

Hélio Pedrini

William Robson Schwartz



Análise de Imagens Digitais

Princípios, Algoritmos
e Aplicações



Sumário

LISTA DE ABREVIATURAS E SIGLAS	xv
LISTA DE SÍMBOLOS	xvii
LISTA DE ALGORITMOS	xix
1 INTRODUÇÃO	1
1.1 Áreas de Aplicação	2
1.2 Etapas de um Sistema de Processamento de Imagens	3
1.3 Componentes de um Sistema de Processamento de Imagens	5
1.4 Organização do Livro	7
Referências Bibliográficas	9
2 FUNDAMENTOS DE IMAGENS DIGITAIS	13
2.1 Tipos de Sinais	13
2.2 Modelo de Imagens	14
2.3 Digitalização	15
2.4 Resolução Espacial e Profundidade da Imagem	18
2.5 Representação de Imagens Digitais	19
2.6 Imagem Multibanda ou Multiespectral	19
2.7 Imagem Multidimensional	22
2.8 Sistema Visual Humano	24
2.9 Ruído em Imagens	26
2.10 Entropia em Imagens	28
2.11 Relacionamentos Básicos entre Elementos de Imagem	30
2.11.1 Vizinhança	30
2.11.2 Conectividade	31
2.11.3 Adjacência	31
2.11.4 Caminho	31
2.11.5 Componentes Conexos	32
2.11.6 Borda e Interior	32
2.11.7 Medidas de Distância	32
2.11.8 Operações Lógicas e Aritméticas	34
2.12 Métricas de Qualidade em Imagens	37
2.12.1 Erro Máximo	38

2.12.2	Erro Médio Absoluto	38
2.12.3	Erro Médio Quadrático	38
2.12.4	Relação Sinal-Ruído de Pico	39
2.12.5	Covariância	39
2.12.6	Coeficiente de Correlação	40
2.12.7	Coeficiente de Jaccard	40
2.13	Limites da Imagem	40
2.14	Notas Bibliográficas	41
2.15	Problemas	41
	Referências Bibliográficas	43
3	TRANSFORMADAS DE IMAGENS	45
3.1	Transformações de Coordenadas	46
3.1.1	Núcleo da Transformada	46
3.1.2	Definição de Imagem	49
3.1.3	Representação das Transformadas de Imagens	50
3.2	Transformada Discreta de Fourier	51
3.2.1	Transformada Unidimensional de Fourier	53
3.2.2	Transformada Bidimensional de Fourier	57
3.2.3	Propriedades da Transformada de Fourier	59
3.2.4	Transformada Rápida de Fourier	67
3.3	Transformada Discreta de Hartley	76
3.4	Transformada Discreta do Cosseno	77
3.5	Transformada Discreta do Seno	80
3.6	Transformada de Fourier-Mellin	82
3.7	Transformada de Karhunen-Loève	86
3.8	Transformada Wavelet	90
3.9	Notas Bibliográficas	95
3.10	Problemas	96
	Referências Bibliográficas	97
4	REALCE DE IMAGENS	101
4.1	Brilho e Contraste	101
4.2	Histograma	104
4.3	Transformação da Escala de Cinza	105
4.3.1	Transformações Lineares e Não-lineares	107
4.3.2	Equalização de Histograma	109
4.3.3	Hiperbolização de Histograma	115
4.3.4	Especificação de Histograma	115
4.4	Filtragem de Imagens	118
4.4.1	Filtragem no Domínio Espacial	118
4.4.2	Filtragem no Domínio de Freqüência	133
4.5	Técnica de Meios-Tons	137
4.6	Realce de Imagens Baseado em Cores	141
4.6.1	Transformação Pseudocor	141

4.6.2	Realce com Transformação HSI	142
4.6.3	Realce por Falsa Cor	143
4.7	Notas Bibliográficas	144
4.8	Problemas	144
	Referências Bibliográficas	147
5	SEGMENTAÇÃO DE IMAGENS	151
5.1	Detecção de Descontinuidades	152
5.1.1	Detecção de Pontos	152
5.1.2	Detecção de Retas	153
5.1.3	Detecção de Bordas	153
5.1.4	Detecção de Junções ou Cantos	168
5.1.5	Transformada de Hough	174
5.2	Limiarização	182
5.2.1	Limiarização Global	184
5.2.2	Limiarização Local	191
5.3	Segmentação de Regiões	194
5.3.1	Crescimento de Regiões	196
5.3.2	Divisão de Regiões	197
5.3.3	Divisão e União de Regiões	198
5.3.4	Segmentação por Divisor de Águas	199
5.4	Notas Bibliográficas	200
5.5	Problemas	200
	Referências Bibliográficas	201
6	COMPRESSÃO DE IMAGENS	207
6.1	Redundância de Dados	208
6.1.1	Redundância de Codificação	208
6.1.2	Redundância Interpixel	210
6.1.3	Redundância Psicovisual	211
6.2	Elementos de Teoria da Informação	212
6.3	Métodos de Compressão de Imagens	216
6.3.1	Compressão sem Perdas	216
6.3.2	Compressão com Perdas	230
6.4	Padronizações de Compressão de Imagens	237
6.4.1	Compressão de Imagens Binárias	238
6.4.2	Compressão de Imagens Estáticas Monocromáticas e Coloridas	238
6.4.3	Compressão de Seqüência de Imagens Monocromáticas e Coloridas	242
6.5	Notas Bibliográficas	244
6.6	Problemas	244
	Referências Bibliográficas	246

7 REPRESENTAÇÃO E DESCRIÇÃO	249
7.1 Esquemas de Representação	249
7.1.1 Código da Cadeia	250
7.1.2 Aproximações Poligonais	252
7.1.3 Assinatura	254
7.1.4 Fecho Convexo	254
7.1.5 Esqueleto de um Objeto	255
7.2 Descritores de Bordas	259
7.2.1 Diâmetro	260
7.2.2 Perímetro	261
7.2.3 Curvatura	262
7.2.4 Energia de Deformação	263
7.2.5 Descritores de Fourier	263
7.2.6 Modelos Deformáveis	265
7.3 Descritores de Região	267
7.3.1 Área	267
7.3.2 Compacidade ou Circularidade	268
7.3.3 Projeções Horizontal e Vertical	269
7.3.4 Propriedades Topológicas	270
7.3.5 Momentos	271
7.3.6 Eixos Maior e Menor	276
7.3.7 Excentricidade	276
7.3.8 Retângulo Envolvente	277
7.3.9 Convexidade	277
7.3.10 Retangularidade	278
7.3.11 Solidez	278
7.4 Descritores Sintáticos ou Relacionais	279
7.5 Notas Bibliográficas	280
7.6 Problemas	281
Referências Bibliográficas	282
8 ANÁLISE DE TEXTURAS	287
8.1 Abordagem Estatística	288
8.1.1 Medidas Baseadas na Distribuição de Níveis de Cinza	289
8.1.2 Matriz de Co-ocorrência	291
8.1.3 Matrizes de Comprimento de Corridas de Cinza	297
8.1.4 Função de Autocorrelação	300
8.1.5 Matriz da Diferença dos Tons de Cinza da Vizinhança	301
8.2 Abordagem Baseada em Processamento de Sinais	304
8.2.1 Decomposição de Valor Singular	304
8.2.2 Espectro de Fourier	305
8.2.3 Transformada Wavelet	306
8.3 Abordagem Geométrica	307
8.3.1 Unidade de Textura	308
8.3.2 Codificação de Características de Texturas	314

8.4	Abordagem Baseada em Modelos Paramétricos	316
8.4.1	Campos Aleatórios de Markov	317
8.5	Notas Bibliográficas	322
8.6	Problemas	322
	Referências Bibliográficas	324
9	MORFOLOGIA MATEMÁTICA	327
9.1	Fundamentos Matemáticos	327
9.2	Operadores Morfológicos em Imagens Binárias	331
9.2.1	Dilatação e Erosão	332
9.2.2	Abertura e Fechamento	336
9.2.3	Transformada Acerto-ou-Erro	337
9.3	Aplicações da Morfologia Matemática em Imagens Binárias	339
9.3.1	Extração de Bordas	340
9.3.2	Preenchimento de Regiões	341
9.3.3	Extração de Componentes Conexos	342
9.3.4	Fecho Convexo	343
9.3.5	Afinamento e Espessamento	345
9.3.6	Extração do Esqueleto de Objetos	347
9.3.7	Poda	348
9.4	Operadores Morfológicos em Imagens Monocromáticas	350
9.4.1	Dilatação e Erosão	350
9.4.2	Abertura e Fechamento	352
9.4.3	Realce de Contraste	353
9.4.4	Granulometria	354
9.4.5	Atenuação de Ruído	355
9.5	Notas Bibliográficas	356
9.6	Problemas	357
	Referências Bibliográficas	359
10	REGISTRO DE IMAGENS	361
10.1	Transformações Geométricas em Imagens	361
10.1.1	Transformações Espaciais	362
10.1.2	Interpolação de Intensidade	376
10.2	Formação Geométrica da Imagem	379
10.2.1	Modelos de Câmeras	379
10.2.2	Calibração de Câmera	386
10.2.3	Visão Estereoscópica	387
10.3	Técnicas para Registro de Imagens	389
10.4	Notas Bibliográficas	392
10.5	Problemas	393
	Referências Bibliográficas	394

11 CLASSIFICAÇÃO DE PADRÕES	397
11.1 Abordagens Aplicadas à Classificação de Padrões	399
11.1.1 Abordagem Sintática	399
11.1.2 Abordagem Estatística	400
11.2 Estimação de Parâmetros	404
11.2.1 Máxima Verossimilhança	405
11.2.2 Algoritmo EM	407
11.3 Histograma e Janela de Parzen	408
11.4 Classificador Bayesiano	410
11.4.1 Modelo de Análise Discriminante Gaussiana	412
11.4.2 Naïve Bayes	414
11.5 Redes Neurais Artificiais	417
11.5.1 Unidade Perceptron	417
11.5.2 Algoritmo Retropropagação	420
11.6 Regra dos k-vizinhos mais Próximos	421
11.7 Agrupamento de Dados	423
11.7.1 Tendência de Agrupamento	425
11.7.2 Algoritmos Hierárquicos	427
11.7.3 Algoritmos Particionais	436
11.8 Avaliação do Desempenho de Métodos de Classificação	438
11.8.1 Contagem Simples	439
11.8.2 Conjunto de Teste	440
11.8.3 Matriz de Confusão	440
11.8.4 Validação Cruzada	441
11.9 Notas Bibliográficas	442
11.10 Problemas	442
Referências Bibliográficas	444
A FUNDAMENTOS MATEMÁTICOS	447
A.1 Números Complexos	447
A.2 Álgebra Matricial	448
A.2.1 Espaços Vetoriais	449
A.2.2 Transformações Lineares	451
A.2.3 Transformações de Coordenadas	451
B PROBABILIDADE E ESTATÍSTICA	453
B.1 Conceitos de Probabilidade e Estatística	453
B.2 Estatística Multivariada	457
Referências Bibliográficas	459
C RADIOMETRIA E FOTOMETRIA	461
C.1 Radiometria	461
C.1.1 Ângulo Sólido	461
C.1.2 Energia Radiante	462
C.1.3 Fluxo Radiante	462

C.1.4	Irradiância	463
C.1.5	Intensidade Radiante	463
C.1.6	Radiânci a	463
C.2	Fotometria	463
C.2.1	Intensidade luminosa	464
C.2.2	Fluxo luminoso	464
C.2.3	Iluminânci a	464
C.2.4	Luminânci a	464
C.3	Unidades Radiométricas e Fotométricas	464
D	MODELOS DE CORES	467
D.1	Fundamentos de Cor	467
D.2	Modelos de Cores	469
D.2.1	Modelo de Munsell	471
D.2.2	Modelo RGB	471
D.2.3	Modelo XYZ	472
D.2.4	Modelo CMY	472
D.2.5	Modelo YIQ	473
D.2.6	Modelo YUV	473
D.2.7	Modelo YCbCr	474
D.2.8	Modelo HSV	474
D.2.9	Modelo HSL	475
D.2.10	Modelo HSI	476
D.2.11	Modelos L*u*v* e L*a*b*	477
REFERÊNCIAS BIBLIOGRÁFICAS	479	
ÍNDICE REMISSIVO	503	

Listas de Algoritmos

3.1	Núcleo para a transformada unidimensional	48
3.2	Núcleo para a transformada bidimensional	49
3.3	Transformadas de imagens	51
3.4	Transformadas separáveis	52
3.5	Transformada rápida de Fourier (FFT)	74
3.6	Permutação dos índices de um vetor	74
3.7	Método linha-coluna para a FFT bidimensional	76
3.8	Núcleo da transformada de Karhunen-Loève	89
4.1	Cálculo do histograma de uma imagem	105
4.2	Equalização de histograma	113
4.3	Processo de convolução	121
4.4	Filtragem com preservação de bordas	129
4.5	Técnica de pontilhado com difusão de erro de Floyd e Steinberg	139
5.1	Determinação de pontos de borda em uma imagem	156
5.2	Detecção de retas pela transformada de Hough	176
5.3	Detecção de circunferências pela transformada de Hough	180
5.4	Transformada de Hough generalizada	182
5.5	Limiarização global iterativa	186
5.6	Divisão e união de regiões	199
6.1	Codificação de Huffman	218
6.2	Codificação de Shannon-Fano	220
6.3	Codificação LZW	223
6.4	Decodificação LZW	224
7.1	Afinamento de Stentiford e Mortimer	260
11.1	Algoritmo EM	409
11.2	Classificação baseada no modelo de análise discriminante Gaussiano	413
11.3	Naïve Bayes	416
11.4	Perceptron	419
11.5	Retropropagação	422
11.6	k-vizinhos mais próximos	422
11.7	Agrupamentos hierárquicos aglomerativos	429
11.8	Agrupamentos hierárquicos divisivos	435
11.9	Agrupamento seqüencial básico	437
11.10	Algoritmo k-means	439
11.11	Validação cruzada	441

2

FUNDAMENTOS DE IMAGENS DIGITAIS

O propósito deste capítulo é introduzir os principais conceitos relacionados a imagens digitais e a terminologia utilizada ao longo deste livro. Um modelo físico de imagens é apresentado por meio de dois componentes, a iluminância e a reflectância. O processo de digitalização, composto pelos passos de amostragem e quantização, é descrito para sinais unidimensionais e bidimensionais. Os conceitos de resolução espacial e profundidade da imagem são introduzidos e discutidos por meio de exemplos.

Representações de imagens digitais são descritas e ilustradas. As principais características do sistema visual humano são apresentadas, permitindo a compreensão das relações existentes entre a visão humana e a visão computacional. Os conceitos de ruído e entropia em imagens são definidos e ilustrados para facilitar sua compreensão.

Os relacionamentos básicos entre elementos de imagens são apresentados, destacando-se os critérios de conectividade, as medidas de distância e as operações lógicas e aritméticas sobre imagens. Finalmente, métricas para avaliar a qualidade de imagens são apresentadas.

2.1 Tipos de Sinais

O conceito de *sinal* varia com respeito ao contexto no qual ele está sendo utilizado. De um ponto de vista geral, um sinal é a manifestação de um fenômeno que pode ser expresso de forma quantitativa.

Um sinal pode ser representado como uma função de uma ou mais variáveis independentes e, tipicamente, contém informação acerca da natureza ou comportamento do fenômeno físico sob consideração¹. Por exemplo, o sinal de voz pode ser definido por uma função de uma variável (tempo), enquanto o sinal correspondente a uma imagem pode ser definido por uma função de duas variáveis (espaço).

No domínio temporal, pode-se analisar como as variações do sinal evoluem com o decorrer do tempo. Um sinal pode ser *contínuo* ou *discreto*. Em um sinal contínuo, seus estados podem ser definidos em qualquer instante de tempo, ou seja, sem interrupção. Por outro lado, um sinal discreto é definido por um conjunto de valores enumeráveis ou inteiros, cujo intervalo depende da natureza do sinal.

Sinais podem ainda ser classificados como *analógicos* ou *digitais*. Sinais analógicos podem variar continuamente no tempo, enquanto um sinal digital pode assumir apenas

¹Há situações em que um sinal não pode ser modelado por uma equação, por exemplo, sinais aleatórios.

valores discretos. Uma onda sonora é um exemplo de sinal analógico, enquanto o código Morse, utilizado em telegrafia, é um exemplo de sinal digital. Circuitos eletrônicos podem converter um tipo de sinal em outro. Uma onda sonora capturada por um microfone deve ser transformada em sinal digital para manipulação em um computador ou para transmissão como qualquer outro tipo de dado. Em contrapartida, para reproduzir o som em um alto-falante, o sinal de áudio digital deve ser convertido em sinal analógico.

2.2 Modelo de Imagens

A representação e manipulação de uma imagem em computador requer a definição de um modelo matemático adequado da imagem.

Uma *imagem* pode ser definida como uma função de intensidade luminosa, denotada $f(x, y)$, cujo valor ou amplitude nas coordenadas espaciais (x, y) fornece a intensidade ou o brilho da imagem naquele ponto. A figura 2.1 mostra uma imagem e a orientação do sistema de coordenadas utilizada neste livro. Por convenção, a origem da imagem está localizada no canto superior esquerdo da imagem.

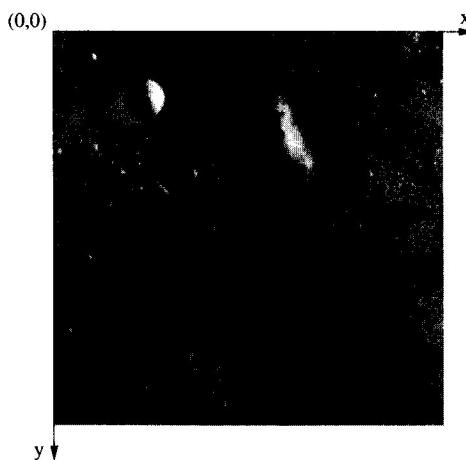


Figura 2.1: Convenção do sistema de coordenadas para representação de imagens digitais.

Um modelo físico para a intensidade de uma cena sob observação pode ser expresso em termos do produto entre dois componentes, a quantidade de luz incidente na cena e a quantidade de luz refletida pelos objetos presentes na cena. Esses componentes são chamados de *iluminância* e *reflectância*, respectivamente, e são representados por $i(x, y)$ e $r(x, y)$. Assim, a função $f(x, y)$ pode ser representada como

$$f(x, y) = i(x, y) r(x, y) \quad (2.1)$$

para

$$0 < i(x, y) < \infty \quad \text{e} \quad 0 < r(x, y) < 1 \quad (2.2)$$

A natureza de $i(x, y)$ é determinada pela fonte de luz, enquanto $r(x, y)$ é determinada pelas características dos objetos na cena. Os valores para os componentes $i(x, y)$ e $r(x, y)$

da equação 2.2 são limites teóricos. A iluminância é medida em lúmen/m² ou lux, enquanto a reflectância é medida em valores percentuais ou no intervalo entre 0 e 1.

Valores médios ilustram alguns intervalos típicos de $i(x, y)$. Em um dia claro, o Sol pode produzir aproximadamente 900000 lúmen/m² de iluminância na superfície da Terra. Esse valor decresce para menos de 10000 lúmen/m² em um dia nublado. O nível de iluminância típico em um escritório é de aproximadamente 1000 lúmen/m². Em uma noite clara, a lua cheia gera aproximadamente 0.1 lúmen/m². Por outro lado, valores típicos de $r(x, y)$ são 0.93 para a neve, 0.80 para parede branca, 0.65 para aço inoxidável e 0.01 para veludo preto (Gonzalez e Woods 2002).

Para melhor compreensão das grandezas associadas à interação da luz em uma cena, o leitor pode consultar o apêndice C.

2.3 Digitalização

Desde que a maioria das técnicas de análise de imagens é realizada por meio de processamento computacional, a função $f(x, y)$ deve ser convertida para a forma discreta. Uma *imagem digital* pode ser obtida por um processo denominado *digitalização*, o qual envolve dois passos, a amostragem e a quantização.

A *amostragem* (*sampling*) consiste em discretizar o domínio de definição da imagem nas direções x e y , gerando uma matriz de $M \times N$ amostras, respectivamente. A *quantização* consiste em escolher o número inteiro L de níveis de cinza (em uma imagem monocromática) permitidos para cada ponto da imagem.

Cada elemento $f(x, y)$ dessa matriz de amostras é chamado pixel (acrônimo do inglês *picture element*), com $0 \leq x \leq M - 1$ e $0 \leq y \leq N - 1$. A imagem contínua $f(x, y)$ é aproximada, portanto, por uma matriz de dimensão M pixels na horizontal (eixo x da figura 2.1) e N pixels na vertical (eixo y da figura 2.1), como mostrado em 2.3.

$$f(x, y) \approx \begin{bmatrix} f(0, 0) & f(1, 0) & \dots & f(M - 1, 0) \\ f(0, 1) & f(1, 1) & \dots & f(M - 1, 1) \\ \vdots & \vdots & \ddots & \vdots \\ f(0, N - 1) & f(1, N - 1) & \dots & f(M - 1, N - 1) \end{bmatrix} \quad (2.3)$$

O conceito de dimensão de um pixel ao longo do eixo x , ou do eixo y , está relacionado com o espaçamento físico entre as amostras. Cada pixel tem associado um valor $L_{\min} \leq f(x, y) \leq L_{\max}$, tal que o intervalo $[L_{\min}, L_{\max}]$ é denominado *escala de cinza*. A intensidade f de uma *imagem monocromática* nas coordenadas (x, y) é chamada de *nível de cinza* da imagem naquele ponto. Uma convenção comum é atribuir a cor preta ao nível de cinza mais escuro (por exemplo, valor 0) e atribuir a cor branca ao nível de cinza mais claro (por exemplo, valor 255).

A digitalização adequada de uma imagem (ou outro sinal) requer alguns cuidados para que nenhuma informação seja perdida no processo de amostragem. Um desses cuidados inclui a escolha correta do espaçamento entre as amostras tomadas da imagem contínua. Reformulando o problema de maneira diferente, o objetivo é encontrar condições de amostragem sob as quais a imagem contínua possa ser completamente recuperada a partir de um

conjunto de valores amostrados. Inicialmente, a análise do problema será realizada por meio de um sinal unidimensional e, posteriormente, estendido para o caso bidimensional.

Considere um sinal $f(x, y)$ com banda limitada² no domínio $[-B, B]$ do espaço de freqüências, sendo B um número real. A freqüência de amostragem, F_a , é a freqüência espacial com que as amostras do sinal são tomadas no processo de amostragem e está relacionada com o intervalo de amostragem Δx , na direção x , pela seguinte equação

$$F_a = \frac{1}{\Delta x} \quad (2.4)$$

A escolha da freqüência de amostragem adequada é determinada pelo *teorema da amostragem de Whittaker-Shannon*, o qual estabelece que um sinal contínuo pode ser completamente reconstruído a partir de um conjunto de amostras, se

$$\Delta x \leq \frac{1}{2B} \quad (2.5)$$

ou, de forma similar, se

$$F_a \geq 2B \quad (2.6)$$

Em outras palavras, o teorema afirma que pelo menos uma amostra a cada meio período do sinal deve ser tomada para que o sinal possa ser completamente reconstruído ou, de forma similar, que a freqüência de amostragem seja no mínimo duas vezes a freqüência máxima do sinal a ser amostrado.

O limite de freqüência de amostragem $\frac{1}{2B}$ é conhecido como *limite de Nyquist*, em homenagem a Harry Nyquist (1928), que demonstrou a importância do limite nas áreas de telefonia e telegrafia. Seus experimentos mostraram que não era necessário transmitir o sinal de voz completo para que a conversação fosse compreendida, bastando enviar pequenas amostras do sinal elétrico correspondente à voz, tomadas a intervalos regulares. Um circuito eletrônico poderia ser utilizado para transformar as amostras em um sinal analógico muito semelhantes ao sinal original.

Se a equação 2.5 não for satisfeita, um fenômeno denominado *aliasing* ocorrerá, comprometendo a completa recuperação do sinal. O fenômeno de *aliasing* pode ser observado por meio de um sinal periódico simples. Seja o sinal $f(t) = a \sin(2\pi f_0 t)$, com freqüência f_0 e amplitude a que varia no tempo t . Um gráfico do sinal $f(t)$ é mostrado na figura 2.2(a). Neste caso, a banda limitada de $f(t)$ é $[-f_0, f_0]$. O limite de Nyquist é, portanto, dado por

$$\Delta t \leq \frac{1}{2f_0} \quad (2.7)$$

Assim, pelo menos uma amostra a cada meio período do sinal deve ser tomada para que o sinal possa ser completamente reconstruído, conforme ilustrado na figura 2.2(b). Na figura 2.2(c), a taxa de amostragem é quatro vezes superior ao limite de Nyquist. Pode-se notar que as amostras representam corretamente o sinal.

A figura 2.3 mostra um sinal periódico (curva contínua). Uma taxa de amostragem inferior ao limite de Nyquist é utilizada para reconstruir o sinal. Como resultado, um

²Um sinal $f(x)$ com banda limitada possui a propriedade de que sua transformada de Fourier $F(u)$ possui valores muito baixos para u fora do intervalo $[-B, B]$. A transformada de Fourier é descrita em detalhes no capítulo 3.

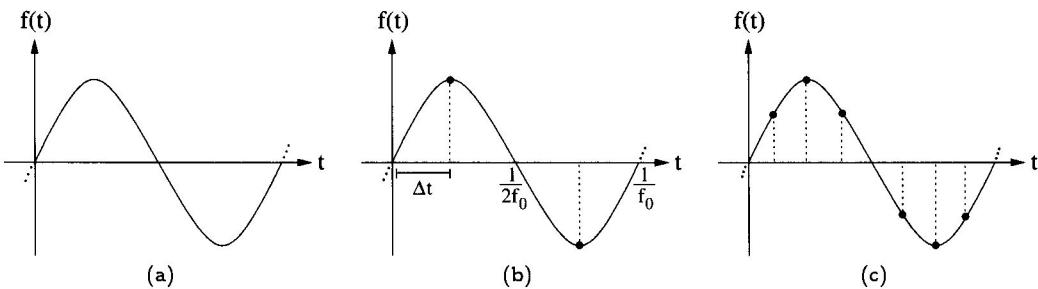
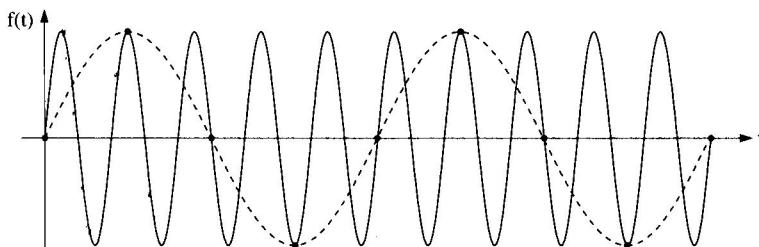


Figura 2.2: Taxa de amostragem de um sinal periódico.

sinal completamente distinto (mostrado na curva tracejada) do sinal original é obtido. Na reconstrução do sinal, obtém-se um sinal de freqüência muito inferior à freqüência do sinal original. As altas freqüências do sinal original aparecem como baixas freqüências no sinal reconstruído, caracterizando o fenômeno de *aliasing*.

Figura 2.3: Fenômeno de *aliasing*.

Exemplo 2.1 Sabe-se que o ser humano é capaz de ouvir sons cujas freqüências variam entre 20 Hz e 20 kHz. Portanto, segundo o teorema da amostragem, para que todas as freqüências audíveis possam ser registradas, deve-se utilizar uma taxa de amostragem mínima de 40 kHz. Sinais de áudio são tipicamente amostrados à taxa de 44.1 kHz, um pouco superior ao limite de Nyquist para assegurar a recuperação de freqüências próximas do limite da audição. ■

A extensão do teorema de Whittaker-Shannon pode ser realizada para sinais n-dimensionais. Seja um sinal $f(x_1, x_2, \dots, x_n)$. O limite de Nyquist deve ser satisfeito considerando-se cada coordenada x_i , $i = 1, 2, \dots, n$, ou seja, existe um vetor $B = (B_1, B_2, \dots, B_n)$ de forma que

$$\Delta x_1 \leq \frac{1}{2B_1}, \dots, \Delta x_n \leq \frac{1}{2B_n} \quad (2.8)$$

Para o caso bidimensional, supondo um sinal $f(x, y)$ com banda limitada $2W_x$ e $2W_y$ nas direções x e y, respectivamente, o sinal pode ser completamente reconstruído se

$$\Delta x \leq \frac{1}{2W_x} \quad \text{e} \quad \Delta y \leq \frac{1}{2W_y} \quad (2.9)$$

É usual em processamento digital de imagens assumir que as dimensões da imagem e o número de níveis de cinza sejam potências inteiras de 2. No caso em que o número de níveis de cinza é igual a 2, a imagem é chamada *binária*. Imagens binárias possuem grande importância prática, pois ocupam menos espaço de armazenamento e podem ser manipuladas por meio de operadores lógicos que estão disponíveis diretamente nas instruções dos computadores.

Considerando que o processo de digitalização envolve parâmetros de amostragem e quantização, uma questão é saber quantas amostras $N \times M$ e níveis de cinza L são necessários para gerar uma boa imagem digital. Isso depende, fundamentalmente, da quantidade de informação contida na imagem e do grau de detalhes dessa informação que é perceptível ao olho humano. Tais parâmetros levam aos conceitos de resolução espacial e profundidade da imagem.

2.4 Resolução Espacial e Profundidade da Imagem

A *resolução espacial* está associada à densidade de pixels da imagem. Quanto menor o intervalo de amostragem entre os pixels da imagem, ou seja, quanto maior a densidade de pixels em uma imagem, maior será a resolução da imagem. É importante notar que uma imagem contendo um grande número de pixels não necessariamente possui resolução maior do que outra contendo menor número de pixels. A resolução de uma imagem deve ser escolhida de modo a atender ao grau de detalhes que devem ser discerníveis na imagem.

Seja, por exemplo, uma imagem $f(x, y)$ representando uma região de 400 cm^2 , consistindo em 20 amostras uniformemente espaçadas na direção x e 20 amostras uniformemente espaçadas na direção y . Cada pixel da imagem possui dimensão de $1 \text{ cm} \times 1 \text{ cm}$. Uma resolução maior para a mesma região poderia consistir em 40 amostras na direção x e 40 amostras na direção y , cada pixel agora correspondendo a $0.5 \text{ cm} \times 0.5 \text{ cm}$. Uma imagem de resolução menor poderia ter 10 amostras na direção x e 10 amostras na direção y , em que cada pixel corresponderia a $2 \text{ cm} \times 2 \text{ cm}$.

Como mencionado na seção 2.3, o número de níveis de quantização da imagem $f(x, y)$ é normalmente uma potência de 2, ou seja, $L = 2^b$, onde L é o número de níveis de cinza da imagem e b é chamado de *profundidade da imagem*. Assim, a profundidade de uma imagem corresponde ao número de bits necessários para armazenar a imagem digitalizada. Por exemplo, seja $L = 256$. Isso significa que cada pixel pode ter associado um valor de cinza entre 0 e 255. A profundidade da imagem nesse caso é de 8 bits por pixel.

A tabela 2.1 mostra o número de bytes empregados na representação de uma imagem monocromática para diferentes dimensões $M \times N$ pixels, com 2, 8, 32, 128 e 512 níveis de cinza.

As figuras 2.4 e 2.5 ilustram, respectivamente, os efeitos da redução da resolução espacial e dos níveis de cinza em uma imagem. As figuras 2.4(a)-(f) mostram os resultados da redução da resolução espacial de uma imagem em seis resoluções diferentes. Todas as imagens são apresentadas com as mesmas dimensões, ampliando-se o tamanho do pixel de forma a tornar mais evidente a perda de detalhes nas imagens de baixa resolução.

A figura 2.5(a) representa uma imagem de 512×512 pixels com 64 níveis de cinza ($b = 6$). As figuras 2.5(b)-(f) foram obtidas reduzindo-se o número de bits de $b = 5$ até $b = 1$ e mantendo as dimensões das imagens com 512×512 pixels.

Tabela 2.1: Número de bytes para imagens monocromáticas.

M	N	Número de bytes				
		L=2	L=8	L=32	L=128	L=512
320	256	4000	12000	20000	28000	36000
480	320	19200	57600	96000	134400	172800
640	400	32000	96000	160000	224000	288000
800	600	60000	180000	300000	420000	540000
1024	720	92160	276480	460800	645120	829440
1280	800	128000	384000	640000	896000	1152000

2.5 Representação de Imagens Digitais

Uma imagem digital pode ser representada por meio de uma matriz bidimensional, na qual cada elemento da matriz corresponde a um pixel da imagem. A figura 2.6 mostra a representação matricial de uma imagem. Uma pequena região da imagem é destacada, sendo formada por números inteiros correspondendo aos níveis de cinza dos pixels da imagem.

Há várias vantagens associadas ao uso de matrizes para representar imagens. Matrizes são estruturas simples para armazenar, manipular e visualizar dados. Uma desvantagem da matriz é sua inerente invariabilidade espacial, já que a estrutura não é adaptativa a eventuais irregularidades que possam existir na imagem. Isso pode produzir uma grande quantidade de redundância de dados. Métodos de compressão, abordados no capítulo 6, podem fornecer ganhos significativos em termos de espaço de armazenamento e tempo para transmissão de imagens.

Imagens podem ser representadas em múltiplas resoluções por meio de representações hierárquicas. Uma estrutura muito utilizada é a *pirâmide*. A representação piramidal de uma imagem com $N \times N$ pixels contém a imagem e k versões reduzidas da imagem. Normalmente, N é uma potência de 2 e as outras imagens possuem dimensões $N/2 \times N/2$, $N/4 \times N/4, \dots, 1 \times 1$. Nessa representação, o pixel no nível l é obtido pela combinação de informação de vários pixels na imagem no nível $l+1$. A imagem inteira é representada como um único pixel no nível superior, o nível 0, e o nível inferior é a imagem original (não reduzida). Um pixel em um nível representa informação agregada de vários pixels no nível seguinte.

A figura 2.7(a) ilustra uma seqüência de imagens representadas em diferentes resoluções. Uma imagem e suas versões reduzidas obtidas pela média dos valores de cinza em vizinhanças 2×2 pixels e dispostas em uma estrutura piramidal são mostradas na figura 2.7(b).

2.6 Imagem Multibanda ou Multiespectral

Em uma imagem digital monocromática, o valor do pixel é um escalar entre L_{\min} e L_{\max} . Imagens multibandas ou multiespectrais podem ser vistas como imagens nas quais cada pixel tem associado um valor vetorial $f(x, y) = (L_1, L_2, \dots, L_n)$, onde $L_{\min} \leq L_i \leq L_{\max}$ e $i = 1, 2, \dots, n$. Em geral, L_i pode representar grandezas diferentes, tais como temperatura, pressão ou freqüência, amostradas em pontos (x, y) e com intervalos de valores distintos. Uma imagem multiespectral também pode ser representada como uma seqüência de n ima-

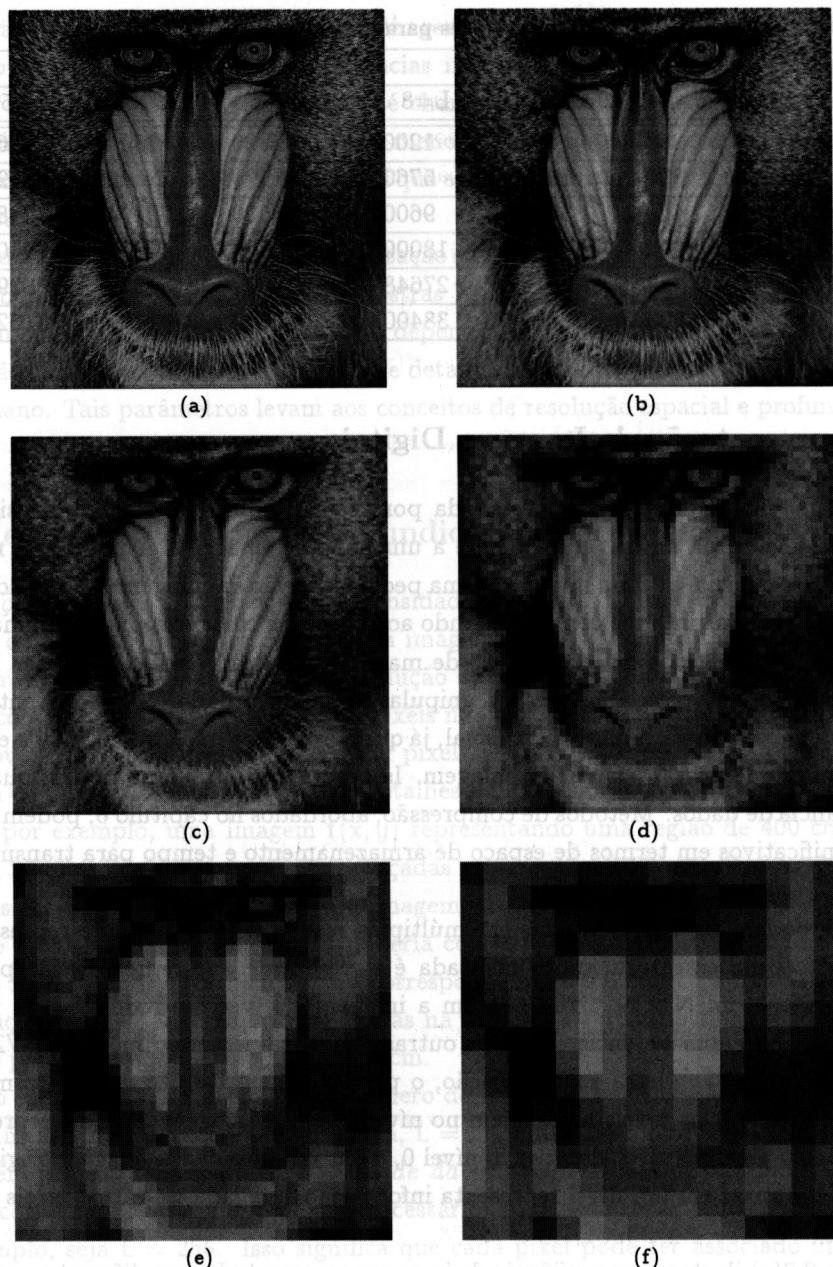
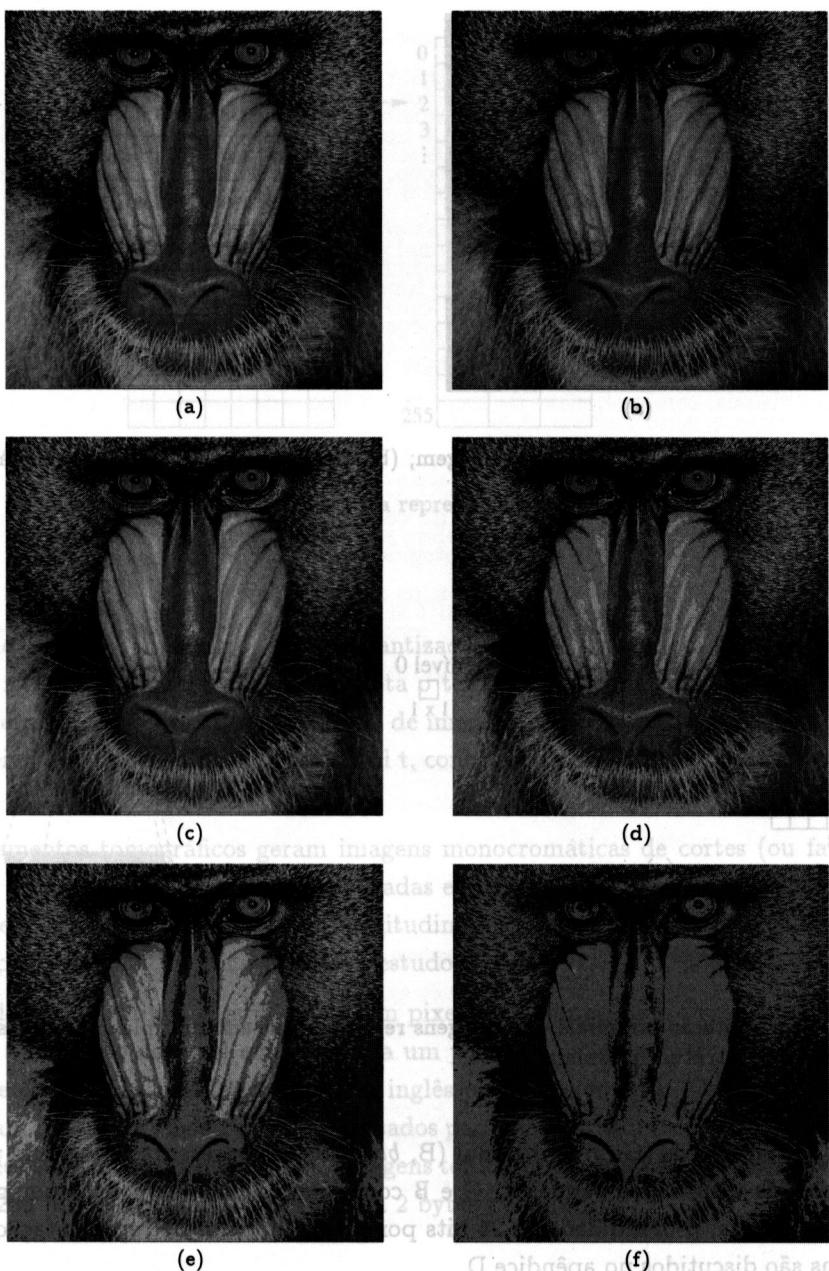


Figura 2.4: Efeitos de redução da resolução espacial (em pixels). (a) 512×512 ; (b) 256×256 ; (c) 128×128 ; (d) 64×64 ; (e) 32×32 ; (f) 16×16 . As imagens são apresentadas com as mesmas dimensões para ilustrar a perda de detalhes.

Imagens monocromáticas, tal que cada imagem é conhecida como *banda*, em que $f_i(x, y) = L_i$, $i = 1, 2, \dots, n$.

Imagens multiespectrais são muito utilizadas em sensoriamento remoto, no qual sensores operam em diferentes faixas do espectro eletromagnético, denominadas bandas espectrais. Dependendo do alvo, tal como vegetação, água ou solo, a interação da radiação eletro-



magnética produz menor ou maior resposta espectral, cujo valor está associado à posição espacial de um pixel da banda em particular.

Uma imagem colorida é uma imagem multibanda ou multiespectral, em que a cor em cada ponto (x, y) é definida por meio de três grandezas: luminância, matiz e saturação. A luminância está associada com o brilho da luz, o matiz com o comprimento de onda dominante e a saturação com o grau de pureza do matiz. A maioria das cores visíveis pelo olho humano pode ser representada como uma combinação de bandas das cores primárias

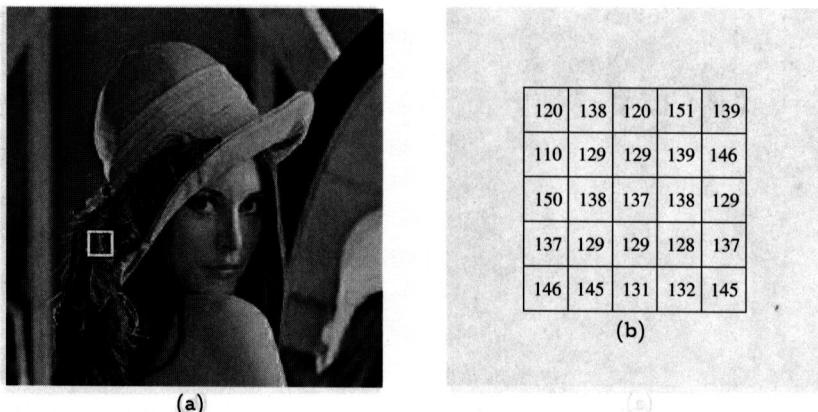


Figura 2.6: Representação matricial. (a) imagem; (b) níveis de cinza correspondentes à região da imagem em destaque.

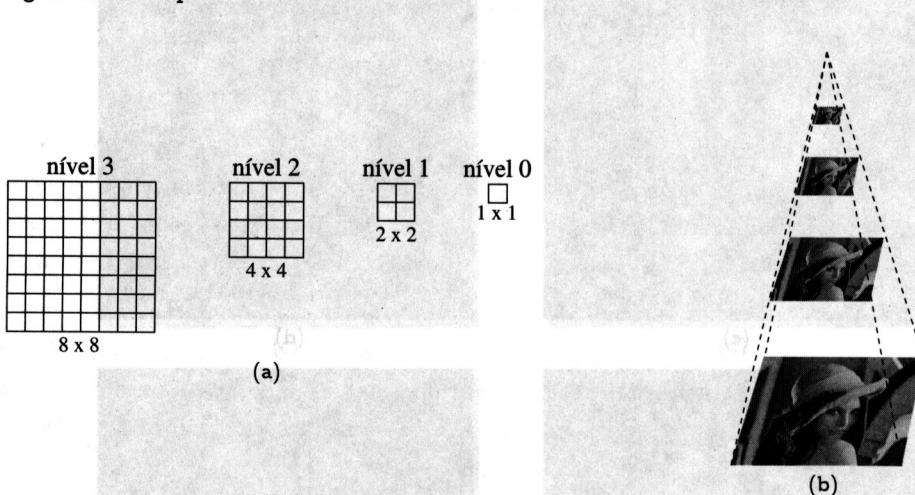


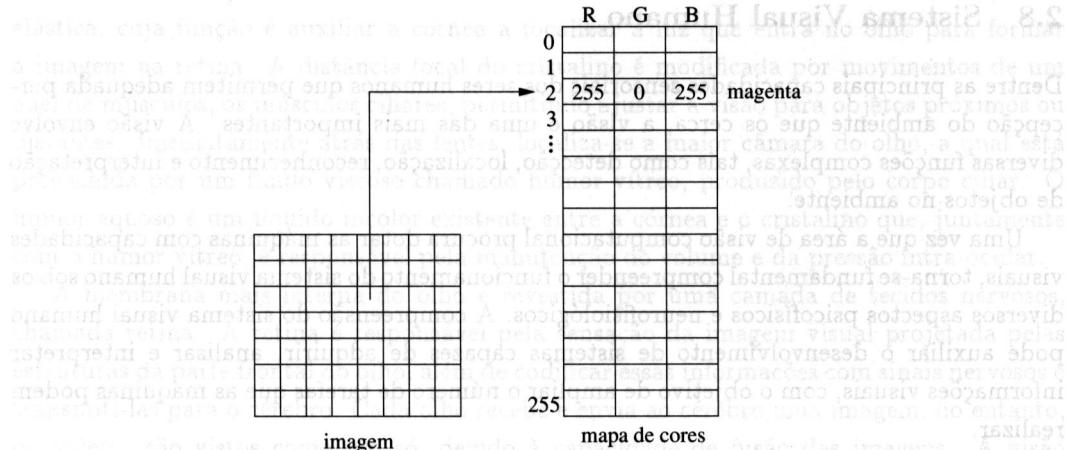
Figura 2.7: Representação piramidal. (a) imagens representadas em múltiplas resoluções; (b) imagens dispostas em estrutura piramidal.

vermelha (R, *red*), verde (G, *green*) e azul (B, *blue*). Uma representação comum para uma imagem colorida utiliza três bandas R, G e B com profundidade de 1 byte por pixel para cada banda, ou seja, profundidade de 24 bits por pixel. Modelos de representação de cores em imagens são discutidos no apêndice D.

Uma imagem colorida também pode ser armazenada por meio de uma imagem monocromática e um mapa de cores (figura 2.8). Nesse caso, o valor de cinza de cada pixel na imagem é um índice para uma entrada do mapa de cores, enquanto a entrada do mapa de cores contém o valor das componentes R, G e B referentes à cor do pixel. A quantidade de entradas do mapa de cores determina o número de cores utilizadas para representar a imagem.

2.7 Imagem Multidimensional

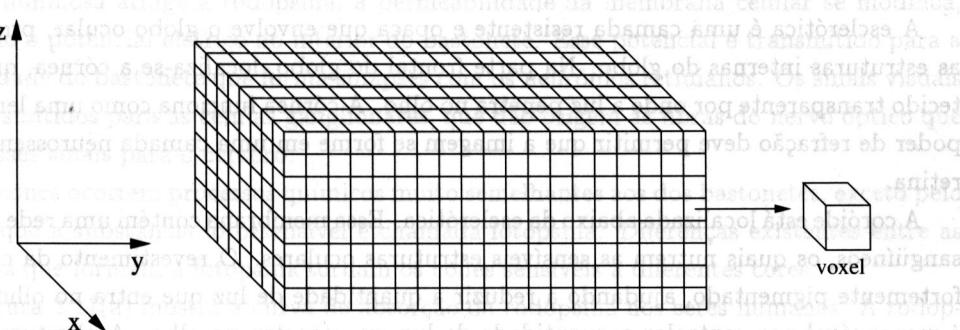
Há situações em que é necessária uma extensão dos conceitos de amostragem e quantização para uma terceira dimensão, a qual representa, em geral, o espaço ou o tempo. Isso é o



mesmo que dizer que a amostragem e a quantização podem ocorrer em (x, y, z) ou (x, y, t) , onde x, y, z representam o espaço e t denota o tempo. Portanto, uma imagem digital 3D pode ser representada como uma seqüência de imagens monocromáticas ou multibandas ao longo do eixo espacial z ou do eixo temporal t , conhecida como *imagem multidimensional*.

Equipamentos tomográficos geram imagens monocromáticas de cortes (ou fatias) normalmente paralelas e uniformemente espaçadas em uma dada região 3D. Na medicina, por exemplo, esses cortes são tipicamente longitudinais, cobrindo uma determinada região do corpo do paciente que contém o órgão em estudo.

Considerando as dimensões $p \times p$ de um pixel nessas imagens e o espaçamento d entre os cortes, a extensão do pixel em 3D forma um pequeno paralelepípedo de dimensões $p \times p \times d$, que é chamado *voxel* (acrônimo do inglês *volume element*). Os voxels (figura 2.9) representam pontos de amostragem e são usados para reconstruir no computador a forma ou a função de estruturas tridimensionais. Imagens tomográficas possuem tipicamente 512×512 ou 256×256 pixels e profundidade de 1 ou 2 bytes por pixel.



2.8 Sistema Visual Humano

Dentre as principais capacidades sensoriais dos seres humanos que permitem adequada percepção do ambiente que os cerca, a visão é uma das mais importantes. A visão envolve diversas funções complexas, tais como detecção, localização, reconhecimento e interpretação de objetos no ambiente.

Uma vez que a área de visão computacional procura dotar as máquinas com capacidades visuais, torna-se fundamental compreender o funcionamento do sistema visual humano sob os diversos aspectos psicofísicos e neurofisiológicos. A compreensão do sistema visual humano pode auxiliar o desenvolvimento de sistemas capazes de adquirir, analisar e interpretar informações visuais, com o objetivo de ampliar o número de tarefas que as máquinas podem realizar.

A figura 2.10 mostra um corte transversal do olho humano. O globo ocular possui formato aproximadamente esférico, com diâmetro de cerca de 20 mm, situado no interior de uma cavidade óssea, chamada órbita, sendo protegido pelas pálpebras e cílios. O globo ocular é aderido à órbita pelos músculos extrínsecos, os quais lhe dão capacidade de movimentação. O globo é envolvido por três membranas, uma camada externa formada pela esclerótica e pela córnea, uma camada intermediária formada pela coróide, íris e corpo ciliar, e uma camada interna formada pela retina.

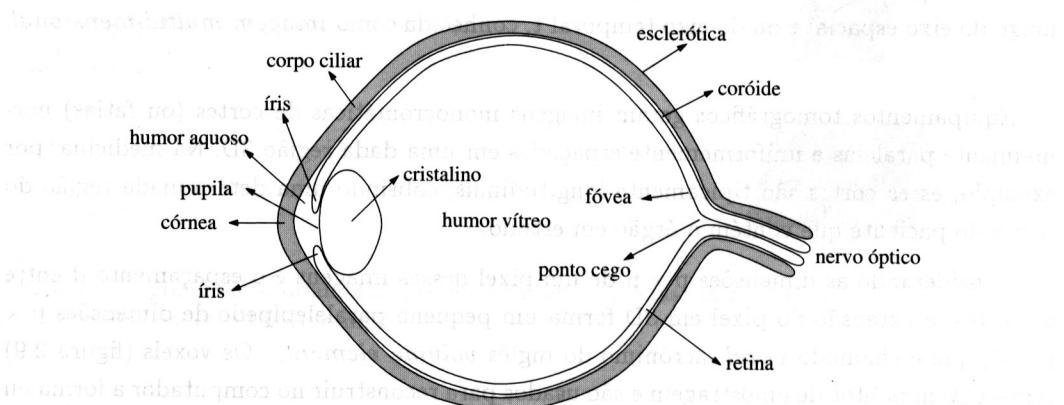


Figura 2.10: Sistema visual humano.

A esclerótica é uma camada resistente e opaca que envolve o globo ocular, protegendo as estruturas internas do globo. Na parte frontal do globo, localiza-se a córnea, que é um tecido transparente por onde a luz penetra no olho. A córnea funciona como uma lente, cujo poder de refração deve permitir que a imagem se forme em uma camada neurossensorial, a retina.

A coróide está localizada abaixo da esclerótica. Essa membrana contém uma rede de vasos sanguíneos, os quais nutrem as sensíveis estruturas oculares. O revestimento da coróide é fortemente pigmentado, ajudando a reduzir a quantidade de luz que entra no olho. A íris é responsável por controlar a quantidade de luz que penetra no olho. A abertura central da íris é conhecida como pupila, cujo diâmetro varia de 2 a 8 mm. A pupila expande ou contrai seu tamanho de acordo com a luminosidade do ambiente, regulando assim a entrada de luz no olho. Exatamente atrás da íris está o cristalino, que é uma lente gelatinosa e

elástica, cuja função é auxiliar a córnea a focalizar a luz que entra no olho para formar a imagem na retina. A distância focal do cristalino é modificada por movimentos de um anel de músculos, os músculos ciliares, permitindo ajustar a visão para objetos próximos ou distantes. Imediatamente atrás das lentes, localiza-se a maior câmara do olho, a qual está preenchida por um fluido viscoso chamado humor vítreo, produzido pelo corpo ciliar. O humor aquoso é um líquido incolor existente entre a córnea e o cristalino que, juntamente com o humor vítreo, é responsável pela manutenção do volume e da pressão intra-ocular.

A membrana mais interna do olho é revestida por uma camada de tecidos nervosos, chamada retina. A retina é responsável pela sensação da imagem visual projetada pelas estruturas da parte frontal do olho, além de codificar essas informações com sinais nervosos e transmiti-las para o cérebro. Cada olho recebe e envia ao cérebro uma imagem, no entanto, os objetos são vistos como um só, devido à capacidade de fusão das imagens. A visão binocular (com os dois olhos) proporciona maior campo visual e a noção de profundidade. O ponto cego, uma pequena região da retina onde está localizado o nervo óptico, não possui fotorreceptores.

A retina é composta por células sensíveis à luz, os cones e os bastonetes. Essas células transformam a energia luminosa das imagens em impulsos elétricos que são transmitidos ao cérebro pelo nervo óptico.

Os cones são altamente sensíveis à cor e responsáveis pela capacidade do olho em discernir detalhes nas imagens. Os cones são em número de 6 a 7 milhões em cada olho e estão localizados na porção central da retina, chamada mácula lútea. Ao centro da mácula lútea está a fóvea, uma região onde as células nervosas estão afastadas para o lado, permitindo que a luz atinja diretamente os receptores. Na fóvea, portanto, a acuidade visual é máxima. A visão pelos cones é chamada de *fotópica*.

O número de bastonetes é muito maior, cerca de 75 a 150 milhões, distribuídos na superfície periférica da retina. Os bastonetes são mais sensíveis à baixa intensidade de luz e permitem uma percepção geral da imagem captada no campo de visão. A visão pelos bastonetes é chamada de *escotópica*.

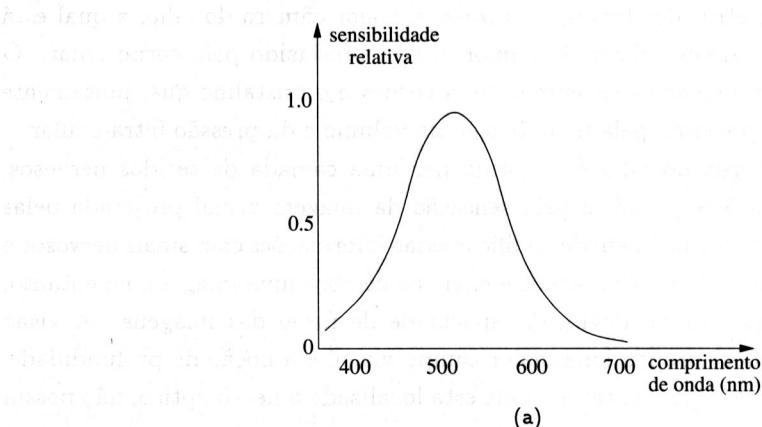
Alterações químicas ocorrem nos bastonetes e cones quando a luz atinge a retina. A vitamina A é o composto químico utilizado tanto pelos cones quanto pelos bastonetes para a síntese de substâncias fotossensíveis.

Nos bastonetes, a substância química fotossensível é chamada rodopsina. Quando a energia luminosa atinge a rodopsina, a permeabilidade da membrana celular se modifica, alterando o potencial elétrico no interior do bastonete. Esse potencial é transmitido para a extremidade do bastonete que forma sinapses com os neurônios retinianos. Os sinais visuais são transmitidos para as células ganglionares, que dão origem às fibras do nervo óptico que levam esses sinais para o cérebro.

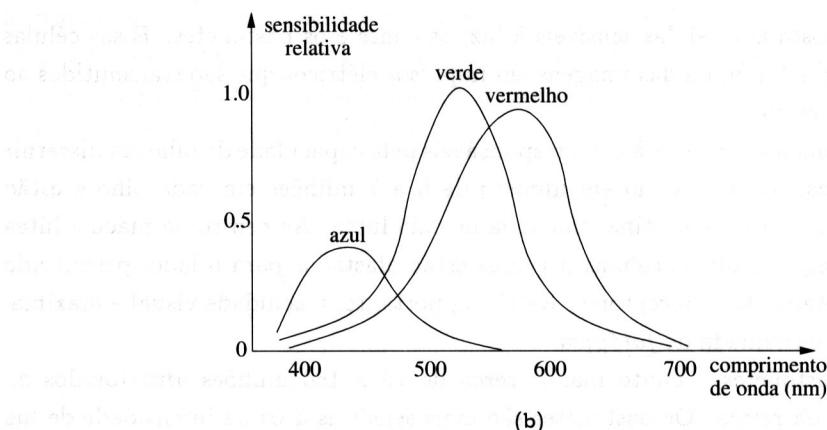
Nos cones ocorrem processos químicos muito semelhantes aos dos bastonetes, exceto pelo fato de que a substância fotossensível é chamada fotopsina. Diferenças existentes entre as proteínas que formam a fotopsina tornam os cones sensíveis a diferentes cores.

A figura 2.11(a) mostra a curva de absorção da rodopsina nos seres humanos. A rodopsina absorve fôtons mais eficientemente quando seus comprimentos de onda estão próximos a 500 nm. A retina contém três tipos de cones, cada um deles respondendo a um espectro cromático específico. A figura 2.11(b) ilustra esse fato, na qual são apresentados os comprimentos de onda da luz correspondentes aos três tipos de cones. A resposta apresenta três

valores máximos nos comprimentos de onda 430, 575 e 600 nm, correspondendo às cores azul, verde-amarela e vermelha, respectivamente. Mais detalhes sobre os fundamentos de cor são apresentados no apêndice D.



(a)



(b)

Figura 2.11: Sensibilidade espectral dos receptores. (a) resposta espectral dos bastonetes; (b) resposta espectral dos cones.

Várias semelhanças podem ser destacadas entre o sistema visual humano e um sistema de sensores, tal como uma câmera fotográfica. O obturador da câmera possui função similar à da pálpebra do olho. O diafragma de uma câmera controla a quantidade de luz que atravessa as lentes, similar à iris no olho humano.

As lentes da câmera são análogas ao conjunto formado pelo cristalino e córnea, cujo objetivo comum é focalizar a luz para tornar nítidas as imagens que serão formadas em uma superfície sensível. No olho humano, esta superfície sensível é a retina. Nas câmeras fotográficas, películas fotossensíveis ou filmes fotográficos são utilizados para registrar as imagens em câmeras analógicas, enquanto cartões de memória são utilizados em câmeras digitais.

2.9 Ruído em Imagens

Imagens reais freqüentemente sofrem degradações durante seu processo de aquisição, transmissão ou processamento. Essa degradação é normalmente chamada de *ruído*. O ruído

pode ser considerado uma variável aleatória z , caracterizada por uma função densidade de probabilidade $p(z)$. Os tipos de ruído mais comumente modelados são o ruído *impulsivo*, *Gaussiano*, *uniforme*, *Erlang*, *exponencial*, *Rayleigh* e *Poisson*.

O ruído impulsivo é caracterizado pela ocorrência aleatória de pixels cujos valores de luminosidade diferem significativamente dos valores de seus pixels vizinhos. Um tipo de ruído impulsivo em que a imagem é degradada pela ocorrência de pixels brancos e pretos é conhecido como ruído *sal-e-pimenta*. A função densidade de probabilidade do ruído impulsivo é dada por

$$p(z) = \begin{cases} P_a, & \text{para } z = a \\ P_b, & \text{para } z = b \\ 0, & \text{caso contrário} \end{cases} \quad (2.10)$$

O ruído Gaussiano é caracterizado pela ocorrência de pixels com valores de intensidade que variam conforme a distribuição Gaussiana. Esse ruído é uma boa aproximação da degradação que ocorre em muitas aplicações práticas, sendo utilizado, por exemplo, para modelar ruído gerado por componentes eletrônicos de um sistema de aquisição digital de imagens. Uma variável aleatória com distribuição Gaussiana possui sua densidade de probabilidade dada pela curva Gaussiana. No caso unidimensional, a função densidade de probabilidade é

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (2.11)$$

em que μ é a média e σ^2 é a variância da variável aleatória z .

O ruído uniforme segue a função densidade de probabilidade dada por

$$p(z) = \begin{cases} \frac{1}{b-a}, & \text{para } a \leq z \leq b \\ 0, & \text{caso contrário} \end{cases} \quad (2.12)$$

A média μ e a variância σ^2 da função de probabilidade uniforme são definidas como

$$\mu = \frac{a+b}{2} \quad \sigma^2 = \frac{(b-a)^2}{12} \quad (2.13)$$

A função densidade de probabilidade do ruído Erlang (ou gama) é dada por

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} e^{-az}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases} \quad (2.14)$$

em que $a > 0$ e b é um valor inteiro positivo. A média μ e a variância σ^2 da função de probabilidade de Erlang são definidas como

$$\mu = \frac{b}{a} \quad \sigma^2 = \frac{b}{a^2} \quad (2.15)$$

O ruído exponencial segue a função densidade de probabilidade dada por

$$p(z) = \begin{cases} a e^{-az}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases} \quad (2.16)$$

em que $a > 0$. A média μ e a variância σ^2 da função de probabilidade exponencial são definidas como

$$\mu = \frac{1}{a} \quad \sigma^2 = \frac{1}{a^2} \quad (2.17)$$

O ruído exponencial segue a função densidade de probabilidade do ruído Erlang para o caso particular em que $b = 1$.

A função densidade de probabilidade do ruído Rayleigh é dada por

$$(2.18) \quad p(z) = \begin{cases} \frac{2}{b}(z-a)e^{-(z-a)^2/b}, & \text{para } z \geq a \\ 0, & \text{para } z < a \end{cases}$$

A média μ e a variância σ^2 da função densidade de Rayleigh são definidas como

$$\mu = a + \sqrt{\pi b/4} \quad \sigma^2 = \frac{b(4-\pi)}{4} \quad (2.19)$$

O ruído Poisson segue a função densidade de probabilidade dada por

$$(2.20) \quad p(z) = \begin{cases} \frac{e^{-\mu} \mu^z}{z!}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases}$$

tal que a média da função de probabilidade é dada por μ e variância $\sigma^2 = \mu$.

Exemplos de imagens corrompidas por ruído impulsivo, Gaussiano e Poisson são ilustrados na figura 2.12. Técnicas para atenuação de ruído em imagens são discutidas na seção 4.4.

2.10 Entropia em Imagens

O conceito de *entropia* ou *incerteza* foi introduzido por Shannon (1948) para medir a quantidade de informação transferida por um canal ou gerada por uma fonte. Quanto maior for o valor de entropia, mais incerteza e, portanto, mais informação estará associada ao canal.

O princípio fundamental da teoria de informação estabelece que a geração de informação pode ser modelada como um processo probabilístico. Uma imagem pode ser considerada como o resultado de um processo aleatório, no qual a probabilidade p_i corresponde à probabilidade de um pixel em uma imagem digital assumir um valor de intensidade i , $i = 0, 1, \dots, L_{\max}$. A distribuição dos níveis de intensidade da imagem pode ser transformada em uma função densidade de probabilidade, dividindo-se o número de pixels de intensidade i , denotado n_i , pelo número total n de pixels na imagem, ou seja

$$(2.21) \quad p_i = \frac{n_i}{n}$$

em que $\sum_{i=0}^{L_{\max}} p_i = 1$.

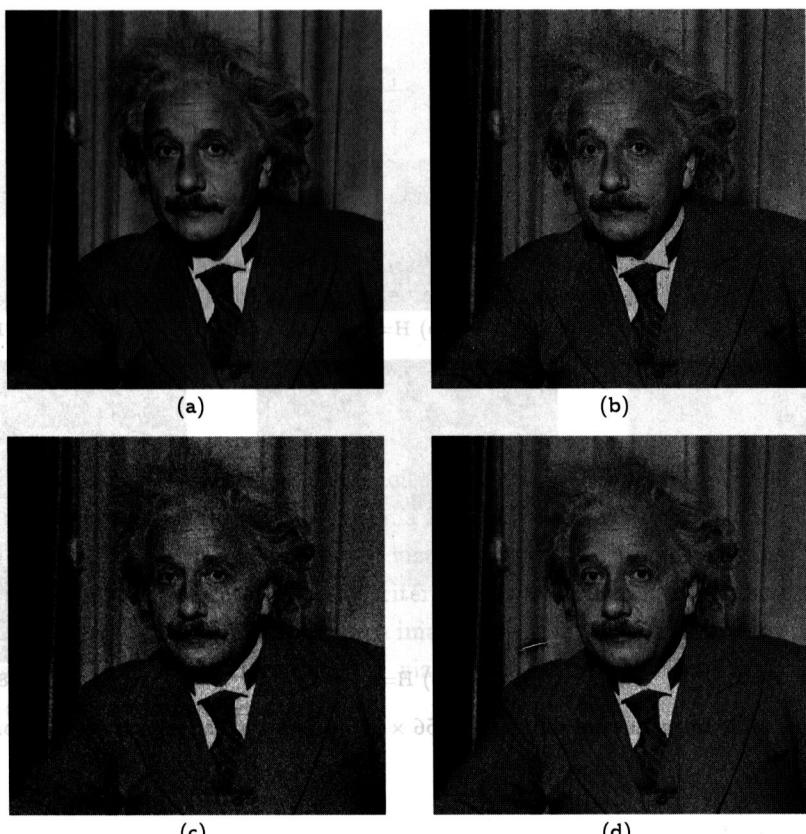


Figura 2.12: Ruído em imagens. (a) imagem original; (b) imagem com ruído impulsivo sal-e-pimenta; (c) imagem com ruído Gaussiano; (d) imagem com ruído Poisson.

A entropia H de uma imagem pode ser calculada como

$$H = - \sum_{i=0}^{L_{\max}} p_i \log p_i \quad (2.22)$$

A entropia de uma imagem é uma medida positiva e, quando a base do logaritmo for dois, a unidade resultante é dada em bits. O menor valor para a entropia é zero, ocorrendo quando todos os pixels possuem uma mesma intensidade de cinza. Por outro lado, a máxima entropia ocorre quando uma imagem contém a mesma quantidade de pixels para todas as intensidades.

A figura 2.13 mostra exemplos de imagens com dimensões 256×256 pixels apresentando diferentes valores de entropia. Valores mínimo e máximo de entropia são ilustrados nas figuras 2.13(a) e (f), respectivamente. A imagem da figura 2.13(f) possui todos os 256 níveis de cinza possíveis distribuídos com a mesma quantidade de pixels. Valores intermediários de entropia são mostrados nas figuras 2.13(b) a (e).

A entropia não está relacionada com a disposição espacial da informação, conforme ilustrado nas figuras 2.13(b) e (c), em que as duas imagens possuem igual quantidade de pixels com as mesmas intensidades, porém, distribuídos espacialmente de maneira diferente.

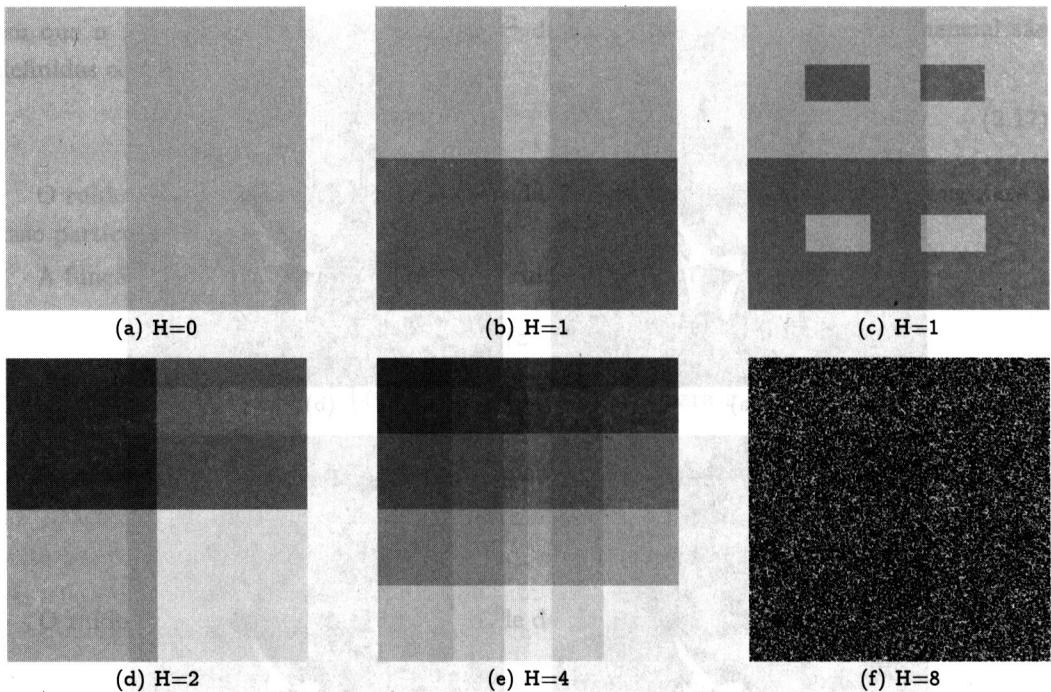


Figura 2.13: Imagens com dimensões 256×256 pixels e suas respectivas entropias.

2.11 Relacionamentos Básicos entre Elementos de Imagem

Esta seção apresenta relacionamentos básicos entre elementos de uma imagem. Um elemento f em uma matriz bidimensional é denotado pelo pixel $f(x, y)$, enquanto em uma matriz tridimensional é denotado pelo voxel $f(x, y, z)$.

2.11.1 Vizinhança

Um pixel f de coordenadas (x, y) possui quatro vizinhos horizontais e verticais, cujas coordenadas são $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$ e $(x, y - 1)$. Esses pixels formam a chamada *vizinhança-4* de f , denotada $N_4(f)$ e ilustrada na figura 2.14(a). Os quatro vizinhos diagonais de f são os pixels de coordenadas $(x - 1, y - 1)$, $(x - 1, y + 1)$, $(x + 1, y - 1)$ e $(x + 1, y + 1)$, que constituem o conjunto $N_d(f)$.

A *vizinhança-8* de f , denotada $N_8(f)$ e ilustrada na figura 2.14(b), é definida como

$$N_8(f) = N_4(f) \cup N_d(f) \quad (2.23)$$

O conceito de vizinhança pode ser estendido para imagens tridimensionais, sendo que vizinhos podem ser definidos de acordo com o número de voxels compartilhando faces, arestas ou vértices em comum. Essas definições resultam nas vizinhanças 6, 18 e 26, respectivamente, mostradas nas figuras 2.14(c)-(e). A figura 2.14(c), por exemplo, ilustra o caso de vizinhança-6, em que os vizinhos do voxel (x, y, z) são $(x - 1, y, z)$, $(x + 1, y, z)$, $(x, y - 1, z)$, $(x, y + 1, z)$, $(x, y, z - 1)$ e $(x, y, z + 1)$.

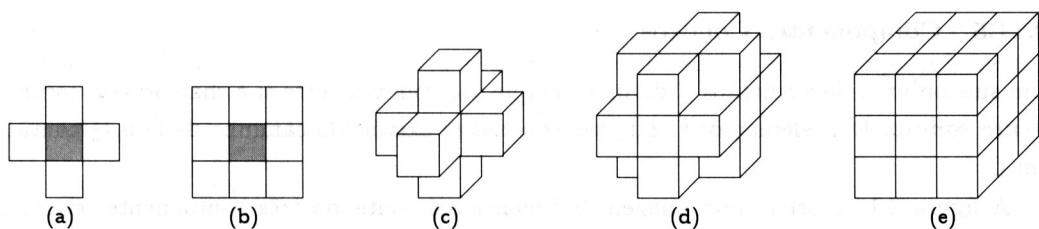


Figura 2.14: Tipos de vizinhança. (a) vizinhança-4 de um pixel; (b) vizinhança-8 de um pixel; (c) vizinhança-6 de um voxel; (d) vizinhança-18 de um voxel; (e) vizinhança-26 de um voxel.

2.11.2 Conectividade

A *conectividade* entre elementos é um conceito importante utilizado para estabelecer limites de objetos e componentes de regiões em uma imagem. Para verificar se dois elementos são conexos é necessário determinar se eles são vizinhos segundo o tipo de vizinhança adotado e se os elementos satisfazem determinados critérios de similaridade, tais como intensidade de cinza, cor ou textura. Por exemplo, em uma imagem binária, em que os pixels podem assumir os valores 0 ou 1, dois pixels podem ter vizinhança-4, mas somente serão considerados conexos se possuírem o mesmo valor.

2.11.3 Adjacência

Um elemento f_1 é *adjacente* a um elemento f_2 se eles forem conexos de acordo com o tipo de vizinhança adotado. Dois subconjuntos de pixels, S_1 e S_2 , são adjacentes se pelo menos um elemento em S_1 for adjacente a algum elemento em S_2 .

2.11.4 Caminho

Um *caminho* na imagem do pixel (x_1, y_1) a um pixel (x_n, y_n) é uma seqüência de pixels distintos com coordenadas $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, onde n é o comprimento do caminho, e (x_i, y_i) e (x_{i+1}, y_{i+1}) são adjacentes, tal que $i = 1, 2, \dots, n - 1$.

Se a relação de conectividade considerar vizinhança-4, então existe um caminho-4; para vizinhança-8, tem-se um caminho-8. Exemplos de caminhos são mostrados na figura 2.15, sendo que o caminho-4 possui comprimento 10 e o caminho-8 possui comprimento 7. O conceito de caminho também pode ser estendido para imagens tridimensionais.

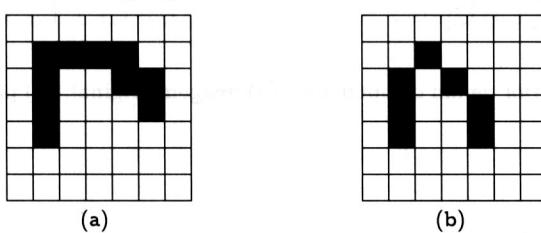


Figura 2.15: Exemplos de caminhos. (a) caminho-4; (b) caminho-8.

2.11.5 Componentes Conexos

Um subconjunto de elementos C da imagem que são conexos entre si é chamado de *componente conexo*. Dois elementos f_1 e f_2 são conexos se existir um caminho de f_1 a f_2 contido em C.

A figura 2.16 mostra uma imagem bidimensional contendo três componentes conexos caso seja considerada a vizinhança-4 ou, então, dois componentes conexos se considerada a vizinhança-8.

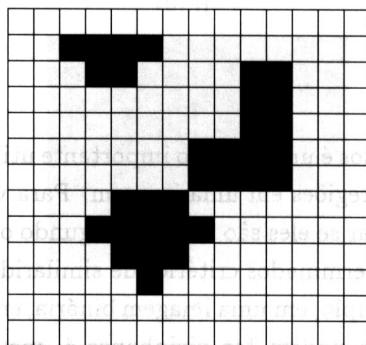


Figura 2.16: Componentes conexos em uma imagem bidimensional.

2.11.6 Borda e Interior

A *borda* de um componente conexo S em uma imagem bidimensional é o conjunto de pixels pertencentes ao componente e que possuem vizinhança-4 com um ou mais pixels externos a S. Intuitivamente, a borda corresponde ao conjunto de pontos no contorno do componente conexo. O *interior* é o conjunto de pixels de S que não estão em sua borda. A figura 2.17 mostra um exemplo de uma imagem binária com sua borda e interior.

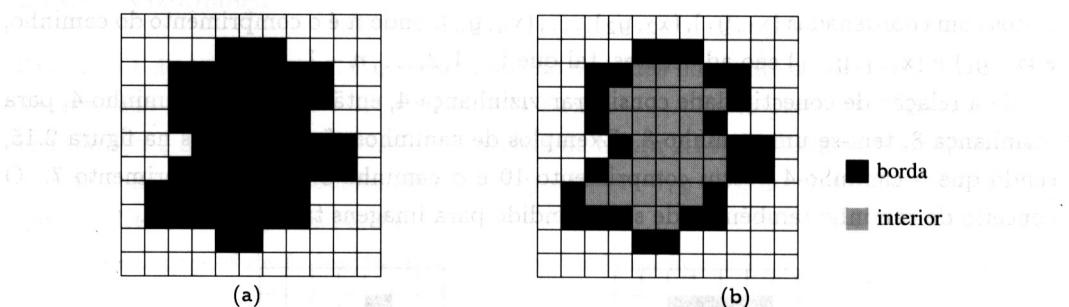


Figura 2.17: Borda e interior de um componente. (a) imagem original; (b) pixels da borda e interior.

2.11.7 Medidas de Distância

Muitas aplicações requerem o cálculo da distância entre dois pixels ou dois componentes de uma imagem. Entretanto, não há uma única forma para se definir distância em imagens

digitais. Dados os pixels f_1 , f_2 e f_3 , com coordenadas (x_1, y_1) , (x_2, y_2) e (x_3, y_3) , respectivamente, qualquer métrica de distância D deve satisfazer todas as seguintes propriedades:

$$(i) \quad D(f_1, f_2) \geq 0 \quad (D(f_1, f_2) = 0 \text{ se, e somente se, } f_1 = f_2)$$

$$(ii) \quad D(f_1, f_2) = D(f_2, f_1)$$

$$(iii) \quad D(f_1, f_3) \leq D(f_1, f_2) + D(f_2, f_3)$$

Existem diversas funções de distância que satisfazem as propriedades acima. Algumas medidas freqüentemente utilizadas em análise de imagens são descritas a seguir.

A *distância Euclidiana* entre f_1 e f_2 é definida como

$$D_E(f_1, f_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2.24)$$

Na distância Euclidiana, os pixels com uma distância menor ou igual a algum valor d formam um disco de raio d centrado em f_1 . Por exemplo, os pontos com distância $D_E \leq 3$ de um ponto central (x, y) formam o conjunto mostrado na figura 2.18.

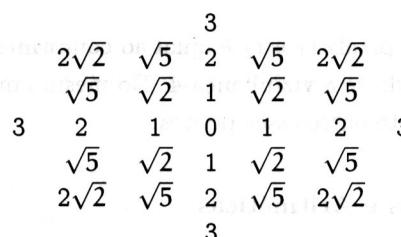


Figura 2.18: Conjunto de pontos com distância Euclidiana menor ou igual a 3 de um ponto central.

A distância Euclidiana está mais próxima do caso contínuo, entretanto, requer mais esforço computacional e pode produzir valores fracionários.

A *distância D_4* entre f_1 e f_2 , também denominada *city-block*, é definida como

$$D_4(f_1, f_2) = |x_1 - x_2| + |y_1 - y_2| \quad (2.25)$$

Os pixels com uma distância D_4 de f_1 menor ou igual a algum valor d formam um losango centrado em f_1 . Em particular, os pontos com distância 1 são os pixels com vizinhança-4 do ponto central. Por exemplo, os pontos com distância $D_4 \leq 3$ de um ponto central (x, y) formam o conjunto mostrado na figura 2.19.

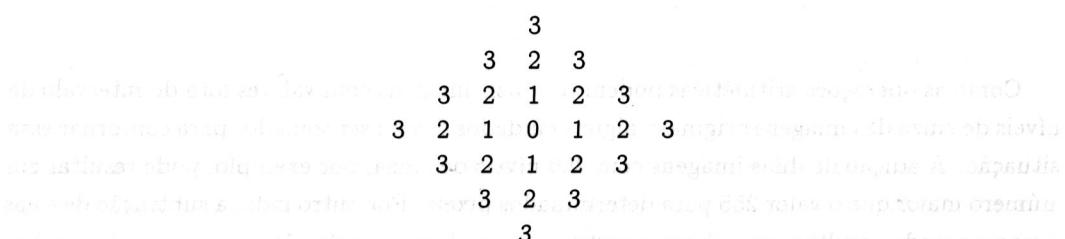


Figura 2.19: Conjunto de pontos com distância D_4 menor ou igual a 3 de um ponto central.

A distância D_8 entre f_1 e f_2 , também denominada *chessboard*, é definida como

$$D_8(f_1, f_2) = \max(|x_1 - x_2|, |y_1 - y_2|) \quad (2.26)$$

Os pixels com uma distância D_8 de f_1 menor ou igual a algum valor d formam um quadrado centrado em f_1 . Os pontos com distância 1 são os pixels com vizinhança-8 do ponto central. Por exemplo, os pontos com distância $D_8 \leq 3$ de um ponto central (x, y) formam o conjunto mostrado na figura 2.20.

3	3	3	3	3	3	3	3
3	2	2	2	2	2	2	3
3	2	1	1	1	2	3	
3	2	1	0	1	2	3	
3	2	1	1	1	2	3	
3	2	2	2	2	2	3	
3	3	3	3	3	3	3	3

Figura 2.20: Conjunto de pontos com distância D_8 menor ou igual a 3 de um ponto central.

A distância D_4 entre dois pixels f_1 e f_2 é igual ao comprimento do caminho mais curto entre esses pixels, considerando-se a vizinhança-4. Do mesmo modo, a distância D_8 corresponde ao caminho-8 mais curto entre esses pontos.

2.11.8 Operações Lógicas e Aritméticas

Operações lógicas e aritméticas podem ser utilizadas para modificar imagens. Embora essas operações permitam uma forma simples de processamento, há uma grande variedade de aplicações em que tais operações podem produzir resultados de interesse prático.

Dadas duas imagens, f_1 e f_2 , as operações aritméticas mais comuns entre dois pixels $f_1(x, y)$ e $f_2(x, y)$ são a adição, subtração, multiplicação e divisão, definidas de acordo com a tabela 2.2.

Tabela 2.2: Operações aritméticas.

Adição	$f_1(x, y) + f_2(x, y)$
Subtração	$f_1(x, y) - f_2(x, y)$
Multiplicação	$f_1(x, y).f_2(x, y)$
Divisão	$f_1(x, y)/f_2(x, y)$

Como as operações aritméticas podem produzir imagens com valores fora do intervalo de níveis de cinza das imagens originais, alguns cuidados devem ser tomados para contornar essa situação. A adição de duas imagens com 256 níveis de cinza, por exemplo, pode resultar em número maior que o valor 255 para determinados pixels. Por outro lado, a subtração de duas imagens pode resultar em valores negativos para alguns pixels. Uma maneira de resolver esse problema é, após a aplicação do operador aritmético, realizar uma transformação da escala de cinza na imagem resultante para manter seus valores dentro do intervalo adequado.

A divisão de imagens pode produzir valores fracionários, os quais devem ser convertidos para valores inteiros. Além disso, divisão por zero deve ser evitada. Uma maneira simples de evitar esse problema é adicionar o valor 1 a todos os valores de intensidade dos pixels, tal que o intervalo de níveis de cinza passa a ser interpretado de 1 a 256, ao invés de 0 a 255.

A adição de imagens pode ser utilizada para sobrepor o conteúdo de uma imagem em outra. A figura 2.21 mostra a combinação de uma imagem contendo três objetos com o mapa de bordas extraído a partir dos objetos. Uma outra aplicação do uso da adição é a remoção de ruídos pelo cálculo da média das imagens (mais detalhes sobre filtragem de imagens são apresentados na seção 4.4).

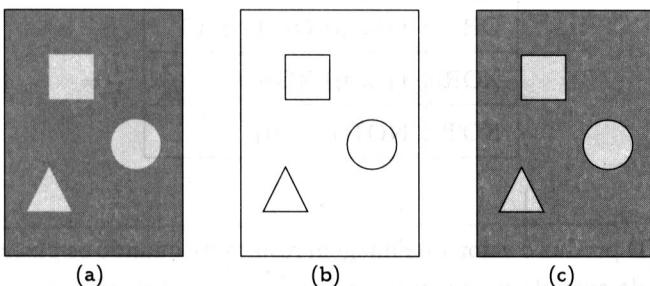


Figura 2.21: Adição de imagens. (a) imagem original contendo três objetos; (b) mapa de bordas da imagem; (c) sobreposição do mapa de bordas à imagem original.

A subtração de imagens possui vários usos interessantes, sendo uma maneira de identificar diferenças entre imagens. A figura 2.22 mostra duas imagens e o resultado da subtração entre elas. As regiões dos objetos que permanecem inalteradas nas imagens são eliminadas. Os objetos que se moveram são claramente mostrados. O movimento de objetos pode ser medido utilizando-se subtração de imagens, caso as imagens sejam adquiridas em taxas suficientemente rápidas. O comprimento da região que sofreu alteração dividido pelo tempo decorrido resulta na velocidade; a direção pode ser determinada pela orientação da região.

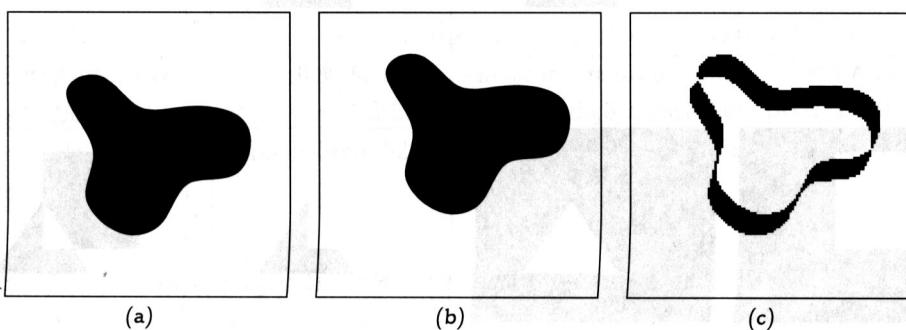


Figura 2.22: Subtração de imagens. (a) imagem f_1 ; (b) imagem f_2 ; (c) diferença $|f_1 - f_2|$.

Uma das principais aplicações da multiplicação ou divisão de imagens é no ajuste de iluminação, eventualmente necessário para corrigir problemas que possam surgir durante o processo de aquisição de imagens. Outras utilizações desses operadores incluem a filtragem de domínio de frequência (mais detalhes na seção 4.4.2) e na modelagem de ruído.

As principais operações lógicas utilizadas em processamento de imagens são mostradas na tabela 2.3. Essas operações podem ser combinadas para formar expressões lógicas mais complexas. Operações lógicas podem ser aplicadas apenas a imagens binárias, enquanto operações aritméticas podem ser usadas em pixels com valores diversos. A terminologia adotada é que pixels com valores iguais a 0 (preto) pertencem aos objetos e pixels com valores iguais a 1 (branco) correspondem ao fundo.

Tabela 2.3: Operações lógicas.

AND	$f_1(x, y) \text{ AND } f_2(x, y)$
OR	$f_1(x, y) \text{ OR } f_2(x, y)$
XOR	$f_1(x, y) \text{ XOR } f_2(x, y)$
NOT	$\text{NOT}(f_1(x, y))$

A operação AND produz o valor 1 na imagem resultante quando os pixels correspondentes nas duas imagens de entrada possuem valor igual a 1. A operação XOR produz 1 quando apenas um dos pixels (mas não ambos) possui valor 1, caso contrário, produz 0. O resultado da operação OR é 1 quando pelo menos um dos pixels das imagens é igual a 1. A operação NOT inverte o valor do pixel na imagem.

As operações lógicas podem ser utilizadas para combinar informação entre as imagens ou extraír regiões de interesse. Alguns exemplos de aplicação de operadores lógicos são mostrados na figura 2.23.

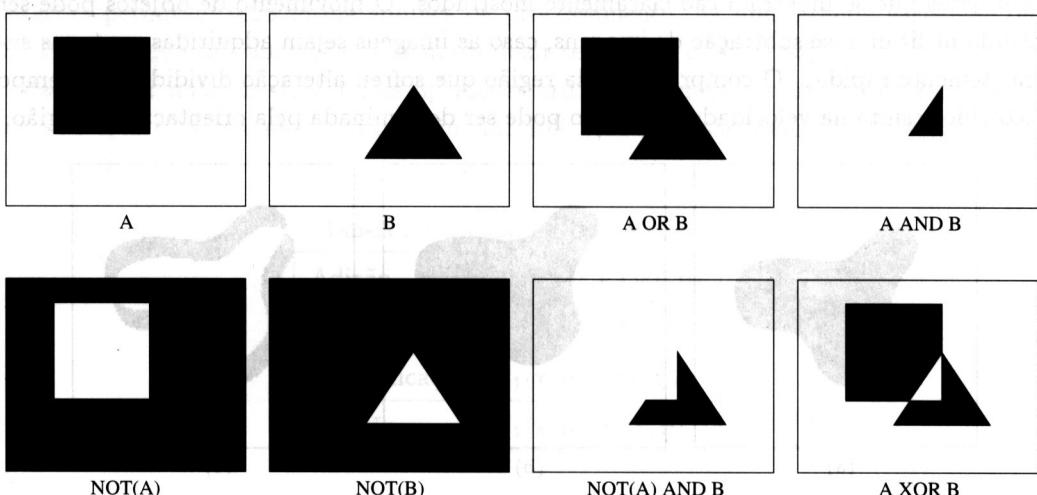


Figura 2.23: Alguns exemplos de operações lógicas em imagens binárias.

Além de processamentos pixel a pixel, as operações lógicas e aritméticas podem ser utilizadas em processamentos orientados à vizinhança. Tipicamente, esse tipo de processamento utiliza as operações com máscaras, em que os termos janela e filtros são freqüentemente utilizados como sinônimos de máscara. A idéia principal na utilização de máscaras é fazer

com que o valor atribuído a um pixel seja uma função do seu próprio valor e dos valores de seus vizinhos.

Para a região da imagem mostrada na figura 2.24, por exemplo, se o valor de $f(x, y)$ fosse substituído pela média $\bar{f}(x, y)$ dos valores dos pixels adjacentes em uma vizinhança de 3×3 pixels, seria necessário realizar uma operação aritmética de acordo com a equação 2.27.

$$\begin{aligned}\bar{f}(x, y) &= \frac{1}{9} \left[f(x-1, y-1) + f(x, y-1) + f(x+1, y-1) + \right. \\ &\quad f(x-1, y) + f(x, y) + f(x+1, y) + \\ &\quad \left. f(x-1, y+1) + f(x, y+1) + f(x+1, y+1) \right] = \\ &= \frac{1}{9} \sum_{m=-1}^1 \sum_{n=-1}^1 f(x+m, y+n)\end{aligned}\quad (2.27)$$

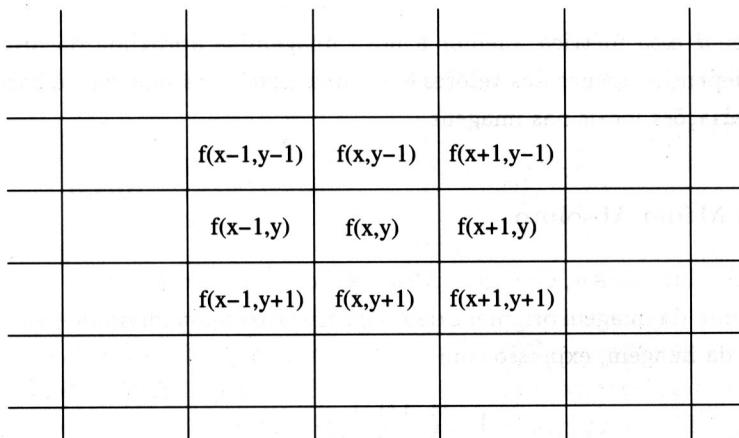


Figura 2.24: Região de uma imagem.

A aplicação de uma máscara em cada pixel da imagem é uma tarefa de alto custo computacional. Por exemplo, a aplicação de uma máscara de dimensões 3×3 em uma imagem de 256×256 pixels requer nove multiplicações e oito adições para cada pixel, resultando em um total de 589 824 multiplicações e 524 288 adições (desconsiderando efeitos de borda da imagem).

2.12 Métricas de Qualidade em Imagens

A qualidade de uma imagem é bastante dependente da aplicação na qual a imagem é utilizada. Imagens podem ser utilizadas, por exemplo, para entretenimento em jogos eletrônicos ou em televisão, assim como em aplicações que requerem grande precisão, tais como na medicina ou automação industrial. Nesses exemplos, os tipos e graus de degradação que uma imagem pode sofrer, em geral, são bem distintos.

Uma imagem pode sofrer degradações durante o processo de aquisição, transmissão ou processamento. Métricas de qualidade ou fidelidade podem ser utilizadas para avaliar a similaridade de uma imagem transformada g em relação à original f .

Algumas medidas são voltadas a avaliações subjetivas, as quais se baseiam em análises realizadas por observadores humanos. Por outro lado, avaliações objetivas procuram medir a qualidade das imagens por meio de funções entre a imagem original e a imagem transformada.

Os métodos de avaliação objetiva mais comuns são baseados em medidas de similaridade ou diferenças entre as imagens. Considerando duas imagens, f e g , ambas com dimensões $M \times N$ pixels, algumas métricas para medir a similaridade entre as imagens são listadas a seguir.

2.12.1 Erro Máximo

O erro máximo (ME, do inglês, *Maximum Error*) é a maior diferença absoluta entre cada par de pontos na imagem original e na imagem aproximada, expresso como

$$ME = \max |f(x, y) - g(x, y)| \quad (2.28)$$

Quanto menor essa métrica, melhor a nova imagem se aproxima da original. O erro máximo, por depender apenas dos valores entre dois pixels nas imagens, é bastante sensível a ruído ou a variações locais nas imagens.

2.12.2 Erro Médio Absoluto

O erro médio absoluto (MAE, do inglês, *Mean Absolute Error*) é a soma da diferença absoluta de cada ponto da imagem original e da imagem aproximada, dividido pela multiplicação das dimensões da imagem, expresso como

$$MAE = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |f(x, y) - g(x, y)| \quad (2.29)$$

Quanto menor essa métrica, melhor a nova imagem se aproxima da original.

2.12.3 Erro Médio Quadrático

O erro médio quadrático (MSE, do inglês, *Mean Square Error*) é a soma do quadrado das diferenças de cada ponto da imagem original e da imagem aproximada, dividido pela multiplicação das dimensões da imagem, expresso como

$$MSE = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2 \quad (2.30)$$

Quanto menor essa métrica, melhor a nova imagem se aproxima da original. Uma variação muito utilizada dessa métrica é conhecida como raiz do erro médio quadrático (RMSE, do inglês, *Root Mean Square Error*), expressa como

$$RMSE = \sqrt{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2} \quad (2.31)$$

Outra variação dessa métrica é o erro médio quadrático normalizado (NMSE, do inglês, *Normalized Mean Square Error*), definido como

$$\text{NMSE} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y)]^2} \quad (2.32)$$

cujos valores variam entre 0 e 1, o que torna possível a avaliação de imagens com dimensões diferentes.

2.12.4 Relação Sinal-Ruído de Pico

A relação sinal-ruído de pico (PSNR, do inglês, *Peak Signal to Noise Ratio*) é utilizada para avaliar a diferença global entre duas imagens, expressa como

$$\text{PSNR} = 20 \log_{10} \frac{L_{\max}}{\text{RMSE}} = 10 \log_{10} \frac{(L_{\max})^2}{\text{MSE}} = 10 \log_{10} \frac{MNL_{\max}^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2} \quad (2.33)$$

sendo L_{\max} o valor máximo de intensidade de cinza. Tipicamente, $L_{\max} = 255$ para imagens representadas por 8 bits de profundidade. A métrica PSNR é expressa em *decibel* (dB), unidade originalmente definida para medir intensidade sonora em escala logarítmica. Valores típicos de PSNR variam entre 20 (para RMSE = 25.5) e 40 (para RMSE = 2.55). Quanto maior essa métrica, melhor a nova imagem se aproxima da original.

Uma métrica relacionada é a relação sinal-ruído (SNR, do inglês, *Signal to Noise Ratio*), definida como

$$\text{SNR} = 10 \log_{10} \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y)]^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2} \quad (2.34)$$

2.12.5 Covariância

A covariância entre duas imagens f e g é dada por

$$\sigma_{fg} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f][g(x, y) - \mu_g]}{MN} \quad (2.35)$$

em que μ_f e μ_g representam o nível de cinza médio nas imagens f e g , respectivamente.

2.12.6 Coeficiente de Correlação

O coeficiente de correlação entre duas imagens f e g é dado por

$$r = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f][g(x, y) - \mu_g]}{\sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f]^2 \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [g(x, y) - \mu_g]^2}} \quad (2.36)$$

Observando-se a equação 2.36, a correlação pode ser vista como a covariância dividida por um fator que depende da distribuição dos níveis de cinza (variância da imagem) de cada uma das imagens. O valor de r varia entre -1 a 1 , sendo que valores próximos de zero representam um relacionamento linear mais fraco entre as duas imagens.

2.12.7 Coeficiente de Jaccard

O coeficiente de Jaccard entre duas imagens f e g é dado por

$$J = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \begin{cases} = 1, & \text{se } f(x, y) = g(x, y) \\ = 0, & \text{caso contrário} \end{cases}}{MN} \quad (2.37)$$

em que a igualdade $f(x, y) = g(x, y)$ é obtida considerando-se um determinado valor de tolerância δ , ou seja, $|f - g| \leq \delta$.

O coeficiente de Jaccard é igual a 0 para duas imagens que não apresentam nenhuma similaridade. Duas imagens que apresentam todos os elementos idênticos possuem coeficiente de Jaccard igual a 1 .

2.13 Limites da Imagem

Durante algumas operações, por exemplo a filtragem espacial descrita na seção 4.4.1, parte da máscara pode se localizar fora das dimensões $M \times N$ da imagem, como ilustrado na figura 2.25. Quando isso ocorrer, a operação requer cuidados especiais.

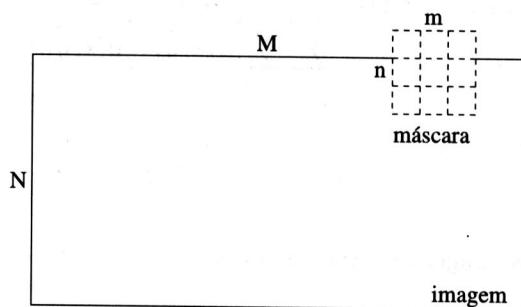


Figura 2.25: Parte da máscara localizada fora das dimensões da imagem.

Há várias estratégias diferentes para tratar desse problema. Uma abordagem é simplesmente ignorar os pixels para os casos em que a operação em questão não possa ser realizada.

Outra abordagem é copiar o valor do pixel correspondente da imagem original, caso não seja possível realizar a operação. A imagem terá uma borda de pixels não processados. Outra estratégia é utilizar uma máscara modificada para realizar a operação, tratando os pixels da borda de maneira especial. Isso pode aumentar consideravelmente a complexidade da operação.

Uma outra alternativa é verificar se o pixel possui coordenadas dentro das dimensões da imagem. Se não possuir, as coordenadas podem ser refletidas na imagem original ou repetidas de forma circular, conforme ilustrado nas figuras 2.26(a) e (b), respectivamente.

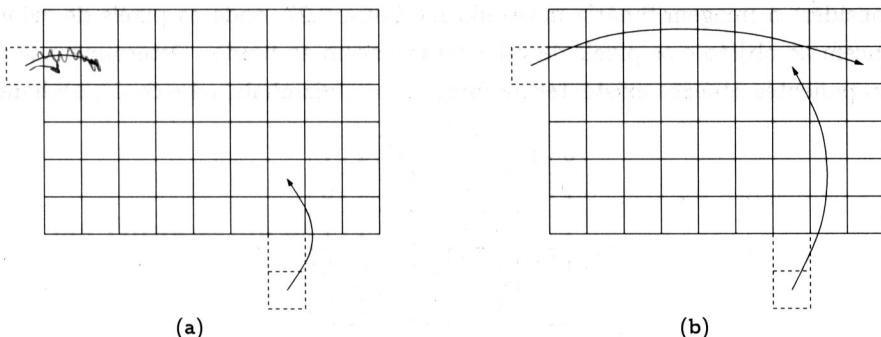


Figura 2.26: Abordagens para tratar dos limites da imagem. (a) indexação por reflexão; (b) indexação circular.

2.14 Notas Bibliográficas

Uma descrição detalhada dos princípios de percepção visual humana pode ser encontrada nos trabalhos de Davson (1963), Graham (1965), Cornsweet (1970), Dodwell (1970), Winston (1975), Hall e Hall (1977), Marr (1982b), Levine (1985) e Wandell (1995). Material adicional sobre a adaptação do olho para discriminar diferentes níveis de brilho pode ser encontrado em Hecht (1924), Sheppard Jr. (1968) e Sekuler e Blake (1990). Referências sobre os relacionamentos básicos entre elementos de imagem podem ser encontradas em Toriwaki et al. (1979), Yokokawa e Kunii (1980), Gonzalez e Woods (2002), Sonka et al. (1999), Jain et al. (1995) e Pitas (1993). Para mais detalhes sobre medidas de qualidade em imagens, o leitor pode consultar Wang et al. (2004), Wang et al. (2002) e Wang e Bovik (2002). Ruídos em imagens digitais são abordados em Gonzalez et al. (2004) e Schalkoff (1989). Referências sobre os processos de amostragem e quantização podem ser encontradas em Jain (1989), Papoulis (1969), Castleman (1996), Landau (1967) e Peterson e Middleton (1962). Entropia em imagens digitais é discutida em Cover e Thomas (1991) e Gray (1990).

2.15 Problemas

1. Qual a diferença entre resolução espacial e profundidade da imagem?
2. Qual o tamanho de uma imagem gerada pela amostragem de uma região de $200 \times 300 \text{ cm}^2$ em intervalos de 0.1 mm na direção x e 0.2 mm na direção y ?

3. Qual a profundidade em bits de uma imagem com 8192 níveis de cinza?
4. Considere um protocolo de transmissão de dados consistindo em pacotes com um bit de início, 8 bits de informação e um bit de parada. Qual o tempo (em segundos) necessário para se transmitir uma imagem de 1024×1024 pixels com 256 níveis de cinza à taxa de transmissão de 9600 bits/segundo?
5. Mostre que a distância D_4 (*city-block*) entre dois pontos p e q é igual ao caminho-4 mais curto entre esses pontos. Esse caminho é único?
6. Considere a imagem binária mostrada na figura 2.27, onde os pixels de valor 1 pertencem ao objeto e os pixels de valor 0 pertencem ao fundo. Determine o número de componentes conexos existentes na imagem com vizinhança-4 e com vizinhança-8.

0	1	0	0	0	0	1	1
1	0	0	0	0	0	1	0
1	0	0	1	0	1	1	1
1	0	1	0	0	1	0	0
0	0	1	0	0	1	1	0
0	1	1	0	1	1	1	0
0	1	0	0	1	1	0	1

Figura 2.27: Componentes conexos em uma imagem binária.

7. Desenvolva um algoritmo para identificar componentes conexos 3D em uma imagem binária tridimensional, considerando-se o conceito de vizinhança-6.
8. Dadas as matrizes x e y mostradas na figura 2.28, correspondentes a regiões de 5×5 pixels em duas imagens binárias, calcule:
 - (a) o resultado da operação lógica $x \text{ OR } y$.
 - (b) o resultado da operação lógica $x \text{ AND } y$.
 - (c) o resultado da operação lógica $x \text{ XOR } y$.

$$x = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Figura 2.28: Regiões de 5×5 pixels em imagens binárias.

9. Descreva os principais tipos de ruído que podem ser modelados em uma imagem.
10. A distância semi-Euclidiana entre dois pontos bidimensionais f_1 e f_2 é definida como

$$D_{(f_1, f_2)} = \begin{cases} |x_1 - x_2| + (\sqrt{2} - 1)|y_1 - y_2|, & \text{se } |x_1 - x_2| > |y_1 - y_2| \\ (\sqrt{2} - 1)|x_1 - x_2| + |y_1 - y_2|, & \text{caso contrário} \end{cases} \quad (2.38)$$

- Compare a distância semi-Euclidiana com as distâncias Euclidiana, *city-block* e *chessboard*.
11. Descreva o conceito de entropia em imagens de níveis de cinza.
 12. Descreva as principais utilizações dos operadores aritméticos de adição e subtração em imagens digitais.
 13. Quais as funções dos cones e dos bastonetes no sistema visual humano?
 14. Diferencie os conceitos de amostragem e quantização no processo de digitalização de imagens.
 15. Descreva as principais formas de representação de imagens digitais.

Referências Bibliográficas

- Castleman, K. R. (1996). *Digital Image Processing*, Prentice Hall, Inc., Englewood Cliffs, NJ, Estados Unidos.
- Cornsweet, T. N. (1970). *Visual Perception*, Academic Press, Nova York, NY, Estados Unidos.
- Cover, T. M. e Thomas, J. A. (1991). *Elements of Information Theory*, John Wiley and Sons, Inc., Nova York, NY, Estados Unidos.
- Davson, H. (1963). *Physiology of the Eye*, Little, Brown & Co., Boston, MA, Estados Unidos.
- Dodwell, P. C. (1970). *Visual Pattern Recognition*, Holt, Rinehart and Winston Inc., Nova York, NY, Estados Unidos.
- Gonzalez, R. C. e Woods, R. E. (2002). *Digital Image Processing*, Prentice Hall, Inc., Upper Saddle River, NJ, Estados Unidos.
- Gonzalez, R. C., Woods, R. E. e Eddins, S. S. (2004). *Digital Image Processing Using MATLAB*, Prentice Hall, Inc., Upper Saddle River, NJ, Estados Unidos.
- Graham, C. H. (1965). *Vision and Visual Perception*, John Wiley & Sons, Inc., Nova York, NY, Estados Unidos.
- Gray, R. (1990). *Entropy and Information Theory*, Springer-Verlag, Nova York, NY, Estados Unidos.
- Hall, C. F. e Hall, E. L. (1977). A Nonlinear Model for the Spatial Characteristics of the Human Visual System, *IEEE Transactions on Systems, Man, and Cybernetics* 7(3): 161-170.
- Hecht, S. (1924). The Visual Discrimination of Intensity and the Weber-Fechner Law, *The Journal of General Physiology* 7: 235-267.
- Jain, A. K. (1989). *Fundamentals of Digital Image Processing*, Prentice Hall, Inc., Englewood Cliffs, NJ, Estados Unidos.
- Jain, R., Kasturi, R. e Schunck, B. G. (1995). *Machine Vision*, McGraw-Hill, Inc., Nova York, NY, Estados Unidos.
- Landau, H. J. (1967). Sampling, Data Transmission and the Nyquist Rate, *Proceedings of IEEE* 55(10): 1701-1706.
- Levine, M. D. (1985). *Vision in Man and Machine*, McGraw-Hill, Nova York, NY, Estados Unidos.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco, CA, Estados Unidos.
- Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory, *Transactions of the American Institute of Electrical Engineers* pp. 617-644.

- Papoulis, A. (1969). *Systems and Transforms with Applications in Optics*, McGraw-Hill, Nova York, NY, Estados Unidos.
- Peterson, D. e Middleton, D. (1962). Sampling and Reconstruction of Wave-number Limited Functions in N-dimensional Euclidean Spaces, *Information and Control* 5: 279–323.
- Pitas, I. (1993). *Digital Image Processing Algorithms*, Prentice Hall, Inc., Englewood Cliffs, NJ, Estados Unidos.
- Schalkoff, R. J. (1989). *Digital Image Processing and Computer Vision*, John Wiley & Sons, Inc., Nova York, NY, Estados Unidos.
- Sekuler, R. e Blake, R. (1990). *Perception*, McGraw-Hill Publishing Company, Nova York, NY, Estados Unidos.
- Shannon, C. E. (1948). A Mathematical Theory of Communication, *Bell System Technical Journal* 27: 379–423.
- Sheppard Jr., J. J. (1968). *Human Color Perception: A Critical Study of the Experimental Foundation*, Elsevier, Nova York, NY, Estados Unidos.
- Sonka, M., Hlavac, V. e Boyle, R. (1999). *Image Processing, Analysis, and Machine Vision*, Brooks/Cole Publishing Company, Pacific Grove, CA, Estados Unidos.
- Toriwaki, J., Kato, N. e Fukumura, T. (1979). Parallel Local Operations for a New Distance Transformation of a Line Pattern and Their Applications, *IEEE Transactions on Systems, Man and Cybernetics* 9: 628–643.
- Wandell, B. A. (1995). *Foundations of Vision*, Sinauer Press, Sunderland, MA, Estados Unidos.
- Wang, Z. e Bovik, A. C. (2002). A Universal Image Quality Index, *IEEE Signal Processing Letters* 9(3): 81–84.
- Wang, Z., Bovik, A. C. e Lu, L. (2002). Why is Image Quality Assessment So Difficult?, *IEEE International Conference on Acoustics, Speech & Signal Processing*, Vol. 4, pp. 3313–3316.
- Wang, Z., Bovik, A., Sheikh, H. e Simoncelli, E. (2004). Image Quality Assessment: From Error Measurement to Structural Similarity, *IEEE Transactions on Image Processing* 13(1).
- Winston, P. H. (1975). *The Psychology of Computer Vision*, McGraw-Hill, Nova York, NY, Estados Unidos.
- Yokokawa, K. e Kunii, T. L. (1980). A Definition of Neighborhood of a Region for Picture Processing, *Computer Graphics and Image Processing* 14(2): 112–144.