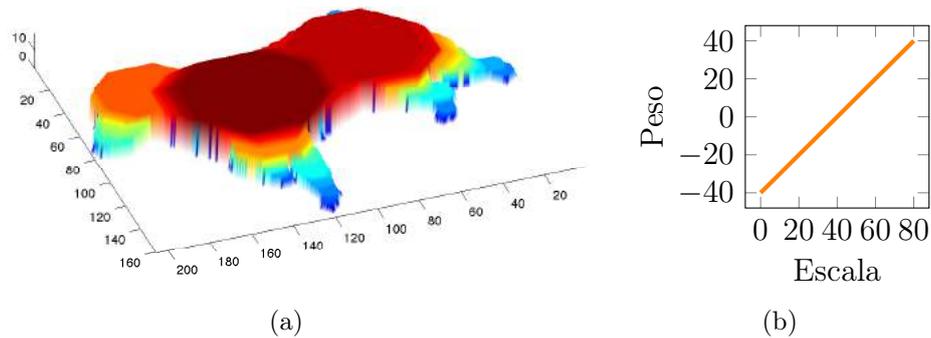


Figura 5.9: (a) Transformação de estruturas com peso diretamente proporcional às escalas. (b) Evolução do tamanho do elemento estruturante e seus respectivos pesos.

o conhecimento *a priori* está disponível ou pode ser inferido a partir da imagem, a função de ponderação torna-se, convenientemente, um meio para aceleração do processamento porque permite reduzir drasticamente a quantidade de operações necessárias.

5.2.4 Decomposição de imagens estruturais por limiar

Um conceito comumente empregado na MM, seja na elaboração de teorias ou nas aplicações práticas é a decomposição por limiar. Este conceito, obviamente, só pode ser aplicado a imagens em níveis de cinza, o que abrange as imagens resultantes da transformação TEM.

Seja $\mathbf{R} = \varphi(\mathbf{I})$, uma imagem estrutural, o procedimento consiste em decompor

\mathbf{R} por meio de operações de corte Φ , tal que

$$\Phi_n(\mathbf{R}) = \{p \in \mathbf{R} : \mathbf{R}(p) \geq n\}, \quad (5.8)$$

em que n é o nível de corte.

O resultado da decomposição por cortes da imagem estrutural \mathbf{R} da Figura 5.8, ilustrado na Figura 5.10. Os cortes foram obtidos a partir da equação 5.8.

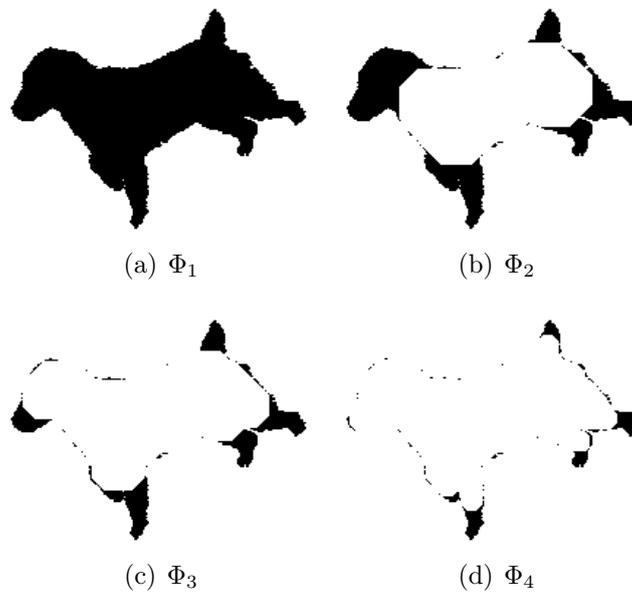
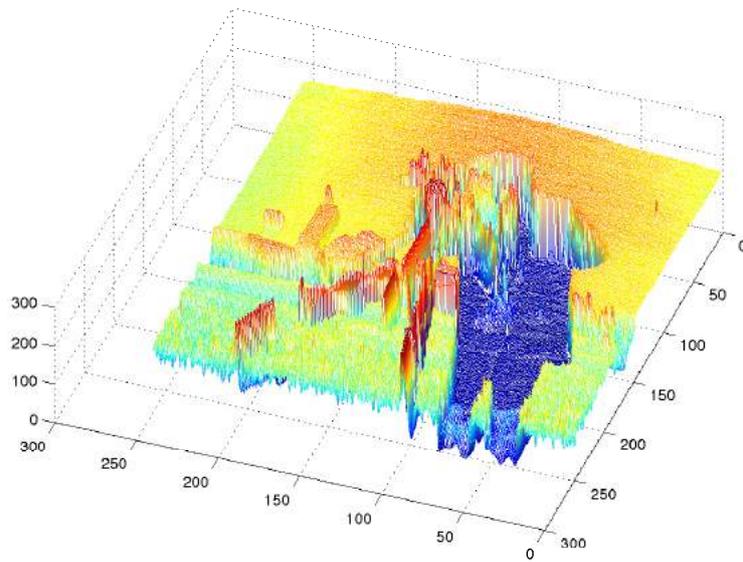


Figura 5.10: Cortes da transformação de “dog-1” da base MPEG-7, usando a TEM.

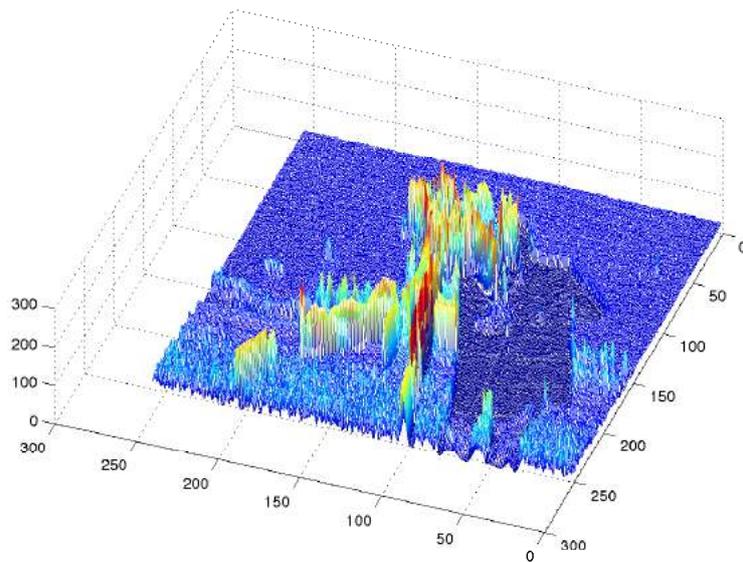
5.3 Realce estrutural de imagens em níveis de cinza

Quando a TEM é aplicada em imagens em níveis de cinza, o efeito obtido é o realce de estruturas em escalas pré-determinadas. O realce obtido pela TEM, ilustrado na Figura 5.11, destaca saliências presentes nas imagens em níveis de cinza. O exemplo da figura é resultado do realce das estruturas de escalas menores. Neste exemplo, é possível perceber que a imagem resultante da transformação preserva a topologia, já que a conectividade entre as estruturas não é afetada. Entretanto, dependendo da função de ponderação de escala escolhida, é possível que algumas estruturas desapareçam da imagem, como no caso da filtragem exemplificada na Figura 5.3.

Características da imagem original são preservadas pela transformação como, por exemplo, a borda dos objetos. A preservação de bordas da transformação com $f_p(n) = 1$ pode ser observada no exemplo da Figura 5.12. A imagem “cameraman” foi



(a)



(b)

Figura 5.11: Visualização tridimensional da imagem “cameraman”. (a) Original e (b) realçada. As áreas mais claras (em tons de amarelo e vermelho) representam intensidades maiores.

escolhida por conter estruturas de diversos formatos, escalas e variação de contraste. Neste exemplo foi realizada a detecção de bordas pelo método de Canny (1986) nas imagens realçadas pela TEM para diferentes níveis. O resultado mostra que a transformação aplicada destaca estruturas de menor escala sem afetar as bordas dos objetos de maior escala. Isto é importante em aplicações que se baseiam na detecção

do contorno dos objetos (GIBSON *et al.*, 2013).

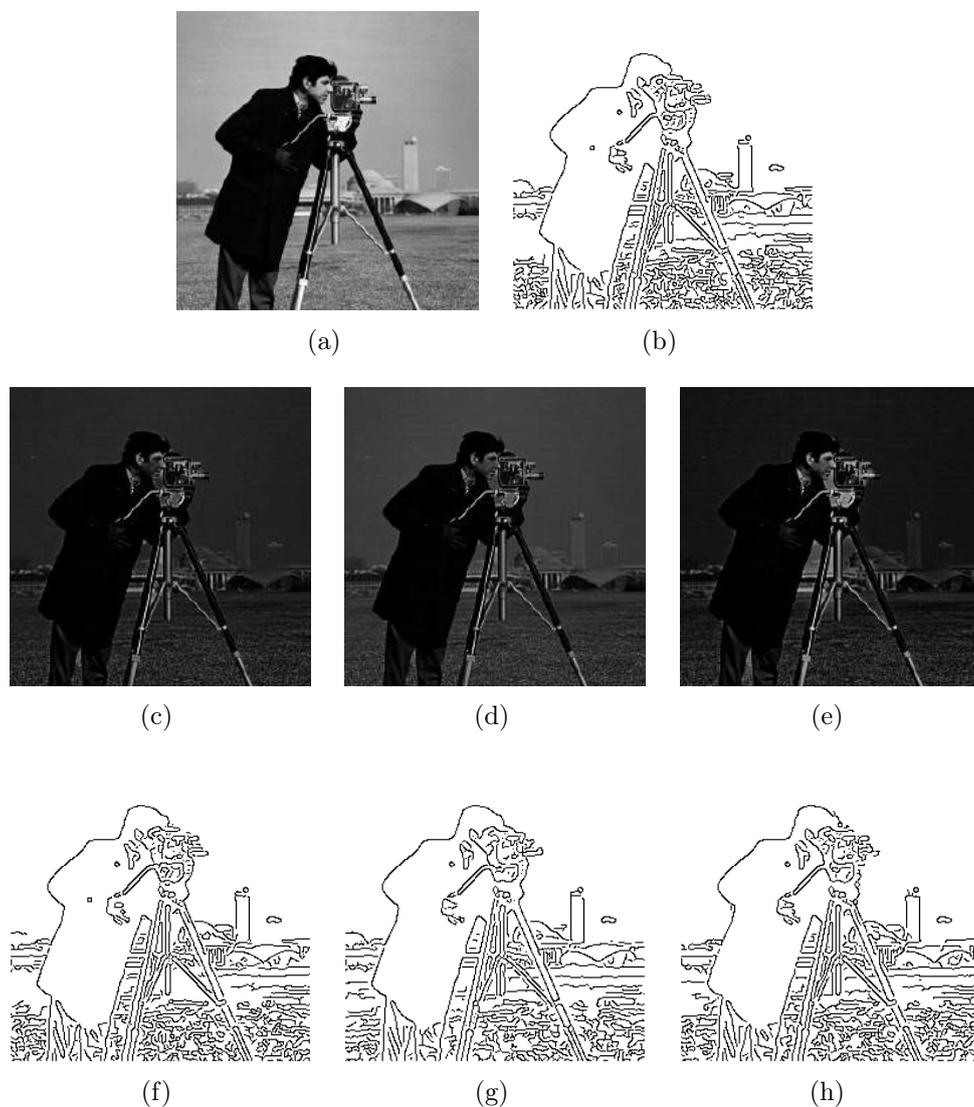


Figura 5.12: Preservação de bordas na transformação TEM. (a) Imagem original “cameraman”. (c,d,e) Imagens realçadas para $N = 2$, $N = 4$ e $N = 8$. (b,f,g,h) Detecção de bordas pelo método de Canny (1986).

5.4 Mapeamento de imagens binárias para níveis de cinza

Em geral, as transformações morfológicas em imagens binárias estão voltadas para a decomposição multiescala, ou granulometria. Quando se deseja preservar a topologia original da imagem binária, as operações morfológicas de preservação topológica são aplicadas. Usando a TEM, obtém-se uma imagem em níveis de cinza \mathbf{R} a partir da imagem binária \mathbf{I} na qual estão preservadas as propriedades topológicas da imagem original.

A preservação da topologia em imagens binárias transformadas pode ser observada no exemplo da Figura 5.13. A imagem “shark” da base MPEG-7 Core Experiment Shape-1¹ foi escolhida por conter estruturas de diversos formatos e escalas. Observa-se que a forma original é preservada, embora as estruturas componentes dessa forma possuam valores de nível de cinza diferentes da forma original. Neste exemplo, foi dado destaque para as estruturas maiores, que constituem a parte central da forma. Os tons mais claros nas imagens das Figuras 5.13 (b) e (c) representam as intensidades menores.

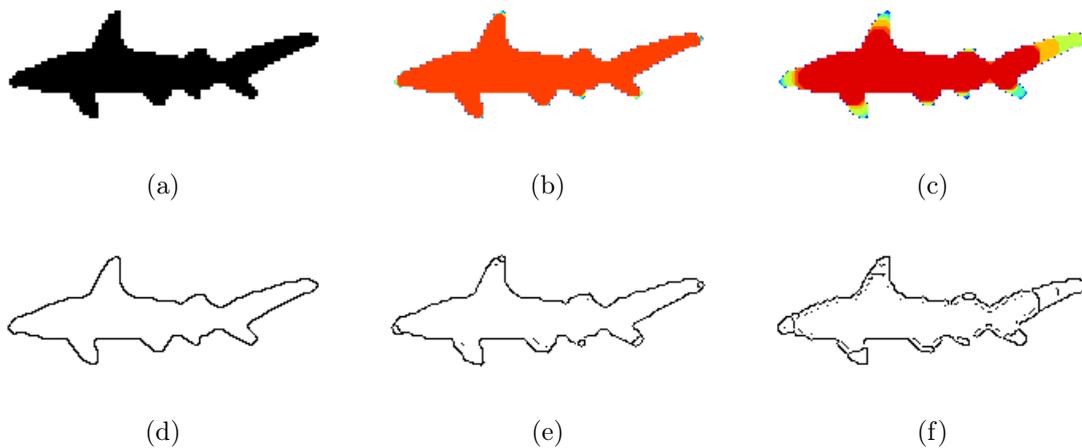


Figura 5.13: Preservação da topologia da TEM. (a) Imagem original “shark” MPEG-7. (b-c) Imagem realçada com 4 e 8 níveis. (d-e) detecção de bordas pelo método de Canny (1986).

Na Figura 5.14, é possível notar que a modificação causada pela TEM na imagem binária permite realçar partes salientes da imagem, que neste exemplo são as estruturas pequenas do contorno.

¹<http://www.cis.temple.edu/~latecki/TestData/mpeg7shapeB.tar.gz>

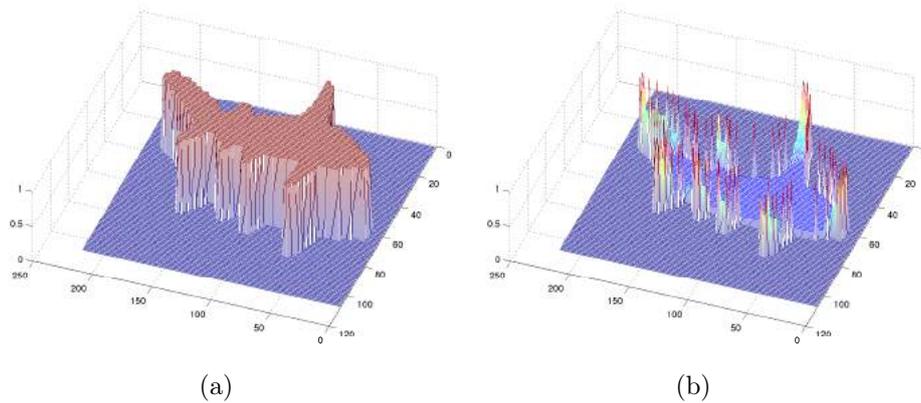


Figura 5.14: Visualização tridimensional da imagem MPEG-7 “shark”. (a) original e (b) realçada, na qual somente as saliências possuem valores maiores do que zero.

5.5 Complexidade computacional da TEM

A TEM possui complexidade assintótica $\mathcal{O}(M_1 M_2 N^2)$, sendo $N = \min(M_1, M_2)$ o tamanho máximo do elemento estruturante para uma imagem de dimensões $M_1 \times M_2$. Essa complexidade corresponde à operação de abertura definida na equação 2.15, que é utilizada na operação *top hat*, conforme equação 2.17. Adotando-se a otimização proposta por Boomgaard e Balen (1992), a complexidade da TEM pode ser reduzida para $\mathcal{O}(M_1 M_2 N \log(N))$. Além disso, quando é possível definir uma função de ponderação de escala ou um critério de convergência, a realização computacional envolve um número de iterações menor do que o valor assintótico complexidade. Por isso, embora a TEM seja um processo que envolve um elevado custo computacional quando $N \rightarrow \min(M_1, M_2)$, é possível que o resultado desejado seja alcançado com um pequeno número de iterações, como no exemplo ilustrado na Figura 5.13.

5.6 Conclusão

Neste capítulo foi apresentada uma proposta de transformação morfológica multiescala. Essa transformação produz como resultado uma imagem em tons de cinza mapeada para um domínio em que os valores dos pixels correspondem à escala das estruturas. O método pode ser aplicado para imagens binárias ou em níveis de cinza, produzindo diferentes interpretações.

Quando aplicada a imagens em níveis de cinza, a Transformação Estrutural Multiescala (TEM) promove um realce das estruturas, o que facilita a segmentação. Um caso particular da transformação proposta possibilita obter cortes de uma imagem nas escalas desejadas. A presença de objetos em uma determinada escala é

controlada pelo tamanho do elemento estruturante que, por sua vez, é uma função do número de escalas desejado. Este processo está relacionado a tarefas de filtragem para remoção de estruturas.

A TEM realiza o mapeamento de imagens binárias em imagens em tons de cinza com destaque para detalhes ou saliências, o que pode ser obtido utilizando-se como parâmetros N e $f_p(n) = 1$, sendo N a quantidade máxima de níveis de cinza.

No próximo capítulo são apresentadas propostas de uso da avaliação estrutural por meio da Matriz de Interdependência Espacial (MIDE) e da TEM.

Capítulo 6

Análise Estrutural Multiescala por meio da MIDE e TEM

Neste capítulo são apresentados exemplos de utilização da Transformação Estrutural Multiescala (TEM) e Matriz de Interdependência Espacial (MIDE), que demonstram sua aplicabilidade em extração de características, avaliação da qualidade de imagens, segmentação e detecção de saliências. Adicionalmente, este capítulo apresenta brevemente as bases de imagens utilizadas nos experimentos.

Os algoritmos para análise de imagens por meio dos atributos extraídos da MIDE envolvem aplicações de extração de características estruturais multiescala e medidas de avaliação referenciada da qualidade de imagens. Nos casos em que a imagem de entrada é uma imagem binária, é conveniente aplicar uma transformação para torná-la uma imagem em níveis de cinza, possibilitando o uso de ferramentas específicas para esse tipo de imagem. Neste contexto, a TEM é empregada para obter uma representação estrutural multiescala da imagem binária, o que permite extrair informações por meio da MIDE.

6.1 Introdução

O emprego da MIDE e da TEM na análise de imagens pode ocorrer separadamente ou combinadas em uma sequência de operações. Quando a imagem a ser analisada é uma imagem em níveis de cinza, a MIDE é capaz de extrair a informação estrutural contida nas imagens. Isto é possível porque imagens em níveis de cinza carregam informação estrutural nos valores dos pixels, através da relação de intensidade desses pixels com a sua vizinhança. No caso de imagens binárias, entretanto, emprega-se a TEM para produzir uma imagem estrutural que possa ser

analisada por meio da MIDE. Ao comparar diferentes segmentações binárias de um mesmo objeto, muitas vezes é importante destacar as diferenças estruturais, ao invés de simplesmente comparar o número de acertos e erros pixel a pixel.

Um modelo geral do uso da TEM e da MIDE é apresentado na Figura 6.1. Várias alternativas de análise são possíveis, dependendo do tipo de imagem de entrada e da informação que se deseja obter como saída. As etapas de transformação somente são empregadas quando o destaque de informações estruturais é necessário, como quando ocorre ao realçar saliências ou segmentar determinadas estruturas em escalas pré-estabelecidas. Para a aplicação de avaliação referenciada da qualidade de imagens, a transformação não é requerida, pois a informação estrutural é detetada pela MIDE na análise comparativa entre a imagem de referência e a sua versão degradada. As linhas descontínuas indicam as diversas possibilidades de obtenção da saída (índice de avaliação) a partir de uma ou mais imagens de entrada. Nos casos em que não há uma imagem de referência, entretanto, utiliza-se um algoritmo de degradação para alterar artificialmente as estruturas da imagem a ser avaliada. A etapa de degradação artificial indicada pelo retângulo pontilhado da Figura 6.1 não deve ser confundida com a degradação referente ao processo de aquisição das imagens. A cópia degradada é comparada com a imagem de entrada, sendo o tipo de degradação escolhido conforme a informação que se deseja extrair da imagem. Por exemplo, quando se deseja avaliar a intensidade de ruído em uma imagem, uma cópia suavizada da imagem pode ser utilizada como referência. Quando a origem do ruído é conhecida, a MIDE pode ser empregada na avaliação da qualidade de filtros, utilizando como entradas a imagem ruidosa e a imagem filtrada.

6.1.1 Base de imagens

Uma das dificuldades inerentes ao desenvolvimento de algoritmos de análise de imagens é a forma de avaliar o seu desempenho. O uso de imagens amplamente utilizadas em diversos trabalhos consiste na forma usual de avaliar algoritmos de Processamento Digital de Imagens (PDI) ou de Sistema de Visão Computacional (SVC). As bases utilizadas nos experimentos desta tese são descritas como segue.

LIVE

O banco de imagens LIVE (*Laboratory for Image and Video Engineering*)(SHEIKH *et al.*, 2005) é formado por imagens de cenas do cotidiano contendo cópias degradadas por compressão (JPEG2k e JPEG), ruído branco (RB), borramento gaussiano (BG) e distorção de fase (FF, do inglês *fast fading*). Estas

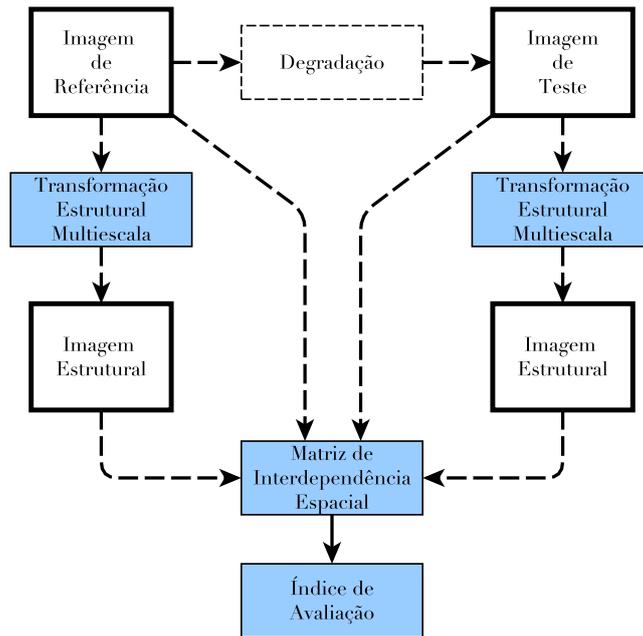


Figura 6.1: Modelo genérico de aplicação da MIDE e TEM na análise estrutural de imagens.

fontes de distorção representam as formas mais comuns de degradação estrutural em imagens digitais. A compressão é um recurso comum em armazenamento e transmissão de imagens. O ruído branco (ou ruído gaussiano) é inerente aos sistemas de captura de imagens, enquanto o borramento gaussiano é causado por distorções do sistema ótico e falhas de foco. A transmissão sem fio também é responsável pelo *fast fading*, causado pela combinação de múltiplos sinais que chegam no receptor devido a fenômenos como reflexão ou difração do sinal em obstáculos.

Os subconjuntos de imagens da LIVE estão listados na Tabela 6.1. Os experimentos para obtenção de um escore referência, o escore médio de opiniões (DMOS, do inglês *Difference Mean Opinion Score*), foram conduzidos em duas sessões para o JPEG2k, duas sessões para JPEG e uma sessão para as demais. O escore DMOS foi obtido a partir de uma média dos escores dos observadores, seguindo a metodologia detalhada em (SHEIKH; SABIR; BOVIK, 2006). Algumas imagens estão ilustradas na Figura 6.2.

DRIVE

A base DRIVE (*Digital Retinal Images for Vessel Extraction*) (STAAL *et al.*, 2004) contém 40 retinografias coloridas registradas em imagens de tamanho 565×584 pixels. As retinografias foram obtidas em um programa de exame de retinopatia

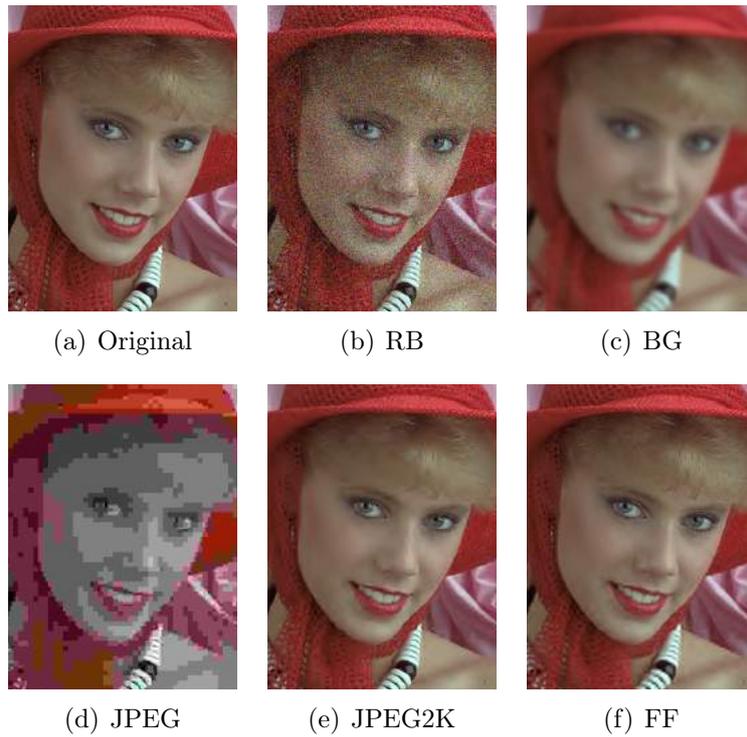


Figura 6.2: (a) Exemplo de imagem não degradada da base LIVE e (b-f) diferentes tipos de degradação.

Tabela 6.1: Subconjuntos da base LIVE, adaptado de (SHEIKH; SABIR; BOVIK, 2006)

Subconjunto	Número de imagens	Número de observadores
JPEG2k 1	116	29
JPEG2k 2	111	29
JPEG 1	116	29
JPEG 2	117	29
Ruído branco (RB)	174	23
Borramento gaussiano (BG)	174	24
<i>Fast fading</i> (FF)	174	20
Total	982	—

diabética nos Países Baixos. A base é disponibilizada publicamente e vem sendo amplamente utilizada em trabalhos de visão computacional para a detecção da rede de vasos, microaneurismas, mácula e disco ótico.

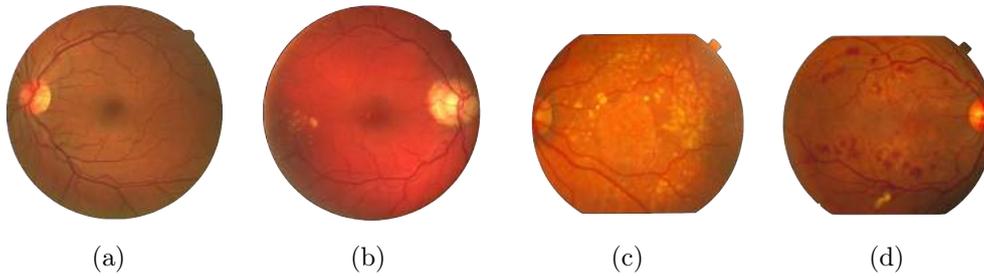


Figura 6.3: Exemplos de retinografias da base (a-b) DRIVE e (c-d) STARE.

STARE

A base STARE (*STructured Analysis of the REtina*) (HOOVER; KOUZNETSOVA; GOLDBAUM, 2000) é formada por 40 imagens de tamanho 700×605 , das quais 20 estão disponíveis para a segmentação da rede de vasos. Estas imagens estão separadas em dois grupos de mesmo tamanho contendo 10 imagens de retinas saudáveis e 10 imagens de retinopatia (FRAZ *et al.*, 2012). Esta base de imagens difere da base DRIVE por conter imagens de retinas bastante prejudicadas pela doença e, portanto, constitui um maior desafio para os métodos automáticos de segmentação de vasos.

Corrosão atmosférica

O banco de imagens corrosão atmosférica (MEDEIROS *et al.*, 2010) é um conjunto de 28 fotografias em cores de superfícies metálicas de equipamentos expostos a ambiente corrosivo. Cada imagem possui resolução espacial 1024×1024 pixels e registram tubulações e reservatórios de uma planta industrial de produção de lubrificantes. Três recortes de 120×120 pixels foram extraídos de cada imagem, contendo regiões corroídas e não corroídas. Assim, o banco de imagens está dividido em um subconjunto de 51 recortes de superfícies atacadas pela corrosão atmosférica (Classe 2) e outro subconjunto com 33 regiões similares (Classe 1). As regiões similares são superfícies não corroídas que contêm manchas ou apresentam algum tipo de rugosidade que pode ser confundida com a corrosão. Um especialista em manutenção industrial rotulou as imagens para geração do padrão ouro (PO).

A imagem da Figura 6.4 (a) ilustra as regiões monitoradas, ou cortes. Um subconjunto ilustrativo desses cortes é apresentado no mosaico da Figura 6.4 (b), no qual observa-se a variedade de superfícies com diferentes texturas e cores.

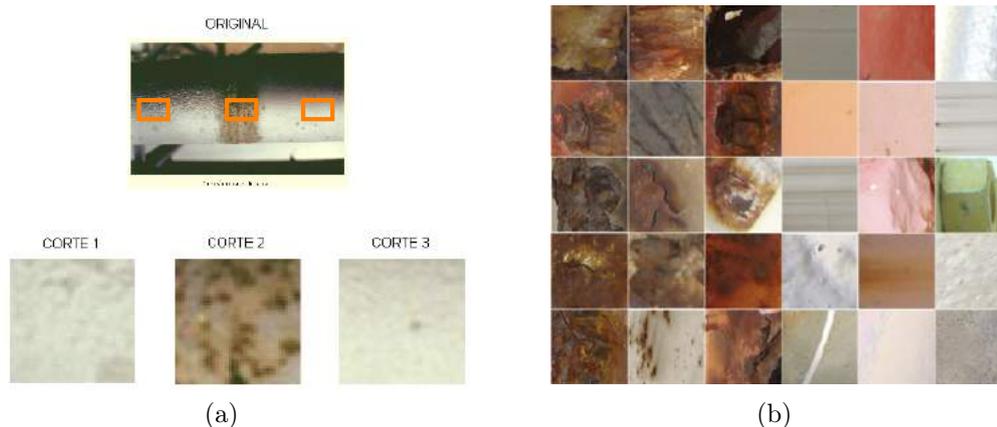


Figura 6.4: (a) Tubulação exposta à corrosão atmosférica destacando três regiões monitoradas. (b) Mosaico dos recortes de superfícies de equipamentos contendo corrosão atmosférica (três primeiras colunas) e livre de corrosão (demais colunas à direita).

Doenças pulmonares

O banco de imagens de doenças pulmonares contendo tomografia computadorizada (TC) do tórax foi utilizado em trabalhos de segmentação e análise de pulmões em trabalhos de Rebouças Filho (2013) e Felix, Cortez e Holanda (2010). As imagens foram obtidas em parceria com o Hospital Walter Cantídio em um estudo aprovado pelo Comitê de Ética da Universidade Federal do Ceará - COMEPE, sob Protocolo nº 35/06, seguindo a Resolução nº 196/96 do Conselho Nacional de Saúde que trata de pesquisas com seres humanos. Os pulmões foram segmentados manualmente por um especialista para obter um PO utilizado como referência na avaliação dos algoritmos. Cada imagem registra dois pulmões, esquerdo e direito, que podem ou não conter algum tipo de doença. Na maioria dos casos ambos os pulmões são saudáveis ou apresentam a mesma doença. Entretanto, existem imagens na qual somente um dos pulmões é afetado pela doença. As imagens contêm amostras de pulmão saudável (PS), fibrose cística (FC) e doença pulmonar obstrutiva crônica (DPOC), conforme exemplos da Figura 6.5. Importante destacar que as imagens foram adquiridas em 16 bits e, embora sejam representadas em 8 bits nas ilustrações, o processamento de análise é realizado na quantização original.

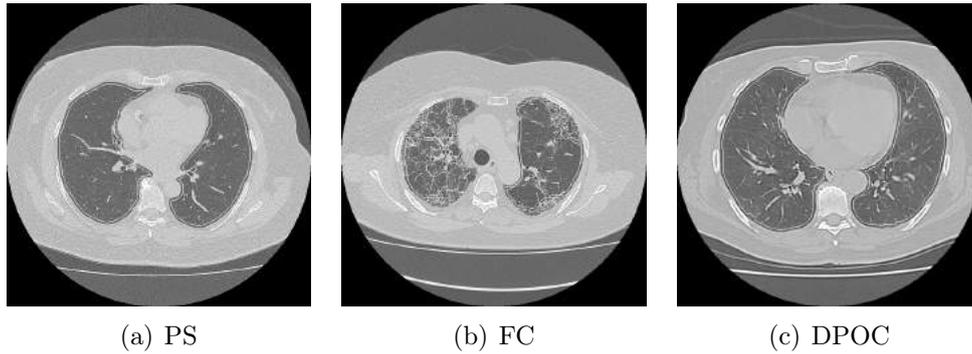


Figura 6.5: Exemplos da base de imagens de TC de pulmões.

6.2 Caso I: detecção de corrosão atmosférica

Equipamentos expostos às condições atmosféricas podem ter acelerado o fenômeno da corrosão, especialmente aqueles cujos parâmetros operacionais como temperatura e acidez do material armazenado ou transportado são elevados.

A detecção automática de superfícies afetadas por corrosão atmosférica é uma tarefa não trivial, porque envolve a detecção de diferenças sutis de rugosidade e também a necessidade de abstrair regiões escuras que podem ser confundidas com a corrosão. O diagrama da Figura 6.6 ilustra o modelo de análise estrutural com base na MIDE e na TEM para a detecção automática da corrosão atmosférica.

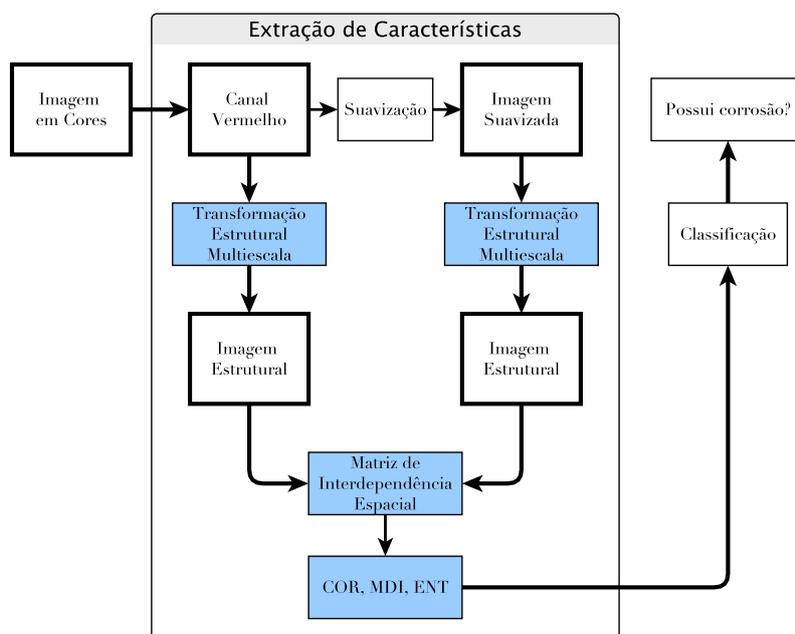


Figura 6.6: Modelo da detecção da corrosão atmosférica.

A extração de atributos no método proposto é realizada pela MIDE sobre a imagem realçada do canal vermelho, quando comparada com a imagem realçada do canal verde, após processo de suavização. A suavização atenua estruturas características da rugosidade de uma superfície corroída e a transformação tem por objetivo destacar essas estruturas. Os descritores da MIDE são utilizados para detectar as diferenças estruturais entre essas duas imagens.

O canal vermelho foi escolhido porque preserva mais informação sobre as regiões que contêm corrosão, já que estas possuem matiz predominante na faixa laranja e vermelho. A MIDE foi calculada para $N = 16$ níveis e o vetor de atributos é formado por $\mathbf{A}_{MIDE} = \{\text{COR}, \text{MDI}, 1 - \text{CHI}\}$, obtido sobre a TEM das imagens de entrada. Sejam $\mathbf{R}_I = \varphi(\mathbf{I})$ e $\mathbf{R}_J = \varphi(\mathbf{J})$, as transformações do canal vermelho e da sua versão suavizada, a MIDE é calculada como $\mathfrak{M}_n(\sum_n \Phi_n(\mathbf{R}_I), \sum_n \Phi_n(\mathbf{R}_J))$.

Para o cálculo da matriz de coocorrência de níveis de cinza (GLCM, do inglês *gray level cooccurrence matrix*) foram utilizados os mesmos $N = 16$ níveis. O vetor de atributos da GLCM foi obtido a partir do valor médio de correlação, contraste, homogeneidade e entropia, nas direções $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ para a distância $d = 1$, ou seja, $\mathbf{A}_{GLCM} = \{\overline{Cor}, \overline{Con}, \overline{Hom}, \overline{Ent}\}$.

A observação da Matriz-U (Apêndice A) da Figura 6.7 permite concluir que o espaço de características da MIDE apresenta uma melhor discriminação entre as duas classes. Isto é perceptível pela existência de duas regiões melhor definidas (superior e inferior) no mapa da MIDE do que no mapa da GLCM.

Para avaliar a capacidade de discriminação dos atributos da MIDE e da GLCM, foi realizado um experimento de classificação com o algoritmo KNN (WEBB, 2002). Foram adotados 2 vizinhos, um para cada classe. Os resultados deste experimento são mostrados na Figura 6.8. Os resultados de classificação para ambos conjuntos de atributos são apresentados, respectivamente, nas matrizes de confusão (Apêndice A) das Figuras 6.8 (a) e (c) e o desempenho geral é avaliado a partir da área sob a curva característica de operação do receptor (ROC, do inglês *receiver operating characteristic*), em função da taxa de verdadeiros positivos (TVP) e da taxa de falsos positivos (TFP), registradas na Figura 6.8 (b) e (d).

Usando os três atributos extraídos da MIDE, foi alcançado um bom resultado de classificação, quando comparado com atributos extraídos da GLCM. O algoritmo proposto obteve 90,5% de acerto com um percentual de falsos negativos 6,0%. O percentual de falsos negativos obtido com a GLCM que foi de 9,5%, com uma taxa de acerto de 88,1%. Embora a vantagem obtida com os descritores da MIDE seja

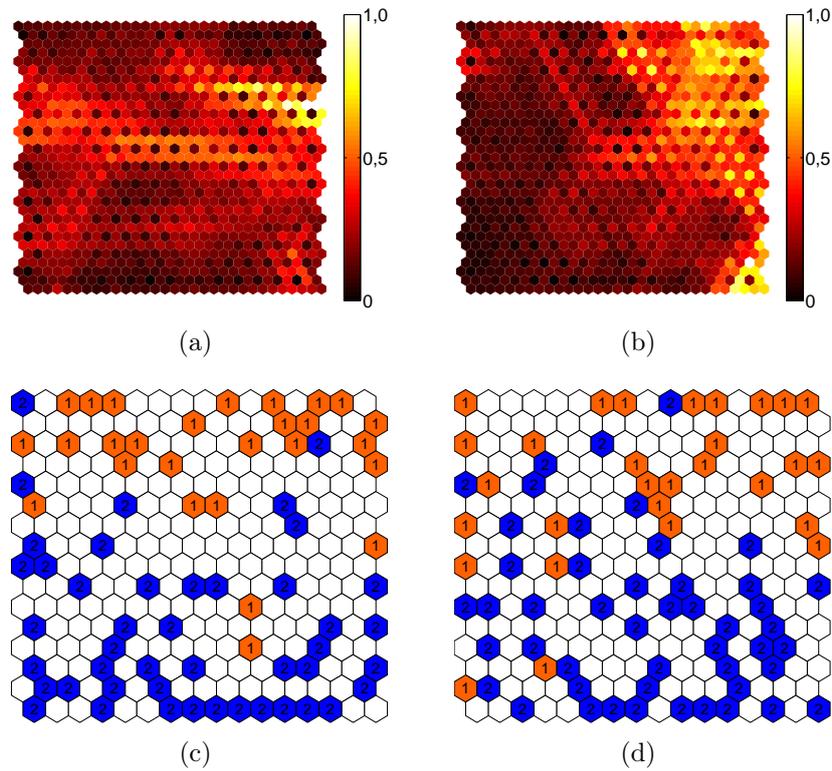


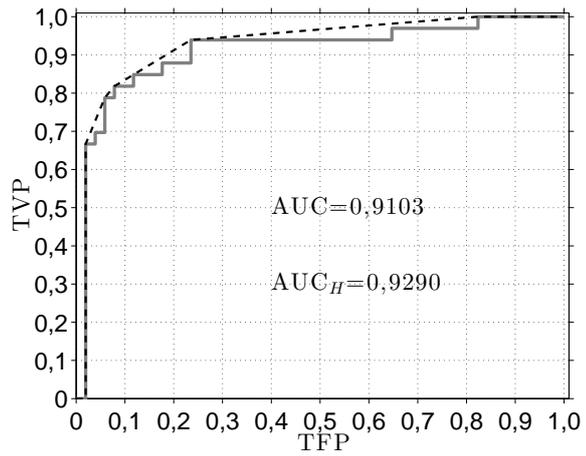
Figura 6.7: Mapa de descritores da (a,c) MIDE e (b,d) GLCM para os atributos de corrosão. As cores identificam as imagens sem corrosão (laranja) e com corrosão (azul).

pequena, o desempenho médio de classificação é superior, como pode ser comprovado pela área sob a curva ROC e área sob o fecho convexo da curva ROC, ilustrados nas Figuras 6.8 (b) e (d). As curvas ROC foram obtidas a partir de experimentos aleatórios de classificação, utilizando o discriminante linear de Fisher (WEBB, 2002). Foram realizadas 100 repetições, adotando-se o método de validação cruzada com a seleção aleatória de 70% das amostras na fase de treinamento e o restante na fase de teste.

A razão pela qual os atributos da MIDE discriminam melhor os dados se deve ao fato da mesma detectar a atenuação da rugosidade da superfície corroída. Por outro lado, a GLCM detecta evidências de textura, mesmo que esta não represente uma área atacada por corrosão. Este fato se observa em algumas das imagens classificadas incorretamente, as quais são exibidas na Figura 6.9.

Classe estimada	1	30 (35,7%)	5 (6,0%)	85,7% (14,3%)
	2	3 (3,6%)	46 (54,8%)	93,9% (6,1%)
		90,9% (9,1%)	90,2% (9,8%)	90,5% (9,5%)
		1	2	
		Classe conhecida		

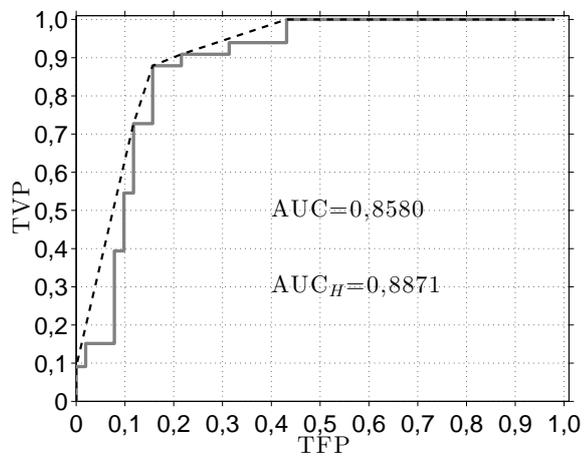
(a) MIDE



(b) MIDE

Classe estimada	1	31 (36,9%)	8 (9,5%)	79,5% (20,5%)
	2	2 (2,4%)	43 (51,2%)	95,6% (4,4%)
		93,9% (6,1%)	84,3% (15,7%)	88,1% (11,9%)
		1	2	
		Classe conhecida		

(c) GLCM



(d) GLCM

Figura 6.8: (a,c) Matrizes de confusão e (b,d) curvas ROC ilustrativos do desempenho da detecção de corrosão atmosférica (Classe 2) utilizando atributos extraídos da MIDE e da GLCM.



Figura 6.9: Exemplo de imagens de corrosão classificadas incorretamente usando o conjunto de descritores da (a-d) MIDE e (e-g) GLCM.

6.3 Caso II: detecção de doenças em TC de pulmões

Um único exame de TC pulmonar pode conter centenas de imagens. A busca visual por evidências de doenças é um processo desgastante e, muitas vezes, só pode ser realizado por especialistas com grande experiência (REBOUÇAS FILHO, 2013). Esse processo pode ser bastante simplificado, quando um algoritmo computacional é capaz de detetar e classificar as imagens de exames em grupos, discriminando o tipo de doença ou a ausência de enfermidades.

A extração de características de imagens de TC requer que os pulmões estejam previamente segmentados. Nesta tese, a segmentação manual realizada pelo especialista foi utilizada como PO. Entretanto, algoritmos de segmentação automática podem ser aplicados (REBOUÇAS FILHO, 2013).

O procedimento de extração de atributos das imagens está ilustrado na Figura 6.10, contudo sem a necessidade da etapa de transformação. Neste experimento foram utilizados os atributos COR, MDI e $1 - CHI$ da MIDE aplicada na imagem original e sua cópia suavizada com uma função gaussiana em uma janela 3×3 com $\sigma = 1$. Para a escolha da função de suavização foi adotado o critério do menor custo computacional, daí o uso do menor tamanho de janela possível e do menor desvio padrão capazes de causar um nível significativo de degradação estrutural. Para o cálculo da MIDE foi utilizado $N = 64$ sobre a região da imagem que corresponde ao PO. Cada pulmão foi avaliado individualmente. O procedimento está ilustrado no diagrama da Figura 6.10.

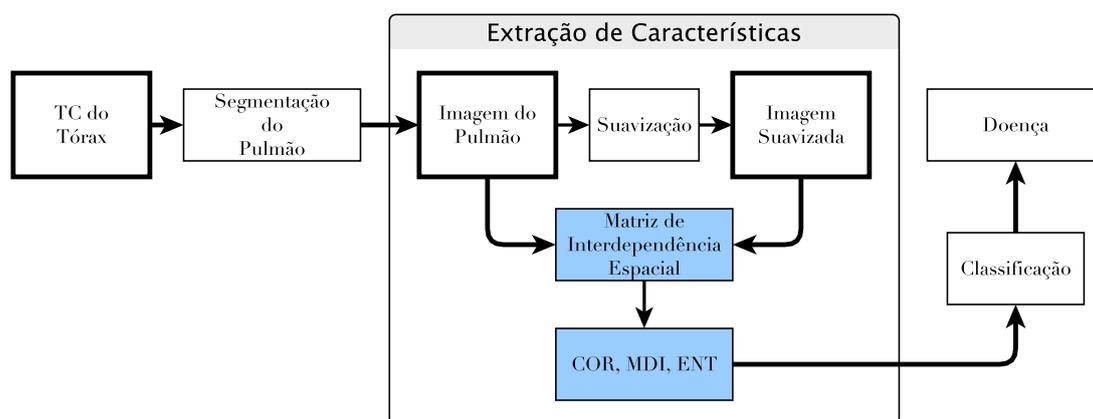


Figura 6.10: Modelo da classificação de doenças pulmonares.

As imagens da DPOC apresentam um nível de intensidade médio baixo com a presença de estruturas claras maiores. Por isso os valores obtidos com os atributos da

MIDE são significativamente maiores, já que não existem muitas estruturas salientes afetadas pela suavização.

As imagens com FC, por sua vez, apresentam uma textura característica da doença e, portanto, é mais susceptível a alterações estruturais causadas pela suavização, podendo ser detetado pelos atributos MDI e CHI. Por outro lado, as estruturas da fibrose estão distribuídas pela maior parte dos pulmões e possuem um tamanho relativo maior, quando comparado com estruturas salientes como os vasos. Por isso o valor COR é bastante elevado com relação aos outros dois atributos.

Nas imagens do PS, existe mais uniformidade, porém alguns vasos salientes são bastante degradados pela suavização. Portanto, os valores dos atributos da MIDE são mais próximos de zero, em geral, do que nos demais casos.

O conjunto de atributos extraídos com a MIDE foi utilizado para classificar um conjunto de imagens contendo 27 amostras de pulmões sadios (PS), 24 amostras contendo doença pulmonar obstrutiva crônica (DPOC) e 21 amostras de pulmões com diferentes níveis de fibrose cística (FC), totalizando 72 imagens de TC contendo pulmões esquerdo e direito. O conjunto foi previamente avaliado por um especialista em doenças pulmonares, que atribuiu rótulos para cada imagem, identificando o tipo de doença.

A análise da Figura 6.11 permite comparar o desempenho dos descritores do método proposto e com dois métodos alternativos, utilizando descritores extraídos com a GLCM e o índice de qualidade de imagens do método Informação de Fidelidade Visual (VIF, do inglês *Visual Information Fidelity*). O aspecto visual das imagens apresenta uma padrão de textura que a GLCM é capaz de detetar. Pelo mesmo motivo, optou-se por testar também a hipótese de que o índice VIF poderia descrever a perda de qualidade visual referente à etapa de suavização. Entretanto, a análise da projeção dos seus descritores, conforme ilustram as matrizes-U das Figuras 6.11 (a-c), torna perceptível que ambas GLCM e VIF não conseguem discriminar corretamente os dados em três classes, fato evidenciado no espaço de descritores da Figuras 6.11 (d-f).

Na Figura 6.12 é apresentado um diagrama que destaca as imagens e suas respectivas matrizes \mathfrak{M} dos elementos que estão na fronteira entre as classes de doenças. As matrizes exibem padrões característicos para cada pulmão e é possível perceber similaridades entre os pulmões de mesma classe. Por exemplo, observa-se um padrão de dispersão em torno da diagonal além da forma retangular para a classe FC, o formato de “X” para a classe DPOC e a forma de “8” para a classe

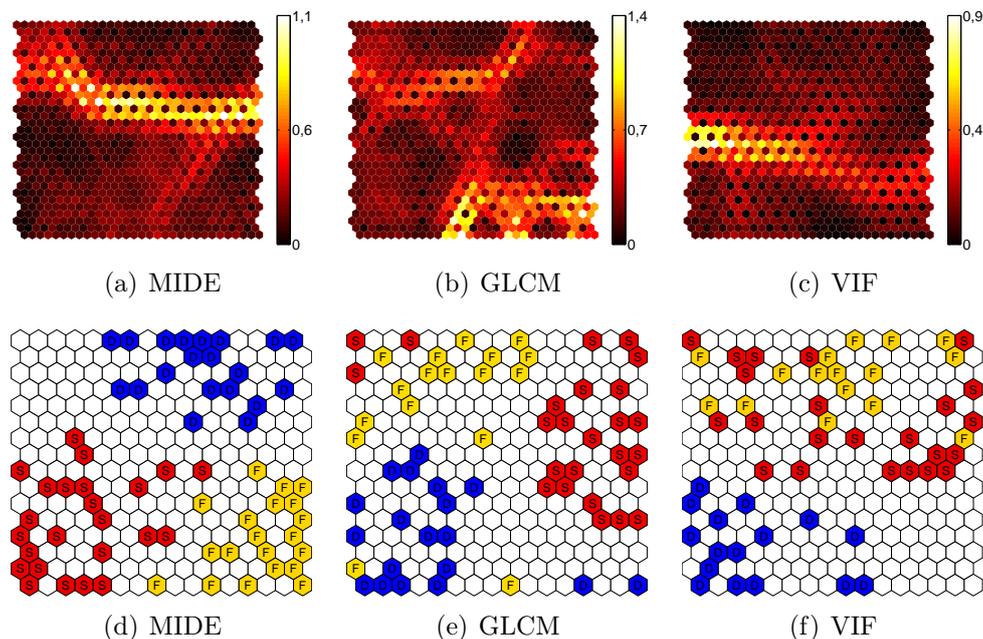


Figura 6.11: (a-c) Matriz-U da MIDE, GLCM e VIF para classificação de TC de pulmão. (d-f) Espaço de descritores em que as cores identificam as classes PS (vermelho), FC (amarelo) e DPOC (azul).

PS. Cada padrão representa uma relação característica entre as estruturas claras e escuras presentes na imagem original com relação à imagem degradada. A maior dispersão em torno da diagonal mostra que estruturas de alto contraste com sua vizinhança foram degradadas, como nas imagens DPOC, ou que elas são pequenas e distribuídas, como nas imagens PS e FC. Ambas matrizes das Figuras 6.12 (a) e (e) apresentam uma região com padrão semelhante em forma de “V”, mostrando que os respectivos pulmões possuem alto contraste entre estruturas vizinhas. Contudo, no caso da imagem do pulmão com FC da Figura 6.12 (e), esse contraste está relacionado à presença de ruído causado pela quantização, enquanto na Figura 6.12 (a) o contraste é inerente à anatomia do pulmão com DPOC, cujas estruturas características são maiores. Também vale destacar que os pulmões da classe FC que estão na fronteira com a classe PS apresentam uma quantidade pequena de lesões. Já os pulmões da classe PS que estão próximos da classe FC apresentam vasos que podem ser confundidos morfologicamente com a anatomia da fibrose.

Para avaliar o poder discriminante desse conjunto de atributos, as imagens foram rotuladas utilizando uma rede Máquina de Aprendizagem Extremo (ELM, do inglês *Extreme Learning Machine*) (Apêndice A) em um experimento de classificação. O vetor de atributos extraídos da MIDE, $\mathbf{A} = \{\text{COR}, \text{MDI}, 1 - \text{CHI}\}$, foi utilizado nas

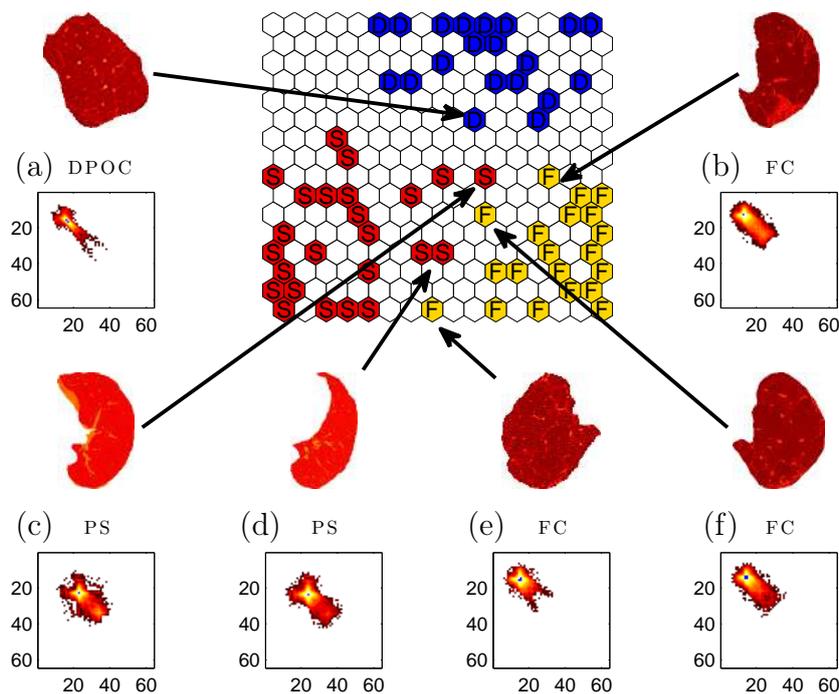


Figura 6.12: Exemplos de pulmões presentes na fronteira das classes e suas respectivas matrizes MIDE.

etapas de treinamento e teste da rede. A rede foi treinada usando 70% dos vetores no treinamento e 30% para teste em um procedimento de validação cruzada com 100 rodadas treinamento-teste. Em cada rodada os vetores de atributos são selecionados aleatoriamente sem repetição. A arquitetura da rede contém HN neurônios na camada oculta e 3 neurônios na camada de saída, uma para cada classe. O índice do neurônio da saída da rede com maior valor é utilizado para rotular a amostra de teste. Os melhores resultados estão registrados na Tabela 6.2 para uma arquitetura com $HN = 4$. O número de neurônios da camada escondida foi determinado por meio de busca exaustiva. Observa-se que os pulmões com DPOC foram classificados corretamente em 100% dos casos. Já para a doença FC, existe uma pequena taxa de falso negativo (Apêndice A).

Os dados da matriz de confusão da Tabela 6.2 revelam uma boa discriminação da doença DPOC, sendo que todas as imagens dessa classe foram corretamente classificadas. Para as classes FC e DPOC, todas as amostras foram corretamente identificadas. Cerca de 13% dos pulmões saudáveis foram classificados como sendo da classe fibrose (falsos positivos), o que comprometeu o desempenho medido pela sensibilidade e especificidade, ficando abaixo de 70% e 90%, respectivamente. Embora o algoritmo não evite falsos alarmes, não houve confusão entre as doenças, o

Tabela 6.2: Matriz de confusão para os atributos da MIDE

Classe	Predição		
	PS	FC	DPOC
PS	0,2500	0,0000	0,0000
FC	0,1250	0,2917	0,0000
DPOC	0,0000	0,0000	0,3333
Acurácia	0,9608		
Especificidade	0,8333	1,0000	1,0000
Sensitividade	0,6667	1,0000	1,0000

que é um fator importante a ser considerado. O resultado global próximo de 96% de acurácia (Apêndice A) e a correta identificação das doenças (verdadeiros positivos) é bastante favorável ao uso da metodologia como ferramenta de pré-diagnóstico, triagem ou rastreamento de doenças pulmonares a partir de imagens de TC.

Para efeito de comparação, o mesmo algoritmo de classificação e procedimento de treinamento e validação foram utilizados para o conjunto de atributos extraídos da GLCM com os parâmetros $\delta = 1$ e $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Os resultados estão registrados na Tabela 6.3. O desempenho global, indicado pela acurácia, comprova que os descritores da MIDE permitem uma melhor discriminação das classes do que os descritores da GLCM. Porém, é importante destacar que um fator de análise mais relevante deve levar em consideração a quantidade de falsos positivos e falsos negativos para a detecção de doenças. Neste caso, a sensibilidade dos atributos da GLCM mostra que são perfeitos para indicar a presença de DPOC.

Tabela 6.3: Matriz de confusão para os atributos da GLCM

Classe	Predição		
	PS	FC	DPOC
PS	0,2917	0,0417	0,0417
FC	0,0833	0,2500	0,0000
DPOC	0,0000	0,0000	0,2917
Acurácia	0,8962		
Especificidade	0,8667	0,9333	0,9286
Sensitividade	0,7778	0,8571	0,8750

6.4 Caso III: avaliação da qualidade de imagens

A avaliação da qualidade de imagens é uma tarefa comum para as pessoas, sendo que o sistema visual humano (HVS, do inglês *human visual system*) é bastante adaptado para essa tarefa (WANG *et al.*, 2004). O algoritmo proposto por Sheikh e Bovik (2006) é capaz de aproximar essa função do HVS com excelentes resultados e boa capacidade de generalização como medida de avaliação da qualidade de imagens para degradações comuns em imagens digitais, a saber, borramento, ruído, compressão e *fast fading*, decorrentes da captura, armazenamento e transmissão.

A avaliação da qualidade de imagens com a MIDE emprega o procedimento proposto na Figura 6.13. Neste processo, a imagem de referência livre de degradação é comparada com sua versão degradada usando a MIDE. Os atributos extraídos da MIDE são combinados para estimar o nível de degradação da imagem ou aproximar a percepção de qualidade visual do HVS. A etapa de transformação para obtenção de uma imagem estrutural tanto da imagem de referência, quanto da imagem degradada é opcional. De fato, esta etapa somente é empregada quando se deseja avaliar uma faixa específica das escalas das estruturas. Para o estudo apresentado a seguir esta etapa não foi necessária.

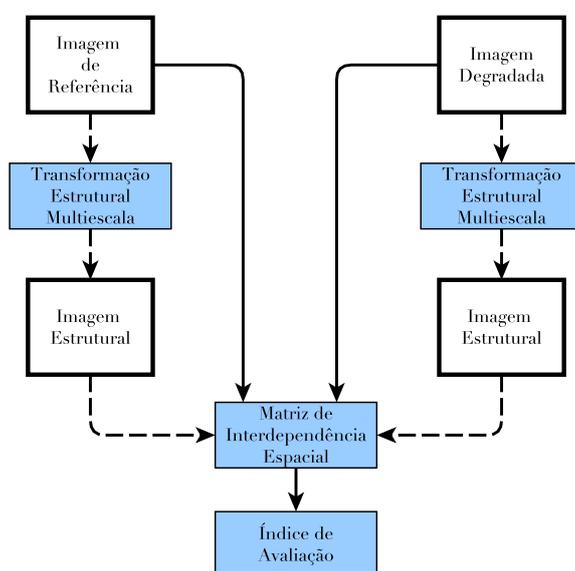


Figura 6.13: Modelo para avaliação da qualidade de imagens. A linha pontilhada indica um procedimento opcional.

Neste estudo foram avaliadas 982 imagens em cores da base LIVE, das quais 779 contêm algum tipo de degradação causada por ruído, borramento ou compressão,

conforme apresentado na Tabela 6.1. Para obtenção de um escore de avaliação objetiva da qualidade de imagens, foi utilizada a MIDE com $N = 32$ níveis. A avaliação foi realizada sobre a luminância das imagens, ou seja, sobre a imagem convertida para nível de cinza. Não há variação significativa nos resultados, quando o teste é feito sobre cada canal individualmente de cor (R,G ou B) e depois é extraída a média dos escores dos três canais.

O exemplo ilustrado na Figura 6.14 permite comparar graficamente os métodos de estimação da qualidade de imagens com os atributos extraídos da MIDE. As degradações combinadas são: ruído gaussiano, nível médio, quantização e borramento. Para comparação, os gráficos mostram o resultado do método de avaliação pela similaridade estrutural (SSIM, do inglês *structural similarity*) e pela GLCM, baseado em Gadkari (2004). Observa-se que, quando diferentes tipos de degradação são combinados, a avaliação da qualidade pela GLCM e pela SSIM apresentam mudança de sinal na sua primeira derivada. Em outras palavras, essas medidas podem apresentar incoerências na comparação da qualidade de imagens degradadas gradualmente, fornecendo valores próximos para níveis de degradação significativamente diferentes. Isto pode ser observado na Figura 6.14 (a), entre os níveis 2 e 6 e na Figura 6.14 (b), entre os níveis 1 e 5. Na avaliação realizada com a MIDE, por sua vez, apenas a medida COR apresentou discordância na representação do valor gradual da degradação, entre os níveis 3 e 4 da Figura 6.14 (b). Observa-se, portanto, que os atributos da MIDE representam melhor a relação entre a perda de qualidade visual e o aumento do nível de degradação.

Índice de qualidade estrutural

As medidas COR e MDI, extraídas da MIDE, apresentam comportamentos distintos para diferentes níveis de degradação estrutural. Com isso é possível obter resultados diferentes quando é realizada uma combinação entre seus valores. A combinação entre os índices COR e MDI é obtida por meio da operação de soma, ponderada por λ , dando origem ao índice de qualidade estrutural *iqe*, conforme definido pela equação 4.8.

O valor de λ deve ser ajustado para a base de imagens em estudo. Nos experimentos realizados, os melhores resultados foram obtidos para $0,90 < \lambda < 1$. A maior parte da contribuição para o índice de qualidade *iqe* é determinado pelo atributo COR. Contudo, os resultados experimentais mostram que o desempenho global para a base LIVE diminui quando $\lambda > 0,98$.

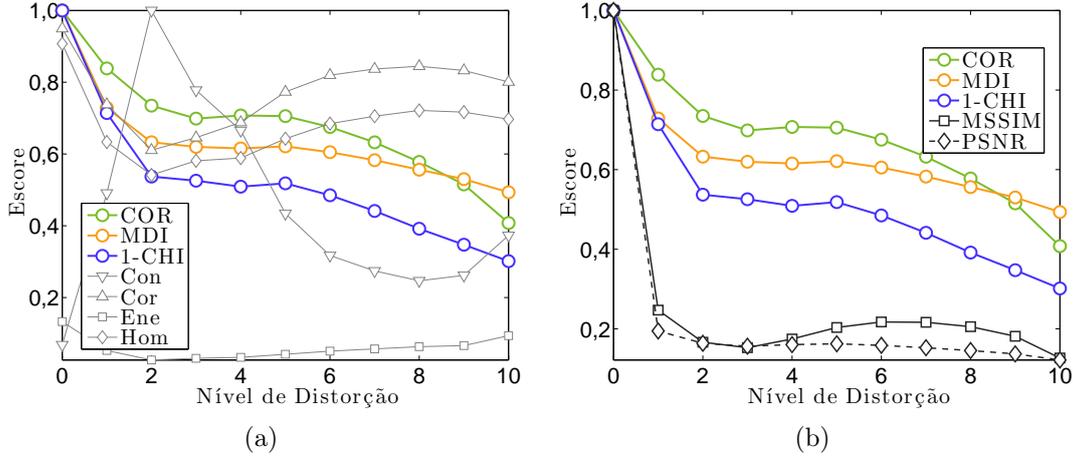


Figura 6.14: Exemplo da análise de qualidade com MIDE, GLCM (GADKARI, 2004) e MSSIM (WANG *et al.*, 2004), para efeitos combinados de 4 diferentes tipos de degradação. (a) GLCM \times MIDE. (b) MSSIM \times MIDE. A medida de referência PSNR está normalizada como $\text{PSNR}/\max(\text{PSNR})$.

Índice de qualidade estrutural multiescala

O índice de qualidade estrutural multiescala iqe_m é obtido quando a equação 4.8 do iqe é aplicada aos cortes das imagens de entrada transformadas pela TEM. Considere $\mathbf{R}_I = \varphi(\mathbf{I})$ e $\mathbf{R}_J = \varphi(\mathbf{J})$ imagens estruturais obtidas por meio da TEM para a imagem de entrada \mathbf{I} e sua versão degradada \mathbf{J} . Para cada par de cortes $\Phi_n(\mathbf{R}_I)$ e $\Phi_n(\mathbf{R}_J)$ é obtida uma MIDE, ou seja, $\mathfrak{M}_n(\Phi_n(\mathbf{R}_I), \Phi_n(\mathbf{R}_J))$. O índice de qualidade multiescala é calculado como a média do iqe_n em cada escala n e é definido como

$$\text{iqe}_m = \frac{1}{N} \sum_{n=1}^N \text{iqe}_n, \quad (6.1)$$

em que N é o número de escalas da transformação TEM.

Os valores listados na Tabela 6.4 apresentam a correlação obtida entre uma medida de qualidade de imagem (IQM, do inglês *image quality measure*) e escores DMOS assinalados por observadores para as imagens da base LIVE. Conforme a metodologia de avaliação dos métodos IQM adotada por Sheikh e Bovik (2006), foram utilizados três métodos de comparação, a saber: i) a correlação linear de Pearson aplicada sobre a regressão não linear dos IQM, utilizando uma função de regressão logística; ii) a correlação de Spearman, que avalia a dependência entre variáveis ordenadas em faixas; iii) a correlação de Kendall, que mede a similaridade

entre variáveis ordenadas com relação a seus pares concordantes e discordantes. A regressão não linear utilizada é uma função logística com termo linear (SHEIKH; BOVIK, 2006), definida pela expressão

$$\text{IQM}_{\text{aj}} = \beta_1 \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(\text{IQM} - \beta_3))} + \beta_4 x + \beta_5, \quad (6.2)$$

com os parâmetros $\beta_1 = \max(\text{DMOS})$, $\beta_2 = \min(\text{DMOS})$, $\beta_3 = \overline{\text{IQM}}$ e $\beta_4 = \beta_5 = 0,1$, sendo $\overline{\text{IQM}}$ o valor médio do IQM. Essa função realiza o ajuste IQM_{aj} da medida IQM para o espaço do valor de referência DMOS utilizado na avaliação de desempenho dos algoritmos.

Tabela 6.4: Comparação entre medidas de avaliação da qualidade de imagem sobre a base LIVE

Medida	EQM	Correlação		
		Pearson	Spearman	Kendall
VIF	6,8427	0,9758	0,9765	0,8667
MS-SSIM	7,9843	0,9669	0,9746	0,8689
MSSIM	10,4773	0,9423	0,9538	0,8207
<i>iqe</i>	11,1090	0,9349	0,9407	0,7934
<i>iqe_m</i>	11,9918	0,9237	0,9328	0,7876
MSVD	14,4361	0,8872	0,8891	0,7302

Os dados da Tabela 6.5 revelam detalhes sobre o desempenho de *iqe* e *iqe_m*, quando comparado aos índices de qualidade MSSIM, MS-SSIM e VIF, para cada subconjunto de degradação da base LIVE. Observa-se que o *iqe_m* apresenta um desempenho próximo ao dos demais índices para as degradações por compressão (JPEG2k e JPEG). Além disso, o *iqe_m* supera o desempenho da medida MSSIM e aproxima a da VIF a ponto de ser indistinguível, para as degradações por borramento (BG) e *fast fading* (FF). O desempenho do *iqe_m* só não é comparável aos outros índices para a degradação por ruído aditivo (RB).

A Tabela 6.6 contém um resumo que revela quais métodos o *iqe* melhor aproxima para os diferentes tipos de degradação. Em geral, o *iqe* apresenta resultados similares ao MSSIM para a compressão e borramento gaussiano, enquanto aproxima o MS-SSIM para ruído branco e *fast fading*. A mesma análise foi realizada para o *iqe_m* na Tabela 6.7. Embora o desempenho global do *iqe* seja melhor do que o

Tabela 6.5: Análise comparativa do desempenho de iqe e iqe_m com MSSIM, MS-SSIM e VIF para os subconjuntos da base LIVE

Grupo	Medida	EQM	Correlação		
			Pearson	Spearman	Kendall
JPEG2k	iqe	9,5087	0,9444	0,9498	0,8127
	iqe_m	6,4546	0,9748	0,9743	0,8656
	MSSIM	7,3609	0,9670	0,9729	0,8636
	MS-SSIM	5,4767	0,9819	0,9843	0,8966
	VIF	4,7595	0,9864	0,9790	0,8815
JPEG	iqe	10,7948	0,9519	0,9489	0,8104
	iqe_m	7,1483	0,9792	0,9757	0,8791
	MSSIM	8,5763	0,9699	0,9700	0,8623
	MS-SSIM	5,3228	0,9885	0,9858	0,9155
	VIF	4,4494	0,9920	0,9799	0,8959
RB	iqe	4,1371	0,9922	0,9861	0,9019
	iqe_m	8,1377	0,9693	0,9697	0,8550
	MSSIM	8,0396	0,9701	0,9784	0,8834
	MS-SSIM	4,7593	0,9896	0,9845	0,9037
	VIF	3,7947	0,9934	0,9894	0,9153
BG	iqe	9,8021	0,9126	0,8884	0,7247
	iqe_m	4,6062	0,9814	0,9802	0,8831
	MSSIM	8,7178	0,9315	0,9386	0,7958
	MS-SSIM	6,2840	0,9650	0,9734	0,8641
	VIF	4,2027	0,9845	0,9819	0,8880
FF	iqe	10,7820	0,9389	0,9436	0,8023
	iqe_m	7,1583	0,9735	0,9753	0,8705
	MSSIM	8,5165	0,9623	0,9659	0,8449
	MS-SSIM	8,6277	0,9613	0,9692	0,8647
	VIF	6,3682	0,9791	0,9774	0,8740

iqe_m , a análise da Tabela 6.7 permite concluir que o iqe_m aproxima melhor o VIF para degradações por borramento e *fast fading*. Para as outras degradações, o iqe_m apresenta um padrão mais próximo ao MSSIM.

Tabela 6.6: Medidas de avaliação da qualidade de imagem das quais o iqe mais se aproxima considerando os tipos de degradação da base LIVE

Grupo	EQM	Correlação		
		Pearson	Spearman	Kendall
JPEG2k	MSSIM	MSSIM	MSSIM	MSSIM
JPEG	MSSIM	MSSIM	VIF	MSSIM
RB	VIF	VIF	MS-SSIM	MS-SSIM
BG	MSSIM	MSSIM	VIF	MSSIM
FF	MS-SSIM	MS-SSIM	MS-SSIM	MS-SSIM

Tabela 6.7: Medidas de avaliação da qualidade de imagem das quais o iqe_{ms} mais se aproxima considerando os tipos de degradação da base LIVE

Grupo	EQM	Correlação		
		Pearson	Spearman	Kendall
JPEG2k	MSVD	MSVD	MSSIM	MSSIM
JPEG	MSSIM	MSSIM	VIF	MSSIM
RB	MSSIM	MSSIM	MSSIM	MSSIM
BG	VIF	VIF	VIF	VIF
FF	VIF	VIF	VIF	VIF

A maior dispersão em torno da curva de regressão, observada na Figura 6.15 para as medidas iqe e MSVD, contrapõe a das medidas iqe_m , MSSIM e MS-SSIM e, principalmente, o bom desempenho global da VIF. Para certos tipos de degradação, os valores estimados são mais próximos de 1 (baixo nível de degradação), mesmo que a percepção da qualidade avaliada pelo HVS indique um nível elevado de degradação visual. Portanto, exceto para a VIF, a diferença de interpretação dos valores estimados pelo IQM para diferentes tipos de degradação aumenta a dispersão em torno da curva de ajuste, o que resulta na diminuição dos coeficientes de correlação.

O tempo médio de processamento para uma imagem está registrado na Tabela 6.8. O tempo foi medido para os algoritmos originais, disponibilizados pelos respectivos autores, implementados na plataforma MATLAB. As medições de tempo foram realizadas em um único computador de quatro núcleos de 3GHz e 8GB de memória RAM. O método de avaliação de imagens iqe é, em média, duas vezes mais

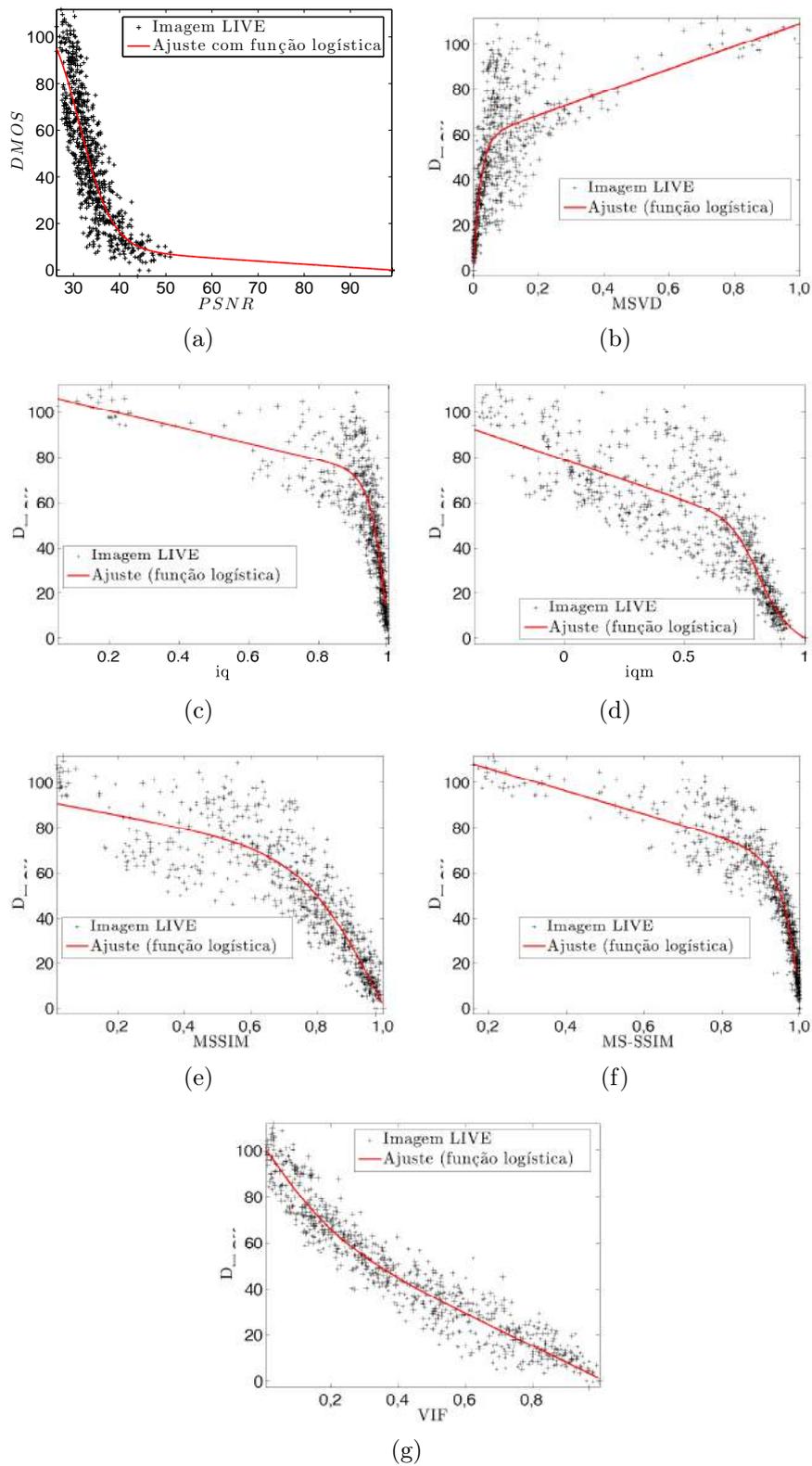


Figura 6.15: Regressão logística das medidas de qualidade de imagem.

rápido do que o MSSIM, que é o mais rápido dentre os métodos de comparação. O método de avaliação de imagens iqe_m é mais rápido, em média, do que o MSVD e o VIF. O tempo de processamento do iqe_m se divide entre a transformação TEM (0,59s) e o cálculo da MIDE para todas as 8 escalas da transformação (0,52s). Importante observar que estes são valores médios obtidos para as imagens da base LIVE.

Tabela 6.8: Tempo de processamento das medidas de avaliação da qualidade de imagens

iqe	iqe_m	MSSIM	MS-SSIM	MSVD	VIF
0,04s	1,11s	0,08s	0,13s	1,28s	1,38s

O tempo médio de processamento para as 982 imagens da base LIVE é de 0,08 segundos para o algoritmo SSIM e 0,04 para o iqe , respectivamente. Isto representa um tempo de processamento 43% menor, o que pode ser bastante significativo para aplicações que envolvem vídeo. O iqe_m também apresentou um tempo de processamento menor do que os métodos MSVD e VIF. Entretanto, dentre as versões multiescala, somente o MS-SSIM é rápido o suficiente para aplicações de tempo real. Estes tempos foram obtidos em um processador de 4 núcleos de 3,2GHz com 8GB de memória, rodando algoritmos implementados no **MATLAB**. O tempo gasto com operações de entrada e saída não foi computado. Uma comparação mais detalhada é exibida no gráfico *boxplot* (Apêndice A) da Figura 6.16. Cada retângulo representa a distribuição dos tempos medidos para cada imagem da base LIVE. Observa-se que o maior tempo de processamento de ambos iqe e iqe_m , registrado na extremidade superior do *boxplot*, é significativamente menor do que o maior tempo obtido com os demais métodos, exceto quando comparado ao PSNR. Além disso, o gráfico indica que o tempo de 75% das imagens processadas pelos métodos propostos está abaixo do percentil 50 (segundo quartil) das medidas MSSIM, MS-SSIM e VIF. Isto representa um ganho significativo de desempenho computacional para aplicações em tempo real.

A análise dos resultados das medidas de avaliação da Tabela 6.4, detalhados na Tabela 6.5, permite concluir que o VIF é a medida que melhor se aproxima globalmente do HVS para as diferentes degradações, porque apresentam menor EQM e maiores coeficientes de correlação. Em sequência aparecem as medidas MS-SSIM e MSSIM, em termos de desempenho global. Essa conclusão é possível ao analisar os gráficos de regressão logística da Figura 6.15. As medidas que apresentam menor dispersão em torno da curva de regressão logística são aquelas com melhor

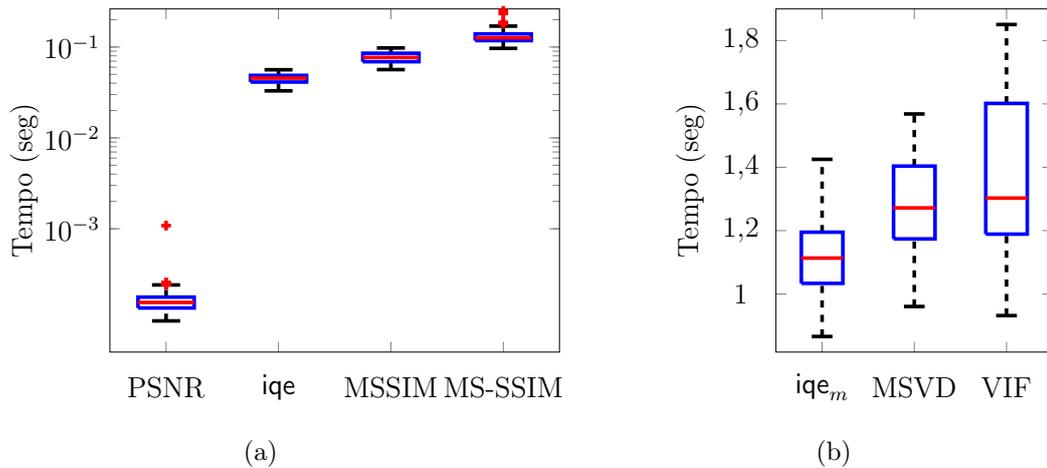


Figura 6.16: Tempo de processamento por imagem para a base LIVE. (a) Métodos mais rápidos (b) Métodos mais lentos.

resultado global. A medida VIF apresenta valores mais fortemente correlacionados independente do tipo de degradação.

Contudo, a medida de avaliação da qualidade VIF, desenvolvida com base em um modelo do HVS, é bastante especializada. Importante destacar que a medidas propostas *iqe* e *iqe_m*, derivadas da MIDE, aproximam o resultado da VIF para casos específicos de degradação. Portanto, pode-se afirmar que a MIDE oferece uma alternativa à avaliação da qualidade de imagens com um menor número de parâmetros de ajustes e também com um custo computacional menor.

6.5 Caso IV: segmentação de vasos de retina

A análise visual da rede de vasos de retina é uma tarefa que requer treinamento e conhecimento técnico especializado. Entretanto, mesmos os especialistas discordam da correta identificação da rede de vasos (HOOVER; KOUZNETSOVA; GOLDBAUM, 2000) e, conseqüentemente, do diagnóstico. A comunidade médica aceita que a detecção automática da rede de vasos é o primeiro passo para o desenvolvimento de um Sistema de Diagnóstico Assistido por Computador (CAD) para desordens oftálmicas (FRAZ *et al.*, 2012).

A segmentação da rede de vasos a partir de retinografias digitais é uma tarefa que vem sendo abordada por diferentes métodos na literatura, incluindo a Morfologia Matemática (MM), curvatura e mistura gaussiana com ferramentas de análise estrutural (SOARES *et al.*, 2006; ZANA; KLEIN, 2001; MENDONCA; CAMPILHO, 2006; FRAZ *et al.*, 2011; FRAZ *et al.*, 2012; VILLALOBOS-CASTALDI;

FELIPE-RIVERÓN; SÁNCHEZ-FERNÁNDEZ, 2010). A retinografia é uma imagem colorida formada pelos canais vermelho \mathbf{I}_R , verde \mathbf{I}_G e azul \mathbf{I}_B no sistema RGB de representação de cores. Alguns métodos utilizam o canal verde da retinografia, porque ele apresenta um maior contraste da rede de vasos (FRAZ *et al.*, 2012),(SOARES *et al.*, 2006), (ZANA; KLEIN, 2001).

A imagem da Figura 6.17 é uma representação tridimensional do canal verde de uma retinografia. Um dos problemas da segmentação desse tipo de imagem é que os objetos de interesse (os vasos) estão inseridos em um fundo ruidoso que, em determinadas regiões, possui a mesma intensidade dos vasos, especialmente os de menor calibre. Por isso, é comum adotar a estratégia do realce do canal verde como etapa preparatória da imagem para a fase de segmentação (FRAZ *et al.*, 2012). Diversos trabalhos abordam essa questão utilizando a superposição de efeitos por operações *top hat* (MENDONCA; CAMPILHO, 2006; ZANA; KLEIN, 2001).

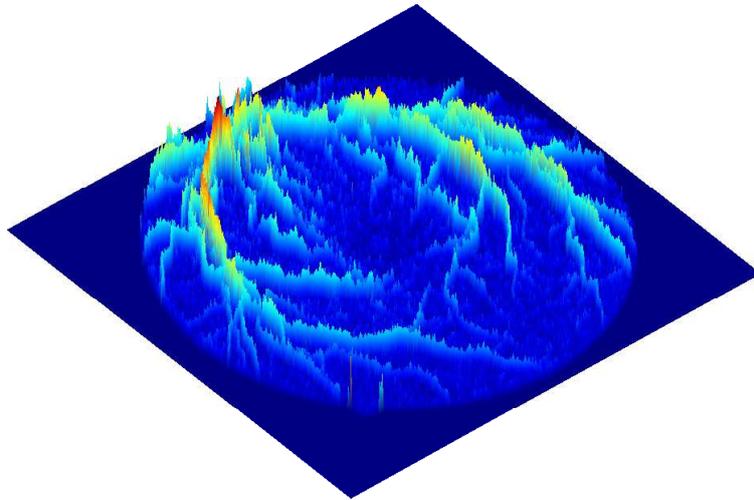


Figura 6.17: Representação 3D da detecção da rede de vasos em retinografias.

A característica multiescala da TEM a torna uma ferramenta bastante útil para detetar estruturas específicas em determinadas escalas, especialmente quando a TEM generalizada é utilizada na forma da equação 5.3.

O método de segmentação de vasos proposto é apresentado na Figura 6.18. A partir do canal verde \mathbf{I}_G da retinografia, é obtida uma representação estrutural $\mathbf{R} = \varphi(\mathbf{I}_G)$ usando a transformação TEM com uma função de ponderação $f_p = \{12,11,\dots,1\}$. Esta função foi escolhida com base nos seguintes critérios:

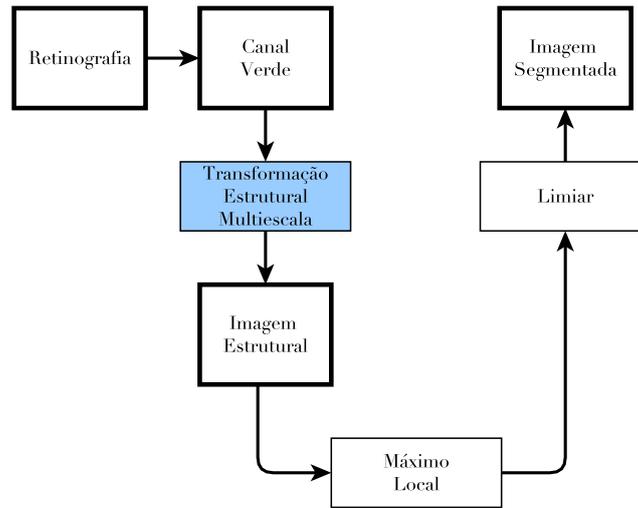


Figura 6.18: Segmentação da rede de vasos de retina baseada na TEM.

- ▶ Assume-se que o diâmetro médio dos vasos possui de 6 a 12 pixels (SOARES *et al.*, 2006; ZANA; KLEIN, 2001; MENDONCA; CAMPILHO, 2006; FRAZ *et al.*, 2011);
- ▶ As estruturas de diâmetro menor ou igual a 12 pixels devem ser realçadas, porque podem conter vasos de menor calibre.

A escolha adequada da função de ponderação de escala define a qualidade do resultado da segmentação. Outras funções de ponderação podem ser aplicadas como, por exemplo, uma função com decaimento exponencial. Para segmentar estruturas pequenas, adota-se uma função de ponderação cujos valores são mais altos para as escalas menores e decaem para escalas maiores. Portanto, o resultado depende da escolha do critério de seleção de escala das estruturas que devem ser destacadas na imagem. A operação de máximo local é utilizada para regularizar o peso das escalas antes de aplicar uma limiarização.

Observa-se na Figura 6.19 que a regularização obtida com o algoritmo proposto resulta em uma imagem com menos ruído, da qual se destaca significativamente a rede de vasos do fundo mais homogêneo. Isto permite segmentar a rede de vasos por um processo de limiarização global com limiar $\tau = 0,1$. A segmentação obtida por este método está ilustrada na Figura 6.20 (a). O resultado da comparação da rede de vasos segmentada com o PO fornecido pelo primeiro observador da base DRIVE está ilustrado na Figura 6.20 (b). Os vasos não detetados, em vermelho, correspondem àqueles de menor calibre. As regiões em azul correspondem à sobresegmentação de

vasos, ou seja, aquelas regiões que excedem o diâmetro definido no PO. Importante destacar que existe discordância entre os dois observadores da base DRIVE quanto aos vasos finos e que o primeiro observador é mais criterioso, incluindo-os no PO. Contudo, o resultado para esta imagem apresenta boas medidas de desempenho, a saber, $Ac = 0,9808$, $Se = 0,8670$, $Es = 0,9896$.

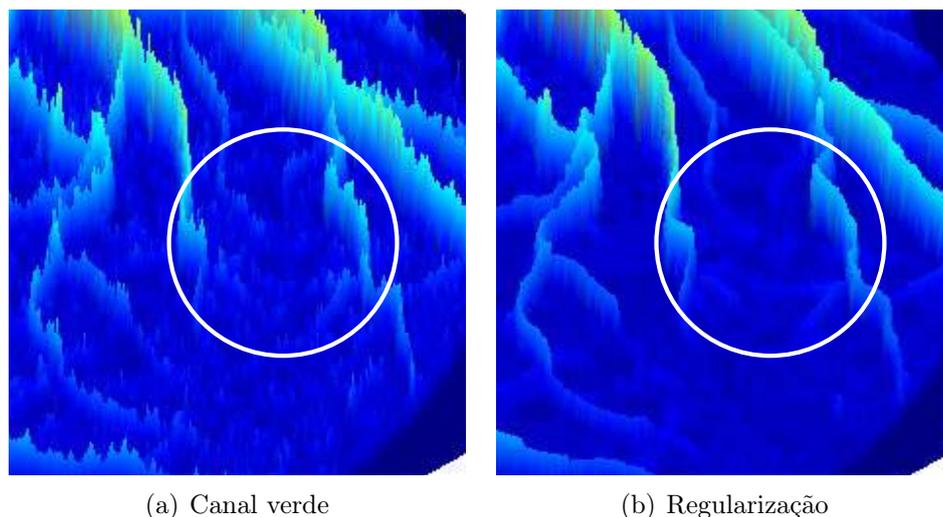


Figura 6.19: Comparação entre retinografias. A região destacada contém vasos de pequeno calibre. (a) Canal verde da retinografia original. (b) Retinografia regularizada com base na TEM.

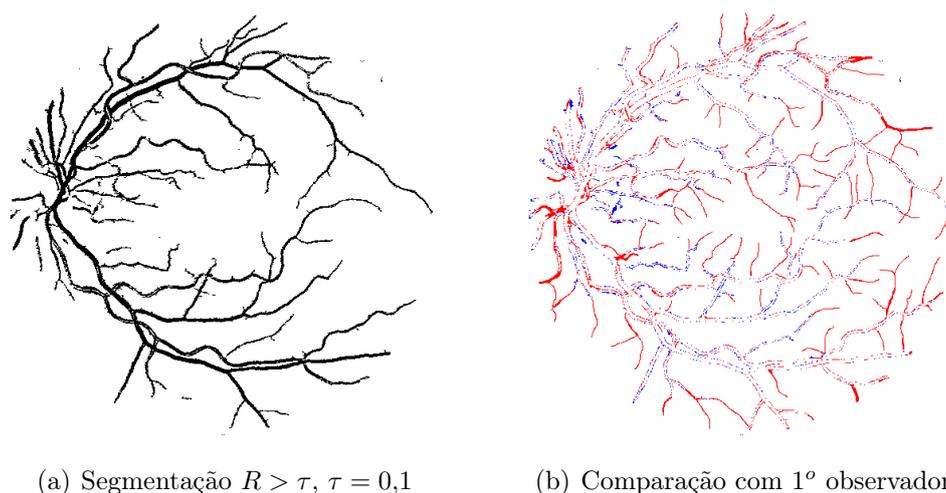


Figura 6.20: (a) Detecção da rede de vasos em retinografia da base DRIVE, sem FOV (sem máscara). (b) Comparação com o 1º observador da base DRIVE. Os pixels em vermelho e em azul representam os erros por ausência e excesso, respectivamente.

Os resultados obtidos para as bases DRIVE e STARE estão resumidos nas Tabelas 6.9 e 6.10, respectivamente. Estes resultados foram comparados aos obtidos pelo algoritmo proposto por Zana e Klein (2001). Este algoritmo foi escolhido por utilizar uma etapa de realce baseada em operações *top hat* da MM. Este realce emprega a superposição de efeitos de sucessivas operações *top hat* utilizando elementos estruturantes direcionais $\theta \in \{\theta_1, \theta_2, \dots, \theta_N\}$ de comprimento $L = 10$. Em uma etapa posterior, é utilizada uma análise de curvatura e, finalmente, a obtenção da rede de vasos segmentada.

Para avaliação dos resultados, foram utilizadas medidas clássicas (acurácia, sensibilidade e especificidade), o índice de confiabilidade κ (Apêndice A) da Tabela A.1 e um método recente de avaliação da segmentação de vasos proposta por Gegúndez-Arias *et al.* (2012).

O método de avaliação proposto por Gegúndez-Arias *et al.* (2012) é o produto CAL das medidas obtidas das equações 3.5, 3.6 e 3.7. Os parâmetros α e β da medida CAL , conforme definido nas equações 3.6 e 3.7, definem a tolerância do método a variações no diâmetro dos vasos em pixels. Neste estudo, foram adotadas duas tolerâncias: $\alpha = \beta = 0$ e $\alpha = \beta = 2$.

O desempenho dos algoritmos foi avaliado no conjunto de testes da base DRIVE e da base STARE. O tempo médio de processamento do algoritmo proposto para cada imagem 512×512 da base DRIVE foi de 1,5s. Para a base STARE, que possui imagens com diferentes resoluções espaciais, o tempo médio por imagem foi 2,1s. Ambos algoritmos foram implementados no MATLAB e executados em um computador de 2 núcleos, Core i5 2,50 GHz, com 8GB de memória RAM.

Os resultados estão listados nas Tabelas 6.9 e 6.10. Observa-se que, para alguns valores médios, não existe diferença estatística significativa (Apêndice A) entre os métodos, admitindo-se o intervalo de confiança de 95% para uma distribuição normal. Nestes casos, o valor p_n aparece indicado, em que n identifica o observador. O maior valor médio para cada medida aparece destacado, por observador.

A análise das medidas de desempenho das Tabelas 6.9 e 6.10 permite concluir que o algoritmo de segmentação baseado na TEM se equipara e até supera o algoritmo proposto por Zana e Klein (2001) para as imagens da base DRIVE e STARE. Se for considerado que o algoritmo baseado na TEM usa um limiar global para a segmentação, os resultados são bastante satisfatórios, embora os valores de acurácia (Ac) e especificidade (Es) (Apêndice A) sejam menores para o observador 2 da base DRIVE e para o observador 1 da base STARE.

Tabela 6.9: Base de retinografias DRIVE

Algoritmo	Obs	Ac	Se	Es	κ ⁽²⁾	CAL	
						$\alpha, \beta = 0$	$\alpha, \beta = 2$
Proposto	1	0,9331 ⁽¹⁾	0,7593	0,9591	0,7055	0,4155	0,6880
	2	0,9381	0,7886	0,9596	0,7226	0,4282	0,6925
Zana e Klein (2001)	1	0,9377 ⁽¹⁾	0,6548	0,9803	0,6961	0,3109	0,5825
	2	0,9435	0,6824	0,9813	0,7190	0,3258	0,5794

⁽¹⁾ $p_1 = 0,63$, ⁽²⁾ $p_1 = 0,38$, $p_2 = 0,71$

Tabela 6.10: Base de retinografias STARE

Algoritmo	Obs	Ac	Se	Es	κ ⁽²⁾	CAL	
						$\alpha, \beta = 0$ ⁽³⁾	$\alpha, \beta = 2$
Proposto	1	0,9234	0,8217	0,9358	0,6460	0,3784	0,5992
	2	0,9098 ⁽¹⁾	0,6872	0,9519	0,6310	0,3491	0,5739 ⁽⁴⁾
Zana e Klein (2001)	1	0,9459	0,7084	0,9733	0,6970	0,4150	0,7173
	2	0,9206 ⁽¹⁾	0,5651	0,9836	0,6278	0,3276	0,6316 ⁽⁴⁾

⁽¹⁾ $p_2 = 0,12$, ⁽²⁾ $p_1 = 0,07$, $p_2 = 0,91$ ⁽³⁾ $p_1 = 0,32$, $p_2 = 0,54$ ⁽⁴⁾ $p_1 = 0,14$

Observa-se pelo valor de κ nas Tabelas 6.9 e 6.10, que a concordância dos dois algoritmos com os observadores é substancial para as duas bases, conforme classificação pela Tabela A.1. Importante salientar que não existe diferença significativa do índice κ entre os algoritmos, considerando um intervalo de confiança de 95%. Em outras palavras, pode-se afirmar que os algoritmos apresentam o mesmo desempenho médio de concordância com os observadores de cada base.

A regularização do método de segmentação proposto também pode ser utilizada como etapa de realce de imagens para a entrada de algoritmos de segmentação de retina (SOARES *et al.*, 2006; ZANA; KLEIN, 2001; MENDONCA; CAMPILHO, 2006; FRAZ *et al.*, 2011).

Vale notar que o objetivo da f_p nesta aplicação é destacar estruturas com escalas no intervalo conhecido do diâmetro dos vasos. Embora a f_p adotada seja linear, a escolha por uma função quadrática, por exemplo $\sqrt{f_p}$, resulta em um ganho significativo de sensibilidade na detecção de vasos. De fato, a sensibilidade aumenta de $Se = 0,7593$ para $Se = 0,7701$ na base DRIVE, sem redução significativa da acurácia. A avaliação pelo índice CAL também apresenta ligeira

melhoria ($CAL_{\alpha,\beta=2} = 0,4185$ e $CAL_{\alpha,\beta=2} = 0,6912$). A mesma conclusão foi obtida para a base STARE com ambos observadores. Importante destacar que a utilização de segmentos de reta direcionais como elementos estruturantes foi abordada por Zana e Klein (2001). Contudo, o resultado obtido com esse tipo de elemento estruturante resulta em uma rede de vasos desconectados. O emprego de um elemento estruturante isotrópico, por outro lado, evita esse problema e ainda permite desenvolver um algoritmo significativamente mais rápido.

6.6 Caso V: avaliação da qualidade da segmentação de vasos de retina

Conforme foi apresentado no Capítulo 3, os métodos clássicos de avaliação supervisionada da segmentação adotam uma comparação orientada a pixel (*pixelwise*) entre a segmentação de referência, ou PO, e a imagem obtida por meio de algum método de segmentação automática.

Para exemplificar o uso combinado da MIDE e da TEM nesse tipo de aplicação, foi realizado um experimento controlado de avaliação da segmentação da rede de vasos de retina, utilizando imagens sintéticas obtidas a partir de degradações incrementais (diferentes níveis de degradação) aplicadas a uma segmentação manual da base DRIVE, tomada como padrão ouro (PO). As degradações compreendem aspectos relevantes do resultado de algoritmos de segmentação de vasos, a saber: a sobresegmentação (SOB), a subsegmentação (SUB), a ausência de vasos finos (AVF), falhas devido ao reflexo da linha central dos vasos (LC) e descontinuidades (DES) da rede de vasos. Exemplos dessas degradações são ilustrados na Figura 6.21. Para fins ilustrativos, a imagem simulada da Figura 6.21 (g) aproxima o resultado de métodos de segmentação automática baseados em morfologia matemática com elemento estruturante direcional (ZANA; KLEIN, 2001). Os detalhes de cada tipo de degradação podem ser observados nos destaques da Figura 6.22.

O algoritmo de avaliação da qualidade da segmentação de vasos da retina baseia-se no uso da MIDE para avaliar as diferenças estruturais entre um PO e uma segmentação automática. A TEM é utilizada para obter uma imagem estrutural em níveis de cinza a partir da rede de vasos binária. A medida *iqs* é determinada pela COR obtida com a MIDE calculada entre a imagem segmentada \mathbf{S} e o padrão ouro \mathbf{P} , ou seja,

$$iqs = COR_{\mathfrak{M}(\mathbf{R}_S, \mathbf{R}_P)}, \quad (6.3)$$

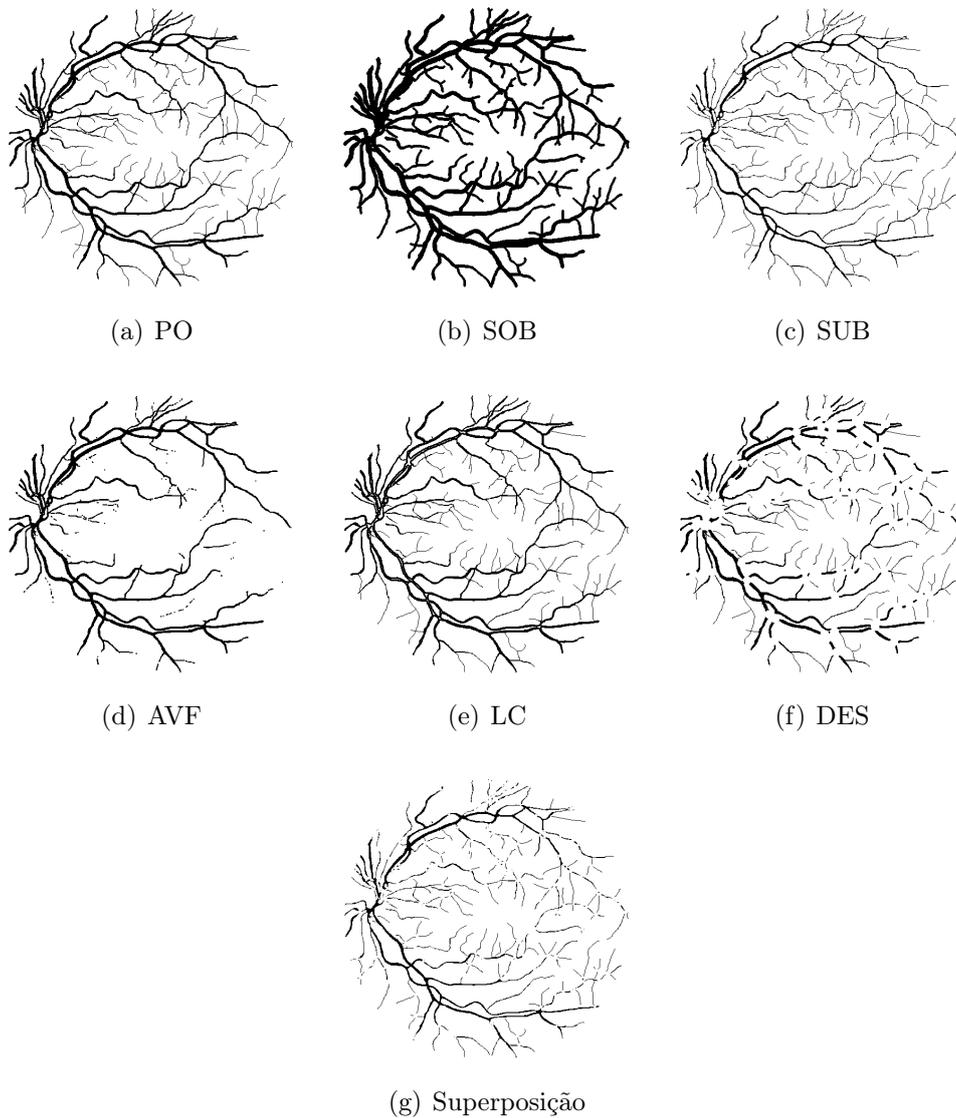


Figura 6.21: Imagens utilizadas na avaliação supervisionada da segmentação de vasos de retina. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Superposição das distorções mais comuns: SUB + DES.

em que $\mathbf{R}_S = \varphi(\mathbf{S})$ e $\mathbf{R}_P = \varphi(\mathbf{P})$ são as imagens estruturais da segmentação automática e da referência (PO), respectivamente, obtidas com a TEM. O modelo ilustrado na Figura 6.23 apresenta as etapas desse algoritmo de avaliação referenciada da segmentação.

O resultado da transformação $\mathbf{R}_P = \varphi(\mathbf{P})$ pode ser observado na Figura 6.24. Os níveis de cinza mais claros correspondem às estruturas de maior calibre, enquanto os níveis de cinza mais escuros correspondem às estruturas mais finas. Os valores mais

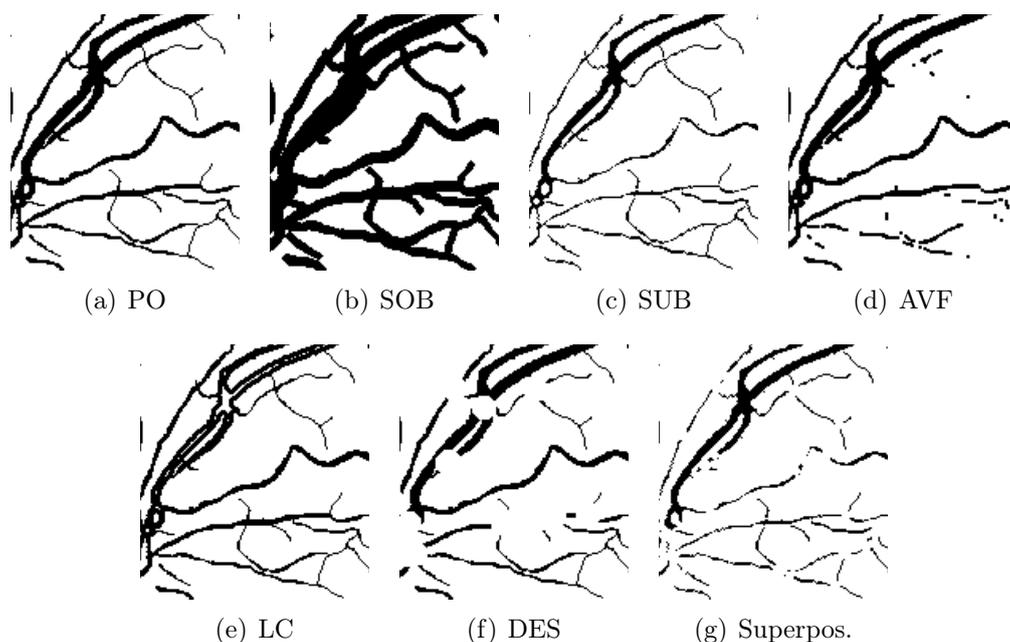


Figura 6.22: Detalhe das imagens utilizadas na avaliação supervisionada da segmentação de vasos de retina. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Espelhamento. (h) Rotação. (i) Superposição das distorções mais comuns: SUB + DES.

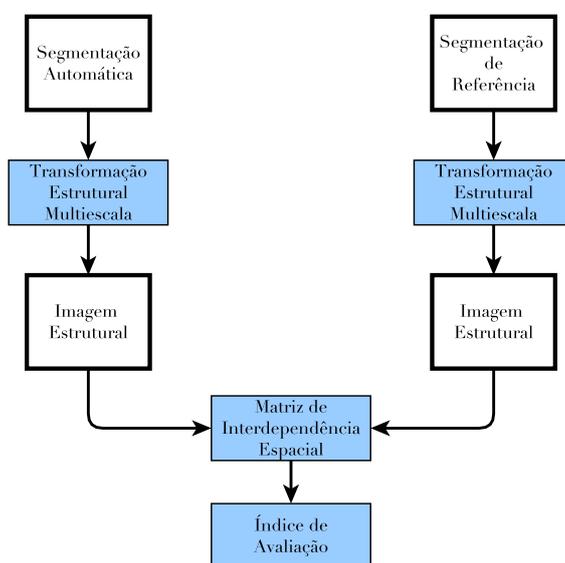


Figura 6.23: Modelo do algoritmo para cálculo do índice de qualidade da segmentação.

altos são atribuídos aos níveis de cinza mais escuros. Essa transformação foi realizada até a convergência. A correspondência entre as estruturas da imagem transformada pode, então, ser avaliada através dos atributos da MIDE.

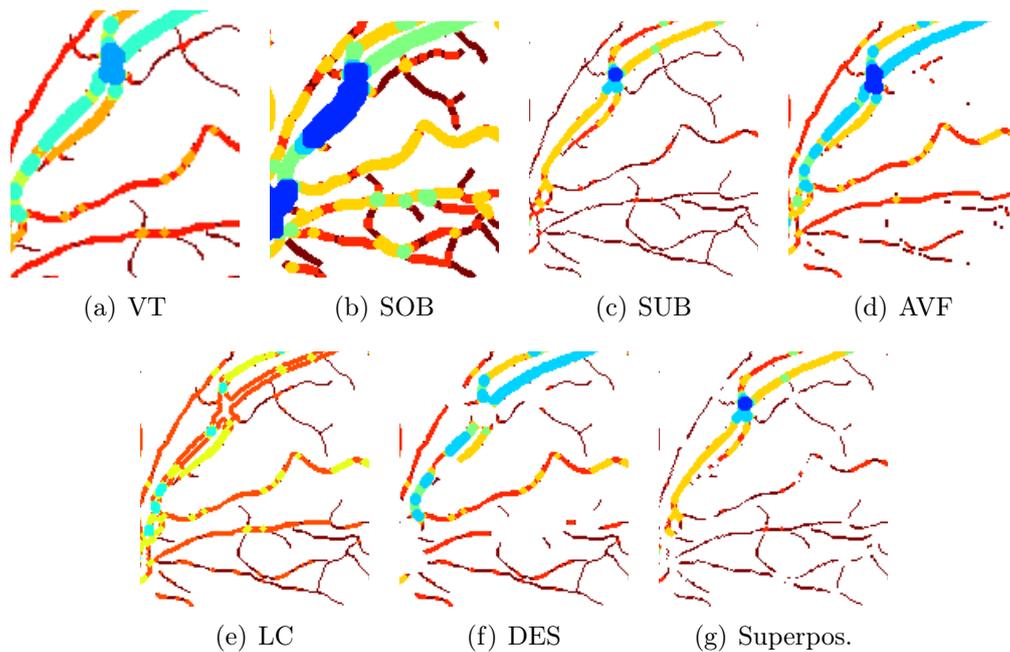


Figura 6.24: Imagens após TEM. (a) Padrão ouro da base DRIVE. (b) Sobresegmentação. (c) Subsegmentação. (d) Ausência de vasos finos. (e) Linha central. (f) Descontinuidades. (g) Espelhamento. (h) Rotação. (i) Superposição das distorções mais comuns: SUB + DES.

O índice de avaliação obtido com a MIDE foi comparado com a avaliação clássica pela medida da acurácia, sensibilidade e especificidade (ZANA; KLEIN, 2001; SOARES *et al.*, 2006; FRAZ *et al.*, 2011) e com o método especializado proposto por Gegúndez-Arias *et al.* (2012). Para a medida CAL do método de avaliação de segmentação de vasos de retina de Gegúndez-Arias *et al.* (2012), foi utilizado o parâmetro de tolerância zero ($\alpha = 0$ e $\beta = 0$).

O resultado é apresentado nos gráficos da Figura 6.25 por meio da distribuição dos índices de avaliação da qualidade para diferentes níveis de degradação, variando de 0 (sem degradação) a 9 (nível máximo de degradação). Para segmentações com nível de degradação 0, os índices retornam valor unitário, o que corresponde a 100% de acerto. Os gráficos registram a mediana (a linha vermelha central que indica o valor que corresponde ao percentil 50) e o retângulo azul delimita os percentis 25 e 75, ou primeiro e terceiro quartis, das medidas iqs , CAL e acurácia Ac . A diferença entre o quartil superior e inferior fornece uma indicação sobre a sensibilidade da medida ao nível de distorção estrutural da imagem segmentada. Quanto maior essa diferença, maior a dispersão do índice de qualidade e, portanto, mais sensível é a medida às falhas estruturais na rede de vasos. Adotando-se como limiar arbitrário

uma amplitude inter-quartil mínima de 0,1, representado pelo retângulo na Figura 6.25, observa-se que a medida *iqs* é bastante sensível às variações de segmentação para os casos SOB, SUB e AVF. Para uma amplitude inter-quartil abaixo de 0,1, considera-se que a medida tem pouca sensibilidade. Embora seja um fator importante da segmentação da rede de vasos de retina para diagnóstico automático de doenças, a menor sensibilidade acontece no caso de descontinuidades dos vasos, conforme pode ser observado na Figura 6.25 (d). Contudo, todas as medidas apresentaram baixa sensibilidade. Isso mostra que, mesmo que a medida especializada *CAL* seja a mais sensível dentre as avaliadas, este tipo de distorção é difícil de ser detetada. Importante notar que, para todos os casos deste experimento, a medida de acurácia *Ac* é a que apresenta menor sensibilidade às alterações estruturais da rede de vasos.

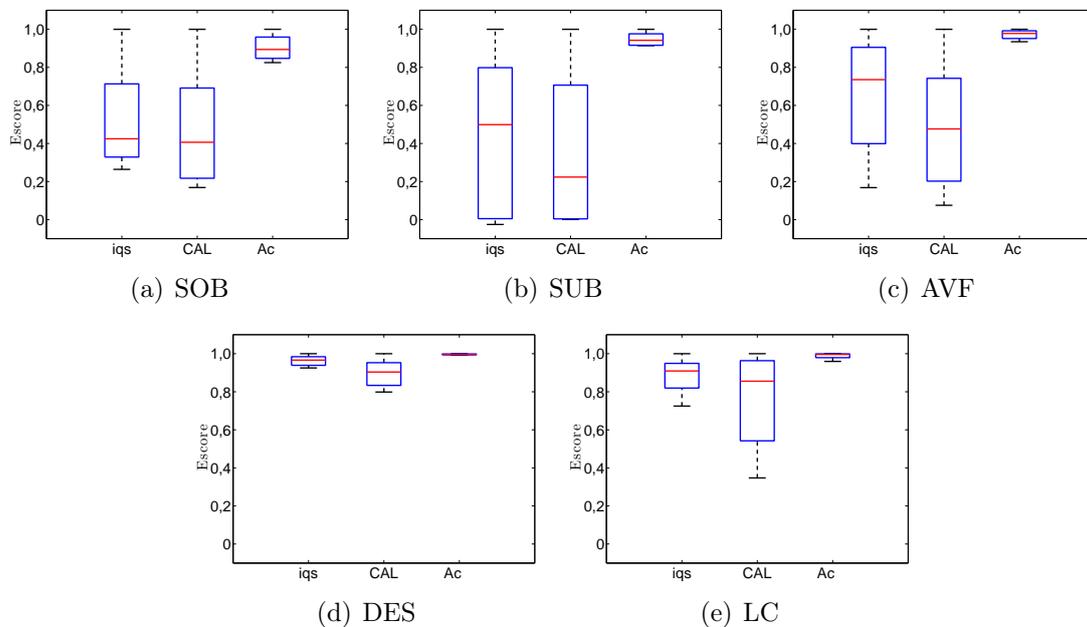


Figura 6.25: Sensibilidade das medidas de avaliação de segmentação de vasos quanto ao tipo de distorção. (a) Sobresegmentação. (b) Subsegmentação. (c) Ausência de vasos finos. (d) Descontinuidades. (e) Linha central.

Pode-se concluir que o índice obtido com a MIDE apresenta uma boa aproximação da qualidade da segmentação da rede de vasos de retina, quando comparada ao método especializado de Gegúndez-Arias *et al.* (2012) e um melhor desempenho quando comparado ao método clássico de avaliação, baseado em medidas de acurácia, sensibilidade e especificidade.

6.7 Conclusão

Neste capítulo foram apresentadas propostas para aplicação prática da Matriz de Interdependência Espacial (MIDE) e da Transformação Estrutural Multiescala (TEM) em problemas de análise estrutural de imagens.

Os estudos de casos apresentados neste capítulo demonstram a aplicabilidade da MIDE e da TEM em análise de imagens, em um contexto estrutural e multiescala. A melhoria de desempenho obtida com os métodos de análise propostos, quando comparados a métodos clássicos, é significativa para uma parte dos casos apresentados, sejam em termos de descritores ou de tempo de processamento.

Dois casos de classificação de imagens utilizando atributos extraídos da MIDE foram propostos. Um deles apresenta uma proposta alternativa ao uso da matriz de coocorrência de níveis de cinza (GLCM, do inglês *gray level cooccurrence matrix*) em uma aplicação de classificação de imagens de corrosão atmosférica. O resultado obtido com os atributos da MIDE foi significativamente melhor do que o método baseado nos atributos de textura da GLCM.

O segundo caso é uma aplicação original da MIDE para a detecção de doenças em imagens de TC dos pulmões. O algoritmo proposto em três diferentes versões empregou descritores da MIDE, da GLCM e a medida Informação de Fidelidade Visual (VIF, do inglês *Visual Information Fidelity*). Os descritores da MIDE mostraram ser capazes de discriminar melhor as doenças, permitindo um bom desempenho de classificação, significativamente melhor do que as outras versões.

Os atributos da MIDE podem ser utilizados como medida de avaliação da qualidade de imagens, sendo capaz de superar o desempenho do MS-SSIM, embora não apresente o mesmo poder de generalização do VIF. Em contrapartida, o método de avaliação baseado na MIDE é significativamente mais rápido do que ambos MS-SSIM e VIF. Este fato o torna uma escolha mais adequada, para aplicações de tempo real, do que os outros dois métodos.

O uso da TEM, foi ilustrado em um experimento de segmentação de vasos de retina. O método de realce da retinografia baseado na TEM mostrou ser eficiente em destacar os vasos, facilitando a sua segmentação. Um algoritmo simples de limiarização global da imagem realçada foi comparado com um método específico da literatura. Os resultados mostram que o método proposto é capaz de obter resultados similares, porém com um custo computacional menor.

Também foi apresentada uma aplicação de uso combinado da TEM e da MIDE

em um experimento de avaliação de algoritmos de segmentação de vasos de retina. O método de avaliação proposto foi comparado com um método especializado. Os resultados obtidos para imagens simuladas mostraram similaridades entre os dois métodos, entretanto com um significativo ganho de desempenho com relação ao tempo de processamento.

Conclusões

As principais contribuições desta pesquisa são apresentadas na forma de um método de realce multiescala e de um método de análise de imagens, ambos aplicáveis tanto em imagens binárias quanto em níveis de cinza.

O método de análise e extração de características proposto, a Matriz de Interdependência Espacial (MIDE), emprega a estatística de coocorrência para extrair características estruturais das imagens. Este método também pode ser empregado na extração de medidas visando a avaliação da qualidade de imagens. A análise por meio da MIDE oferece uma alternativa à avaliação da qualidade de imagens com um menor número de parâmetros de ajustes e também com um custo computacional menor do que o apresentado pelos métodos de comparação estudados.

Também é possível realizar a análise estrutural a partir da extração direta ou indireta de características presentes na imagem. Algumas dessas características, ao contrário da forma e escala de objetos, podem não ser facilmente mensuráveis como a textura e as distorções estruturais causadas por degradação. Entretanto, os atributos da MIDE podem descrever imagens ou regiões estruturalmente degradadas.

O método de transformação proposto, a Transformação Estrutural Multiescala (TEM), baseia-se no arcabouço da morfologia matemática e apresenta uma solução para problemas de realce multiescala com preservação de topologia, bordas e forma dos objetos. Além disso, esta transformação permite criar uma representação estrutural de imagens binárias, transformando-as em imagens em níveis de cinza, porém preservando suas características. As imagens transformadas pela TEM possibilitam o uso da MIDE sem adaptações. A transformação proposta realça saliências visuais, preservando características da imagem como o contorno de objetos. Com o ajuste de uma função de ponderação de escala, a TEM permite regularizar

imagens destacando as estruturas de interesse e suavizando as estruturas indesejadas, como o ruído.

Foram apresentados experimentos utilizando tanto imagens sintéticas quanto imagens utilizadas em diversos problemas práticos de análise estrutural. Os estudos de casos mostraram que os métodos de análise propostos permitem aproximar a capacidade do sistema visual humano (HVS, do inglês *human visual system*) de extrair informação estrutural relevante das cenas. Isto envolve o realce de estruturas pequenas e delgadas, além da detecção da degradação dessas estruturas.

Os resultados apresentados neste trabalho permitem concluir que a metodologia proposta tem aplicabilidade na solução de uma variedade de problemas de visão computacional.

7.1 Produção científica e tecnológica

Ao longo da pesquisa, foram produzidos trabalhos que registram o arcabouço teórico e tecnológico que deu origem a esta tese, a saber:

- ▶ *On the Evaluation of Texture and Color Features for Nondestructive Corrosion Detection. EURASIP Journal of Advances in Signal Processing* (MEDEIROS *et al.*, 2010).
- ▶ *Retinal Vessel Enhancement and Segmentation using Mathematical Morphology* – Artigo submetido ao periódico *Pattern Recognition Letters*, out/2013.
- ▶ *Lung Disease Detection Within Active Contour 2D Crisp Images* – Artigo submetido ao *Brazilian Journal of Biomedical Engineering*, nov/2013.
- ▶ Aparato e Método para Obtenção e Interpretação de Imagens Digitais de Superfície e Método de Monitoramento de Efeitos da Corrosão em Superfícies – Patente depositada, processo n° BR 10 2013 004141 6, fev/2013.

7.2 Perspectivas de trabalhos futuros

No decorrer da pesquisa foi possível observar a aplicabilidade dos métodos desenvolvidos na solução de diversos problemas práticos. Como perspectiva de desenvolvimento da metodologia, pretende-se desenvolver algoritmos de visão computacional para investigar a sua aplicabilidade em:

- ▶ avaliação de sinais unidimensionais para detecção de falhas em motores (ver Apêndice B);
- ▶ avaliação não referenciada da qualidade de imagens a partir da classificação de padrões utilizando descritores estruturais;
- ▶ avaliação não referenciada da segmentação baseada na análise estrutural da imagem original em níveis de cinza sem a necessidade de um padrão ouro;
- ▶ avaliação da qualidade da compressão de mamografias (ver Apêndice B);
- ▶ avaliação referenciada da qualidade de filtros;
- ▶ detecção de mudanças em imagens (RADKE *et al.*, 2005) para fins forenses ou biomédicos;
- ▶ recuperação de imagens baseada em contexto, por meio da detecção de saliências (ver Apêndice B) e de uma função de similaridade estrutural;
- ▶ análise de imagens em cores, avaliando as estruturas presentes em cada canal de um modelo de cores através da incorporação de uma relação de vizinhança entre essas estruturas;
- ▶ seleção ou fusão de bandas em imagens em cores ou multiespectrais;
- ▶ processamento de vídeo com o objetivo de restaurar quadros corrompidos.

Ferramentas de Avaliação

Neste apêndice são apresentadas ferramentas de reconhecimento de padrões e estatísticas utilizadas na realização dos experimentos e avaliação dos resultados, respectivamente.

Algoritmos de reconhecimento de padrões

O Reconhecimento de Padrões (RP) envolve diferentes classes de algoritmos, a saber, os classificadores paramétricos e não paramétricos, lineares e não lineares, supervisionados e não supervisionados (HAYKIN, 2001; WEBB, 2002; KOHONEN, 2001; DUDA; HART; STORK, 2000).

As Redes Neurais Artificiais (RNAs) do tipo Perceptron Multi-Camadas (MLP, do inglês *Multi-Layer Perceptron*) (HAYKIN, 2001) e o Mapa de Kohonen ou Mapa Auto-Organizável (SOM, do inglês *Self-Organizing Map*) (KOHONEN, 2001) são exemplos de algoritmos comumente utilizados em RP para classificação supervisionada e agrupamento de dados por aprendizagem competitiva, respectivamente. O discriminante de Fisher (FISHER, 1936) é um exemplo de classificador paramétrico linear ou não linear, enquanto o algoritmo k-Vizinhos mais Próximos (KNN, do inglês *k-Nearest Neighbor*) é um exemplo de algoritmo não paramétrico (DUDA; HART; STORK, 2000; WEBB, 2002). Mais recentemente tem sido adotada a Máquina de Aprendizado Extremo (ELM, do inglês *Extreme Learning Machine*) (HUANG; ZHU; SIEW, 2004; HUANG; ZHU; SIEW, 2006; LIU; LOH; TOR, 2005), uma Rede Neural Artificial (RNA) extremamente rápida e mais efetiva do que a MLP se alguns cuidados forem tomados (WANG; CAO; YUAN, 2011) como, por exemplo, a exigência de um maior número de neurônios. Nesta tese, estes algoritmos são utilizados como ferramentas de teste e validação em etapas das

aplicações da metodologia proposta.

Na Figura A.1 estão ilustrados modelos de uma rede MLP, uma rede ELM e de uma rede SOM. Dado um conjunto de vetores de características $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, o objetivo do algoritmo ELM é adaptar uma rede de pesos sinápticos w para aproximar uma função $f(x)$ dada a informação de saída desejada d . A forma como os pesos são atualizados, entretanto, é diferente para os dois algoritmos, sendo o segundo mais rápido. No caso da rede SOM, os neurônios estão no mesmo espaço dos dados de entrada. Depois de treinados, os neurônios da rede SOM são protótipos que representam esses dados mantendo a relação de vizinhança que existia antes do treinamento.

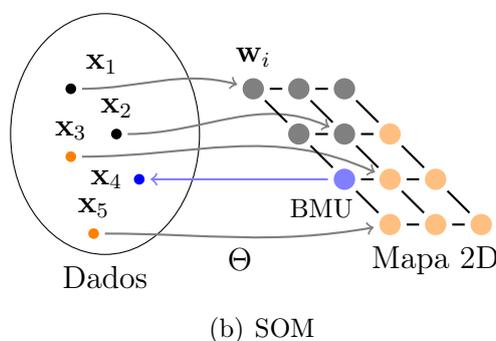
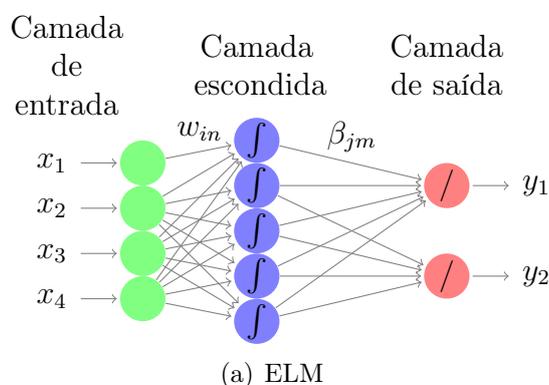


Figura A.1: (a) Modelo de uma rede ELM com 3 neurônios na camada escondida e 2 neurônios na camada de saída. As setas representam os pesos sinápticos da camada escondida w_{in} e da camada de saída β_{jm} . (b) Modelo de uma rede SOM com 12 neurônios em um mapa 4×3 . As linhas representam a relação de vizinhança entre os neurônios.

As redes ELM e MLP possuem arquitetura e regra de aprendizagem similares. A principal diferença entre ambas reside no algoritmo de atualização dos pesos que na ELM é realizada em um único passo, da entrada para a saída. Por isso a sua característica principal é a rapidez de treinamento. Em geral, a ELM também alcança desempenho superior em classificação sobre a MLP, embora necessite de

mais neurônios na sua camada escondida (HUANG; ZHU; SIEW, 2004; HUANG; ZHU; SIEW, 2006; WANG; CAO; YUAN, 2011) para um bom desempenho.

A regra geral de atualização dos pesos \mathbf{w} dos neurônios de uma ELM é definida por

$$\sum_{i=1}^N \beta_i g(\mathbf{w}_i x_j + b_i) = d_j; j = 1, \dots, N, \quad (\text{A.1})$$

em que N é a quantidade de neurônios da camada escondida. $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^\top$ são os pesos da camada escondida e $\beta = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^\top$ são os pesos da camada de saída, cujo valor desejado é d_j . $g(\cdot)$ é uma função de ativação, em geral, do tipo logística. Essa regra de aprendizagem pode ser escrita de forma compacta como $\mathbf{H}\beta = \mathbf{T}$. A solução pode ser determinada por mínimos quadrados

$$\beta = \mathbf{H}^\dagger \mathbf{T}, \quad (\text{A.2})$$

em que \mathbf{H}^\dagger é a pseudo inversa generalizada de Moore-Penrose (HUANG; ZHU; SIEW, 2004). Importante notar que a rede independe dos pesos da camada escondida, os quais podem ser inicializados com valores aleatórios.

A rede SOM é um algoritmo não supervisionado que possui características de preservação da topologia e redução de dimensionalidade. Além disso, o algoritmo SOM é flexível o bastante para ser transformado em um classificador de padrões, quantização vetorial e aprendizagem ativa (KOHONEN, 2001). A forma mais comum de garantir a convergência de uma rede SOM é decrementar as taxas de aprendizagem e a taxa de atualização dos vizinhos ao longo do tempo. Quando essa atualização é reduzida somente ao vencedor, a SOM se torna uma quantização vetorial. Em uma rede SOM bidimensional cada protótipo n -dimensional $\mathbf{w}_i = \{w_1, w_2, \dots, w_n\}$, $i = 1, 2, \dots, m$, é associado a um nó da malha. A malha de visualização tem correspondência um-para-um com os índices associados com os vetores dos protótipos, ou seja, possui m nós. Na rede SOM, é encontrado o vetor vencedor que melhor representa a entrada.

A expressão geral que define a adaptação dos pesos da rede SOM, ou seja do vetor de protótipos $\mathbf{w}_i = \{w_1, w_2, \dots, w_n\}$, é determinada por

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \eta(t) h_{ci}(t) [\mathbf{x}(t) - \mathbf{w}_i(t)], \quad (\text{A.3})$$

em que $t = 0, 1, 2, \dots$ é um inteiro que identifica a época atual de treinamento no

tempo discreto; $h_{cl}(t) = \exp(-\frac{\|r_c(t)-r_l(t)\|^2}{2\sigma^2(t)})$ é uma função de vizinhança comumente utilizada, que define quais os pesos serão atualizados; $\sigma(t)$ é um raio que define a abrangência da vizinhança. Os vetores $r_c(t)$ e $r_l(t)$ correspondem às posições dos nós da malha. A taxa de aprendizagem $0 < \eta(t) < 1$ deve decrescer gradualmente para garantir a convergência. O operador $\|\cdot\|$ representa a norma. Para acelerar a convergência, faz-se $h_{cl}(t) \rightarrow 0$ quando $t \rightarrow \infty$ e os vetores \mathbf{w}_i são inicializados com valores aleatórios antes do treinamento.

Uma rede SOM treinada também pode ser utilizada como classificador. Neste caso, um vetor desconhecido recebe a classe da Unidade Melhor Correspondente (BMU, do inglês *Best-Matching Unit*). Em outras palavras, o vetor desconhecido receberá o rótulo do protótipo de melhor correspondência segundo algum critério. Caso o critério adotado seja a distância euclidiana, por exemplo, o vetor recebe o rótulo do protótipo mais próximo.

Mapa estrutural de atributos

Na análise de imagens, é frequente o uso de extratores de atributos utilizados em tarefas de análise, classificação e reconhecimento de padrões (CHOI; KIM, 2005). Em casos de reconhecimento de padrões, por exemplo, são utilizados vetores de atributos de dimensões elevadas, o que dificulta a sua visualização gráfica e, portanto, a compreensão do que esses atributos representam como discriminantes de características da imagem.

Uma rede SOM com grande número de neurônios torna aparente as características estruturais intrínsecas do espaço de dados. Isto pode ser expresso por um critério de vizinhança dos dados armazenados nesses neurônios. Fazendo modificações nos algoritmos de aprendizagem da SOM, Ultsch e Siemon (1990) desenvolveram um método de projeção bidimensional de dados de grande dimensão, conhecido como matriz de distância unificada (em inglês, *unified distance matrix*) ou simplesmente matriz-U.

Uma matriz-U exhibe a densidade local de estruturas em uma topologia que preserva a projeção do espaço de alta dimensão em uma mapa bidimensional. Considere o conjunto de dados sintéticos de dimensão 3, ilustrado na Figura A.2 (a). Observa-se que existem duas classes bem definidas, a saber, classe A em cor laranja e classe B em cor azul. O algoritmo de aprendizagem da SOM preserva a relação de vizinhança do espaço original dos dados no mapa resultante (KOHONEN, 2001), observado na Matriz-U da Figura A.2 (b), que exhibe em tons escuros uma

representação de vizinhança intraclasse e em tons claros a relação de vizinhança interclasse. Cada lugar do mapa é um neurônio cujo valor representa a distância de si mesmo para os seus vizinhos de mesma classe. Os tons variam do vermelho, para distâncias pequenas, até o amarelo, para distâncias maiores. No mapa da Figura A.2 (c) pode-se observar a existência de uma fronteira bem definida separando as classes. Os elementos mais próximos da fronteira correspondem aos tons mais claros da Figura A.2 (b). Quanto mais próximo da fronteira, maior é o valor da distância para o elemento mais próximo de sua classe.

Para a geração da Matriz-U é utilizada a biblioteca **SOM Toolbox**¹ para **MATLAB**².

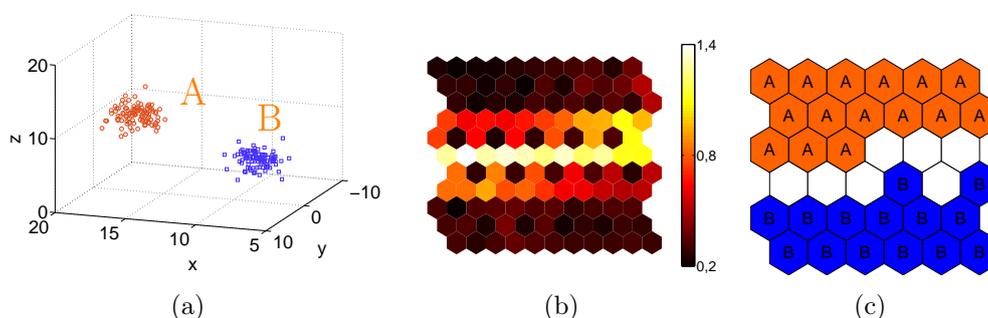


Figura A.2: (a) Gráfico de espalhamento dos dados simulados com 3 dimensões. (b) Mapa bidimensional (Matriz-U). (c) Mapa rotulado.

A análise individual de cada variável por meio da sua própria matriz torna possível interpretar o poder discriminatório dessa variável, como ilustram os mapas da Figura A.3. A diferença de tonalidade demonstra que as variáveis x e y possuem um poder discriminatório maior do que a variável z , embora todas desempenhem um papel importante na separabilidade das classes A e B. De fato, a ausência de uma dessas variáveis inviabiliza a correta separação das classes.

¹<http://www.cis.hut.fi/projects/somtoolbox/>

²<http://www.mathworks.com>

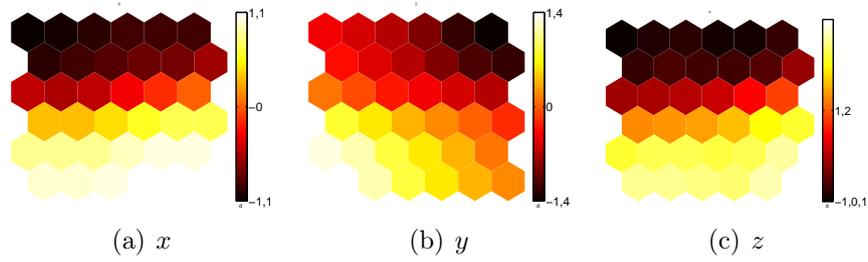


Figura A.3: Contribuição das variáveis x , y e z para o mapa da Figura A.2 (b).

Medidas de avaliação de desempenho de algoritmos

Para avaliação dos resultados dos métodos propostos, foram utilizadas medidas clássicas, a saber, acurácia, sensibilidade e especificidade, além do índice de confiabilidade κ e um método recente de avaliação da segmentação de vasos proposta por Gegúndez-Arias *et al.* (2012).

Matriz de confusão

A matriz de confusão (MC) é formada a partir do registro das frequências de concordâncias (ou coocorrências) entre duas ou mais classes de dados. Trata-se de um recurso bastante utilizado para comparar a saída de algoritmos de classificação com uma referência conhecida. Para n classes, a MC é representada por

$$\text{MC} = \begin{array}{c} \text{Esperado} \\ \left\{ \begin{array}{c} \overbrace{\left(\begin{array}{cccc} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & & c_{2n} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{array} \right)}^{\text{Predição}} \end{array} \right. , \end{array} \quad (\text{A.4})$$

em que as linhas representam o valor esperado e as colunas o valor obtido. Um sistema de classificação perfeito retorna uma MC diagonal, ou seja, cujos elementos $c_{ij} = 0, \forall i \neq j$.

Quando um sistema de classificação binária é avaliado, $n = 2$. Neste caso particular,

$$MC = \begin{pmatrix} VP & FP \\ FN & VN \end{pmatrix}, \quad (A.5)$$

em que VP e VN representam a quantidade de elementos classificados corretamente nas classes 1 e 2, respectivamente; FP representa a quantidade de elementos da classe 1 classificados como classe 2; FN representa a quantidade de elementos da classe 2 classificados como classe 1.

As seguintes medidas são derivadas da matriz de confusão:

► Acurácia

Mede a taxa total de acerto na classificação de elementos de todas as classes.

$$Ac = \frac{\#\{\text{elementos classificados corretamente}\}}{\#\{\text{total de elementos}\}} = \frac{\sum_{ij} \{c_{ij} | i = j\}}{\sum_{ij} c_{ij}} \quad (A.6)$$

► Sensitividade (TVP)

Mede o percentual de acerto para uma classe i , ou taxa de verdadeiros positivos (TVP).

$$Se_i = \frac{\#\{\text{elementos classificados corretamente na classe } i\}}{\#\{\text{total de elementos da classe } i\}} \quad (A.7)$$

► Especificidade (TVN)

Mede o percentual de acerto por não pertencer à classe i , ou taxa de verdadeiros negativos (TVN).

$$Es_i = \frac{\#\{\text{elementos classificados corretamente na classe complementar}\}}{\#\{\text{total de elementos estimados da classe complementar}\}} \quad (A.8)$$

► Área sob a curva ROC

A curva característica de operação do receptor (ROC, do inglês *receiver operating characteristic*) é uma forma de representação gráfica do desempenho

de um sistema de predição binário, ou seja, no qual apenas duas classes de dados são observadas. A curva $Se \times 1 - Es$ é construída a partir dos valores de sensibilidade (Se) e especificidade (Es) obtidos da matriz de confusão para diferentes limiares de separação entre as classes. Essa curva é uma representação sumária dos resultados do classificador e que permite uma interpretação visual do seu desempenho. Um bom desempenho do classificador corresponde a curvas afastadas da diagonal.

A área sob a curva ROC (AUC, do inglês *area under curve*) é uma medida utilizada para comparar numericamente o desempenho de dois sistemas de classificação. Quanto mais a curva se afasta da diagonal, maior é a área sob a curva e melhor é o desempenho do classificador. O valores da AUC estão no intervalo $[-1,1]$, sendo o valor máximo aquele que caracteriza um classificador ideal. O valor 0,5 corresponde ao pior caso, quando um classificador apenas aproxima a chance aleatória de 50%. Valores negativos caracterizam uma predição invertida com relação à classificação esperada.

Quando o número de amostras é pequeno, a curva ROC assume o formato de degraus e a comparação de modelos por meio da AUC pode ser imprecisa. Nestes casos, emprega-se o uso de uma *convex hull*, curva que liga os pontos extremos dos degraus, suavizando a curva ROC. Assim, em geral, adota-se a AUC_H como medida de comparação mais sensível.

► Precisão e Revocação

Precisão Pre e revocação Rev são medidas do desempenho de um sistema de busca de informação. A medida precisão representa a proporção das instâncias recuperadas de uma classe i que são relevantes, ou seja, a taxa de acerto dentro de uma classe. A medida de revocação representa a proporção das instâncias relevantes de uma classe i que de fato são recuperadas. A revocação tem um significado semelhante ao da sensibilidade. Os seus respectivos valores são determinados pelas expressões extraídas da matriz de confusão, segundo as expressões

$$Pre_i = \frac{\#\{\text{instâncias relevantes} \cap \text{instâncias recuperadas}\}}{\#\{\text{instâncias recuperadas}\}} = \frac{c_{ii}}{\sum_j c_{ij}}, e \quad (\text{A.9})$$

$$Rev_i = \frac{\#\{\text{instâncias relevantes} \cap \text{instâncias recuperadas}\}}{\#\{\text{instâncias relevantes}\}} = \frac{c_{ii}}{\sum_j c_{ji}}, \quad (\text{A.10})$$

em que, $\sum_j c_{ij}$ é o somatório das colunas e $\sum_j c_{ji}$ é o somatório das linhas da matriz de confusão, para uma classe i .

Para medir o desempenho de sistemas de recuperação de informação adota-se a medida F . Essa medida combina a precisão e revocação por meio de uma média harmônica:

$$F = (1 + \beta^2) \cdot \frac{Pre \cdot Rec}{\beta^2 Pre + Rec}, \quad (\text{A.11})$$

em que β é uma valor da importância da precisão sobre a revocação. Quando a importância atribuída a ambos é a mesma, usa-se $\beta = 1$ e a medida é conhecida como escore F_1 .

► Índice Kappa

O índice Kappa κ é uma medida estatística de concordância entre dois observadores, definida como

$$\kappa = \frac{\text{percentual de acertos} - \text{percentual de acerto esperado}}{1 - \text{percentual de acerto esperado}} = \frac{Ac - P_e}{P_e}, \quad (\text{A.12})$$

em que $P_e = \sum_{i=1}^n \frac{\sum_{i=1}^n MC_{ij} \cdot \sum_{j=1}^n MC_{ij}}{(\sum_{i,j=1}^n MC_{ij})^2}$ é a probabilidade de concordância hipotética. Em geral, o índice kappa é utilizado para avaliar qualitativamente o

resultado de dois observadores usando as faixas da Tabela A.1 como indicativo do grau de concordância.

Tabela A.1: Índice κ

κ	Concordância
<0	nenhuma
0,00 a 0,19	pobre
0,20 a 0,39	justa
0,40 a 0,59	moderada
0,60 a 0,79	substancial
0,80 a 1,00	quase perfeita

► Coeficiente de correlação de Matthews

O coeficiente de correlação de Matthews (Mcc) (MATTHEWS, 1975) é uma medida utilizada para avaliar o resultado de algoritmos de classificação binária a partir de uma $MC_{2 \times 2}$. São levados em conta os verdadeiros e falsos positivos e negativos, sendo uma medida equilibrada que pode ser empregada mesmo se as classes possuem tamanhos muito diferentes. O coeficiente também pode ser utilizado para avaliar o resultado de algoritmos de segmentação binária, sendo determinado pela expressão

$$Mcc = \frac{(VP \times VN) - (FP \times FN)}{\sqrt{(VP + FP) \times (VP + FN) \times (VN + FP) \times (VN + FN)}}, \quad (A.13)$$

em que $Mcc = 1$ representa uma predição perfeita, $Mcc = 0$ representa uma predição aleatória e $Mcc < 0$ representa uma predição inversa.

Teste de hipótese (ANOVA)

O teste ANOVA é realizado para comparar duas variáveis aleatórias e determinar se elas pertencem ou não a um mesmo conjunto amostral. Também é utilizado para comparar resultados de algoritmos diferentes, a fim de determinar se existem diferenças significativas que permitam concluir se esses resultados são similares ou não.

Nesta tese, a hipótese nula (H_0) é aquela em que os conjuntos de dados estudados pertencem a uma mesma população. Assim, a hipótese nula é aquela em que não existe evidência estatística, $p < 0,05$ para um intervalo de confiança de 95%, de que os conjuntos estudados pertençam a classes diferentes de dados.

Boxplot

O *boxplot*, ou gráfico de caixas, é uma representação visual que resume diversas informações sobre um conjunto de dados ordenados. Na Figura A.4 encontram-se representados numa caixa o primeiro quartil Q1, a mediana M, ou segundo quartil Q2 e o terceiro quartil Q3. Esses três elementos representam, respectivamente, os valores que dividem o espaço amostral em 25%, 50% e 75%. O elemento IQR (*interquartile range*) mede a diferença entre o terceiro e primeiro quartis, que contém 50% dos dados. O valor médio do espaço amostral é representado por \bar{x} e os valores máximo e mínimo são representados pelas linhas nas extremidades da caixa. Os elementos ‘o’ e ‘*’ que aparecem além desses limites, representam os valores atípicos, ou *outliers*. O gráfico pode ser apresentado na horizontal, como ilustrado na Figura A.4, ou na vertical.

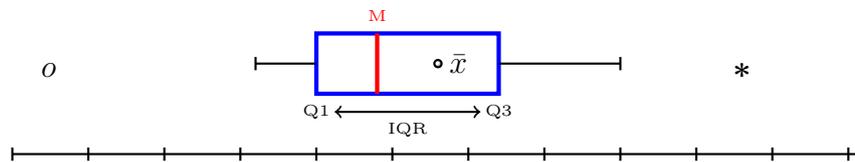


Figura A.4: Elementos de um *boxplot*.

Medidas de correlação

A correlação é uma medida da dependência entre os resultados de dois sistemas para o mesmo conjunto de entrada.

► Correlação linear paramétrica

Supõe que a relação entre as variáveis é linear e requer que as variáveis estejam em um mesmo intervalo.

Pearson: coeficiente de correlação linear, assume que as variáveis estão distribuídas segundo uma função normal e são homoscedásticas (dados distribuídos normalmente ao longo da linha de regressão).

O coeficiente de correlação linear de Pearson entre duas variáveis x e y , é definido por

$$r = \frac{\sum_{n=0}^N (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\left[\sum_{n=0}^N (x_n - \bar{x})^2\right] \left[\sum_{n=0}^N (y_n - \bar{y})^2\right]}}, \quad (\text{A.14})$$

em que N é a dimensão de x e de y ; \bar{x} e \bar{y} representam o valor médio de x e de y , respectivamente.

Para aplicar a correlação de Pearson em uma relação não linear entre variáveis, primeiro deve ser obtida uma curva de regressão não linear para mapear as variáveis para um espaço em que a relação entre ambas seja linear.

► Correlação não paramétrica

Não é necessário conhecer a função densidade de probabilidade (PDF, do inglês *probability density function*) das variáveis medidas. Por empregar uma ordenação, as variáveis podem representar valores em faixas diferentes.

Spearman: é a correlação de Pearson medida sobre as variáveis ordenadas.

$$\rho = 1 - \frac{6 \sum d_i^2}{n^3 - n}, \quad (\text{A.15})$$

em que $d_i = x_i - y_i$ é a diferença entre cada posto de valor correspondentes das variáveis x e y . n é o número de amostras dos pares dos valores.

Kendall: mede o grau de concordância entre as variáveis ordenadas de duas amostras.

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}, \quad (\text{A.16})$$

em que n é o número total de elementos, n_c é o número de pares concordantes e n_d é o número de pares discordantes.

Apêndice **B**

Aplicações Suplementares

Neste apêndice são apresentados exemplos suplementares da aplicação da Matriz de Interdependência Espacial (MIDE) e da Transformação Estrutural Multiescala (TEM).

Detecção de falhas elétricas em motores

Detetar sinais de falhas de curto circuito entre bobinas nos motores elétricos trifásicos previne a inutilização do equipamento e, portanto, reduz significativamente os custos de operação e manutenção.

Base de dados de falhas elétricas

A base de dados utilizada neste experimento consiste de sinais de corrente e vibração de um equipamento submetido a falhas de curto circuito entre espiras. Um motor trifásico de 0,5 cv foi submetido a diferentes níveis de falhas. As bobinas do motor foram submetidas a curto circuito em média impedância, em diferentes níveis de intensidade. O sinal das três correntes foi medido utilizando uma taxa de amostragem de 10kHz. Ao todo foram coletadas 12 amostras de 10 segundos de duração para cada condição, com ou sem falha. Para avaliar a resposta do teste em diferentes condições de operação, o sistema foi submetido a 3 níveis de carga, a saber, 0%, 50% e 100%, considerando 100% a carga máxima para o limite da corrente nominal do motor.

Classificação da falha

Esta é uma aplicação da MIDE na análise de um sinal unidimensional. Os atributos foram extraídos a partir de duas das três correntes, assumindo que as correntes estão balanceadas na condição sem falha. O vetor de atributos $\mathbf{A} =$

{COR, MDI, 1 – CHI} foi extraído dos sinais de corrente utilizando uma MIDE com 100 níveis de intensidade.

Utilizou-se uma rede Perceptron Multi-Camadas (MLP, do inglês *Multi-Layer Perceptron*) para classificação dos dados. A taxa média de acerto para 100 testes foi de aproximadamente 98% com desvio de 0,02. Foram utilizados 3 descritores, extraídos diretamente do sinal de corrente (baixa impedância) e classificados com uma MLP com 30 neurônios na camada escondida. Foram consideradas 4 classes equiprováveis, a saber: sem falha (A), curto circuito na bobina 1 (B), curto circuito na bobina 1 e 2 (C) e curto circuito nas bobinas 1, 2 e 3 (D). A matriz de confusão média dos 100 testes é exibida na Tabela B.1.

Tabela B.1: Matriz de confusão média de 100 testes de classificação

Classe	Predição			
	A	B	C	D
A	0,2492	0,0073	0,0026	0,0021
B	0,0008	0,2426	0,0049	0,0017
C	0,0000	0,0001	0,2423	0,0013
D	0,0000	0,0000	0,0002	0,2449

Acurácia=0,9791

Os resultados para dados de corrente em alta impedância ou dados de corrente alta+baixa juntas também foram avaliados, contudo os resultados ficaram próximos a 90% de acerto médio.

O espaço de descritores exibido na Figura B.1 ilustra o poder de discriminação das falhas. A existência de limites tênues também ilustra a capacidade de separação em subclasses por intensidade de carga. Isto é mais evidente para a falha C, no canto inferior esquerdo dos mapas. Contudo, também é possível observar subclasses de cada falha por intensidade de carga.

O espaço dos descritores ilustrado no gráfico da Figura B.2 exhibe exagramas em azul que identificam as amostras na condição sem falha (A). Os demais marcadores representam as amostras na condição de falha B, C e D, respectivamente representados por marcadores na forma de quadrado, triângulo e estrela. A intensidade de carga está representada pelo tamanho do marcador (mais escuro e menor significa menor carga).

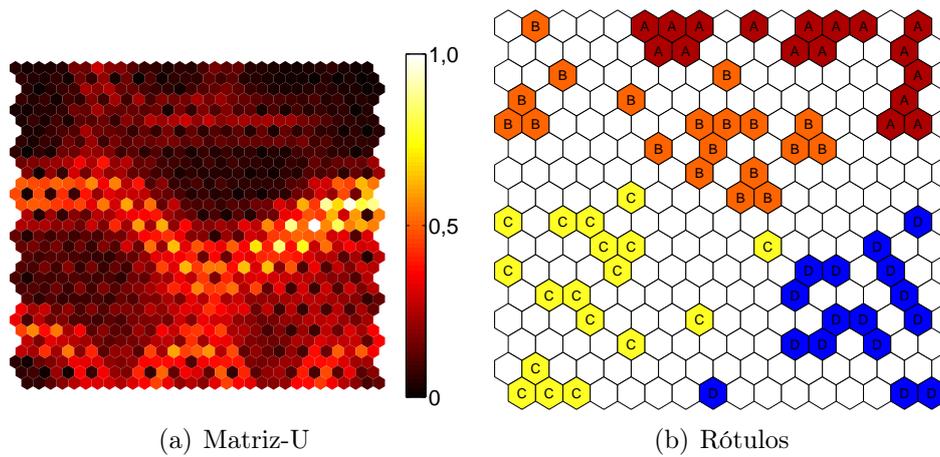


Figura B.1: Mapa de descritores da MIDE para falhas de curto circuito. As cores identificam as classes A (vermelho), B (laranja), C (amarelo) e D (azul).

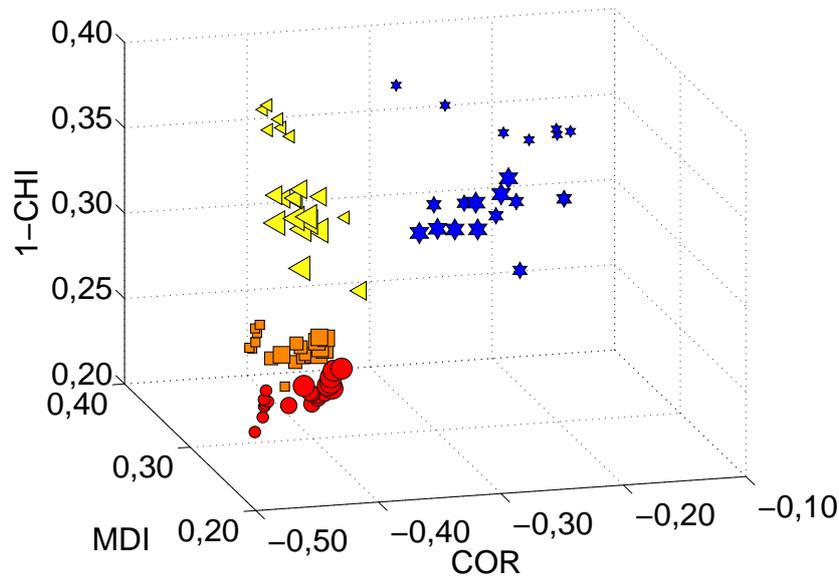


Figura B.2: Espaço dos descritores da MIDE para falhas de curto circuito. As cores identificam as classes A (círculo vermelho), B (retângulo laranja), C (triângulo amarelo) e D (estrela azul).

Deteção de saliências em imagens em níveis de cinza

Nas aplicações de deteção de saliências, sabe-se que estruturas de alto contraste com relação à vizinhança e de pequena escala estão presentes em regiões onde existe alguma informação relevante. Esse conhecimento *a priori* pode ser modelado pela TEM utilizando uma função de ponderação f_p que atribui maior importância para escalas pequenas. Uma boa escolha é, por exemplo, a da Figura B.3 que ilustra uma função $f_p = \exp(N_{max} - n)$, em que n é a escala e N_{max} é o valor máximo da escala.

O valor de N_{max} deve ser grande o suficiente para incluir todas as escalas desejadas, mas pequeno o suficiente para não acarretar um custo computacional elevado.

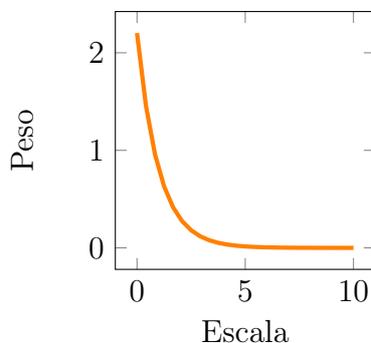


Figura B.3: Gráfico da função de ponderação utilizada na detecção de saliências.

A detecção de saliências em imagens em níveis de cinza segue a mesma linha de raciocínio para as imagens binárias. Contudo, ao invés de detetar estruturas compatíveis com os cantos de um contorno serão detetadas as estruturas salientes dos objetos da imagem. Estas estruturas salientes, em geral de alto contraste, contêm pontos de interesse (em inglês, *spots*). Pontos de interesse são bastante subjetivos, mas existe um entendimento de que, em um rosto humano, esses pontos estão concentrados principalmente nos olhos, boca, nariz (ULLMAN, 1996).

As imagens da Figura B.4 ilustram o resultado de um experimento de detecção de pontos de interesse em uma imagem de um rosto utilizando o algoritmo da Figura B.5. Conforme ilustrado na Figura B.4 (c), o algoritmo retorna pontos de interesse concentrados nos olhos, boca, cabelo e detalhes do vestuário.



Figura B.4: Detecção de saliências baseada na TEM. (a) Imagem original “Lena”. (b) Imagem estrutural com 5 níveis. (c) Regiões relevantes na imagem.

O algoritmo proposto consome um tempo médio de 1,2s para a busca por correlação nos pontos de interesse. Para efeito de comparação, o tempo da busca

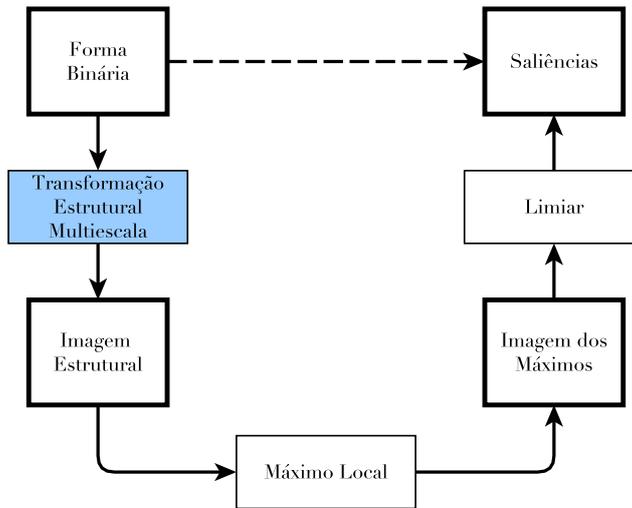


Figura B.5: Algoritmo de detecção de pontos de interesse.

por correlação em cada pixel da imagem, mesmo desconsiderando as regiões de bordas, é da ordem de minutos, dependendo do tamanho da imagem. Quanto maior a imagem, maior a diferença de tempo entre o método proposto e um algoritmo que realiza a busca em toda a imagem. Além disso, neste último caso, a chance de encontrar falsos positivos também aumenta.

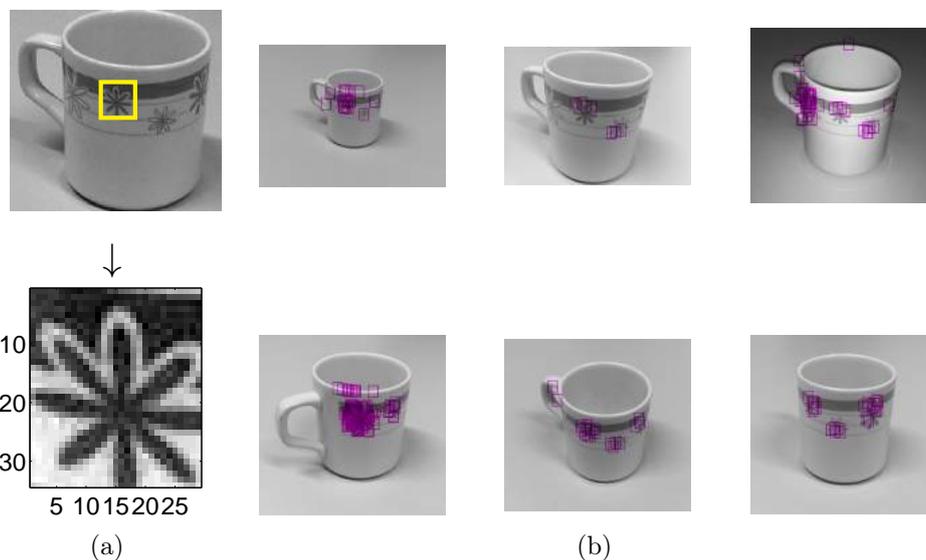


Figura B.6: Detecção de saliências com TEM e MIDE. (a) Imagem original e o recorte de 34×29 pixels utilizado na busca. (b) Regiões de busca resultantes da detecção de pontos de interesse para variações geométricas e de luminosidade.

Detecção de cantos em formas binárias

A característica multiescala da TEM a torna uma ferramenta bastante útil para detetar estruturas em determinadas escalas, especialmente quando a TEM generalizada é utilizada na forma da equação 5.3. O uso de uma função de ponderação de escala auxilia no destaque de determinadas estruturas, caso exista algum conhecimento *a priori* da escala dessas estruturas.

Base MPEG7-CE Shape-1 classe Device

Esta base contém 14 formas binárias geométricas. Os contornos das formas contêm ruído, o que torna mais difícil obter todos os cantos verdadeiros. Os cantos dessas formas foram identificados e registrados por Pedrosa e Barcelos (2010), formando o padrão ouro (PO) utilizado como referência para avaliação dos algoritmos.

A detecção de cantos em imagens binárias transformadas pela TEM pode ser observada no exemplo da Figura B.7, que utiliza a imagem “shark” da base MPEG-7. Foi utilizada a função de ponderação $f_p = \exp(10 - n)$. Para fins de comparação, a Figura B.7 (c) ilustra o resultado obtido com o método clássico de detecção de cantos de Harris e Stephens (1988). O algoritmo utilizado para a detecção de cantos está ilustrado na Figura B.8. Este algoritmo não incorpora nenhum pré-processamento da imagem de entrada e nenhum pós-processamento para eliminação de cantos falsos ou duplicados.

Para validar esta aplicação, foi realizado um experimento de detecção de cantos nas formas da classe “Device” da base MPEG-7. Alguns resultados estão ilustrados na Figura B.9. Os pontos em verde representam os cantos detetados que coincidem com o PO adotado por Pedrosa e Barcelos (2010). Obteve-se um valor médio para toda a base MPEG-7 com precisão $Pre = 0,8731$ ($\sigma = 0,1892$) (Apêndice A). O valor da precisão é compatível com o encontrado em métodos recentes de detecção de cantos da literatura (PEDROSA; BARCELOS, 2010). Neste exemplo, entretanto, os cantos detetados (em azul) excederam a quantidade referente ao PO, principalmente devido à presença de ruído no contorno das formas. O resultado da detecção de cantos em imagens binárias transformadas pela TEM pode ser observado na Figura B.9. Um refinamento neste método permite remover boa parte dos cantos falsos. Adotar uma regra de eliminação de cantos vizinhos próximos aumenta significativamente o valor da revocação.

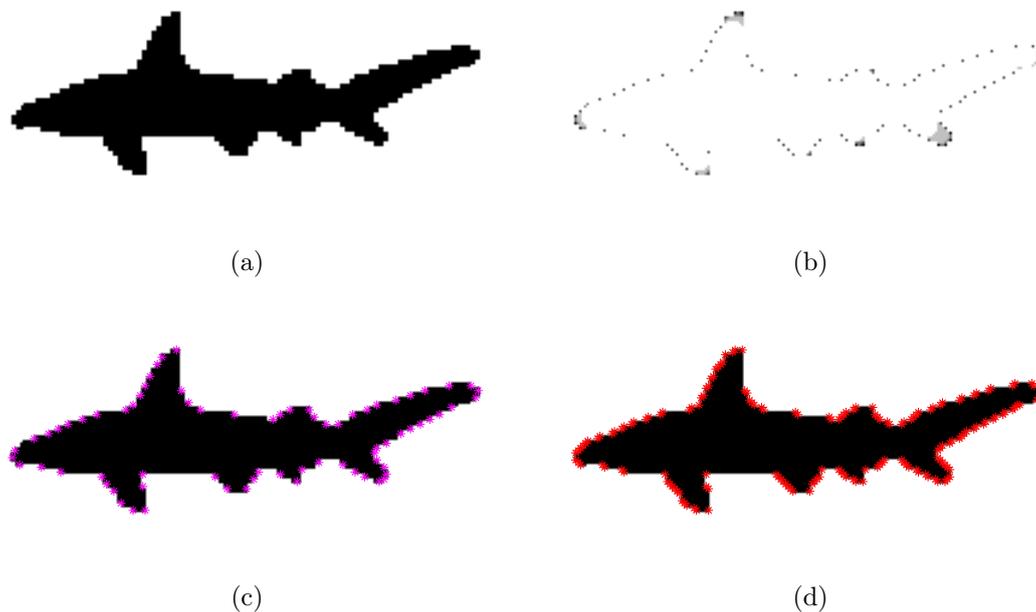


Figura B.7: Detecção de cantos com a TEM. (a) Imagem original “shark” da base MPEG-7. (b) Imagem realçada com 4 níveis. (c) Cantos detetados pelo método de Harris e Stephens (1988). (d) Cantos detetados com $\tau > 1$.

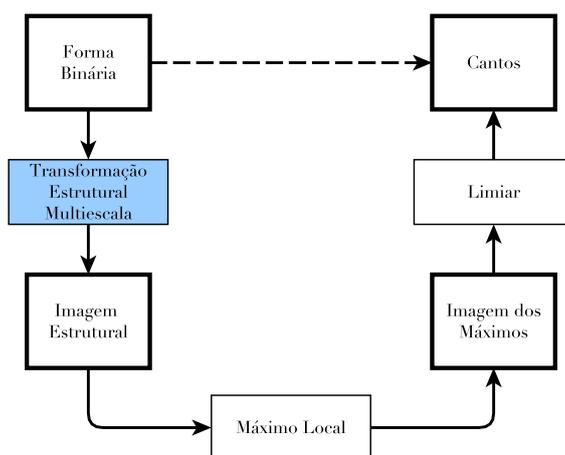


Figura B.8: Algoritmo para deteção de cantos.

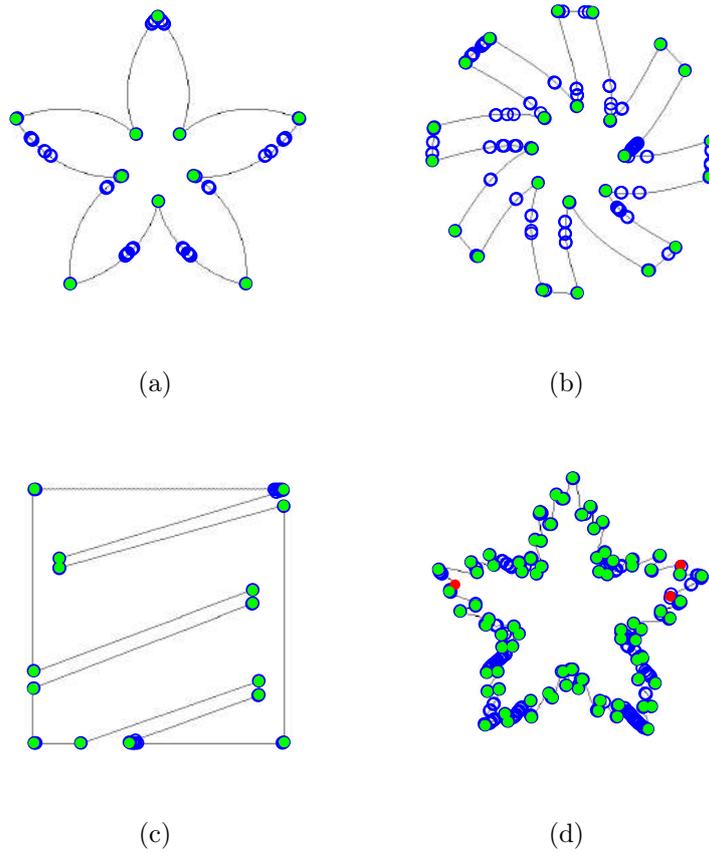


Figura B.9: Detecção de cantos com a TEM para formas da base MPEG-7 CE Shape-1 da classe “Device”. (a-c) Em verde os cantos que coincidem com o padrão outro. (d) Em vermelho os cantos não detetados ou não coincidentes com a verdade terrestre.

Detecção de microcalcificações em mamografias

As microcalcificações são aglomerados pequenos e de baixo contraste que aparecem em imagens radiográficas da mama. A detecção de microcalcificações é de extrema importância para que o médico seja capaz de identificar casos de câncer de mama nos estágios iniciais da doença, quando o tratamento é eficaz.

Base MIAS

Base de imagens mamográficas (SUCKLING, 1994) contendo mamografias de mamas sadias e diagnosticadas com microcalcificações ou tumores. As imagens possuem 1024×1024 pixels e apresentam uma resolução espacial de $200\mu m$. Três tipos de mamas estão presentes na base: Densa glandular (DG), Gordurosa (F) e Gordurosa glandular (FG). Para cada tipo existem casos benignos e malignos contendo as anomalias: CALC - microcalcificação, CIRC - massas circulares bem definidas, SPIC - massas em forma de agulha, MISC - massas características de alguma doença, ARCH - distorção arquitetônica, ASYM - assimetria e NORM - normal.

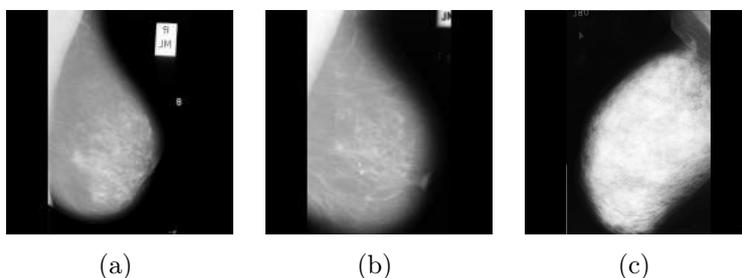


Figura B.10: Base MIAS. Densa, Fatty, Densa.

Realce de microcalcificações

O realce das mamografias é obtido pela aplicação direta da TEM na imagem original usando um função de ponderação definida pela expressão

$$f_p(r) = \frac{1}{a} \exp\left(-\frac{(r+s)^2}{a^2}\right), \quad (\text{B.1})$$

em que s é o tamanho médio de uma microcalcificação e a é um fator de incerteza desse tamanho. O tamanho médio de uma microcalcificação foi determinado como tendo aproximadamente 5 pixels de diâmetro (MELLOUL; JOSKOWICZ, 2002). A função de ponderação escolhida baseia-se na função da curva normal centrada em

s. Utilizando os parâmetros $s = 5$ e $a = 1/10$, a função de ponderação tem a forma ilustrada na Figura B.11.

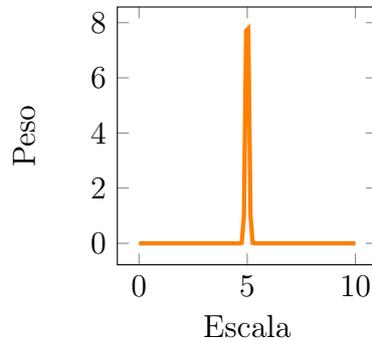


Figura B.11: Função de ponderação para realce de microcalcificações com a TEM.

As imagens da Figura B.12 ilustram o uso da TEM na tarefa de realce estrutural das microcalcificações. A imagem da Figura B.12 (a) é uma mamografia de mama densa contendo um aglomerado de microcalcificações. Por ser bastante densa, é difícil observar a presença desse aglomerado. O destaque ilustrado na Figura B.12 (c) deixa evidente que o baixo contraste e pequeno tamanho das microcalcificações são o principal motivo pelos quais sua detecção não é uma tarefa trivial. As imagens realçadas pela TEM são mostradas na Figura B.12 (b) e o detalhe na Figura B.12 (c). O realce com a TEM não apenas destaca a presença do aglomerado de microcalcificações do fundo denso da mama, mas torna perceptível o formato dessas microcalcificações e o arranjo espacial entre elas.

A Figura B.13 ilustra as mesmas regiões da Figura B.12 (c) e (d) em uma visualização tridimensional. Esta visualização mostra que os pixels da imagem realçada na Figura B.13 (b) que fazem parte das microcalcificações apresentam valores mais elevados e, portanto, um algoritmo de detecção de picos tem melhor desempenho nesta imagem do que na imagem original representada na Figura B.13 (a).

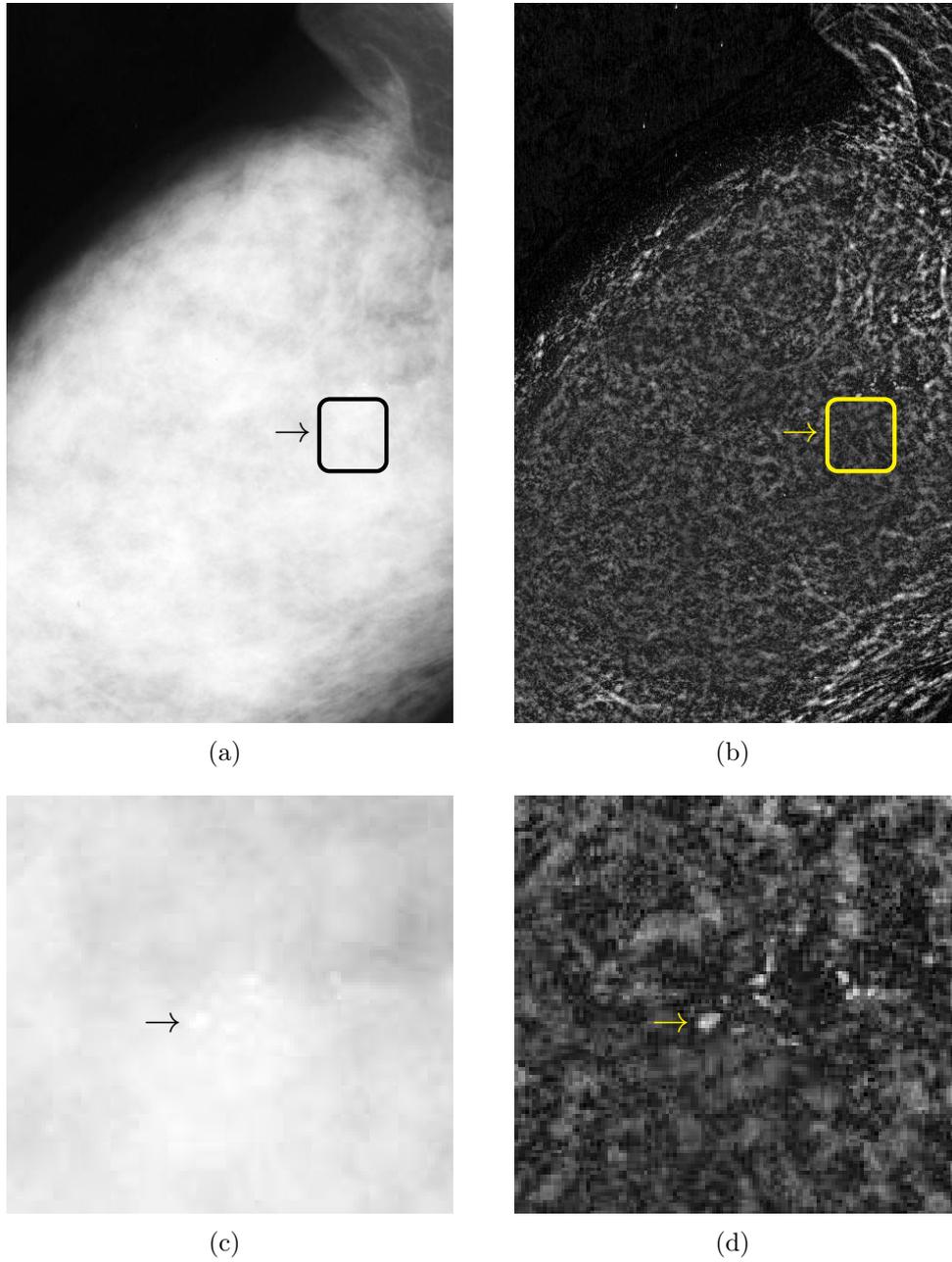


Figura B.12: Realce de estruturas que contêm microcalcificações em uma mamografia 1024 x 1024, $200\mu m$. Imagem 253 da base MIAS. (a) Imagem original com relevo suave. (b) Imagem realçada \mathbf{R} . (c) Detalhe da imagem original em uma região de microcalcificação. (d) Detalhe da imagem realçada em uma região de microcalcificação.

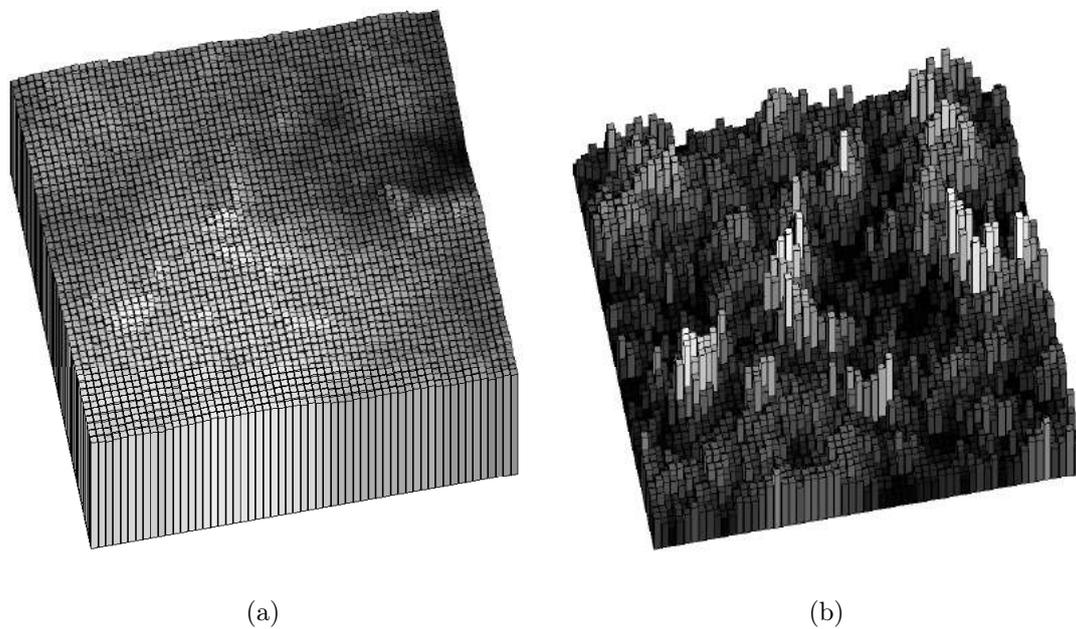


Figura B.13: Detalhe do relevo de uma região de mamografia contendo um aglomerado de microcalcificações. Imagem 253 da base MIAS (a) antes e (b) após realce estrutural.

Estimação do grau de sensitização em aços inoxidáveis austeníticos

Uma das principais características dos aços inoxidáveis austeníticos é a sua boa resistência à corrosão, determinada pela grande quantidade de Cr presente na liga, dentre outros fatores. O Cr atua como elemento de sacrifício no processo de oxidação natural dos metais. Durante a oxidação, o Cr livremente diluído na matriz do grão sofrerá preferencialmente a oxidação formando uma camada de óxido de cromo (OCr). Esta camada tem a propriedade de dificultar a extração de Fe da matriz, dificultando a formação de óxido de ferro (OFe).

A corrosão intergranular (IGC, do inglês *intergranular corrosion*) é um fenômeno que, dentre uma diversidade de fatores, ocorre devido à precipitação de carboneto de cromo ($M_{23}C_6$) nos contornos de grão da microestrutura cristalina dos aços inoxidáveis austeníticos. A precipitação provoca o enfraquecimento da concentração de cromo (Cr) na vizinhança do contorno de grão, o que aumenta a susceptibilidade do material à IGC. Esse fenômeno, também conhecido por sensitização do aço, geralmente ocorre quando o material está submetido à faixa de temperatura que vai de 450°C a 650°C quando em ambiente ácido, ou até 850°C sob determinadas condições. A IGC depende de muitas variáveis, incluindo a composição da liga e o meio no qual o material está submetido. Contudo, duas variáveis regem o comportamento da precipitação, capaz de tornar um aço inoxidável austenítico mais suscetível à IGC: a temperatura e o tempo de exposição.

As Figuras B.14 (a-c) ilustram exemplos de micrografias de aço inoxidável austenítico, com classificações AISI 304 e AISI 316, em diferentes níveis de sensitização conforme a norma ASTM-262: *Step* (não sensitizado), *Dual* (nível intermediário de sensitização) e *Ditch* (completamente sensitizado). O processo de preparação das amostras envolve etapas de tratamento térmico, lixamento, polimento e ataque químico com ácido oxálico. O ataque é mais profundo onde há maior deposição de $M_{23}C_6$. Por isso, a micrografia de um aço sensitizado apresenta regiões escuras e alongadas, cuja área é proporcional ao grau de sensitização. Para detectar essas regiões, e estimar com boa precisão o grau de sensitização por meio de processamento de imagem, é necessário realçar as áreas escuras localizadas nos contornos de grão. O algoritmo de realce está ilustrado no diagrama da Figura B.15. A função do realce é destacar os ataques mais profundos, os quais assume-se que possuem maior chance de serem os precipitados $M_{23}C_6$ que evidenciam a

sensitização. As micrografias realçadas pela TEM são segmentadas por um processo de limiarização. O resultado é exibido nas Figuras B.14 (d-f).

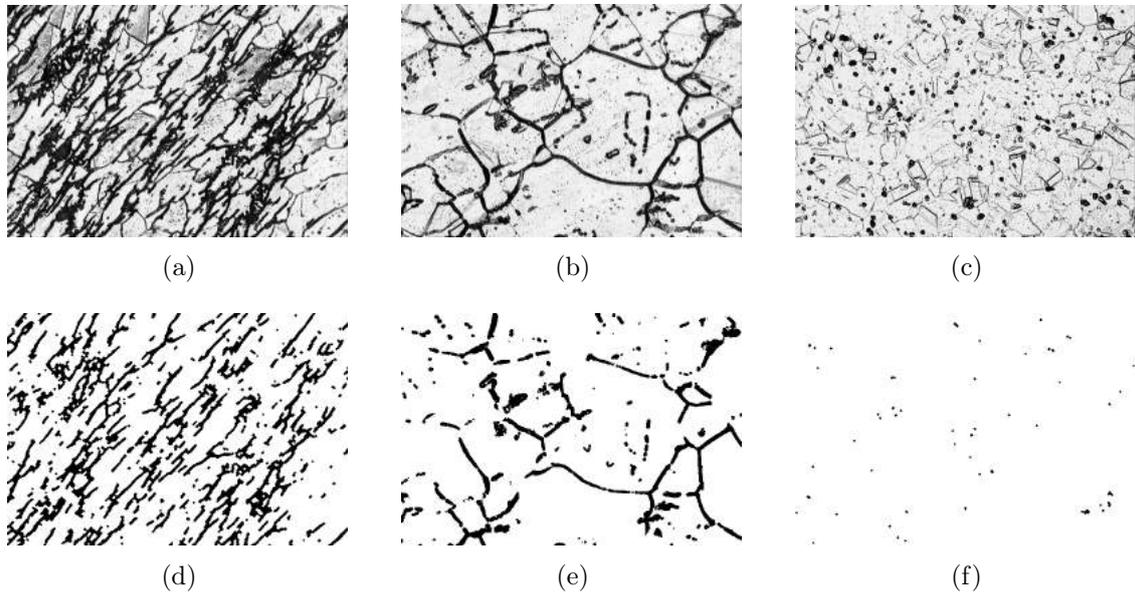


Figura B.14: Realce de precipitados em micrografias. (a,d) Ditch. (b,e) Dual. (c,f) Step.

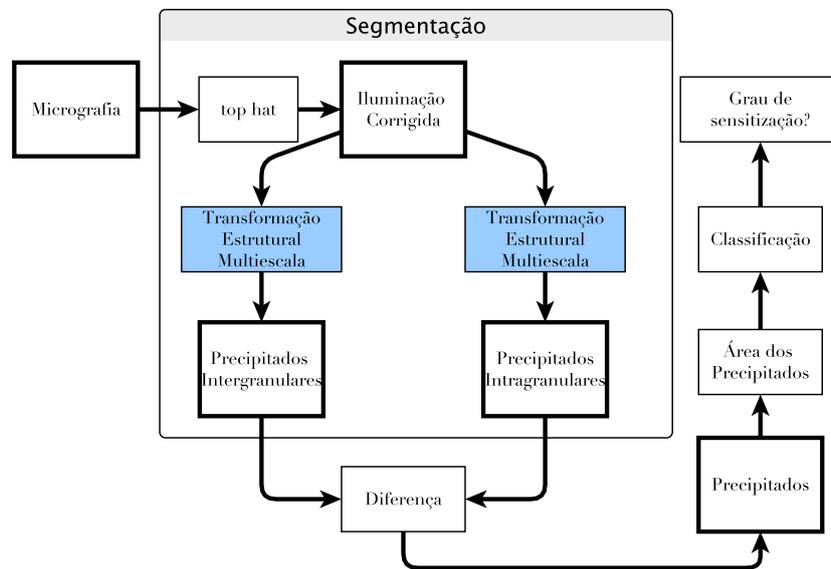


Figura B.15: Algoritmo para segmentação de precipitados intergranulares.

Sabe-se que o grau de sensitização, ou grau de susceptibilidade à IGC, é proporcional à área dos precipitados $M_{23}C_6$ nos contornos de grão (MAJIDI; STREICHER, 1984). Contudo, os precipitados $M_{23}C_6$ na matriz dos grãos também exercem papel importante no empobrecimento de Cr e, por isso, a área deles também

deve ser contabilizada. Isto pode ser observado na micrografia da Figura B.14 (e). O resultado do experimento é ilustrado na Figura B.16. Percebe-se que o valor estimado é suficiente para separar as amostras nas classes de sensitização, exceto para os três casos particulares que aparecem indicados com marcadores claros.

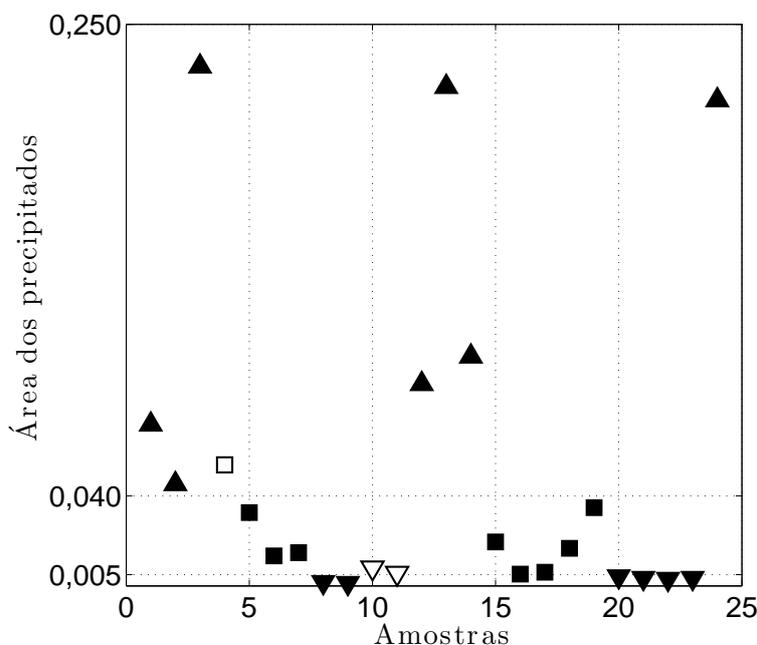


Figura B.16: Classificação das amostras de acordo com a área normalizada dos precipitados (ordenados por tipo de aço 304 e 316). Os marcadores escuros indicam as amostras classificadas corretamente.

Para avaliar o uso desse método de estimação da sensitização, foi realizado um experimento de classificação de amostras de aços inoxidáveis austeníticos tratados sob diferentes condições de tempo e temperatura. Adotou-se como limiar de separação entre classes o valor da área de precipitados obtido no experimento anterior da Figura B.16. Portanto, atribuiu-se a classe *Step* para área A de precipitados menores do que 0,005, *Dual* para o intervalo $0,005 < A \leq 0,04$ e *Ditch* para valores acima de 0,04. O resultado é apresentado na Figura B.17. Observa-se na matriz de confusão que foi possível alcançar uma taxa de classificação correta próxima a 88%. A curva ROC mostra que existe uma boa confiabilidade, já que a classificação apresenta taxas de verdadeiros positivos próxima a 1, enquanto a taxa de falsos positivos é pequena para as três classes. O relevante número de 12,5% (4,2%+8,3%) de falsos positivos está relacionado à variabilidade da preparação das amostras, incluindo o tratamento térmico, a preparação metalográfica e o ataque químico. Contudo, para efeito de seleção de amostras visando a manutenção

preventiva de equipamentos da indústria de petróleo e gás, o método proposto apresenta indicativos que podem auxiliar procedimentos de análise de falhas e inspeção.

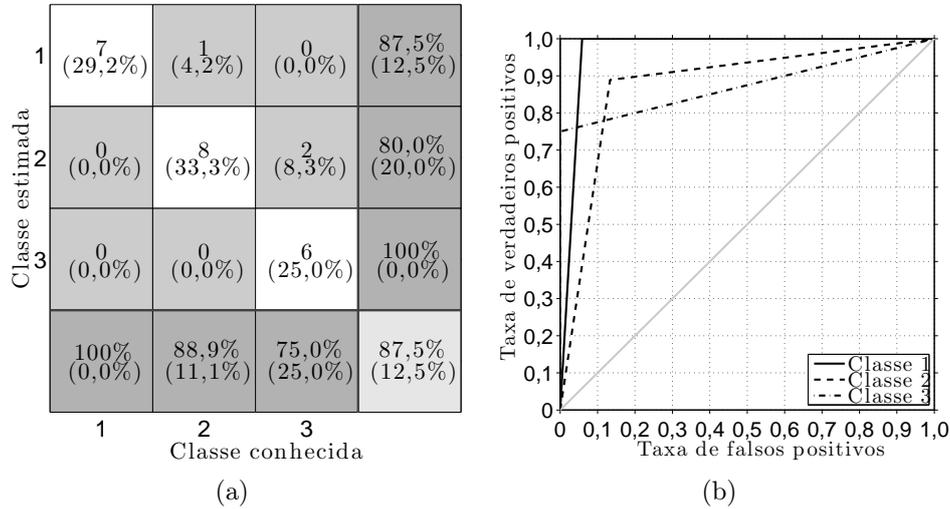


Figura B.17: Classificação de microestruturas com relação ao grau de susceptibilidade à IGC. (a) Matriz de confusão para as classes *Step* (1), *Dual* (2) e *Ditch* (3). (b) Curva ROC.

Detecção de nanopartículas de testosterona

A análise de nanopartículas de testosterona em imagens de microscopia eletrônica de varredura (MEV) é um dos diversos processos utilizados na avaliação de desempenho de fármacos no tratamento de curto-prazo da andropausa (BOTELHO *et al.*, 2013). A imagem da Figura B.19 (a) foi obtida em um MEV 30kV. A dificuldade em segmentar esse tipo de imagem se deve a dois fatores: i) o fundo da imagem não é uniforme, apresentando alto contraste entre regiões vizinhas; ii) a borda das nanopartículas possui alto contraste, enquanto o seu centro tem a mesma intensidade que o fundo da imagem ao seu redor. Estes fatores podem ser observados nas Figuras B.19 (a) e no seu detalhe em (b). Métodos clássicos de limiarização não abordam essas dificuldades, como ilustra o resultado da limiarização por Otsu (OTSU, 1979) na Figura B.19 (c). O histograma da Figura B.19 (e) não permite estimar corretamente o diâmetro das nanopartículas.

O método de segmentação proposto emprega duas transformações TEM com o objetivo de realçar as nanopartículas antes de segmentar a imagem. A primeira transformação emprega a função de ponderação $f_p(n) = n - 15$, que permite realçar as partículas de menor tamanho, com diâmetros em torno de 10 pixels. A função

de ponderação $f_p(n) = \exp(1 - (n - 15)^2/5)$, da segunda transformação, é utilizada para realçar as partículas de diâmetro próximo a 15 pixels. As funções de ponderação estão ilustradas na Figura B.18. A segmentação é obtida por um procedimento de limiarização dos pixels das imagens realçadas, utilizando como limiar o valor médio adicionado do desvio padrão das intensidades.

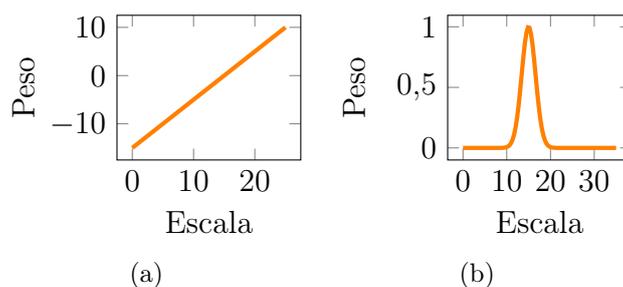
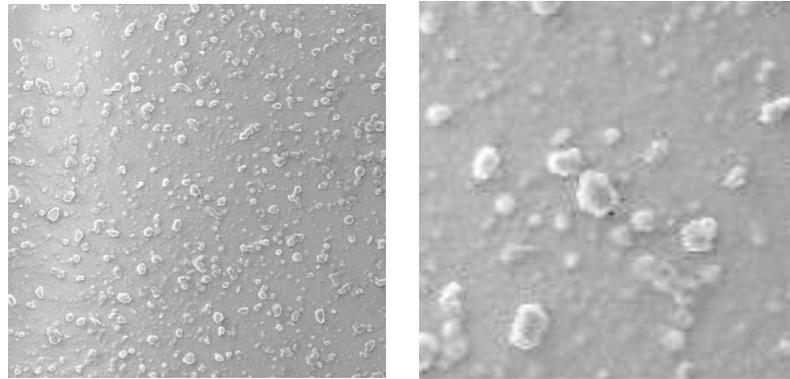


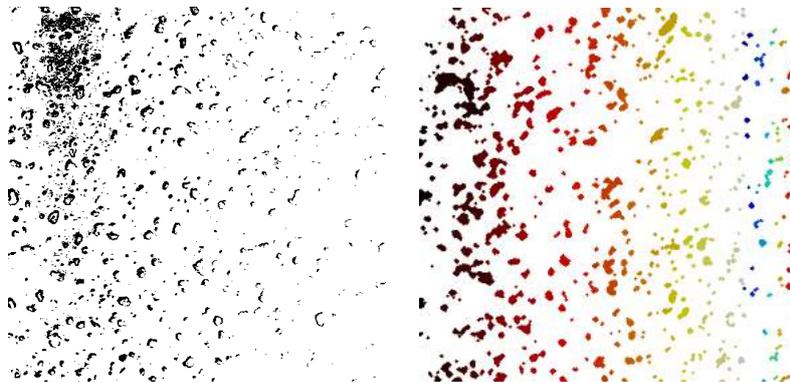
Figura B.18: Função de ponderação para realce de microcalcificações com a TEM.

Para exemplificar o emprego deste método, nas Figuras B.19 (f) e (g) são exibidos histogramas da estimativa de diâmetro das nanopartículas, em pixels. A análise da geometria dos objetos da imagem segmentada permite inferir, de forma automática, informações sobre as nanopartículas para auxiliar estudos de desempenho de fármacos.



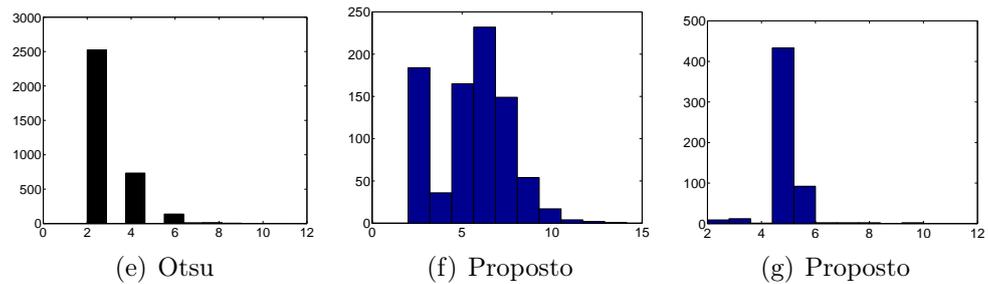
(a) Original

(b) Detalhe



(c) Otsu

(d) Proposto



(e) Otsu

(f) Proposto

(g) Proposto

Figura B.19: (a) Imagem original (gentilmente cedida pelo Prof. Dr. Marco Botelho, RENORBIO). (b) Detalhe de uma nanopartícula. O contraste foi alterado para melhor exibição. (c) Limiarização pelo método de Otsu (1979). (d) Rótulos da segmentação pelo método proposto. (e) Histograma do diâmetro em pixels das nanopartículas da imagem (c). (f) Histograma do diâmetro em pixels das nanopartículas maiores da imagem (d). (g) Histograma do diâmetro em pixels das nanopartículas menores da imagem (d).

Referências Bibliográficas

- ASHA, V.; BHAJANTRI, N.; NAGABHUSHAN, P. GLCM-based chi-square histogram distance for automatic detection of defects on patterned textures. **International Journal of Computational Vision and Robotics**, v. 2, n. 4, p. 302, fev. 2011.
- ASSILZADEH, H.; MANSOR, S. B. Early warning system for oil spill using SAR images. **22nd Asian Conference on Remote Science**, v. 1, p. 460–465, nov. 2001.
- BEVK, M.; KONONENKO, I. A statistical approach to texture description of medical images: a preliminary study. **Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems**, p. 239–244, jun. 2002.
- BOOMGAARD, R. v.; BALEN, R. van. Methods for fast morphological image transforms using bitmapped. **Graphical Models and Image Processing**, v. 54, n. 3, p. 252–258, maio 1992.
- BOTELHO, M. A.; QUEIROZ, D. B.; FREITAS, A.; GUERREIRO, S.; UMBELINO, S.; BARROS, G.; BORGES, B.; RUELA, R.; ALMEIDA, J. G.; JR, L. Q. Effects of a new testosterone transdermal delivery system, biolipid b2-testosterone in healthy middle aged men: A confocal Raman spectroscopy study. **Journal of Pharmaceutical and Scientific Innovation**, v. 2, p. 1–7, abr. 2013.
- BRAGA-NETO, U.; GOUTSIAS, J. Object-based image analysis using multiscale connectivity. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 6, p. 892–907, jun. 2005.
- BREKKE, C.; SOLBERG, A. H. S. Oil spill detection by satellite remote sensing. **Remote Sensing of Environment**, v. 95, n. 1, p. 1–13, mar. 2005.
- CAI-XIANG, Y.; SHU-BO, Q. Image retrieval based on texture and color features. **WASE International Conference on Information Engineering**, v. 1, p. 125–128, jul. 2009.
- CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 8, n. 6, p. 679–698, nov. 1986.

CHETOUANIA, A.; BEGHADIA, A.; DERICHE, M. A hybrid system for distortion classification and image quality evaluation. **Signal Processing: Image Communication**, v. 27, n. 9, p. 948–960, out. 2012.

CHOI, H.; LEE, C. No-reference image quality metric based on image classification. **EURASIP Journal on Advances in Signal Processing**, v. 2011, n. 65, p. 1–11, set. 2011.

CHOI, K. Y.; KIM, S. S. Morphological analysis and classification of types of surface corrosion damage by digital image processing. **Corrosion Science**, v. 47, n. 1, p. 1–15, jan. 2005.

COLOMÉ-SERRA, M. F.; SERÓN, D.; MORESO, F.; CARRERA, M.; GRINÓ, J. M.; SERRA, J. Image analysis: utility of grey level granulometry to measure renal interstitial chronic damage. **Engineering in Medicine and Biology Society, 1992 14th Annual International Conference of the IEEE**, v. 5, p. 1934–1935, nov. 1992.

COSTA, L. F.; CESAR, J. R. M. **Shape Classification and Analysis - Theory and Practice**. 2a. ed. Nova Iorque: CRC Press, 2009. 612 p.

DAUBECHIES, I. **Ten Lectures on Wavelets**. 1a. ed. Filadelfia: Society for Industrial and Applied Mathematics (SIAM), 1992. 360 p.

DELVES, L. M.; WILKINSON, R.; OLIVER, C. L.; WHITE, R. G. Comparing the performance of SAR image segmentation algorithms. **International Journal of Remote Sensing**, v. 13, n. 11, p. 2121–2149, jan. 1992.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2a. ed. Inglaterra: Wiley, 2000.

ELEYAN, A.; DEMIREL, H. Co-occurrence matrix and its statistical features as a new approach for face recognition. **Turk J Elec Eng & Comp Sci.**, v. 19, n. 1, p. 97–107, jan. 2011.

FELIX, J. H. da S.; CORTEZ, P. C.; HOLANDA, M. A. Sistema automático para quantificação e visualização da aeração pulmonar em imagens de tomografia computadorizada de tórax: Sistema de análise de imagens pulmonares – SAIP. **Revista Brasileira de Engenharia Biomédica**, v. 26, n. 3, p. 195–208, dez. 2010.

FISHER, R. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, p. 179–188, set. 1936.

FRATE, F. D.; PETROCCHI, A.; LICHTENEGGER, J.; CALABRESI, G. Neural networks for oil spill detection using ERS-SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 38, n. 5, p. 2282–2287, set. 2000.

- FRAZ, M.; BARMAN, S.; REMAGNINO, P.; HOPPE, A.; BASIT, A.; UYYANONVARA, B.; RUDNICKA, A.; OWEN, C. An approach to localize the retinal blood vessels using bit planes and centerline detection. **Computer Methods and Programs in Biomedicine**, v. 108, n. 2, p. 600—616, nov. 2011.
- FRAZ, M.; REMAGNINO, P.; HOPPEA, A.; UYYANONVARA, B.; RUDNICKA, A.; OWEN, C.; BARMAN, S. Blood vessel segmentation methodologies in retinal images – a survey. **Computer Methods and Programs in Biomedicine**, v. 108, n. 1, p. 407–433, abr. 2012.
- GADKARI, D. **Image Quality Analysis Using GLCM**. Dissertação (Mestrado) — College of Arts and Sciences, Universidade da Flórida Central, Orlando, dez. 2004.
- GAO, X.; LU, W.; TAO, D.; LI, X. Image quality assessment based on multiscale geometric analysis. **IEEE Transactions on Image Processing**, v. 18, n. 7, p. 1409–1423, maio 2009.
- GEGÚNDEZ-ARIAS, M. E.; AQUINO, A.; BRAVO, J. M.; MARÍN, D. A function for quality evaluation of retinal vessel segmentations. **IEEE Transactions on Medical Imaging**, v. 31, n. 2, p. 231–239, fev. 2012.
- GIBSON, R. M.; AHMADINIA, A.; MCMEEKIN, S. G.; STRANG, N. C.; MORISON, G. A reconfigurable real-time morphological system for augmented vision. **EURASIP Journal on Advances in Signal Processing**, v. 134, ago. 2013.
- GINESU, G.; MASSIDDA, F.; GIUSTO, D. D. A multi-factors approach for image quality assessment based on a human visual system model. **Signal Processing: Image Communication**, v. 21, n. 4, p. 316—333, abr. 2006.
- GONZALEZ, R. C.; WOODS, R. E. **Image Processing**. 2a. ed. Nova Jérsei: Prentice Hall, 2001.
- HANBURY, J. S. A. Analysis of oriented textures using mathematical morphology. **Proceedings of the Annual Workshop of the Austrian Association for Pattern Recognition (OAGM)**, p. 201–208, set. 2002.
- HARALICK, K. S. R. M.; DINSTEIN, I. Textural features for image classification. **IEEE Transactions on Systems, Man and Cybernetics**, v. 3, n. 6, p. 610–621, nov. 1973.
- HARALICK, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, v. 67, n. 5, p. 786–804, maio 1979.
- HARALICK, R. M.; SHAPIRO, L. G. **Computer and Robot Vision**. 1a. ed. Nova Iorque: Addison-Wesley, 1992. 682 p.
- HARDING, P.; ROBERTSON, N. M. Visual saliency from image features with application to compression. **Cognitive Computation**, p. 76–98, mar. 2013.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. **Proceedings of the 4th Alvey Vision Conference**, p. 147–151, ago. 1988.

HAYKIN, S. **Redes Neurais Artificiais**. 2a. ed. Porto Alegre: Bookman, 2001.

HOOVER, A. D.; KOUZNETSOVA, V.; GOLDBAUM, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. **IEEE Transactions on Medical Imaging**, v. 19, p. 203–210, mar. 2000.

HORN, R. A.; JOHNSON, C. R. **Topics in Matrix Analysis**. 1a. ed. Nova Iorque: Cambridge University Press, 1991. 612 p.

HU, D.; TIAN, X. A multi-directions algorithm for edge detection based on fuzzy mathematical morphology. **Proceedings of the 16th International Conference on Artificial Reality and Telexistence**, p. 361 – 364, nov. 2006.

HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. **IEEE International Joint Conference on Neural Networks. Proceedings**, v. 2, p. 985–990, jul. 2004.

_____. Extreme learning machine: theory and applications. **Neurocomputing**, v. 70, n. 1-3, p. 489–501, dez. 2006.

ITTI, L.; KOCH, C. Computational modeling of visual attention. **Nature Reviews Neuroscience**, v. 2, n. 3, p. 194–203, mar. 2001.

JACCARD, P. The distribution of the flora in the alpine zone. **New Phytologist**, v. 11, n. 2, p. 37–50, fev. 1912.

KIM, D. H.; YUN, I. D.; LEE, S. U. Shape decomposition scheme by combining mathematical morphology and convex partitioning. **ACCV2002: The 5th Asian Conference on Computer Vision**, p. 418–423, jan. 2002.

KOENDERINK, J. J. The structure of images. **Biological Cybernetics**, v. 50, n. 5, p. 363–370, ago. 1984.

KOENDERINK, J. J.; DOORN, A. J. van. Invariant features of contrast detection: an explanation in terms of self-similar detector arrays. **Journal of Optical Society of America**, v. 72, n. 1, p. 83–87, jan. 1982.

KOHONEN, T. **Self-Organizing Maps**. 3a. ed. Berlin: Springer, 2001. 501 p.

LAZEBNIK, S.; SCHMID, C.; PONCE, J. A sparse texture representation using local affine regions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 8, p. 1265–78, ago. 2005.

LEITE, T. P. C. **Detecção de estruturas finas e ramificadas em imagens usando campos aleatórios de Markov e informação perceptual**. Tese (Doutorado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, ago. 2012.

LEYDESDORFF, L.; VAUGHAN, L. Co-occurrence matrices and their applications in information science: extending ACA to the web environment. **Journal of the American Society for Information Science and Technology**, v. 57, n. 12, p. 1616–1628, out. 2006.

LI, C.; BOVIK, A. C. Content-weighted video quality assessment. **Journal of Electronic Imaging**, v. 19, n. 1, p. 1–9, jan. 2010.

LIU, Y.; LOH, H. T.; TOR, S. B. Comparison of extreme learning machine with support vector machine for text classification. **Innovations in Applied Artificial Intelligence - Lecture Notes in Computer Science**, v. 3533, p. 390–399, jun. 2005.

MACKAY, D. J. **Information Theory, Inference, and Learning Algorithms**. 4a. ed. Cambridge: Cambridge University Press, 2005. 640 p.

MAJIDI, A. P.; STREICHER, M. The double loop reactivation method for detecting sensitization in AISI 304 stainless steels. **Corrosion**, v. 40, n. 11, p. 584–593, nov. 1984.

MARAGOS, P. Pattern spectrum and multiscale shape representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 701–716, jul. 1989.

MARÍN, D.; AQUINO, A.; Gegúndez-Arias, M. E.; BRAVO, J. M. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. **IEEE Transactions on Medical Imaging**, v. 30, n. 1, p. 146–158, jan. 2011.

MARR, D. **Vision - A Computational Investigation into the Human Representation and Processing of Visual Information**. 1a. ed. New York: MIT Press, 1982. 429 p.

MASCARO, A.; MELLO, C.; SANTOS, W.; CAVALCANTI, G. D. Mammographic images segmentation using texture descriptors. **Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society**, v. 2009, p. 3653–6, jan. 2009.

MASCHAL, J. R. A.; YOUNG, S. S.; REYNOLDS, J. P.; KRAPELS, K.; FANNING, J.; CORBIN, T. New image quality assessment algorithms for CFA demosaicing. **IEEE Sensors Journal**, v. 13, n. 1, p. 371–378, jan. 2013.

MATTHEWS, B. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. **Biochimica et Biophysica Acta**, v. 405, n. 2, p. 442–451, out. 1975.

MEDEIROS, F. N. S.; RAMALHO, G. L. B.; BENTO, M. P.; MEDEIROS, L. C. L. On the evaluation of texture and color features for nondestructive corrosion

detection. **EURASIP Journal on Advances in Signal Processing**, v. 2010, n. 817473, p. 1–8, jul. 2010.

MELLOUL, M.; JOSKOWICZ, L. Segmentation of microcalcification in X-ray mammograms using entropy thresholding. **CARS 2002 Computer Assisted Radiology and Surgery**, p. 671–676, jun. 2002.

MENDONCA, A.; CAMPILHO, A. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. **IEEE Transactions on Medical Imaging**, v. 25, n. 9, p. 1200–1213, set. 2006.

MERCIER, G.; GIRARD-ARDHUIN, F. Unsupervised oil slick detection by SAR imagery using kernel expansion. **Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. IEEE International Proceedings.**, v. 1, jul. 2005.

MITREA, D.; NEDEVSCHI, S.; SOCACIU, M.; BADEA, R. The role of the superior order GLCM in the characterization and recognition of the liver tumors from ultrasound images. **Radio Engineering**, v. 21, n. 1, p. 79–85, abr. 2012.

MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. No-reference image quality assessment in the spatial domain. **IEEE Transactions on Image Processing**, v. 21, n. 12, p. 4695–4708, dez. 2012.

MUKHOPADHYAY, S.; CHANDA, B. A multiscale morphological approach to local contrast enhancement. **Signal Processing**, v. 80, n. 4, p. 685–696, abr. 2000.

NGUYEN-DUC, H.; DO-HONG, T.; LE-TIEN, T.; BUI-THU, C. A new descriptor for image retrieval using contourlet co-occurrence. **2010 Third International Conference on Communications and Electronics (ICCE)**, p. 169–174, ago. 2010.

OTSU, N. A threshold selection method from gray level histograms. **IEEE Transactions on Systems, Man and Cybernetics**, v. 9, p. 62–66, jan. 1979.

PEDROSA, G. V.; BARCELOS, C. a.Z. Anisotropic diffusion for effective shape corner point detection. **Pattern Recognition Letters**, Elsevier B.V., v. 31, n. 12, p. 1658–1664, set. 2010.

PEDROSA, G. V.; BATISTA, M. a.; BARCELOS, C. a.Z. Image feature descriptor based on shape salience points. **Neurocomputing**, Elsevier, v. 120, p. 156–163, nov. 2013.

PÉREZ, J.; MAS, D.; ESPINOSA, J.; VÁZQUEZ, C.; ILLUECA, C. Propagation, structural similarity and image quality. **Proceedings of the SPIE 8427, Biophotonics: Photonic Solutions for Better Health Care III**, v. 8427, p. 1–10, jun. 2012.

PIANPIAN, W.; BEIZHAN, W.; CHENG, Q.; LIYAN, C. Improved generalized co-occurrence matrix for image retrieval using MPEG-7 texture edge. **2008 International Symposium on Information Science and Engineering**, p. 205–208, dez. 2008.

RADKE, R. J.; ANDRA, S.; AL-KOFAHI, O.; ROYSAM, B. Image change detection algorithms: A systematic survey. **IEEE Transactions on Image Processing**, v. 14, n. 3, p. 294–307, mar. 2005.

RAMALHO, G. L. B.; MEDEIROS, F. N. S. de. Using boosting to improve oil spill detection in SAR images. **18th International Conference on Pattern Recognition**, Hong Kong, v. 2, p. 1066–1069, ago. 2006.

_____. Improving Reliability of Oil Spill Detection Systems Using Boosting for High-Level Feature Selection. **International Conference on Image Analysis and Recognition 2007. Lecture Notes in Computer Science**, Montreal, v. 4633, p. 1172–1181, ago. 2007.

REBOUÇAS FILHO, P. P. **Métodos de contornos ativos Crisp adaptativo 2D e 3D aplicados na segmentação dos pulmões em imagens de tomografia computadorizada do tórax**. Tese (Doutorado) — Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Fortaleza, mar. 2013.

ROOMI, M. M.; S.SARANYA. Bayesian classification of fabrics using binary co-occurrence matrix. **International Journal of Information Sciences and Techniques**, v. 2, n. 2, p. 1–9, mar. 2012.

ROTSHTEIN, P.; GENG, J. J.; DRIVER, J.; DOLAN, R. J. Role of features and second-order spatial-relations in face discrimination, face recognition, and individual face skills: behavioral and fMRI data. **Journal of Cognitive Neuroscience**, v. 19, n. 9, p. 1435–1452, dec. 2008.

ROUSE, D. M.; HEMAMI, S. S. Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM. **SPIE: Human Vision and Electronic Imaging XIII**, v. 6806, p. 1–14, fev. 2008.

SAHOO, P. K.; ARORA, G. A thresholding method based on two-dimensional Renyi's entropy. **Pattern Recognition**, v. 37, n. 6, p. 1149–1161, jun. 2004.

SAMPAT, M. P.; WANG, Z.; GUPTA, S.; BOVIK, A. C.; MARKEY, M. K. Complex wavelet structural similarity: A new image similarity index. **IEEE Transactions on Image Processing**, v. 18, n. 11, p. 2385–2401, nov. 2009.

SERRA, J. **Image Analysis and Mathematical Morphology**. 1a. ed. Londres: Academic Press Inc., 1982. 610 p.

SETHIAN, J. A. **Level Set Methods and Fast Marching Methods**. 2a. ed. Cambridge: Cambridge, 1999. 378 p.

- SHEIKH, H. R.; BOVIK, A. C. Image information and visual quality. **IEEE Transactions on Image Processing**, v. 15, n. 2, p. 430–444, fev. 2006.
- SHEIKH, H. R.; BOVIK, A. C.; VECIANA, G. de. An information fidelity criterion for image quality assessment using natural scene statistics. **IEEE Transactions on Image Processing**, v. 14, n. 12, p. 2117–2128, dez. 2005.
- SHEIKH, H. R.; SABIR, M. F.; BOVIK, A. C. A statistical evaluation of recent full reference image quality assessment algorithms. **IEEE Transactions on Image Processing**, v. 15, n. 11, p. 3441–3452, nov. 2006.
- SHEIKH, H. R.; WANG, Z.; CORMACK, L.; BOVIK, A. C. **Live Image Quality Assessment Database Release 2**. 2005. Disponível em: <<http://live.ece.utexas.edu/research/quality>>.
- SHI, Z.; YANG, Z.; ZHANG, G.; CUI, G.; XIONG, X.; LIANG, Z.; LU, H. Characterization of texture features of bladder carcinoma and the bladder wall on MRI. **Academic Radiology**, v. 20, n. 8, p. 930–938, ago. 2013.
- SHINHA, D.; DOUGHERTY, E. R. Fuzzy mathematical morphology. **Journal of Visual Communication and Image Representation**, v. 3, n. 3, p. 286–302, set. 1992.
- SHNAYDERMAN, A.; GUSEV, A. An SVD-based grayscale image quality measure for local and global assessment. **IEEE Transactions on Image Processing**, v. 15, n. 2, p. 422–429, fev. 2006.
- SOARES, J. V. B.; LEANDRO, J. J. G.; CESAR, R. M.; JELINEK, H. F.; CREE, M. J. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. **IEEE Transactions on Medical Imaging**, v. 25, n. 9, p. 1214–1222, set. 2006.
- SOILLE, P. **Morphological Image Analysis**. 2a. ed. Berlin: Springer, 2002. 391 p.
- SOLBERG, A. S.; BREKKE, C.; SOLBERG, R.; HUSOY, P. O. Algorithms for oil spill detection in Radarsat and ENVISAT SAR images. **Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International**, v. 7, p. 4909–4912, set. 2004.
- STAAL, J.; ABRAMOFF, M.; NIEMEIJER, M.; VIERGEVER, M.; GINNEKEN, B. van. Ridge based vessel segmentation in color images of the retina. **IEEE Transactions on Medical Imaging**, v. 23, n. 4, p. 501–509, abr. 2004.
- STOJIC, T.; RELJIN, B. Enhancement of microcalcifications in digitized mammograms: multifractal and mathematical morphology approach. **Faculty of Mechanical Engineering Transactions**, v. 38, n. 1, p. 1–9, 2010.

SUCKLING, J. The mammographic image analysis society digital mammogram database. **Exerpta Medica. International Congress Series 1069**, p. 375–378, jan. 1994.

SUK, T.; FLUSSER, J. Refined morphological methods of moment computation. **20th International Conference on Pattern Recognition**, p. 966–970, ago. 2010.

SUK, T.; HOSCHL, I. C.; FLUSSER, J. Decomposition of binary images—a survey and comparison. **Pattern Recognition**, v. 45, n. 12, p. 4279–4291, dec. 2012.

THURLEY, M. J.; DANELL, V. Fast morphological image processing open-source extensions for GPU processing with CUDA. **IEEE Journal of Selected Topics in Signal Processing**, v. 6, n. 7, p. 849–855, nov. 2012.

ULLMAN, S. **High-Level Vision**. 1a. ed. Massachusetts: MIT Press, 1996. 357 p.

ULTSCH, A.; SIEMON, H. Kohonen’s self-organizing feature maps for exploratory data analysis. in **Proc. of INNC’90, International Neural Network Conference, Dordrecht, Holanda**, p. 305–308, jul. 1990.

VILLALOBOS-CASTALDI, F.; FELIPE-RIVERÓN, E. M.; SÁNCHEZ-FERNÁNDEZ, L. P. A fast, efficient and automated method to extract vessels from fundus images. **Journal of Visualization**, v. 13, n. 3, p. 263–270, ago. 2010.

VINCENT, L. Morphological transformations of binary images with arbitrary structuring elements. **Signal Processing**, v. 22, n. 1, p. 3–23, jan. 1991.

_____. Morphological gray scale reconstruction in image analysis: applications and efficient algorithms. **IEEE Transactions on Image Processing**, v. 2, n. 2, p. 176–201, abr. 1993.

VIZIREANU, D. N. Generalizations of binary morphological shape decomposition. **Journal of Electronic Imaging**, v. 16, n. 1, mar. 2007.

WANG, Y.; CAO, F.; YUAN, Y. A study on effectiveness of extreme learning machine. **Neurocomputing**, Elsevier, v. 74, n. 16, p. 2483–2490, set. 2011.

WANG, Z.; BOVIK, A. C. A universal image quality index. **IEEE Signal Processing Letters**, v. 9, n. 3, p. 81–84, mar. 2002.

_____. Why is image quality assessment so difficult. **IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)**, v. 4, p. 3313–3316, maio 2002.

_____. Mean squared error: love it or leave it? - a new look at signal fidelity measures. **IEEE Signal Processing Magazine**, p. 98–117, jan. 2009.

- WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. **IEEE Transactions on Image Processing**, v. 13, n. 4, p. 600–612, abr. 2004.
- WANG, Z.; LUC, L.; BOVIK, A. C. Video quality assessment based on structural distortion measurement. **Signal Processing: Image Communication**, v. 19, n. 2, p. 121–132, fev. 2004.
- WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multi-scale structural similarity for image quality assessment. **Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2003**, v. 2, p. 1398–1402, nov. 2003.
- WEBB, A. R. **Statistical Pattern Recognition**. 2a. ed. Inglaterra: Wiley, 2002.
- WELFER, D.; SCHARCANSKI, J.; KITAMURA, C. M.; PIZZOL, M. M. D.; LUDWIG, L. W.; MARINHO, D. R. Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach. **Computers in Biology and Medicine**, v. 40, n. 2, p. 124–137, fev. 2010.
- WU, H.; SUN, T.; WANG, J.; LI, X.; WANG, W.; HUO, D.; LV, P.; HE, W.; WANG, K.; GUO, X. Combination of radiological and gray level co-occurrence matrix textural features used to distinguish solitary pulmonary nodules by computed tomography. **Journal of Digital Imaging**, v. 26, n. 4, p. 797–802, ago. 2013.
- XU, C. L.; CHEN, Y. Q. Statistical landscape features for texture classification. **Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004**, v. 1, p. 676–679, ago. 2004.
- XU, J. Morphological decomposition 2D binary shapes into modestly overlapped disk components. **IEEE International Conference on Image Processing, ICIP 2005**, v. 2, p. 470–473, set. 2007.
- YAO, A.; YU, S. Spatial feature interdependence matrix (SFIM): a robust descriptor for face recognition. **Lecture Notes in Computer Science: Advances in Image and Video Technology**, v. 7087, p. 132–143, nov. 2011.
- _____. Robust face representation using hybrid spatial feature interdependence matrix. **IEEE Transactions on Image Processing**, v. 22, n. 8, p. 3247–3259, ago. 2013.
- ZANA, F.; KLEIN, J. C. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. **IEEE transactions on image processing: a publication of the IEEE Signal Processing Society**, v. 10, n. 7, p. 1010–9, jan. 2001.
- ZHANG, F.; LIN, W.; CHEN, Z.; NGAN, K. N. Additive log-logistic model for networked video quality assessment. **IEEE Transactions on Image Processing**, v. 22, n. 4, p. 1536–1547, abr. 2013.

ZHANG, H.; FRITTS, J. E.; GOLDMAN, S. A. Image segmentation evaluation: a survey of unsupervised methods. **Computer Vision and Image Understanding**, v. 110, n. 2, p. 260—280, jul. 2008.

ZHANG, Y. J. A survey on evaluation methods for image segmentation. **Pattern Recognition**, v. 29, n. 8, p. 1335–1346, ago. 1996.

ZIMERMANN, H.-J. **Fuzzy Set Theory – and its applications**. 3a. ed. Boston: Kluwer Academic Publisher, 1992. 441 p.