

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELETRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELETRICA**

MARCO TÚLIO FAISSOL TANNÚS

**Comparação de técnicas para a
determinação de semelhança entre
imagens digitais**

MARCO TÚLIO FAISSOL TANNÚS

Comparação de técnicas para a determinação de semelhança entre imagens digitais

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Processamento da Informação

Uberlândia, 25 de maio de 2008

Banca Examinadora

Profa. Dra. Edna Lúcia Flores – Orientadora (UFU)

Prof. Dr. Sandrerley Ramos Pires (ALFA)

Prof. Dr. Gilberto Arantes Carrijo (UFU)

Prof. Dr. Antonio Claudio Pachoaelli Veiga (UFU)

Dados Internacionais de Catalogação na Publicação (CIP)

T167c Tannús, Marco Túlio Faissol, 1963-

Comparação de técnicas para a determinação de semelhança entre
imagens digitais / Marco Túlio Faissol Tannús. - 2008.

290 f. : il.

Orientadora: Edna Lúcia Flores.

Dissertação (mestrado) - Universidade Federal de
Uberlândia, Pro-
grama de Pós-Graduação em Engenharia Elétrica.

Inclui bibliografia.

1. Processamento de imagens - Técnicas digitais - Teses. 2. Banco de
dados - Teses. I. Flores, Edna Lúcia. II. Universidade Federal de Uber-
lândia. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 621.397.331

MARCO TÚLIO FAISSOL TANNÚS

Comparação de técnicas para a determinação de semelhança entre imagens digitais

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Processamento da Informação

Profa. Dra. Edna Lúcia Flores
Orientadora

Prof. Dr. Darizon Alves de Andrade
Coordenador da Pós-Graduação

**Uberlândia
2008**

RESUMO

A recuperação de imagens semelhantes em bancos de dados é um campo de pesquisa amplo, complexo e que apresenta grande demanda por aplicativos que apresentem bons resultados. O volume crescente de informações disponibilizadas ao público e o sucesso das ferramentas de busca textuais na Internet motivam a criação de utilitários que possibilitem a busca de imagens por semelhança de conteúdo. Podem-se utilizar várias características para a determinação da semelhança entre imagens digitais, tais como tamanho, cor, forma, variação de cores, textura, objetos e sua disposição espacial, entre outras. A textura e a cor são as duas características mais importantes que permitem uma análise preliminar da semelhança. Este trabalho apresenta várias técnicas constantes da literatura, que analisam textura e cor. Algumas dessas técnicas foram implementadas, seus desempenhos foram analisados e comparados e os resultados foram apresentados detalhadamente. Esse comparativo amplo permite determinar as melhores técnicas, possibilita a análise da aplicabilidade de cada uma delas e pode ser utilizada como referência em estudos futuros. As análises quantitativas de desempenho foram realizadas utilizando a métrica ANMRR, definida no padrão MPEG-7, e as matrizes de confusão, apresentadas para cada técnica testada. Dois grupos de testes quantitativos foram realizados: o primeiro utilizando um banco de imagens de texturas em tons de cinza e o segundo utilizando um banco de imagens coloridas. Os resultados dos testes com o banco de texturas em tons de cinza mostraram que as técnicas PBLIRU16, MCNC e sua combinação apresentaram os melhores desempenhos. Para o banco de imagens coloridas, os melhores desempenhos foram observados com a utilização das técnicas SCD, HDCIG e CSD.

Palavras-chaves: semelhança, histograma, MPEG-7, matriz de confusão, taxa média de retorno modificada e normalizada.

ABSTRACT

The retrieval of similar images in databases is a wide and complex research field that shows a great demand for good performance applications. The increasing volume of information available in the Internet and the success of textual search engines motivate the development of tools that make possible image searches by content similarity. Many features can be applied in determining the similarity between images, such as size, color, shape, color variation, texture, objects and their spatial distribution, among others. Texture and color are the most important features which allow a preliminary analysis of image similarity. This dissertation presents many techniques introduced in the literature, which analyze texture and color. Some of them were implemented, their performances were compared and the results were presented. This comparison allows the determination of the best techniques, making possible the analysis of their applicability and can be used as a reference in future works. The quantitative performance analyses were done using the ANMRR metric, defined in the MPEG-7 standard, and the confusion matrices were presented for each of the tested techniques. Two groups of quantitative tests were realized: the first one was applied upon a gray scale texture database and the second one, upon a color image database. For the experiment with the gray scale texture images, the techniques PBLIRU16, MCNC and their combination presented the best performances. For the experiment with the color images, SCD, HDCIG and CSD techniques performed best.

Keywords: similarity, histogram, MPEG-7, confusion matrix, average normalized modified retrieval rate.

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Introdução.....	19
1.2	Objetivos e Justificativa.....	21
1.3	Estrutura desta Dissertação	23
1.4	Considerações Finais deste Capítulo	24
2	HISTOGRAMAS	25
2.1	Introdução.....	25
2.2	Definição de Histograma.....	25
2.3	Tipos de Medida	31
2.4	Comparação entre Histogramas.....	37
2.5	Projeções Suavizantes.....	49
2.6	Assinaturas.....	56
2.7	Considerações Finais deste Capítulo	61
3	SEMELHANÇA EM IMAGENS DIGITAIS	62
3.1	Introdução.....	62
3.2	Revisão Bibliográfica	65
3.2.1	Conceitos básicos.....	65
3.2.2	Alguns métodos de extração de características texturais	67
3.2.2.1	Autocorrelação.....	68
3.2.2.2	Densidade de bordas	71
3.2.2.3	Densidade de extremos.....	72
3.2.2.4	Modelos de imagens	75
3.2.2.5	Estimação de energia.....	77
3.3	Matriz de Co-ocorrência de Níveis de Cinza	91
3.3.1	O cálculo das matrizes.....	92
3.3.2	As principais características propostas	94
3.3.3	Os histogramas de soma e diferença	106
3.4	Matriz de Tamanho de Corridas de Níveis de Cinza.....	110
3.4.1	O cálculo da matriz MTCNC.....	110
3.4.2	As características propostas	113
3.4.3	Alguns aprimoramentos	119
3.5	A Representatividade da Percepção das Características e a Matriz da Diferença de Níveis de Cinza da Vizinhança	122
3.6	Espectro de Textura e seus Aprimoramentos.....	130
3.6.1	A unidade de textura	130
3.6.2	O espectro de textura	133
3.6.3	As características propostas	134
3.6.4	Alguns aperfeiçoamentos – os padrões binários locais.....	137
3.6.5	Padrões binários locais invariantes à rotação	144
3.6.6	Padrões binários locais uniformes.....	149
3.7	Tópicos sobre o Padrão MPEG-7.....	154
3.7.1	Os procedimentos de testes.....	156
3.7.2	Os descritores de cor.....	160

3.7.2.1	O descritor escalável de cores - SCD.....	161
3.7.2.2	O descritor de estrutura de cores - CSD	164
3.7.2.3	O descritor de cores dominantes - DCD.....	169
3.7.2.4	O descritor de leiaute de cores - CLD.....	177
3.7.3	Os descritores de textura	180
3.7.3.1	O descritor de exploração de textura - TBD	180
3.7.3.2	O descritor de texturas homogêneas.....	182
3.7.3.3	O descritor de histograma de bordas - EHD	183
3.8	Algumas Técnicas Recentes.....	185
3.8.1	O histograma da direção e magnitude do gradiente.....	186
3.8.2	Os histogramas de cores e de variação da magnitude do gradiente	192
3.8.3	O descritor de padrões de bordas locais	200
3.8.4	A distância normalizada da informação	208
3.9	Considerações Finais deste Capítulo	210
4	OS RESULTADOS OBTIDOS	212
4.1	Introdução.....	212
4.2	Algumas Informações Contextuais	212
4.2.1	As bases de imagens utilizadas nos testes.....	212
4.2.2	As técnicas testadas.....	217
4.2.3	Os instrumentos de avaliação de desempenho utilizados.....	218
4.2.4	Outras informações contextuais	220
4.3	Os Testes Quantitativos da Base de Texturas de Brodatz	221
4.3.1	A média e variância da intensidade	221
4.3.1.1	Base de texturas Brodatz 512 x 512	221
4.3.1.2	Base de texturas Brodatz 128 x 128	225
4.3.2	A matriz de co-ocorrência de níveis de cinza (MCNC).....	228
4.3.2.1	Base de texturas Brodatz 512 x 512	229
4.3.2.2	Base de texturas Brodatz 128 x 128	231
4.3.3	Os padrões binários locais uniformes (PBLIRU16).....	232
4.3.3.1	Base de texturas Brodatz 512 x 512	233
4.3.3.2	Base de texturas Brodatz 128 x 128	234
4.3.3.3	Teste combinando PBLIRU16 com MCNC e com a Média e a Variância ...	236
4.3.4	O histograma da direção e magnitude do gradiente.....	239
4.3.4.1	Base de texturas Brodatz 512 x 512	241
4.3.4.2	Base de texturas Brodatz 128 x 128	242
4.3.5	O descritor de texturas homogêneas (HTD).....	243
4.3.5.1	Base de texturas Brodatz 512 x 512	244
4.3.5.2	Base de texturas Brodatz 128 x 128	245
4.3.6	O descritor de histograma de bordas (EHD).....	247
4.3.7	O conjunto de histogramas de cores e da direção da variação da magnitude do gradiente.....	255
4.3.7.1	Base de texturas Brodatz 512 x 512	255
4.3.7.2	Base de texturas Brodatz 128 x 128	256
4.3.8	Comentários sobre os testes de semelhança de texturas.....	258
4.4	Os Testes Quantitativos da Base de Imagens Coloridas SIMPLIcity.....	264
4.4.1	O descritor de leiaute de cores (CLD)	264
4.4.2	O descritor de estrutura de cores (CSD).....	266

4.4.3	O descritor de cores dominantes (DCD)	267
4.4.4	O descritor escalável de cores (SCD)	268
4.4.5	O conjunto de histogramas de cores e da direção de variação da magnitude do gradiente	269
4.4.6	Comentários sobre os testes de semelhança de imagens coloridas	271
4.5	Os Testes Qualitativos da Base de Imagens Coloridas VisTex	276
4.6	Conclusões	280
5	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS.....	283
5.1	Introdução.....	283
5.2	Conclusões	283
5.3	Contribuições	284
5.4	Trabalhos Futuros.....	285
	REFERÊNCIAS BIBLIOGRÁFICAS.....	287

LISTA DE FIGURAS

Figura 2.1 - Um conjunto amostral e duas características de seus elementos.	27
Figura 2.2 - Histogramas de cor e de peso de um conjunto de bolas coloridas.	28
Figura 2.3 - Duas imagens distintas que apresentam o mesmo histograma.	30
Figura 2.4 - Cálculo do histograma de ângulos de tangentes.	33
Figura 2.5 - Exemplo de cálculo da distância linear e circular entre os pontos 45° e 270°	37
Figura 2.6 - Dois conjuntos de bolas coloridas e seus histogramas.	40
Figura 2.7 - Histograma de peso de bolas para três conjuntos A, B e C.	41
Figura 2.8 - Representação pictórica dos histogramas H_A e H_B	46
Figura 2.9 - Duas formas de movimentação de blocos para transformar o histograma H_A no histograma H_B	47
Figura 2.10 - Representação de dois histogramas de medida modular.	48
Figura 2.11 - Movimentos para transformar H_A em H_B , considerando a medida modular. (a) Movimento 1, mínimo; (b) Movimento 2, não mínimo; (c) Um ciclo completo.	49
Figura 2.12 - Histogramas de uma grandeza hipotética, calculados para diferentes número de níveis. (a) Histograma com 256 níveis; (b) histograma com 26 níveis; (c) histograma suavizado por projeção de bancos vizinhos.	51
Figura 2.13 - Resultado da aplicação de duas técnicas para aproximar um conjunto de pontos por uma curva. Curva contínua: interpolação. Curva tracejada: aproximação por polinômio de grau menor.	52
Figura 2.14 - Seqüência de quatro bancos utilizando cosseno quadrado levantado.	55
Figura 2.15 - (a) Histograma H_A de um conjunto A; (b) assinatura S_A desse histograma.	57
Figura 2.16 - Obtenção das assinaturas estendidas a partir dos histogramas originais.	59
Figura 3.1 - Exemplos de imagens que apresentam característica predominante de textura e de cor. (a) Textura D84 (BRODATZ, 1968). (b) Colcha colorida. (c) Cor e textura estão igualmente presentes.	67
Figura 3.2 - Texturas e sua auto-correlação. (a) Uma textura fina: D55 (BRODATZ, 1968). (b) Uma textura grossa: D56 (BRODATZ, 1968). (c) e (d) As regiões centrais da auto-correlação das respectivas texturas. (e) e (f) O gráfico da linha central horizontal da auto-correlação das respectivas texturas.	70
Figura 3.3 - (a) e (b) Magnitude do gradiente para as texturas apresentadas na Figura 3.2; (c) e (d) histograma normalizado da magnitude do gradiente para essas respectivas texturas.	72
Figura 3.4 - (a) e (b) Resultado da marcação dos extremos para as texturas da Figura 3.2; (c) e (d) densidade média dos extremos calculada em um vizinhança 8×8 ; (e) e (f) histograma normalizado da distribuição da densidade de bordas.	74
Figura 3.5 - Processo de extração de características de imagens por meio de filtragem.	78
Figura 3.6 - Resposta em freqüência dos filtros separáveis de Laws h_1, h_2, h_3, h_4 e h_5 na ordem dos picos.	79

Figura 3.7 – Divisão do espectro de frequências em 25 regiões. Os algarismos no interior de cada região indicam os índices dos filtros aplicados na horizontal e na vertical.	79
Figura 3.8 – Resultado da filtragem da textura D56 da Figura 3.2(b) pelo banco de filtros de Laws.	80
Figura 3.9 – Resultado da filtragem da textura D55 da Figura 3.2(a) pelo banco de filtros de Laws.	81
Figura 3.10 – Filtros de Jain (a) em anel; (b) em cunha.	82
Figura 3.11 – (a) Resposta impulsiva de um filtro de Gabor normalizado unidimensional, partes real (linha sólida) e imaginária (linha tracejada); (b) resposta em frequência do mesmo filtro.	85
Figura 3.12 – (a) e (b) Partes real e imaginária de um filtro de Gabor bidimensional; (c) resposta em frequência desse filtro.	86
Figura 3.13 – Espectro de um banco de filtros de Gabor.	87
Figura 3.14 – Resposta do padrão D55 da Figura 3.2(a) ao utilizar um banco de filtros de Gabor de 4 frequências e 5 rotações.	88
Figura 3.15 – Resposta do padrão D55 da Figura 3.2(b) ao utilizar um banco de filtros de Gabor de 4 frequências e 5 rotações.	89
Figura 3.16 – (a) Uma imagem; (b) suas matrizes de co-ocorrência.	93
Figura 3.17 – Texturas com baixa (D105) e alta (D75) energia, suas respectivas matrizes MCNC e valores de características.	100
Figura 3.18 – Texturas com baixa (D44) e alta (D105) entropia, suas respectivas matrizes MCNC e valores de características.	101
Figura 3.19 – Texturas com baixo (D48) e alto (D106) contraste, suas respectivas matrizes MCNC e valores de características.	101
Figura 3.20 – Texturas com baixa (D32) e alta (D101) variância, suas respectivas matrizes MCNC e valores de características.	102
Figura 3.21 – Texturas com baixa (D106) e alta (D48) correlação, suas respectivas matrizes MCNC e valores de características.	102
Figura 3.22 – Texturas com baixa (D102) e alta (D102) homogeneidade, suas matrizes MCNC e valores de características.	103
Figura 3.23 – Um exemplo de imagem e as quatro matrizes principais de comprimento de corrida de níveis de cinza.	111
Figura 3.24 – Valores das ênfases em corridas curtas e longas para quatro distribuições.	116
Figura 3.25 – Valores da desuniformidade para quatro distribuições.	117
Figura 3.26 – As texturas D55 e D56, suas respectivas matrizes de corrida de cinza e as características calculadas a partir delas.	118
Figura 3.27 – Exemplo de cálculo do número de unidade de textura.	132
Figura 3.28 – Sequências distintas para o cálculo do N_{UT}	132
Figura 3.29 – A textura D106 e dois histogramas de textura.	134
Figura 3.30 – Exemplo de cálculo do valor do padrão binário local.	139
Figura 3.31 – Textura D106 e seu histograma PBL.	139
Figura 3.32 – Textura D76 e seu histograma PBL.	140
Figura 3.33 – Textura D79 e seu histograma PBL.	140
Figura 3.34 – Textura D86 e seu histograma PBL.	141
Figura 3.35 – Realce da micro-textura de regiões das texturas D76, D79, D86 E D106.	142
Figura 3.36 – Texturas palha e lã, normais e rotacionadas e seus correspondentes histogramas PBLIR.	146

Figura 3.37 – Texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIR normalizados.	147
Figura 3.38 – Texturas madeira rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIR normalizados.	148
Figura 3.39 – Exemplos de padrões binários locais uniformes e não uniformes e seus respectivos números de transições.....	151
Figura 3.40 – Texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIRU2 normalizados.....	153
Figura 3.41 – Texturas madeira rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIRU2 normalizados.....	154
Figura 3.42 – Imagens coloridas e seus histogramas SCD.....	162
Figura 3.43 – Duas imagens diferentes que possuem o mesmo histograma SCD.....	165
Figura 3.44 – Duas imagens distintas e seus histogramas CSD.....	166
Figura 3.45 – Imagens coloridas e seus histogramas CSD.....	169
Figura 3.46 – Imagens coloridas e uma representação livre de seus histogramas DCD.	176
Figura 3.47 – Imagens coloridas e seus histogramas CLD.	179
Figura 3.48 – Divisão da imagem e máscaras detectoras de bordas.	185
Figura 3.49 – Uma imagem e seu gradiente apresentado nas formas polar e cartesiana.	187
Figura 3.50 – Histograma de direção do gradiente da textura <i>raffia.000</i> e suas versões suavizadas.	189
Figura 3.51 – Duas texturas rotacionadas e seus histogramas de direção do gradiente.	190
Figura 3.52 – Diagrama que ilustra o cálculo do gradiente.	193
Figura 3.53 – A contabilização dos <i>bins</i> do histograma HDCIG.....	193
Figura 3.54 – Quatro imagens coloridas que são comparadas pelos histogramas HDCIG, HUCUB e HBCNB.	194
Figura 3.55 – Histogramas HDCIG das imagens da Figura 3.54.	195
Figura 3.56 – Histogramas HUCUB das imagens da Figura 3.54.....	197
Figura 3.57 – Histogramas HBCNB das imagens da Figura 3.54.	198
Figura 3.58 – Cálculo do padrão local de bordas, LEP, de um elemento da imagem.....	201
Figura 3.59 – O cálculo do LEP modificado.	205
Figura 4.1 – Versões rotacionadas da textura <i>bark</i> e os 16 retalhos 128×128 <i>pixels</i> da versão rotacionada de 200°	214
Figura 4.2 – As outras 12 texturas originais de dimensão 512×512 <i>pixels</i> utilizadas para compor a base de texturas dos testes deste trabalho.	215
Figura 4.3 – Duas amostras de cada classe da base SIMPLIcity: indígenas, praia, construção, ônibus, dinossauro, elefante, flor, cavalo, montanha e comida.	216
Figura 4.4 – (a) Matriz de confusão simples; (b) matriz de confusão percentual.	220
Figura 4.5 – Gráfico de variação da ANMRR em função do número de coeficientes para a base de busca Brodatz 512×512	240
Figura 4.6 – Gráfico de variação da ANMRR em função do número de coeficientes para a base de busca Brodatz 128×128	240
Figura 4.7 – Gráfico comparativo dos valores médios dos histogramas de bordas para as texturas Brodatz 128×128	249
Figura 4.8 – Exemplo de resultados de busca de texturas levemente direcionais utilizando o descritor de histograma de bordas.	252
Figura 4.9 – Exemplo de resultados de busca de texturas com direcionalidade mais acentuada utilizando o descritor de histograma de bordas.....	253

Figura 4.10 – Outro exemplo de resultados de busca de texturas direcionais utilizando o descritor de histograma de bordas. As confusões de ângulos adjacentes estão marcadas.....	254
Figura 4.11 – Gráfico comparativo dos resultados dos testes realizados nas imagens da base de busca Brodatz 512 x 512.	259
Figura 4.12 – Gráfico comparativo dos resultados dos testes realizados nas imagens da base de busca Brodatz 128 x 128.	261
Figura 4.13 – Gráfico dos valores da ANMRR em função dos valores dos pesos das componentes de cor e de textura na métrica da técnica dos histogramas HDCIG + HUCUB + HBCNB.	270
Figura 4.14 – Gráfico comparativo dos resultados obtidos nos testes realizados nas imagens coloridas da base de busca SIMPLIcity.	273
Figura 4.15 – Primeiro exemplo de busca por semelhança na base VisTex utilizando a técnica HDCIG.	277
Figura 4.16 – Segundo exemplo de busca por semelhança na base VisTex utilizando a técnica HDCIG.	278
Figura 4.17 – Terceiro exemplo de busca por semelhança na base VisTex utilizando a técnica HDCIG.	279

LISTA DE TABELAS

Tabela 2.1 - Dados de cor e peso para um conjunto de oito bolas coloridas.....	27
Tabela 2.2 - Distâncias entre os elementos de M	35
Tabela 3.1 - Comparativo entre os valores de contraste MCNC, variância MCNC e contraste da imagem para algumas texturas do álbum de Brodatz. As linhas estão ordenadas pela Variância MCNC.	105
Tabela 3.2 - Distâncias nominais dos histogramas das texturas D76, D79, D86 e D106.....	141
Tabela 3.3 - Propostas de distribuições multidimensionais de diferença de níveis de cinza entre <i>pixels</i> vizinhos.....	143
Tabela 3.4 - A lista dos padrões binários locais invariantes à rotação.....	145
Tabela 3.5 - Distâncias nominais entre histogramas PBLIR das texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90°	147
Tabela 3.6 - Distâncias nominais entre histogramas PBLIR das texturas madeira rotacionadas de 0° , 30° , 60° e 90°	148
Tabela 3.7 - Distâncias nominais entre histogramas PBLIRU2 das texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90°	153
Tabela 3.8 - Distâncias nominais entre histogramas PBLIRU2 das texturas madeira rotacionadas de 0° , 30° , 60° e 90°	154
Tabela 3.9 - Distâncias normalizadas entre os histogramas SCD das imagens da Figura 3.42.	163
Tabela 3.10 - Esquema de quantização do espaço de cores HMMD para o histograma CSD de 184 <i>bins</i>	168
Tabela 3.11 - Distâncias entre os histogramas CSD da Figura 3.45.....	169
Tabela 3.12 - Os DCD das imagens da Figura 3.46.....	177
Tabela 3.13 - Distâncias normalizadas entre os histogramas DCD das imagens da Figura 3.46.	177
Tabela 3.14 - Distâncias entre os histogramas CLD das respectivas imagens da Figura 3.47.	179
Tabela 3.15 - Os descritores para cada uma das versões rotacionadas da textura <i>raffia</i> e da textura <i>straw</i>	191
Tabela 3.16 - Distâncias entre as versões rotacionadas da textura <i>raffia</i> e a textura <i>straw</i>	191
Tabela 3.17 - Distâncias entre as imagens da Figura 3.54.	200
Tabela 4.1 - Matriz de confusão para Brodatz 512 x 512, utilizando a técnica da média e da variância sem normalização.	222
Tabela 4.2 - Matriz de confusão para Brodatz 512 x 512, utilizando técnica da média e da variância com normalização.	222
Tabela 4.3 - Matriz de confusão para Brodatz 512 x 512, utilizando apenas a média, sem normalização.....	224
Tabela 4.4 - Matriz de confusão para Brodatz 512 x 512, utilizando apenas a variância, sem normalização.....	225
Tabela 4.5 - Matriz de confusão para Brodatz 128 x 128, técnica da média e da variância sem normalização.....	226
Tabela 4.6 - Matriz de confusão para Brodatz 128 x 128, técnica da média e da variância com normalização.....	226
Tabela 4.7 - Matriz de confusão para Brodatz 128 x 128, utilizando apenas a média, sem normalização.....	227

Tabela 4.8 – Matriz de confusão para Brodatz 128 x 128, utilizando apenas a variância, sem normalização.....	228
Tabela 4.9 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica da MCNC com normalização.....	230
Tabela 4.10 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica da MCNC com normalização.....	231
Tabela 4.11 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica da MCNC com normalização, considerando-se apenas as características f_1 a f_3	232
Tabela 4.12 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica da PBLIRU16 sem normalização.	233
Tabela 4.13 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica da PBLIRU16 com normalização.	234
Tabela 4.14 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica da PBLIRU16 sem normalização.	235
Tabela 4.15 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica da PBLIRU16 com normalização.	235
Tabela 4.16 – Matriz de confusão para Brodatz 128 x 128, utilizando a combinação de PBLIRU16 e da média e da variância.	237
Tabela 4.17 – Matriz de confusão para Brodatz 128 x 128, utilizando a combinação de PBLIRU16 e MCNC.	237
Tabela 4.18 – Matriz de confusão para Brodatz 128 x 128, utilizando a combinação de PBLIRU16, MCNC e da média e da variância.....	238
Tabela 4.19 – Resumo dos resultados obtidos nos testes realizados com a média e a variância, MCNC e PBLIRU16 e suas combinações.....	238
Tabela 4.20 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica do histograma de direção e magnitude do gradiente com 3 coeficientes.	241
Tabela 4.21 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica do histograma de direção e magnitude do gradiente com 1 coeficiente.....	242
Tabela 4.22 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica do histograma de direção e magnitude do gradiente com 3 coeficientes.	243
Tabela 4.23 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica do descritor de texturas homogêneas.....	244
Tabela 4.24 – Matriz de confusão reduzida para Brodatz 128 x 128, utilizando a técnica do descritor de texturas homogêneas.	245
Tabela 4.25 – Bloco de confusão para as rotações da textura <i>brick</i> , apresentando baixo grau de confusão.....	246
Tabela 4.26 – Bloco de confusão para as rotações da textura <i>bark</i> , apresentando confusão para as rotações adjacentes.....	246
Tabela 4.27 – Bloco de confusão para as rotações da textura <i>grass</i> , apresentando alto grau de confusão.	247
Tabela 4.28 – Matriz de confusão reduzida para Brodatz 128 x 128, utilizando a técnica do descritor de histograma de bordas.	248
Tabela 4.29 – Matriz de confusão reduzida para Brodatz 128 x 128, utilizando a técnica do descritor de histograma de bordas. Os blocos de confusão são apresentados esquematicamente.	250
Tabela 4.30 – Matriz de confusão para Brodatz 512 x 512, utilizando a técnica dos histogramas HDCIG + HUCUB + HBCNB.	256
Tabela 4.31 – Matriz de confusão para Brodatz 128 x 128, utilizando a técnica dos histogramas HDCIG + HUCUB + HBCNB.	257
Tabela 4.32 – Comparação dos resultados obtidos nos testes efetuados nas imagens da base de busca Brodatz 512 x 512.....	259

Tabela 4.33 – Comparação dos resultados dos testes efetuados nas imagens da base de busca Brodatz 128 x 128.....	260
Tabela 4.34 – Comparação dos resultados dos testes não invariantes à rotação realizados nas imagens da base de busca Brodatz 128 x 128.....	262
Tabela 4.35 – Comparação dos valores percentuais das matrizes de confusão para as técnicas testadas como invariantes à rotação nas imagens da base de busca Brodatz 128 x 128.	263
Tabela 4.36 – Matriz de confusão para SIMPLIcity, utilizando a técnica do descritor de leiaute de cores.	265
Tabela 4.37 – Matriz de confusão para SIMPLIcity, utilizando a técnica do descritor de estrutura de cores.	266
Tabela 4.38 – Matriz de confusão para SIMPLIcity, utilizando a técnica do descritor de cores dominantes.	267
Tabela 4.39 – Matriz de confusão para SIMPLIcity, utilizando a técnica do descritor escalável de cores.	268
Tabela 4.40 – Matriz de confusão para SIMPLIcity, utilizando a técnica dos histogramas HDCIG + HUCUB + HBCNB.	269
Tabela 4.41 – Comparação dos resultados obtidos nos testes realizados nas imagens coloridas da base de busca SIMPLIcity.	272
Tabela 4.42 – Comparação dos valores percentuais das matrizes de confusão para as técnicas testadas nas imagens da base SIMPLIcity.	275

LISTA DE ALGORITMOS

Listagem 2.1 - Algoritmo para o cálculo da distância ordinal entre as assinaturas.	60
Listagem 2.2 - Algoritmo para o cálculo da distância modular entre as assinaturas.	60

LISTA DE ABREVIACÕES

ANMRR	Posição de retorno média, modificada e normalizada
ARR	Posição de retorno média
AVR	Posição média de retorno
CCS	Covariâncias Centrais Simétricas
CGTCNC	Característica Geral de Tamanho de Corrida de Nível de Cinza
CLD	Descritor de leiaute de cores
CSD	Descritor de estrutura de cores
DCD	Descritor de cores dominantes
DEC	Distância Eucilidiana Cumulativa
DLM	Distância Land Mover
DNC	Desuniformidade de Níveis de Cinza
DTC	Desuniformidade de Tamanhos de Corrida
ECA	Ênfase de Corridas de nível de cinza Alto
ECB	Ênfase de Corridas de nível de cinza Baixo
ECC	Ênfase em Corridas Curtas
ECCNA	Ênfase em Corridas Curtas com Níveis de cinza Altos
ECCNB	Ênfase em Corridas Curtas com Níveis de cinza Baixos
ECL	Ênfase em Corridas Longas
ECLNA	Ênfase em Corridas Longas com Níveis de cinza Altos
ECLNB	Ênfase em Corridas Longas com Níveis de cinza Baixos
EHD	Descritor de Histogramas de Bordas
FFT	Transformada rápida de Fourier
GD	Grau de direcionalidade
HBCNB	Histograma de blocos de duas cores
HDCIG	Histograma de direção de variação de magnitude do gradiente
HistGrad	Histograma de direção e magnitude do gradiente
HTD	Descritor de texturas homogêneas
HUCUB	Histograma de blocos de cor uniforme
JPEG	<i>Joint Photographic Experts Group</i>
LEP	Padrões de bordas locais
MCNC	Matriz de Co-ocorrência de Níveis de Cinza

MDA	<i>Multiple Discriminant Analysis</i>
MedVar	Características combinadas de média e variância de intensidade
MPEG	<i>Moving Picture Experts Group</i>
MRR	Posição de retorno modificada
MTCNC	Matriz de Tamanhos de Corridas de Níveis de cinza
NMRR	Posição de retorno modificada, normalizada
NR	Número de corridas em uma imagem
PBL	Padrões Binários Locais
PBLIR	Padrões Binários Locais Invariantes à Rotação
PBLIRU16	Padrões Binários Locais Invariantes à Rotação, Uniformes de vizinhança 16
PC	Percentual de Corridas
PCA	<i>Principal Component Analysis</i>
RR	Taxa de retorno
SCD	Descritor escalável de cores
SG	Simetria Geométrica
SPB	Simetria preto/branco

1 INTRODUÇÃO

1.1 Introdução

A determinação da semelhança entre imagens digitais constitui um problema que motiva as pesquisas desde que as imagens de satélite começaram a ser utilizadas para o mapeamento do relevo e dos acidentes geográficos da superfície terrestre. Esse mapeamento, também chamado de anotação, tinha de ser executado manualmente, por especialistas em interpretação das imagens de satélite. O incremento do volume dessas imagens, bem como de sua resolução, começou a inviabilizar a anotação manual. Tornou-se necessário buscar alternativas que possibilitassem a automação ao menos parcial do processo de anotação.

Dentre as primeiras técnicas desenvolvidas, destaca-se a extração de características das chamadas matrizes de co-ocorrência de níveis de cinza, proposta por Robert Haralick no início da década de 1970 (HARALICK, 1973). A essa técnica seguiram-se várias outras que possibilitaram aplicações em diversas áreas como na medicina, na segurança, na agricultura, na economia e no entretenimento.

O desenvolvimento crescente da eletrônica e da informática permitiu a criação e disseminação de equipamentos mais potentes, mais precisos e mais rápidos. Isso proporcionou a possibilidade de obter, armazenar e processar imagens digitais cada vez melhores e maiores. A consequência disso foi o aumento do número e da diversidade de imagens disponíveis. Considerando-se ainda o crescimento da utilização da Internet, o que se observa é uma facilidade incrível de distribuição dos conteúdos digitais, neles inclusos as imagens, vídeos e sons.

A necessidade de se encontrar os conteúdos desejados nesse volume inimaginável de dados que hoje se apresentam dirigiram as pesquisas de determinação para a área de anotação automática das imagens digitais. Novos padrões de anotação e de busca foram desenvolvidos, destacando-se o MPEG-7, disponibilizado no início desta década de 2000.

Este trabalho enquadra-se no contexto exposto acima e tem como objetivo estudar e comparar algumas das principais técnicas de determinação de semelhança entre imagens digitais. Essas técnicas podem ser utilizadas em três tipos básicos de aplicações: segmentação, classificação e recuperação de imagens por semelhança em bancos de dados. A segmentação tem como finalidade dividir a imagem original em segmentos que apresentem internamente características visuais homogêneas, mas que sejam distintos dos demais segmentos obtidos. Essa aplicação é útil para destacar objetos, para separar regiões que destoem do padrão desejado, ou para preparar uma imagem para posterior processamento. A classificação é utilizada para reconhecer determinados objetos, para destacar e anotar padrões conhecidos, ou para agrupar itens semelhantes em um número inicialmente indeterminado de categorias distintas. Finalmente, a recuperação de imagens em bancos de dados possibilita a localização de informações visuais em um volume muito grande de dados disponíveis.

Este trabalho enquadra-se no terceiro tipo de aplicação: a recuperação de imagens por semelhança em bancos de dados. A busca de imagens em bancos de dados tem aplicação variada e utilidade incontestável. Duas vertentes se abrem nesse ramo. A busca por padrões exatos, ou muito próximos, e a busca por padrões mais genéricos, por assim dizer, de um número maior de imagens semelhantes. O primeiro caso, a busca por imagens quase iguais tem aplicação principalmente no campo da segurança, mais especificamente no campo da identificação. Bancos de impressões digitais, bancos de íris, ou banco de faces são utilizados para identificação e autenticação de pessoas.

Já o caso de busca por padrões mais genéricos impõe requisitos de identificação menos rígidos. Da mesma forma que um dicionário de sinônimos, ou como a memória humana, o aplicativo de recuperação deve selecionar em um banco de dados muito grande aquelas imagens que possuam algumas características comuns com a imagem de busca. As imagens retornadas não precisam, e não devem restringir-se às imagens idênticas, ou muito parecidas. Ao invés, o conjunto retornado pode ser bastante heterogêneo, desde que seus componentes apresentem algumas características comuns que possibilitem a ligação entre eles.

Ainda na área de busca por semelhança, vários caminhos podem ser percorridos. Um caminho possível é a determinação de objetos, ou principais regiões de interesse presentes na imagem de busca e a conseqüente procura por imagens que os contenham. Esse paradigma mostra-se bastante complexo e sua aplicação pode resultar em tempos de processamento bastante elevados. Uma alternativa mais viável é tentar determinar características mais genéricas, de mais baixo nível, que envolvam propriedades da imagem como um todo. Essa alternativa permite executar uma filtragem prévia mais rápida das imagens da base de busca e a conseqüente redução significativa do universo de imagens semelhantes. Posteriormente, buscas detalhadas e mais demoradas podem ser executadas sobre o conjunto reduzido, possibilitando um refinamento do resultado.

As técnicas de busca de imagens semelhantes baseadas na determinação de características da imagem como um todo abrangem, invariavelmente, conceitos estatísticos. Mesmo que seja necessário determinar padrões estruturais básicos que componham a imagem, os procedimentos de avaliação da semelhança compreendem a determinação de médias, variâncias, ou inter-relações espaciais desses componentes estruturais elementares. Assim, quando se fala em conceitos estatísticos, os histogramas surgem como ferramenta básica cujo emprego pode ser de grande utilidade.

Pode-se prever uma utilização massiva de aplicativos que recuperem imagens de conteúdo semelhante. A analogia com mecanismos de busca textual na Internet é imediata e o sucesso desses mecanismos faz acreditar que os aplicativos de busca de conteúdo de mídia semelhantes apresentarão o mesmo comportamento.

Este capítulo apresenta os objetivos, a justificativa e a estrutura deste trabalho. Finalmente, são realizadas considerações finais deste capítulo.

1.2 Objetivos e Justificativa

O objetivo principal deste trabalho é comparar algumas técnicas que utilizam os histogramas como ferramenta importante na determinação de semelhança entre imagens digitais, voltada à busca preliminar e eficiente de imagens correlatas em bancos de dados. Essa determinação de semelhança visa proporcionar aos aplicativos de busca de conteúdo

visual a possibilidade de redução do universo inicial, permitindo posterior refinamento pela aplicação de técnicas mais complexas, que extrapolam o escopo deste estudo.

Como objetivos secundários, os experimentos conduzidos neste trabalho visam proporcionar uma melhor compreensão da aplicabilidade das técnicas comparadas aos problemas reais que se apresentam, bem como determinar o desempenho dessas técnicas para os bancos de imagens testados com a finalidade de obter um conjunto de dados que poderão ser utilizados como referência para trabalhos futuros que venham a propor novas técnicas de determinação de semelhança entre imagens digitais.

Os argumentos expostos na seção introdutória deste capítulo justificam a escolha do tema deste estudo. De fato, a previsão de crescente demanda por aplicações de mineração de conteúdo multimídia é motivação suficiente para direcionar os esforços de pesquisa com a finalidade de desenvolver técnicas que possibilitem soluções mais eficazes para o problema em questão.

Entretanto, nenhuma técnica apresenta desempenho ótimo para todas as aplicações. Essa afirmação é ainda mais incisiva quando analisada no contexto deste trabalho porque o conceito da semelhança envolve um alto grau de subjetividade e relatividade. A determinação da semelhança entre imagens está condicionada a aspectos objetivos mais elementares tais como forma, cor, textura, tamanho disposição espacial. Por isso, o seu grau elevado de relatividade. Condiciona-se também a aspectos subjetivos como memória, reminiscências e estética, o que lhe confere alto grau de dependência à perspectiva do observador.

Essas condicionantes fazem do problema da determinação de semelhança entre imagens um desafio complexo e intrincado, cuja solução está longe do óbvio. Entretanto, como ocorre nas mais diversas áreas do conhecimento técnico, o desenvolvimento de soluções mais adequadas passa, invariavelmente, pela melhor compreensão dos elementos fundamentais que constituem o problema.

Nesse sentido, o estudo comparativo realizado justifica-se por que possibilita delimitar com maior precisão o campo de aplicação de cada uma das técnicas testadas. A compreensão dos principais detalhes de cada técnica bem como a análise dos resultados de

sua aplicação permitem conhecer seus pontos fortes e suas deficiências. Esse conhecimento auxilia, tanto na seleção da técnica mais apropriada a cada caso, quanto no desenvolvimento de técnicas híbridas que possam vir a apresentar desempenhos melhores.

1.3 Estrutura desta Dissertação

Este capítulo apresenta os objetivos, a justificativa e a estrutura deste trabalho. Finalmente, são realizadas considerações finais deste capítulo.

O capítulo 2 apresenta o conceito de histograma, suas características, os tipos de histogramas e as formas de comparação entre eles. Finalmente, são realizadas considerações finais desse capítulo.

O capítulo 3 apresenta o conceito de semelhança entre imagens digitais, acompanhado pelo resumo do levantamento bibliográfico sobre o tema. Desse levantamento, são expostas as principais técnicas que vêm sendo utilizadas ao longo dos anos para a determinação de características estruturais e estatísticas das texturas e das características relacionadas às tonalidades de cor. A análise desses dois componentes, cor e textura, é fundamental para a caracterização das imagens. As técnicas foram apresentadas em ordem cronológica, quando possível, para deixar claro o caminho percorrido pelas investigações e os progressos que foram sendo conquistados. A seguir, foram expostas algumas técnicas desenvolvidas mais recentemente. Finalmente, são realizadas considerações finais desse capítulo.

O capítulo 4 apresenta os testes realizados, bem como os resultados obtidos. O objetivo principal dos testes foi o de avaliar quantitativamente o desempenho das técnicas experimentadas e também o de avaliar qualitativamente os resultados das buscas sob o aspecto visual. Um objetivo secundário, mas não menos importante, foi a avaliação da complexidade de implementação e a estimativa da viabilidade de aperfeiçoamento das técnicas testadas com vistas à sua melhoria de desempenho. Finalmente, são realizadas as conclusões sobre esses resultados.

O capítulo 5 apresenta as conclusões e as contribuições desta dissertação e as sugestões para trabalhos futuros.

1.4 Considerações Finais deste Capítulo

Este capítulo apresentou os objetivos do trabalho, procurou justificar a escolha do tema e a relevância do assunto e apresentou a estrutura desta dissertação.

O próximo capítulo apresenta o conceito de histogramas, suas características, os tipos de histogramas e a forma de comparação entre eles. Finalmente, são realizadas as considerações finais desse capítulo.

2 HISTOGRAMAS

2.1 Introdução

Os histogramas são ferramentas flexíveis que podem ser utilizadas em uma grande variedade de aplicações, abrangendo desde a análise estatística, econômica e social até o processamento de imagens digitais. Neste campo, os histogramas são usualmente empregados no realce e na segmentação de imagens e também na classificação de padrões (GONZALEZ e WOODS, 2000). A simplicidade do cálculo aliada à riqueza de informações que os histogramas proporcionam, fazem deles um instrumento apropriado para a determinação de semelhanças entre imagens.

Este capítulo apresenta inicialmente a definição de histograma e de alguns conceitos correlatos, bem como exemplos que ilustram sua aplicação. Segue-se um tópico importante, de aplicação recorrente quando se deseja comparar imagens por meio de seus histogramas: os métodos de comparação entre eles. O conceito de distância é apresentado, três tipos de distâncias são mencionados, doze diferentes medidas de distância entre histogramas são listadas e os algoritmos mais importantes para seu cálculo são analisados de forma detalhada. A seguir, apresentam-se dois métodos de cálculo de distâncias recentemente propostos na literatura, as projeções suavizantes e as assinaturas. Finalmente, são realizadas as considerações finais deste capítulo.

2.2 Definição de Histograma

Histograma é a representação gráfica de uma distribuição de frequência. Essa definição pode ser ampliada para abranger, além da representação gráfica, a representação numérica dessa distribuição. Gonzalez (GONZALEZ e WOODS, 2000) define histograma de uma imagem digital com níveis de cinza no intervalo $[0, L-1]$ como uma função discreta da probabilidade de ocorrência do nível de cinza r_k , $p(r_k)$, como mostrado na Equação (2.1).

$$p(r_k) = \frac{n_k}{n} \quad k = 0, 1, 2, \dots, n-1$$

onde:

(2.1)

r_k : k -ésimo nível de cinza;

n_k : número de pixels na imagem com valor r_k ; e

n : número total de pixels na imagem.

Falando grosseiramente, $p(r_k)$ fornece uma estimativa da probabilidade de ocorrência do nível de cinza r_k . O gráfico da função $p(r_k)$ para todos os valores de k fornece uma descrição global da aparência de uma imagem.

Cha e Srihari (CHA e SRIHARI, 2002) apresentaram uma definição mais genérica de histograma, aplicável a qualquer característica mensurável e não somente a níveis de cinza de pixels de uma imagem. Segundo eles, a representação por histograma de um conjunto amostral de uma população com respeito a uma métrica, ou medida, é a frequência de ocorrência dos valores quantizados daquela medida entre as amostras.

É importante destacar alguns conceitos presentes nas definições de Gonzalez (GONZALEZ e WOODS, 2000) e de Cha e Srihari (CHA e SRIHARI, 2002). Os três conceitos principais são: característica, métrica e frequência. O conceito de *característica de um conjunto amostral de uma população* ressalta o fato que o objetivo do histograma é representar uma determinada característica de uma população, e não a população em si. Uma mesma população pode ter várias representações histográficas diferentes associadas a ela. Por exemplo, considerando-se um conjunto de bolas coloridas, pode-se calcular o histograma da cor e do peso. Os dois histogramas são distintos, cada um deles representa uma característica da população, mas nenhum deles representa a população por inteiro. A Figura 2.1 mostra o conjunto de bolas e as características de cor e de peso desse conjunto.

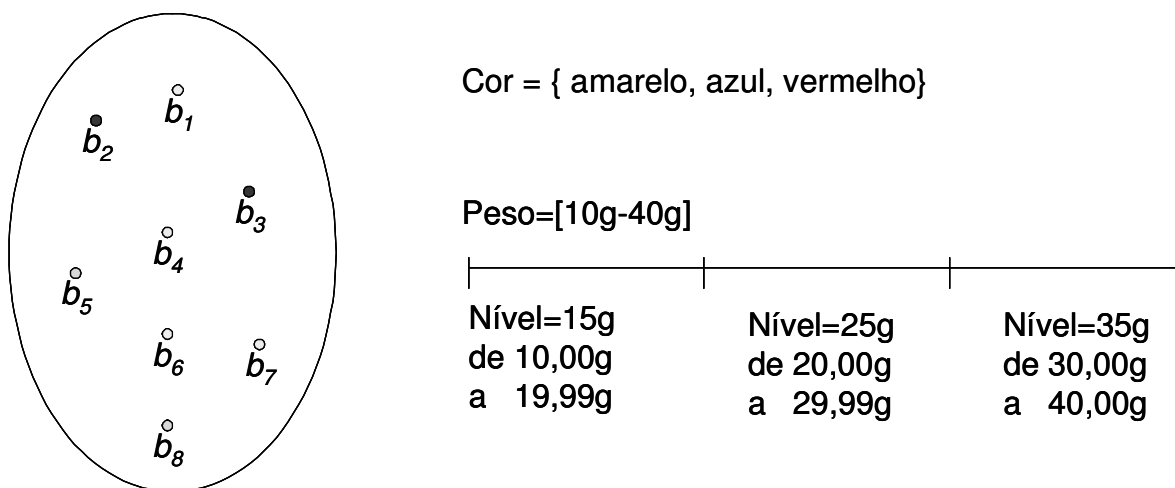


Figura 2.1 - Um conjunto amostral e duas características de seus elementos.

A Tabela 2.1 ilustra os dados para cada uma das bolas. Nessa tabela, a coluna “peso quantizado” mostra os valores do peso das bolas quantizados para as faixas arbitradas.

Tabela 2.1 - Dados de cor e peso para um conjunto de oito bolas coloridas

<i>Elemento</i>	<i>Cor</i>	<i>Peso</i>	<i>Peso quantizado</i>
<i>b1</i>	amarelo	12,475	15
<i>b2</i>	vermelho	27,989	25
<i>b3</i>	vermelho	23,251	25
<i>b4</i>	amarelo	31,000	35
<i>b5</i>	azul	19,300	15
<i>b6</i>	amarelo	29,654	25
<i>b7</i>	amarelo	25,033	25
<i>b8</i>	azul	24,788	25

A Figura 2.2 mostra o histograma das características de *cor* e de *peso*, para as bolas coloridas, conforme os dados da Tabela 2.1. Pode-se observar nessa figura que os dois histogramas são diferentes e referem-se a características distintas para a mesma população.

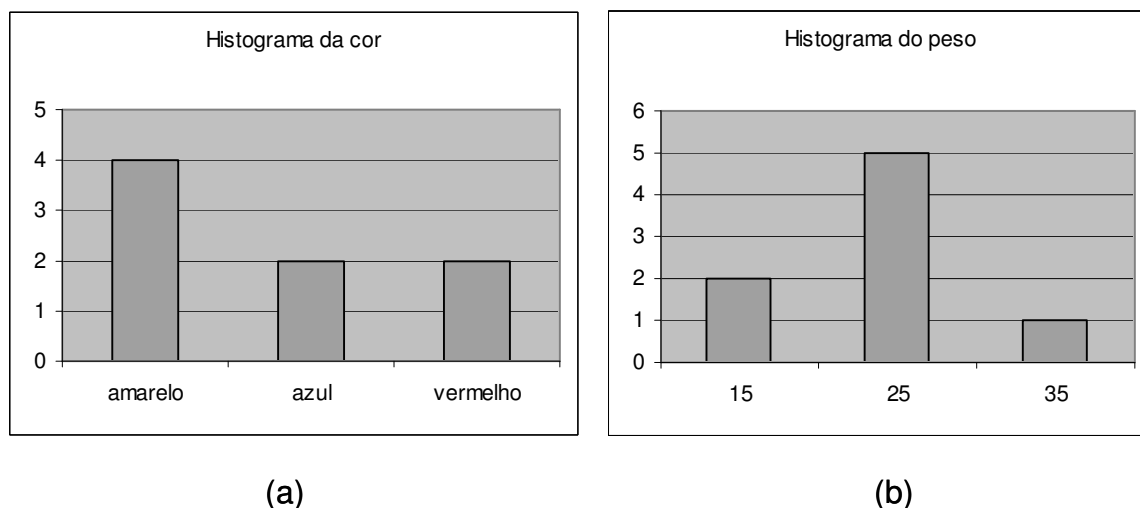


Figura 2.2 - Histogramas de cor e de peso de um conjunto de bolas coloridas.

O conceito de *métrica discreta com valores quantizados* é a necessidade de atribuir à característica que se deseja representar um valor pertencente a um universo enumerável e finito. As etapas do processo de construção do histograma são as operações de classificação e de contagem. Não existe como classificar os valores de medida e contar o número de ocorrências desses valores quando o intervalo de classificação não é enumerável. Assim, a medida subjacente pode ter como universo um sub-intervalo do conjunto dos reais. Entretanto, para se construir o histograma, deve-se arbitrar um conjunto finito de níveis em que os resultados da medida devem ser mapeados. Dessa forma, consegue-se associar os valores da medida ao nível correspondente, contabilizando sua ocorrência e construindo o histograma. Isso pode ser observado na coluna *Peso Quantizado* da Tabela 2.1. A medida do peso resulta em um valor pertencente a um intervalo contínuo. No exemplo, foram definidos três níveis para a medida peso: até 20 g, de 20 a 30 g e superior a 30 g, conforme mostrado na Figura 2.1. A cada elemento da população atribuiu-se como peso o valor central do intervalo ao qual pertencia o valor do peso medido. Ao elemento b_1 associou-se o peso 15 já que o valor medido de seu peso (12,475g) pertence ao intervalo [10-20).

Finalmente, o conceito de *freqüência* surge do processo de contagem do número de ocorrências de cada nível. A freqüência pode ser absoluta ou relativa. Se for absoluta, o

histograma é construído como uma seqüência ordenada de números inteiros. A cada nível é associado o número de ocorrências de medidas correspondentes. Se a freqüência for relativa, o histograma passa por um processo de normalização. O valor associado a cada nível é dividido pelo número total de elementos do conjunto de amostras. Esse procedimento torna o valor de cada nível do histograma uma estimativa da probabilidade de ocorrência daquele valor na população.

Após as considerações realizadas nos parágrafos anteriores deste item, segue uma definição formal de histograma. Sejam uma característica mensurável x e um conjunto enumerável de b níveis de medidas dessa característica $X = \{x_0, x_1, x_2, \dots, x_{b-1}\}$. Considere um conjunto de n elementos para os quais a característica x apresenta o conjunto A de valores quantizados de medida apresentado a seguir: $A = \{a_1, a_2, a_3, \dots, a_n\}$, tal que $a_i \in X$. O histograma do conjunto A com relação às medidas x é $H(x, A)$, uma lista ordenada de dimensão b , daqui por diante chamada *tupla*, cujos elementos são as quantidades com que ocorrem, no conjunto A , cada uma das b medidas x_i . Cada elemento do histograma é chamado de nível ou *bin*. Denota-se por $H_i(A)$ o i -ésimo *bin* do histograma, $0 \leq i \leq b - 1$. Portanto, $H_i(A)$ é o número de elementos de A que apresentam o valor x_i como medida. Assim, $H(x, A) = H(A) = [H_0(A), H_1(A), \dots, H_{b-1}(A)]$ e os valores $H_i(A)$ são obtidos pela Equação (2.2).

$$H_i(A) = \sum_{j=1}^n c_{ij} \quad \text{onde } c_{ij} = \begin{cases} 1 & \text{se } a_j = x_i \\ 0 & \text{caso contrário} \end{cases} \quad (2.2)$$

Definindo $P_i(A)$ como a probabilidade de ocorrência da medida x_i , então, pela definição de probabilidade, $P_i(A) = H_i(A)/n$. Se for calculado $P_i(A)$ para cada um dos b intervalos, $P(A) = [P_0(A), P_1(A), \dots, P_{b-1}(A)]$ é o histograma normalizado. Se x for considerada uma variável aleatória, $P(A)$ será a função densidade de probabilidade dessa variável aleatória para o conjunto A .

É importante ressaltar que o histograma não representa o conjunto de elementos a partir do qual é construído. Isso significa que o conjunto original nem sempre pode ser reconstruído a partir de seu histograma. Nem mesmo o conjunto de medidas de característica pode ser sempre reconstruído. Considere o exemplo das bolas coloridas. Pode-se reconstruir o conjunto das medidas de cor das bolas, mas não se pode reconstruir o conjunto das medidas de peso, já que foi feito um processo de quantização no qual ocorreu perda de informação.

Outro exemplo importante de impossibilidade de reconstrução é o de um histograma de níveis de cinza de uma imagem. Não se pode reconstruir a imagem a partir de seu histograma de níveis de cinza. A rigor, não se pode reconstruir uma imagem a partir do histograma de qualquer das características usualmente utilizadas na prática, como, por exemplo: níveis de cinza, direção do gradiente, matriz de co-ocorrência de níveis de cinza, histogramas de multiresolução ou histogramas de curvatura. Isso é porque ao se construir um histograma unidimensional de uma imagem a informação de posição dos pixels é perdida. Para ilustrar esse fato, a Figura 2.3 (a) apresenta duas imagens de dimensão 3 x 3 *pixels* em que cada pixel pode assumir um dos três níveis de cinza possíveis. Apesar de serem diferentes, ambas as imagens possuem o mesmo histograma, mostrado na Figura 2.3 (b).

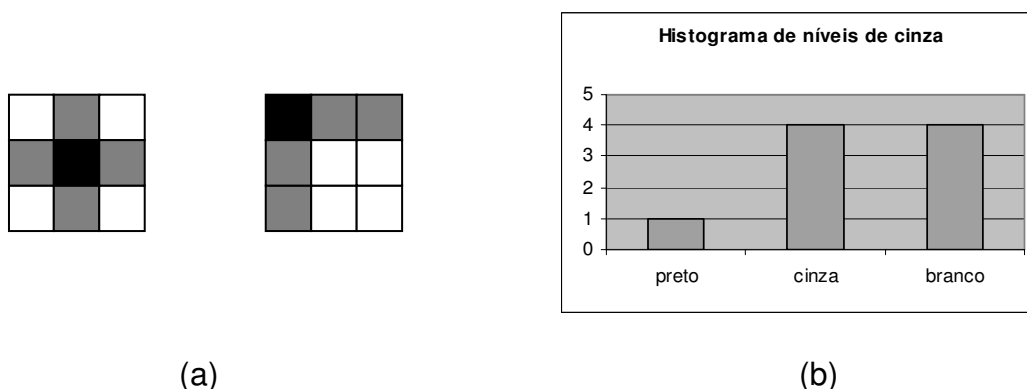


Figura 2.3 - Duas imagens distintas que apresentam o mesmo histograma.

Pode-se observar que não se pode reconstruir uma imagem ou outra a partir de seu histograma. Caso se conheça a dimensão da imagem, por exemplo, na Figura 2.3 é de 3×3 *pixels*, existem 630 imagens diferentes que podem ser construídas a partir do histograma apresentado. O problema equivale a distribuir 4 pixels brancos, 4 cinzas e 1 preto em um vetor de nove posições. Os quatro primeiros valores podem ser distribuídos em qualquer das nove posições. Os quatro valores seguintes podem ser distribuídos em cinco posições e o último valor, o pixel preto, fica na posição restante.

2.3 Tipos de Medida

Pode-se perceber pela definição de histograma que a medida é um ponto crucial para a construção do histograma. Nem todas as medidas apresentam as mesmas propriedades. Retornando ao exemplo das bolas coloridas, as duas características consideradas apresentam medidas com propriedades diferentes. O peso é medido de forma usual. Considerando-se duas bolas, é possível dizer se uma pesa mais ou pesa menos que a outra. Pode-se dizer, ainda, se ambas têm, ou não, o mesmo peso. Entretanto, a característica cor não apresenta a propriedade acima. Pode-se afirmar que duas bolas têm a mesma cor, ou que têm cores diferentes. Contudo, não faz sentido afirmar que a bola b_5 , na cor azul, mede mais ou menos que a bola b_6 , na cor amarela. Conhecer os diferentes tipos de medida é importante, já que as propriedades diferentes dos tipos de medida influenciarão a forma de se calcular a distância entre histogramas. O cálculo da distância indica se dois histogramas são mais parecidos (distância menor), ou pouco parecidos (distância maior) e, por conseguinte, indica se duas imagens são mais semelhantes, ou menos semelhantes.

Medir uma característica de um elemento significa atribuir um valor a esse elemento. Considere um conjunto formado pelos elementos de uma população $P = \{p_1, p_2, \dots, p_n\}$. Considere também uma característica x comum a todos os elementos de P . Considere ainda um conjunto M formado pelos possíveis valores que a característica x pode assumir. Medir a característica x dos elementos de P é definir uma função $f: P \rightarrow M$ que atribui a cada elemento de P (o domínio de f), um valor do conjunto M (a imagem de f).

A composição do conjunto imagem M depende do que se deseja medir. Se são as cores das bolas, $M = \{\text{amarelo, azul, vermelho}\}$, ou se é o peso, $M = \{ m \mid m \in \mathbb{R} \wedge m \in [10,40] \}$. Na primeira medida, o conjunto imagem M é enumerável e finito. Na segunda medida, é infinito e não enumerável, porém não é necessário impor qualquer restrição ao conjunto M . No entanto, pode-se considerar, sem perda de generalidade, que o conjunto M deve ser finito e, portanto, enumerável. Na construção de histogramas, se esse não é o caso, pode-se sempre definir uma nova função que será responsável por quantizar o resultado da medida. Para simplificar a notação e a argumentação, assume-se que M já é a imagem da medição quantizada.

O que determina os diferentes tipos de medida é a existência ou não, sobre o conjunto M , de uma relação de ordem linear. Um conjunto apresenta uma relação de ordem total ou linear quando (KNUTH, 1998):

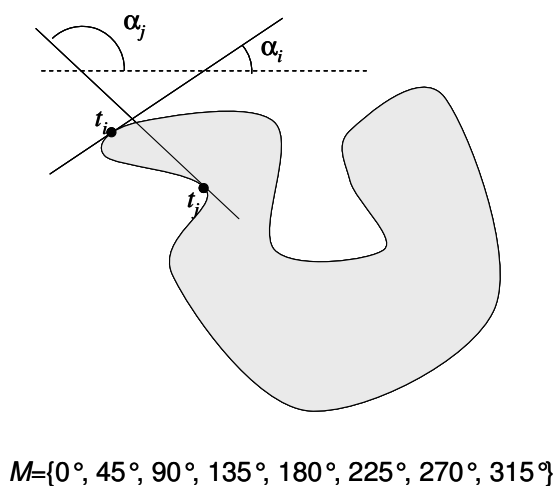
- a) Exatamente uma das possibilidades seguintes é satisfeita para qualquer par de elementos $a, b \in M$: ou $a < b$, ou $a = b$, ou $b < a$. Esta propriedade é conhecida como *lei da tricotomia*; e
- b) Sejam $a, b, c \in M$. Se $a < b$ e $b < c$ então $a < c$. Esta propriedade é conhecida como *lei da transitividade*.

A existência de uma relação de ordem linear sobre um conjunto M permite que seus elementos sejam ordenados de forma unívoca. Além disso, para um conjunto finito, com pelo menos dois elementos, garante a existência de um único mínimo e de um único máximo.

Existe outro tipo de relação que ordena os elementos de um conjunto enumerável, só que de forma circular. Seja a relação definida da seguinte forma: $R \equiv$ é *sucessor imediato de*. Se ela puder ser aplicada a um conjunto enumerável $C = \{c_1, c_2, \dots, c_m\}$ de modo que, para cada elemento c_i de C exista um único elemento $c_j \in C$ tal que $(c_i, c_j) \in R$, tudo se passa como se os elementos desse conjunto estivessem distribuídos ao longo de uma circunferência. Definido o sentido de percurso dessa circunferência, define-se

quem é sucessor imediato de quem. Não se fala em maior, nem menor, muito menos em primeiro, ou último. Sabe-se qual elemento vem imediatamente após um determinado elemento, mas não existe um elemento inicial, nem um elemento final. Esse tipo de relação define uma ordem relativa entre os elementos de um conjunto e não uma ordem absoluta. Ocorre normalmente quando se trabalha com vetores na forma polar. A componente angular do vetor obedece a essa relação de ordem.

A Figura 2.4 mostra um exemplo de um conjunto imagem que obedece a relação de ordem circular. Para esse conjunto, pode-se afirmar que 315° sucede 270° e 0° sucede 315° . Esse conjunto imagem M pode ser ordenado segundo essa relação. Uma maneira de realizar isso é ilustrada nessa figura. Outra forma de ordenar seus elementos é considerar o valor 135° como valor inicial e seguí-lo dos próximos, sempre obedecendo a regra da relação "é sucessor imediato de". Desse modo, $M' = \{135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ, 0^\circ, 45^\circ, 90^\circ\}$. A rigor, existem oito representações ordenadas do conjunto M , cada qual iniciando com um de seus elementos. Qualquer uma delas tem seus elementos ordenados segundo a relação de ordem circular.



CÁLCULO DO HISTOGRAMA DE ÂNGULOS TANGENTES

- 1) Para cada ponto t do contorno da figura, calcule o ângulo que a tangente a este ponto faz com a horizontal.
- 2) Substitua o valor medido pelo ângulo mais próximo a ele pertencente ao conjunto M mostrado ao lado.
Para os pontos t_i e t_j destacados, os valores finais, após a substituição, serão 45° e 135° , respectivamente.
- 3) Construa o histograma totalizando para cada um dos oito ângulos pertencentes ao conjunto M , o número de pontos t do contorno que apresentem aquela medida quantizada.

Figura 2.4 - Cálculo do histograma de ângulos de tangentes.

Conforme mencionado anteriormente neste capítulo, a existência ou não da relação de ordem sobre o conjunto M , bem como o tipo de relação de ordem aplicável a esse conjunto, é o fator que determina o tipo de medida aplicável à característica em

questão. Pode-se distinguir três tipos de medidas diferentes que serão consideradas neste trabalho: as medidas nominais, as medidas ordinais e as medidas modulares.

As medidas nominais são aquelas em que o conjunto imagem M não apresenta uma relação de ordem. A característica cor no exemplo das bolas coloridas é uma medida nominal. Não é interessante definir uma relação de ordem entre as cores. Não há por que dizer que o azul é maior do que o amarelo, ou ainda, que o vermelho sucede o azul. Nesse exemplo, a única relação importante é a de *igualdade*. Azul é *igual a* azul; vermelho *não é igual a* azul. Um outro exemplo de medida nominal é o valor do voto em uma eleição. O voto pode ser dado somente a um candidato pertencente ao conjunto de todos os candidatos ao cargo disputado. Não se fala que o voto para o candidato c_1 é maior do que o voto para o candidato c_2 . A consequência mais importante desse fato é que o conceito de distância para medidas nominais é dicotômico: ou a medida é a mesma, ou a medida é diferente. A distância entre duas medidas nominais é binária e pode ser definida pela Equação (2.3).

$$d_{nom}(x, y) = \begin{cases} 0 & \text{se } x = y \\ 1 & \text{caso contrário} \end{cases} \quad (2.3)$$

A Equação (2.3) define uma distância. Cha e Srihari (CHA e SRIHARI, 2002) demonstraram detalhadamente que $d(x, y)$ obedece às seguintes propriedades:

- a) $d_{nom}(x, y) \geq 0$
- b) $d_{nom}(x, x) = 0$
- c) $d_{nom}(x, y) = d(y, x)$
- d) $d_{nom}(x, z) \leq d(x, y) + d(y, z)$

Ao aplicar o conceito de distância ao conjunto imagem das cores das bolas, a Tabela 2.2 mostra os resultados obtidos do cálculo da distância entre o elemento da linha e o elemento da coluna. Como a distância deve ser simétrica, a matriz é simétrica e como a

distância de um elemento a ele mesmo é nula pode-se observar nessa tabela que a diagonal principal da matriz também é nula.

Tabela 2.2 - Distâncias entre os elementos de M .

	amarelo	azul	vermelho
amarelo	0	1	1
azul	1	0	1
vermelho	1	1	0

O segundo tipo a ser explicado neste capítulo é a medida ordinal. Ela ocorre quando uma relação de ordem linear pode ser definida sobre o conjunto M . Isto significa que os elementos do conjunto imagem podem ser organizados sobre uma linha, existindo um primeiro e um último elemento no conjunto. Nesse caso, pode-se, falar em medida maior e medida menor; existe uma escala absoluta para a comparação de medidas. No exemplo das bolas coloridas, a medida de peso enquadra-se nessa categoria. O peso da bola b_2 é 27,989 g, maior do que o peso da bola b_3 que é 23,351 g. Comparando-se os pesos de b_2 e b_3 com o peso da bola b_1 , que é 12,475 g, pode-se afirmar que tanto b_2 quanto b_3 pesam mais do que b_1 . Mas pode-se ir além e afirmar que b_3 pesa mais do que b_1 , mas b_2 pesa *ainda mais*. Nesse caso, o conceito de distância não precisa restringir-se a apontar a diferença, mas pode ainda expressar o *tamanho* dessa diferença.

Realmente, a distância para um conjunto imagem linearmente ordenável passa a ser definida pela Equação (2.4).

$$d_{ord}(x, y) = |x - y| \quad (2.4)$$

Para as bolas coloridas b_1 , b_2 e b_3 , os cálculos da distância são:
 $d(b_1, b_2) = 15,514$ g; $d(b_1, b_3) = 10,876$ g e $d(b_2, b_3) = 4,638$ g. Pode-se observar claramente

que $d(b_1, b_2) > d(b_1, b_3)$ indicando que b_3 pesa mais do que b_1 , mas b_2 pesa *ainda mais*, como citado anteriormente, neste capítulo.

O terceiro tipo de medida explicado neste capítulo é a medida modular. Utiliza-se essa medida quando se pode aplicar ao conjunto imagem uma relação de ordem circular. Nesse caso, existe uma ordem relativa entre os elementos, mas não existe uma escala absoluta que permite definir que uma medida é maior do que outra. No entanto, um conceito de distância ainda pode ser definido. A Equação (2.5) apresenta esse conceito, que é explicado a seguir.

$$d_{\text{mod}}(x, y) = \begin{cases} |x - y| & \text{se } |x - y| \leq \frac{b}{2} \\ b - |x - y| & \text{caso contrário} \end{cases}$$

onde :

(2.5)

x, y : medidas cuja distância deseja -se calcular; e

b : comprimento total do ciclo.

Quando os elementos de um conjunto estão dispostos em uma linha, a distância entre eles é definida univocamente. Quando eles estão dispostos em um círculo, existem dois possíveis valores para a distância entre dois desses elementos. Isto pode ser observado na Figura 2.5, em que a distância externa, a maior, está ressaltada em cinza escuro e a distância interna, a menor, está ressaltada em preto. A distância modular é sempre a menor distância sobre a circunferência. Analisando-se a definição apresentada na Figura 2.5, pode-se concluir que a distância modular é sempre menor ou igual à metade da circunferência, ou seja, do ciclo completo que o conjunto dos valores determina.

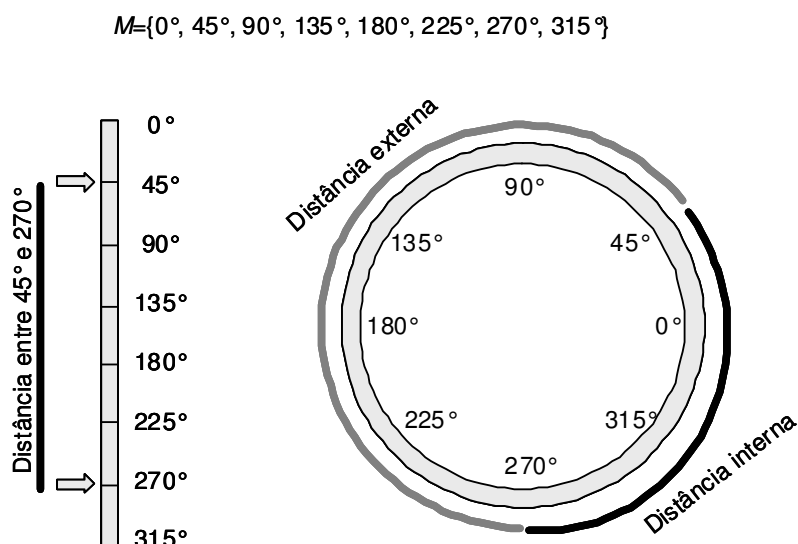


Figura 2.5 - Exemplo de cálculo da distância linear e circular entre os pontos 45° e 270°.

As medidas modulares são utilizadas principalmente quando se trabalha com ângulos pois eles possuem a natureza cíclica. Os histogramas de direção de gradiente ou os histogramas de direção da normal são exemplos clássicos dessa utilização.

2.4 Comparação entre Histogramas

A seção 2.3 deste capítulo mostrou detalhadamente como medir a distância entre duas medidas de uma mesma característica. Essa seção apresenta como medir a distância entre dois conjuntos de medidas de uma característica. De fato, o objetivo principal deste trabalho é analisar a semelhança entre duas imagens por meio de seus histogramas. Para saber se as imagens se parecem mais, ou menos, deve-se poder comparar seus histogramas e, dessa comparação, obter um valor que indique o grau dessa semelhança. Como um histograma é um conjunto de medidas, pode-se observar que medir a distância entre dois histogramas equivale a medir a distância entre dois conjuntos de medidas.

Existem várias maneiras diferentes de se medir a distância entre dois conjuntos. Entretanto, existem duas interpretações principais sobre o que o conjunto representa. A primeira interpretação trata os elementos do conjunto como as componentes de um vetor

multidimensional. A segunda interpretação considera cada elemento como uma medida de uma mesma característica. Cada uma dessas interpretações determina uma categoria específica de medidas que é adequada a ela.

Ao considerar o conjunto como um vetor, cada conjunto representa, na verdade, um elemento da população descrito por várias características diferentes. A distância entre dois conjuntos de características é a distância entre dois elementos da população, ou mais claramente, é a distância entre dois vetores em um espaço multidimensional. A álgebra linear ocupa-se de estudar espaços multidimensionais em que se define tanto produto interno quanto norma. Poole (POOLE, 2004) apresenta a definição de distância baseada na definição de norma, reproduzida na Equação (2.6).

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

onde: (2.6)

$\|\cdot\|$ - significa norma

Alguns tipos de norma para vetores definidos em espaços multidimensionais são apresentados nas Equações de (2.7) a (2.10).

$$\text{Seja um vetor } \mathbf{v} \in R^n \quad \mathbf{v} = [v_1, v_2, \dots, v_n]^T \quad (2.7)$$

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \dots + |v_n| \quad \text{norma squareblock, do taxista ou } L_1$$

$$\|\mathbf{v}\|_2 = \sqrt{|v_1|^2 + |v_2|^2 + \dots + |v_n|^2} \quad \text{norma euclidiana ou } L_2 \quad (2.8)$$

$$\|\mathbf{v}\|_\infty = \max\{|v_1|, |v_2|, \dots, |v_n|\} \quad \text{norma do máximo, uniforme ou norma } -\infty \quad (2.9)$$

Todas as normas mostradas nas Equações de (2.7) a (2.9) são casos especiais da norma de Minkowsky (L_p), mostrada na Equação (2.10). Elas são aplicáveis quando as componentes do vetor são descorrelacionadas.

$$\|\mathbf{v}\|_p = (v_1^p + v_2^p + \dots + v_n^p)^{1/p} \quad \text{norma de Minkowsky} \quad (2.10)$$

Ao contrário, quando as componentes do vetor apresentam correlação, ou quando se deve considerar uma escala diferente de ponderação para cada um delas, devem ser utilizadas normas estatísticas. A norma de Mahalanobis, mostrada na Equação (2.11), leva em conta a densidade de probabilidade conjunta entre as várias componentes de um vetor e é muito utilizada nas classificações baseadas na teoria Bayesiana. Nesse caso, supõe-se que as componentes obedecem à distribuição de probabilidade normal e que as classes ω apresentam probabilidade de ocorrência a priori $P(\omega)$ conhecida (DUDA *et alii*, 2000).

$$D(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

onde: (2.11)

$\mathbf{x}, \boldsymbol{\mu}$: vetores cuja distância deve ser calculada; e

$\boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{-1}$: matriz de covariância dessa classe e sua inversa.

Neste capítulo, as distâncias e normas mostradas nas Equações (2.6) a (2.11) apresentam uma característica comum bastante significativa para o problema em questão: a inversão da ordem das componentes v_i não altera o valor da norma. Essa é a propriedade da invariância ao embaralhamento (*shuffling invariance*) mencionada por Cha e Srihari (CHA e SRIHARI, 2002).

Em várias situações a invariância ao embaralhamento é uma característica desejável. Por exemplo, quando se trabalha com histogramas de medidas nominais isso é verdadeiro. Nesse caso, a comparação entre os níveis é binária – níveis iguais ou diferentes. A Figura 2.6 mostra um exemplo de dois conjuntos de bolas coloridas com o mesmo número total de bolas.

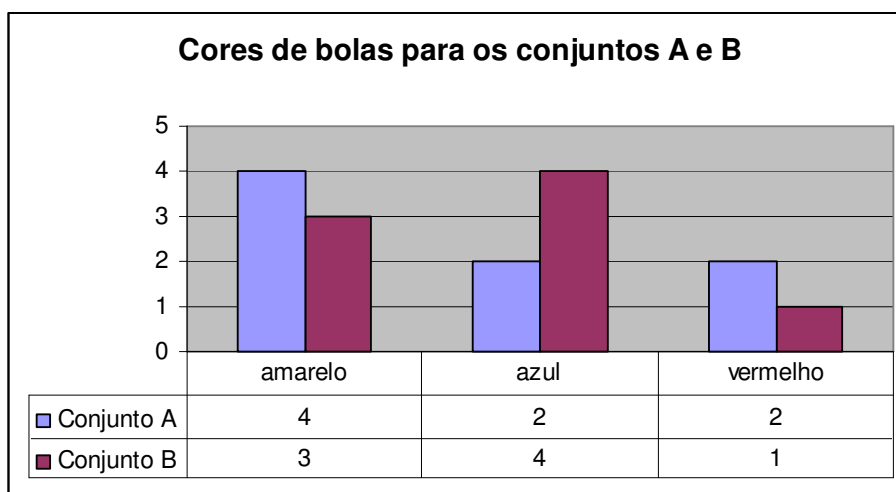


Figura 2.6 - Dois conjuntos de bolas coloridas e seus histogramas.

A distância entre os dois conjuntos da Figura 2.6 pode ser calculada pela distância L_1 . Aplicando a Equação (2.7), $(|4-3|+|2-4|+|2-1|)$ obtém-se a distância igual a 4. Alterar a ordem dos níveis considerados não altera o resultado, já que a adição é comutativa.

A aplicação da distância L_2 (distância Euclidiana) resulta em um valor diferente:

$$(|4-3|^2+|2-4|^2+|2-1|^2)^{1/2}=\sqrt{6} \approx 2,45.$$

A aplicação da distância L_∞ (distância uniforme) resulta no valor igual a 2.

Não importa qual dos cálculos de distância seja utilizado, a ordem dos níveis não altera o resultado final, já que todas apresentam a invariância ao embaralhamento mencionada anteriormente neste capítulo.

Se para as medidas nominais a propriedade da invariância ao embaralhamento é desejável, isso não é verdade, nem para as medidas ordinais, nem para as medidas

modulares. Isso porque os níveis vizinhos de uma medida ordinal ou modular guardam uma correlação maior do que os níveis mais separados, ou distantes. O cálculo da distância para as medidas ordinais e modulares, portanto, deve levar em conta esse fato para ser mais representativo. Essa é a situação em que o conjunto de medidas pode ser considerado como uma distribuição e não como um vetor.

Utilizando-se ainda o exemplo das bolas coloridas, agora considerando-se como referência seu peso e não sua cor, considere três conjuntos de bolas cujos pesos são dados pelos conjuntos $A=\{15, 15, 15, 15, 15\}$, $B=\{25, 25, 25, 25, 25\}$ e $C=\{35, 35, 35, 35, 35\}$. Os histogramas correspondentes são $H_A=[5, 0, 0]$, $H_B=[0, 5, 0]$ e $H_C=[0, 0, 5]$, como mostrado na Figura 2.7.

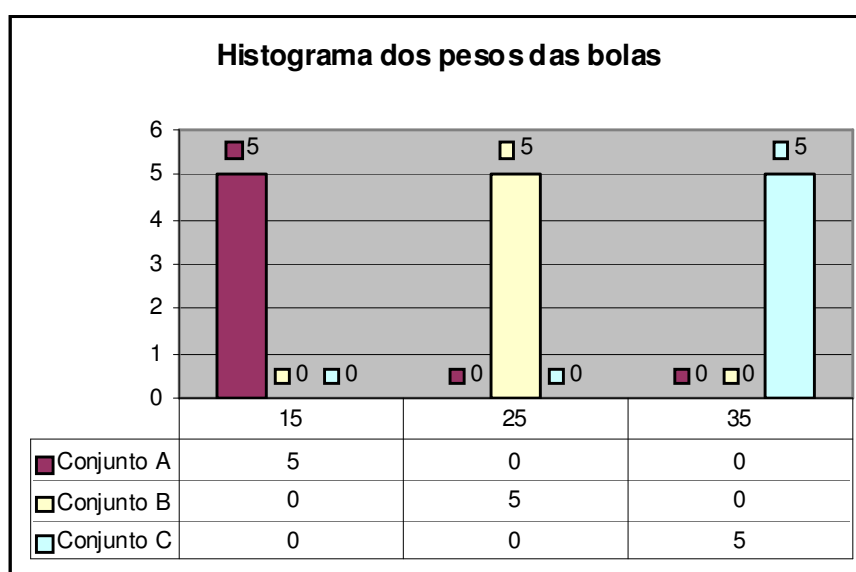


Figura 2.7 - Histograma de peso de bolas para três conjuntos A, B e C.

Utilizar qualquer das distâncias testadas para o caso da medida nominal, anteriormente apresentada neste capítulo, resulta em uma conclusão inusitada: o cálculo das distâncias entre os histogramas H_A , H_B e H_C , considerados dois a dois, leva ao mesmo resultado. Assim, $d(H_A, H_B) = 5$, $d(H_B, H_C) = 5$ e $d(H_A, H_C) = 5$ se a distância L_1 for utilizada. Isso conduz à seguinte afirmação: os histogramas H_A , H_B e H_C são eqüidistantes. Essa conclusão não condiz com a realidade. Se as distâncias entre H_A e H_B , e entre H_B e H_C

parecem iguais, a distância entre H_A e H_C parece ser bem maior. De fato, o peso total do conjunto A é 45, do conjunto B é 75 e do conjunto C é 105. Sob esse aspecto, $d_T(H_A, H_B) = 30$, $d_T(H_B, H_C) = 30$ e $d_T(H_A, H_C) = 60$. Esse resultado reflete melhor a realidade do que o resultado obtido quando foram utilizadas as distâncias invariantes ao embaralhamento: L_1 e L_2 .

Duas medidas de distância são bastante adequadas para o problema proposto. São elas a distância euclidiana cumulativa (DEC), e a distância *land mover* (DLM). Ambas levam em consideração a distribuição dos valores (o histograma) e não apenas as componentes isoladamente.

As Equações (2.12) e (2.13) mostram o cálculo das distâncias DEC e DML, respectivamente.

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=0}^{b-1} \left(\sum_{j=0}^i u_j - v_j \right)^2} \quad \text{Distância Euclidiana Cumulativa (DEC)} \quad (2.12)$$

$$d(\mathbf{u}, \mathbf{v}) = \sum_{i=0}^{b-1} \left| \sum_{j=0}^i u_j - v_j \right| \quad \text{Distância Land Mover (DLM)} \quad (2.13)$$

Pode-se verificar nas Equações (2.12) e (2.13) que, a cada passo da somatória externa, acumulam-se as diferenças adjacentes na somatória interna. Isso faz com que a ordem das componentes influa no resultado final. A troca de posição entre duas componentes pode alterar o resultado final, como desejado.

Aplicando a DLM ao exemplo apresentado na Figura 2.7, obtém-se os seguintes resultados:

$$d(H_A, H_B) = |5-0| + |5-0 + 0-5| + |5-0 + 0-5 + 0-0| = 5$$

$$d(H_B, H_C) = |0-0| + |0-0 + 5-0| + |0-0 + 5-0 + 0-5| = 5$$

$$d(H_A, H_C) = |5-0| + |5-0 + 0-0| + |5-0 + 0-0 + 0-5| = 10$$

Esses resultados são compatíveis com o que se esperava e estão em consonância com a realidade. Para perceber que a troca de posição entre os componentes afeta o resultado, considere os histogramas H_A e H_C , trocando de lugar o primeiro e o segundo componentes. Os histogramas tornam-se $H_A' = [0, 5, 0]$ e $H_C' = [0, 0, 5]$. A nova distância entre H_A e H_C torna-se $d(H_A', H_C') = 5$, diferente do valor anterior, que era 10.

Visando uniformizar uma definição de distância para histogramas de medida nominal, ordinal e modular, Cha e Srihari (CHA e SRIHARI, 2002) propuseram a diferença mínima entre as combinações de pares (*minimum difference of pair assignments*) que é obtida pela Equação (2.14):

$$D(A, B) = \min_{A, B} \left(\sum_{i, j=0}^{n-1} d(a_i, b_i) \right)$$

onde :

(2.14)

$D(A, B)$: diferença entre os conjuntos A e B; e

$d(a_i, b_i)$: diferença entre os elementos dos conjuntos A e B.

A distância $d(a_i, b_i)$ é $d_{nom}(a_i, b_i)$, $d_{ord}(a_i, b_i)$, ou $d_{mod}(a_i, b_i)$ conforme a medida aplicável aos conjuntos seja nominal, ordinal ou modular.

A Equação (2.14) define a distância entre dois conjuntos A e B como o valor da soma das distâncias entre os pares de elementos para a combinação que retorne o menor desses valores. Explicando mais claramente, tudo se passa como se fossem exaustivamente testadas todas as possíveis combinações entre os pares de elementos do conjunto A e do conjunto B. Para cada combinação, acumulam-se as distâncias entre os elementos dos pares. A distância entre os conjuntos é a menor das distâncias acumuladas para cada combinação.

Para exemplificar, considere os conjuntos $A = \{1, 3, 4, 8\}$ e $B = \{1, 3, 3, 5\}$ e a medida ordinal. Segundo a definição, montam-se os pares para a primeira combinação.

$C_1 = \{(1,1), (3,3), (4,3), (8,5)\}$. $D_1(A,B) = 0+0+1+3 = 4$. A segunda combinação resulta $C_2 = \{(1,1), (3,3), (4,5), (8,3)\}$. $D_1(A,B) = 0+0+1+5 = 6$. Todas as 24 combinações possíveis devem ser testadas. Uma delas produzirá o menor valor para a soma das distâncias entre os pares, que é a distância entre A e B. O resultado é 4, a distância calculada para a combinação C_1 .

Para conjuntos grandes, observa-se que a aplicação da definição de $D(A,B)$ é inviável. Por exemplo, conjuntos com 10 elementos produzem $10! = 3.638.800$ combinações diferentes que devem ser testadas. Esse algoritmo tem ordem $O(n!)$ o que torna impossível sua aplicação para valores grandes de n . Cha e Srihari (CHA e SRIHARI, 2002) apresentaram algoritmos eficientes que podem ser executados em ordem $O(n)$ para as medidas nominais e ordinais e em ordem $O(n^2)$, no pior caso, para medidas modulares.

Para histogramas de medida nominal, a solução elegante utiliza a distância baseada na não interseção, obtida pela Equação (2.15).

$$D_{nom}(A,B) = n - \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) \quad (2.15)$$

A somatória na Equação (2.15) equivale à sobreposição mínima entre os histogramas. Considera-se que ambos os conjuntos possuem o mesmo número n de elementos. Se isso não ocorrer, normalizam-se os histogramas para que isso ocorra. A distância entre os conjuntos A e B pode ser definida como a área de não sobreposição entre os conjuntos. A prova de que a distância $D_{nom}(A,B)$ é igual à metade da distância L_1 é apresentada na Equação (2.16).

$$n - \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = \frac{1}{2} \sum_{i=0}^{b-1} |H_i(A) - H_i(B)|$$

Como :

$$\min(H_i(A), H_i(B)) = \frac{1}{2} [H_i(A) + H_i(B) - |H_i(A) - H_i(B)|]$$

$$\sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = \frac{1}{2} \sum_{i=0}^{b-1} [H_i(A) + H_i(B) - |H_i(A) - H_i(B)|]$$

$$2 \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = \sum_{i=0}^{b-1} H_i(A) + \sum_{i=0}^{b-1} H_i(B) - \sum_{i=0}^{b-1} |H_i(A) - H_i(B)|$$

$$2 \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = 2n - \sum_{i=0}^{b-1} |H_i(A) - H_i(B)|$$

$$2n - 2 \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = \sum_{i=0}^{b-1} |H_i(A) - H_i(B)| \quad (2.16)$$

Portanto :

$$n - \sum_{i=0}^{b-1} \min(H_i(A), H_i(B)) = \frac{1}{2} \sum_{i=0}^{b-1} |H_i(A) - H_i(B)|$$

A distância entre dois conjuntos de medida nominal pode ser calculada como a metade da distância L_1 de seus histogramas. Esse é, claramente, um cálculo de ordem $O(n)$.

Para histogramas de medida ordinal, o algoritmo apresentado por Cha e Srihari (CHA e SRIHARI, 2002) baseia-se no algoritmo de remoção de terra (*land mover*, devido a isso o nome da distância DLM). Este algoritmo compara dois histogramas baseado no trabalho necessário para transformar um em outro. Essa analogia supõe dois histogramas normalizados de conjuntos de n elementos. Considera que o primeiro histograma seja uma distribuição linear de terra e o segundo histograma seja uma distribuição linear de "buracos". Como os histogramas são normalizados, toda a terra do primeiro cabe nos buracos do segundo. A distância entre eles é obtida pelo esforço necessário para mover a

terra do primeiro para os buracos do segundo. Mover uma unidade de terra entre níveis adjacentes custa uma unidade de esforço. Movê-la por dois níveis custa duas unidades de esforço e assim por diante.

Como exemplo, pode-se calcular a distância entre os histogramas $H_A = [5, 2, 1, 0, 2, 0, 0, 0]$ e $H_B = [0, 0, 1, 1, 0, 1, 3, 4]$ supondo medida ordinal. Ambos têm 10 unidades e oito níveis. Supõe-se, sem perda de generalidade, os níveis 0, 1, 2, ..., 7. Pode-se imaginar que cada histograma seja uma bancada com oito posições adjacentes. Cada uma delas tem uma quantidade de blocos empilhados, como mostra a Figura 2.8. Para transformar H_A em H_B , devem-se movimentar os blocos pelas posições adjacentes e contabilizar o esforço, até que se consiga montar H_B . Nesse caso, mover um bloco, uma posição à direita, custa uma unidade de esforço. Mover um bloco uma unidade à esquerda custa uma unidade negativa de esforço. Pode-se, dessa forma, contabilizar o número de esforços positivos e negativos efetuados em cada fronteira de posição.

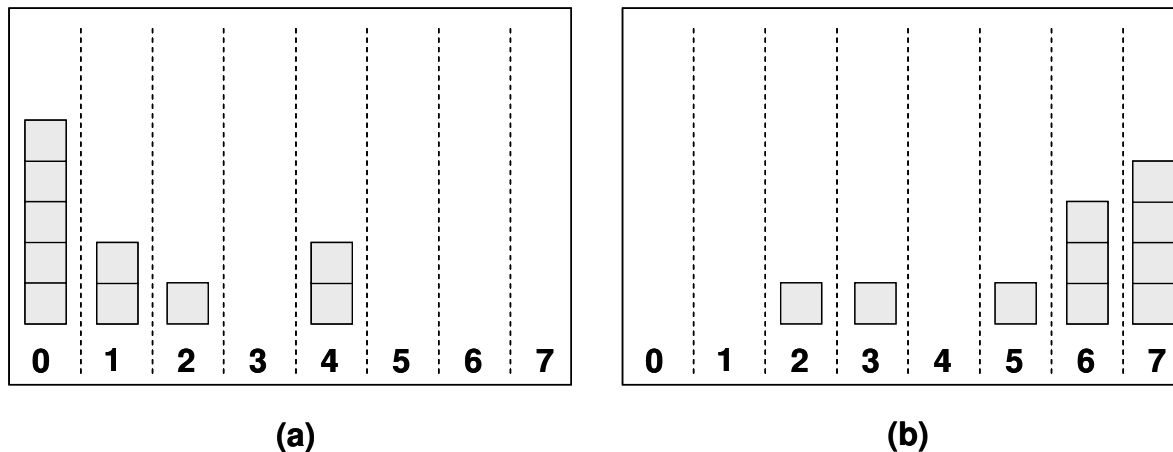


Figura 2.8 - Representação pictórica dos histogramas H_A e H_B .

O esforço total é obtido pela soma líquida dos esforços de fronteira, como mostram as duas possibilidades ilustradas na Figura 2.9. Nessa figura, as setas representam os esforços necessários para mover os blocos. Esforços de sinais opostos se anulam e de mesmo sinal se somam.

A Figura 2.9(a) apresenta um movimento que utiliza esforços positivos e negativos. Para totalizar o esforço, basta que se totalizem os esforços em cada fronteira e depois acumulem-se os resultados. Nesse caso, a distância soma $D(H_A, H_B) = 5+7+7+6+8+7+4 = 44$. A Figura 2.9(b) apresenta outro movimento, agora sem esforços negativos. Pode-se verificar que, apesar de diferente, esse movimento resulta no mesmo valor para a distância. Essa distância pode ser calculada na ordem $O(n)$ pela aplicação da distância DLM, que é uma distância cumulativa.

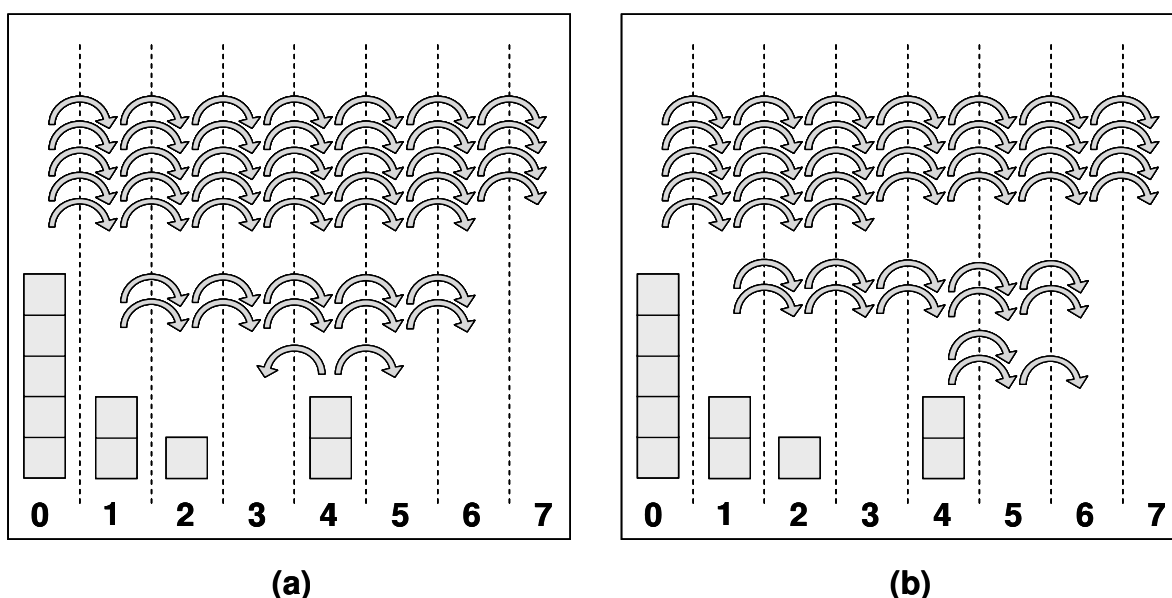


Figura 2.9 - Duas formas de movimentação de blocos para transformar o histograma H_A no histograma H_B .

Finalmente, o algoritmo para calcular a distância modular assemelha-se bastante ao DLM, apenas adaptado para levar em consideração os atalhos característicos da medida modular. Todo movimento maior do que meio ciclo pode ser substituído por seu suplemento para 360° . Os exemplos da Figura 2.10 (a) e (b) ilustram dois histogramas modulares. A tarefa é calcular a distância entre eles. Isso pode ser resolvido movendo-se os blocos do histograma H_A ao longo da circunferência e contabilizando-se o esforço total, exatamente da mesma forma que se fez no caso da distância ordinal, considerando-se, porém, que o histograma não possui extremidades.

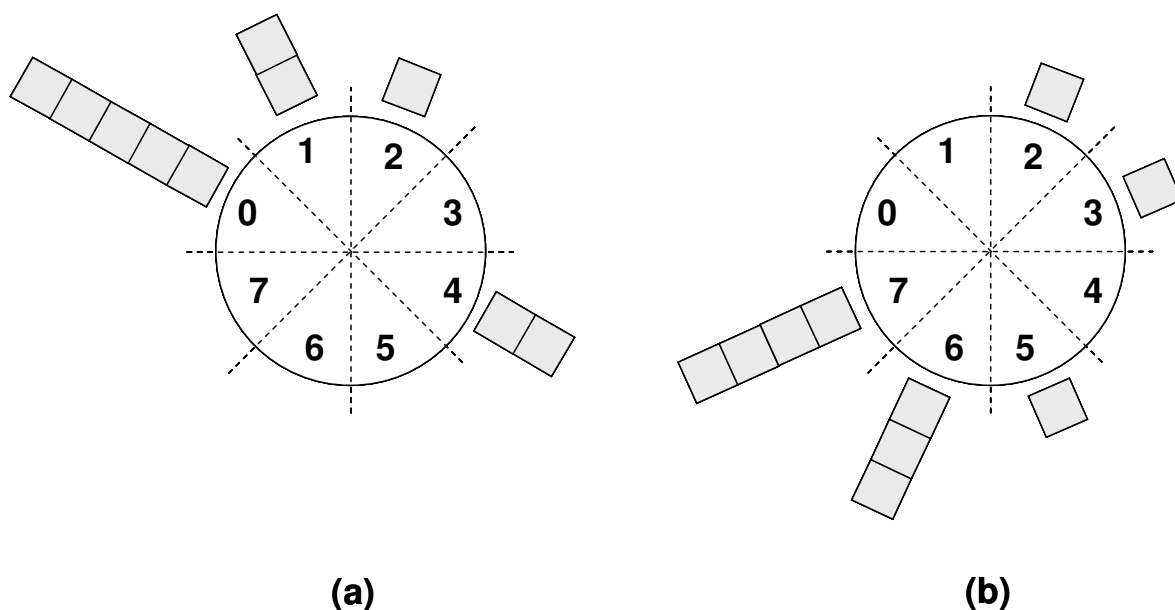


Figura 2.10 - Representação de dois histogramas de medida modular.

A distância modular é calculada quando, ao final dos movimentos, obtém-se o histograma H_B . O cálculo dessa distância é realizado contabilizando-se o esforço líquido nas fronteiras entre os níveis.

O princípio dos esforços que se cancelam também é aplicável no cálculo da distância modular. As letras (a) e (b) da Figura 2.11 mostram um esforço mínimo e um movimento que resulta em uma distância maior, respectivamente. No entanto, deve-se ressaltar a possibilidade de se adicionar um ciclo completo ao diagrama sem que as posições se alterem. O exemplo disso é mostrado na Figura 2.11 (c). Sem grande esforço, pode-se mostrar que os movimentos das letras (a) e (b) da Figura 2.11 resultam no mesmo histograma. Para isso, basta adicionar ao diagrama de movimentos da letra (b) dessa figura, sete ciclos completos no sentido anti-horário. Transforma-se o Movimento 2 no Movimento 1, que é mínimo. O algoritmo que permite chegar a esse resultado apresenta ordem $O(n^2)$. Outra observação notável é que, também no cálculo da distância modular, a alteração de posições de componentes altera o resultado final.

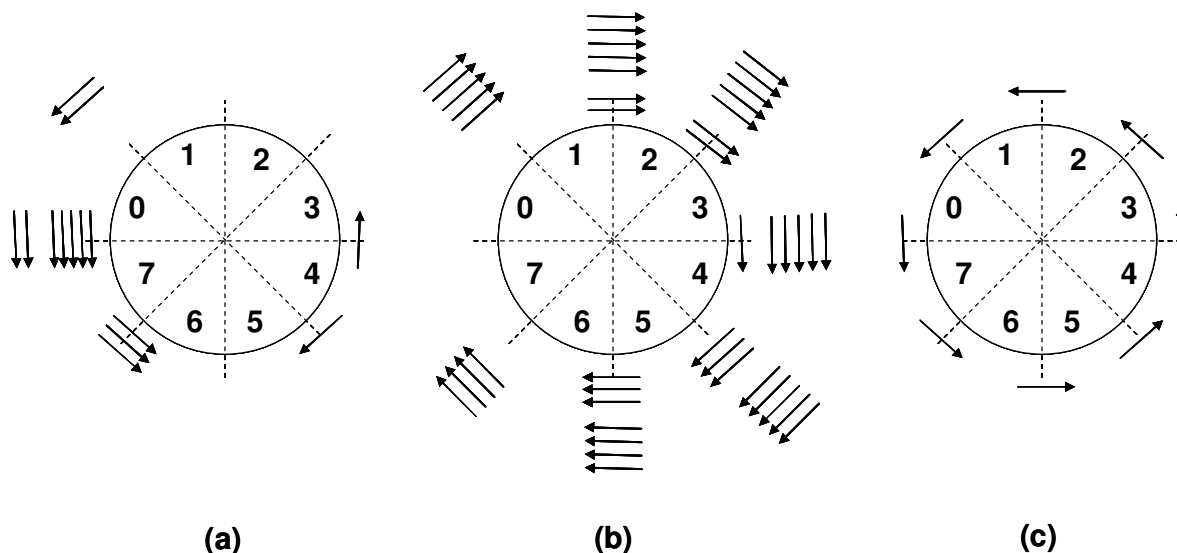


Figura 2.11 - Movimentos para transformar H_A em H_B , considerando a medida modular. (a) Movimento 1, mínimo; (b) Movimento 2, não mínimo; (c) Um ciclo completo.

2.5 Projeções Suavizantes

Histogramas de características de imagens possuem, geralmente, um grande número de níveis. Isso faz com que as comparações entre eles requeiram um grande número de operações. Além disso, quando são utilizadas técnicas estatísticas, pode surgir o problema conhecido como a maldição da dimensionalidade (*dimensionality curse*) descrito detalhadamente por Duda e outros (DUDA *et alii*, 2000). Esta seção aborda esses problemas, bem como uma técnica utilizada quer seja para solucioná-los ou então para minimizar seus efeitos.

Para se utilizar a métrica de Mahalanobis, é necessário o cálculo da inversa da matriz de covariância dos níveis do histograma, como mostrado na Equação (2.11). Cada nível do histograma equivale a uma dimensão na tarefa de classificação, ou na determinação da distância. A matriz de covariância mede a correlação entre todos os pares de níveis dos histogramas de uma classe. Essa matriz é estimada utilizando-se um determinado número de histogramas como amostras de uma determinada classe. Quanto maior o número de histogramas-amostra, mais precisa será a estimativa e melhor será o resultado no processo de classificação.

Na prática, o aumento significativo do número de dimensões no processo de classificação tende a diminuir a precisão dos resultados. Os classificadores bayesianos conseguem aprimorar seu desempenho quando é adicionada às dimensões existentes uma nova característica que discrimine detalhes não considerados pelas demais. Isso é verdade tanto teoricamente quanto nas implementações práticas desses classificadores. Quando o número de dimensões torna-se razoavelmente grande, pode-se observar um comportamento inverso do esperado: a adição de uma nova característica pode levar à piora do desempenho do classificador. Isso pode ser devido à precisão numérica dos cálculos envolvidos, ou à subestimativa dos elementos da matriz de covariância. Quanto maior o número de dimensões envolvidas, mais valores são necessários para estimar a matriz. Nem sempre é possível dispor do número apropriado de valores para essa estimativa.

Outro ponto importante a ser considerado é que o elevado número de dimensões pode ocultar padrões globais do histograma, que ficam disfarçados entre o grande número de detalhes locais. Isso pode ser observado nos histogramas das letras (a) e (b) da Figura 2.12. Ambos representam o mesmo conjunto, com a diferença que o primeiro foi quantizado em níveis com a faixa de largura de uma unidade e o segundo com a faixa de largura dez vezes maior. Portanto, a resolução do primeiro é dez vezes maior do que a resolução do segundo. Pode-se observar uma taxa de variação muito alta na Figura 2.12(a), bastante semelhante a um ruído sobreposto a um sinal. Já na Figura 2.12(b), tudo se passa como se o ruído tivesse sido suprimido. Esse efeito é causado pela suavização da média e equivale a uma filtragem passa-baixas. Pode-se concluir disso que, se por um lado, o excesso de detalhes pode enriquecer o processo, por outro lado, pode encobrir um padrão importante para a classificação.

Duda (DUDA *et alii*, 2000) explicou o efeito mencionado no parágrafo anterior com uma analogia bastante ilustrativa, a interpolação de um conjunto de n pontos no plano por um polinômio. Uma maneira de resolver esse problema é encontrar um polinômio de ordem n que passe por todos os pontos. O problema se reduz à resolução de um sistema

linear de n equações e n incógnitas. O resultado é uma curva semelhante à curva contínua da Figura 2.13.

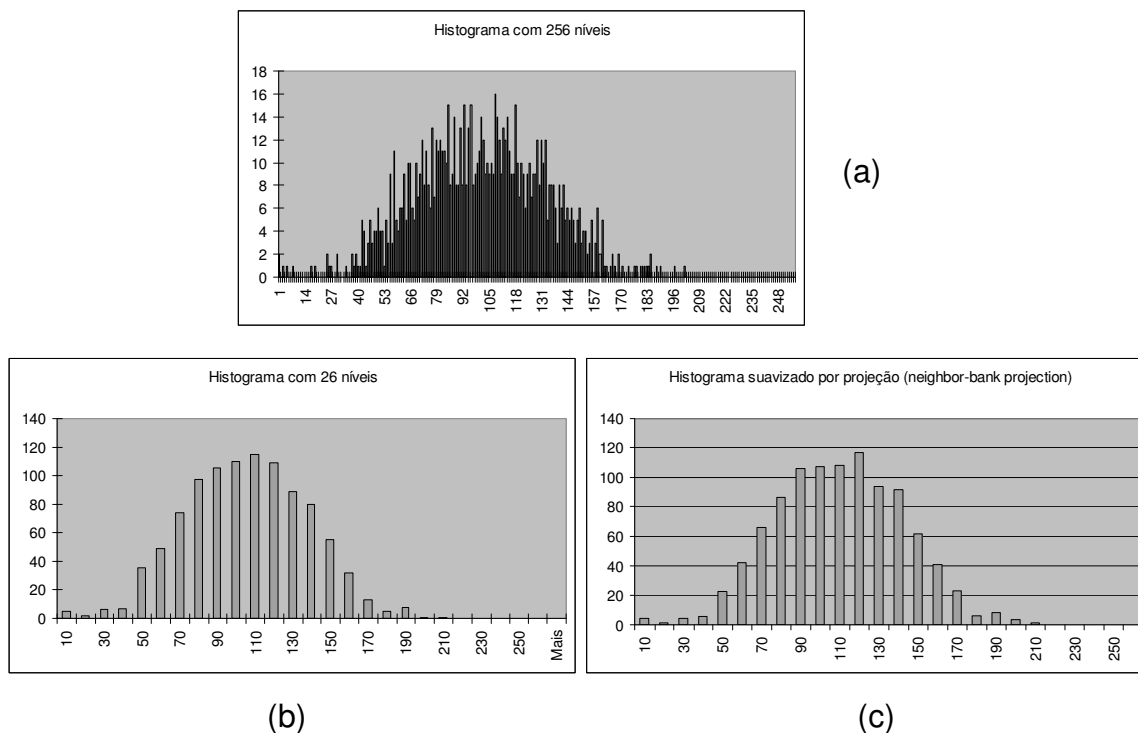


Figura 2.12 - Histogramas de uma grandeza hipotética, calculados para diferentes número de níveis. (a) Histograma com 256 níveis; (b) histograma com 26 níveis; (c) histograma suavizado por projeção de bancos vizinhos.

Outra maneira de resolver o problema da interpolação de um conjunto de n pontos é encontrar um polinômio de grau menor do que n que melhor se encaixe no conjunto de pontos. Isso pode ser feito pela minimização do erro médio quadrático, obtendo-se como resultado algo semelhante à curva tracejada da Figura 2.13.

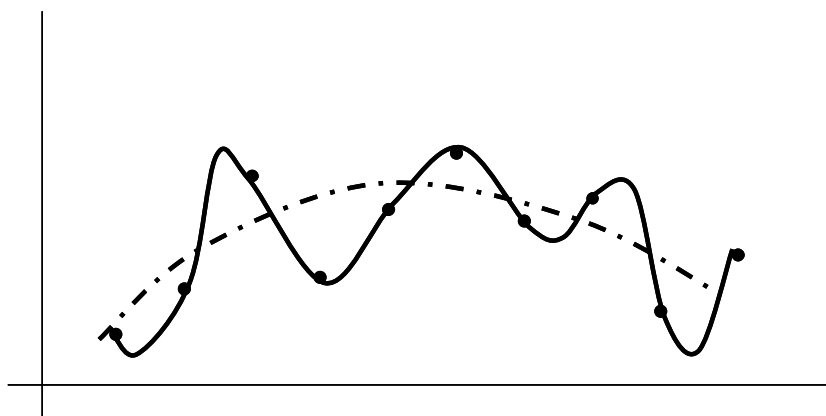


Figura 2.13 - Resultado da aplicação de duas técnicas para aproximar um conjunto de pontos por uma curva. Curva contínua: interpolação. Curva tracejada: aproximação por polinômio de grau menor.

Pode-se observar na Figura 2.13 que a aplicação da primeira técnica (interpolação) produz uma curva bastante complexa, enquanto que a aplicação da segunda (minimização do erro médio quadrático) resulta em uma curva simples que melhor representa o conjunto de pontos. Quando se conhece a priori a forma que a curva deve obedecer — uma reta ou uma parábola, na maioria dos casos — a segunda técnica é a mais apropriada. Na realidade, utilizar a interpolação de n -ésimo grau distorce o resultado, levando a uma conclusão errônea.

Da mesma forma, no domínio da classificação bayesiana, nem sempre superestimar a dimensionalidade conduz a resultados melhores. Por vezes é aconselhável uma redução do número de características. Isso faz com que a complexidade do classificador diminua, o tempo de classificação seja reduzido e o índice de acertos aumente.

Várias técnicas podem ser utilizadas para reduzir a dimensionalidade no processo de classificação, entre elas a análise dos componentes principais (*Principal Component Analysis: PCA*), e a análise de múltiplos discriminantes (*Multiple Discriminant Analysis: MDA*).

A técnica PCA, como o próprio nome indica, busca determinar quais são os principais componentes — aquelas dimensões que concentram a maior energia na representação do sinal. Essa técnica permite que uma amostra seja representada por um número menor de componentes desconsiderando-se aqueles componentes que carregam

uma menor quantidade de informação, ou energia, conforme é o caso. A finalidade dessa técnica é encontrar um conjunto de autovetores para os maiores autovalores de uma matriz de representação. Os autovetores formam uma base para a decomposição da amostra e a solução consiste em desconsiderar os autovetores correspondentes aos autovalores de menor valor. Dessa forma, reduz-se a dimensão de representação sem que se note uma distorção sensível no resultado final.

A técnica PCA é bastante semelhante à transformada de Hotelling (GONZALEZ e WOODS, 2000). Essa transformada pode ser utilizada com sucesso em problemas de compressão de imagem.

Ao contrário, em sistemas de classificação, a otimização da base de representação da amostra não necessariamente conduz a melhores resultados. Duda (DUDA *et alii*, 2000) enfatizou que, apesar da PCA encontrar componentes que são úteis para a representação dos dados, não existe razão para presumir que esses componentes devam ser úteis para a discriminação de dados entre classes diferentes. Se todas as amostras forem consideradas, as direções desconsideradas por ela podem ser exatamente as direções necessárias para distinguir essas classes. Duda segue ilustrando seu argumento com um exemplo de classificador que reconhece letras maiúsculas impressas a partir de um determinado conjunto de dados. Ao se utilizar a técnica PCA com dados das letras O e Q, as características mais importantes para caracterizar essas letras seriam determinadas, mas a pequena cauda que distingue um O de um Q poderia ser desconsiderada por não conter uma quantidade de informação significativa. Enquanto PCA busca por direções que sejam eficientes para a representação, a análise de discriminantes busca por direções que sejam eficientes para a discriminação.

Ao invés de buscar uma maior compactação na representação, a técnica MDA, procura por uma base de menor dimensão que seja mais apropriada para discriminar as diferentes classes que compõem o problema. O objetivo dessa técnica é encontrar um posicionamento da base que consiga maximizar as distâncias entre os centróides das classes, duas a duas, permitindo uma maior seletividade, ou separabilidade entre elas.

Ambas as técnicas, PCA e MDA, são gerais e seus cálculos são complexos, nem sempre conduzindo a resultados ótimos no sentido de desempenho, velocidade, ou espaço de armazenamento necessário à sua implementação.

Uma proposta alternativa para reduzir o número de dimensões nos problemas que utilizam histogramas foi apresentada por Kamarainen e outros (KAMARAINEN *et alii*, 2003). A projeção suavizante de um histograma em um espaço com dimensão menor é realizada pela utilização de bancos de vizinhos (*neighbor-bank*). A justificativa desse nome é porque níveis vizinhos do histograma são combinados com a finalidade de reduzir a dimensão final do histograma. Cada banco de vizinhos do histograma original produz um nível do histograma suavizado.

A suavização do histograma é realizada multiplicando-se os níveis do banco por uma janela que pode ser triangular ou cosseno quadrado levantado. Pode-se verificar que essa operação é muito semelhante à quantização com um número menor de níveis. Isso pode ser observado nas letras (b) e (c) da Figura 2.12. A Figura 2.12(b) mostra o efeito da quantização para um número reduzido de níveis. A Figura 2.12(c) ilustra a suavização do histograma utilizando banco de vizinhos com janela cosseno levantado. Pequenas diferenças nessas figuras podem ser observadas, principalmente devido à ponderação dos níveis dentro de cada banco. Os bancos se entrelaçam com os bancos vizinhos de forma a conseguir uma suavização apropriada no conjunto de dados.

A Figura 2.14 mostra quatro bancos de ponderação cosseno quadrado levantado. A cor do primeiro banco é branca, a do segundo é preta, a do terceiro é cinza e a do quarto banco é branca. Pode-se observar o entrelaçamento dos bancos vizinhos que fornece a suavização desejada. Pode-se verificar também o formato do banco que atribui mais peso aos pontos centrais e um peso menor aos pontos dos bancos vizinhos.

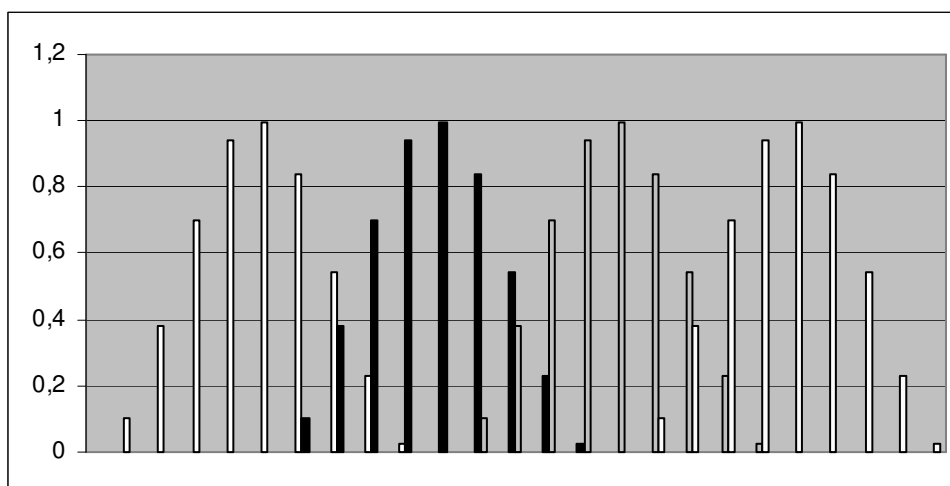


Figura 2.14 - Seqüência de quatro bancos utilizando cosseno quadrado levantado.

Uma questão importante é determinar o motivo por que a suavização do histograma utilizando bancos de vizinhos, que é um procedimento bastante simples, funciona melhor que as outras técnicas mais complexas de redução de dimensionalidade apresentadas (PCA e MDA). Uma boa explicação é porque essa suavização é aplicada a histogramas ordinais ou modulares. Os níveis vizinhos desses histogramas apresentam maior correlação quanto maior é a proximidade entre eles. Isso pode ser observado nos histogramas de níveis de cinza ou nos histogramas de ângulo de gradiente de imagens naturais. Utilizar bancos de vizinhos para reduzir a dimensionalidade, ainda mais quando se ponderam os níveis vizinhos inversamente com a proximidade ao nível central, é um método empírico, com base na teoria de análise de componentes apresentada.

Kamarainen e outros (KAMARAINEN *et alii*, 2003) avaliaram a acurácia da classificação em diversas situações, medindo o desempenho para diferentes métricas, para diferentes classificadores, para várias reduções na dimensionalidade e para quatro tipos de bancos de imagem. Eles concluíram que o método é bastante eficaz para a utilização com classificadores estatísticos, melhorando o desempenho da classificação para as métricas mais utilizadas, com ênfase na distância de Mahalanobis. Mostraram ainda que uma redução considerável na dimensionalidade do problema pode ser obtida, com exatidão da classificação ainda superior àquela obtida com a utilização do histograma original.

O método de projeções suavizantes proposto por Kamarainen e outros (KAMARAINEN *et alii*, 2003) pode ser utilizado para a melhoria da medição da similaridade entre histogramas ordinais e modulares. Por ser uma técnica de simples aplicação e por proporcionar uma redução significativa no número de dimensões, permite tanto uma redução no espaço de armazenamento, quanto uma redução no tempo de processamento da comparação entre os histogramas.

2.6 Assinaturas

Algumas aplicações exigem a utilização de histogramas esparsos. Esses histogramas apresentam grande número de níveis sendo que apenas alguns deles possuem um valor diferente de zero. Esse grande número de níveis necessita de um maior espaço de armazenamento e de maior tempo de processamento. Para superar essas necessidades, Serratosa e Sanfeliu (SERATOSA e SANFELIU, 2006) propuseram uma nova estrutura de dados que denominaram de assinatura.

Assinaturas são representações equivalentes de histogramas. Isso significa que é possível representar qualquer histograma por uma assinatura, bem como obter o histograma original a partir de sua assinatura equivalente. Os objetivos dessa representação são reduzir o espaço necessário para o armazenamento do histograma esparsos e também processar de maneira mais rápida a comparação entre os histogramas. Para isso, na assinatura, os níveis que possuam valor igual a zero não são representados explicitamente. Em contrapartida, os níveis diferentes de zero são representados pelo par (*nível*, *valor*), em que o *nível* é o índice do nível representado e o *valor* é o valor do nível correspondente.

Serratosa e Sanfeliu (SERATOSA e SANFELIU, 2006) definiram assinatura formalmente como se segue. Sejam $H(A) = [H_1(A), \dots, H_T(A)]$ e $S(A) = [S_1(A), \dots, S_Z(A)]$ o histograma e a assinatura do conjunto A, respectivamente. Cada componente $S_k(A)$, em que $1 \leq k \leq Z \leq T$, é definido como $S_k(A) = (w_k, m_k)$. Nesse par ordenado, se $w_k = i$, então $m_k = H_i(A)$, ou seja, w_k é o nível e m_k é o valor associado ao nível. Resumidamente,

$S_k(A) = (w_k, m_k)$, onde $m_k = H_{w_k}(A)$. Duas restrições importantes completam a definição: a) $w_k < w_t \Leftrightarrow k < t$; b) $m_k > 0$ para todo valor de k . A primeira impõe uma ordenação aos elementos da assinatura e a segunda assegura que a representação é mínima.

A Figura 2.15 ilustra a representação da assinatura de um histograma de um conjunto de dados. O conjunto original A é apresentado juntamente com o domínio dos valores de seus elementos, o conjunto X . A partir disso, o histograma pode ser calculado e é mostrado na Figura 2.15(a). O histograma resultante é uma tupla ordenada composta de oito elementos. Pode-se observar nesse histograma que os níveis 3, 5, 6 e 7 possuem valor zero. O histograma resultante é $H = \{5, 2, 1, 0, 2, 0, 0, 0\}$. Com o objetivo de compactar a representação sem perda de informação, suprime-se a listagem dos níveis com valor igual a zero. Para isso, deve-se adicionar a cada informação de valor, o nível a que ela corresponde. O exemplo acima não traz benefício aparente por conter um número pequeno de valores nulos. Ao contrário, no caso dos histogramas esparsos, a representação por assinatura é muito vantajosa.

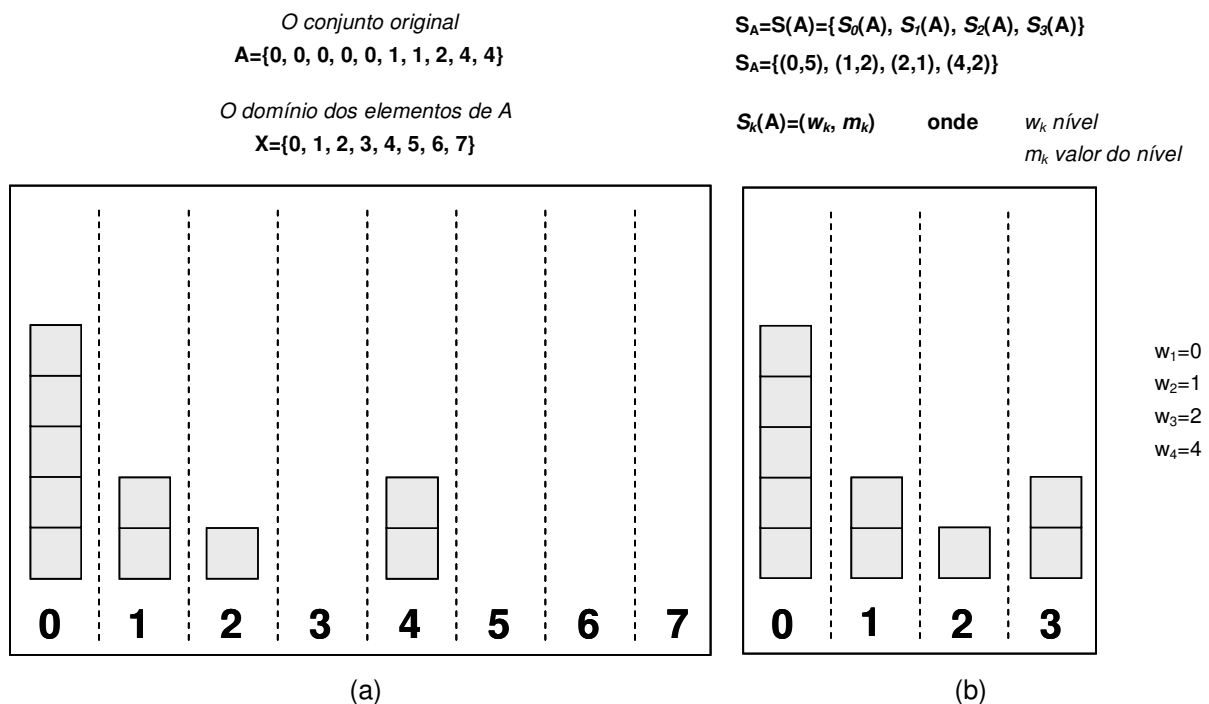


Figura 2.15 - (a) Histograma H_A de um conjunto A ; (b) assinatura S_A desse histograma.

Pode-se verificar que a reconstrução do histograma a partir da assinatura e do conjunto domínio é direta. Basta inicializar com zero uma tupla ordenada correspondente ao domínio e, para cada elemento $S_k(A) = (w_k, m_k)$ da assinatura, atribuir ao correspondente w_k do histograma o valor m_k . Seguindo o exemplo, inicializa-se $H = \{0,0,0,0,0,0,0,0\}$. Para o primeiro elemento da assinatura, $S_0 = (0,5)$, o valor 5 é atribuído ao nível de índice 0 para obter $H = \{5,0,0,0,0,0,0,0\}$. Repete-se o procedimento para $S_1 = (1,2)$ resultando $H = \{5,2,0,0,0,0,0,0\}$ e assim por diante, para todos os demais elementos da assinatura, resulta em $H = \{5,2,1,0,2,0,0,0\}$ que é o histograma original.

Serratos e Sanfeliu (SERATOSA e SANFELIU, 2006), após provarem a adequação de sua proposta para as medidas nominal, ordinal e modular, apresentaram os algoritmos que fazem a comparação dos histogramas para cada um desses tipos. O conceito chave para o entendimento desses algoritmos é a assinatura estendida.

A assinatura estendida é uma assinatura na qual foi adicionado um número mínimo de níveis vazios (que possuem valor zero) com a finalidade de assegurar que, para um par de assinaturas a serem comparadas, o número de níveis de ambas seja o mesmo. Além disso, cada nível em cada uma das assinaturas representa o mesmo nível no histograma.

Para calcular a distância entre duas assinaturas, obtêm-se, primeiramente as assinaturas estendidas. Isso garante que cada elemento de uma assinatura possui um elemento correspondente na outra assinatura. Sobre as assinaturas estendidas aplicam-se os algoritmos apropriados à medida utilizada, seja nominal, ordinal, ou modular. O algoritmo utilizado para o cálculo da assinatura estendida é bastante simples e é executado em ordem $O(n)$. Como os elementos das assinaturas são ordenados segundo os níveis do histograma subjacente, basta percorrer ambas as assinaturas elemento a elemento. Quando os elementos atuais em ambas as assinaturas correspondem ao mesmo nível, avançam-se os ponteiros de ambas as assinaturas para o próximo elemento. Quando não existe uma correspondência, um dos níveis é maior do que o outro. Na assinatura que contiver o maior

nível, insere-se um nível vazio correspondente ao menor deles. O mesmo procedimento é realizado no caso de final prematuro de uma das assinaturas.

A Figura 2.16 exemplifica a obtenção da assinatura estendida. As letras (a) e (b) da Figura 2.16 apresentam os histogramas originais e a Figura 2.16(c) mostra as assinaturas correspondentes. A Figura 2.16(d) apresenta o resultado da aplicação do algoritmo de Serratosa e Sanfeliu (SERATOSA e SANFELIU, 2006) que calcula as assinaturas estendidas. Pode-se observar na Figura 2.16(d) que na assinatura S_A foram inseridos os níveis vazios (3,0) e (5,0) e na assinatura S_B inseriram-se os níveis vazios (1,0) e (4,0). Com isso, cada nível da assinatura estendida S'_A passou a possuir um nível correspondente na assinatura estendida S'_B e vice-versa.

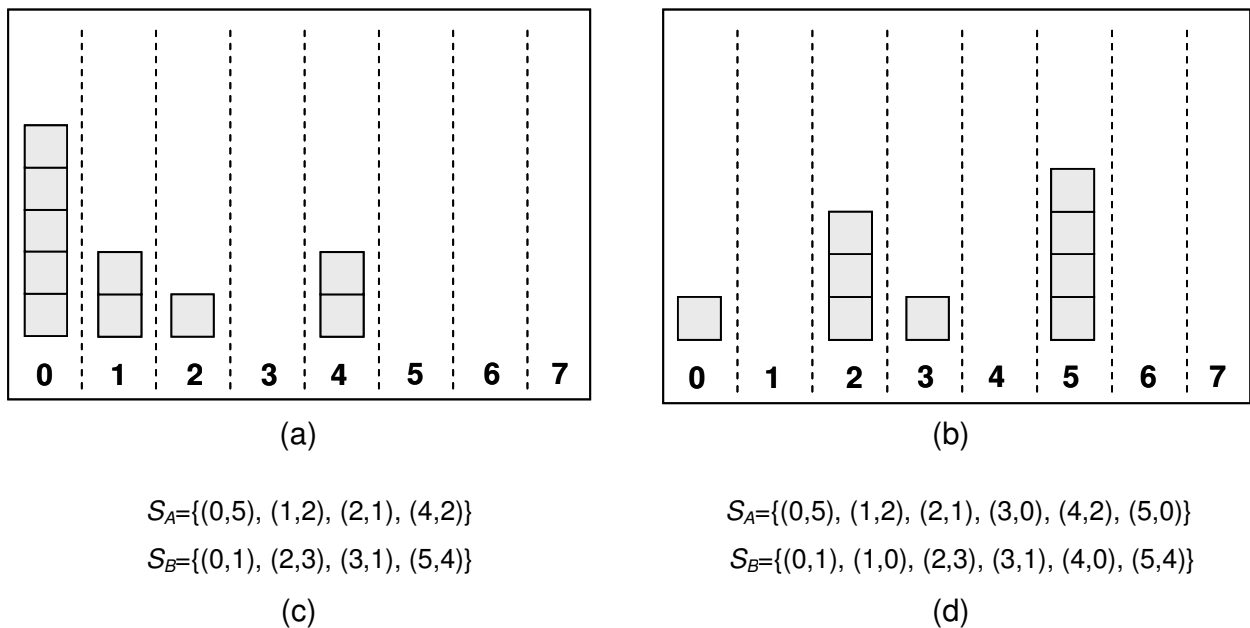


Figura 2.16 - Obtenção das assinaturas estendidas a partir dos histogramas originais.

Os algoritmos de cálculo da distância entre as assinaturas propostos por Serratosa e Sanfeliu (SERATOSA e SANFELIU, 2006) são basicamente os mesmos propostos por Cha e Srihari (CHA e SRIHARI, 2002) para o cálculo de distância entre os histogramas. Pequenas adaptações foram realizadas para levar em conta a ausência dos níveis vazios nas assinaturas estendidas. Na medida nominal, nenhuma alteração foi realizada. Nas medidas

ordinal e modular, uma única alteração foi proposta: a operação que incrementa a distância com a diferença entre os valores dos histogramas foi ponderada pela diferença entre os níveis consecutivos da assinatura. Isso é suficiente para levar em conta os níveis vazios não representados. Os algoritmos que calculam a distância ordinal e a distância modular entre as assinaturas são apresentados nas Listagem 2.1 e Listagem 2.2, respectivamente. Nessas listagens, as linhas que implementam as modificações realizadas foram destacadas.

Listagem 2.1 - Algoritmo para o cálculo da distância ordinal entre as assinaturas.

```

1   $D_{ord} = \text{Distância\_Ordinal}(S_A, S_B)$ 
2   $S'_A, S'_B, z' \leftarrow \text{Assinatura\_Estendida}(S_A, S_B);$ 
3   $D_{ord} \leftarrow 0; p \leftarrow 0;$ 
4  for  $i \leftarrow 0$  to  $z'-2$ 
5       $p \leftarrow p + S'_A.m_i - S'_B.m_i;$ 
6   $D_{ord} \leftarrow (S'_A.w_{i+1} - S'_A.w_i) * \text{abs}(p);$ 
7  end

```

Listagem 2.2 - Algoritmo para o cálculo da distância modular entre as assinaturas.

```

1   $D_{mod} = \text{Distância\_Modular}(S_A, S_B)$ 
2   $S'_A, S'_B, z' \leftarrow \text{Assinatura\_Estendida}(S_A, S_B);$ 
3   $D_{mod} \leftarrow 0; p[0] \leftarrow S'_A.m_0 - S'_B.m_0;$ 
4  for  $i \leftarrow 1$  to  $z'-1$ 
5       $p[i] \leftarrow S'_A.m_i - S'_B.m_i + p[i-1];$ 
6  for  $i \leftarrow 0$  to  $z'-2$ 
7   $D_{mod} \leftarrow D_{mod} + (S'_A.w_{i+1} - S'_A.w_i) * \text{abs}(p[i]);$ 

8  do
9       $D2 \leftarrow 0;$ 
10      $c \leftarrow \text{min positive}(p[i] \text{ for } 0 \leq i \leq z'-1);$ 
11      $\text{temp}[i] \leftarrow p[i] - c \text{ for } 0 \leq i \leq z'-1;$ 
12     for  $i \leftarrow 0$  to  $z'-2$ 
13      $D2 \leftarrow D2 + (S'_A.w_{i+1} - S'_A.w_i) * \text{abs}(\text{temp}[i]);$ 
14     if  $D_{mod} > D2$ 
15          $p[i] \leftarrow \text{temp}[i] \text{ for } 0 \leq i \leq z'-1;$ 
16     while  $D_{mod} > D2;$ 

17  do
18      $D2 \leftarrow 0;$ 
19      $c \leftarrow \text{max negative}(p[i] \text{ for } 0 \leq i \leq z'-1);$ 
20      $\text{temp}[i] \leftarrow p[i] - c \text{ for } 0 \leq i \leq z'-1;$ 
21     for  $i \leftarrow 0$  to  $z'-2$ 
22      $D2 \leftarrow D2 + (S'_A.w_{i+1} - S'_A.w_i) * \text{abs}(\text{temp}[i]);$ 
23     if  $D_{mod} > D2$ 
24          $p[i] \leftarrow \text{temp}[i] \text{ for } 0 \leq i \leq z'-1;$ 
25     while  $D_{mod} > D2;$ 

```

Nas medidas nominal e ordinal, a ordem $O(n)$ do algoritmo leva a uma redução linear do tempo de processamento, equivalente à supressão dos níveis nulos comuns. Já na medida modular, a ordem $O(n^2)$ do algoritmo permanece, causando uma redução bastante sensível no tempo de execução. Supondo que a metade dos níveis do histograma são suprimidos na assinatura, o tempo de execução passa a ser de um quarto do tempo original. A redução é muito mais sensível quando se trabalha com histogramas esparsos.

Pelo que foi apresentado neste item, pode-se concluir que quando se trabalha com histogramas esparsos a utilização de assinaturas é bastante vantajosa tanto do ponto de vista de economia de espaço de armazenamento quanto do aspecto de economia no tempo de execução dos algoritmos de medida de distância.

2.7 Considerações Finais deste Capítulo

Este capítulo apresentou os histogramas como uma importante ferramenta de análise de dados. Foi mostrado que os histogramas podem ser considerados como uma estimativa da densidade de probabilidade de ocorrência de uma grandeza, ou como um vetor de características relacionado a determinado objeto. Mostrou-se também que existem três tipos de medidas de histogramas: ordinal, nominal e modular. Esses tipos de medida determinam a forma de comparação entre eles. Algumas métricas utilizadas para o cálculo da distância entre histogramas foram definidas e uma breve comparação entre elas foi discutida. Além desses tópicos, foram também apresentados os conceitos de projeções suavizantes e de assinaturas.

No próximo capítulo os conceitos apresentados neste capítulo serão utilizados e desenvolvidos quando forem apresentadas as técnicas mais utilizadas na determinação de semelhança entre imagens digitais. Grande parte delas fundamenta-se em cálculos estatísticos e todas, sem exceção, buscam definir as melhores medidas de similaridade ou dessemelhança, que recaem em algum caso das métricas de distância mostradas no decorrer deste capítulo.

3 SEMELHANÇA EM IMAGENS DIGITAIS

3.1 Introdução

A diminuição do preço dos equipamentos de registro de imagens digitais tem provocado uma crescente disseminação na utilização desse meio nas mais diversas áreas e para as mais diversas finalidades. É comum que o usuário médio de uma câmara fotográfica digital produza, em um período de poucos meses, milhares de fotografias que devem ser descarregadas em um computador pessoal, ou armazenadas em algum *bureau* de serviços. Empresas de comunicação como jornais e canais de televisão, devem prover algum tipo de arquivamento de fotos, filmes ou gravações de programas de variedades para posterior utilização. Museus, teatros e casas de cultura podem guardar em um acervo digital cópias de suas telas, fotos de suas obras, ou registros de suas montagens e instalações. Órgãos de pesquisa, de segurança, de prospecção de recursos, podem e fazem uso das imagens digitais como um instrumento corriqueiro para o desempenho de suas funções.

A crescente utilização de imagens digitais é justificada não apenas pela facilidade de produção do material fotográfico, como também pela praticidade de seu arquivamento. As cópias materiais são volumosas e necessitam de cuidado em sua manipulação, além de demandarem um esforço físico razoável para sua localização e recuperação. Esse esforço é tanto maior quanto maior é o acervo, existindo casos em que os equipamentos especiais de arquivamento instalados em grandes ambientes são fundamentais para que as consultas sejam factíveis. Nos casos de exemplares únicos, o compartilhamento da consulta torna-se impossível, dificultando o acesso à informação. Finalmente, não se pode ignorar que o material arquivado existe em sua forma física e, portanto, é suscetível à degradação. Na maioria dos casos, a cópia com a finalidade de preservação tem sua qualidade degradada além de demandar recurso material, financeiro e dispêndio de tempo para que possa ser efetuada.

Todas as características negativas citadas anteriormente nesta seção são superadas pela adoção do arquivamento digital em meio eletrônico. O volume gasto no armazenamento eletrônico é várias ordens de grandeza inferior ao armazenamento físico. A

localização e recuperação, respeitadas as mesmas técnicas de indexação e pesquisa, não demandam esforço físico e são realizadas instantaneamente quando comparadas ao caso material. O compartilhamento da informação é facilitado, sem necessidade de cópias. Não se fala em degradação do conteúdo. Se o meio de armazenamento pode apresentar erros, degradação ou corrupção, técnicas de replicação existem para que se minimize a perda de informação mesmo na presença de falhas nos equipamentos ou de eventos catastróficos externos. É importante enfatizar que essas técnicas de replicação não impõem degradação do conteúdo e, apesar de demandarem recursos financeiros consideráveis, bem como tempo de replicação mensurável, ganham na comparação com seus equivalentes físicos.

A despeito das vantagens mencionadas no parágrafo anterior, o aumento no volume de material fotográfico em formato digital cria novas demandas para o seu processamento, classificação, armazenamento e recuperação. Tomando-se como exemplo uma coleção de fotografias de um usuário comum, alguns aplicativos permitem a anotação manual de instantâneos. Essa anotação, em conjunto com as técnicas de arquivamento usuais, permite que se possa localizar uma foto, ou um conjunto delas, com grande facilidade e rapidez. Mas nem sempre existe a disponibilidade de se anotar as imagens com o detalhamento necessário. Assim, tanto a possibilidade de anotação automática, como a recuperação de imagens por semelhança surgem como alternativas altamente desejáveis, tendo como objetivo a recuperação de fotos por meio de consultas mais complexas, mais abrangentes ou mais restritivas.

Uma categoria de aplicações similares à anotação e que também apresenta grande aplicabilidade prática é a classificação de imagens. Nela, uma amostra deve ser categorizada, considerando um conjunto de classes previamente determinadas. Nesse contexto, pode-se tomar como exemplo uma imagem médica como a tomografia computadorizada do cérebro que deve ser esquadrinhada em busca de características especiais indicativas de alguma moléstia ou anormalidade. Outro exemplo é a rotulação de imagens de satélites, ou provenientes de levantamentos aerofotogramétricos. Nesses casos, é desejável que se possa segmentar automaticamente a imagem em regiões homogêneas que podem ser classificadas e rotuladas segundo um conjunto de possibilidades

previamente estabelecidas, tais como matas, regiões de cultura, regiões urbanas, cursos d'água, lagos, ou oceanos, de forma a permitir uma melhor utilização das imagens por outros aplicativos mais especializados. Essas aplicações são conhecidas, na literatura como segmentação, ou classificação, supervisionada.

Finalmente, pode-se ressaltar a possibilidade de segmentar uma imagem, dividindo-a em regiões uniformes mesmo sem que exista sobre elas qualquer conhecimento prévio. É a segmentação não supervisionada. Essa tarefa é bastante útil ao se analisar as diferentes possibilidades de agrupamento para um conjunto de dados desconhecidos com a finalidade de que se possa estabelecer uma estrutura inicial que permita uma análise mais detalhada.

Todos os exemplos citados anteriormente nesta introdução, apesar de serem diversos, compartilham uma característica comum: a determinação da semelhança entre imagens diferentes, ou entre regiões diferentes da mesma imagem. O caso de recuperação de imagens semelhantes é bastante óbvio: deseja-se procurar imagens do banco de dados que se assemelhem com uma imagem referência. Na anotação automática, ou na classificação supervisionada, procura-se determinar, para cada região da imagem considerada, com qual dos padrões previamente definidos ela mais se assemelha. Na segmentação não supervisionada, busca-se aglutinar regiões de dados uniformes que apresentam características semelhantes, separando-as de outras regiões uniformes para as quais aquelas características não são observadas.

O conceito de semelhança é bastante complexo porque apresenta caracteres de subjetividade. Apesar da grande quantidade de trabalhos desenvolvidos nessa área, não foi possível, ainda, estabelecer um critério absoluto pelo qual se possa afirmar que duas imagens são pouco ou muito semelhantes. Duas imagens podem se assemelhar pelas cores principais, pela existência de objetos de forma semelhante, por apresentarem textura parecida, pela organização de suas partes mais chamativas, por serem mais escuras, sóbrias, ou por serem claras e brilhantes. Não existe uma característica única, nem mesmo um conjunto definido de características que, quando utilizados, permitem a medição da

semelhança entre duas imagens. Por isso, os trabalhos publicados abrangem um largo espectro de técnicas, conceitos e modelos que buscam obter progressos no assunto.

Como a questão da semelhança é importante para definir critérios de classificação, segmentação e busca, e conforme mencionado anteriormente nesta introdução, por se tratar de um tema de relativa complexidade, o objetivo deste capítulo é apresentar as principais linhas de pesquisa no campo da determinação de semelhanças em imagens digitais.

Este capítulo apresenta, inicialmente, uma revisão bibliográfica sobre semelhança entre imagens digitais. A seguir, apresenta as técnicas que utilizam matrizes de tamanho de corridas de níveis de cinza, a matriz de co-ocorrência de níveis de cinza e as unidades de textura. Apresenta ainda algumas técnicas mais recentes aplicadas à determinação de semelhança em imagens digitais e resalta alguns tópicos sobre o padrão MPEG-7. Finalmente, são realizadas as considerações finais deste capítulo.

3.2 Revisão Bibliográfica

3.2.1 Conceitos básicos

Uma imagem digital pode ser considerada como um vetor bidimensional cujas células armazenam informação de cor ou de níveis de cinza que se referem a alguma gradação do espectro eletromagnético na região da luz visível e sua vizinhança. Sejam $L_x = \{1, 2, \dots, N_x\}$ e $L_y = \{1, 2, \dots, N_y\}$ os domínios espaciais unidimensionais X e Y . Seja também $G = \{1, 2, \dots, N_g\}$ um conjunto imagem de tonalidades discretas. Uma imagem pode ser considerada como uma função $I : L_x \times L_y \rightarrow G$, que associa a cada célula de resolução bidimensional, um valor de tonalidade. Apesar dessa definição conceituar imagens compostas por níveis de cinza, sua extensão para imagens coloridas pode ser realizada com a simples adequação do conjunto imagem G que passará a conter um conjunto discreto de cores, e não mais de tonalidades monocromáticas.

Robert Haralick e outros (HARALICK *et alii*, 1973) afirmaram que a tarefa mais difícil na classificação de imagens digitais é a definição de um conjunto adequado de

características que permitam sua análise e que, portanto, traduzam grandezas familiares para a interpretação humana. Essas características enquadram-se em três categorias: características espectrais, texturais e contextuais. No grupo das características espectrais, analisam-se as variações tonais nas várias bandas do espectro visível e de sua vizinhança. As características texturais referem-se às inter-relações entre as tonalidades em uma vizinhança restrita. Finalmente, as características contextuais buscam descrever as relações entre blocos vizinhos de uma região e seus objetos. O estudo das características contextuais extrapola os objetivos deste trabalho, que tem como objeto de estudo a utilização das cores e das texturas na determinação da semelhança entre as imagens.

Cor e textura são conceitos distintos, porém são relacionados intimamente. A cor diz respeito às variações tonais das células de resolução. A textura refere-se à distribuição espacial das tonalidades pelos *pixels* de uma micro-região. A cor é uma propriedade do *pixel* enquanto que a textura é uma propriedade de uma região. Apesar de serem conceitos diferentes, cor e textura estão intimamente ligadas e não se pode falar em uma sem considerar a outra. Cada região de uma imagem apresenta ambas as características. É comum observar que algumas regiões caracterizam-se por extensões monótonas e homogêneas de tonalidades predominantes, enquanto outras regiões apresentam grande e rápida variação tonal. No primeiro caso, diz-se que a característica predominante da região é a cor, ao passo que, na segunda, prevalecem as características texturais (HARALICK, 1979). No entanto, essa divisão de prevalência entre cor e textura não é determinável rigidamente. As imagens naturais normalmente apresentam uma gradação que ora é mais suave, ora é mais abrupta, como pode ser observado na Figura 3.1.

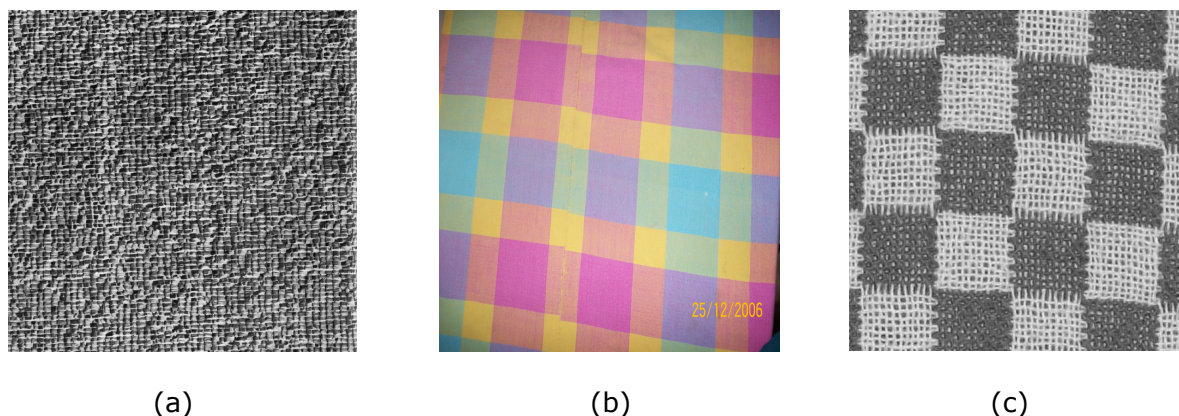


Figura 3.1 - Exemplos de imagens que apresentam característica predominante de textura e de cor. (a) Textura D84 (BRODATZ, 1968). (b) Colcha colorida. (c) Cor e textura estão igualmente presentes.

Pode-se observar na Figura 3.1 (a) a predominância de textura, enquanto que na Figura 3.1 (b) a predominância é de cor. Na Figura 3.1 (c), cor e textura estão presentes de maneira equilibrada. Pode-se verificar que, conforme a resolução considerada, ocorre a predominância de textura (resolução fina), ou de cor (resolução grosseira).

3.2.2 Alguns métodos de extração de características texturais

Várias revisões bibliográficas sobre as técnicas de extração de características texturais estão disponíveis na literatura. Robert Haralick (HARALICK, 1979) reviu algumas técnicas, então utilizadas, classificando-as como estatísticas ou estruturais. Entre as abordagens apresentadas por ele, vale mencionar as seguintes: a função de auto-correlação, os métodos que utilizam transformadas, a abordagem que analisa a distribuição de bordas na imagem (*textural edgeness*), a análise de co-ocorrência espacial de níveis de cinza e a análise do comprimento de corridas (*run length*). Além dessas, a medição da densidade de bordas por unidade de área, bem como a densidade de extremos são também mencionadas com maior detalhe por ele.

3.2.2.1 Autocorrelação

A autocorrelação de uma função unidimensional visa determinar o grau de auto-similaridade apresentado por essa função, ou seja, em que medida essa função apresenta regiões que se repetem exata ou aproximadamente. No caso discreto, o coeficiente de autocorrelação de uma seqüência real unidimensional pode ser calculado pela Equação (3.1). Esse coeficiente assume valores no intervalo $[-1,1]$. Quanto mais próximo da unidade está o valor absoluto de $\rho(m)$, maior é a auto-similaridade entre a imagem e sua versão deslocada de uma distância m . De forma análoga, quanto mais próximo de zero, menor é a similaridade. O caso de funções bidimensionais é a extensão direta do caso unidimensional.

$$\rho_x(m) = \frac{\sum_i x(i+m)x(i)}{\sum_i x^2(i)}$$

onde :

(3.1)

$x(i)$: termos da seqüência

Quando se imagina que uma textura pode ser entendida como a repetição periódica de uma configuração particular de níveis de cinza, espera-se que a autocorrelação de uma região que tenha uma textura desse tipo, apresente um comportamento periódico. De fato, a Figura 3.2 apresenta duas texturas semelhantes. Na Figura 3.2 (a) é ilustrada uma textura com padrões menores e mais próximos e na Figura 3.2 (b) é mostrada outra com padrões maiores e mais separados. No primeiro caso, diz-se que a textura é fina (*fine texture*) e no segundo caso, diz-se que a textura é grossa (*coarse texture*). Nas letras (c) e (d) da Figura 3.2, são apresentados os detalhes centrais da autocorrelação dessas respectivas texturas. Podem-se observar nessas figuras picos claros periódicos que denotam as posições de maior auto-similaridade. Pode-se verificar também que na textura fina os picos estão mais próximos, ao passo que na textura grossa eles estão mais separados, o que está de acordo com a periodicidade dos elementos texturais.

Um ponto importante ressaltado por Robert Haralick (HARALICK, 1979) é que a medida da granularidade (*coarseness*) da textura, mais fina ou mais grossa, pode ser medida pela taxa de queda da autocorrelação em torno da origem, ou seja, para pequenos deslocamentos. Ele explica que para texturas finas, como os picos estão mais próximos, a queda é mais abrupta, enquanto que para texturas grossas, como a variação de níveis de cinza são mais suaves, a queda também é.

Para confirmar essa explicação pode-se observar nas letras (e) e (f) da Figura 3.2, o gráfico da linha horizontal central das autocorrelações das texturas das letras (a) e (b) da Figura 3.2, respectivamente. Pode-se verificar claramente na origem desses gráficos, situadas em suas regiões centrais evidenciadas pelas setas, que a queda é mais abrupta para a textura mais fina.

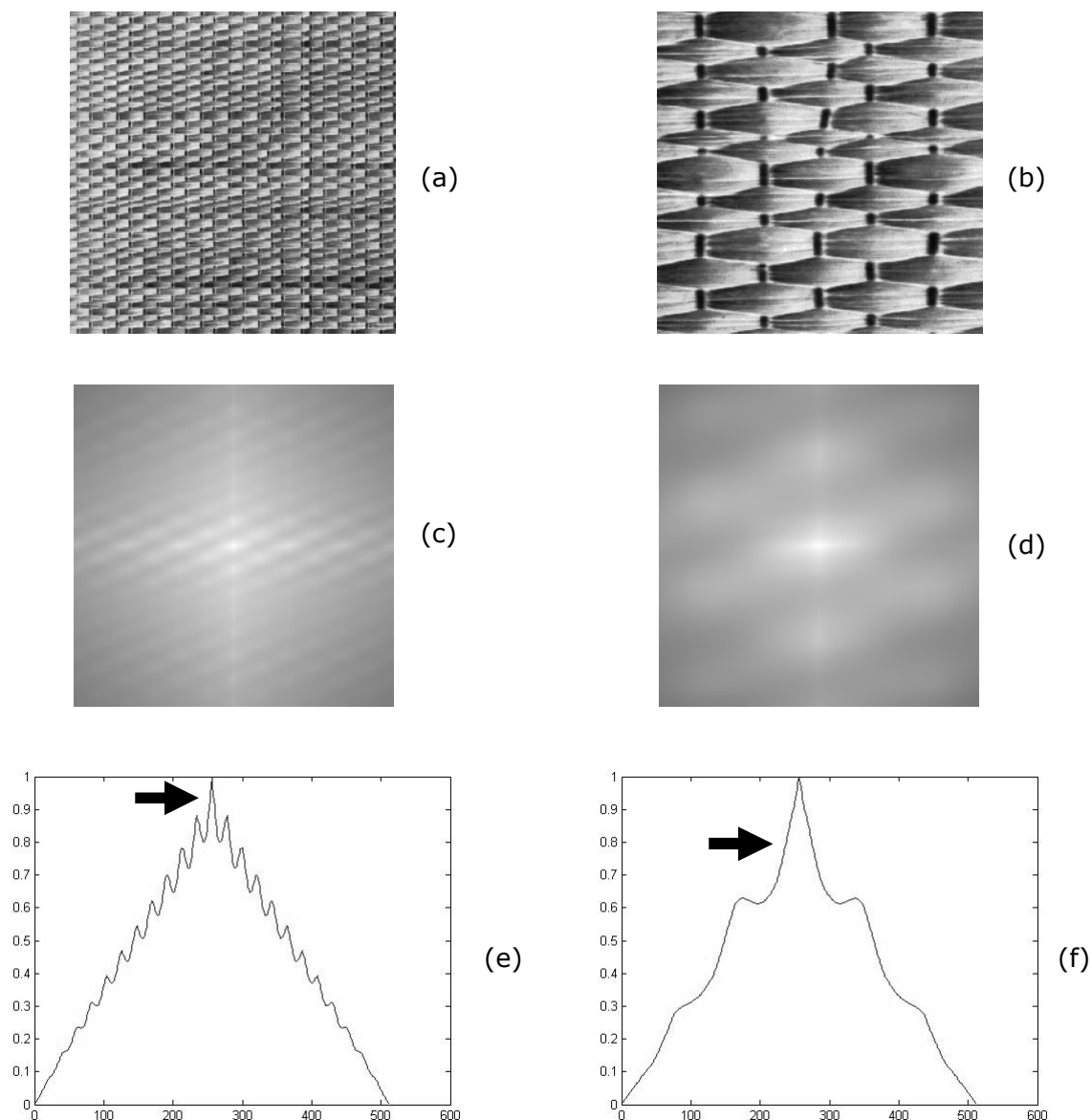


Figura 3.2 - Texturas e sua auto-correlação. (a) Uma textura fina: D55 (BRODATZ, 1968). (b) Uma textura grossa: D56 (BRODATZ, 1968). (c) e (d) As regiões centrais da auto-correlação das respectivas texturas. (e) e (f) O gráfico da linha central horizontal da auto-correlação das respectivas texturas.

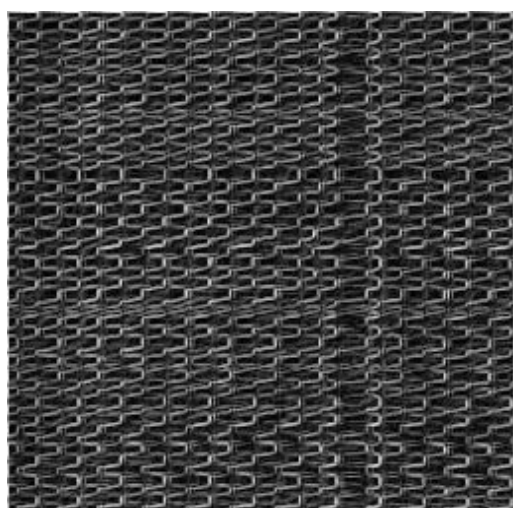
Apesar dos resultados da auto-correlação refletirem uma concordância com o comportamento esperado, Robert Haralick (HARALICK, 1979) ressalta que essa medida não capta algumas variações devidas à granularidade de certas regiões mais homogêneas da imagem. Outras técnicas podem ser utilizadas para a obtenção de melhores resultados na determinação da granularidade das texturas. A densidade de bordas e a densidade de extremos citadas anteriormente neste capítulo são dois exemplos importantes.

3.2.2.2 Densidade de bordas

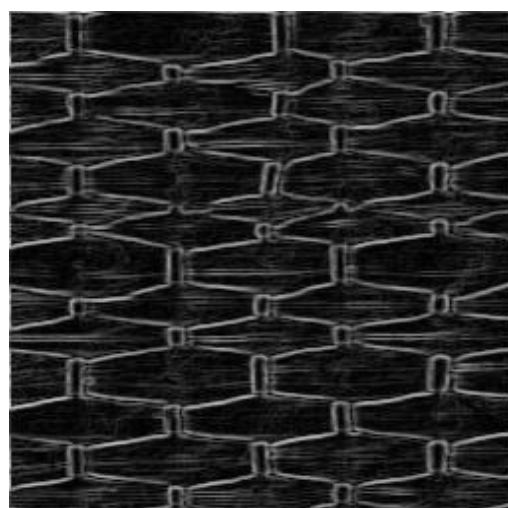
Calcula-se a densidade de bordas de uma imagem por meio do cálculo da magnitude do gradiente em cada um dos seus *pixels*. O gradiente em um *pixel* é aproximado pela aplicação de qualquer dos operadores usuais na vizinhança apropriada. Roberts, Prewitt, Sobel e Canny são exemplos de operadores de detecção de bordas que se baseiam na estimativa do gradiente (GONZALES e WOODS, 2000). Calculada a magnitude do gradiente em cada ponto da imagem, obtém-se uma estimativa de sua densidade de probabilidade pelo cálculo de seu histograma. Assim, pode-se calcular a “quantidade” média de borda por unidade de área da figura.

A utilidade da medida da densidade de bordas pode ser entendida por meio do seguinte raciocínio. A magnitude do gradiente indica a medida da maior transição no nível de cinza para um determinado ponto. Quanto maior essa magnitude, maior é a variação observada no nível de cinza entre os pixels vizinhos. A média da magnitude do gradiente para uma imagem, ou região, indica a presença de muita variação nas tonalidades de cinza para essa imagem ou região. Ao contrário, valores médios baixos para essa propriedade tendem a indicar regiões mais monótonas.

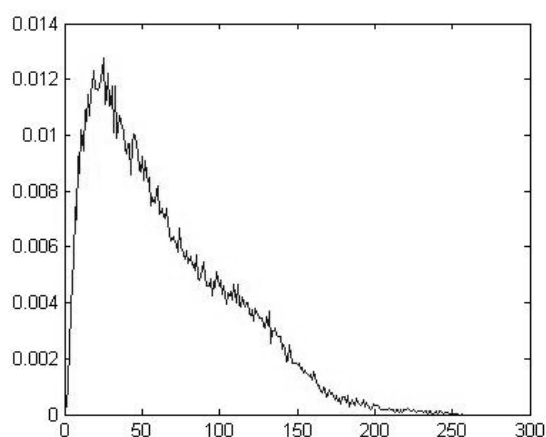
Ligando o comportamento do gradiente mencionado no parágrafo anterior àquele apresentado na discussão da granularidade, pode-se intuir que uma alta densidade de bordas, ou seja, um grande valor para a magnitude média do gradiente pode implicar em uma textura de granularidade fina, enquanto que um valor baixo para a média pode significar uma textura com granularidade grossa. Na Figura 3.3, as letras (a) e (b) mostram os gráficos da magnitude do gradiente para as texturas correspondentes da Figura 3.2. As letras (c) e (d) da Figura 3.3 apresentam os respectivos histogramas normalizados segundo os valores máximos da magnitude do gradiente para cada caso. O cálculo da densidade média de bordas por unidade de área resulta em 63,08 para a textura fina e em 37,52 para a textura grossa. Esses resultados confirmam a suposição apresentada anteriormente neste parágrafo, de que quanto maior é a densidade de bordas, mais fina é a textura.



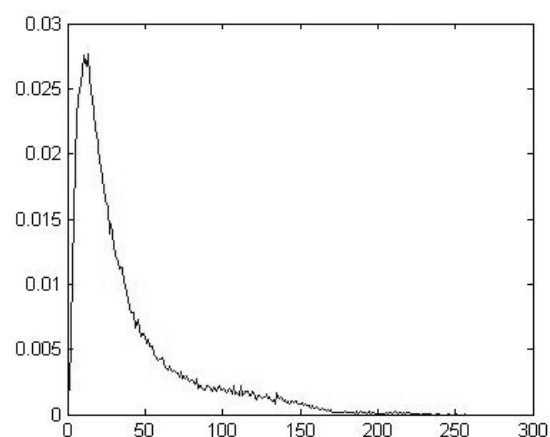
(a)



(b)



(c)



(d)

Figura 3.3 - (a) e (b) Magnitude do gradiente para as texturas apresentadas na Figura 3.2; (c) e (d) histograma normalizado da magnitude do gradiente para essas respectivas texturas.

3.2.2.3 Densidade de extremos

O cálculo da densidade de extremos apresenta aplicabilidade semelhante à da densidade de bordas e é bastante simples. O objetivo é encontrar e marcar todos os extremos locais, máximos e mínimos, de uma imagem. A partir dessa informação, calcula-se, para cada *pixel*, o número de extremos existentes dentro de uma região quadrangular centralizada nele. Assim como ocorre com a densidade de bordas, regiões que apresentam um valor alto para a densidade de extremos parecem ser formadas por uma textura mais

finas, enquanto regiões que apresentam um valor menor tendem a apresentar uma textura mais monótona com um padrão tonal mais uniforme.

Uma alternativa à simples marcação dos extremos é atribuir a cada ponto extremo sua altura, definindo o valor dos demais como zero. Essa técnica permite, além do cálculo da densidade média de extremos, o cálculo da altura média dos extremos em uma região. Essa última medida sugere uma aproximação para a rugosidade da região.

A definição de extremo pode conter algumas armadilhas que devem ser consideradas cuidadosamente. Por simplicidade, são apresentadas as definições para o caso unidimensional. A extensão para o caso bidimensional é simples e direta.

Um pixel na posição i apresentando um nível de cinza $g(i)$ é considerado um extremo se uma das Equações (3.2) é obedecida.

$$\begin{aligned} g(i) \geq g(i-1) \quad \text{e também} \quad g(i) \geq g(i+1) \quad \text{para um máximo} \\ g(i) \leq g(i-1) \quad \text{e também} \quad g(i) \leq g(i+1) \quad \text{para um mínimo} \end{aligned} \tag{3.2}$$

A utilização dos operadores maior ou igual e menor ou igual, na Equação (3.2), faz com que sejam considerados como extremos, não apenas os extremos reais, como também todos os pontos de um platô. Esse fato pode trazer distorções para o cálculo da densidade de extremos, atribuindo valores altos a regiões de platô, o que é um contrassenso. Para superar esse problema, os operadores “menor que” e “maior que” podem ser utilizados alternativamente. Outra maneira é marcar como extremo apenas um dos pontos do platô, ou ainda, ao invés de atribuir aos pontos do platô a sua altura real, atribui-se a média de altura a cada um deles. Nesse caso, se existem N pontos com altura h em um platô, cada um deles é marcado com o valor h/N .

A determinação dos extremos no caso bidimensional é um pouco mais difícil, já que uma vizinhança de 8 deve ser, em princípio, considerada. Como o objetivo principal da técnica é a estimação da rugosidade ou da granularidade, pode-se simplificar o cálculo considerando-se apenas uma vizinhança de 4 na determinação dos extremos. Nesse caso,

pode-se marcar os extremos em cada linha da imagem pela utilização da Equação (3.2) e, posteriormente aplicar a mesma marcação para as colunas da imagem. São considerados extremos aqueles pixels que apresentam duas marcações. Os pixels que apresentam apenas uma marcação são pontos de inflexão ou pontos de sela e, portanto, não são considerados como extremos.

Para ilustrar a aplicação da técnica da densidade dos extremos, as letras (a) e (b) da Figura 3.4, mostram o resultado da marcação dos extremos das texturas apresentadas na Figura 3.2.

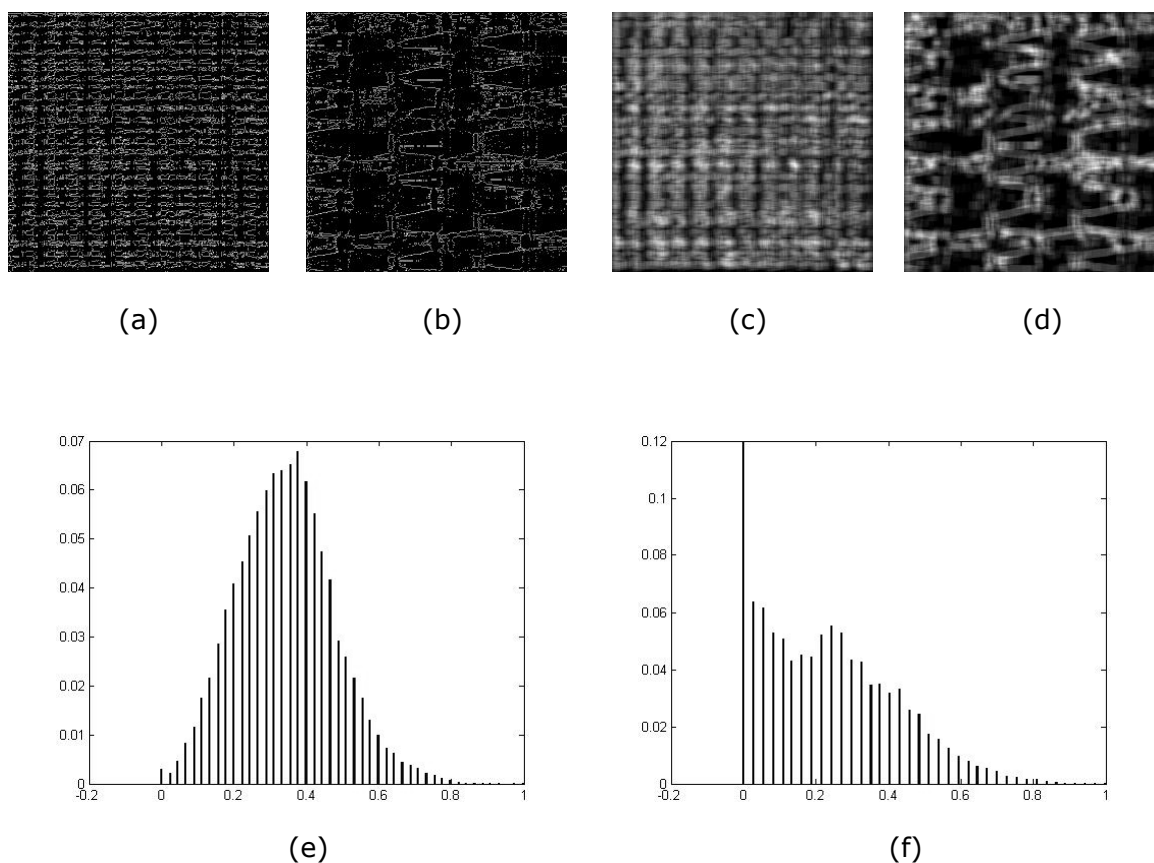


Figura 3.4 – (a) e (b) Resultado da marcação dos extremos para as texturas da Figura 3.2; (c) e (d) densidade média dos extremos calculada em um vizinhança 8×8 ; (e) e (f) histograma normalizado da distribuição da densidade de bordas.

As letras (c) e (d) da Figura 3.4 correspondem ao cálculo da densidade dos extremos em uma região 8×8 em torno do *pixel*. Finalmente, as letras (e) e (f) da Figura 3.4 são os histogramas normalizados da densidade dos extremos. O cálculo da densidade

média dos extremos para cada textura resulta nos valores 0,3407 para a textura fina e 0,2309 para a textura grossa, de pleno acordo com o exposto na argumentação sobre a densidade dos extremos.

3.2.2.4 Modelos de imagens

Outra abordagem sobre as características texturais baseia-se na definição dos modelos de imagem. Azriel Rosenfeld e Larry Davis (ROSENFELD e DAVIS, 1979) apresentaram vários modelos utilizados na análise de imagens digitais. Também nesse caso, os modelos foram caracterizados como estatísticos ou espaciais. Fazem parte do primeiro grupo aqueles modelos que descrevem as texturas por meio das propriedades estatísticas que os valores de tonalidade apresentam nas regiões da imagem. Desse grupo participam propriedades de primeira ordem, tais como brilho e contraste, bem como estatísticas de ordem superior, que analisam a densidade de probabilidade conjunta de múltiplas grandezas ou propriedades locais.

Ainda na categoria dos modelos estatísticos, aparecem os modelos de campos aleatórios (*random field models*) e a aplicação do conceito de séries temporais para a modelagem de texturas. Segundo o modelo de campos aleatórios, cada pixel da imagem é considerado uma variável aleatória. Esse modelo tem por objetivo descrever as relações espaciais das tonalidades de cinza por meio das suas funções de densidades de probabilidade individual e conjunta para os *pixels* da imagem. Cabe notar que essas funções não podem ser derivadas do histograma dos níveis de cinza da própria imagem. Considerando que as propriedades estatísticas dependem da localização do *pixel* na imagem, as estimativas dessas funções devem ser computadas sobre um conjunto de imagens semelhantes, ou seja, imagens descritas pelo mesmo modelo.

A utilização de séries temporais visa modelar a dependência espacial dos níveis de cinza da imagem como uma seqüência de variáveis aleatórias em que cada uma delas depende de um conjunto de variáveis precedentes. Esses modelos podem ser auto-regressivos, de média móvel ou mistos. Modelos auto-regressivos são descritos pela

Equação (3.3). O valor de uma variável aleatória depende das n variáveis que a precedem e de um termo aleatório que desempenha o papel análogo ao de um ruído.

$$\tilde{z}_i = \sum_{j=1}^n \alpha_j \tilde{z}_{i-j} + \tilde{v}(i)$$

onde :

(3.3)

$\tilde{z}_i$: i -ésima variável aleatória da seqüência;

α_j : peso arbitrário; e

$\tilde{v}$: variável aleatória Gaussiana de média 0 e variância σ^2 .

Os modelos de média móvel são descritos pela Equação (3.3) e consideram que uma variável aleatória depende apenas de um conjunto de n observações anteriores do termo aleatório, e não dos valores anteriores das variáveis. Esse modelo trata a série como sendo resultado de uma filtragem de um sinal aleatório.

$$\tilde{z}_i = \sum_{j=0}^n \beta_j \tilde{v}(i-j)$$

onde :

(3.4)

$\tilde{z}_i$: i -ésima variável aleatória da seqüência;

β_j : peso arbitrário; e

$\tilde{v}$: variável aleatória Gaussiana de média 0 e variância σ^2 .

O modelo misto é descrito por uma equação que combina as Equações (3.3) e (3.4). A grande desvantagem da utilização das séries temporais é que elas apresentam um comportamento unilateral, considerando a dependência das variáveis anteriores. Representar uma imagem por esse modelo certamente pode não permitir uma modelagem adequada, já que os *pixels* de uma imagem correlacionam-se com uma vizinhança que engloba tanto os precedentes quanto os subseqüentes.

Quanto aos modelos espaciais, eles pressupõem que imagens são compostas por regiões e buscam descrever basicamente três aspectos distintos: a interação entre regiões, a transição entre elas e seu tamanho e forma. O primeiro aspecto busca verificar se as regiões são compostas de pontos espalhados esparsamente, se elas se interpenetram ou se formam um mosaico, semelhante a um mapa. Quanto às transições, se elas são abruptas, ou graduais, ou ainda, se são observadas como uma mancha desfocada. Finalmente, quanto ao tamanho e forma, valores que medem a curvatura da borda, ou indiquem se são rombudas ou alongadas ou ainda, se são pequenas ou grandes, todos esses são descritores que caracterizam cada uma das regiões importantes da imagem. (ROSENFELD e DAVIS, 1979). Essas características se enquadram no aspecto contextual mencionado anteriormente neste capítulo e, por isso, não são aprofundadas neste trabalho.

3.2.2.5 Estimação de energia

Em um estudo comparativo sobre técnicas de classificação de texturas, Trygve Randem e John Håkon Husøy (RANDEN e HUSØY, 1999) apresentaram uma revisão das técnicas de extração de características baseadas na utilização de bancos de filtros. Nesse trabalho, eles expuseram e compararam uma extensa série de técnicas que se baseiam, fundamentalmente, na estimação da energia local em várias regiões do espectro de frequências da imagem. Segundo essa abordagem, o conjunto dessas estimativas da energia local para cada imagem forma um vetor de característica que pode ser utilizado para a classificação, podendo ser útil também para a determinação da semelhança entre as imagens.

Segundo Randem e Husøy (RANDEN e HUSØY, 1999), todas as técnicas comparadas compartilham o mesmo conjunto de etapas para a extração do vetor de características a partir da imagem. As diferenças entre as técnicas está na implementação de cada etapa, principalmente na escolha do banco de filtros e na escolha do filtro não linear retificador. O esquema das etapas principais do processo de extração das características é apresentado na Figura 3.5. Na primeira etapa a imagem é submetida a um banco de filtros sintonizados que enfatizam determinada região do espectro. A saída dessa etapa é um conjunto de imagens filtradas, $o_1, o_2, \dots, o_n$, de mesmo tamanho que a imagem

original, uma para cada filtro do banco. A etapa seguinte submete as imagens filtradas a um filtro não linear. O objetivo desse filtro é retificar a imagem, suprimindo seus valores negativos de alguma forma. A seguir, as imagens filtradas e retificadas passam por um filtro suavizador cujo propósito é calcular o valor médio, eliminando as flutuações das estimativas de energia $e_1, e_2, \dots e_n$. Finalmente aplica-se, opcionalmente, um processo de normalização das estimativas obtidas, produzindo-se as componentes do vetor de características $c_1, c_2, \dots c_n$.

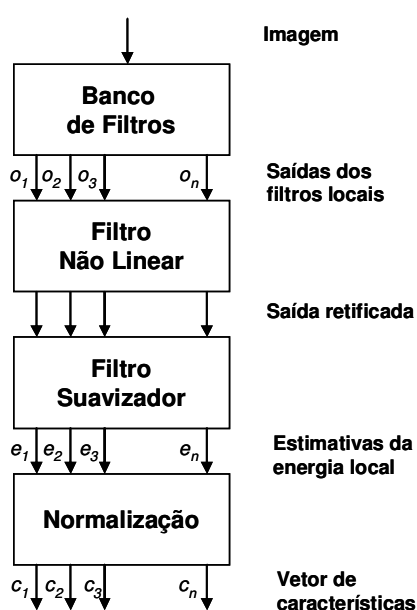


Figura 3.5 – Processo de extração de características de imagens por meio de filtragem.

Randem e Husøy (RANDEN e HUSØY, 1999) citaram Laws (LAWS, 1980) como um dos pioneiros da utilização de filtragem para identificação de texturas. Ele propôs a divisão do espectro de freqüências em 25 regiões aproximadamente quadradas pela utilização de cinco filtros separáveis de cinco amostras cada, cujas respostas impulsionais são: $h_1 = [1, 4, 6, 4, 1]$, $h_2 = [-1, -2, 0, 4, 1]$, $h_3 = [-1, 0, 2, 0, -1]$, $h_4 = [-1, 2, 0, -2, 1]$ e $h_5 = [1, -4, 6, -4, 1]$. As respostas em freqüência normalizadas para esses filtros podem ser observadas na Figura 3.6. Pode-se perceber a divisão do espectro em cinco intervalos conforme os picos de cada uma das respostas. Como os filtros são separáveis e aplicados às

linhas e depois às colunas da imagem, o espectro bidimensional é dividido nas 25 regiões mostradas na Figura 3.7.

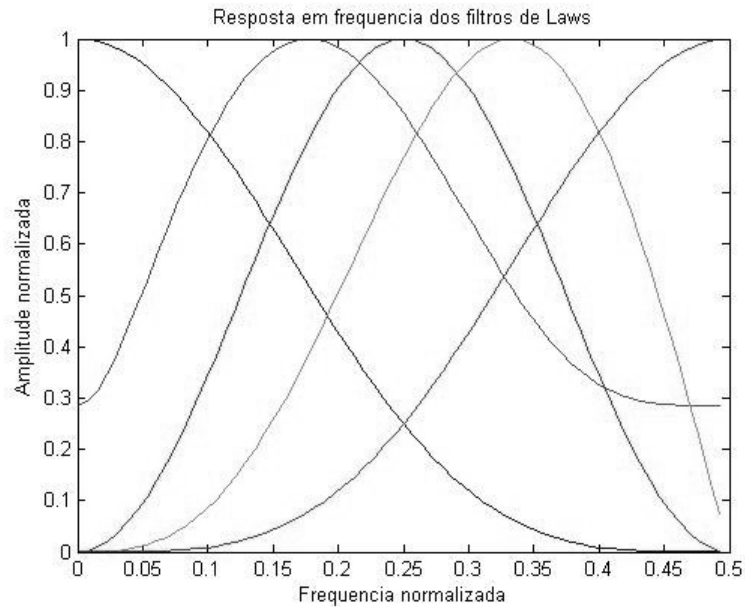


Figura 3.6 – Resposta em frequência dos filtros separáveis de Laws h_1 , h_2 , h_3 , h_4 e h_5 na ordem dos picos.

Espectro de frequências vertical	11	12	13	14	15
	21	22	23	24	25
	31	32	33	34	35
	41	42	43	44	45
	51	52	53	54	55
Espectro de frequências horizontal					

Figura 3.7 – Divisão do espectro de frequências em 25 regiões. Os algarismos no interior de cada região indicam os índices dos filtros aplicados na horizontal e na vertical.

A Figura 3.8 mostra o resultado da filtragem da textura D56 da Figura 3.2(b) por cada um dos filtros de Laws. As respostas de uma mesma linha foram submetidas ao

mesmo filtro horizontal. De forma análoga, as respostas de uma mesma coluna foram submetidas ao mesmo filtro vertical. As freqüências centrais dos filtros crescem da esquerda para a direita e de cima para baixo. Considerando-se a primeira linha, pode-se notar que as respostas mais à direita enfatizam as componentes verticais de maior freqüência, já que as raias horizontais aparecem em grande número. Considerando-se a primeira coluna, notam-se raias verticais preponderantes nas respostas localizadas mais abaixo. Nas respostas da diagonal principal, foi aplicado o mesmo filtro, tanto na horizontal quanto na vertical.

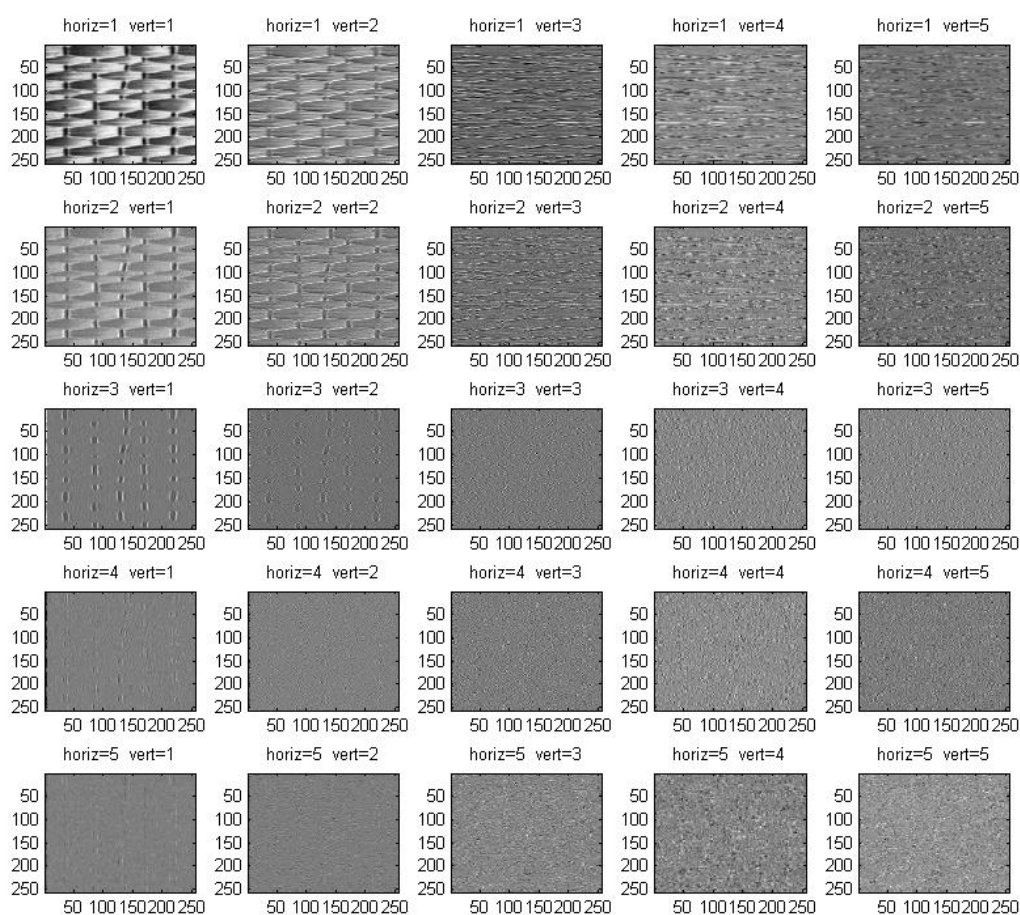


Figura 3.8 – Resultado da filtragem da textura D56 da Figura 3.2(b) pelo banco de filtros de Laws.

A Figura 3.9 apresenta os resultados da filtragem da textura D55 da Figura 3.2 (a) pelo mesmo banco de filtros. A diferença mais evidente para essa D55 é que ela possui componentes de maior freqüência mais fortes, já que sua granularidade é mais fina. Isso pode ser observado comparando-se, por exemplo, as respostas na posição 52, ou na

posição 25. Para a textura D56, de granularidade mais grossa, as respostas nessas posições apresentam maior monotonia, enquanto que para a textura D55, alguma textura na resposta em frequência ainda é observável.

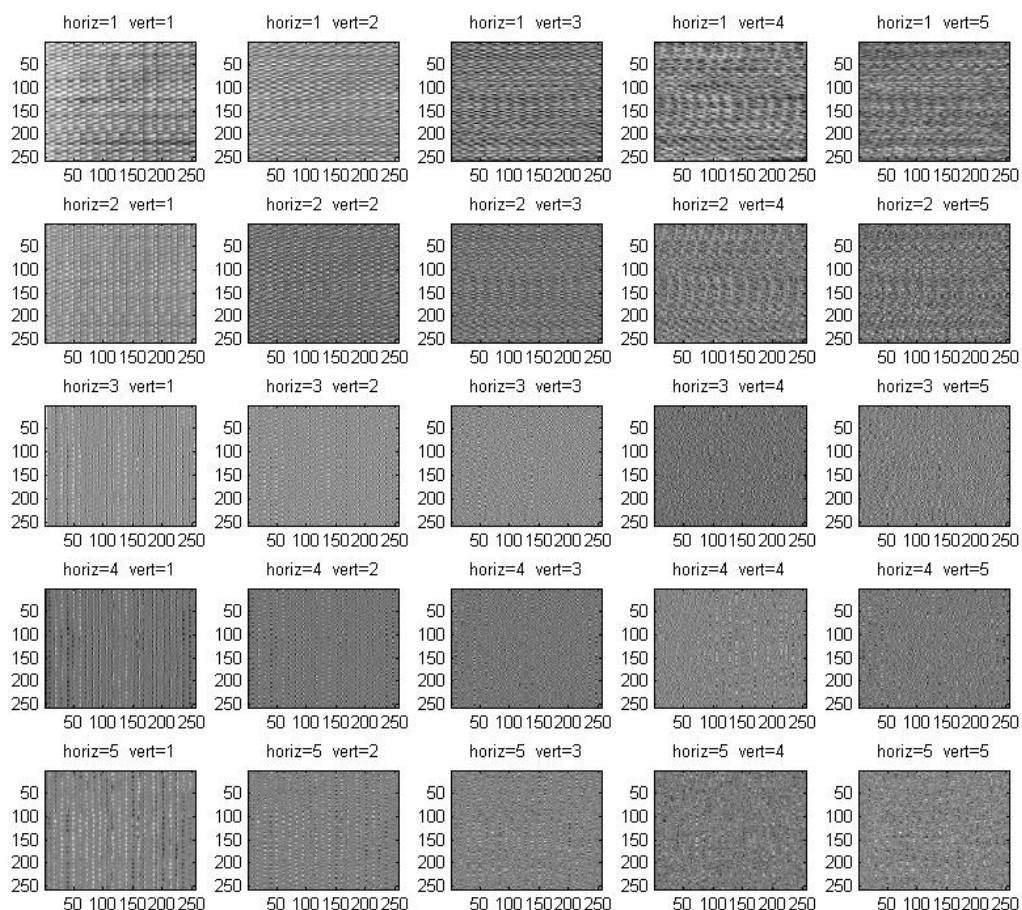


Figura 3.9 - Resultado da filtragem da textura D55 da Figura 3.2(a) pelo banco de filtros de Laws.

Outra abordagem relacionada à filtragem, mencionada por Randem e Husøy (RANDEN e HUSØY, 1999) e também analisada por Richard Connors e Charles Harlow (CONNERS e HARLOW, 1980) é o chamado método do espectro de potência (*Power Spectral Method*). Ele assemelha-se ao proposto por Laws, entretanto, a divisão do espectro é realizada de modo diferente. Coggins e Jain (COGGINS e JAIN, 1985) propuseram uma série de filtros, sete em formato de anel e outros quatro em formato de cunha que visavam levar em conta tanto a frequência quanto a orientação espacial. Eles são mostrados na Figura 3.10 (a) e (b). As bandas passantes dos filtros em anel são mostradas em tonalidades

gradativas de cinza. As bandas passantes dos filtros em forma de cunha são os quadrantes claros, que selecionam as direções de 0° , 45° , 90° e 135° .

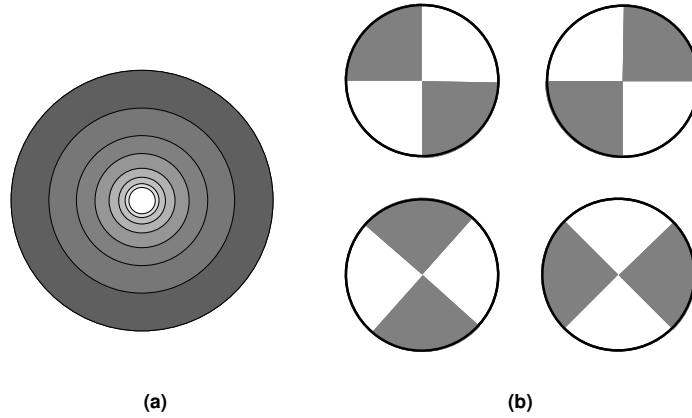


Figura 3.10 – Filtros de Jain (a) em anel; (b) em cunha.

Resultados semelhantes podem ser obtidos pela integração do espectro de potência da imagem em uma região em forma de anel ou de cunha (CONNERS e HARLOW, 1980). O espectro de potência de uma imagem é obtido pela Equação (3.5).

$$\Phi(u, v) = F(u, v) \cdot F^*(u, v) = |F(u, v)|^2$$

onde : (3.5)

$F(u, v)$: transformada de Fourier da imagem; e

$F^*(u, v)$: complexo conjugado da transformada de Fourier da imagem.

A integração em um conjunto de regiões anelares é obtida pela Equação (3.6).

$$c_j = \int_0^\pi \int_{\rho_j}^{\rho_j + \Delta\rho_j} \Phi(\rho, \theta) \rho d\rho d\theta, \quad j = 1, 2, \dots, n_a$$

onde :

$$\begin{aligned} n_a & : \text{número de aneis;} \\ \Delta\rho_j & : \text{largura de faixa do } j\text{-ésimo anel;} \\ \rho & = (u^2 + v^2)^{1/2}; e \\ \theta & = \tan^{-1} \frac{v}{u} \end{aligned} \tag{3.6}$$

A integração em um conjunto de regiões em formato de cunha é obtida pela Equação (3.7).

$$c_j = \int_{\rho_{\min}}^{\rho_{\max}} \int_{\theta_j}^{\theta_j + \Delta\theta_j} \Phi(\rho, \theta) \rho d\rho d\theta \quad j = 1, 2, \dots, n_c$$

onde :

$$\begin{aligned} n_c & : \text{número de regiões em formato de cunha;} \\ \Delta\theta_j & : \text{ângulo de abertura da } j\text{-ésima cunha;} \\ \rho & = (u^2 + v^2)^{1/2}; e \\ \theta & = \tan^{-1} \frac{v}{u} \end{aligned} \tag{3.7}$$

Além da utilização dos bancos de filtros de Laws e dos filtros anelares e em forma de cunha, outro tipo de processamento largamente empregado utiliza os bancos de filtros de Gabor para a extração de características (REED e WECHSLER, 1990), (MANJUNATH e MA, 1996), (DIMAI, 1999), (RANDEN e HUSØY, 1999), (MANJUNATH *et alii*, 2001). Os filtros de Gabor são filtros sintonizados em frequência e orientação (ILONEN *et alii*, 2005). Um banco desses filtros apropriadamente projetado permite cobrir o espectro da imagem com um conjunto de regiões elipsoidais que selecionam tanto a faixa de frequência quanto a orientação desejadas. Os filtros de Gabor são separáveis e permitem

uma série de otimizações baseadas em simetria. São aplicáveis tanto no domínio espacial quanto no domínio da frequência. J. Ilonen, J.-K. Kämäräinen e H. Kälviäinen (ILONEN *et alii*, 2005) discorreram detalhadamente sobre os filtros de Gabor e sobre as técnicas de otimização que permitem uma implementação eficiente desses filtros.

A resposta impulsiva de um filtro de Gabor unidimensional normalizado é uma onda senoidal complexa modulada por uma gaussiana, definida pela Equação (3.8).

$$\psi(t) = \frac{|f_0|}{\gamma\sqrt{\pi}} e^{-\frac{|f_0|^2}{\gamma} t^2} e^{j2\pi f_0 t} \quad (3.8)$$

onde :

f_0 : frequência de sintonia do filtro; e
 γ : banda do filtro.

A frequência de sintonia e a largura de faixa determinam a faixa do espectro que será realçada. Quanto menor essa largura, mais seletivo é o filtro. A Figura 3.11(a) apresenta as partes real (linha contínua) e imaginária (linha tracejada) da resposta impulsiva de um filtro, para valores de $f_0 = 5$ e $\gamma = 2$. A Figura 3.11(b) mostra a resposta em frequência do mesmo filtro. Pode-se observar nessa figura, a seletividade do filtro e sua frequência central.

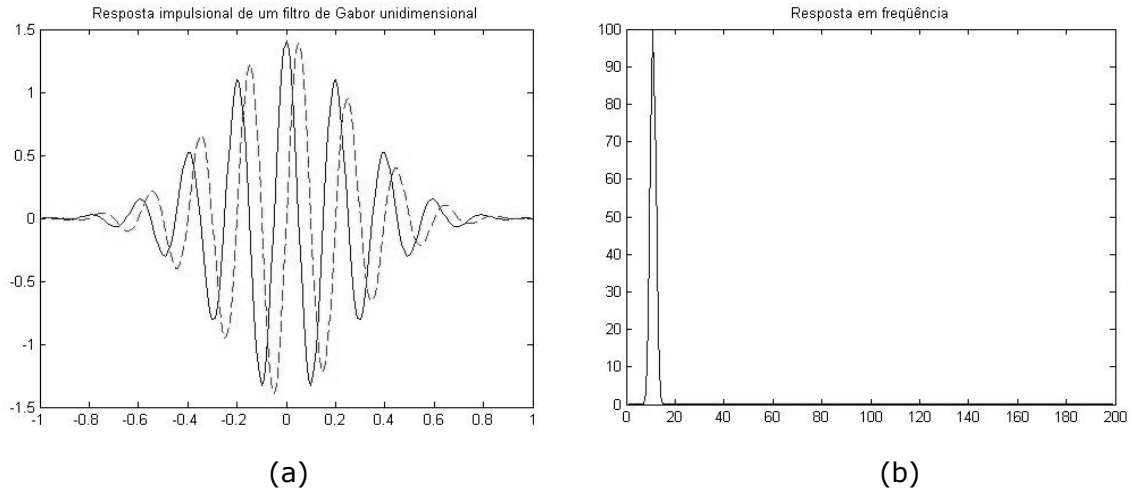


Figura 3.11 – (a) Resposta impulsiva de um filtro de Gabor normalizado unidimensional, partes real (linha sólida) e imaginária (linha tracejada); (b) resposta em frequência do mesmo filtro.

A extensão bidimensional do filtro de Gabor é imediata e sua definição normalizada é mostrada na Equação (3.8). Como se pode verificar nessa equação, a mudança de variáveis equivale a uma rotação por um ângulo θ da resposta impulsiva em relação ao eixo principal x' . Ao longo desse eixo, apresenta-se uma onda senoidal modulada por uma gaussiana. Ao longo do eixo secundário, y' , observa-se uma gaussiana pura. A razão entre os desvios padrão γ , ao longo do eixo principal, e η , ao longo do eixo secundário, define o aspecto mais circular ou mais elíptico da gaussiana modulante.

$$\psi(x, y; f_0, \theta) = \frac{f_0^2}{\pi\gamma\eta} e^{-\left(\frac{f_0^2}{\gamma^2}x'^2 + \frac{f_0^2}{\eta^2}y'^2\right)} e^{i2\pi f_0 x'}$$

onde :

$$\begin{aligned} x' &= x \cos \theta + y \sin \theta; \\ y' &= -x \sin \theta + y \cos \theta; \\ f_0 &: \text{frequência central;} \\ \theta &: \text{ângulo entre o eixo principal da gaussiana e o plano de onda;} \\ \gamma &: \text{largura de faixa ao longo do eixo principal; e} \\ \eta &: \text{largura de faixa ao longo do eixo secundário.} \end{aligned} \tag{3.9}$$

As letras (a) e (b) da Figura 3.12 mostram as partes real e imaginária da resposta impulsiva de um filtro de Gabor bidimensional, respectivamente. A Figura 3.12(b) mostra sua parte imaginária. A Figura 3.12(c) apresenta a resposta em frequência desse mesmo filtro.

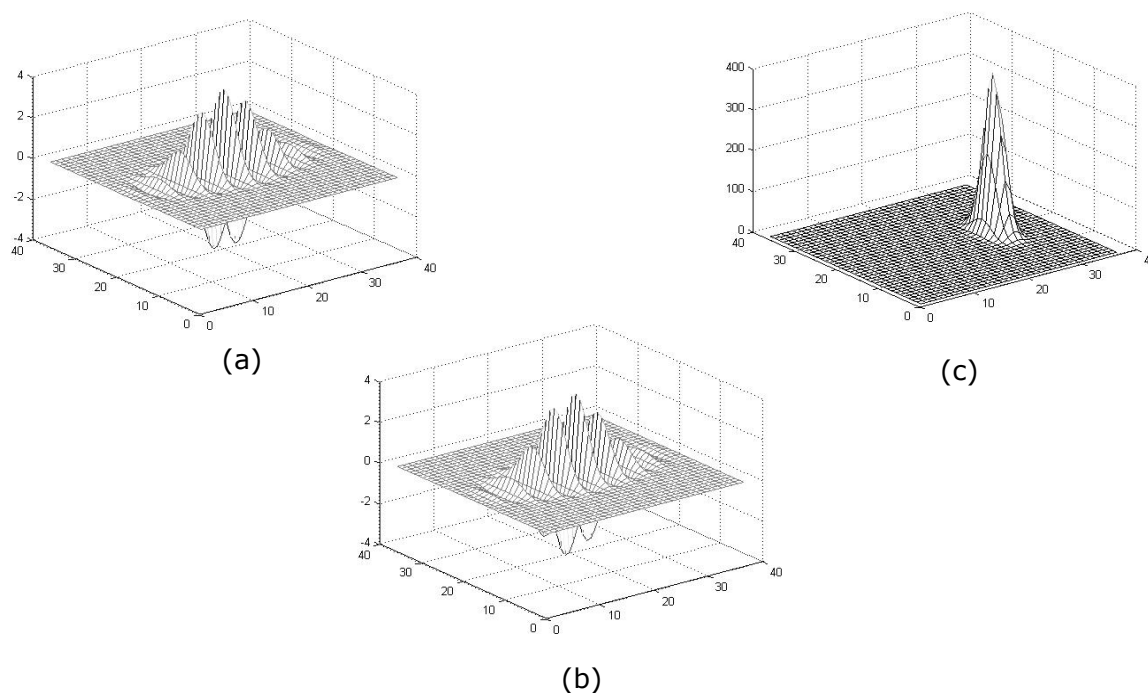


Figura 3.12 – (a) e (b) Partes real e imaginária de um filtro de Gabor bidimensional; (c) resposta em frequência desse filtro.

Os gráficos mostrados na Figura 3.12 são referentes a um filtro de Gabor para o qual o ângulo θ da Equação (3.8) é nulo. Para valores não nulos desse ângulo, obtém-se gráficos de mesma forma, porém rotacionados em relação ao eixo principal da gaussiana, o eixo x . A resposta em frequência também é rotacionada pelo mesmo ângulo, já que a transformada de Fourier apresenta tal propriedade (GONZALES e WOODS, 2000). Outra propriedade da transformada de Fourier, o teorema da modulação, também explica a localização do lóbulo da resposta em frequência na direção do eixo principal. Essa localização é determinada pela frequência da onda senoidal f_0 complexa modulada pela gaussiana, mais distante da origem para grandes valores de f_0 . Levando em conta essas

propriedades, pode-se concluir que é possível dividir o plano espectral utilizando-se um banco de filtros de Gabor obtido pela variação do ângulo θ e da frequência f_0 , ajustando-se simultaneamente os parâmetros de seletividade γ e η com a finalidade de se obter uma cobertura completa e com pouca sobreposição entre os espectros dos vários filtros. A técnica para fazê-lo é apresentada com detalhes no relatório de Ilonen e outros (ILONEN *et alii*, 2005). Um exemplo de cobertura obtida com quatro frequências e cinco ângulos de rotação é apresentado na Figura 3.13.

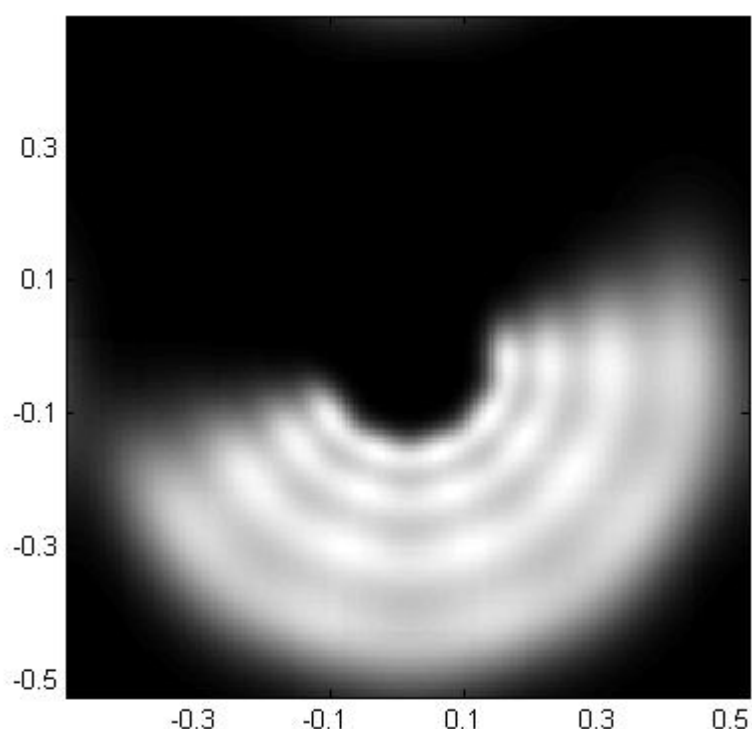


Figura 3.13 – Espectro de um banco de filtros de Gabor.

Pode-se verificar na Figura 3.13 os quatro anéis e, em cada um deles, cinco regiões aproximadamente elipsoidais, cada uma delas representando a resposta em frequência de um dos vinte filtros do banco considerado. Pode-se ressaltar que, por questão de simetria, apenas um dos semi planos do espectro é apresentado nessa figura. As coordenadas são apresentadas na frequência normalizada.

A Figura 3.14 mostra a saída de cada um dos filtros do banco de 4 frequências e 5 rotações (espectro da Figura 3.13) quando a entrada é a textura D55 da Figura 3.2(a).

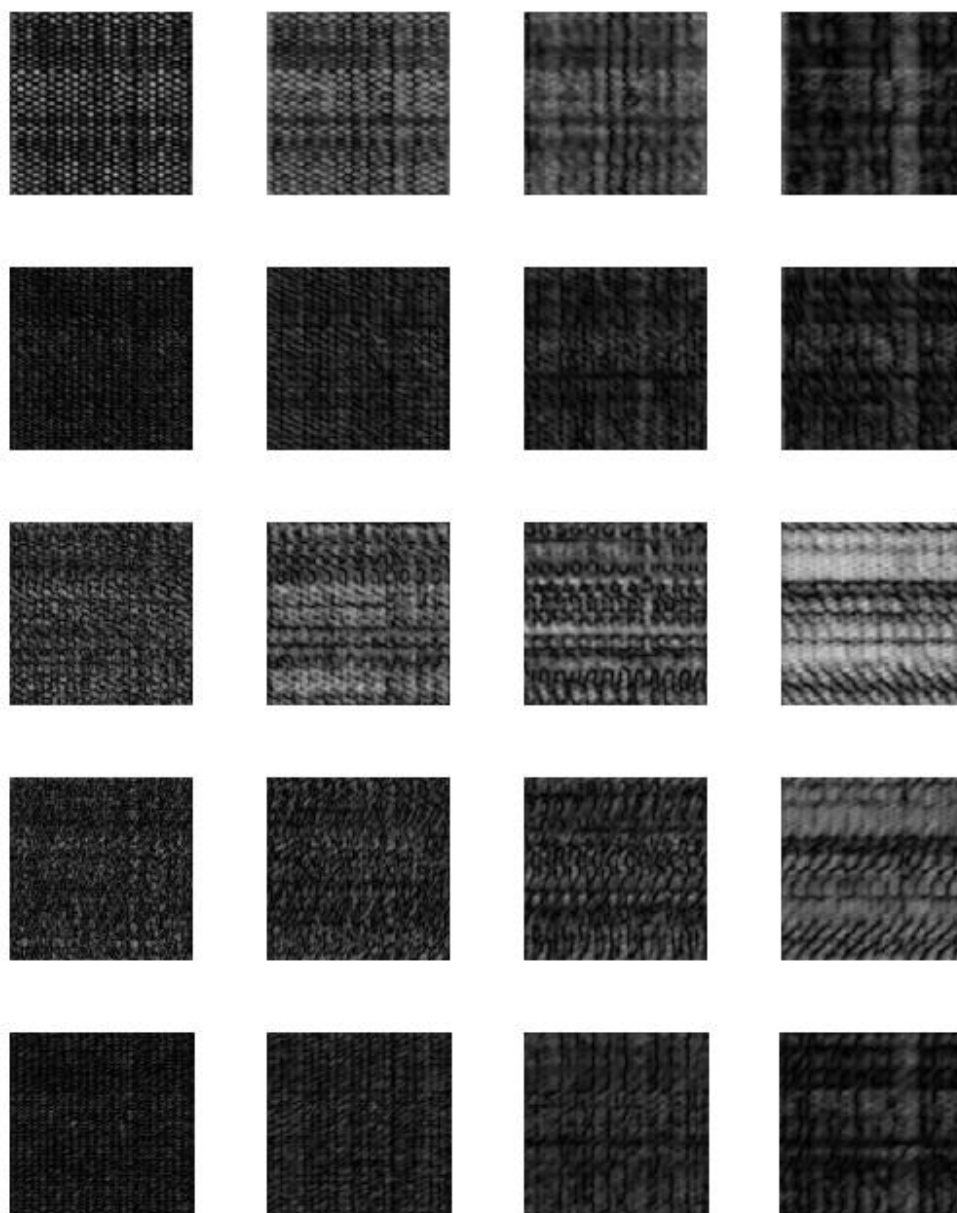


Figura 3.14 – Resposta do padrão D55 da Figura 3.2(a) ao utilizar um banco de filtros de Gabor de 4 frequências e 5 rotações.

Na Figura 3.14, as colunas apresentam filtros de mesma frequência f_0 e as linhas apresentam filtros de mesma rotação θ . Os valores da frequência normalizada f_0 são, aproximadamente: 0,14; 0,20; 0,28 e 0,40. Os valores do ângulo de rotação θ em graus são: 0° , 26° , 72° , 108° e 144° .

Para efeito de comparação, a Figura 3.15 mostra as saídas para o mesmo banco de filtros quando a entrada é a textura D56, da Figura 3.2(b).

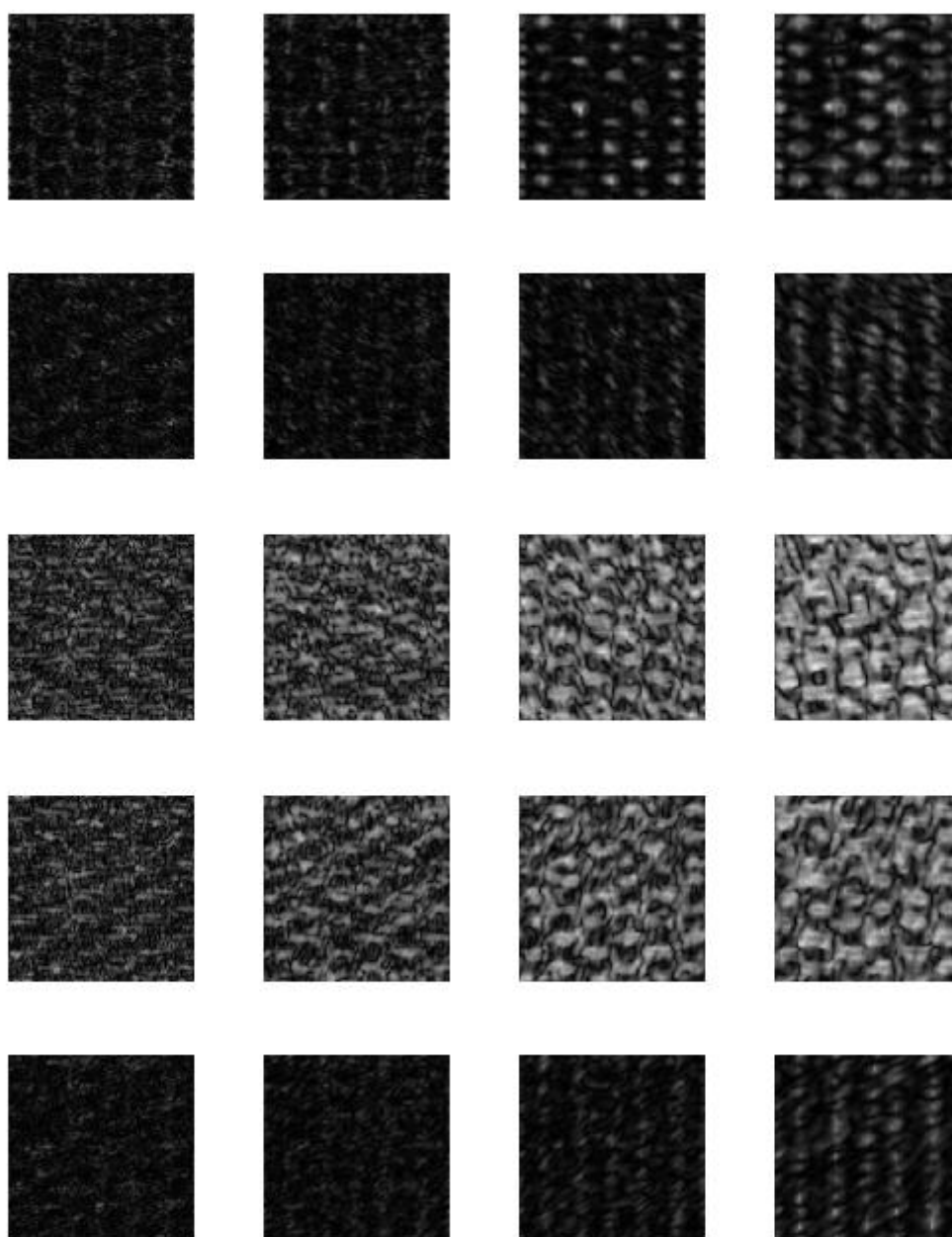


Figura 3.15 – Resposta do padrão D55 da Figura 3.2(b) ao utilizar um banco de filtros de Gabor de 4 frequências e 5 rotações.

Manjunath e Ma (MANJUNATH e MA, 1996) descreveram uma aplicação dos bancos de filtros de Gabor na determinação de semelhança entre imagens compostas por regiões de textura uniforme. Eles propuseram o seguinte procedimento. A imagem é submetida a um banco de filtros de Gabor de 4 frequências e 6 rotações. Para cada uma das 24 saídas, calcula-se a média e o desvio padrão dos valores absolutos da saída. O resultado

é um vetor de características composto de 48 elementos. A métrica utilizada para a comparação de similaridade é mostrada na Equação (3.10).

$$D(i, j) = \sum_f \sum_{\theta} d_{f\theta}(i, j)$$

$$d_{f\theta}(i, j) = \left| \frac{\mu_{f\theta}^{(i)} - \mu_{f\theta}^{(j)}}{\alpha(\mu_{f\theta})} \right| + \left| \frac{\sigma_{f\theta}^{(i)} - \sigma_{f\theta}^{(j)}}{\alpha(\sigma_{f\theta})} \right|$$

onde :

(3.10)

$\mu_{f\theta}^{(i)}$: média da saída do filtro $f\theta$ para a imagem i ;

$\sigma_{f\theta}^{(i)}$: desvio padrão da saída do filtro $f\theta$ para a imagem i ;

$\alpha(\bullet)$: desvio padrão da grandeza referida calculado sobre todas as imagens do banco de dados; e

f, θ : frequência e rotação do filtro do banco.

A técnica de Manjunath e Ma (MANJUNATH e MA, 1996) é utilizada na recuperação de imagens semelhantes em um banco de imagens. Para sua aplicação, é necessário que se calcule o desvio padrão global de cada um dos 48 elementos do vetor de características para as imagens do banco de dados. A razão para isso é possibilitar a normalização dos elementos do vetor sobre o banco de dados, permitindo sua adição sem que existam distorções causadas por um ou mais elementos que porventura apresentem valores muito discrepantes dos demais.

Manjunath e Ma (MANJUNATH e MA, 1996) reportaram resultados superiores a 70% para a taxa média de recuperação, calculada pela razão entre o número de imagens corretas recuperadas e o total de imagens recuperadas. Esses resultados são bastante significativos pela simplicidade da técnica proposta e foram superiores aos resultados obtidos quando outras técnicas mais complexas foram utilizadas (MANJUNATH e MA, 1996).

3.3 Matriz de Co-ocorrência de Níveis de Cinza

As técnicas de análise de texturas apresentadas até este ponto, neste capítulo, utilizam, ou medidas de distribuições de uma grandeza, ou resultados de filtragem e estimação de energia, ou ainda a estimação de parâmetros de modelos de imagem. No primeiro caso enquadram-se as distribuições de densidade de bordas ou dos extremos. As técnicas que utilizam filtragem, bem como as que utilizam modelos de imagem foram apresentadas nas seções anteriores deste capítulo. Esta seção tem por objetivo apresentar uma técnica que introduz as estatísticas de segunda ordem na análise de texturas. Essa técnica baseia-se na construção, a partir das imagens analisadas, das estruturas chamadas por seus propositores de matrizes de dependência espacial angular de níveis de cinza entre os vizinhos próximos (*angular nearest-neighbor gray-tone spatial-dependence matrices*) e foi apresentada por Robert Haralick, K. Shanmugam e Its'hak Dinstein em um importante artigo de 1973 (HARALICK *et alli*, 1973).

A partir do cálculo das matrizes de dependência espacial angular de níveis de cinza entre os vizinhos próximos, 14 diferentes medidas de textura foram propostas com o objetivo de fornecer as características utilizadas na classificação de três conjuntos diferentes de imagens: microfotografias de rochas sedimentares, imagens aerofotogramétricas e dados derivados de imagens de satélite.

O papel das características propostas é extrair das matrizes originais, valores que buscam sintetizar determinados aspectos da imagem ou de uma região nela contida. Grande parte dessas características traz um nome que tende a ser significativo quando considerado sob o aspecto sensível. Assim, o segundo momento angular, ou energia, o contraste, ou a entropia já expõem antecipadamente as noções sobre que grandeza a característica pretende medir. Em trabalhos posteriores, vários autores tentaram traduzir determinadas grandezas sensíveis por métricas que tomaram por base o trabalho de Haralick e outros semelhantes. Grandezas tais como a granularidade, direcionalidade, regularidade e aspereza e homogeneidade, entre outras, surgiram com o desenvolvimento das pesquisas na área (TAMURA *et alii*, 1978), (AMADASUN e KING, 1989), (PARKER, 1997). Espera-se que os valores dessas características para determinada imagem estejam

de acordo com a noção que seu nome implica. Espera-se também que eles sejam capazes de classificar as imagens em classes pré-determinadas, ou ainda de medir o grau de semelhança entre imagens diferentes.

Haralick e seus colaboradores apresentaram, em seu trabalho, resultados bastante significativos para a classificação, superando o valor de 80% de acerto, segundo eles (HARALICK *et alli*, 1973).

Esta seção apresenta o método de cálculo das matrizes e das características propostas por Haralick e seus colaboradores. Apresenta também uma alternativa mais vantajosa para o cálculo de algumas das características, que são os histogramas da soma e da diferença. Finalmente, comenta sobre o trabalho de alguns pesquisadores que buscaram aferir o paralelo entre as grandezas representadas por algumas características e suas contrapartidas sensíveis, obtidas da pesquisa de opinião de observadores sobre as imagens testadas.

3.3.1 O cálculo das matrizes

Conhecidas também como matrizes de co-ocorrência de níveis de cinza, (MCNC), as estruturas apresentadas neste capítulo visam estimar a densidade de probabilidade conjunta da ocorrência de um par de valores de níveis de cinza em dois *pixels* separados por uma distância d medida em uma direção que forma um ângulo θ com a horizontal. As razões que embasam a proposta são, primeiramente, o fato que texturas são propriedades regionais, distribuições de níveis de cinza que obedecem a um padrão específico e em segundo lugar, o fato que os valores de cor para *pixels* próximos em imagens naturais apresentam um alto nível de correlação espacial que tende a diminuir com a distância.

Para cada par de distância d e ângulo θ considerados, uma matriz deve ser calculada. A matriz é indexada por valores de níveis de cinza tanto nas colunas quanto nas linhas. Cada elemento m_{ij} da matriz de co-ocorrência $M(d, \theta)$ conterá o número de ocorrências, na imagem, de pares de *pixels* separados pela distância d calculada na direção θ que apresentam valores de nível de cinza i e j , respectivamente. Para exemplificar, considera-se a direção horizontal e a distância unitária. Cada elemento m_{ij} da matriz

$M(1,0^\circ)$ indica quantas vezes a combinação de pixels i e j , vizinhos na horizontal aparecem na imagem.

A Figura 3.16 (a) mostra os valores dos pixels de uma imagem 8 x 8. A faixa de valores para os pixels é o intervalo $[0,3]$. A dimensão da matriz de co-ocorrência $M(1,0^\circ)$ é 4 x 4, já que existem 4 valores de *pixel* distintos. A Figura 3.16 (b) mostra as matrizes de co-ocorrência $M(1,0^\circ)$, $M(1,45^\circ)$, $M(1,90^\circ)$ e $M(1,135^\circ)$, nas quatro direções principais.

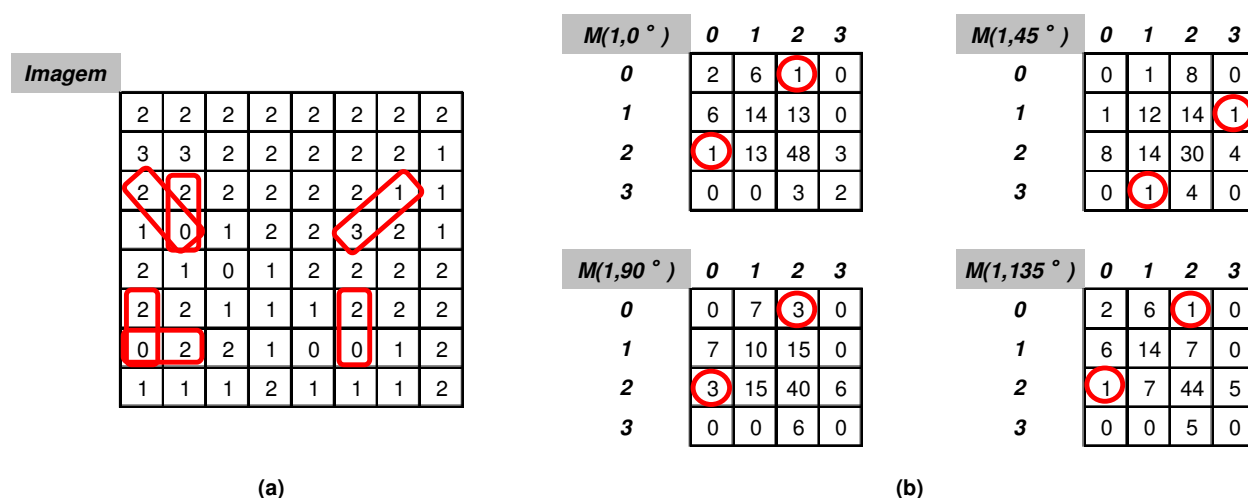


Figura 3.16 – (a) Uma imagem; (b) suas matrizes de co-ocorrência.

Os retângulos na imagem da Figura 3.16 (a) ressaltam os elementos que compõem os valores circulados das respectivas matrizes de co-ocorrência. Assim, existe apenas um par (0,2) na horizontal, um par (1,3) na direção 45° , um par (0,2) na direção 135° e três pares (0,2) na vertical. Vale observar que as matrizes de co-ocorrência, como definidas, são simétricas. Tanto o par (1,3) quanto o par (3,1) significam dois pixels com valores 3 e 1 separados pela distância unitária na direção 45° , ou seja, para o cálculo da distância, não se leva em consideração o sentido em que ela é medida.

A dimensão de uma MCNC não depende do tamanho da imagem sobre a qual ela é calculada. Ela depende do número de níveis de cinza considerados. Uma imagem que apresenta 256 níveis de cinza, portanto, origina uma MCNC de dimensão 256 x 256. Um tamanho tão grande de matriz pode inviabilizar a análise de texturas. Assim, o que normalmente se faz para superar esse inconveniente é quantizar os níveis de cinza em um

número de níveis menor. Deve-se sempre ter em mente que a quantização pode eliminar detalhes importantes da textura, por isso, ela deve ser utilizada com cautela.

Um ponto positivo que a MCNC apresenta é o baixo requisito de memória para as operações. Se o resultado exige espaço, o processamento da imagem para o cálculo exige que apenas duas de suas linhas sejam armazenadas simultaneamente na memória. Elas são as linhas que contêm os pontos a analisar. Ao final de seu processamento, duas novas linhas são carregadas na memória e processadas, até que toda a imagem tenha sido percorrida. Isso é uma grande vantagem quando se trabalha com imagens muito grandes e conserva em muito os requisitos de memória.

Quanto à velocidade, o cálculo dos elementos da matriz depende unicamente e de forma linear da dimensão da imagem o que é outro ponto favorável das matrizes de co-ocorrência de níveis de cinza.

Uma questão muito importante diz respeito à sensibilidade da matriz à variação dos níveis de cinza. Haralick ressalta que a aquisição da imagem está sujeita a vários fatores que podem fazer com que duas imagens diferentes de um mesmo objeto possam apresentar tonalidades distintas. Normalmente, essas duas imagens podem ser relacionadas por uma transformação monotônica. É desejável que duas imagens que sejam transformações monotônicas uma da outra produzam a mesma MCNC. Para isso, uma operação de equalização de histogramas é aconselhável como pré-processamento (HARALICK *et alli*, 1973).

3.3.2 As principais características propostas

O pressuposto fundamental para a utilização das matrizes MCNC é que toda a informação referente à textura da imagem geradora está contida nelas. Assim, baseando-se nessas matrizes Haralick (HARALICK *et alli*, 1973) definiu as medidas de textura ou de complexidade de variação dos níveis de cinza reproduzidas nas Equações (3.11) a (3.24). A notação utilizada nas definições é mostrada a seguir.

N : número de níveis de cinza;

R : constante de normalização;

$p(i, j)$: cada elemento da matriz normalizada;

$p_x(i) = \sum_{j=1}^N p(i, j)$ e $p_y(j) = \sum_{i=1}^N p(i, j)$: distribuições marginais x e y ;

$\mu_x, \mu_y, \sigma_x, \sigma_y$: médias e variâncias das distribuições marginais;

$\sum_i \bullet = \sum_{i=1}^N \bullet$: notação simplificada para o somatório em i ;

$\sum_j \bullet = \sum_{j=1}^N \bullet$: notação simplificada para o somatório em j .

$$p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^N \sum_{j=1}^N p(i, j), \quad k = 2, 3, \dots, 2N$$

$$p_{x-y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^N \sum_{j=1}^N p(i, j), \quad k = 0, 1, \dots, N-1$$

As características texturais são:

1) Segundo Momento Angular: Energia ou Uniformidade

$$f_1 = \sum_i \sum_j \{p(i, j)\}^2 \quad (3.11)$$

2) Contraste

$$f_2 = \sum_{n=0}^{N-1} n^2 p_{x-y}(n) \quad (3.12)$$

3) Correlação

$$f_3 = \frac{\sum_i \sum_j [(ij)p(i, j)] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.13)$$

4) Soma dos Quadrados: Variância

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (3.14)$$

5) Momento do Inverso da Diferença: Homogeneidade

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \quad (3.15)$$

6) Média da Soma

$$f_6 = \sum_{i=2}^{2N} i p_{x+y}(i) \quad (3.16)$$

7) Variância da Soma

$$f_7 = \sum_{i=2}^{2N} (i - f_6)^2 p_{x+y}(i) \quad (3.17)$$

8) Entropia da Soma

$$f_8 = - \sum_{i=2}^{2N} p_{x+y}(i) \log(p_{x+y}(i)) \quad (3.18)$$

9) Entropia

$$f_9 = - \sum_i \sum_j p(i, j) \log(p(i, j)) \quad (3.19)$$

10) Variância da diferença

$$f_{10} = \text{variância de } p_{x-y} \quad (3.20)$$

11) Entropia da Diferença

$$f_{11} = - \sum_{i=0}^{N-1} p_{x-y}(i) \log(p_{x-y}(i)) \quad (3.21)$$

12), 13) Medidas de Informação da Correlação

Considerando – se :

H_x, H_y : entropias de p_x e p_y , respectivamente;

$$H_{xy} = -\sum_i \sum_j p(i, j) \log(p(i, j))$$

$$H_{xy}^1 = -\sum_i \sum_j p(i, j) \log(p_x(i) p_y(j))$$

$$H_{xy}^2 = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

Define – se :

$$f_{12} = \frac{H_{xy} - H_{xy}^1}{\max\{H_x, H_y\}} \quad (3.22)$$

$$f_{13} = \sqrt{1 - e^{-2(H_{xy}^2 - H_{xy})}} \quad (3.23)$$

14) Máximo Coeficiente de Correlação

$$f_{14} = \sqrt{\text{segundo maior autovalor de } Q}$$

$$\text{onde :} \quad (3.24)$$

$$Q(i, j) = \sum_k \frac{p(i, k) p(j, k)}{p_x(i) p_y(k)}$$

Dessas 14 medidas apresentadas nas Equações (3.11) a (3.24), algumas foram mais usualmente empregadas nas pesquisas utilizando MCNC. Andrea Baraldi e Flavio Parmiggiani (BARALDI e PARMIGGIANI, 1995) fizeram uma investigação sobre as características texturais associadas a seis dessas medidas: segundo momento angular,

entropia, contraste, variância, correlação, e homogeneidade, ou momento do inverso da diferença. Nesse trabalho, aplicaram cada uma dessas seis medidas como métrica no processo de aglomeração não supervisionada de pequenos blocos de uma imagem aérea de uma região da Antártida. Como resultado, para cada medida obteve-se uma imagem segmentada em 5 e em 10 classes de terreno. Essas imagens resultantes foram comparadas à imagem original com a finalidade de se intuir a que tipo de percepção visual cada medida se associa. As conclusões principais do trabalho deles serão apresentadas a seguir.

O segundo momento angular (f_1) é também citado na literatura como energia, ou uniformidade. Os valores dessa característica pertencem ao intervalo $[0,1]$. Deve-se lembrar que, para uma matriz MCNC normalizada, a soma de todos os seus elementos é igual a 1. Como é uma medida de energia, calculada pela soma dos quadrados dos elementos da MCNC, valores altos ocorrem quando a matriz apresenta um pequeno número de elementos com valores elevados e o restante com valores baixos ou nulos. O menor valor para a uniformidade ocorre quando todos os elementos da MCNC apresentam o mesmo valor. Texturas monótonas, ou texturas que apresentam um grande número de ocorrências do mesmo par de cores, apresentam um valor alto para o segundo momento angular. Baraldi e Parmiggiani concluíram que as áreas monótonas ou que apresentam um padrão periódico conduzem a um valor elevado para essa medida (BARALDI e PARMIGGIANI, 1995).

A entropia (f_9) apresenta alta correlação inversa com o segundo momento angular. Ela mede a desordem de uma textura, resultando em valores mais elevados se a matriz MCNC possui elementos com valores semelhantes e, por conseguinte, probabilidade de ocorrência baixa. Isso significa que a imagem original apresenta muitos pares de transição diferentes, ou seja, tem comportamento semelhante ao ruído. Uma desvantagem dessa medida em relação ao segundo momento angular é que esse último é normalizado (sua faixa de valores é o intervalo $[0,1]$) enquanto que a primeira apresenta valores que se estendem pelo intervalo $(0, \infty)$.

A expressão que descreve o contraste (f_2) apresentada na Equação (3.12) pode ser reescrita de forma diferente, conforme a Equação (3.25).

$$f_2 = \sum_{n=0}^{N-1} n^2 p_{x-y}(i, j) = \sum_i \sum_j (i-j)^2 p(i, j) \quad (3.25)$$

Pode-se perceber que o contraste, conforme definido, é a variância das diferenças de níveis de cinza dos pontos separados pela distância d , na direção θ . Valores baixos dessa medida ocorrem quando a matriz MCNC apresenta elementos significativos próximos à sua diagonal principal. Isso leva à conclusão que a imagem apresenta muitos pares de pixel com valores próximos. Se os pares considerados compõem-se de pixels adjacentes, isso implica em uma textura com frequências espaciais baixas, predominantemente. Altos níveis de contraste implicam em uma MCNC com valores significativos mais distantes da diagonal principal. Se o par de pixels for adjacente, conclui-se que as texturas apresentam componentes de frequências espaciais predominantemente altas.

A variância (f_4) mede a variação dos níveis de cinza em relação ao nível de cinza médio da textura. Ela tem forte correlação com a medida da variância dos níveis de cinza (que se conhece como contraste), e não apresenta correlação forte com a medida do contraste da MCNC explicada nos parágrafos anteriores deste capítulo.

A correlação (f_3) mede a dependência linear das cores dos pares de pixel considerados, na direção θ para a qual a matriz é calculada.

O momento inverso da diferença, ou homogeneidade (f_5), assume valores mais altos, próximos à unidade quando a GLMC apresenta valores mais significativos próximos à diagonal principal. Isso ocorre quando o contraste (f_2) é baixo para texturas que apresentam o mesmo nível de energia. Baraldi e Parmiggiani afirmaram que a homogeneidade é inversamente correlacionada simultaneamente ao contraste e à energia (BARALDI e PARMIGGIANI, 1995).

As Figuras 3.17 a 3.22 apresentam algumas texturas extraídas do álbum de Brodatz, a representação de suas respectivas matrizes $M(1,0^\circ)$ e os valores calculados para as características mencionadas. Para efeito de clareza, são apresentados os valores logarítmicos dos elementos das matrizes, e não seus valores originais.

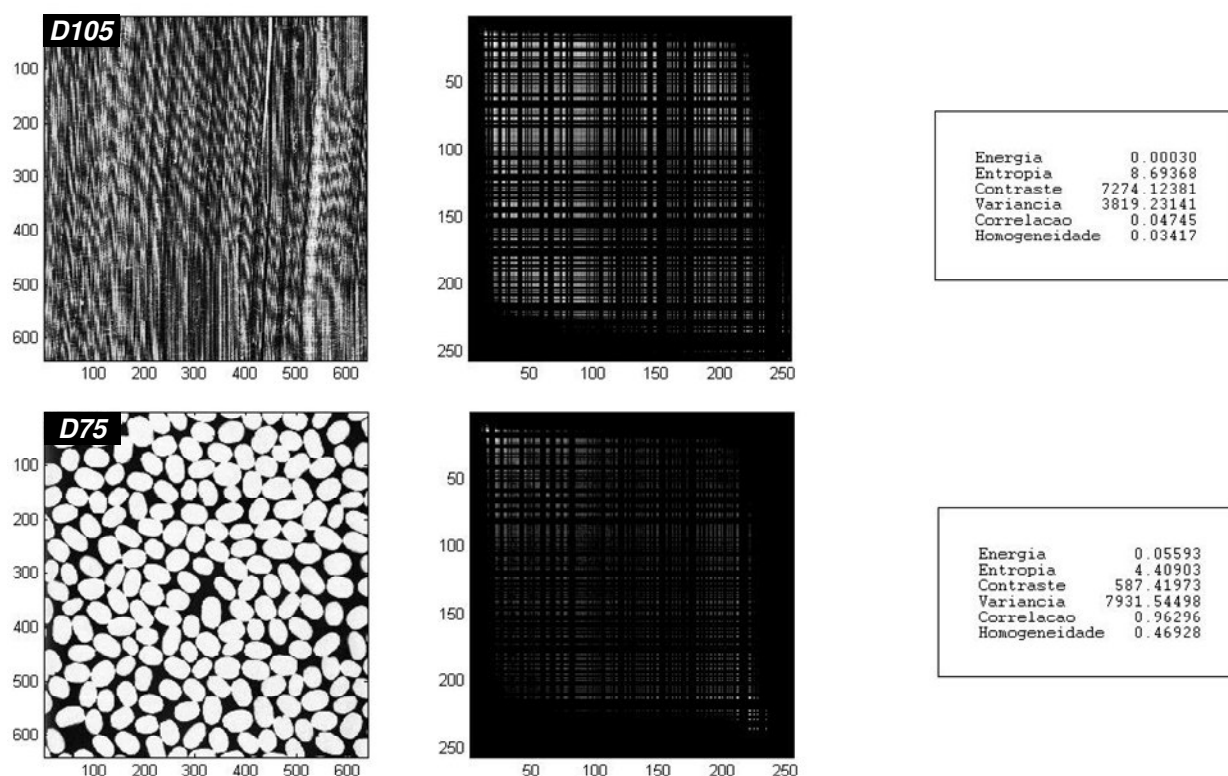


Figura 3.17 – Texturas com baixa (D105) e alta (D75) energia, suas respectivas matrizes MCNC e valores de características.

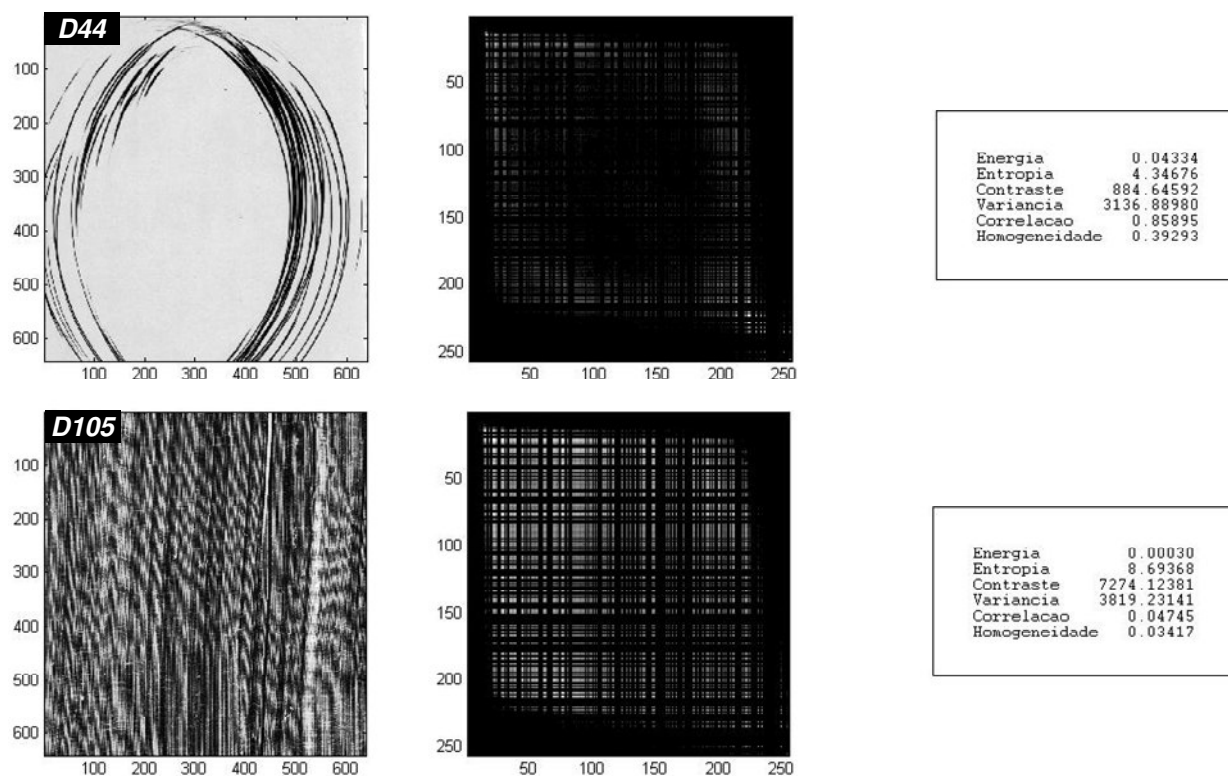


Figura 3.18 – Texturas com baixa (D44) e alta (D105) entropia, suas respectivas matrizes MCNC e valores de características.

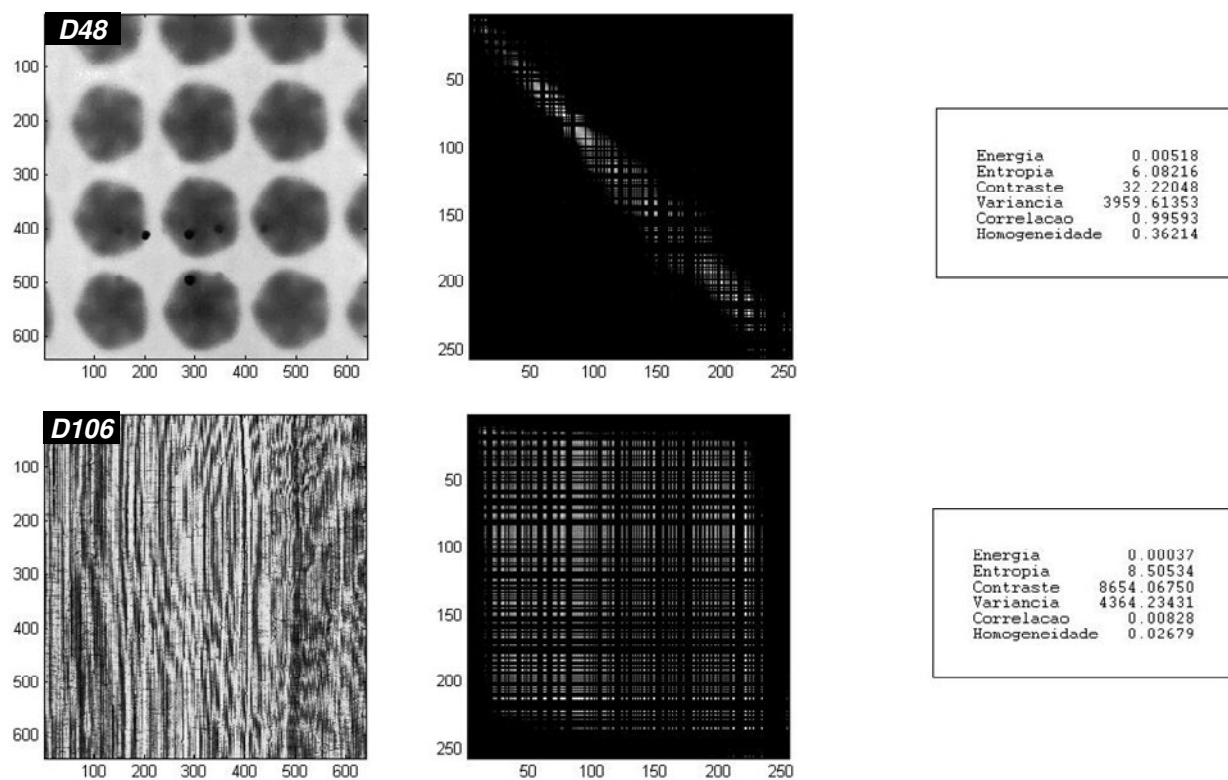


Figura 3.19 – Texturas com baixo (D48) e alto (D106) contraste, suas respectivas matrizes MCNC e valores de características.

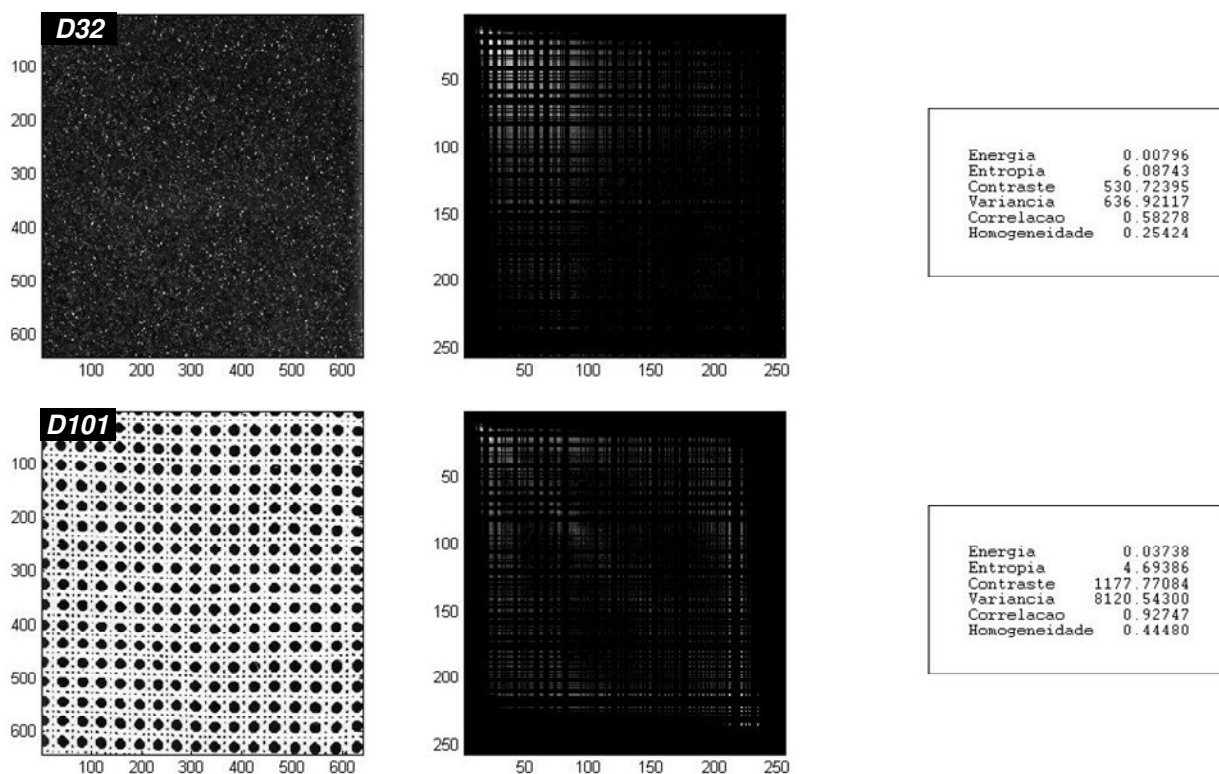


Figura 3.20 – Texturas com baixa (D32) e alta (D101) variância, suas respectivas matrizes MCNC e valores de características.

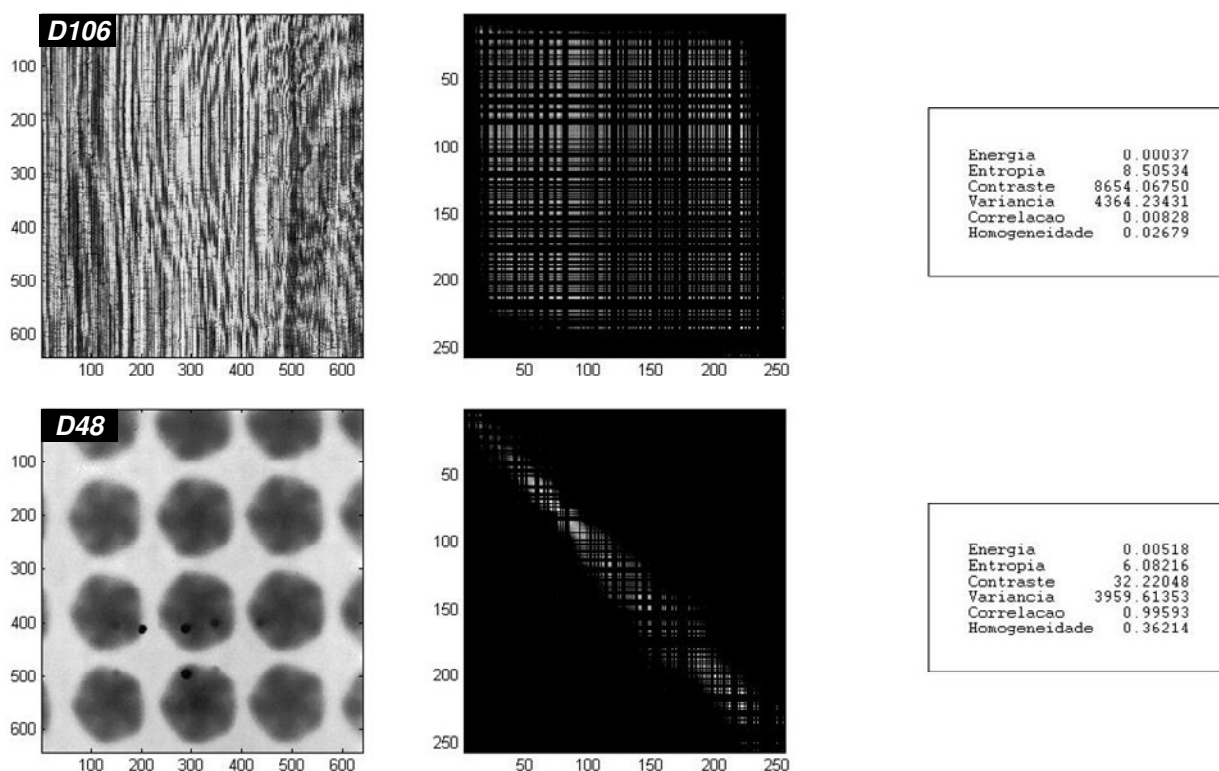


Figura 3.21 – Texturas com baixa (D106) e alta (D48) correlação, suas respectivas matrizes MCNC e valores de características.

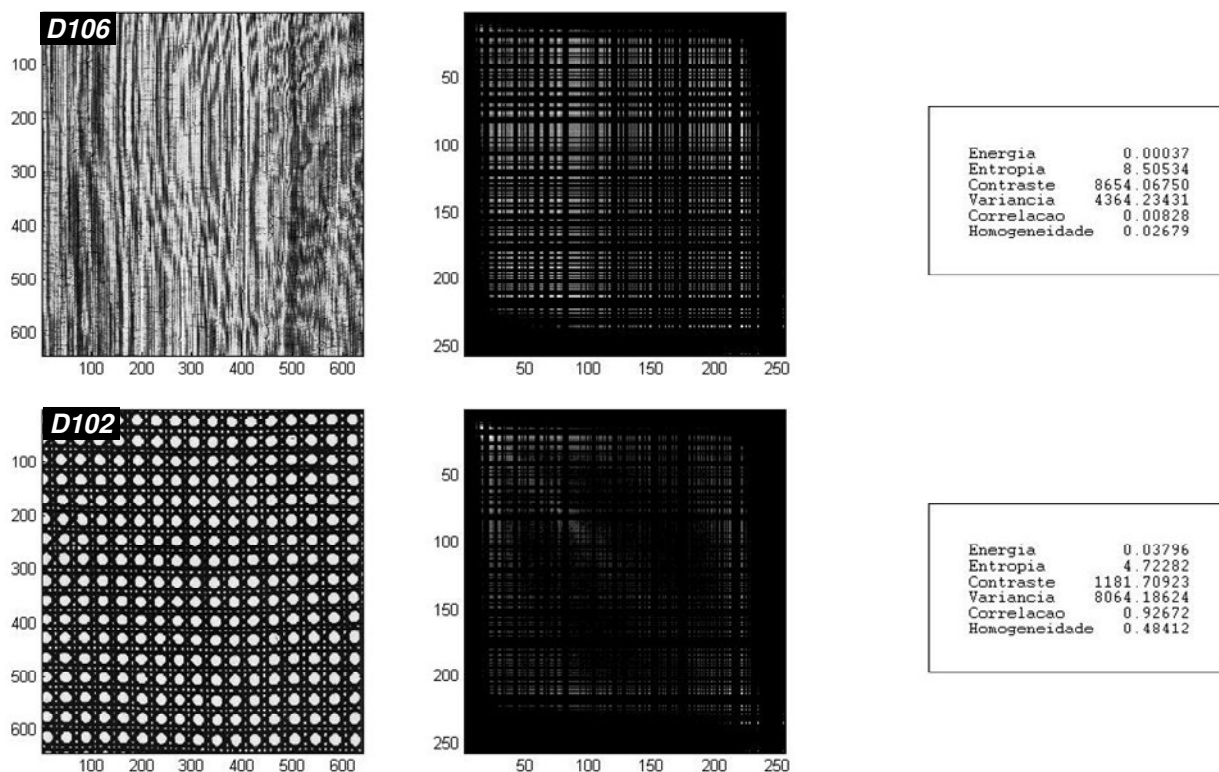


Figura 3.22 – Texturas com baixa (D102) e alta (D102) homogeneidade, suas matrizes MCNC e valores de características.

Pode-se perceber um efeito reticulado nas matrizes MCNC. Esse efeito deve-se ao fato das texturas terem sido submetidas à equalização de histogramas, o que elimina determinados valores de pixel do intervalo $[0,255]$. Se não existem pixels com determinado valor a , por exemplo, conseqüentemente os valores de $p(i,j)$ serão iguais a zero para $i = a$ e para $j = a$, determinando linhas negras horizontais e verticais nas posições com índice a .

A Figura 3.17 apresenta duas texturas com baixo e alto valor para a característica energia. Conforme observado anteriormente nesta seção, a textura D105 possui uma matriz MCNC com elementos de valor pequeno e distribuídos por toda sua extensão e, por conseguinte, baixo valor de energia. Já a textura D75 apresenta elementos da MCNC com valores elevados nas extremidades da diagonal principal. Isso faz com que seu valor de energia seja mais elevado.

A Figura 3.18 mostra as texturas D44 e D105, que apresentam valor baixo e alto de entropia, respectivamente. Conforme foi ressaltado, a entropia correlaciona-se

fortemente e de forma inversa com a energia, devido a isso o aparecimento da textura D105, mencionada no parágrafo anterior como tendo valor baixo de energia e alto de entropia.

A Figura 3.19 mostra as texturas D48, com baixo nível de contraste, e D106, com alto nível de contraste. Observando-se a MCNC da textura D48, pode-se verificar que seus elementos significativos concentram-se próximos à diagonal principal. Como o contraste mede a diferença média entre os níveis de cinza dos pares de pixel considerados, isso faz com que termos da diagonal principal contribuam com valores baixos de contraste. Já a MCNC da textura D106 apresenta elementos igualmente espalhados pela matriz, o que leva a valores maiores de contraste, o que está de acordo com o comportamento esperado.

É importante ressaltar a diferença mencionada anteriormente nesta seção entre o contraste da matriz MCNC e o contraste da imagem, que é medido também pela variância da matriz MCNC. A variância MCNC e o contraste da imagem medem o erro médio quadrático dos níveis de cinza em relação ao nível de cinza médio da imagem. O contraste MCNC mede o quadrático da diferença entre níveis de cinza dos pixels dos pares considerados, valor este que estima uma média da frequência espacial da imagem. Isso pode ser melhor observado na Tabela 3.1 que mostra os valores de contraste MCNC, da variância MCNC e do contraste da imagem calculados para as texturas D32, D48, D106, D75 e D101. As linhas estão ordenadas por variância MCNC, que é correlacionada com o contraste da imagem. Pode-se observar que as texturas D48 e D106, que apresentam o menor e o maior valor de contraste MCNC do conjunto considerado, apresentam valores próximos para as outras duas características. Uma análise visual dessas duas texturas levam à conclusão que a textura D48 apresenta predominantemente frequências baixas, enquanto que a textura D106 apresenta frequências preponderantemente altas na direção horizontal. As demais texturas apresentam frequências de valores intermediários, confirmando a argumentação anteriormente feita nesta seção.

Tabela 3.1 – Comparativo entre os valores de contraste MCNC, variância MCNC e contraste da imagem para algumas texturas do álbum de Brodatz. As linhas estão ordenadas pela Variância MCNC.

Textura	Contraste MCNC	Variância MCNC	Contraste
D32	531	637	1,91
D48	32	3960	2,29
D106	8654	4364	2,31
D75	587	7932	2,45
D101	1178	8121	2,46

A Figura 3.20 mostra as texturas D32 e D101, que apresentam valores baixos e altos para a variância MCNC. Conforme apresentado no parágrafo anterior, variância MCNC e contraste da imagem são fortemente correlacionados, o que pode ser observado pela observação das texturas. A primeira (D32) apresenta baixo nível de contraste, enquanto a segunda (D106) é composta de valores extremos da escala de níveis de cinza, ilustrando de forma clara o conceito de contraste.

A Figura 3.21 apresenta as texturas D106 e D48. A primeira possui um valor baixo de correlação na direção horizontal enquanto a segunda apresenta um valor alto para essa característica. A correlação está relacionada, aqui, com a previsibilidade do nível de cinza do segundo pixel do par, sendo que se sabe o nível de cinza do primeiro pixel do par. Analisando-se as linhas matriz MCNC da textura D106, que correspondem às cores do primeiro pixel do par, pode-se notar que elas apresentam valores significativos para uma ampla faixa de colunas, que correspondem às cores do segundo pixel dos pares. Isso significa que a cada cor do primeiro pixel estão associadas várias cores igualmente (ou aproximadamente) prováveis para o segundo pixel do par. Assim, a incerteza na estimativa é bastante grande.

Ao contrário, a textura D48 apresenta matriz MCNC com elementos significativos próximos à diagonal principal. Realizando-se análise semelhante, considerando uma linha correspondente à cor do primeiro pixel, as cores prováveis do segundo pixel estão contidas em uma faixa bastante mais estreita, o que torna a estimativa menos incerta.

Finalmente, a Figura 3.22 apresenta duas texturas com valores de homogeneidade discrepantes. A textura D106 possui um valor baixo de homogeneidade e a textura D102 possui um valor alto para essa característica. A homogeneidade ou momento

do inverso da diferença é mais sensível para texturas que apresentam MCNC com valores significativos mais próximos da diagonal principal. Por isso, relaciona-se inversamente com o contraste, como pode ser concluído pela observação das fórmulas de cálculo para ambas as características.

3.3.3 Os histogramas de soma e diferença

A utilização das características extraídas da matriz de co-ocorrência de níveis de cinza permite a obtenção de resultados interessantes nos problemas de classificação de texturas. No entanto, a dimensão dessas matrizes é uma desvantagem marcante dessa técnica, conforme mencionado nas observações finais da seção 3.3.1 deste capítulo. Uma alternativa bastante razoável para contornar esse problema foi proposta por Michael Unser e consiste em calcular algumas das características propostas por Haralick a partir dos histogramas da soma e da diferença entre níveis de cinza para os pares de *pixels* separados pela distância d na direção θ . Esses pares de *pixels* são exatamente os mesmos utilizados na construção da MCNC. Contudo, na técnica proposta por Michael Unser (UNSER, 1986), essa matriz MCNC não precisa ser calculada, o que permite uma redução dos requisitos de armazenamento da aplicação.

Os histogramas da soma e da diferença são estruturas bem mais compactas do que a matriz MCNC. Considerando-se uma imagem que apresente N_g níveis de cinza, uma matriz MCNC possui $N_g \times N_g$ elementos, enquanto que os histogramas da soma e da diferença possuem, cada um $2N_g$ elementos. Isso faz com que a utilização desses histogramas seja vantajosa sob o aspecto do armazenamento para situações em que se consideram imagens com mais de quatro níveis de cinza. Por exemplo, quando se consideram 64 níveis de cinza, o espaço de armazenamento necessário é da ordem de 16 vezes menor e para 256 níveis de cinza, 64 vezes menor quando os histogramas são utilizados em substituição da matriz MCNC.

Outra vantagem bastante notável decorre da dimensão dos histogramas. Sendo eles unidimensionais, o algoritmo que calcula as características apresenta ordem $O(N)$ em relação ao número de níveis de cinza considerados. No entanto, as matrizes MCNC

são estruturas bidimensionais e, portanto, conduzem à ordem $O(N^2)$ para os algoritmos que calculam as características. Assim, a utilização dos histogramas leva à economia tanto de espaço de armazenamento quanto de tempo de execução.

O cálculo dos histogramas da soma e da diferença é bastante simples e é efetuado de forma similar ao cálculo da matriz de co-ocorrência de níveis de cinza. Para cada par de *pixels* da imagem, calcula-se a soma e a diferença dos valores dos seus níveis de cinza, conforme a Equação (3.26).

$$s_{k,l} = f(k,l) + f(k+d_1, l+d_2)$$

$$d_{k,l} = f(k,l) - f(k+d_1, l+d_2)$$

(3.26)

onde :

$f(k,l)$: valor do nível de cinza do *pixel* localizado no ponto (k,l) ; e

(d_1, d_2) : deslocamento de tamanho d e direção θ .

Os histogramas da soma, h_s e da diferença, h_d , são definidos como o número de ocorrências, na imagem, dos pares de *pixels* que possuem valores $s_{k,l}$ e $d_{k,l}$, para cada um dos possíveis valores $s_{k,l}$ e $d_{k,l}$. Se na imagem existem N_g níveis de cinza, conclui-se que existirão $2N_g$ valores possíveis para soma e $2N_g$ valores possíveis para a diferença entre os níveis de cinza de um par de *pixels*.

A Equação (3.27) define os histogramas da soma e da diferença.

$$h_s(i, d_1, d_2) = h_s(i) = \text{Card}\{(k,l) \in D \mid s_{k,l} = i\}$$

$$h_d(i, d_1, d_2) = h_d(i) = \text{Card}\{(k,l) \in D \mid d_{k,l} = i\}$$

(3.27)

onde :

$\text{Card}\{\bullet\}$: número de ocorrências; e

D : domínio das coordenadas dos pontos da imagem.

Os histogramas normalizados, tanto o da soma quanto o da diferença são definidos pela Equação (3.28).

$$\hat{P}_s(i) = h_s(i)/N \quad (i = 2, 3, \dots, N_g)$$

$$\hat{P}_d(i) = h_d(i)/N \quad (i = -N_g + 1, \dots, N_g - 1)$$

onde : (3.28)

$$N = \sum_i h_s(i) = \sum_i h_d(i)$$

Unser demonstrou em seu trabalho (UNSER, 1986), que os histogramas normalizados da soma e da diferença podem ser utilizados para aproximar a densidade de probabilidade conjunta de segunda ordem evidenciada pela matriz de co-ocorrência de níveis de cinza. De fato, caso se parta do pressuposto que o nível de cinza de cada *pixel* de uma imagem qualquer possa ser modelado como uma variável aleatória e que a imagem como um todo seja uma realização de um processo estocástico estacionário e ergódico, pode-se afirmar que as variáveis aleatórias $s_{k,l}$ e $d_{k,l}$, correspondentes à soma e à diferença das variáveis originais, são descorrelatadas, pela suposição de estacionaridade. Se as variáveis forem Gaussianas, elas são, por consequência, independentes. Assim, a densidade de probabilidade conjunta pode ser calculada pelo produto das densidades dadas pelos histogramas da soma e da diferença. Entretanto, caso as variáveis obedeçam a outra distribuição, essa afirmação não é verdadeira. Todavia, Unser argumenta que, mesmo nesse caso, o produto dos histogramas ainda aproxima razoavelmente a densidade conjunta.

Unser (UNSER, 1986) mostrou que nove das 14 características propostas por Haralick podem ser bem aproximadas por cálculos efetuados a partir dos histogramas da soma e da diferença sem a necessidade da construção da matriz MCNC. Conforme argumentado anteriormente nessa seção, isso acarreta um ganho sensível tanto nos requisitos de memória quanto no tempo de processamento. As Equações de (3.29) a (3.37) apresentam as características que podem ser computadas pela utilização da técnica proposta.

$$Média = \frac{1}{2} \sum_i i \cdot \hat{P}_s(i) = \mu \quad (3.29)$$

$$Variância = \frac{1}{2} \left(\sum_i (i - 2\mu)^2 \cdot \hat{P}_s(i) + \sum_j j^2 \cdot \hat{P}_d(j) \right) \quad (3.30)$$

$$Energia = \sum_i \hat{P}_s(i)^2 \cdot \sum_j \hat{P}_d(j)^2 \quad (3.31)$$

$$Correlação = \frac{1}{2} \left(\sum_i (i - 2\mu)^2 \cdot \hat{P}_s(i) - \sum_j j^2 \cdot \hat{P}_d(j) \right) \quad (3.32)$$

$$Entropia = - \sum_i \hat{P}_s(i) \cdot \log(\hat{P}_s(i)) - \sum_j \hat{P}_d(j) \cdot \log(\hat{P}_d(j)) \quad (3.33)$$

$$Contraste = \sum_j j^2 \cdot \hat{P}_d(j) \quad (3.34)$$

$$Homogeneidade = \sum_j \frac{1}{1 + j^2} \cdot \hat{P}_d(j) \quad (3.35)$$

$$Cluster\ Shade = \sum_i (i - 2\mu)^3 \cdot \hat{P}_s(i) \quad (3.36)$$

$$Cluster\ Prominence = \sum_i (i - 2\mu)^4 \cdot \hat{P}_s(i) \quad (3.37)$$

Unser (UNSER, 1986) comparou o desempenho de sua técnica com o desempenho da técnica proposta por Haralick em ensaios de classificação de texturas. Para isso, selecionou um conjunto de características derivadas da MCNC usualmente utilizadas e selecionou o conjunto de características correspondentes derivadas dos histogramas da soma e da diferença. Os resultados obtidos por ambas as técnicas foram comparáveis, o

que corrobora sua proposta e evidencia a melhoria obtida tanto nos requisitos de memória quanto no tempo de processamento.

3.4 Matriz de Tamanho de Corridas de Níveis de Cinza

Ao analisar o problema da segmentação de imagens compostas por várias classes de terreno, Mary M. Galloway propôs uma técnica alternativa à utilização de matrizes de co-ocorrência de níveis de cinza. A técnica proposta por ela baseou-se também na construção de uma matriz que relacionou, para cada nível de cinza considerado, o número de corridas de *pixels* daquele nível para cada tamanho de corrida presente na imagem. A partir da construção da matriz de tamanhos de corridas de níveis de cinza (MTCNC), cinco medidas numéricas foram propostas e sua utilização na segmentação foi analisada (GALLOWAY, 1975). O objetivo desta seção é expor a técnica proposta por Mary Galloway, bem como apresentar aprimoramentos propostos por outros autores.

3.4.1 O cálculo da matriz MTCNC

Uma corrida é definida como uma seqüência colinear de *pixels* adjacentes que possuem o mesmo tom de cinza. Considerando-se uma seqüência de *pixels* colineares de mesma cor $A_i, A_{i+1}, A_{i+2}, \dots, A_{i+n}$, essa seqüência forma uma corrida se os *pixels* A_{i-1} ou A_{i+n+1} ou não existem porque a borda da figura foi atingida, ou possuem cor diferente dos demais. Nesse caso, o tamanho da corrida é $n+1$, ou seja, é o número de *pixels* de mesma cor. Uma corrida pode ter direção vertical, horizontal ou qualquer inclinação entre esses dois extremos. Para delimitar o problema, definem-se quatro direções principais que apresentam ângulos com a horizontal de 0° , 45° , 90° e 135° . Essas direções facilitam o cálculo da matriz MTCNC proposta, sendo que, para cada direção principal, é construída uma matriz de tamanhos de corrida.

Para exemplificar, a Figura 3.23 apresenta os valores dos *pixels* de uma pequena imagem 4×4 *pixels*. Os valores dos *pixels* em questão variam entre 0 e 3. As quatro matrizes principais também são mostradas nessa figura e têm dimensão 4×4 , já que tanto os tamanhos de corrida podem variar entre 1 e 4, como as cores podem variar entre 0 e 3.

Cada linha é composta pelo número de corridas de uma mesma cor para cada tamanho possível. Cada coluna é composta pelo número de corridas de mesmo tamanho para cada cor possível.

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

		Tamanhos de corrida				
0 grau			1	2	3	4
Cores	0	2	0	0	0	0
	1	2	0	0	0	0
	2	1	0	0	1	
	3	2	2	0	0	

		Tamanhos de corrida				
45 graus			1	2	3	4
Cores	0	2	0	0	0	0
	1	2	0	0	0	0
	2	6	0	0	0	
	3	4	1	0	0	

		Tamanhos de corrida				
90 graus			1	2	3	4
Cores	0	2	0	0	0	0
	1	2	0	0	0	0
	2	4	1	0	0	
	3	2	2	0	0	

		Tamanhos de corrida				
135 graus			1	2	3	4
Cores	0	2	0	0	0	0
	1	2	0	0	0	0
	2	4	1	0	0	
	3	4	1	0	0	

Figura 3.23 – Um exemplo de imagem e as quatro matrizes principais de comprimento de corrida de níveis de cinza.

Ao se observar a terceira linha da imagem mostrada na Figura 3.23, pode-se verificar uma corrida horizontal de quatro *pixels* de valor 2. Essa é a única corrida horizontal de tamanho 4 contendo pixels de valor 2 e esse fato está registrado na matriz 0°, na posição correspondente à cor 2 e ao tamanho 4.

O exemplo mostrado na Figura 3.23 é bastante simples. No entanto, quando se analisa uma imagem real, algumas considerações devem ser realizadas para que a técnica possa ser efetivamente utilizada com sucesso. Uma imagem real comum pode possuir muitos níveis de cinza mais que a imagem desse exemplo. Um valor comum para a resolução tonal é da ordem de 256. Imagens reais são também razoavelmente maiores do que a imagem desse exemplo, podendo atingir tamanhos da ordem de 256 x 256 *pixels*, ou mais. Isso faz com que os tamanhos de corrida possam variar de 1 a 256. Portanto,

assumindo uma imagem de dimensões e resolução tonal como as mencionadas, as matrizes de tamanho de corrida atingem a dimensão 256×256 , ou a extrapolam, o que as tornam proibitivas para a utilização com imagens reais. Além disso, quanto maior a resolução tonal da imagem, maior a probabilidade que pequenas e quase irrelevantes variações de cor interrompam uma corrida que, perceptivelmente, poderia ser maior, tornando os tamanhos médios de corridas artificialmente pequenos. Outra observação decorre do fato de que a matriz de tamanhos de corridas visa extrair da imagem características texturais, e não características tonais. Por isso, ao se considerar duas texturas semelhantes, mas que se distingam pelo nível médio de tonalidade, suas matrizes serão distintas, o que é indesejável.

Os problemas levantados no parágrafo anterior apresentam soluções bastante simples. Para que o efeito da diferença de nível médio de tonalidade, ou mesmo da diferença do desvio padrão seja resolvido, basta que a imagem seja submetida previamente a uma equalização de histograma. Isso faz com que todas as imagens apresentem distribuições de níveis de cinza próximas à uniforme, compensando as variações que, de outra forma, prejudicariam os resultados.

No que diz respeito à dimensão proibitiva dessas matrizes, ou mesmo com relação à grande resolução tonal, a solução é reduzir tanto o número de níveis de cinza quanto o número de tamanhos de corrida diferentes. Isso é feito quantizando-se a faixa de cores distintas bem como a faixa de tamanho de corrida em uma resolução menor, pelo agrupamento de valores semelhantes. Uma faixa de 256 níveis de cinza pode ser convertida em outra faixa com apenas oito tons distintos pelo mapeamento linear de cada grupo de 32 tons consecutivos em apenas um tom. Os tamanhos de corrida, da mesma forma, podem ser quantizados tanto de maneira uniforme quanto não uniforme para que as matrizes resultantes tenham, por exemplo, dimensão final de 8×8 , ao invés de 256×256 , como inicialmente estimado. Com isso, tanto o problema do tamanho excessivo, quanto o problema da falsa interrupção das corridas são mais bem tratados.

Galloway utilizou imagens de dimensão 64×64 *pixels*, com 64 níveis de cinza diferentes. Os níveis de cinza foram agrupados em oito conjuntos distintos de oito valores

adjacentes cada. Já os tamanhos de corrida foram agrupados não uniformemente nas seis seguintes faixas: 1, 2-3, 4-7, 8-15, 16-31 e 32-64. Percebe-se que uma quantização logarítmica foi utilizada. Portanto, as matrizes utilizadas apresentaram dimensão 8 x 6, o que é bastante razoável (GALLOWAY, 1975).

3.4.2 As características propostas

As matrizes de tamanho de corrida permitem analisar uma imagem do ponto de vista da granularidade. Uma imagem que apresente uma textura de granularidade grossa tenderá a apresentar um número grande de corridas grandes. Isso porque uma granularidade grossa implica em elementos de textura grandes ou largamente espaçados, o que leva à suposição de prevalência de corridas maiores. Já uma imagem de granularidade fina apresenta elementos de textura pequenos e bastante próximos, o que permite supor a predominância de corridas curtas. Esse raciocínio leva à possibilidade de determinação de medidas a partir das matrizes de tamanho de corrida que sejam significativas sob o aspecto sensível.

Galloway (GALLOWAY, 1975) propôs cinco medidas que visam expor algumas características importantes: a ênfase de corridas curtas (*short runs emphasis*), a ênfase de corridas longas (*long runs emphasis*), a desuniformidade de níveis de cinza (*gray level nonuniformity*), a desuniformidade de tamanhos de corrida (*run length nonuniformity*) e o percentual de corridas (*run percentage*). Esta seção descreve cada uma dessas características e outras características propostas posteriormente por outros autores visando ao refinamento da técnica.

Para estabelecer uma nomenclatura coerente, a Equação (3.38) define os termos utilizados.

N_g : número de níveis de cinza;

N_r : número de tamanhos de corrida;

$N_g \times N_r$: dimensão da matriz;

$p(i, j)$: cada elemento da matriz;

(3.38)

i : nível de cinza;

j : tamanho da corrida; e

P : número de pontos da imagem.

O número de corridas em uma imagem é calculado pela Equação (3.39).

$$NR = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j) \quad (3.39)$$

O objetivo da ênfase de corridas curtas (ECC) é realçar a presença de corridas curtas em uma imagem. Isso é realizado pela soma ponderada das corridas, na qual os pesos dados às corridas curtas são maiores do que os pesos dados às corridas longas. O valor de ECC é calculado pela Equação (3.40).

$$ECC = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{j^2} \bigg/ \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j) \quad (3.40)$$

Ao contrário da ECC, a ênfase em corridas longas (ECL) visa realçar a presença de corridas longas na imagem. Portanto, os pesos individuais para cada tamanho de corrida são maiores para as corridas de maior comprimento. O valor de ECL é obtido pela Equação (3.41).

$$ECL = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j) \bigg/ \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j) \quad (3.41)$$

O valor da desuniformidade de níveis de cinza (DNC) é calculado pela Equação (3.42).

$$DNC = \sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} p(i,j) \right)^2 \bigg/ \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j) \quad (3.42)$$

O valor da desuniformidade de tamanhos de corrida (DTC) é calculado pela Equação (3.43).

$$DTC = \sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} p(i,j) \right)^2 \bigg/ \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j) \quad (3.43)$$

O percentual de corridas (PC) visa medir a quantidade de corridas em relação ao número de pontos da imagem. Quanto menor esse valor, mais monótona é a textura. O valor de PC é obtido pela Equação (3.44).

$$PC = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j) \bigg/ P \quad (3.44)$$

As medidas de ênfase em corridas curtas ou longas podem ser melhor avaliadas comparando-se os valores dessas medidas calculados para quatro distribuições artificialmente compostas para esse fim, que são: distribuição decrescente, uniforme, crescente e aleatória. A Figura 3.24 mostra uma simulação do cálculo dessas ênfases para cada uma das quatro distribuições. Os valores entre parênteses na legenda referem-se à ênfase em corridas curtas (o primeiro) e em corridas longas (o segundo). Pode-se verificar claramente nessa figura que a ECC para a distribuição que apresenta preponderância de

corridas curtas (distribuição decrescente) é maior do que as demais. Ao invés, a ECL é maior na distribuição crescente, que apresenta maior número de corridas longas.

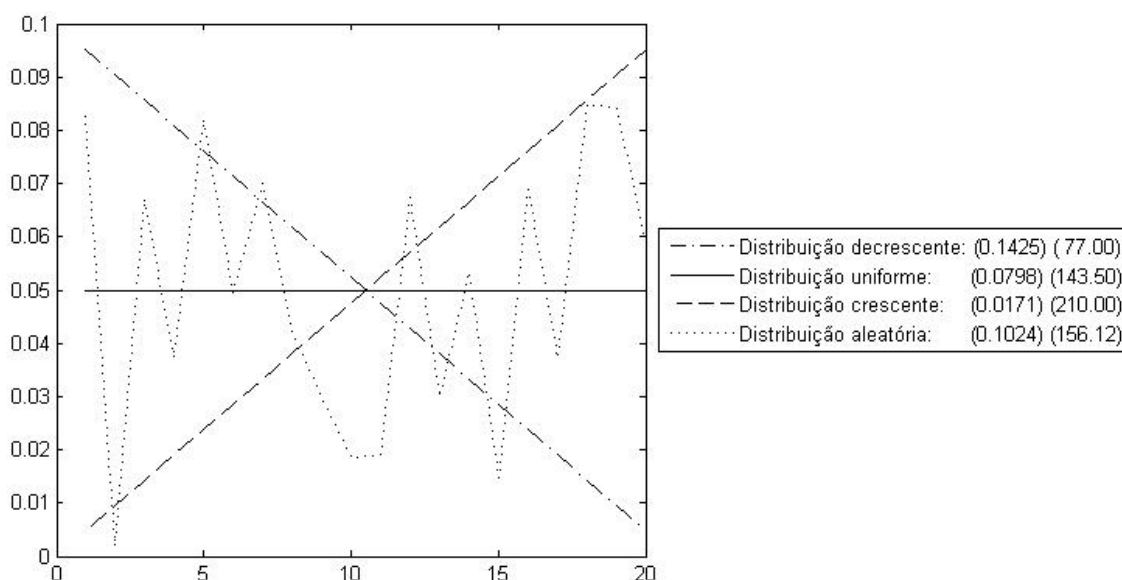


Figura 3.24 – Valores das ênfases em corridas curtas e longas para quatro distribuições.

As medidas de desuniformidade, tanto de níveis de cinza quanto de tamanhos de corrida, visam comparar a distribuição dessas características com uma situação padrão em que a distribuição é uniforme. Os menores valores para a desuniformidade são obtidos quando a distribuição da característica é uniforme. Se uma imagem apresenta números de corrida iguais para cada nível de cinza, a distribuição de corridas por nível de cinza é uniforme. De forma análoga, se a imagem apresenta números de corrida iguais para cada tamanho de corrida, a distribuição de corridas por tamanho também é uniforme.

A Figura 3.25 mostra uma simulação do cálculo da desuniformidade para quatro distribuições de características. A primeira é uma distribuição decrescente, ou seja, a imagem apresenta maior presença de valores menores da característica. A segunda é uma distribuição uniforme, a terceira é uma distribuição decrescente e a última é uma distribuição aleatória. Os valores entre parênteses nos itens da legenda correspondem ao

cálculo da desuniformidade para cada caso. Pode-se notar que o menor valor corresponde à distribuição uniforme, como ressaltado anteriormente neste parágrafo.

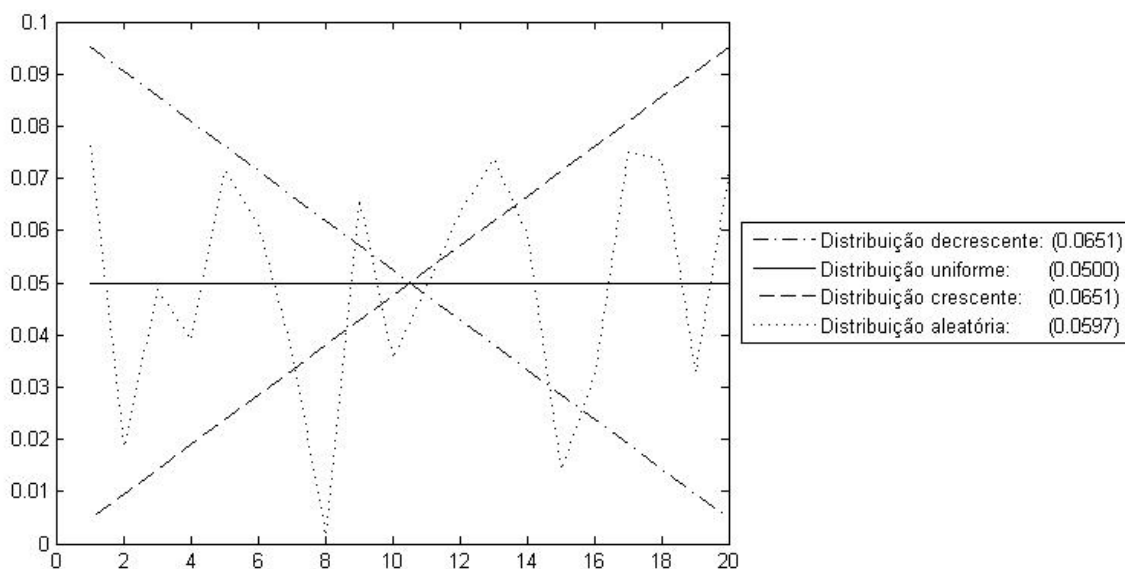
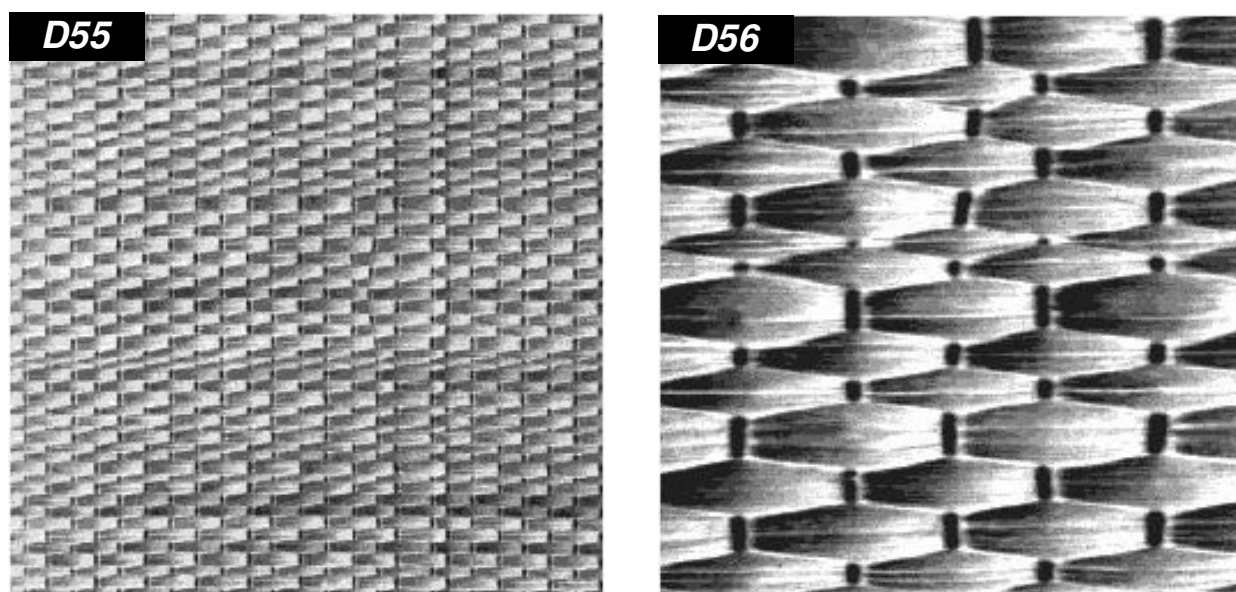


Figura 3.25 – Valores da desuniformidade para quatro distribuições.

Os valores de medida de textura propostos anteriormente nessa seção são utilizados isoladamente, ou em conjunto, como características nas aplicações de segmentação, classificação ou recuperação de imagens.

Para ilustrar a técnica das matrizes MTCNC, a Figura 3.26 apresenta as texturas D55 e D56 equalizadas e quantizadas, suas respectivas matrizes de tamanhos de corrida de cinza e os respectivos valores de ECC, ECL, DNC e DTC.



D55 0 grau		Tamanhos de corrida									
Cores		1	2	3	4	5	6	7	8		
	1	0	807	12	2	0	0	0	0	ECC	0.2040
	2	0	1320	238	58	0	0	0	0	ECL	6.0789
	3	0	2591	1338	607	0	0	0	0	DNC	4680.8000
	4	0	3797	1614	222	0	0	0	0	DTC	14024.0000
	5	0	4214	1587	155	1	0	0	0		
	6	0	3725	1246	71	0	0	0	0		
	7	0	1331	1534	365	0	0	0	0		
	8	0	240	61	1	0	0	0	0		

D56 0 grau		Tamanhos de corrida									
Cores		1	2	3	4	5	6	7	8		
	1	0	296	483	168	187	4	0	0	ECC	0.1775
	2	0	1166	533	353	73	0	0	0	ECL	8.0715
	3	0	1430	635	302	35	0	0	0	DNC	2389.9000
	4	0	1605	687	316	33	0	0	0	DTC	6204.8000
	5	0	1849	760	277	16	0	0	0		
	6	0	1874	822	218	7	0	0	0		
	7	0	1707	832	184	6	0	0	0		
	8	0	518	317	330	176	17	0	0		

Figura 3.26 – As texturas D55 e D56, suas respectivas matrizes de corrida de cinza e as características calculadas a partir delas.

Dos dados apresentados na Figura 3.26, pode-se observar que a textura D55 apresenta valores maiores para a ECC e menores para a ECL, quando comparados com os mesmos dados da textura D56. Isso concorda com a observação, já que a textura D55 apresenta granularidade mais fina que D56, logo, suas corridas devem ser menores.

3.4.3 Alguns aprimoramentos

Ao analisar imagens celulares com o objetivo de classificação, Chu, Sehgal e Greenleaf (CHU *et alli*, 1990) observaram que o conjunto de medidas proposto por Galloway não era tão abrangente como poderia ser. De fato, ECC, ECL e DTC dependem ou podem ser escritas como função apenas da distribuição dos tamanhos de corrida, praticamente desconsiderando a distribuição de níveis de cinza. Isso pode ser constatado pelas Equações (3.38) a (3.44) e é mostrado na Equação (3.45).

Sejam :

$r(j) = \sum_i p(i, j)$: número de corridas de tamanho j ; e

$g(i) = \sum_j p(i, j)$: número de corridas de cor i .

As medidas de Galloway (GALLOWAY, 1975) podem ser escritas como :

$$ECC = \frac{1}{NR} \sum_j \frac{r(j)}{j^2}$$

$$ECL = \frac{1}{NR} \sum_j j^2 r(j)$$

$$DTC = \frac{1}{NR} \sum_j (r(j))^2$$

(3.45)

$$DNC = \frac{1}{NR} \sum_i (g(i))^2$$

O valor $r(j)$ significa a soma de todos os elementos contidos na coluna j da matriz, ou seja, é a soma das quantidades de corrida de comprimento j para todas as cores i . Ao somar as quantidades para todas as cores, é ignorada a informação contida na distribuição de cores. O não aproveitamento dessa informação faz com que seja impossível distinguir duas texturas cujas matrizes apresentam conteúdo diferente, mas o mesmo total de corridas de mesmo tamanho. Isso ocorre porque ambas as matrizes apresentam os

mesmos valores para $r(j)$ e, portanto, os mesmos valores para as medidas ECC, ECL, DTC e PC propostas por Galloway (GALLOWAY, 1975).

Para superar essa fraqueza, Chu e outros (CHU *et alli*, 1990) propuseram duas novas medidas que podem ser consideradas simétricas às medidas ECC e ECL. Eles as chamaram de ênfase de corridas de nível de cinza baixo, ECB, (*low gray level runs emphasis*) e ênfase de corridas de nível de cinza alto, ECA, (*high gray level runs emphasis*). A medida ECB corresponde à medida ECC e a medida ECA corresponde à medida ECL. Essas medidas são obtidas na Equação (3.46).

$$ECB = \frac{1}{NR} \sum_{j=1}^{N_r} \sum_{i=1}^{N_g} \frac{p(i, j)}{i^2} \quad (3.46)$$

$$ECA = \frac{1}{NR} \sum_{j=1}^{N_r} \sum_{i=1}^{N_g} i^2 p(i, j)$$

As medidas ECB e ECA devem ser utilizadas em conjunto com as outras cinco (NR, ECC, ECL, DTC e PC) e trazem melhoria dos resultados. No entanto, Belur Dasarathy e Edwin Holder (DASARATHY e HOLDER, 1991) observaram posteriormente que essas medidas, apesar de gerarem bons resultados não aproveitam todo o potencial das matrizes de tamanho de corridas de níveis de cinza. Eles observaram em seu artigo que as medidas propostas anteriormente não levam em conta a distribuição conjunta das duas grandezas em questão: tamanho de corridas e níveis de cinza. Cada uma das medidas utiliza apenas uma delas de cada vez, desprezando a outra. Dasarathy e Holder propuseram então uma categoria geral de medidas da qual as anteriores são casos especiais. Essa categoria geral, que eles nomearam de característica geral de tamanho de corrida de nível de cinza (CGTCNC), (*general run length gray level feature*) é apresentada na Equação (3.47).

$$CGTCNC = \frac{1}{NR} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^{k_i} j^{k_j} p(i, j)$$

onde : (3.47)

$k_j \neq 1$ para que a soma não seja independente de j ; e

$k_i \neq 0$ e $k_j \neq 0$ para considerar a distribuição conjunta.

Pela Equação (3.47), pode-se concluir que ECC é obtido de CGTCNC fazendo-se $k_i = 0$ e $k_j = -2$. Para se obter ECL, fazem-se $k_i = 0$ e $k_j = 2$ em CGTCNC. Raciocínio semelhante permite derivar também ECB e ECA da característica geral.

As quatro medidas propostas por Dasarathy e Holder (DASARATHY e HOLDER, 1991) para substituir aquelas utilizadas anteriormente são combinações conjuntas das possíveis ênfases: corridas curtas com níveis de cinza baixos (ECCNB), (*short-run low gray-level emphasis*), corridas curtas com níveis de cinza altos (ECCNA), (*short-run high gray-level emphasis*), corridas longas com níveis de cinza baixos (ECLNB), (*long-run low gray-level emphasis*) e corridas longas com níveis de cinza altos (ECLNA), (*long-run high gray-level emphasis*). As Equações (3.48) a (3.51) definem essas medidas:

$$ECCNB = \frac{1}{NR} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{i^2 j^2} \quad (3.48)$$

$$ECCNA = \frac{1}{NR} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 \frac{p(i, j)}{j^2} \quad (3.49)$$

$$ECLNB = \frac{1}{NR} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 \frac{p(i, j)}{i^2} \quad (3.50)$$

$$ECLNA = \frac{1}{NR} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 j^2 p(i, j) \quad (3.51)$$

Ao aplicar essas quatro medidas (ECCNB, ECCNA, ECLNB e ECCNA) ao mesmo conjunto de dados analisado por Chu e outros (CHU *et alli*, 1990), Dasarathy e Holder obtiveram resultados melhores e mais confiáveis do que aqueles obtidos anteriormente (DASARATHY e HOLDER, 1991). Isso leva à conclusão de que a utilização de todo o potencial das distribuições conjuntas acarreta maiores benefícios que a utilização em separado das distribuições marginais.

3.5 A Representatividade da Percepção das Características e a Matriz da Diferença de Níveis de Cinza da Vizinhança

Várias características texturais foram apresentadas nas seções anteriores deste capítulo, conforme a proposição de Haralick e seus colaboradores. Baraldi e Parmiggiani (BARALDI e PARMIGGIANI, 1995) buscaram explicar o significado de algumas dessas características e analisar a correlação entre elas. No entanto, dois outros artigos (TAMURA *et alli*, 1978) e (AMADASUN e KING, 1989) vão mais além e aprofundam a análise da relação entre as características texturais e as percepções visuais associadas a essas características. O tema é bastante interessante, pois a utilização de um conjunto de características de textura cujas percepções visuais sejam bem determinadas é de grande utilidade, visto que permite a construção de aplicativos mais amigáveis para a recuperação de imagens em bancos de dados. Em ambos os artigos, os autores propuseram um conjunto de medidas de textura e as analisam sob a ótica da percepção visual.

No primeiro artigo, de Hideyuki Tamura, Shunji Mori e Takashi Yamawaki (TAMURA *et alli*, 1978), eles definiram seis tipos de características visuais de textura: *coarseness*, *contrast*, *directionality*, *line-likeness*, *regularity* e *roughness*. A partir dessas definições, eles fizeram um experimento psicológico em que várias fotos de texturas são apresentadas ao público duas a duas e cada participante indica qual das fotos apresenta maior valor para cada característica.

As respostas foram trabalhadas e originaram uma tabela que mostra seis colunas, uma para cada característica, contendo a ordenação das texturas segundo aquela característica. Além disso, foi montada uma tabela de coeficientes de correlação entre as características. Esses coeficientes mostraram como cada uma delas se relacionava com as demais.

Em seguida, Tamura e outros (TAMURA *et alii*, 1978) propuseram um conjunto de medidas computacionais, uma para cada característica definida anteriormente. Testaram essas medidas no conjunto de texturas utilizado no experimento psicológico e compararam os resultados dos dois experimentos. Finalmente eles testaram a utilização das medidas propostas para a determinação da semelhança entre as texturas. Um segundo experimento psicológico foi realizado e os dados tabulados e comparados com os resultados obtidos pelo conjunto de medidas computacionais. Os resultados não foram ótimos, porém indicaram um caminho para as pesquisas futuras.

As características definidas por Tamura e outros (TAMURA *et alii*, 1978) são:

Coarseness: indica a densidade da textura. Quanto menores e mais juntos estão os elementos da textura, mais densa ela é. Quanto maiores ou mais espaçados os elementos da textura, menos densa, ou mais esparsa ela é. A medida utilizada foi proposta por Rosenfeld (HAYES *et alii*, 1974, *apud* TAMURA *et alii*, 1978) e adaptada por Tamura e outros. Ela consiste em uma suavização da imagem por um filtro quadrado de lado L seguido do cálculo da diferença entre os pontos separados pela distância L. Esse processo é repetido para vários valores de L, potências de dois, e a medida da densidade (*coarseness*) é encontrada como a dimensão do quadrado correspondente ao maior valor da diferença.

Contrast: indica o contraste da textura. Tamura e outros (TAMURA *et alii*, 1978) definiram o contraste segundo quatro componentes: i) Devido à distribuição de níveis de cinza; ii) Devido à polarização de níveis de cinza no histograma; iii) Devido à agudeza das bordas e iv) Devido ao período de repetição dos elementos da textura. Eles criticaram a medida de contraste de Haralick porque ela não espelha o conceito de contraste da percepção visual. Ao invés de usar a medida proposta por Haralick (HARALICK, 1973),

utilizaram uma combinação entre o desvio padrão e a curtose dos níveis de cinza, obtendo resultados razoáveis.

Directionality: indica se uma textura é direcional ou não. Esse conceito está relacionado tanto à forma do elemento quanto à distribuição dos elementos na textura. Tamura e outros (TAMURA *et alii*, 1978) buscaram medir o grau de direcionalidade e não a direção específica. Duas texturas iguais que diferem por uma rotação devem apresentar o mesmo grau de direcionalidade. A medida proposta por eles deriva do histograma de direção das bordas, ou do gradiente, em que se desprezam os pontos em que o módulo do gradiente é pequeno. A presença de picos acentuados mostra que a textura é mais direcional. Histogramas achatados indicam uma textura pouco ou nada direcional. A medida final relaciona-se à agudeza dos picos, medida pela variância em torno dos picos: quanto mais agudos, mais direcional é a textura.

Line-likeness: indica se um elemento é composto de linhas. Quando as direções de bordas de um ponto e sua vizinhança são iguais, ou parecidas, supõe-se que esses pontos de borda formam uma linha. Ao invés, quando a diferença entre as direções de borda aproxima-se de 90 graus, supõe-se o oposto. A partir da matriz de co-ocorrência de direções de bordas pode-se computar um índice que mede a característica de *line-likeness*.

Regularity: indica a regularidade dos elementos de textura, ou seja, a uniformidade dos padrões elementares e a simetria de posicionamento desses elementos na textura. Tamura e outros (TAMURA *et alii*, 1978) afirmaram a dificuldade de se definir a regularidade e propuseram uma medida indireta que contabiliza quatro características primitivas em sub-imagens não entrelaçadas da imagem original. As primitivas são: *coarseness* (densidade), *contrast* (contraste), *directionality* (direcionalidade) e *line-likeness*. Para cada sub-imagem, o valor das quatro características é computado e a medida de regularidade é calculada como um valor proporcional à soma dos desvios padrão de cada característica na imagem.

Roughness: indica a aspereza ou suavidade da textura, conforme a sensação tátil. Tamura e outros (TAMURA *et alii*, 1978) ressaltaram a dificuldade de se definir a aspereza e propuseram uma medida indireta que combina *coarseness* e contraste.

O segundo trabalho, de Moses Amadasun e Robert King (AMADASUN e KING, 1989), segue uma linha semelhante ao trabalho de Tamura e outros. Amadasun e King propuseram outro conjunto de medidas de textura que defenderam ser mais representativos das propriedades de percepção visual, a saber: *coarseness*, *contrast*, *busyness*, *complexity* e *strength*. Definidas as medidas, submeteram os resultados de sua aplicação a ensaios de ranqueamento, correlacionando os resultados do *ranking* das medidas calculadas com os resultados do *ranking* obtido da opinião de um conjunto de observadores.

O que distingue o trabalho de Amadasun e King (AMADASUN e KING, 1989) do trabalho de Tamura e outros (TAMURA *et alii*, 1978) é o método de cálculo das características. Tamura e outros utilizaram várias técnicas distintas para o cálculo das características, cada uma baseada em um conceito diferente como a distância entre as diferenças máximas das médias de níveis de cinza, ou o relacionamento entre média, variância e curtose da distribuição de níveis de cinza, ou ainda o histograma da magnitude das bordas em relação com sua direção. Ao contrário, Amadasun e King propuseram um procedimento homogêneo a partir da montagem de uma estrutura especial que chamaram de matriz da diferença de níveis de cinza da vizinhança, MDTCV (*neighborhood Gray-tone difference matrix*, NGTDM). As características propostas por eles são calculadas a partir dos valores dos elementos dessa matriz.

Na verdade, a MDTCV é um vetor unidimensional, um histograma, que atribui a cada nível de cinza presente na imagem o valor da soma das diferenças absolutas entre o nível de cinza de um pixel central e o valor médio dos níveis de cinza de sua vizinhança. O tamanho da vizinhança é um parâmetro que pode ser ajustado à aplicação. A Equação (3.52) define a média dos níveis de cinza da vizinhança.

$$\bar{A}_i = \bar{A}(k, l) = \frac{1}{W-1} \left[\sum_{\substack{m=-d \\ (m,n) \neq (0,0)}}^d \sum_{n=-d}^d f(k+m, l+n) \right]$$

onde :

(3.52)

d : tamanho da vizinhança;

$W = (2d+1)^2$; e

$i = f(k, l)$: nível de cinza do *pixel* nas coordenadas k e l da imagem.

Os elementos da matriz MDTCV são obtidos pela Equação (3.53).

$$s(i) = \begin{cases} \sum_{i \in N_i} |i - \bar{A}_i|, & \text{se } N_i \neq \emptyset \\ 0, & \text{caso contrário} \end{cases}$$

(3.53)

onde :

N_i : conjunto de todos os *pixels* que têm cor i ,
exceto as bordas de tamanho d .

Outro conceito importante é o da probabilidade de ocorrência de níveis de cinza na região central da imagem. Essa é a região da imagem que excetua as bordas de largura d , conforme caracterizado nas Equações (3.52) e (3.25). Essa borda é desconsiderada para evitar distorções. A probabilidade de ocorrência de níveis de cinza p_i é calculada pela Equação (3.54).

$$p_i = N_i / n^2$$

onde :

(3.54)

$n = N - 2d$; e

N : tamanho do lado de uma imagem $N \times N$.

A partir da construção da matriz das diferenças de níveis de cinza da vizinhança, MDTCV, pode-se calcular as características propostas por Amadasun e King (AMADASUN e KING, 1989). A medida *coarseness* visa capturar a presença de grandes áreas uniformes decorrentes da presença, na imagem, de primitivas texturais de dimensões maiores, ou mais espaçadas. O conceito de percepção visual dessa característica é próximo ao conceito proposto por Tamura e outros (TAMURA *et alii*, 1978). Amadasun e King (AMADASUN e KING, 1989) ressaltaram que a presença de tais estruturas implica em menores taxas de variação de intensidade da textura. A Equação (3.55) define a fórmula de cálculo para *coarseness*.

$$Coarseness = \frac{1}{\varepsilon + \sum_i p_i \cdot s(i)} \quad (3.55)$$

Uma imagem apresenta altos níveis de contraste, segundo Amadasun e King (AMADASUN e KING, 1989), quando se pode distinguir claramente entre áreas de níveis de cinza diferentes. O contraste relaciona-se também com a escala dinâmica de níveis de cinza e com o nível da taxa de variação espacial da intensidade. Todos esses argumentos estão em consonância com as observações de Tamura e outros (TAMURA *et alii*, 1978) e de Baraldi e Parmiggiani (BARALDI e PARMIGGIANI, 1995). Assim, a medida de contraste apresenta dois termos: o primeiro leva em conta a escala dinâmica dos níveis de cinza e o segundo contabiliza a variação do nível de cinza entre os *pixels* e sua vizinhança, contribuindo para medir a quantidade de variação local de intensidade. A expressão que define a medida de contraste é apresentada na Equação (3.56).

$$Contrast = \left[\frac{1}{N(N-1)} \sum_i \sum_j p_i p_j (i-j)^2 \right] \left[\frac{1}{n^2} \sum_i s(i) \right]$$

$$N = \sum_i Q_i$$

onde : (3.56)

$$Q_i = \begin{cases} 0, & \text{se } p_i \neq 0 \\ 1, & \text{caso contrário} \end{cases}$$

Q_i : número de níveis de cinza presentes efetivamente na imagem.

A característica *busyness* pode ser entendida como a rápida e grande variação entre os níveis de cinza na imagem. Nesse conceito são fundamentais tanto a taxa quanto a intensidade de variação. A presença de altas taxas de variação de pequena intensidade é a percepção de uma área razoavelmente homogênea. Do mesmo modo, pequenas taxas de variação de grande intensidade resultam na percepção de várias regiões quase homogêneas. A medida do *busyness* deve levar em conta ambos esses fatores. É o que se apresenta na Equação (3.57). O numerador contribui com a taxa de variação e o denominador com sua intensidade.

$$Busyness = \frac{\sum_i p_i \cdot s(i)}{\sum_i \sum_j i \cdot p_i - j \cdot p_j}, \quad \text{onde } p_i \neq 0, p_j \neq 0 \quad (3.57)$$

A complexidade relaciona-se com a quantidade de informação apresentada pela imagem. Quanto mais complexa, mais informação contém a imagem. Cabe dizer que a complexidade relaciona-se diretamente com a entropia, que não deixa de ser uma forma de medida da informação. Sob esse aspecto, imagens com grande entropia apresentam muitos níveis de cinza quase que equiprováveis. Assim, pode-se relacionar complexidade diretamente com *busyness* e com o contraste. A medida da complexidade é apresentada na Equação (3.58). Pode-se observar um termo direto da diferença absoluta de níveis de cinza

acumulados aos pares. Esse termo contabiliza a contribuição do contraste. Observa-se também o termo da soma ponderada das entradas da MDTCV, que contribui com a taxa de variação local dos níveis de cinza. O denominador é um fator de normalização que equivale à soma do número de ocorrências dos níveis de cinza considerados.

$$Complexity = \sum_i \sum_j \left[\frac{|i-j|}{n^2(p_i + p_j)} \right] [p_i \cdot s(i) + p_j \cdot s(j)] \quad (3.58)$$

Das características propostas por Amadasun e King (AMADASUN e KING, 1989), *strength*, ou força da textura é a que apresenta conceituação mais difícil. Uma textura é forte se suas primitivas podem ser distinguidas com clareza. Grandes primitivas que possuam pequenas diferenças de tonalidade são facilmente identificáveis. Ao contrário, para se distinguir entre pequenas primitivas, elas devem apresentar diferenças de tonalidade mais marcantes. Assim, a força da textura relaciona-se com *coarseness* e com o contraste. A Equação (3.59) define a medida da característica *strength* e seu numerador contribui com a parcela do contraste enquanto seu denominador contabiliza a contribuição de *coarseness*.

$$Strength = \frac{\sum_i \sum_j (p_i + p_j) \cdot (i-j)^2}{\mathcal{E} + \sum_i s(i)} \quad (3.59)$$

Amadasun e King (AMADASUN e KING, 1989) aplicaram sua técnica na classificação de fotografias aéreas de campos de culturas tendo obtido resultados satisfatórios. A vantagem da técnica proposta é a facilidade de computação da matriz MDTCV (de fato um vetor unidimensional) e das características derivadas dela. Tanto os requisitos de armazenamento quanto o tempo de processamento são bastante modestos, permitindo sua implementação em uma ampla gama de equipamentos.

3.6 Espectro de Textura e seus Aprimoramentos

No final da década de 80, Dong-Chen He e Li Wang (HE e WANG, 1990) propuseram uma técnica de análise de texturas que apresentava uma abordagem diferente em diversos aspectos. Em seus trabalhos, definiram o que chamaram de unidade de textura (*texture unit*), um conceito que associa um valor de fácil obtenção a um bloco de pixels de dimensão 3×3 . Esse valor visa identificar padrões de textura significativamente distintos. Eles definiram também o conceito de espectro de textura (*texture spectrum*), que nada mais é que o histograma de ocorrências das unidades de textura em uma imagem. A partir dessa definição, eles utilizaram esses conceitos na classificação de texturas e na segmentação de imagens com resultados bastante significativos. Em uma de suas publicações, He e Wang (HE e WANG, 1991) abordaram o tema, apresentaram as definições e os métodos de cálculo e propuseram características extraídas do espectro de textura. Os conceitos de unidade e espectro de textura foram desenvolvidos e aprimorados posteriormente por outros autores, permitindo a obtenção de melhores resultados (OJALA *et alii*, 1996), (PIETIKÄINEN e OJALA, 2000), (OJALA *et alii*, 2001), (OJALA *et alii*, 2002).

O objetivo desta seção é apresentar os conceitos de unidade de textura e de espectro de textura, mostrando detalhes de seu cálculo e sua aplicabilidade. Visa também apresentar os aperfeiçoamentos desenvolvidos bem como as vantagens de sua utilização.

3.6.1 A unidade de textura

A unidade de textura, UT, é definida para uma região da imagem correspondente a um quadrado de 3×3 *pixels*. Essa região é composta de um *pixel* central, de índice $i = 0$ e dos oito *pixels* periféricos que compõem a vizinhança do *pixel* central. Aos *pixels* da vizinhança são atribuídos os índices $i = 1, 2, \dots, 8$. O *pixel* superior esquerdo recebe o índice $i = 1$ e os demais são numerados seqüencialmente, a partir do primeiro, no sentido horário. Denota-se por V_i o nível de cinza do *pixel* superior esquerdo e os demais são considerados ordenadamente no sentido horário. A unidade de textura é o conjunto ordenado de oito valores definidos na Equação (3.60).

$$UT = \{E_1, E_2, \dots, E_8\}$$

onde : (3.60)

$$E_i = \begin{cases} 0 & \text{se } V_i < V_0 \\ 1 & \text{se } V_i = V_0 \\ 2 & \text{se } V_i > V_0 \end{cases}$$

Cada elemento da unidade de textura pode assumir um dos três valores, dependendo de como o valor do *pixel* correspondente se relaciona com o valor do *pixel* central. Se o valor do *pixel* da vizinhança é menor do que o do *pixel* central, o elemento correspondente tem valor zero; se os níveis de cinza forem iguais, o elemento terá valor um e se o nível de cinza do *pixel* vizinho é maior do que o do *pixel* central, o elemento terá valor dois. Dessa forma, existem $3^8 = 6561$ possíveis combinações distintas para os elementos do conjunto UT e cada uma delas pode receber como identificador um número do intervalo $[0, 6560]$, computado conforme mostra a Equação (3.61).

$$N_{UT} = \sum_{i=1}^8 E_i \times 3^{i-1} \quad (3.61)$$

É importante observar que o valor do número de unidade de textura N_{UT} não é uma medida, já que não existe escala relacionada às unidades de textura. O valor de N_{UT} é simplesmente um identificador para cada um dos possíveis arranjos de *pixels* para uma vizinhança 3×3 quando é aplicada a definição da Equação (3.60). Outra observação importante é que a unidade de textura não leva em conta os valores exatos das tonalidades de cinza, mas sim a relação entre a tonalidade do *pixel* central e as tonalidades de cada um de seus vizinhos. A Figura 3.27 apresenta um exemplo de cálculo do N_{UT} .

IMAGEM			$UT = \{E_i\}$			POTÊNCIAS DE 3		
4	12	15	0	2	2	1	3	9
5	10	10	0		1	2187		27
6	17	7	0	2	0	729	243	81

$$NUT = 0 \times 1 + 2 \times 3 + 2 \times 9 + 1 \times 27 + 0 \times 81 + 2 \times 243 + 0 \times 729 + 0 \times 2187 = 537$$

Figura 3.27 – Exemplo de cálculo do número de unidade de textura.

As Equações (3.60) e (3.61) definem arbitrariamente uma ordem para o conjunto UT que define os valores do N_{UT} . Contudo, pode-se perceber que essa ordem não é única. De fato, existem oito formas distintas de ordenamento seqüencial dos pixels da vizinhança e, por conseguinte, oito formas distintas para o cálculo do N_{UT} . A primeira é a que foi arbitrada e as demais podem ser calculadas pela atribuição de cada um dos demais *pixels* como sendo o *pixel* inicial da seqüência. Como mostra a Figura 3.28.

Seqüência arbitrada	Seqüências alternativas																												
<table> <tr><td>V_1</td><td>V_2</td><td>V_3</td></tr> <tr><td>V_8</td><td>V_0</td><td>V_4</td></tr> <tr><td>V_7</td><td>V_6</td><td>V_5</td></tr> </table>	V_1	V_2	V_3	V_8	V_0	V_4	V_7	V_6	V_5	<table> <tr><td>V_8</td><td>V_1</td><td>V_2</td></tr> <tr><td>V_7</td><td>V_0</td><td>V_3</td></tr> <tr><td>V_6</td><td>V_5</td><td>V_4</td></tr> </table>	V_8	V_1	V_2	V_7	V_0	V_3	V_6	V_5	V_4	<table> <tr><td>V_7</td><td>V_8</td><td>V_1</td></tr> <tr><td>V_6</td><td>V_0</td><td>V_2</td></tr> <tr><td>V_5</td><td>V_4</td><td>V_3</td></tr> </table>	V_7	V_8	V_1	V_6	V_0	V_2	V_5	V_4	V_3
V_1	V_2	V_3																											
V_8	V_0	V_4																											
V_7	V_6	V_5																											
V_8	V_1	V_2																											
V_7	V_0	V_3																											
V_6	V_5	V_4																											
V_7	V_8	V_1																											
V_6	V_0	V_2																											
V_5	V_4	V_3																											

A posição do pixel inicial está ressaltada com fundo preto e a do pixel central com fundo cinza.

O reticulado da esquerda mostra a ordem arbitrada e os demais mostram duas seqüências alternativas, das oito sepossíveis

Figura 3.28 – Seqüências distintas para o cálculo do N_{UT} .

A Equação (3.61) pode ser alterada para levar em conta a seqüência $k = (1, 2, \dots, 8)$ utilizada para o cálculo do N_{UT} , onde k significa o índice do *pixel* inicial da seqüência. A Equação (3.62) apresenta a nova definição.

$$N_{UT(k)} = \sum_{i=1}^8 E_{i \text{rot } k} \times 3^{i-1}$$

onde: (3.62)

$$i \text{ rot } k = \{[(i-1) - (k-1)] \bmod 8\} + 1$$

Pode-se dizer que a unidade de textura busca capturar a distribuição espacial das variações de níveis de cinza na região analisada, mas não se preocupa em quantificá-las. Esse fato acarreta duas conseqüências imediatas. A primeira é que as informações que podem quantificar exatamente as medidas de contraste são perdidas, já que ficam registradas apenas as relações de ordem e não as diferenças entre níveis de cinza. São perdidas também as informações de média e variância da região analisada. Isso pode ser tomado como uma desvantagem, já que várias técnicas apresentadas anteriormente neste capítulo propõem medidas de contraste, ou de variância. A segunda conclusão, é que, ao ignorar a média e a variância da região da unidade de textura, possíveis variações tonais na imagem devidas ao sombreamento, ou mesmo a variações da escala dinâmica de imagens diferentes, porém semelhantes, são imediatamente compensadas quando os números de unidade de textura N_{UT} são calculados para várias regiões da imagem. Isso é vantajoso, porque torna desnecessário o processamento prévio, como por exemplo, a equalização de histogramas, destinado a uniformizar as escalas de níveis de cinza de imagens diferentes.

3.6.2 O espectro de textura

A unidade de textura revela o aspecto local da textura em um determinado ponto da imagem. He e Wang defenderam que o aspecto textural global de uma região suficientemente grande pode ser descrito pela freqüência com que as unidades de textura aparecem nessa região (HE e WANG, 1991). O espectro de textura de uma imagem é definido como o histograma do número de unidade de textura, N_{UT} , para essa imagem. Portanto, o espectro de textura é um vetor contendo 6561 posições indexadas de 0 a 6560, em que cada posição contém o número de ocorrências, na imagem, do N_{UT} correspondente.

Os N_{UT} são simples identificadores, e não medidas referentes a uma escala. Portanto, não existe relação de ordem entre eles. Isso leva à conclusão que o espectro de textura é um histograma nominal. Além disso, o histograma de texturas pode ser calculado para cada um dos oito $N_{UT(k)}$, o que significa que existem oito histogramas de textura diferentes, denotados por S_k . A rigor todos os histogramas S_k são equivalentes, já que o que os distingue é a ordem entre seus elementos. Para histogramas nominais, a ordem de seus elementos não é importante, mas para o cálculo das características propostas por He e Wang (HE e WANG, 1991), a distinção entre os S_k é fundamental.

A Figura 3.29 mostra a textura D106 e dois histogramas de textura: S_1 e S_2 . Pode-se notar que ambos os histogramas apresentam os mesmos valores, só que reorganizados em ordens diferentes.

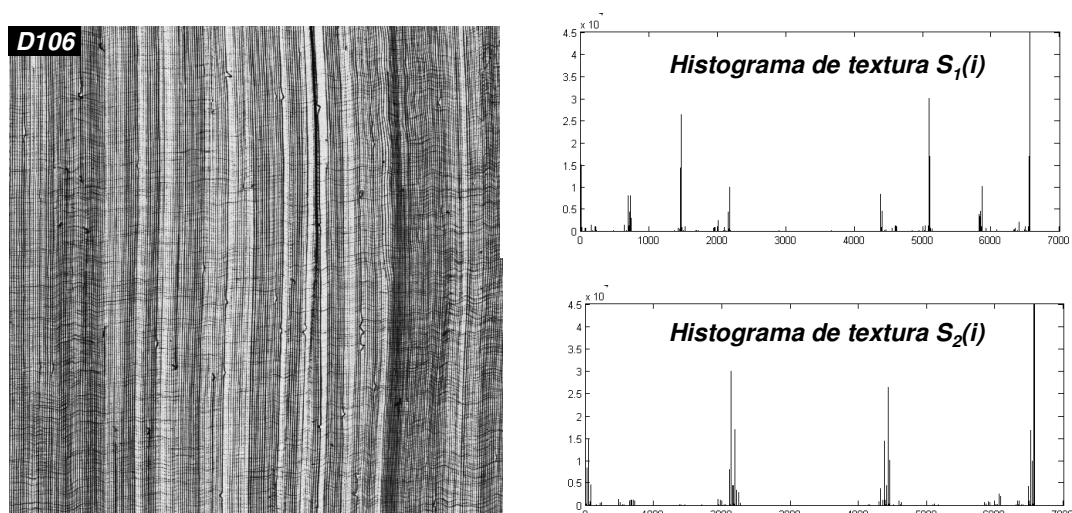


Figura 3.29 – A textura D106 e dois histogramas de textura.

3.6.3 As características propostas

Os histogramas de textura podem ser utilizados para a determinação de semelhança entre texturas. Isso pode ser feito pelo cálculo da distância nominal entre os histogramas de textura de duas texturas diferentes. No entanto, He e Wang (HE e WANG, 1991) propuseram algumas características que podem ser extraídas dos histogramas de textura. Segundo eles, essas características podem ressaltar algumas propriedades

importantes das texturas, podendo ser utilizadas nos experimentos de classificação. Esta seção apresenta três das características propostas por He e Wang e breves considerações sobre elas.

A primeira característica proposta por He e Wang (HE e WANG, 1991) é a simetria preto/branco, *SPB (Black-white symmetry)*. Essa característica mede a simetria entre as metades esquerda e direita do espectro de textura e deve apresentar valores altos quando, ao se inverter os níveis de cinza de uma imagem, a imagem resultante for parecida com a original. O motivo é que inverter os níveis de cinza equivale a inverter os valores dos elementos da unidade de textura: os elementos que apresentam valor 0 passam a apresentar o valor 2 e vice-versa. Os elementos que apresentam valor 1 permanecem inalterados. Essa inversão tem o efeito de transformar o número da unidade de textura N_{UT} em seu simétrico em relação à posição central do histograma. A Equação (3.63) apresenta a medida proposta. Cabe observar que a diferença absoluta subtrai os números de unidade de textura simétricos. Quando uma imagem apresenta números de ocorrência próximos para elementos de textura simétricos, ou seja, um grande grau de simetria preto/branco, a soma das diferenças absolutas fica próxima de zero e o SBP aproxima-se de 100. Caso contrário o valor dessa soma aproxima-se do valor do somatório do denominador, a fração aproxima-se da unidade e o SBP aproxima-se de zero. A equação original apresentada por He e Wang (HE e WANG, 1991) foi corrigida.

$$SPB = \left[1 - \frac{\sum_{i=0}^{3279} |S(i) - S(6560 - i)|}{\sum_{i=0}^{6560} S(i)} \right] \times 100 \quad (3.63)$$

A segunda característica proposta por He e Wang (HE e WANG, 1991) foi a simetria geométrica, *SG (Geometric symmetry)*, cuja fórmula de cálculo é apresentada na Equação (3.64). Essa medida compara os valores dos elementos de dois espectros de textura com ordenamentos da unidade de textura separados de 180° . O termo $S_j(i)$ faz referência ao espectro de frequência em que as unidades de textura têm como elemento

inicial o *pixel* V_j . Seguindo o raciocínio, o termo $S_{j+4}(i)$ faz referência ao espectro cuja *UT* tem como elemento inicial o *pixel* V_{j+4} . Pode-se observar na Figura 3.28, que os *pixels* cujos índices diferem de quatro são diametralmente opostos pelo elemento central, ou seja, estão localizados a 180° um do outro. Assim, a diferença absoluta acumulada na Equação (3.64) é tão menor quanto mais próximos forem os espectros diametralmente opostos. O denominador é um fator de normalização. Quando se calcula a média dos quatro pares de espectros diametralmente opostos, o resultado fica próximo de zero e, conseqüentemente *SG* resulta próximo de 100 quando, ao rotacionar a imagem de 180° , a imagem rotacionada se parece com a imagem original. Quanto menor a semelhança, mais o valor de *SG* se aproxima de zero.

$$SG = \left[1 - \frac{1}{4} \sum_{j=1}^3 \frac{\sum_{i=0}^{6560} |S_j(i) - S_{j+4}(i)|}{2 \times \sum_{i=0}^{6560} S_j(i)} \right] \times 100 \quad (3.64)$$

A outra característica proposta por He e Wang (HE e WANG, 1991) tem nome de grau de direcionalidade, *GD* (*degree of directionality*) e é mostrada na Equação (3.65). Essa característica é calculada pela comparação de histogramas com ordenamentos distintos, dois a dois, com a acumulação das diferenças absolutas de seus elementos. Eles afirmaram que uma imagem que possui alto valor de *GD* apresenta algumas “estruturas lineares de elementos fundamentais da imagem”. Entretanto, a tradução desse conceito em percepção visual não ficou clara quando dos testes com texturas do álbum de Brodatz.

$$GD = \left[1 - \frac{1}{6} \sum_{m=1}^3 \sum_{n=m+1}^4 \frac{\sum_{i=0}^{6560} |S_m(i) - S_n(i)|}{2 \times \sum_{i=0}^{6560} S_m(i)} \right] \times 100 \quad (3.65)$$

Outras características foram propostas no trabalho de He e Wang. Elas foram chamadas de características de orientação e simetria central. Todas essas medidas foram utilizadas em dois ensaios de classificação, e os resultados foram comparados com o

desempenho de um conjunto de características propostas por Haralick (HARALICK *et alii*, 1973). Os resultados foram promissores, segundo eles.

3.6.4 Alguns aperfeiçoamentos – os padrões binários locais

A unidade de textura e o correspondente histograma de textura propostos por He e Wang (HE e WANG, 1991) foram conceitos originais e bastante promissores, como eles ressaltaram. Embora o conceito seja interessante, existe uma desvantagem de ordem prática que deve ser discutida. O domínio do histograma, composto de todas as 6561 possibilidades de combinação de valores dos elementos da unidade de textura, é bastante extenso. Esse número de combinações faz com que o tempo de execução dos cálculos para a determinação de semelhança de texturas diferentes consuma mais tempo e ocupe muito espaço no banco de dados. Como agravante, impõe um limite inferior para o tamanho das regiões comparadas para que os resultados sejam expressivos.

O histograma de textura pode ser considerado como uma aproximação de uma densidade de probabilidade de ocorrência, na região analisada, das 6561 unidades de textura possíveis. Para que essa aproximação seja significativa, é necessário que o número de UT computadas seja grande o suficiente para evitar distorções indesejáveis. Se uma média de dez UT por *slot* for arbitrada, por exemplo, o número total de UT que a região deve conter supera o valor de 65000. Como cada *pixel* da imagem corresponde a uma UT (exceto os *pixels* da borda), conclui-se que a menor região a ser analisada deve ter 65000 *pixels*, ou seja, deve ter dimensão mínima em torno de 256×256 *pixels*. Se para aplicações de recuperação de imagens uma região com tamanho dessa ordem é adequada, para a segmentação, essa ordem de grandeza é demasiado inconveniente porque pode limitar a detecção correta de pequenos retalhos de textura.

Como aperfeiçoamento da técnica, uma proposta alternativa apresentada por Timo Ojala, Matti Pietikäinen e Davis Harwood segue a linha dos conceitos de unidade e histograma de texturas, mas alivia as restrições sobre as dimensões da região analisada pela diminuição do tamanho do histograma (OJALA *et alii*, 1996). Ojala e seus colaboradores definiram os padrões binários locais, PBL (*Local binary patterns*) de modo

semelhante ao exposto na Equação (3.60). A diferença surge no valor dos elementos do PBL. Ao invés de distinguir três valores possíveis, correspondentes às três possibilidades de relação entre cada *pixel* da vizinhança e o *pixel* central, eles propuseram apenas duas relações entre eles, conforme mostrado na Equação (3.66).

$$UT = \{E_1, E_2, \dots, E_8\}$$

onde : (3.66)

$$E_i = \begin{cases} 0 & \text{se } V_i < V_0 \\ 1 & \text{se } V_i \geq V_0 \end{cases}$$

Essa alteração faz com que cada elemento do PBL possa assumir apenas dois valores possíveis. Assim, o universo de PBL foi reduzido para apenas 256 possibilidades diferentes. Cabe observar que a proposta original alterou o ordenamento seqüencial dos elementos no sentido horário dos *pixels*. A nova proposta de Ojala e outros (OJALA *et alii*, 1996) considerou a ordem dos *pixels* variando da esquerda para a direita e de cima para baixo. Em trabalhos posteriores, eles retomaram a ordem cíclica no sentido horário (PIETIKÄINEN *et alii*, 2000), (OJALA *et alii*, 2001). A consequência mais importante da nova proposta foi a redução do tamanho do histograma de textura de 6561 para 256 elementos, reduzindo, portanto, os aspectos negativos da técnica de He e Wang (HE e WANG, 1991).

Os números dos padrões binários locais (PBL) são calculados pela Equação (3.67).

$$N_{UT} = \sum_{i=1}^8 E_i \times 2^{i-1} \quad (3.67)$$

A Figura 3.30 exemplifica o cálculo do valor do PBL segundo a ordem cíclica horária.

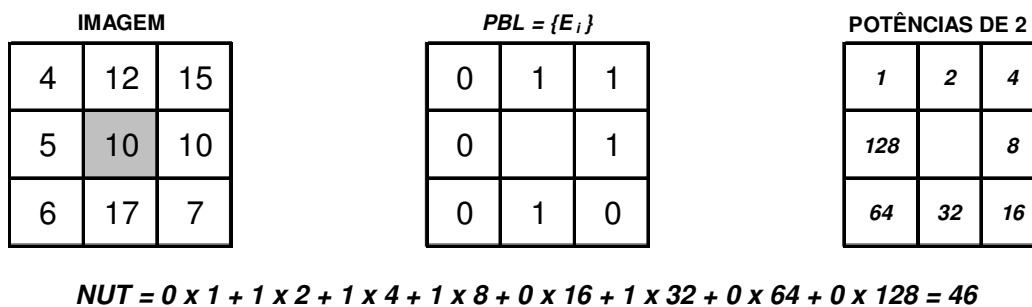


Figura 3.30 – Exemplo de cálculo do valor do padrão binário local.

Como os histogramas de padrões binários locais baseiam-se no conceito de unidades de textura, também valem para eles as observações realizadas para os histogramas de texturas. Os valores dos PBL são meros identificadores e não constituem uma medida, já que não há escala de medida para eles. Assim, os histogramas PBL são histogramas de medida nominal. Além disso, como os valores de PBL são calculados presumindo-se uma ordem arbitrária cíclica e horária de seus elementos, existem oito possíveis casos diferentes, um para cada *pixel* da vizinhança tomado como *pixel* inicial. As Figuras 3.31 a 3.34, mostram quatro texturas diferentes e seus respectivos histogramas PBL.

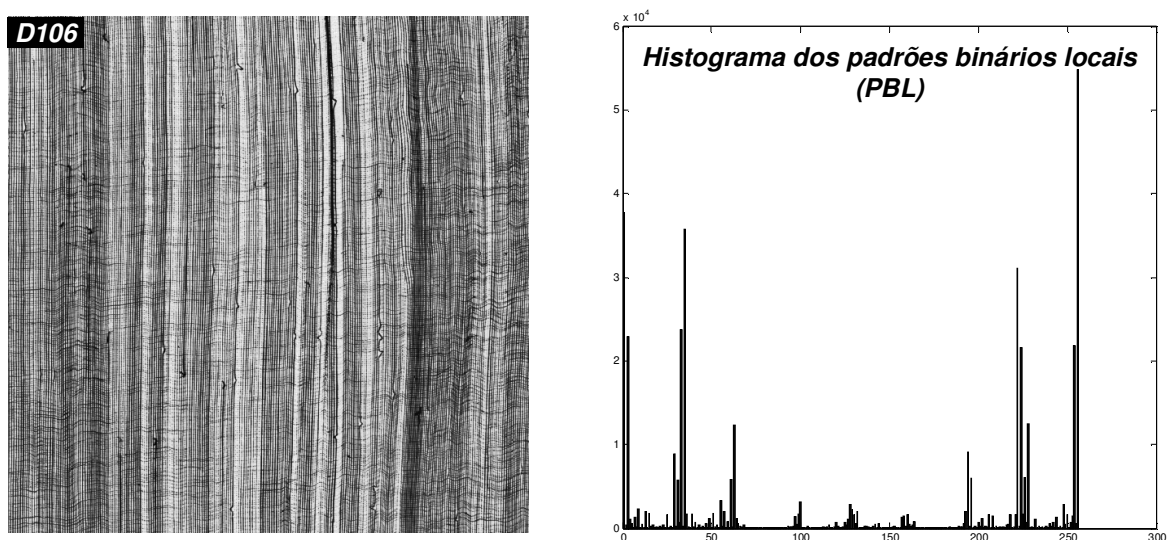


Figura 3.31 – Textura D106 e seu histograma PBL.

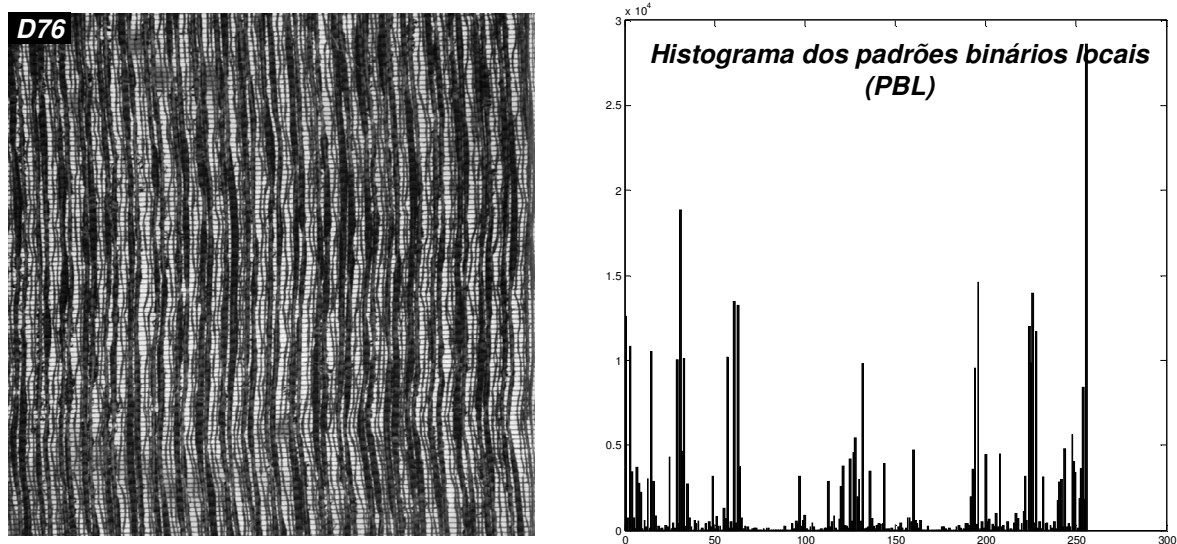


Figura 3.32 – Textura D76 e seu histograma PBL.

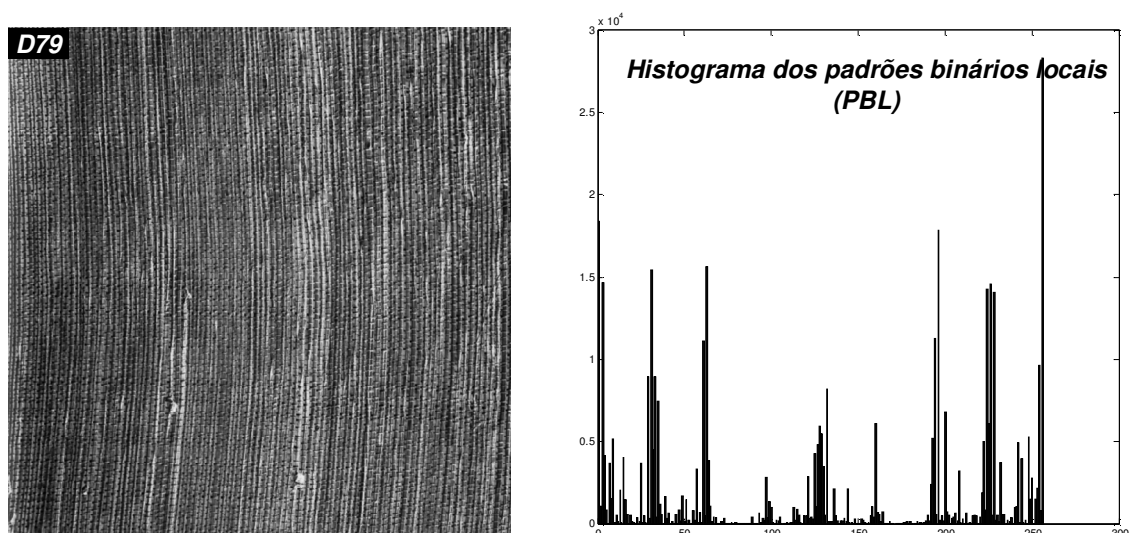


Figura 3.33 – Textura D79 e seu histograma PBL.

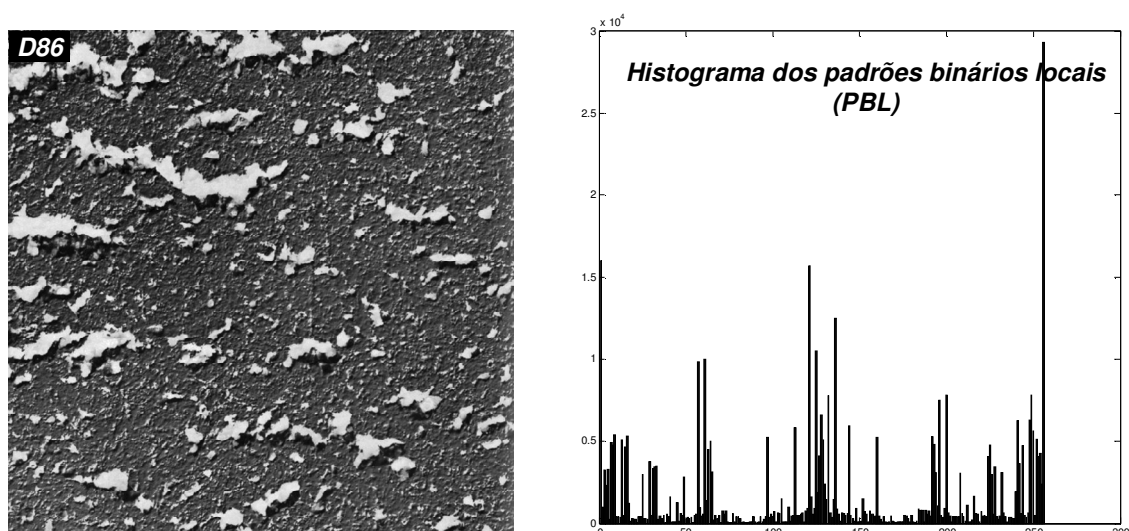


Figura 3.34 – Textura D86 e seu histograma PBL.

Pode-se comparar os histogramas das quatro texturas mostradas nas Figuras 3.31 a 3.34 usando a distância nominal, já que a ordem dos *bins* não é importante. O resultado da comparação é apresentado na Tabela 3.2. Os valores de distância pertencem ao intervalo $[0,1]$, onde o valor zero significa coincidência e um significa maior dessemelhança.

Tabela 3.2 – Distâncias nominais dos histogramas das texturas D76, D79, D86 e D106.

<i>Distância normalizada</i>	<i>D76</i>	<i>D79</i>	<i>D86</i>	<i>D106</i>
<i>D76</i>		0,1473	0,3116	0,4327
<i>D79</i>	0,1473		0,3425	0,3786
<i>D86</i>	0,3116	0,3425		0,5869
<i>D106</i>	0,4327	0,3786	0,5869	

Pode-se observar na Tabela 3.2 que as texturas mais semelhantes, segundo a distância dos histogramas PBL, são D76 e D79. Deve-se ressaltar que os histogramas PBL medem a distribuição de ocorrência das micro-texturas de regiões 3×3 . Visualmente as texturas D106 e D76 são muito semelhantes. Entretanto, os resultados indicaram que a textura D79 é a mais semelhante à D106. A Figura 3.35 mostra micro-regiões ampliadas das quatro texturas mostradas nas Figuras 3.31 a 3.34, para efeito de ilustração.

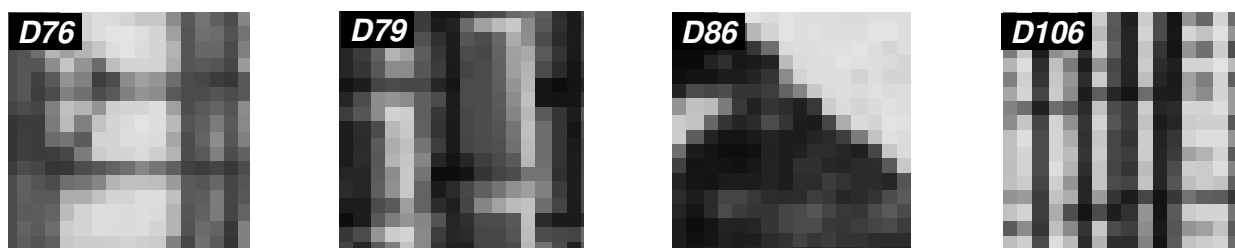


Figura 3.35 – Realce da micro-textura de regiões das texturas D76, D79, D86 E D106.

Ojala e outros (OJALA *et alii*, 1996) relataram resultados bastante positivos nas experiências em que compararam o desempenho dos padrões binários locais com outras técnicas. Os resultados foram melhores quando os PBL foram conjugados com medidas de covariâncias centrais simétricas, CCS, propostas por David Harwood e outros (HARWOOD *et alii*, 1995). Isso porque os PBL são relativamente invariantes à escala dinâmica de níveis de cinza, bem como a variações monotônicas dos níveis de cinza. Já as medidas de covariâncias centrais simétricas computam valores relacionados à escala de níveis de cinza. Como PBL e CCS são medidas pouco ou nada correlacionadas, sua combinação leva a melhores resultados.

Em um trabalho posterior, Ojala e outros (OJALA *et alii*, 2001) retomaram os testes dos PBL, justificaram melhor seu funcionamento e compararam seu desempenho com campos aleatórios gaussianos de Markov (*Gaussian Markov random fields*) e com filtragem de Gabor. Os resultados reportados foram favoráveis aos PBL.

A discussão sobre os PBL realizada por Ojala e outros (OJALA *et alii*, 1996) apresenta-os como uma simplificação de distribuições multidimensionais de diferenças espaciais de níveis de cinza. Retomam-se as propostas de Haralick (HARALICK *et alii*, 1973) sobre as distribuições conjuntas de níveis de cinza em dois pontos e de Unser (UNSER, 1986) sobre os histogramas da soma e da diferença. As matrizes de co-ocorrência de níveis de cinza são poderosas ferramentas para a análise de texturas porque utilizam estatística de segunda ordem. Elas retratam a distribuição conjunta de níveis de cinza em dois pontos separados por uma distância arbitrada em uma direção determinada. Essa distribuição pode ser reescrita baseada na diferença de níveis de cinza, como mostrado na Equação (3.68).

$$p(g_0, g_1) = p(g_0, g_1 - g_0) \cong p(g_0)p(g_1 - g_0)$$

onde : (3.68)

g_0 e g_1 : valores dos níveis de cinza dos *pixels* 0 e 1.

A aproximação de $p(g_0, g_1) \approx p(g_0)p(g_1 - g_0)$ é justificável pelos argumentos levantados por Unser (UNSER, 1986) e mencionados na seção 3.3.3 deste capítulo. Em seu trabalho, Unser ressaltou que a aproximação da MCNC pelos histogramas da soma e da diferença é bastante satisfatória. Ele observou também que o histograma da diferença é mais potente que o da soma e consegue, sozinho, um desempenho próximo daquele alcançado pelo conjunto dos dois histogramas. A análise do termo $p(g_0)p(g_1 - g_0)$ conduz à verificação de que o fator $p(g_0)$ relaciona-se ao valor médio do nível de cinza, enquanto que o outro fator, $p(g_1 - g_0)$ representa efetivamente a correlação entre os níveis dos *pixels*.

Ojala e outros (OJALA *et alii*, 1996) utilizaram esse argumento, de que o histograma da diferença aproxima a correlação entre os *pixels*, para propor estatísticas multidimensionais que podem melhorar o desempenho. Assim, eles propuseram a distribuição conjunta das diferenças de níveis de cinza de *pixels* adjacentes pertencentes a uma região 3 x 3. A Tabela 3.3 ilustra a proposta. Ressalta-se que a distribuição p_8 é muito semelhante aos PBL. A diferença é que enquanto p_8 é calculada sobre o valor das diferenças entre níveis de cinza, os PBL consideram apenas os sinais da diferença, e não seu tamanho.

Tabela 3.3 – Propostas de distribuições multidimensionais de diferença de níveis de cinza entre *pixels* vizinhos.

Região 3 x 3			Distribuições propostas	
g_4	g_2	g_3	$p_2(g_1 - g_0, g_2 - g_0)$	
g_5	g_0	g_1	$p_4(g_1 - g_0, g_2 - g_0, g_3 - g_0, g_4 - g_0)$	
g_6	g_7	g_8	$p_8(g_1 - g_0, g_2 - g_0, \dots, g_8 - g_0)$	

A conclusão que as experiências de Ojala e outros (OJALA *et alii*, 1996) apresentaram é que os PBL são efetivamente poderosos para a análise de texturas. Além de serem computacionalmente eficientes, os PBL apresentam invariância a transformações monotônicas nos níveis de cinza, o que permite sua aplicação sem a necessidade de processamento prévio visando a equalização da escala dinâmica nem a quantização do espaço de características.

3.6.5 Padrões binários locais invariantes à rotação

Nem as unidades de textura, UT, nem os padrões binários locais, PBL, são invariantes à rotação. Isso porque a ordem dos elementos das UT e dos PBL deve ser arbitrada, o que define a posição dos padrões nos histogramas. Se um padrão sofre uma rotação, sua posição no histograma também se altera.

Para ultrapassar essa limitação, Pietikäinen e outros propuseram um aprimoramento aos PBL para atingir a invariância rotacional (PIETIKÄINEN *et alii*, 2000). O conceito que suporta a técnica é agrupar como um único padrão binário local invariante à rotação, PBLIR, todas as versões rotacionadas do padrão original. Por exemplo, existem oito PBL que apresentam apenas um bit 1 em sua constituição. São eles: 00000001, 00000010, 00000100, 00001000, 00010000, 00100000, 01000000 e 10000000. Seus valores são, respectivamente, 1, 2, 4, 8, 16, 32, 64 e 128. Todos esses oito PBL serão mapeados em um mesmo PBLIR. Esse mesmo raciocínio pode ser aplicado aos demais padrões para obter a tabela dos 36 PBLIR possíveis. A Tabela 3.4 apresenta o universo dos LBPIR.

Tabela 3.4 – A lista dos padrões binários locais invariantes à rotação.

Índice do LBPIR	Padrão do LBPIR	Exemplos de LBP mapeados			
0	00000000	00000000			
1	00000001	00000010	00000100	00010000	10000000
2	00000011	00011000	01100000	11000000	10000001
3	00000101	00010100	01010000	01000001	10000010
4	00001001	00100100	10010000	00100001	01000010
5	00010001	00100010	01000100	10001000	
6	00000111	00111000	11100000	11000001	10000011
7	00001011	00101100	10110000	01100001	11000010
8	00010011	00100110	01001100	10011000	00110001
9	00001101	00110100	11010000	10100001	01000011
10	00010101	10101000	01010001	10100010	01000101
11	00100101	01001010	10010100	00101001	01010010
12	00011001	00110010	01100100	11001000	10010001
13	00001111	00011110	00111100	11000011	10000111
14	00010111	01011100	10111000	11000101	10001011
15	00011011	01101100	10110001	11000110	10001101
16	00011101	01110100	11101000	11010001	10100011
17	00100111	10011100	00111001	11100100	11001001
18	00101011	01010110	10101100	01011001	10010101
19	00101101	10110100	01101001	11010010	10100101
20	00110011	01100110	11001100	10011001	
21	00110101	01101010	11010100	10101001	01010011
22	01010101	10101010			
23	00011111	01111100	11111000	11100011	10001111
24	00101111	01011110	10111100	11110010	11001011
25	00110111	11011100	10111001	11100110	11001101
26	00111011	11101100	11011001	10110011	01100111
27	00111101	01111010	11110100	11010011	10100111
28	01010111	10101110	01011101	10111010	01110101
29	01011011	10110110	10110101	01101011	11010110
30	00111111	01111110	11111100	11111001	11100111
31	01011111	10111110	01111101	11111010	11010111
32	01101111	11011110	10111101	01111011	11110110
33	01110111	11101110	11011101	10111011	
34	01111111	11111110	11101111	11011111	10111111
35	11111111	11111111			

Pode-se observar na Tabela 3.4, que existem apenas 36 padrões PBLIR distintos que são nomeados segundo o índice. Mais uma vez, a atribuição dessa ordem é arbitrária e qualquer outra pode ser escolhida. Como não existe fórmula para o cálculo do índice, pode-se utilizar uma tabela de busca. Esta tabela conterá o menor dos PBL correspondente ao PBLIR. Para computar o índice, primeiro calcula-se o PBL da forma usual. Depois, submete o PBL calculado a oito rotações sucessivas e considera-se apenas o menor dos valores calculados. Esse valor deverá ser o padrão PBLIR constante da tabela. Finalmente, procura-

se na tabela pelo valor calculado e retorna-se o índice desse elemento. O histograma de PBLIR conterá, conseqüentemente, 36 *bins* e deverá ser invariante à rotação.

A Figura 3.36 apresenta duas texturas, palha e lã, em versões normais e rotacionadas de 30° (palha) e 60° (lã). Mostra também os histogramas PBLIR de cada uma delas. Uma inspeção visual permite concluir que os histogramas das duas texturas palha são virtualmente iguais. O mesmo ocorre para os histogramas das duas texturas lã. Entretanto, os histogramas de palha e lã são diferentes, o que permite diferenciar uma da outra.

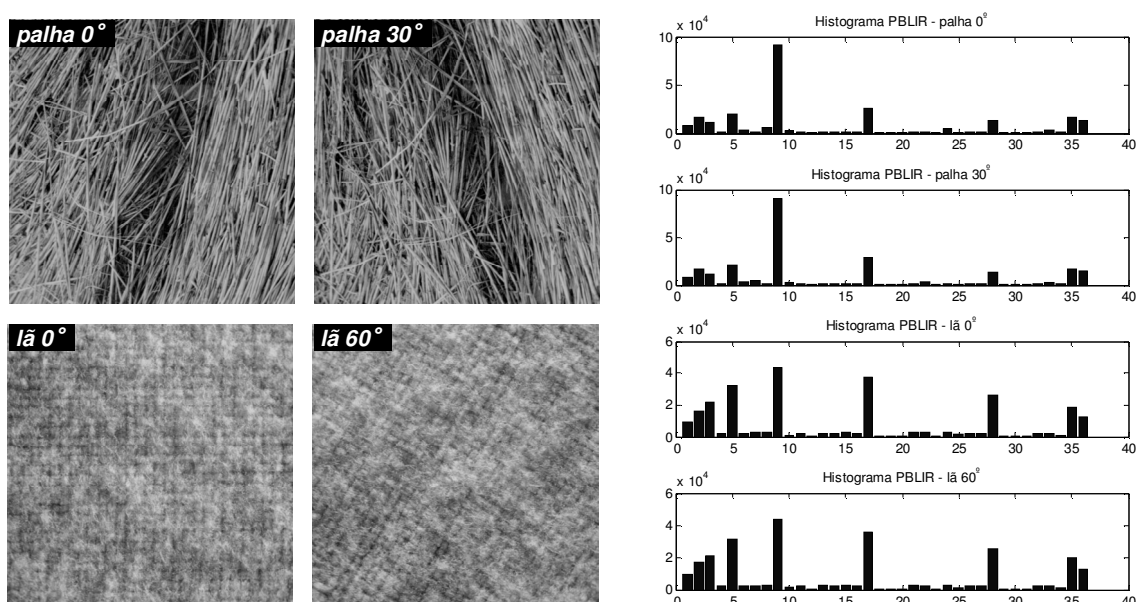


Figura 3.36 – Texturas palha e lã, normais e rotacionadas e seus correspondentes histogramas PBLIR.

A partir deste ponto, neste capítulo, a versão original, não rotacionada de uma textura será referenciada como textura 0°.

Pietikäinen e outros (PIETIKÄINEN *et alii*, 2000) testaram o desempenho do PBLIR e de outras características em experimentos de classificação que tinham como objeto 15 texturas do álbum de Brodatz e suas versões rotacionadas de 30°, 60°, 90°, 120° e 150°. Segundo eles, o PBLIR não apresentou resultados surpreendentes quando utilizado sozinho. Algumas razões para isso foram apresentadas. A primeira é porque as texturas do álbum de Brodatz parecem necessitar de informação de contraste e de tonalidade para uma

melhor discriminação. A outra porque a região 3×3 do PBLIR considera efetivamente apenas as rotações múltiplas de 45° . As rotações que apresentam ângulos no entremeio desses valores não são mapeadas exatamente, gerando erros de classificação. Esse efeito é mais acentuado em texturas que apresentam um alto grau de direcionalidade.

A Figura 3.37 apresenta versões rotacionadas de 0° , 30° , 60° e 90° da textura casca de árvore. Essa é uma textura pouco direcional. Os histogramas PBLIR normalizados são bastante semelhantes. A Tabela 3.5 mostra as distâncias nominais entre as texturas, tomadas duas a duas. Pode-se perceber a pequena distância entre elas.

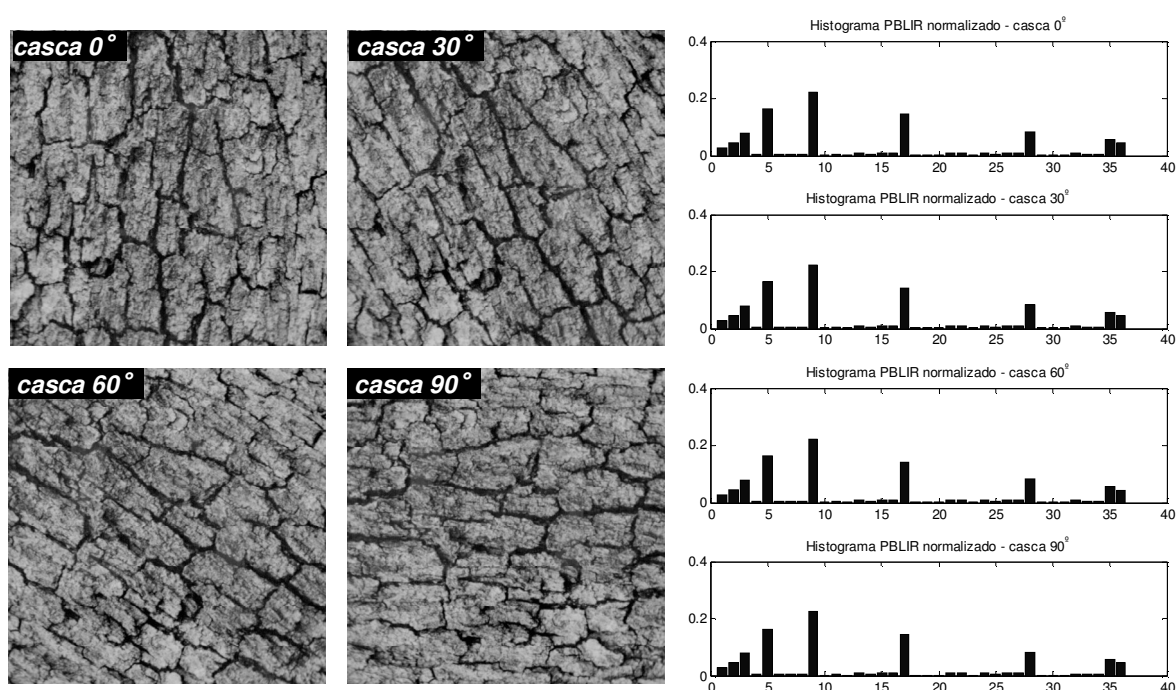


Figura 3.37 – Texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIR normalizados.

Tabela 3.5 – Distâncias nominais entre histogramas PBLIR das texturas casca de árvore rotacionadas de 0° , 30° , 60° e 90° .

<i>Distâncias nominais normalizadas</i>	<i>0°</i>	<i>30°</i>	<i>60°</i>	<i>90°</i>
<i>0°</i>		0,0068	0,0084	0,0067
<i>30°</i>	0,0068		0,0055	0,0097
<i>60°</i>	0,0084	0,0055		0,0123
<i>90°</i>	0,0067	0,0097	0,0123	

A Figura 3.38 apresenta versões rotacionadas de 0° , 30° , 60° e 90° da textura madeira. Essa é uma textura muito direcional, como se pode perceber pelas diversas raias na mesma direção. Nesse caso, os histogramas PBLIR apresentam algumas divergências. A Tabela 3.6 mostra as distâncias nominais entre as texturas, tomadas duas a duas. Pode-se perceber que a distância entre algumas delas apresenta-se maior. As distâncias são pequenas para os pares de textura cujo ângulo de rotação relativo entre elas é de 90° . As distâncias são maiores para os pares de textura cujo ângulo de rotação relativo entre elas é diferente de 90° .

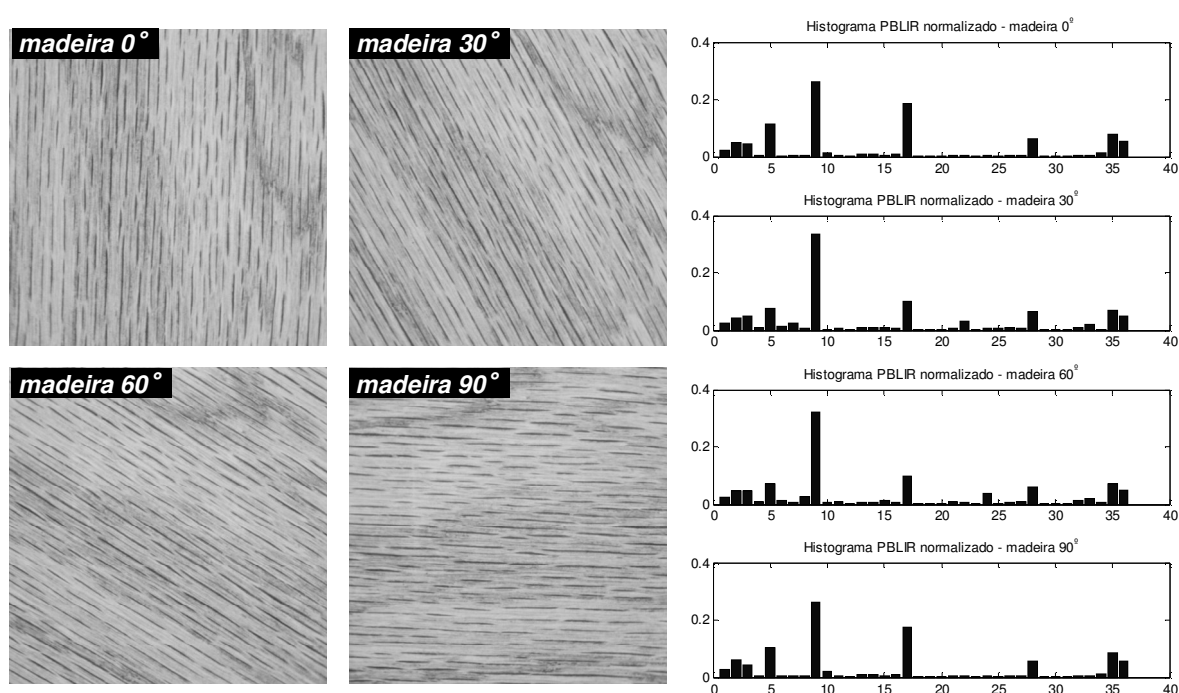


Figura 3.38 - Texturas madeira rotacionadas de 0° , 30° , 60° e 90° e seus correspondentes histogramas PBLIR normalizados.

Tabela 3.6 - Distâncias nominais entre histogramas PBLIR das texturas madeira rotacionadas de 0° , 30° , 60° e 90° .

Distâncias nominais normalizadas	0°	30°	60°	90°
0°		0,1691	0,1741	0,0395
30°	0,1691		0,0808	0,1781
60°	0,1741	0,0808		0,1790
90°	0,0395	0,1781	0,1790	

Apesar do baixo desempenho obtido por Pietikäinen (PIETIKÄINEN *et alii*, 2000) e outros quando da utilização exclusiva do PBLIR nos experimentos de classificação, quando eles utilizaram os PBLIR combinados com outra característica, principalmente com aquelas que levam em conta a escala ou o nível de cinza, os resultados destacaram-se das demais técnicas aplicadas, provando que os PBLIR são de grande aplicabilidade.

3.6.6 Padrões binários locais uniformes

Ojala e outros prosseguiram no aperfeiçoamento da técnica proposta dos padrões binários locais invariantes à rotação. Conforme mencionado na seção 3.6.5 deste capítulo, o histograma dos PBLIR comporta-se de modo indesejado quando se comparam duas imagens de texturas semelhantes que apresentam um ângulo de rotação entre elas de magnitude diferente de 90° . Com a finalidade de superar essa deficiência, eles propuseram uma reestruturação para a vizinhança. Ao invés de considerar uma vizinhança quadrada, foi sugerida uma vizinhança circular de raio arbitrado, medido a partir do *pixel* central. Agora, essa vizinhança circular passa a poder apresentar raios variados, dependendo da resolução angular desejada (OJALA *et alii*, 2001).

Para esclarecer os detalhes da técnica de Ojala e outros (OJALA *et alii*, 2001), deve-se supor que se deseja calcular a vizinhança de um *pixel* central localizado nas coordenadas $(0,0)$. Essa vizinhança deve ser composta por P *pixels* distantes de um raio R do *pixel* central. Se o *pixel* p_0 está localizado à direita do *pixel* central, suas coordenadas são $(0,R)$. Assim, as coordenadas dos P *pixels* da vizinhança são obtidas pela Equação (3.69).

$$p_i = \left(-R \sin\left(\frac{2\pi i}{P}\right), R \cos\left(\frac{2\pi i}{P}\right) \right)$$

onde:

(3.69)

i : i -ésimo pixel de vizinhança;

P : número de *pixels* da vizinhança; e

R : raio da vizinhança.

É provável que as coordenadas de alguma das P localidades não coincidam com o centro de um *pixel* da imagem. Nesse caso o valor do nível de cinza correspondente deve ser calculado por interpolação.

As etapas seguintes são semelhantes àquelas propostas para os PBLIR. Primeiramente calculam-se os sinais das diferenças entre os níveis de cinza da vizinhança e do *pixel* central. Esses sinais são traduzidos para os valores 0 e 1 conforme a Equação (3.66). O valor associado à vizinhança passa a ser designado por PBLIRU2_{P,R} em que P informa o número de vizinhos e R informa o raio (medido em *pixels*) da vizinhança. A abreviação IR informa que o padrão é invariante à rotação e o sufixo U2 significa que se trata de padrões uniformes que apresentam duas transições.

O conceito de padrão uniforme está relacionado à distribuição dos sinais das diferenças pela vizinhança. Ojala e outros (OJALA *et alii*, 2001) investigaram a distribuição dos padrões em diversas imagens e chegaram à conclusão que alguns deles predominaram sobre os demais. Esses padrões predominantes têm uma característica comum: quando se consideram as seqüências dos sinais das diferenças traduzidas pelas seqüências de 0 e 1 que os identificam, aqueles padrões que apresentam até duas transições entre corridas de 0 e corridas de 1 correspondem a até 90% dos padrões totais das imagens. Esses padrões foram chamados de uniformes de medida 2. Sua importância reside no fato deles traduzirem padrões recorrentes de textura, como bordas, cantos, pontos e planos. Os padrões que apresentam mais transições traduzem micro-texturas menos comuns e, por isso, são rotulados por um identificador comum.

A Figura 3.39 apresenta exemplos de padrões uniformes e não uniformes para o caso $P = 8$ e $R = 1$. O número de transições U_i é calculado percorrendo-se a vizinhança no sentido horário e contando o número de transições 0/1 e 1/0 que ela apresenta.

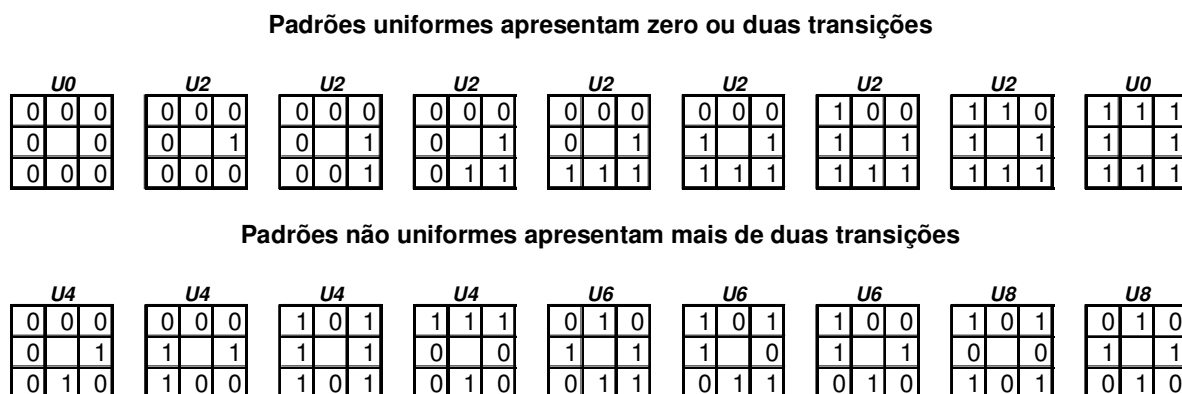


Figura 3.39 – Exemplos de padrões binários locais uniformes e não uniformes e seus respectivos números de transições.

O valor do padrão é calculado de forma bastante simples. Cada padrão uniforme, U_0 ou U_2 , recebe como valor o número de bits 1 que o compõem. Todos os demais padrões não uniformes recebem o mesmo valor igual a $P + 1$. Essa regra é geral e vale para quaisquer valores de P e R . Para o caso $P = 8$ e $R = 1$, os padrões uniformes, correspondentes a linha superior da Figura 3.39, recebem os valores 0, 1, 2, até 8, respectivamente. Todos os demais padrões não uniformes recebem o valor 9.

A expressão geral para o cálculo do rótulo dos padrões é apresentada na Equação (3.70). Apesar dela ser complexa, o cálculo do valor do padrão é bastante simples, como explicado no parágrafo anterior.

$$PBL_{P,R}^{iru2} = \begin{cases} \sum_{i=0}^{P-1} s(g_i - g_c) & \text{se } U(PBL_{P,R}) \leq 2 \\ P+1 & \text{caso contrário} \end{cases}$$

onde :

(3.70)

$$U(PBL_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{i=1}^{P-1} |s(g_i - g_c) - s(g_{i-1} - g_c)|$$

g_i : nível de cinza do i -ésimo *pixel* da vizinhança; e

$s(g_i - g_c)$: diferença de nível de cinza entre o i -ésimo *pixel* e o *pixel* central.

Uma das características mais marcantes da técnica de Ojala e outros (OJALA *et alii*, 2001) é a redução de *bins* do histograma para $P + 2$. Como consequência, obtém-se uma maior concentração dos aspectos texturais mais marcantes. Essa concentração é ocasionada pela redução da dimensão do espaço de características.

Os resultados dos experimentos reportados por Ojala e outros (OJALA *et alii*, 2001) mostraram que foi melhorado o desempenho da classificação com a utilização unicamente do padrão PBLIRU2, principalmente para os casos $P = 16$ e $R = 2$ e $P = 24$ e $R = 3$. A combinação de PBLIRU2 com uma característica que mede a escala dinâmica dos níveis de cinza não acarretou melhorias significativas.

A Figura 3.40 apresenta versões rotacionadas de 0° , 30° , 60° e 90° da textura casca de árvore e seus histogramas de padrões uniformes. A Tabela 3.7 mostra as distâncias nominais entre as texturas, tomadas duas a duas. Comparando-se esses dados com a Figura 3.38 e a Tabela 3.5, pode-se perceber um desempenho semelhante para as técnicas PBLIR e PBLIRU2.

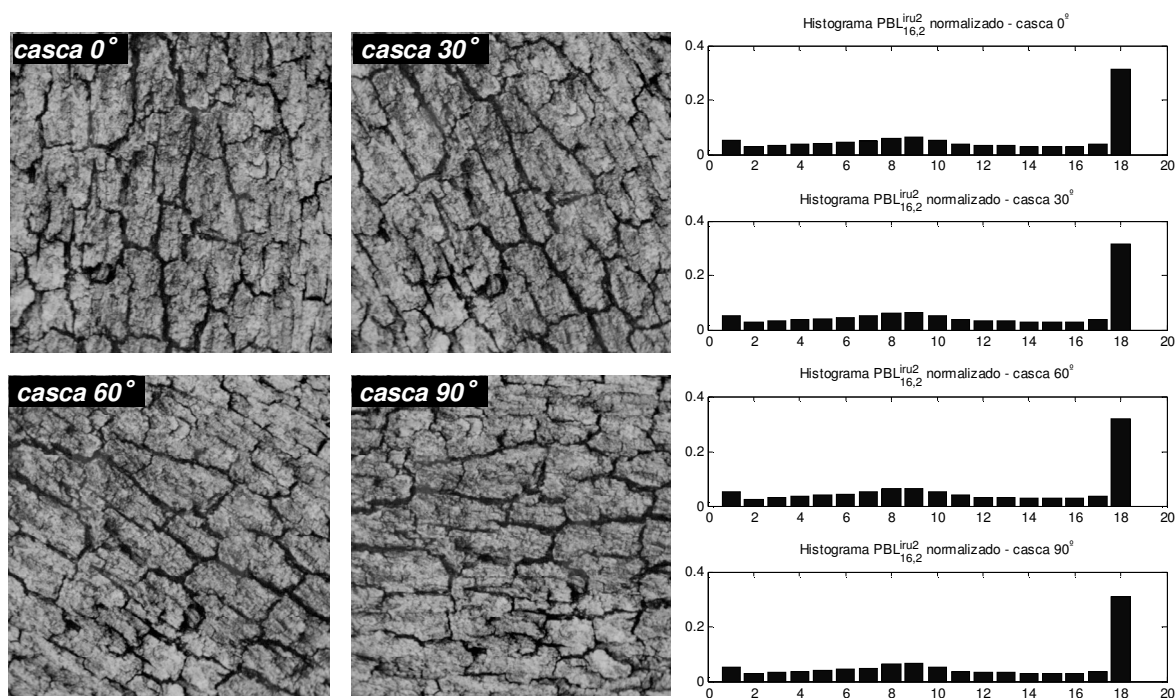


Figura 3.40 - Texturas casca de árvore rotacionadas de 0°, 30°, 60° e 90° e seus correspondentes histogramas PBLIRU2 normalizados.

Tabela 3.7 – Distâncias nominais entre histogramas PBLIRU2 das texturas casca de árvore rotacionadas de 0°, 30°, 60° e 90°.

<i>Distâncias nominais normalizadas</i>	<i>0°</i>	<i>30°</i>	<i>60°</i>	<i>90°</i>
<i>0°</i>		0,0059	0,0087	0,0073
<i>30°</i>	0,0059		0,0047	0,0102
<i>60°</i>	0,0087	0,0047		0,0114
<i>90°</i>	0,0073	0,0102	0,0114	

A Figura 3.41 apresenta versões rotacionadas de 0°, 30°, 60° e 90° da textura madeira. A Tabela 3.8 mostra as distâncias nominais entre as texturas, tomadas duas a duas. Essa é uma textura muito direcional e o desempenho da técnica PBLIR foi baixo, conforme pode ser observado na Figura 3.38 e na Tabela 3.6. No entanto, a utilização da técnica PBLIRU2 teve melhor desempenho, conforme pode ser constatado pela análise da Tabela 3.8. Apesar de um leve desvio do histograma da textura rotacionada de 60°, as distâncias normalizadas entre os histogramas agora são bem menores.

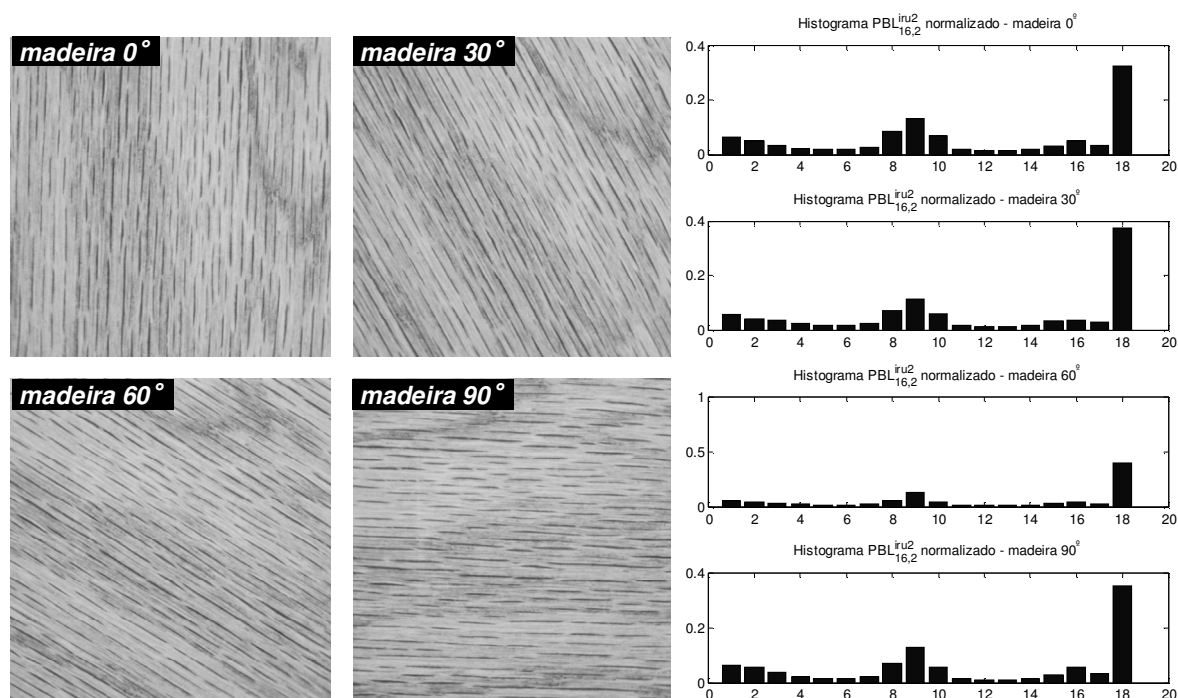


Figura 3.41 - Texturas madeira rotacionadas de 0°, 30°, 60° e 90° e seus correspondentes histogramas PBLIRU2 normalizados.

Tabela 3.8 - Distâncias nominais entre histogramas PBLIRU2 das texturas madeira rotacionadas de 0°, 30°, 60° e 90°.

Distâncias nominais normalizadas	0°	30°	60°	90°
0°		0,0596	0,0828	0,0436
30°	0,0596		0,0424	0,0560
60°	0,0828	0,0424		0,0687
90°	0,0436	0,0560	0,0687	

3.7 Tópicos sobre o Padrão MPEG-7

As seções anteriores deste capítulo apresentaram algumas das principais técnicas de determinação de semelhança entre imagens baseadas em características texturais. Apesar da cor ser um importante fator de distinção e caracterização de imagens, nenhuma das técnicas citadas anteriormente leva em consideração sua utilização como característica de descrição e busca. Isso porque o arquivamento e o processamento de imagens coloridas implicam em requisitos computacionais e de espaço de armazenamento que os equipamentos comercialmente disponíveis não possuíam até alguns anos atrás. Cada

imagem colorida pode requerer até três vezes mais espaço em disco para ser arquivada e uma quantidade de memória semelhante para ser processada quando comparada com uma imagem em níveis de cinza de resolução equivalente (MEHTRE *et alii*, 1995). Com a comercialização de máquinas mais potentes a preços mais atrativos, as pesquisas que incluem a cor no conjunto de características de descrição e classificação de imagens tornaram-se factíveis e permitiram o surgimento de técnicas mais potentes na área de classificação, segmentação e busca de imagens semelhantes em bancos de dados.

Nesse contexto, foi desenvolvido pelo grupo MPEG da ISO/IEC o padrão MPEG-7 (ISO-15938), que possibilita a descrição de conteúdo multimídia em vários níveis de granularidade e em diversas áreas, visando, sobretudo, apoiar o desenvolvimento de aplicativos de busca e recuperação, por semelhança, de informações de som, imagem e vídeo. O padrão MPEG-7 é bastante amplo e se ocupa em definir regras que permitam a interoperabilidade entre os sistemas e aplicativos diversos que produzam, gerenciem, distribuam e consumam descrições de conteúdos áudio-visuais (CHANG *et alii*, 2001). Para permitir a interoperabilidade desejada, o padrão deve ser abrangente e rigoroso. Entretanto, deve apresentar simultaneamente um grau de flexibilidade razoável com a finalidade de permitir que novas tecnologias sejam integradas de forma natural. Para isso, o padrão define os requisitos que alguns instrumentos descritores devem atender, contudo, não especifica rigidamente os métodos ou técnicas de extração das características subjacentes a esses descritores (SIKORA, 2001).

O objetivo desta seção é apresentar os aspectos principais de alguns descritores definidos pelo padrão MPEG-7 que podem ser utilizados para a determinação de semelhança entre imagens, tais como os descritores visuais gerais de cor e textura que ele propõe. Será também apresentado o corpo de procedimentos de teste que integram o padrão e que têm como objetivo a aferição de desempenho dos aplicativos que pretendam uma conformação com suas diretrizes.

3.7.1 Os procedimentos de testes

Quando se pretende integrar naturalmente diferentes tecnologias, ou diferentes soluções para um mesmo problema, é importante que se definam padrões que possam comparar o desempenho dessas diferentes soluções e produzir resultados lógicos e coerentes. O paradigma adotado pelo padrão MPEG-7 foi o da busca por exemplo (*query by example*). Nesse paradigma, uma imagem de busca é processada e tem suas características extraídas. Essas características são comparadas com as características de um conjunto de imagens que compõem o banco de dados por meio de uma medida numérica objetiva que quantifica o grau de similaridade entre elas.

Para determinar coerentemente o resultado dos testes de desempenho é necessário que se defina o conjunto de imagens de busca, o conjunto de imagens do banco de dados e os conjuntos-verdade correspondentes (*ground-truth data*). Estes conjuntos-verdade são os conjuntos de imagens pertencentes ao banco de dados que sejam similares a cada imagem do conjunto de imagens de busca. Assim, é possível determinar o grau de acerto e, por conseguinte, o desempenho de uma determinada técnica de busca (MANJUNATH *et alii*, 2001). Esse procedimento pode ser bastante penoso e consumir um tempo razoável. Em certas situações, para facilitar o trabalho, como, por exemplo, na análise de descritores de texturas, é possível subdividir uma imagem de textura maior em imagens menores. Aplicam-se os testes ao conjunto de imagens menores e avaliam-se os resultados. Essa abordagem é muito utilizada nas pesquisas relatadas na literatura e apresenta a vantagem de que o conjunto-verdade é diretamente determinado – as imagens menores devem casar com a imagem mãe da qual foram extraídas. Isso evita que procedimentos subjetivos de medida de semelhança tenham que ser adotados. Esses procedimentos, ou são excessivamente custosos em termos de tempo, ou não apresentam boa confiabilidade, se não forem tomados os cuidados devidos.

Definido o paradigma, resta a proposição das métricas para avaliar o desempenho. Uma medida bastante utilizada é a taxa de recuperação, RR (*retrieval rate*), ou taxa de acertos, que mede a taxa de acertos obtidos nas $\alpha \times NV(q)$ primeiras imagens recuperadas. O parâmetro α deve ser maior ou igual a 1 e $NV(q)$ é o número de imagens do

conjunto-verdade para a imagem buscada. A taxa é calculada sobre o número de imagens que constituem o conjunto-verdade correspondente à imagem de busca. A taxa de acertos assume valores no intervalo $[0,1]$ em que o valor zero corresponde a nenhum acerto e o valor um corresponde à recuperação de todas as imagens semelhantes. Quanto maior é o valor de α , mais tolerante é o valor da taxa porque se permite um número maior de tentativas, o que pode aumentar o número de acertos.

A Equação (3.71) define a taxa de acertos.

$$RR(q) = \frac{NA(\alpha, q)}{NV(q)}$$

onde :

(3.71)

q : imagem de busca;

$NA(\alpha, q)$: número de acertos nas $\alpha \times NV(q)$ primeiras imagens recuperadas; e

$NV(q)$: número de imagens do conjunto verdade da imagem q .

Quando uma série de buscas é realizada, pode-se definir a taxa média de acertos, ARR (*average retrieval rate*), que nada mais é que a média aritmética das taxas de acertos individuais.

A Equação (3.72) define a taxa média de acertos.

$$ARR = \frac{1}{NB} \sum_{i=1}^{NB} RR(q)$$

onde :

(3.72)

NB : número de buscas realizadas.

A taxa de acertos proposta depende tanto do tamanho do conjunto-verdade, quanto da escolha adequada do parâmetro α . Se os tamanhos dos conjuntos-verdade variam muito para as imagens de teste, a taxa de acerto penaliza muito os casos de

conjuntos-verdade pequenos e é muito complacente com os casos de conjuntos-verdade grandes. O mesmo argumento é válido para os valores de α .

Para superar essa deficiência, o padrão MPEG-7 propõe uma medida de desempenho que se baseia nas posições em que as imagens pertencentes ao conjunto-verdade são retornadas dentro da seqüência de imagens recuperadas. Define-se assim a posição média, AVR (*average rank*) da busca, que é a média da posição de retorno calculada para todas as imagens do conjunto-verdade. Uma posição média pequena significa que as imagens do conjunto-verdade foram retornadas nas posições iniciais da seqüência de retorno, o que indica um melhor desempenho. Para contemplar graciosamente os casos em que algumas imagens do conjunto-verdade não sejam retornadas, ou sejam ocorram no final de uma longa seqüência de retorno, deve-se definir um limite K para as posições relevantes, bem como uma penalidade para os acertos que ocorram posteriormente a esse limite.

A Equação (3.73) define a posição de retorno de uma imagem k do conjunto-verdade levando em consideração o limite K e a penalidade mencionada no parágrafo acima.

$$Posição^*(k) = \begin{cases} Posição(k) & \text{se } Posição(k) \leq K \\ 1,25 \times K & \text{se } Posição(k) > K \end{cases}$$

onde :

(3.73)

k : imagem do conjunto - verdade;

$Posição(k)$: posição da imagem k na seqüência de retorno; e

K : número limite de posições relevantes.

A Equação (3.74) define a posição média para uma busca.

$$AVR(q) = \frac{1}{NV(q)} \sum_{k=1}^{NV(q)} Posição^*(k) \quad (3.74)$$

Para minimizar os efeitos das variações de tamanho nos conjuntos-verdade das imagens de busca, o padrão define a posição de retorno modificada, MRR (*modified retrieval rank*) conforme a Equação (3.75). Essa medida retorna valor zero para um retorno perfeito e valores maiores para os casos de menor desempenho.

$$MRR(q) = AVR(q) - 0,5 \times [1 + NV(q)] \quad (3.75)$$

Para normalizar a medida da Equação (3.75), define-se a posição de retorno modificada normalizada, NMRR (*normalized modified retrieval rank*), cuja definição é apresentada na Equação (3.76). A faixa de valores para essa medida é o intervalo $[0,1]$.

$$NMRR(q) = \frac{MRR(q)}{1,25 \times K - 0,5 \times [1 + NV(q)]} \quad (3.76)$$

Finalmente, a Equação (3.77) define a posição média de retorno modificada normalizada, ANMRR (*average normalized modified retrieval rank*). Essa medida nada mais é que a média simples dos valores da posição de retorno modificada normalizada, NMRR, para os vários ensaios de busca executados com as imagens pertencentes ao conjunto de busca.

$$ANMRR = \frac{1}{NB} \sum_{q=1}^{NB} NMRR(q) \quad (3.77)$$

A ANMRR é a medida de desempenho utilizada pelo padrão MPEG-7 em todos os testes de avaliação de desempenho para os descritores de cor. Sua aplicação pode ser estendida para outras características com a finalidade de uniformizar o critério de avaliação de desempenho.

Manjunath e outros observaram que a taxa de acertos na recuperação mensurada pela ANMRR não deve ser o único critério que define o desempenho dos descritores visuais. Embora esse critério seja bastante elucidativo das potencialidades das técnicas propostas, medidas de espaço de armazenamento dos descritores e o tempo de

processamento gasto para seu cálculo, entre outras, devem também ser consideradas, como de fato são (MANJUNATH *et alii*, 2001).

3.7.2 Os descritores de cor

Os descritores de cor propostos pelo padrão MPEG-7 formam um conjunto interessante de medidas que permitem a avaliação da similaridade entre as imagens. Os descritores propostos são: descritor de cor escalável, SCD (*scalable color descriptor*), descritor de estrutura de cores, CSD (*color structure descriptor*), descritor de cores dominantes, DCD (*dominant color descriptor*) e descritor de leiaute de cores, CLD (*color layout descriptor*).

As especificações desses descritores envolvem, tanto sua semântica, quanto determinados detalhes de implementação que levam em conta uma variação da precisão das medidas de características que os compõem. Essa variação de precisão relaciona-se com o formato de expressão dos valores das características extraídas das imagens e armazenadas nos descritores. Um exemplo pode ilustrar melhor esse conceito.

Uma medida que pode ser utilizada como característica de uma região é o valor médio das cores dos *pixels* dessa região. O valor da média pode ser expresso, por exemplo, como um número real efetivamente representado por um conjunto de bits. Esse conjunto define o valor binário mais próximo do valor real da grandeza calculada. Conseqüentemente, essa forma de expressão da média pode envolver um erro que é tanto maior quanto menor for o número de *bits* utilizado para sua representação. O padrão MPEG-7 permite formas de representação escaláveis, ou seja, formas que admitem quantidades diferentes de bits para sua representação. Entretanto, é necessário que a degradação do desempenho ocasionada pela representação que adote um número de *bits* menor, ou pela quantização de intervalos que apresente um número de níveis menor, seja graciosa.

Outra questão importante a ser considerada quando se trabalha com cores é o espaço de cores utilizado para a expressão dos valores de cor dos *pixels*. Uma mesma cor pode ser representada em vários sistemas ou espaços diferentes. Vários são os sistemas de cor disponíveis, dos quais RGB, XYZ, HSV, HSI, YCrCb, Lab, Luv*a, Luv*b e HMMD são os

principais exemplos. A literatura trata extensivamente desses sistemas (GONZALES e WOODS, 2000), (OHM *et alii*, 2000), (MANJUNATH *et alii*, 2001), (PEDRINI e SHWARTZ, 2008). Os sistemas de cor são formas de expressão da grandeza cor. A escolha do sistema de cor não altera a cor em si, mas apenas sua representação. Entretanto, essa escolha pode influir nos cálculos de distância entre cores distintas. Alguns sistemas representam as cores de uma forma mais próxima da maneira com que o ser humano as percebe visualmente. Outros sistemas utilizam uma escala menos comprometida com a percepção visual humana. Assim, o valor da distância entre duas cores distintas calculada em um determinado sistema de cor pode ser consideravelmente diferente do valor calculado em outro sistema. Esse fato tem influência direta em algumas técnicas de extração de características. Por isso, o padrão MPEG-7 determina quais são os sistemas de cor que podem ser utilizados em cada descritor proposto (MANJUNATH *et alii*, 2001).

As definições do padrão MPEG-7 que envolvem a forma de expressão de valores ou a quantização de escalas de medida, ou ainda o sistema de cores utilizado, envolvem muitos detalhes. Esta seção não tratará dessas definições para evitar que essas minúcias obliterem os conceitos principais dos descritores em questão. Por motivo de clareza, e sem perda de generalidade, esta seção enfatizará os aspectos semânticos dos descritores propostos.

3.7.2.1 O descritor escalável de cores - SCD

O descritor escalável de cores, SCD, é utilizado para compor o histograma de cores para a imagem. Esse histograma é uma ferramenta bastante simples que gera resultados aceitáveis. Ele registra o número de ocorrências em uma imagem de cada valor possível de cor. Para uma imagem em três planos de cores, com 256 níveis para cada plano, o número de cores diferentes atinge o total de $256^3 = 2^{24} \approx 4$ milhões. Esse total inviabiliza o armazenamento e o processamento.

Para superar essa inviabilidade, o padrão MPEG-7 determina uma quantização uniforme do espaço de cores do sistema HSV em 256 níveis. O plano H é quantizado em 16 níveis (4 *bits*) e os planos S e V são quantizados em quatro níveis (2 *bits*), cada um. Os

valores de cada *bin* do histograma são codificados como valores inteiros truncados em 11 *bits*. Para economia de espaço, cada valor de 11 *bits* é mapeado em uma representação de 4 *bits* com quantização não uniforme. Posteriormente, o histograma é codificado utilizando a transformada de Haar, que permite um grau de compactação maior. A comparação entre histogramas é efetuada utilizando-se a métrica L1.

A Figura 3.42 apresenta quatro imagens coloridas e seus correspondentes histogramas SCD. As imagens pertencem à coleção VisTex do MediaLab/MIT (VisTex, 2008).

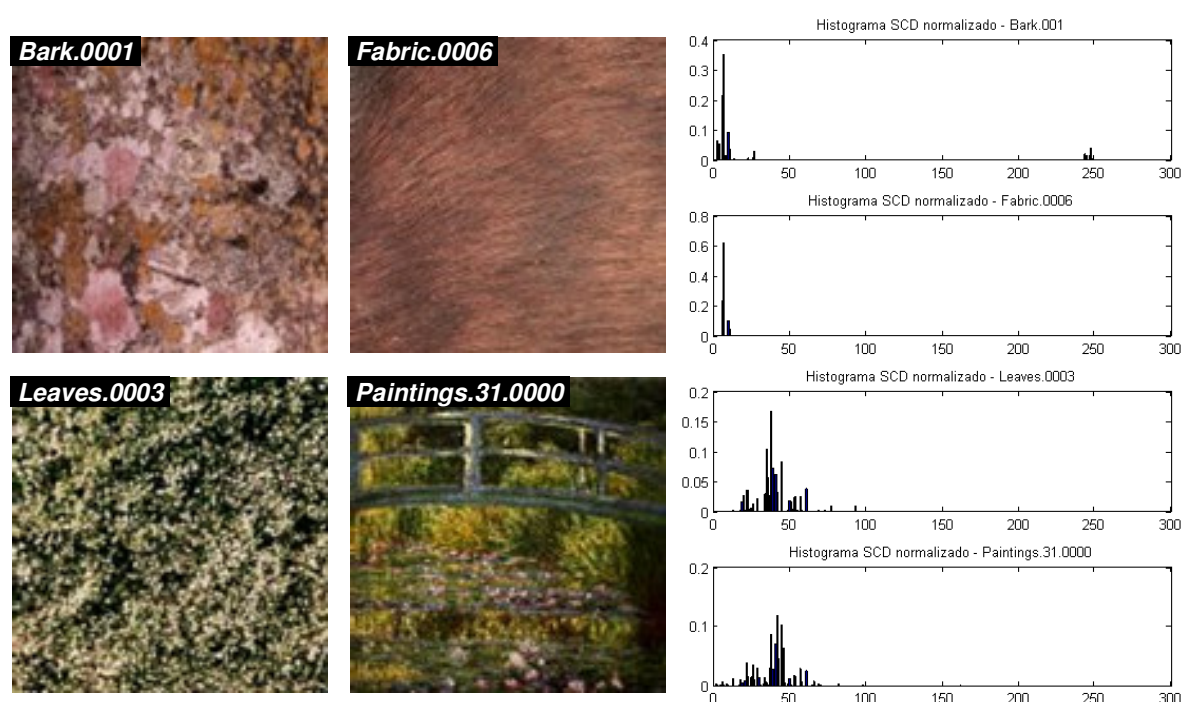


Figura 3.42 – Imagens coloridas e seus histogramas SCD.

A Tabela 3.9 mostra os valores das distâncias nominais normalizadas entre os histogramas das imagens apresentadas na Figura 3.42. Pode-se notar que as imagens Bark e Fabric são mais próximas, assim como as imagens Leaves e Paintings. As distâncias entre os demais arranjos são bastante maiores. Isso é coerente com a percepção visual que se tem dessas imagens.

Tabela 3.9 – Distâncias normalizadas entre os histogramas SCD das imagens da Figura 3.42.

Distâncias nominais normalizadas	Bark.0001	Fabric.0006	Leaves.0003	Paintings.31.0000
Bark.0001		0,3005	0,9795	0,9374
Fabric.0006	0,3005		0,9996	0,9856
Leaves.0003	0,9795	0,9996		0,4211
Paintings.31.0000	0,9374	0,9856	0,4211	

Os histogramas de cores e seus similares, dos quais os histogramas SCD fazem parte, são uma boa ferramenta para a determinação de semelhança entre imagens coloridas. Babu Mehtre, M. Kankanhalli, A. Narasimhalu e Guo Chang Man já haviam proposto anteriormente, duas alternativas simples e viáveis ao histograma de cores: o método da distância e o método da tabela de cores de referência (MEHTRE *et alii*, 1995). Determinadas aplicações lidam com imagens que apresentam grandes regiões de coloração homogênea. Imagens de tecidos, bandeira, logomarcas são exemplos desse tipo de imagem. Os histogramas tridimensionais de cores para essa categoria de imagens tendem a apresentar alguns poucos modos dominantes e uma grande quantidade de *bins* com valores nulos ou pouco significativos. A utilização dos histogramas tridimensionais de cores, para esse tipo de imagens acarreta um desperdício desnecessário de recursos.

O método da distância é o mais simples e de resultados mais modestos. Ele consiste em calcular a média de cada um dos histogramas unidimensionais correspondentes aos três planos de cores do sistema RGB. Outros sistemas similares podem também ser utilizados. A cada imagem associa-se um vetor de características contendo as três médias calculadas. A comparação entre imagens utiliza a medida L_1 ou L_2 .

O método da tabela de cores de referência é um pouco mais complexo, e apresentou resultados bastante significativos. Com a finalidade de reduzir a dimensionalidade do histograma 3D, foi proposta uma tabela de cores de referência. A imagem é percorrida *pixel a pixel* e a cor de cada um deles é substituída pela cor mais próxima da tabela de referência. Nesse ponto, a imagem colorida original transformou-se em uma imagem mapeada. A partir dela produz-se um histograma unidimensional

normalizado em que os *bins* são os índices das cores da tabela de referência. A tabela proposta possui 27 cores apenas, o que reduz drasticamente os requisitos computacionais necessários. A comparação entre duas imagens utiliza uma distância euclidiana modificada. Como na distância euclidiana, calculam-se os quadrados das diferenças dos valores para cada *bin* dos histogramas. Entretanto, se um, ou ambos os valores de um *bin* forem nulos, a diferença para aquele *bin* é arbitrada em 1. Essa é uma forma de penalização que tem como objetivo reduzir a influência do ruído no cálculo da distância, já que as imagens objeto da pesquisa tendem a apresentar poucas cores bastante distintas. A distância corresponde à soma dos quadrados das diferenças, calculadas da forma exposta.

Como mencionado no parágrafo anterior, o método da tabela de cores de referência assemelha-se muito ao histograma SCD. Ambos reduzem a dimensionalidade pela quantização do espaço de cores. A vantagem do histograma SCD está na técnica utilizada para a quantização. Ela utiliza uma quantização uniforme e determinada do espaço HSV, enquanto que a tabela de cores de referência foi obtida por uma divisão empírica do espaço de cores. Além disso, a quantização SCD apresenta menor complexidade computacional, ao passo que o outro método envolve busca em tabelas, o que demanda maior processamento. Finalmente, os descritores de cor escalável permitem uma maior compactação das características quando os requisitos de precisão não forem muito grandes.

3.7.2.2 O descritor de estrutura de cores - CSD

O segundo descritor proposto é o descritor de estrutura de cores, CSD. Ele é um histograma que tem como objetivo retratar a distribuição espacial de cores na imagem, característica que não é captada pelo histograma SCD. É possível que duas imagens apresentem histogramas SCD idênticos, ou muito semelhantes, porém sejam bastante diferentes. Suponha duas imagens compostas por *pixels* de duas cores distintas: verde e vermelho, por exemplo. A primeira imagem é formada por linhas verticais de mesma largura, nas cores verde e vermelha alternadamente. Na segunda, todos os *pixels* de cor verde ocupam o lado esquerdo e todos os de cor vermelha ocupam o lado direito. Essas imagens são apresentadas na Figura 3.43, juntamente com seus histogramas SCD

respectivos. Pode-se observar que os dois histogramas são idênticos. As imagens, no entanto, são bastante diferentes sob o aspecto da percepção visual.

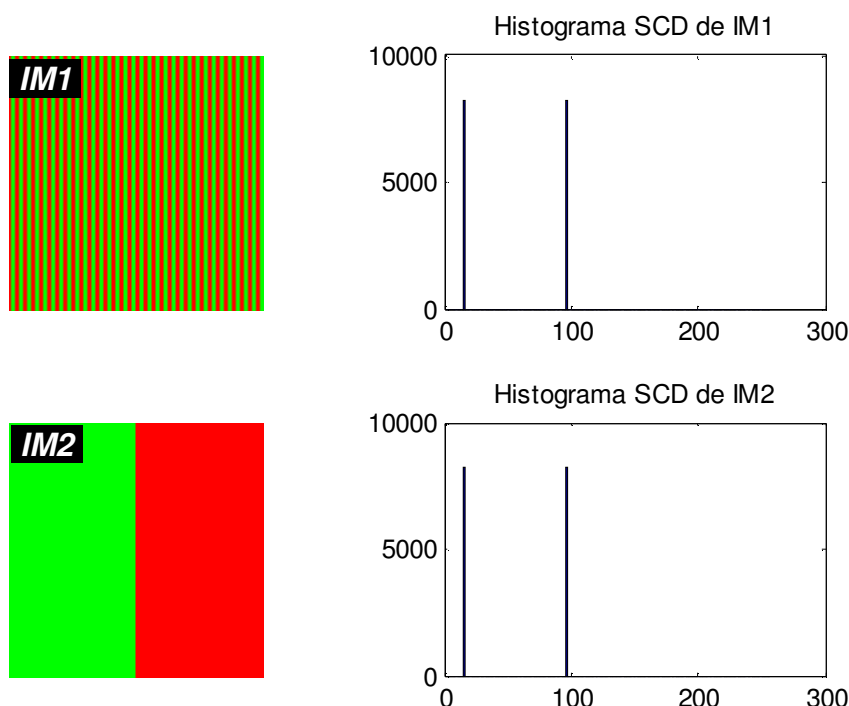


Figura 3.43 – Duas imagens diferentes que possuem o mesmo histograma SCD.

O descritor de estrutura de cores mede, para cada cor presente na imagem, o quanto ela está concentrada, ou o quanto ela está distribuída. Com referência à Figura 3.44, o histograma CSD distingue as duas imagens porque, na imagem IM2, as cores estão concentradas em regiões delimitadas, mas na imagem IM1 elas estão distribuídas por toda a imagem.

Para conseguir estimar essa característica, define-se um elemento de estrutura, um quadrado de 8×8 *pixels* que é deslocado por toda a imagem *pixel a pixel*. Para cada posição que o elemento ocupa, verificam-se quais são as cores presentes na região que ele determina. Para cada uma das cores presentes, computa-se uma ocorrência no *bin* correspondente do histograma CSD. Mesmo que uma cor apareça em um elemento mais de uma vez, apenas uma ocorrência é computada. Ao final, o histograma é normalizado dividindo-se o valor de cada *bin* pelo número de posições que o elemento ocupou.

Cada *bin* do histograma CSD contém o percentual de elementos em que a cor correspondente aparece. Se uma cor aparece concentrada em uma região, o valor correspondente do histograma será menor, já que, certamente, haverá um grande número de elementos nos quais essa cor não estará presente. Se, ao invés, uma cor aparece distribuída por toda a imagem, o valor correspondente do histograma CSD será elevado.

A Figura 3.44 mostra as mesmas imagens da Figura 3.43 e seus correspondentes histogramas CSD. Pode-se observar que as raiais do gráfico estão localizadas nas mesmas posições, já que as cores presentes em ambas as imagens são as mesmas. Entretanto, os valores dos *bins* correspondentes são bem diferentes.

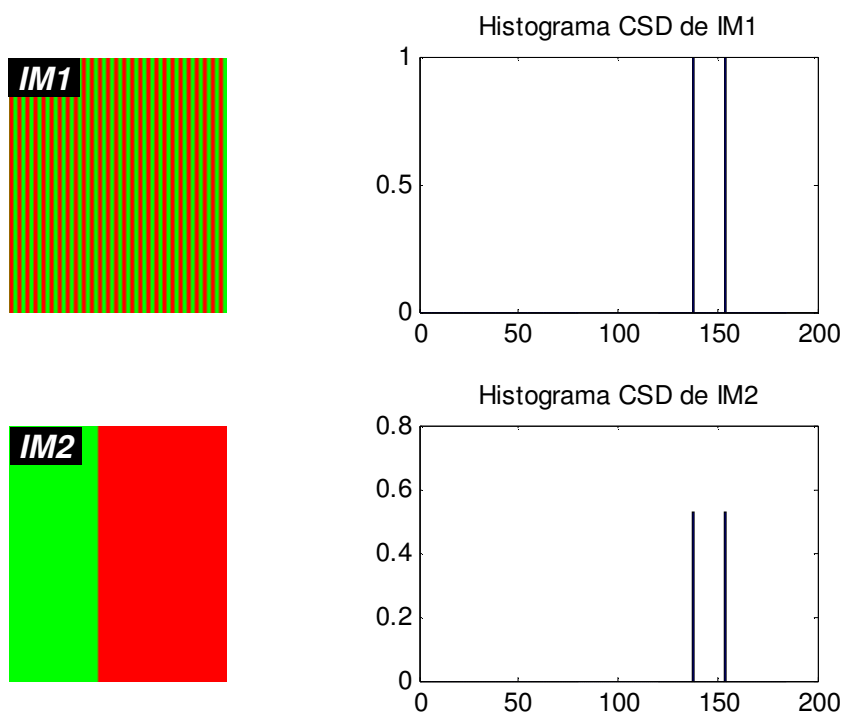


Figura 3.44 – Duas imagens distintas e seus histogramas CSD.

Na imagem IM1, os valores são máximos, iguais à unidade. Para a imagem IM2, os valores correspondem à metade da escala (0,5), indicando que cada cor, nessa imagem, está presente em apenas 50% dos elementos. Se as cores ocorrem na mesma quantidade, pode-se dizer que quanto maior o valor correspondente no histograma CSD, mais essa cor

está distribuída na imagem. Em sentido contrário, quanto menor o valor no histograma, mais concentrada a cor se apresenta.

Para o cálculo do CSD, o padrão MPEG-7 estabelece que a imagem deve estar representada no sistema HMMD. Esse sistema apresenta quatro planos de representação: *hue*, *max*, *min* e *diff*. Apenas três planos são suficientes para a representação: *hue*, *sum*, *diff*. O plano *hue* tem a mesma definição dos sistemas HSV e HSI e carrega a informação da tonalidade. O plano *max* contém o valor da maior componente RGB. O plano *min* contém o valor da menor componente RGB. O plano *diff* contém a diferença entre os valores da maior e da menor componente RGB. Quanto maior o valor *diff*, mais pura, ou saturada, é a cor. O plano *sum* contém a média entre os valores máximo e mínimo. Ele tem relação estreita com a intensidade.

O espaço de cores HMMD é quantizado não uniformemente para a diminuição da dimensão do histograma. Essa quantização divide uniformemente o plano *diff* em cinco níveis. Para cada um dos cinco subespaços, quantiza-se uniformemente os planos *hue* e *sum*. O número de níveis de cada plano varia para cada subespaço, com a finalidade de conseguir um conjunto de cores mais uniformemente distribuídas. Além disso, pode-se obter histogramas de 32, 64, 120 e 184 *bins*, ressaltando o compromisso entre precisão e espaço de armazenamento.

Este trabalho utiliza os histogramas de 184 *bins* e seu esquema de quantização é mostrado na Tabela 3.10.

Tabela 3.10 – Esquema de quantização do espaço de cores HMMD para o histograma CSD de 184 *bins*.

Componente	Subespaço	N.º de níveis de quantização
Hue	0	1
	1	8
	2	12
	3	12
	4	24
Sum	0	8
	1	4
	2	4
	3	4
	4	2

Para que se possam comparar imagens de tamanhos diferentes, o padrão MPEG-7 determina que imagens maiores sejam submetidas a uma sub-amostragem antes que o histograma CSD seja calculado. Imagens de dimensão menor que 256 x 256 *pixels* não sofrem sub-amostragem. Imagens maiores são sub-amostradas por um fator determinado pela expressão apresentada na Equação (3.78).

$$p = \max\{0, \text{round}(0.5 \log_2(WH) - 8)\}$$

$$K = 2^p$$

(3.78)

onde:

W, H : largura e altura da imagem, respectivamente; e

K : fator de sub - amostragem.

A comparação entre histogramas CSD é realizada utilizando-se a distância L_1 . Não é conveniente normalizar essa distância já que as somas dos valores do histograma não são constantes para todos os histogramas. Uma imagem monótona tende a ter apenas alguns *bins* com valores não nulos, ao contrário de uma imagem muito colorida e com cores dispersas espacialmente, que tende a apresentar muitos *bins* com valores próximos a um.

Para ilustrar a aplicação do histograma CSD, a Figura 3.45 mostra as mesmas imagens da Figura 3.42 e seus respectivos histogramas CSD. A Tabela 3.11 apresenta as distâncias entre elas.

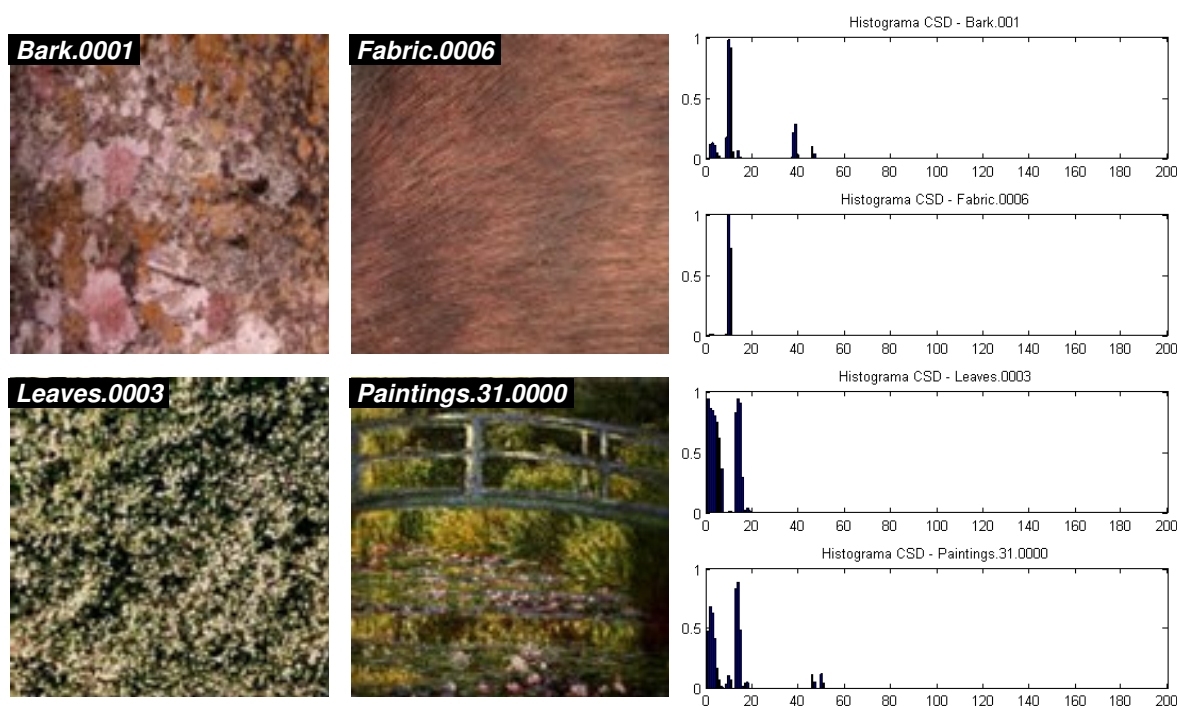


Figura 3.45 – Imagens coloridas e seus histogramas CSD.

Tabela 3.11 – Distâncias entre os histogramas CSD da Figura 3.45

<i>Distâncias nominais</i>	<i>Bark.0001</i>	<i>Fabric.0006</i>	<i>Leaves.0003</i>	<i>Paintings.31.0000</i>
<i>Bark.0001</i>		1,586	10,455	6,906
<i>Fabric.0006</i>	1,586		9,863	6,699
<i>Leaves.0003</i>	10,455	9,863		4,055
<i>Paintings.31.0000</i>	6,906	6,699	4,055	

3.7.2.3 O descritor de cores dominantes - DCD

O objetivo do descritor de cores dominantes, DCD, é listar as cores que melhor representam a imagem. Ele é composto das cores representativas, dos percentuais de ocorrência, das variâncias de cada uma delas e da medida de coerência espacial do conjunto. O número de cores dominantes é limitado a oito, o que torna o DCD um descritor bastante compacto.

A quantidade de cores presentes em uma imagem pode variar em uma faixa bastante extensa. Para reduzir essa quantidade a um número menor ou igual a oito cores, vários algoritmos podem ser utilizados. Entre eles uma modificação do algoritmo de Lloyd generalizado, GLA (*generalized Lloyd algorithm*), (DENG *et alii*, 2001), ou o algoritmo de bloco linear, LBA (*linear block algorithm*), proposto por Yang e outros (YANG *et alii*, 2008). Ambos utilizam o princípio da aglomeração de cores semelhantes.

O algoritmo GLA modificado baseia-se no algoritmo de quantização que minimiza o erro médio quadrático proposto por Stuart Lloyd (LLOYD, 1995), também conhecido como aglomeração de k-médias (*k-means clustering*), descrito por Richard Duda e outros (DUDA *et alii*, 2000). Ele tem como entradas um conjunto inicial pequeno e arbitrário de cores e a imagem a ser processada. O processamento consiste de repetir uma seqüência de operações até que uma condição final seja satisfeita. A saída do algoritmo é o conjunto de cores dominantes. Cada cor membro do conjunto inicial é chamada de centróide porque representa um conjunto de cores próximas. A seqüência de operações iteradas compõe-se de dois procedimentos. O primeiro é a determinação dos *clusters*. Nessa etapa, cada *pixel* da imagem é associado ao centróide mais próximo. Ao final dessa etapa, os *pixels* da imagem foram particionados em *clusters*, cada um em torno de um centróide. O segundo procedimento é o recálculo dos centróides. Nessa etapa, para cada *cluster* obtido na etapa anterior, a média das cores dos *pixels* é calculada. Essa média corresponde à cor que melhor representa as cores dos *pixels* do *cluster* correspondente. Também se calcula nessa etapa a distorção para cada *cluster*. Essa distorção corresponde à variância, simples ou ponderada, das cores do *cluster*. Como o valor do centróide foi modificado, pode ocorrer que um *pixel* que pertencia a um *cluster*, porque estava mais próximo de seu centróide, agora esteja mais próximo ao centróide de outro *cluster*. Assim, o procedimento é repetido, com o reparticionamento dos *pixels* seguido do recálculo dos centróides, até que a variação da soma das distorções seja inferior a um patamar determinado. Nesse ponto o algoritmo finaliza e o conjunto de centróides representa as cores dominantes.

O algoritmo GLA apresenta algumas deficiências. A primeira é que ele gera um esforço computacional muito grande, já que várias iterações são normalmente necessárias

até que a condição de parada seja satisfeita. A segunda é que a escolha do conjunto inicial pode influir no resultado final e no tempo de processamento. A segunda é que as cores obtidas tendem normalmente a se concentrar em torno de regiões mais densas do histograma de cores. Suponha, por exemplo, uma imagem que apresente uma grande região uniforme de cores ligeiramente distintas, porém similares do ponto de vista da percepção visual, e várias regiões menores de cores contrastantes. O algoritmo GLA tende a produzir mais centróides com as cores da região homogênea com a finalidade de diminuir a distorção global e as regiões menores de cores contrastantes podem deixar de ser representadas. No entanto, o desejado é justamente o oposto do obtido, já que, sob o aspecto da percepção visual, a grande região de cores similares poderia ser representada por apenas uma cor, o que permitiria que as demais regiões fossem contempladas por centróides representativos (YANG *et alii*, 2008).

O algoritmo LBA, proposto por Yang e outros (YANG *et alii*, 2008) busca superar as deficiências do GLA expostas no parágrafo anterior. Sua estrutura é composta de quatro etapas. Na primeira, os oito *clusters* iniciais são determinados por partição do espaço RGB em octantes. A segunda etapa determina os oito centróides dos oito *clusters* pela média das cores dos *pixels*. Na terceira etapa ocorre a aglutinação iterativa dos dois *clusters* mais próximos cujos centróides estejam localizados a uma distância menor do que um determinado patamar. Na etapa final, os *clusters* pouco populosos, ou seja, aqueles que possuam uma quantidade percentual de *pixels* menor do que um determinado patamar, são aglutinados com os *clusters* mais próximos.

O algoritmo LBA é mais simples e mais rápido do que o algoritmo GLA. O padrão MPEG-7 determina que o tamanho máximo do DCD é de oito cores dominantes. A aglutinação dos *pixels* pela divisão do espaço RGB em octantes é natural e faz com que o conjunto inicial de cores seja mais significativo. A supressão da etapa iterativa de determinação dos centróides elimina a parte mais demorada do processamento do GLA. Os novos centróides pertencem um a cada octante, diminuindo o efeito da concentração em torno de cores similares muito presentes. As etapas iterativas de aglutinação de *clusters* no LBA não consome muito tempo porque os novos centróides são calculados por ponderação

de percentuais, não havendo necessidade de se repassar os *pixels* da imagem. A Equação (3.79) apresenta as expressões de cálculo dos centróides quando se aglutinam dois *clusters* próximos.

$$c^R = c_1^R \times \left(\frac{p_1}{p_1 + p_2} \right) + c_2^R \times \left(\frac{p_2}{p_1 + p_2} \right)$$

$$c^G = c_1^G \times \left(\frac{p_1}{p_1 + p_2} \right) + c_2^G \times \left(\frac{p_2}{p_1 + p_2} \right)$$

$$c^B = c_1^B \times \left(\frac{p_1}{p_1 + p_2} \right) + c_2^B \times \left(\frac{p_2}{p_1 + p_2} \right)$$

$$p = p_1 + p_2$$

onde :

c^R, c^G, c^B : componentes vermelha, verde e azul do novo centróide;

c_1^R, c_1^G, c_1^B : componentes vermelha, verde e azul do primeiro centróide;

c_2^R, c_2^G, c_2^B : componentes vermelha, verde e azul do segundo centróide;

p : percentual do novo centróide; e

p_1, p_2 : percentuais do primeiro e do segundo centróides.

(3.79)

Apesar do DCD ser tecnicamente um histograma, o cálculo da distância entre duas imagens não pode utilizar as técnicas apresentadas no Capítulo 2. Isso porque os *bins* não são rotulados de forma homogênea. Ao contrário, no DCD, cada *bin* recebe como rótulo o valor da cor dominante. Os valores das cores dominantes não são os mesmos para as imagens, mas sim o resultado da média das cores dos *pixels* de um *cluster*. A comparação entre *bins* de rótulos potencialmente distintos não faz sentido. Para calcular a distância entre os histogramas DCD de duas imagens utiliza-se uma medida de dissimilaridade pseudo-quadrática (*quadratic-like dissimilarity measure*) que calcula a distância de dois vetores, ou histogramas, computando a acumulação ponderada das diferenças dos

componentes tomados dois a dois. A forma geral para medidas desse tipo é mostrada na Equação (3.80).

$$dist^2(A, B) = [\text{hist}(A) - \text{hist}(B)]^T \times W \times [\text{hist}(A) - \text{hist}(B)]$$

onde :

(3.80)

$\text{hist}(\bullet)$: histograma da imagem ; e

W : matriz de pesos w_{ij} .

A Equação (3.80) pode ser manipulada e transformada em uma expressão que permite o cálculo da distância também para histogramas com diferentes números de *bins* (HAFNER *et alii*, 1995). A Equação (3.81) apresenta a expressão modificada equivalente.

$$dist^2(A, B) = \sum_{i=1}^{N_A} p_i^2 + \sum_{j=1}^{N_B} q_j^2 - \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} 2w_{ij} p_i q_j$$

onde :

$$A = \{(c_i, p_i), \quad i = 1, \dots, N_A\};$$

$$B = \{(d_j, p_j), \quad j = 1, \dots, N_B\};$$

c_i, d_j : cores dominantes dos histogramas A e B ;

p_i, q_j : percentuais das cores dominantes dos histogramas A e B ; e

N_A, N_B : números de cores dominantes dos histogramas A e B

e também

$$w_{ij} = \begin{cases} 1 - d_{ij}/d_{\max} & \text{se } d_{ij} \leq T_d \\ 0 & \text{se } d_{ij} > T_d \end{cases}$$

(3.81)

onde :

$$d_{ij} = \|c_i - d_j\|$$

$$d_{\max} = \alpha T_d$$

T_d : maior distância entre duas cores consideradas semelhantes.

O ponto principal da equação (3.81) localiza-se no somatório duplo. Este somatório computa a semelhança entre as duas imagens. Os pesos w_{ij} podem variar de zero a um. Eles valem zero se a distância euclidiana entre um par de cores dos histogramas DCD for maior do que um patamar T_d . Se a distância entre elas for menor do que o patamar T_d , o peso cresce com a distância até atingir o valor um quando as cores forem iguais. O parâmetro α permite um ajuste para o grau de similaridade entre as cores.

Lai-Man Po e Ka-Man Wong observaram que essa medida de distância não é muito eficiente e produz algumas distorções. Como medida alternativa eles propuseram a montagem de um conjunto comum de cores para as duas imagens consideradas. Para isso, concatenam-se os DCD das duas imagens e agrupam-se as cores semelhantes (que distam

menos que T_d), encontrando-se um conjunto de cores comuns a ambos os DCD modificados. Os DCD são expressos no novo conjunto de cores comuns e a distância entre eles agora pode ser calculada aplicando-se a interseção dos histogramas ou a medida $L1$ (PO e WONG, 2004). A desvantagem dessa técnica é a necessidade do cálculo do conjunto de cores comuns a cada comparação, o que torna o processo de busca em grandes bancos de dados mais lento.

O trabalho de Yang e outros aproveita e melhora os conceitos propostos por Po e Wong (PO e WONG, 2004) e corrige as distorções que a medida quadrática acarreta (YANG *et alii*, 2008). A medida proposta por eles é apresentada na Equação (3.82).

$$D^2(A, B) = 1 - SIM(A, B)$$

onde :

(3.82)

$$SIM(A, B) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} 2w_{ij}S_{ij}$$

$$S_{ij} = [1 - |p_i - q_j|] \times \min(p_i, q_j)$$

A medida da distância, ou dissimilaridade, proposta por Yang (YANG *et alii*, 2008) é o complemento a um da medida de similaridade SIM . Essa medida SIM é a soma dos produtos de dois fatores que comparam as cores dominantes duas a duas. O primeiro fator, w_{ij} apresentado na Equação (3.81), leva em conta a distância entre as duas cores. O segundo fator, S_{ij} , leva em conta o percentual de ocorrência das duas cores. A maior diferença entre as medidas das Equações (3.81) e (3.81) é que, na segunda, a soma dos termos quadráticos das porcentagens foi substituída pelo valor 1. Yang e outros observaram que quanto maior é o número de cores dominantes das duas imagens comparadas, menor é o valor da soma dos termos quadráticos e, conseqüentemente, menor é o valor da distância, mesmo que as cores dominantes das imagens sejam bastante diferentes. Esse

resultado contraria a percepção visual. A substituição dessa parcela por 1 reverte esse comportamento, mantendo a distância calculada em valores elevados (YANG *et alii*, 2008).

A Figura 3.46 mostra as quatro imagens da Figura 3.45 e os gráficos de seus histogramas DCD. Para desenhar esses gráficos, foi feita uma quantização uniforme do espaço RGB em que cada plano foi quantizado em 8 níveis. As cores dominantes foram representadas na ordem RGB. Deve-se notar que essa quantização foi feita exclusivamente para efeito de apresentação neste trabalho. A Tabela 3.12 mostra os conteúdos dos histogramas DCD e a Tabela 3.13 mostra as distâncias das imagens, duas a duas, considerando-se a métrica proposta por Yang e outros (YANG *et alii*, 2008), apresentada na Equação (3.82).

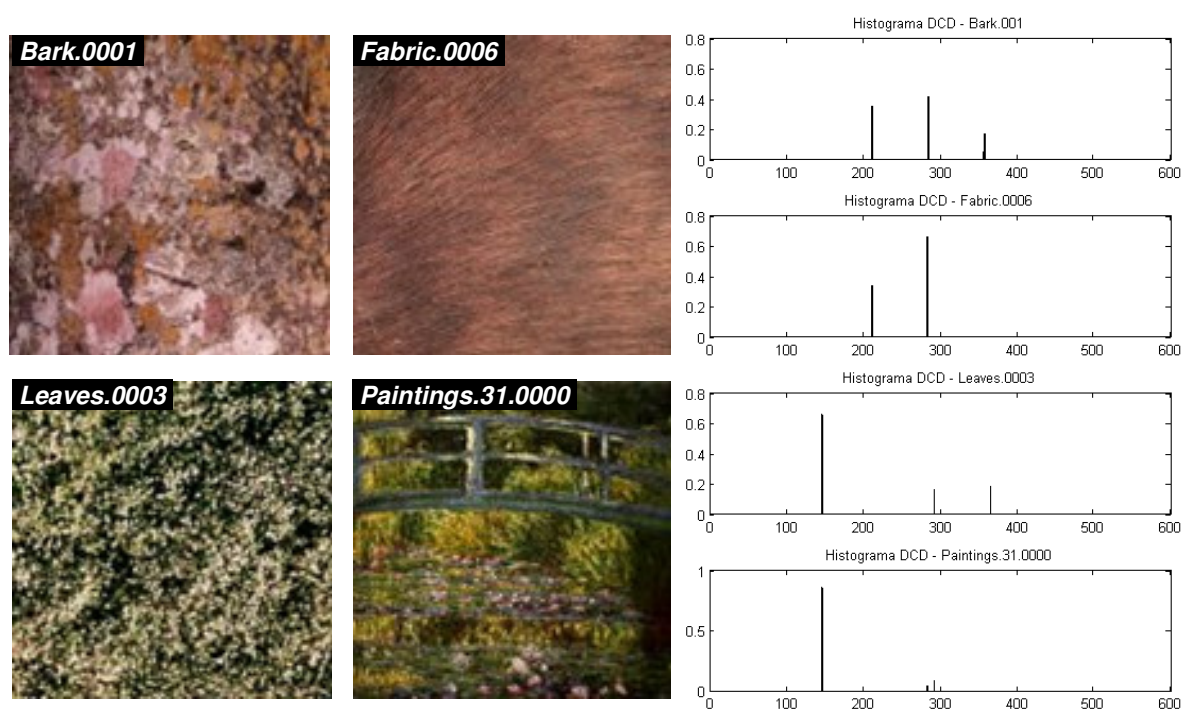


Figura 3.46 – Imagens coloridas e uma representação livre de seus histogramas DCD.

Tabela 3.12 – Os DCD das imagens da Figura 3.46.

Imagem		Valor do componente da cor			Percent.	Variância do componente			Coerência espacial da cor	Coerência espacial da imagem
		RED	GREEN	BLUE		RED	GREEN	BLUE		
Bark	1	0,398	0,258	0,218	0,353	0,006	0,025	0,038	0,708	0,634
	2	0,585	0,410	0,358	0,418	0,003	0,033	0,054	0,622	
	3	0,685	0,520	0,470	0,055	0,001	0,029	0,047	0,206	
	4	0,766	0,610	0,595	0,174	0,005	0,030	0,034	0,649	
Fabric	1	0,436	0,277	0,232	0,340	0,003	0,028	0,045	0,683	0,785
	2	0,577	0,389	0,325	0,660	0,003	0,038	0,066	0,838	
Leaves	1	0,247	0,241	0,142	0,656	0,019	0,019	0,030	0,732	0,573
	2	0,575	0,551	0,421	0,161	0,003	0,003	0,026	0,230	
	3	0,761	0,739	0,620	0,184	0,008	0,008	0,028	0,306	
Paintings	1	0,263	0,245	0,121	0,855	0,014	0,015	0,034	0,905	0,82096
	2	0,535	0,462	0,246	0,050	0,001	0,006	0,085	0,198	
	3	0,643	0,592	0,372	0,095	0,007	0,010	0,081	0,384	

Tabela 3.13 - Distâncias normalizadas entre os histogramas DCD das imagens da Figura 3.46.

Distâncias nominais normalizadas	Bark.0001	Fabric.0006	Leaves.0003	Paintings.31.0000
Bark.0001		0,818	1	1
Fabric.0006	0,818		1	1
Leaves.0003	1	1		0,706
Paintings.31.0000	1	1	0,706	

Pode-se observar na Tabela 3.13 que as imagens dos pares (*Bark*, *Fabric*) e (*Leaves*, *Paintings*) apresentam alguma semelhança já que as distâncias entre elas são menores do que um. A distância entre as imagens do par (*Leaves*, *Paintings*) é menor do que a distância entre as imagens do par (*Bark*, *Fabric*). Isso significa que *Leaves* é mais semelhante a *Paintings* do que *Bark* a *Fabric*. As comparações entre as imagens de um par e de outro resultam distâncias com valor um, ou seja, não são semelhantes. Esses resultados são coerentes.

3.7.2.4 O descritor de leiaute de cores - CLD

O descritor de leiaute de cores, CLD, tem como objetivo capturar a distribuição espacial de cores pela imagem. Ele é um descritor bastante compacto e de cálculo simples ainda que eficaz para permitir uma busca rápida pelo banco de dados. Para o cálculo desse descritor, a imagem deve estar representada no espaço YCrCb. A imagem deve ser

particionada em uma grade de 8 x 8 sub-imagens, resultando em um conjunto de 64 blocos. Para cada bloco, calcula-se a cor média, efetuando-se os cálculos para cada um dos planos: Y , Cr e Cb . O resultado final é uma matriz 8 x 8 de 64 valores médios de cor, para cada plano de cor. Para tornar o descritor mais compacto, efetua-se a transformada discreta de cosseno, DCT, de cada plano e exprime-se os coeficientes em um vetor utilizando-se a mesma técnica zig-zag empregada na compactação JPEG (SALOMON, 2004). O descritor é composto pelos 6 primeiros coeficientes do plano Y , e pelos três primeiros coeficientes de cada um dos planos Cr e Cb . Esses são os coeficientes de menor frequência espacial e que concentram a maior parte da energia da imagem. Esses 12 coeficientes tornam o descritor bastante pequeno. O padrão MPEG-7 ainda especifica que os coeficientes DC sejam quantizados utilizando 6 *bits* e os nove demais coeficientes AC sejam quantizados utilizando 5 *bits*. Isso produz um descritor de apenas 63 *bits* de tamanho.

A distância entre CLD utiliza a métrica L2 com diferenças ponderadas. Calculam-se as distâncias L2 ponderadas para os coeficientes do mesmo plano e somando-se as distâncias para cada plano. Os pesos são maiores para os coeficientes de menor frequência espacial. A Equação (3.83) apresenta a expressão do cálculo da distância entre CLD.

$$\begin{aligned}
 D = & \sqrt{\sum_{i=1}^6 w_{Yi} (DY_i - DY_i')^2} \\
 & + \sqrt{\sum_{i=1}^3 w_{ri} (DCr_i - DCr_i')^2} \\
 & + \sqrt{\sum_{i=1}^3 w_{bi} (DCb_i - DCb_i')^2}
 \end{aligned} \tag{3.83}$$

onde :

DY_i, DCr_i, DCb_i : coeficientes da DCT de cada componente; e
 w_{Yi}, w_{ri}, w_{bi} : pesos para cada diferença de cada componente.

Para ilustrar a utilização do descritor CLD, a Figura 3.47 apresenta as imagens *bark*, *fabric*, *leaves* e *paintings* que vêm sendo utilizadas para os demais descritores e seus

respectivos histogramas CLD na ordem Y , Cr e Cb . A Tabela 3.14 mostra os valores das distâncias cruzadas entre os histogramas CLD para as imagens em questão.

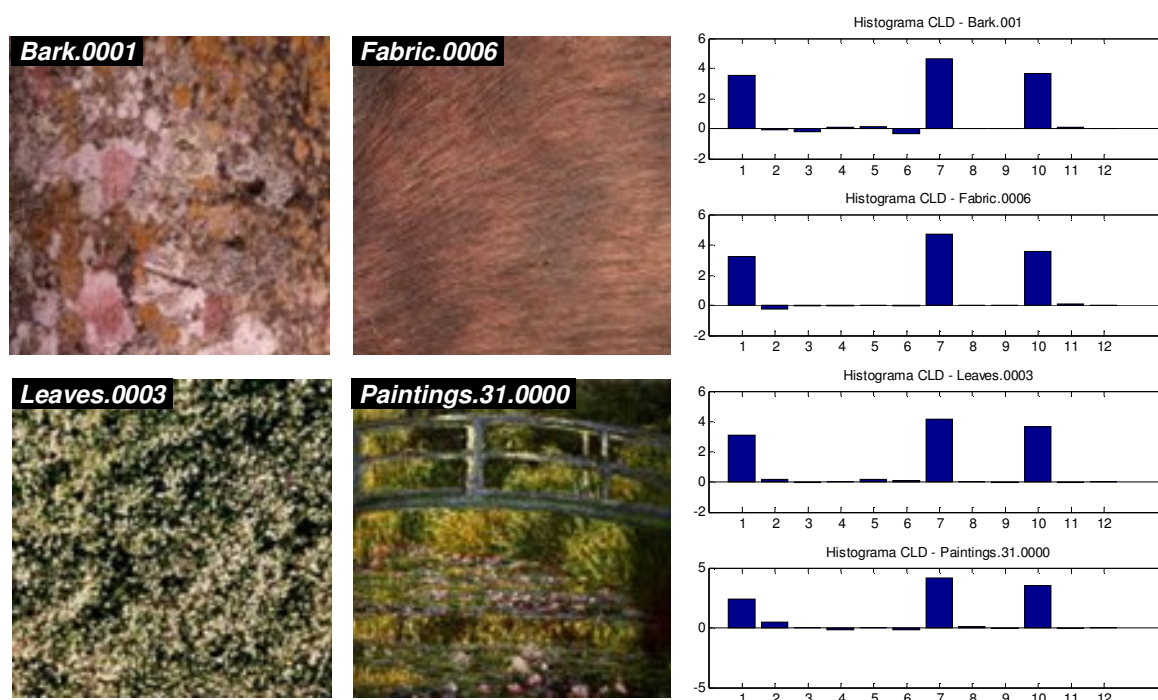


Figura 3.47 – Imagens coloridas e seus histogramas CLD.

Tabela 3.14 – Distâncias entre os histogramas CLD das respectivas imagens da Figura 3.47.

Distâncias nominais	Bark.0001	Fabric.0006	Leaves.0003	Paintings.31.0000
Bark.0001		0,922	1,317	1,671
Fabric.0006	0,922		1,213	1,574
Leaves.0003	1,319	1,213		1,106
Paintings.31.0000	1,671	1,574	1,106	

O comportamento das distâncias mostradas na Tabela 3.14 seguem o comportamento dos demais descritores. No entanto, as diferenças, por não serem normalizadas, não possibilitam uma noção muito clara da semelhança ou dessemelhança entre as imagens. Como já foi mencionado neste capítulo, o descritor CLD não é o melhor dos descritores expostos. Entretanto, por sua facilidade de cálculo e seu tamanho reduzido, pode ser utilizado para uma primeira filtragem do banco de dados.

3.7.3 Os descritores de textura

O padrão MPEG-7 preocupou-se em definir um conjunto de descritores para a análise de textura, além daqueles descritores propostos para a análise de cor, apresentados na seção 3.7.2 deste capítulo. São eles o descritor de exploração de textura, TBD (*texture browsing descriptor*), o descritor de texturas homogêneas, HTD (*homogeneous texture descriptor*) e o descritor de histograma de bordas, EHD (*edge histogram descriptor*). Esses descritores utilizam conceitos e técnicas já apresentados nas seções anteriores deste capítulo. Por isso, esta seção apresentará os descritores de textura propostos pelo padrão MPEG-7 de forma sucinta.

3.7.3.1 O descritor de exploração de textura - TBD

O objetivo do descritor de exploração de textura, TBD, é mensurar três importantes características perceptivas das texturas de uma imagem: regularidade (*regularity*), direcionalidade (*directionality*) e granularidade (*coarseness*). Tamura e outros (TAMURA *et alii*, 1978) definiram essas grandezas perceptivas da seguinte forma. A regularidade está relacionada com a uniformidade dos padrões elementares e a simetria de posicionamento desses elementos na textura. A direcionalidade está relacionada tanto à forma dos elementos primitivos quanto à distribuição desses elementos na textura de forma a criar a percepção de direções principais na imagem. A granularidade relaciona-se tanto ao tamanho quanto à proximidade dos padrões elementares que compõe a textura. Quanto menores e mais próximos são seus elementos constitutivos, mais densa é a textura. Quanto maiores e mais distantes, mais esparsa ela é.

O TBD é um descritor bastante compacto, composto por cinco valores agrupados em três conjuntos. A Equação (3.84) mostra a composição do TBD.

$$TBD = \{r, (d_1, d_2), (c_1, c_2)\}$$

$$r = \begin{cases} 0 & \text{textura irregular} \\ 1, \dots, 3 & \text{para regularidade crescente} \end{cases}$$

(3.84)

$$d_1, d_2 = \begin{cases} 0 & \text{textura não direcional} \\ 1, \dots, 6 & \text{para ângulos de } 0^\circ, 30^\circ, \dots, 150^\circ \end{cases}$$

$$c_1, c_2 = 0, \dots, 3 \quad \text{gradação de granularidade de fina para grossa}$$

A regularidade r recebe valores inteiros de zero a três. O valor zero indica uma textura irregular e os valores não nulos permitem três gradações da regularidade: baixa, média e alta regularidade. O par de valores para a direcionalidade, d_1 e d_2 , permitem registrar até duas direções importantes para a textura. Cada membro do par pode assumir o valor zero, indicando nenhuma direcionalidade, ou valores de um a seis, indicando direções que variam de 0° a 150° em incrementos de 30° . Os ângulos são medidos a partir da horizontal. De forma semelhante, o par de valores para a granularidade, c_1 e c_2 , permitem o registro de até dois valores para essa característica. Cada membro do par recebe valores que variam entre zero e três. Valores menores indicam texturas com granularidade fina e valores maiores indicam texturas com granularidade grossa.

Os cinco valores do descritor TBD são calculados a partir dos resultados da filtragem da imagem por um banco de filtros de Gabor. As respostas de escala e orientação são projetadas nos eixos horizontal e vertical e passam posteriormente por uma operação de autocorrelação. A técnica prossegue com a detecção de picos e vales nessas autocorrelações. Das respostas que apresentam picos e vales regulares, são selecionadas aquelas que apresentam maior contraste. Finalmente, os valores de direcionalidade e granularidade são extraídos a partir dos ângulos e escalas dos bancos de filtro correspondentes às respostas selecionadas. O coeficiente de regularidade é obtido da classificação das respostas selecionadas segundo um critério específico de concentração dessas respostas em torno de um ou mais filtros do banco de filtros de Gabor. A técnica é

apresentada detalhadamente no trabalho de Pu, Manjunath, Newsam e Shin (PU *et alii*, 2000).

O TBD é compacto, ocupa apenas 12 *bits*, e permite a exploração de um banco de imagens pela combinação das três características perceptivas: regularidade, direcionalidade e granularidade.

3.7.3.2 O descritor de texturas homogêneas

O descritor de texturas homogêneas, HTD, permite a busca de imagens por semelhança, mediante uma medida de distância calculada entre os respectivos HTD. A técnica utilizada é, novamente, a filtragem da imagem por um banco de filtros de Gabor, seguida da construção de um descritor constituído por 62 coeficientes. Os dois primeiros são a média e o desvio padrão da intensidade da imagem. Os 60 restantes são calculados a partir das saídas dos 30 filtros do banco. Para cada saída são calculados sua média e seu desvio padrão. Os 30 filtros são projetados pela divisão do espaço de frequências normalizadas em 6 direções e 5 escalas, conforme explicado na seção 3.2.2.5 deste capítulo.

As Equações (3.85) e (3.86) apresentam a constituição do descritor HTD e a métrica de similaridade proposta, respectivamente. Essa métrica baseia-se na Equação (3.10) da seção 3.2.2.5 deste capítulo.

$$HTD = \{\mu_{imagem}, \sigma_{imagem}, \mu_{f_1, \theta_1}, \dots, \mu_{f_5, \theta_6}, \sigma_{f_1, \theta_1}, \dots, \sigma_{f_5, \theta_6}\}$$

onde:

$$\begin{aligned} \mu_{imagem}, \sigma_{imagem} &: \text{média e desvio padrão da intensidade da imagem;} \\ \mu_{f_i, \theta_j}, \sigma_{f_i, \theta_j} &: \text{média e desvio padrão da saída do } ij\text{-ésimo filtro do banco;} \\ f_i \quad i = 1, \dots, 5 &: \text{cada uma das cinco frequências centrais dos filtros do banco; e} \\ \theta_j \quad j = 1, \dots, 6 &: \text{cada uma das seis orientações dos filtros do banco.} \end{aligned} \tag{3.85}$$

$$D(i, j) = \left| \frac{\mu_{imagem}^{(i)} - \mu_{imagem}^{(j)}}{\alpha(\mu_{imagem})} \right| + \left| \frac{\sigma_{imagem}^{(i)} - \sigma_{imagem}^{(j)}}{\alpha(\sigma_{imagem})} \right| + \sum_{f=f_1}^{f_5} \sum_{\theta=\theta_1}^{\theta_6} \left(\left| \frac{\mu_{f\theta}^{(i)} - \mu_{f\theta}^{(j)}}{\alpha(\mu_{f\theta})} \right| + \left| \frac{\sigma_{f\theta}^{(i)} - \sigma_{f\theta}^{(j)}}{\alpha(\sigma_{f\theta})} \right| \right)$$

onde :

$\mu_{f\theta}^{(i)}$: média da saída do filtro $f\theta$ para a imagem i ; (3.86)

$\sigma_{f\theta}^{(i)}$: desvio padrão da saída do filtro $f\theta$ para a imagem i ;

$\alpha(\bullet)$: designa o desvio padrão da grandeza referida calculado;
sobre todas as imagens do banco de dados;

f_i : são as cinco frequências centrais dos filtros do banco; e

$\theta_j = 0^\circ, 30^\circ, 60^\circ, \dots, 0^\circ$: seis orientações dos filtros do banco.

É interessante observar que a técnica HTD permite realizar buscas invariantes à rotação. Para entender o processo, deve-se observar que quando os coeficientes $\mu_{f\theta}$ e $\sigma_{f\theta}$ do descritor HTD_1 da imagem IM_1 são deslocados de uma posição e de forma cíclica, para cada escala f do banco de filtros, o novo descritor obtido é equivalente ao descritor HTD_2 de uma imagem IM_2 , muito semelhante à imagem IM_1 rotacionada de 30° no sentido horário ou anti-horário, dependendo do sentido do deslocamento. Portanto, o deslocamento cíclico dos coeficientes de duas, três ou mais posições, corresponde à rotação da imagem por ângulos de 60° , 90° ou demais múltiplos de 30° . Pode-se calcular a distância invariante à rotação entre duas imagens IM_1 e IM_2 , calculando-se o menor valor das distâncias entre os HDT_1 rotacionados da primeira e o HDT_2 da segunda. A Equação (3.87) ilustra o conceito.

$$D_{IR}(i, j) = \min \{ D(HTD_i |_{m\Delta\theta}, HTD_j) | m = 0, \dots, 5 \text{ e } \Delta\theta = 30^\circ \} \quad (3.87)$$

3.7.3.3 O descritor de histograma de bordas - EHD

O descritor de histograma de bordas, EHD, funciona para as texturas como o descritor de leiaute de cores funciona para as imagens coloridas. Ele busca medir a distribuição espacial de bordas pela imagem. Como já foi dito nas seções iniciais deste

capítulo, as bordas carregam muita informação sobre a textura porque definem regiões de transição de intensidade dos *pixels*.

Os histogramas simples de bordas são utilizados como ferramentas de caracterização de texturas. Entretanto, como todo histograma, eles não conseguem representar a distribuição espacial das bordas na imagem. Para contornar essa característica, o cálculo do EHD determina a divisão da imagem em 16 regiões de mesmo tamanho por meio de uma grade 4 x 4. O histograma de bordas é calculado para cada uma das 16 regiões e a concatenação dos 16 histogramas formam o EHD.

Cada um dos 16 histogramas das regiões da imagem possui cinco *bins*. Um deles computa a ocorrência de regiões isotrópicas, que são aquelas para as quais as bordas não apresentam direção definida. Os outros quatro *bins* contam as ocorrências de bordas de direção 0° (horizontais), 90° (verticais) 45° e 135° (diagonais). Portanto, o EHD é constituído por $16 \times 5 = 80$ coeficientes, cada um deles ocupando três bits, o que totaliza 240 *bits* para os descritores de histogramas de bordas.

O cálculo da direção da borda é bastante simples. Cada região é subdividida em um determinado número de blocos cuja dimensão varia com a dimensão da imagem original. Cada bloco deve ser um quadrado com lado cujo tamanho deve ser uma potência de 2. O número de blocos deve ser o mesmo para todas as regiões da imagem. Cada um desses blocos é subdividido em quatro quadrantes. A intensidade de cada quadrante é calculada como a média da intensidade de seus *pixels*. Assim cada bloco fica resumido a um quadrado de 2 x 2 *pixels*. A esse quadrado são aplicados cinco operadores 2 x 2: um deles isotrópico e os outros quatro direcionais (um para cada uma das quatro direções). Se o maior dos cinco valores correspondentes à aplicação dos operadores é maior do que um patamar determinado, o bloco é considerado um bloco de bordas e sua ocorrência é computada no *bin* de direção correspondente. Caso contrário, se todos os cinco valores forem menores do que o patamar, o bloco não é computado no histograma. Os diagramas da Figura 3.48 ilustram o processo de subdivisão da imagem e apresentam as máscaras detectoras de bordas.

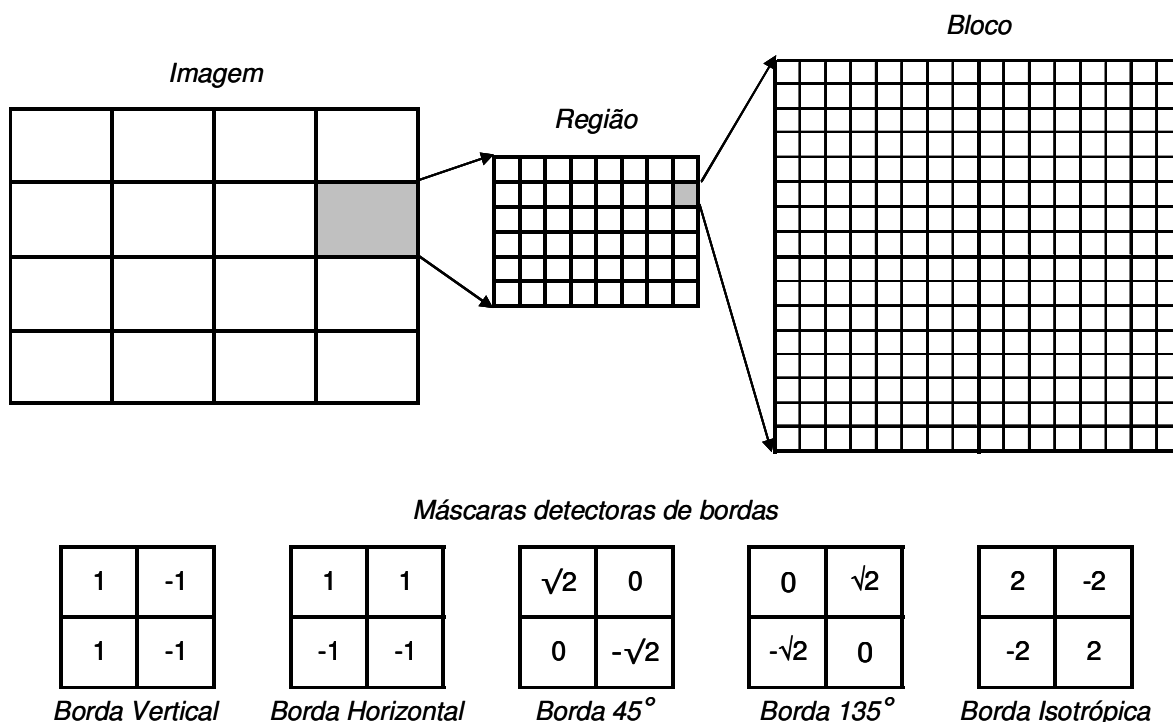


Figura 3.48 – Divisão da imagem e máscaras detectoras de bordas.

Os descritores EHD permitem a comparação de semelhança entre duas imagens pelo cálculo da distância entre seus respectivos descritores. A métrica utilizada é a distância $L1$ calculada como a soma das diferenças absolutas dos componentes do EHD. Os resultados dos testes para as imagens da coleção MPEG-7, cerca de 10000 imagens, reportam ANMRR de aproximadamente 0,34. Esse valor permite considerar os descritores EHD bastante eficientes para a representação de imagens naturais (MANJUNATH *et alii*, 2001).

3.8 Algumas Técnicas Recentes

As seções anteriores deste capítulo apresentaram uma série de técnicas para a determinação de semelhança entre imagens digitais. A maior parte delas se dedica à obtenção de características texturais e algumas outras à determinação de características baseadas na cor. Isso porque o armazenamento e o processamento intensivo de imagens coloridas tornaram-se viáveis apenas recentemente com o surgimento de computadores mais potentes. Todas as técnicas apresentadas são consagradas e já foram submetidas a

testes comparativos de toda ordem, como pode comprovar uma rápida pesquisa pela literatura. No entanto, algumas técnicas mais recentes se destacam, ora pela simplicidade e facilidade de implementação, ora pelo desempenho destacado, ora pela originalidade dos conceitos envolvidos.

Esta seção apresenta algumas dessas técnicas recentes com a finalidade de ilustrar as várias alternativas que se apresentam na busca de novos mecanismos para a determinação de semelhança entre imagens. A primeira delas propõe uma variante dos histogramas de direção de bordas para a obtenção de características invariantes à rotação. A segunda combina histogramas de blocos uni cores, de blocos bicolores e de variação na intensidade do gradiente com a finalidade de compor um aplicativo de exploração de imagens semelhantes em um banco de dados. A terceira técnica propõe o conceito dos padrões de bordas locais, LEP, que estende a proposta dos padrões binários locais (LBP) ao mapa de intensidade de bordas. Finalmente, a quarta técnica propõe a utilização da distância de informação normalizada, um conceito bastante diferente dos apresentados anteriormente neste capítulo para a determinação de semelhança entre imagens.

3.8.1 O histograma da direção e magnitude do gradiente

O gradiente é um operador espacial que mede a taxa de variação da grandeza sobre a qual é aplicado. Por ser um operador vetorial, apresenta duas informações: magnitude e direção. Quando o gradiente é aplicado sobre a intensidade de uma imagem, duas imagens de mesmo tamanho que a original são produzidas: uma delas retrata a taxa de variação dos níveis de cinza e a outra retrata a direção da máxima variação, medida pelo ângulo que o gradiente forma com a horizontal. Essa é a chamada forma polar de expressão de um vetor. Outra forma de exprimir o gradiente de uma imagem é calcular as variações parciais na direção horizontal e vertical separadamente. Para cada *pixel*, estão definidos dois valores que são as componentes horizontal e vertical do vetor gradiente. Essa é a forma cartesiana de expressão de um vetor. A Figura 3.49 apresenta uma imagem e as duas representações do seu gradiente: a forma polar e a forma cartesiana.

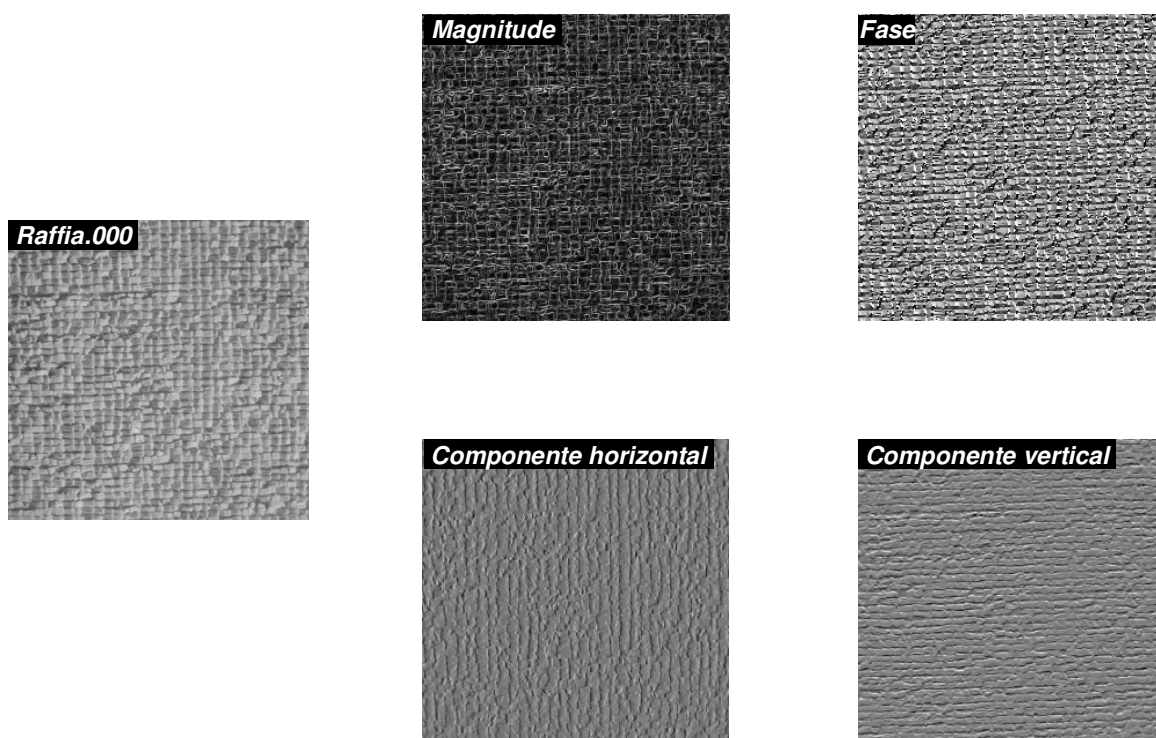


Figura 3.49 – Uma imagem e seu gradiente apresentado nas formas polar e cartesiana.

O gradiente de uma imagem relaciona-se intimamente com as bordas. As bordas são as regiões de transição mais acentuada do nível de cinza da imagem. Como o gradiente mede a taxa de variação do nível de cinza, ele é um dos métodos mais utilizados para a detecção de bordas.

A magnitude do gradiente mede o tamanho da transição. Se uma região da imagem é monótona e uniforme, a magnitude do gradiente apresenta valores pequenos quando calculada para cada *pixel* dessa região. Se, ao contrário, uma região apresenta muitas bordas, a magnitude do gradiente assumirá valores mais elevados. A observação da imagem da magnitude da Figura 3.49 mostra esse comportamento. Os *pixels* mais claros correspondem a magnitudes mais elevadas e os mais escuros, a magnitudes menores. Nota-se que os *pixels* mais claros da imagem da magnitude coincidem com as áreas de maior transição na imagem original.

A direção do gradiente contém bastante informação sobre a imagem original. Observando a Figura 3.49 pode-se verificar que a imagem correspondente à fase apresenta

um padrão visual muito semelhante à imagem original. Essa informação é importante e pode ser utilizada para caracterizar uma imagem.

Stephanie Fountain e Tieneu Tan propuseram a construção de um histograma da direção do gradiente com a finalidade de calcular um vetor de características invariantes à rotação (FOUNTAIN e TAN, 1998). Em linhas gerais, o procedimento é simples. Inicialmente calcula-se a magnitude e a fase do gradiente para cada ponto da imagem. A seguir o histograma de direções é construído percorrendo cada ponto da imagem e somando o valor da magnitude ao *bin* da direção correspondente. Cada *bin* do histograma conterá, não o número de ocorrências da direção correspondente, mas um valor que tenderá a enfatizar as direções que apresentam transições mais significativas. Somente são computados aqueles pontos para os quais a magnitude do gradiente superar um patamar arbitrado.

O histograma de direção do gradiente tende a apresentar muitas transições espúrias. Um processo de suavização por média de três vizinhos foi proposto por Fountain e Tan (FOUNTAIN e TAN, 1998) para corrigir esse problema. Essa suavização é repetida até que se consiga o resultado desejado. O histograma resultante é normalizado. Os picos porventura presentes indicam aquelas direções que preponderam na imagem.

A Figura 3.50 mostra o histograma de bordas original da textura *raffia.000* e cinco versões suavizadas distintas pelo número de vezes que o processo de suavização foi repetido. Os histogramas possuem 360 *bins*.

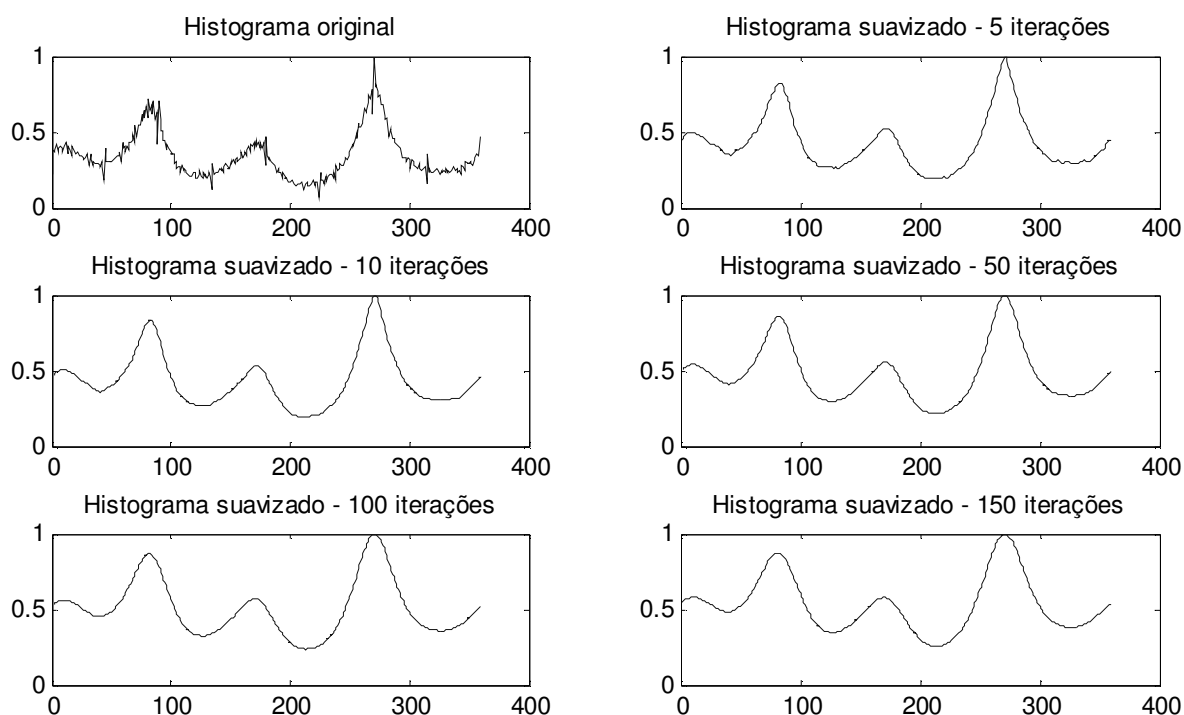


Figura 3.50 – Histograma de direção do gradiente da textura *raffia.000* e suas versões suavizadas.

O histograma de direção do gradiente é claramente cíclico. Seus *bins* são os ângulos formados pelo gradiente com a horizontal. Se uma imagem for rotacionada de um ângulo θ , o histograma da imagem rotacionada será muito semelhante ao histograma da imagem original deslocado do mesmo ângulo θ .

Para ilustrar a afirmação, a Figura 3.51 apresenta as imagens *raffia.000* e *raffia.060*. A textura *raffia.060* corresponde à rotação da textura *raffia.000* por um ângulo de 60° . São também apresentados os histogramas de direção do gradiente para ambas as imagens. Os histogramas foram suavizados por 50 iterações. Pode-se notar que os histogramas são semelhantes. Um é a versão deslocada do outro. Calculando os ângulos correspondentes ao maior pico, verifica-se que a diferença entre eles aproxima-se de 60° .

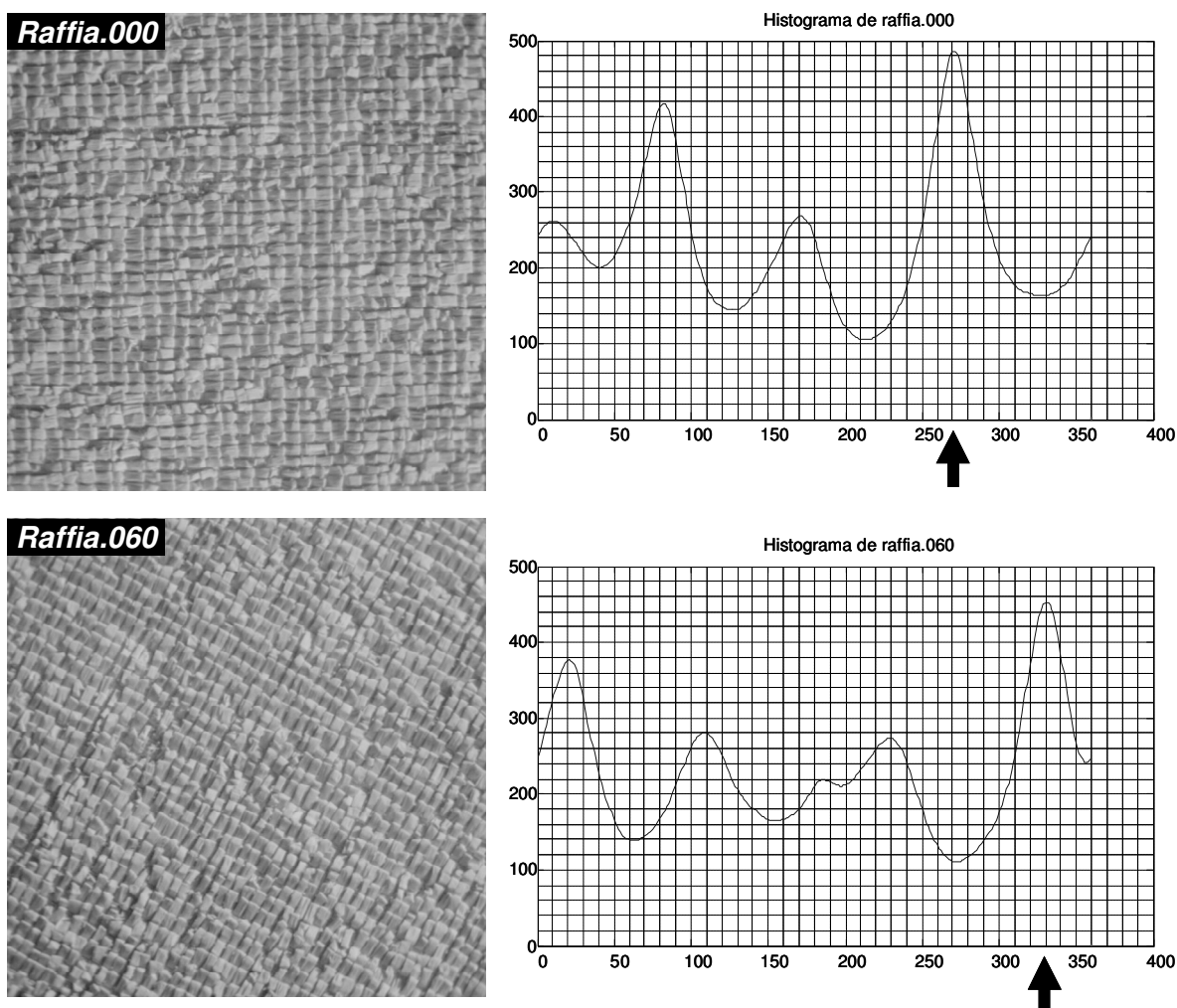


Figura 3.51 – Duas texturas rotacionadas e seus histogramas de direção do gradiente.

Como a rotação da imagem equivale à translação do histograma, Fountain e Tan (FOUNTAIN e TAN, 1998) propuseram aplicar a transformada de Fourier ao histograma. Como essa transformada apresenta invariância à translação, as transformadas dos histogramas de duas imagens em que uma é a rotação da outra são iguais, ou muito semelhantes. Notando-se que a suavização do histograma equivale a uma filtragem passa baixas, basta considerar alguns termos de frequência mais baixa da transformada.

Após exaustivos testes que visaram comparar algumas opções de implementação, Fountain e Tan (FOUNTAIN e TAN, 1998) determinaram os seguintes procedimentos como os melhores: aplicação do operador de Sobel para determinar o gradiente. Compilação do histograma. Suavização do histograma por meio de 100 iterações

do suavizador de três vizinhos. Normalização do histograma pelo valor máximo. Aplicação da transformada de Fourier e retenção das magnitudes dos quatro primeiros coeficientes para formar o descritor invariante à rotação. A métrica recomendada para o cálculo da dessemelhança é a distância euclidiana simples.

A Tabela 3.15 apresenta os valores dos descritores para seis versões rotacionadas da textura *raffia*: 0°, 30°, 60°, 90°, 120°, 150° e 200°. Também apresenta o descritor para a textura *wood* a 0°. O histograma da textura *wood* é bastante diferente do histograma da textura *raffia*. A Tabela 3.16 mostra as distâncias cruzadas entre essas texturas.

Tabela 3.15 – Os descritores para cada uma das versões rotacionadas da textura *raffia* e da textura *straw*.

Coeficientes da FFT	1.º - DC	2.º	3.º	4.º
<i>raffia</i>	0,76560	0,04347	0,10858	0,08235
30	0,77352	0,05634	0,03266	0,13747
60	0,73946	0,07473	0,08641	0,09940
90	0,80562	0,07466	0,08402	0,03570
120	0,78664	0,12455	0,02672	0,06209
150	0,75983	0,00673	0,11744	0,11600
200	0,75679	0,04154	0,04716	0,15451
<i>straw</i>	0,46430	0,08359	0,38485	0,06727

Tabela 3.16 – Distâncias entre as versões rotacionadas da textura *raffia* e a textura *straw*.

Distâncias normalizadas	<i>raffia</i>	30	60	90	120	150	200	<i>straw</i>
<i>raffia</i>	0	0,10	0,05	0,07	0,12	0,05	0,10	0,41
30	0,10	0	0,08	0,12	0,10	0,10	0,03	0,47
60	0,05	0,08	0	0,09	0,10	0,08	0,08	0,41
90	0,07	0,12	0,09	0	0,08	0,12	0,14	0,46
120	0,12	0,10	0,10	0,08	0	0,16	0,13	0,48
150	0,05	0,10	0,08	0,12	0,16	0	0,09	0,41
200	0,10	0,03	0,08	0,14	0,13	0,09	0	0,46
<i>straw</i>	0,41	0,47	0,41	0,46	0,48	0,41	0,46	0

Pode-se observar na tabela Tabela 3.16 que as distâncias entre as versões da textura *raffia* apresentam valores em torno de (0,10) e as distâncias entre a textura *straw* e

as versões da textura *raffia* apresentam valores acima de (0,41). Isso é consistente com o objetivo do descritor.

3.8.2 Os histogramas de cores e de variação da magnitude do gradiente

O trabalho de Nezamabadi-Pour e Kabir procurou combinar cores e variação da magnitude do gradiente para comparar imagens por semelhança, permitindo uma busca por exemplos bastante eficiente. Sua técnica computa três histogramas a partir da imagem: um histograma de probabilidades de direção da variação da magnitude do gradiente da intensidade, HDCIG, um histograma de cor para blocos uniformes, HUCUB, e um histograma de par de cores para blocos desuniformes, HBCNB. O sistema de cores utilizado é o HSV e as distâncias entre os histogramas são calculadas utilizando-se a métrica L1 (NEZAMABADI-POUR e KABIR, 2004).

Para calcular o histograma HDCIG, a imagem é dividida em quatro quadrantes e é calculada a imagem gradiente da intensidade (o plano *V* do sistema HSV). A Figura 3.52 ilustra o processo. Essa divisão da imagem em quadrantes acarreta melhores resultados na sua descrição. Para as imagens naturais, os histogramas dos seus quadrantes podem ser sensivelmente diferentes. Nessa situação, o histograma da imagem inteira compõe uma média que deverá ser razoavelmente diferente dos histogramas dos quadrantes. Ao utilizar regiões de tamanho menor, as diferenças regionais são levadas em conta no cálculo da distância, proporcionando uma maior possibilidade de discriminação entre as imagens.

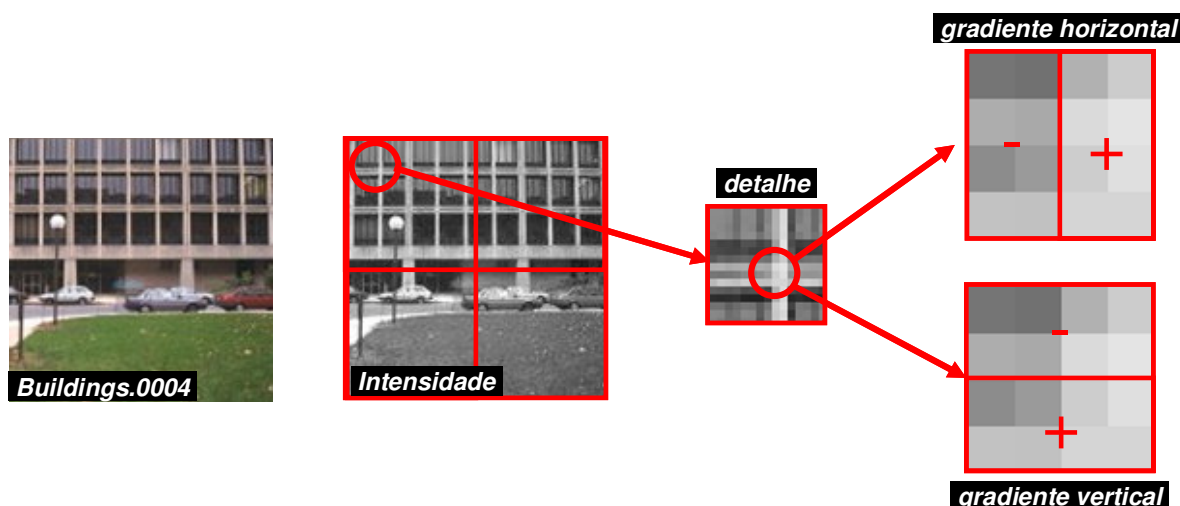


Figura 3.52 – Diagrama que ilustra o cálculo do gradiente.

Conforme esquematizado na Figura 3.52, a imagem gradiente é calculada dividindo cada quadrante em vários blocos 4 x 4 não sobrepostos. Para cada bloco calcula-se a diferença entre o valor médio da metade esquerda e da metade direita do bloco. Essa é a componente horizontal do gradiente. De forma análoga, a diferença entre a metade superior e a metade inferior do bloco determina a componente vertical do gradiente. A intensidade do gradiente é calculada conforme o usual na álgebra vetorial. A dimensão da imagem gradiente é reduzida em uma escala de 16:1 em relação à imagem original.

O passo seguinte é fazer deslizar um elemento 3 x 3 sobre a imagem gradiente de cada quadrante. Para cada posição do elemento, verifica-se se o valor da magnitude do gradiente de cada um de seus oito vizinhos é maior do que a magnitude do gradiente para o *pixel* central. A Figura 3.53 ilustra o procedimento.

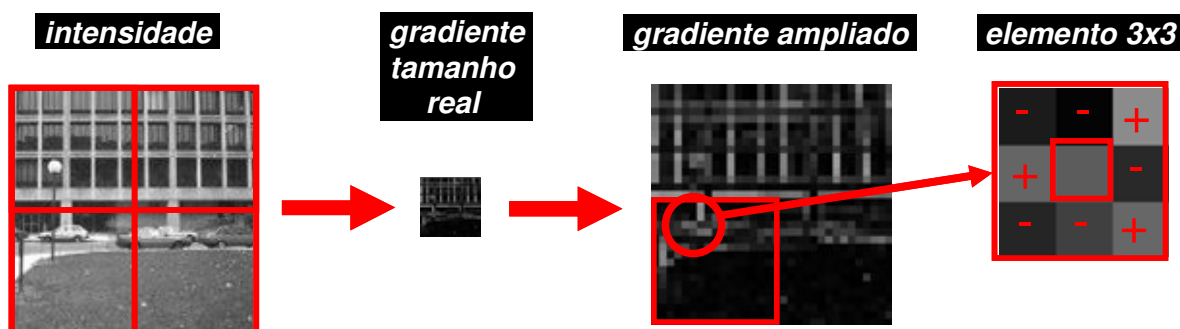


Figura 3.53 – A contabilização dos *bins* do histograma HDCIG.

O histograma HDCIG de cada quadrante tem oito *bins*, cada um correspondendo a um vizinho do *pixel* central em um elemento 3 x 3. Cada *bin* totalizará o número de posições do elemento 3 x 3 para as quais o correspondente vizinho apresenta magnitude maior do que a do *pixel* central. Esses histogramas são normalizados pelo número de elementos do quadrante. O histograma HDCIG da imagem é a concatenação dos histogramas de cada quadrante, sendo composto pelo total de 32 *bins*.

A Figura 3.54 mostra quatro imagens coloridas. Das quatro imagens, as duas *buildings* da linha superior são semelhantes. O mesmo ocorre com as duas imagens *food*. No entanto, não existe semelhança aparente entre as imagens do primeiro e do segundo grupos. A Figura 3.55 mostra os histogramas HDCIG dessas quatro imagens.



Figura 3.54 – Quatro imagens coloridas que são comparadas pelos histogramas HDCIG, HUCUB e HBCNB.

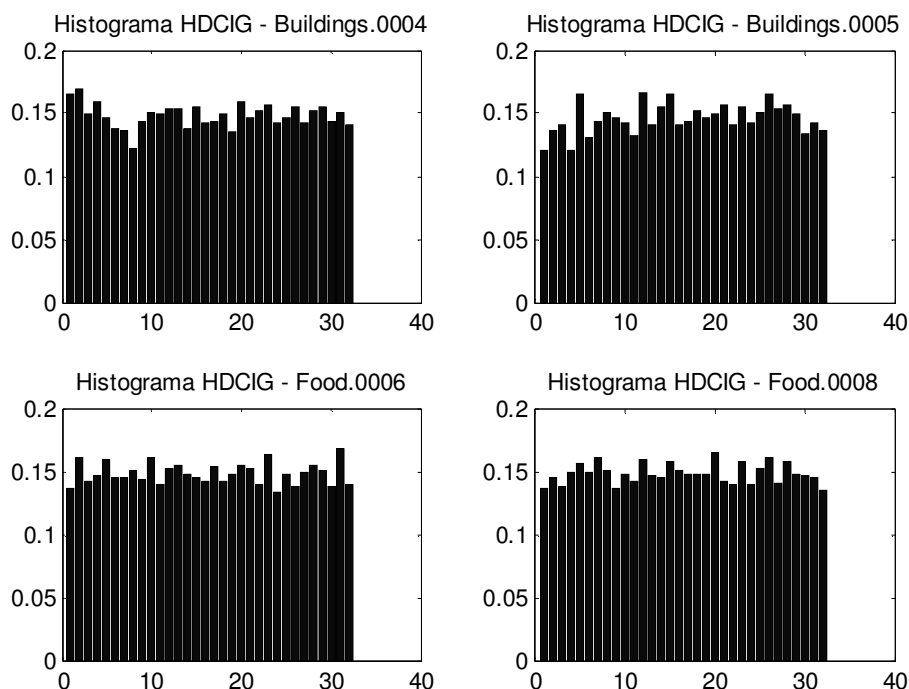


Figura 3.55 – Histogramas HDCIG das imagens da Figura 3.54.

O aspecto visual do histograma HDCIG não é significativo para que se faça qualquer conjectura sobre as imagens retratadas. Entretanto, esse histograma é útil para caracterizar as distâncias entre as imagens.

O histograma de blocos de cores uniformes, HUCUB, é calculado pela identificação dos blocos uniformes e sua posterior contabilização em função da cor média desses blocos. Blocos uniformes são aqueles para os quais a magnitude do gradiente é menor do que um determinado patamar. O gradiente é calculado da mesma forma que no HDCIG.

Quando a magnitude do gradiente de um bloco é baixa, isso significa que a variação das tonalidades dos pixels do bloco também é baixa, ou seja, os *pixels* do bloco têm aproximadamente a mesma cor. Portanto, a cor média dos *pixels* é uma boa característica para descrever esses blocos uniformes.

Um dos objetivos do HUCUB é o de mostrar a distribuição das cores dos blocos na imagem. Já foi mencionado anteriormente neste capítulo que o número de cores disponíveis é muito elevado. Para tornar o histograma efetivo, é fundamental quantizar o

espaço de cores para que o número de *bins* não seja exagerado. A quantização proposta por Nezamabadi-Pour e Kabir (NEZAMABADI-POUR e KABIR, 2004) é dividir o plano H do sistema HSV em seis níveis e os planos S e V em três níveis cada um. O argumento para isso é que o plano H guarda uma variação de cores grandes. Seis níveis são suficientes para representar as três cores primárias e as três cores complementares. Os outros planos podem ser descritos mais grosseiramente por saturação (ou brilho) baixa, média e alta. Essa combinação totaliza 54 níveis distintos que são organizados em um histograma nominal.

Outra função do HUCUB é a de registrar o percentual de blocos uniformes na imagem. Essa função é contabilizada normalizando-se o histograma pela divisão do valor de seus *bins* pelo número total de blocos da imagem. Assim, além da distribuição de cores dos blocos uniformes o HUCUB mostra se uma imagem é mais ou menos uniforme. Isso pode ser observado pela totalização dos *bins* do histograma. Quanto mais esse valor se aproxima de um, mais uniforme é a imagem. Quanto menor o valor total dos *bins*, mais transições e conseqüentemente mais textura a imagem apresenta. É importante ressaltar que grandes áreas monótonas e de uma cor podem ser percebidas como raias de valor mais alto que se destacam no histograma. Efeito semelhante pode ser observado quando existem muitas regiões pequenas de uma mesma cor. Entretanto, se essas regiões tiverem aproximadamente o tamanho do bloco, ou se forem menores, pode ocorrer que se distribuam por blocos vizinhos. Isso faz com que esses blocos contenham mais de uma cor e, portanto, não sejam considerados uniformes, nem sejam contabilizados no HUCUB

A Figura 3.56 mostra os histogramas HUCUB para as imagens exemplo apresentadas na Figura 3.54. Como observado, percebe-se uma raia predominante no HUCUB da imagem *buildings.0004*. Essa raia está na região do verde e se deve à grande área de gramado dessa imagem. As demais cores têm contribuição uniforme. Com relação à imagem *buildings.0005*, pode-se concluir que é uma imagem com mais textura porque os valores dos *bins* são menores. Nas imagens *food.0006* e *food.0008*, predomina a textura. A escala de seus HUCUB é bem menor, na ordem de 100 vezes. Na imagem *food.0006*, as tonalidades avermelhadas (mais à esquerda no HUCUB) são mais presentes. Na imagem

food.0008, as tonalidades verdes e azuis (mais ao centro e à esquerda) ganham importância.

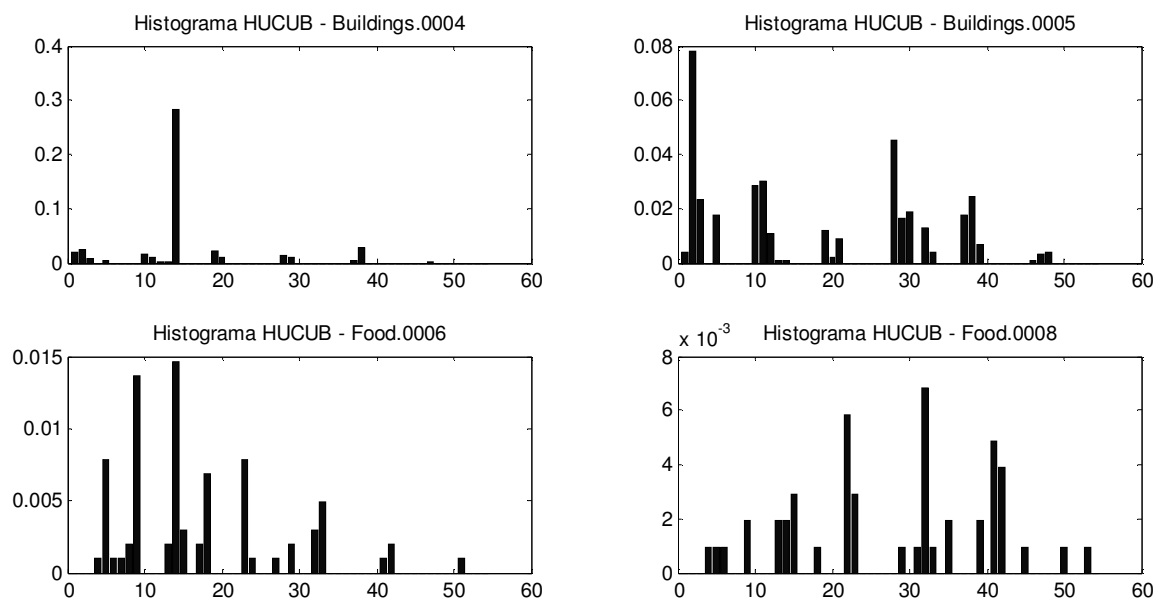


Figura 3.56 – Histogramas HUCUB das imagens da Figura 3.54.

O histograma de duas cores para blocos não uniformes, HBCNB, complementa o conjunto de descritores propostos por Nezamabadi-Pour e Kabir (NEZAMABADI-POUR e KABIR, 2004). Seu objetivo é contabilizar tanto o percentual desses blocos, o que mede a quantidade de textura na imagem, quanto a ocorrência dos pares de cores nesses blocos. Blocos não uniformes são aqueles para os quais a magnitude do gradiente ultrapassa o patamar mencionado anteriormente neste capítulo. Nesse caso, supõe-se que haja uma transição de cores no bloco, definindo uma borda. Não se questiona sobre o formato da borda, ou se existe mais de duas cores presentes no bloco, ou ainda se o bloco possui características ruidosas. O critério da magnitude do gradiente simplesmente particiona os blocos da imagem em duas classes: a dos blocos uniformes e a dos blocos não uniformes. Postula-se que os blocos não uniformes são caracterizados por um par de cores distintas.

O HBCNB é um histograma composto por um grande número de *bins*. Isso porque devem ser contempladas todas as possíveis combinações de duas cores diferentes em um bloco. Pela quantização proposta, o número de pares de cores é de $54 \times 54 = 2916$.

Descontadas as 54 possibilidades de cores iguais para o par, restam 2862. Como o par (cor_1, cor_2) é idêntico ao par (cor_2, cor_1) , o número de combinações distintas cai pela metade. Portanto, o histograma HBCNB tem 1431 *bins*. Esse número é muito grande para os propósitos desejados. Entretanto, é razoável supor que uma imagem não apresenta essa quantidade de combinações. De fato, uma imagem de dimensão 1024×1024 possui $256 \times 256 = 65.523$ blocos. Desse número, pode-se supor que 40% sejam blocos uniformes, restando em torno de 40.000 blocos não uniformes. Uma imagem natural apresenta um alto grau de redundância e normalmente não apresenta regiões em todo o espectro de cores. Isso permite presumir que o HBCNB de uma imagem natural possui uma grande quantidade de *bins* vazios. Assim, a utilização de assinaturas para o armazenamento e cálculo de distâncias poupa espaço de armazenamento e tempo de processamento. A Figura 3.57 mostra que os histogramas HBCNB das imagens exemplo apresentam esse comportamento.

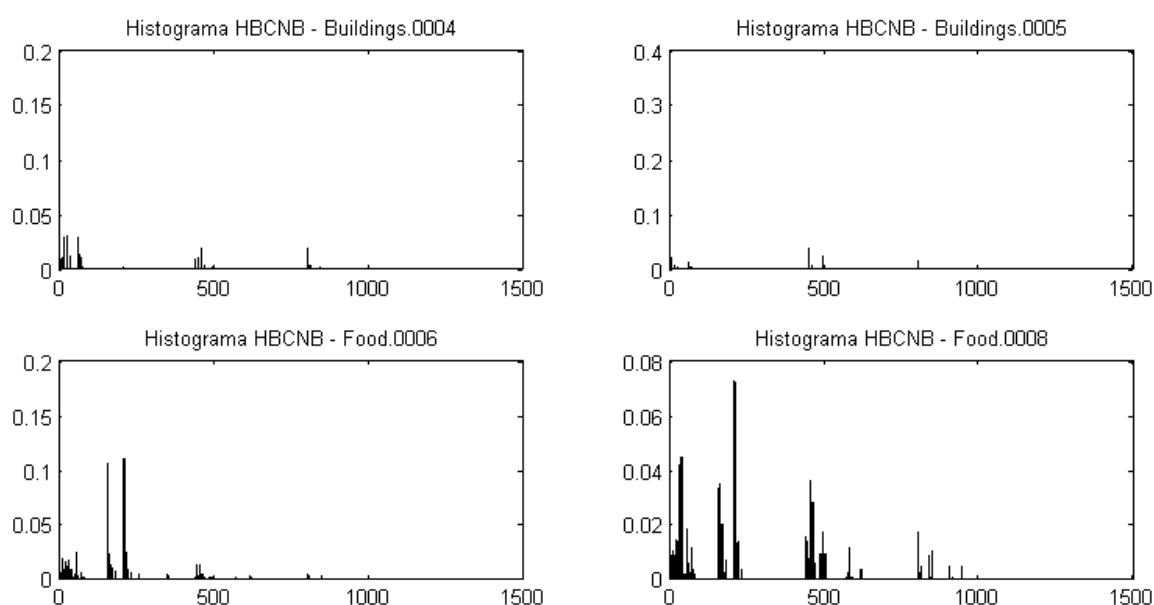


Figura 3.57 - Histogramas HBCNB das imagens da Figura 3.54.

A determinação do par de cores do bloco é bastante simples. Determina-se o brilho médio do bloco pela média dos valores do plano V da imagem. Os *pixels* que possuem brilho maior do que a média são chamados de *pixels* B (*bright*). Os demais são os *pixels* D

(*dim*). Obtém-se a cor média dos *pixels* B e a cor média dos *pixels* D. As duas cores obtidas formam o par de cores do bloco. Devido ao processo de média seguido da quantização, é possível que ambas as cores sejam mapeadas na mesma cor, mesmo que o bloco tenha sido considerado uniforme. Nesse caso, o bloco não é computado no HBCNB. O rótulo do *bin* pode ser computado por consulta a uma tabela para facilitar e agilizar a construção do histograma. A normalização do HBCNB é feita pela divisão dos valores dos *bins* pelo total de blocos da imagem. Esse procedimento tem o mesmo objetivo mencionado para o HUCUB: permitir calcular o percentual de blocos não uniformes na imagem e, por conseguinte, deduzir se nela predomina a cor ou a textura. Assim como no histograma HDCIG, não é fácil determinar outras características da imagem pela simples inspeção visual do HBCNB.

Os três histogramas apresentados nesta seção têm como função determinar a dessemelhança entre as imagens pelo cálculo da distância. A métrica base utilizada é a norma L1. Os detalhes da combinação das distâncias entre esses três histogramas é apresentada na Equação (3.88).

$$D(A, B) = \alpha_1 \left[\alpha_3 \left(\sum_{i=1}^{54} |HUCUB_A(i) - HUCUB_B(i)| + \sum_{i=1}^{1431} |HBCNB_A(i) - HBCNB_B(i)| \right) + \alpha_2 \left(\alpha_4 \sum_{i=1}^{32} |HDCIG_A(i) - HDCIG_B(i)| \right) \right] \quad (3.88)$$

onde :

A, B : são as imagens;

α_1, α_2 : constantes de ajuste de predominância de cor ou textura; e

α_3, α_4 : constantes de normalização.

Na expressão da distância, as constantes α_3 e α_4 têm valor de 1/2 e 1/8, respectivamente. Seu papel é normalizar os termos entre parênteses para que variem no intervalo [0,1]. As constantes α_1 e α_2 têm o papel de permitir um ajuste fino para compensar a distância para mais, ou menos, os histogramas que tratam da cor (HUCUB e