

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Engenharia da Computação e Automação Industrial

ANÁLISE DE TEXTURAS ORIENTADAS E APLICAÇÕES NA MONITORAÇÃO DE MULTIDÕES

Autor: Aparecido Nilceu Marana

Orientador: Prof. Dr. Roberto de Alencar Lotufo

Co-Orientador: Prof. Dr. Luciano da Fontoura Costa

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação, da Universidade Estadual de Campinas, UNICAMP, como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica.

Campinas - São Paulo - Brasil

(Dezembro / 1997)

Este exemplar corresponde a redação final da tese
defendida por Aparecido Nilceu
Marana e aprovada pela Comissão
Julgada em 16/12/97
Orientador

9808087
M325a

33108/BC



UNIDADE	BC
CHAMADA:	unicamp
	325a
Ex.	
DMBO BC/	33108
ROC.	395/98
C	☐
D	☒
RECO	R\$ 11,00
DATA	25/03/98
* CPD	

CM-00108588-1

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

M325a Marana, Aparecido Nilceu
Análise de texturas orientadas e aplicações na
monitoração de multidões / Aparecido Nilceu Marana.--
Campinas, SP: [s.n.], 1997.

Orientadores: Roberto de Alencar Lotufo , Luciano da
Fontoura Costa.

Tese (doutorado) - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Hough, Funções de. 2. Processamento de imagens.
3. Visão por computador. I. Lotufo, Roberto de Alencar, .
II. Costa, Luciano da Fontoura. III. Universidade Estadual
de Campinas. Faculdade de Engenharia Elétrica e de
Computação. IV. Título.

*Aos meus pais,
Lourdes e Nilso.*

*À minha esposa,
Denise.*

*À minha filha,
Fernanda.*

*“Agradeço a Deus,
pelas oportunidade de trabalho
e aprendizado.”*

Agradecimentos

Ao concluir este trabalho, sou grato:

Aos meus orientadores, Prof. Dr. Roberto de Alencar Lotufo e Prof. Dr. Luciano da Fontoura Costa, por terem viabilizado a sua realização, pelo tempo e atenção dedicados, pela segurança na orientação, pelo respeito profissional, pela amizade e pelas reuniões *aos sábados*;

Ao Dr. Sergio Velastin, do *Vision and Robotics Laboratory*, do *Department of Electrical and Electronics Engineering*, do *King's College London*, Londres, Reino Unido, pela amável recepção em Londres e pela orientação nas pesquisas sobre “crowd monitoring”, durante o ano de 1996;

Ao *Grupo de Pesquisa em Visão Cibernética*, especialmente ao Prof. Dr. Luciano da Fontoura Costa, do Instituto de Física de São Carlos, USP, São Carlos, pelo imenso apoio durante a realização dos trabalhos;

À CAPES, através do programa PICD, pelo apoio financeiro concedido para a realização de parte deste trabalho no Brasil;

Ao CNPq, através do programa de *Bolsas Sanduíche*, pelo apoio financeiro concedido para a realização de parte deste trabalho no Reino Unido;

À UNESP, Universidade Estadual Paulista, pelos afastamentos concedidos.

Agradeço também:

À minha querida esposa Denise, pela companhia e apoio em todas as etapas deste trabalho;

Ao membros do *Vision and Robotics Laboratory*, especialmente, ao Kaspar, à Gabriela e ao Tallat, que facilitaram minha adaptação ao laboratório e à Londres;

À Dra. Maria Alicia Vicencio-Silva, da University College London, Londres, Reino Unido, por ter sugerido a utilização de textura para medir densidades de congestionamentos de pessoas;

À Railtrack PLC, Londres, Reino Unido, por ter permitido o acesso a suas instalações;

Ao Jia Hong, pelo auxílio na aquisição das imagens de multidões utilizadas neste trabalho;

Aos amigos brasileiros do Lillian Penson Hall, pela “Brazilian Party”;

Ao João Batista, Milton e Munir, doutorandos do Imperial College, pelas dicas e favores prestados;

Ao “Spiritist Family Group”, pela recepção fraterna, e pelas reuniões londrinas de estudo e confraternização;

Ao amigo, Prof. Dr. Marco Antonio Leonel Caetano, por ter sido um procurador tão eficaz;

Ao amigo, Prof. Dr. João Eduardo Ferreira, por ter assumido o ônus da prorrogação do meu afastamento junto ao DEMAC;

Ao amigo, Prof. Dr. Hélio Aparecido Navarro, por ter dividido comigo as atribuições do PET;

Aos demais amigos do DEMAC - UNESP - Rio Claro, pelo constante apoio.

“O esforço da pessoa para libertar-se de pressuposições e dominar seu desejo de um resultado específico deve acompanhar constantemente a busca da verdade.”

Alfred L. Baldwin

“It is widely believed that texture measurements on images are an important means of classifying and analyzing image fields. Yet, texture measurement remains a highly nebulous subject. Many indicators and measurements have been proposed, but their effectiveness is questionable because the structural characteristics of natural fields are still ill defined.”

W. K. Pratt

Resumo

Um dos objetivos desta tese é propor novos descritores para a caracterização e classificação de texturas orientadas, inspirados nos sistemas biológicos de visão, dentre os quais o sistema de visão humano, que é particularmente sensível a estímulos retilíneos. A aplicação dos novos descritores, que dependem dos segmentos de retas que são extraídos das imagens através da transformada de Hough, na classificação de texturas orientadas reais apresentou bons resultados. Outro objetivo desta tese foi avaliar a eficácia das medidas de texturas e a relevância das informações sobre orientação na estimação de níveis de densidades de congestionamentos de pessoas. Tal estimação é crucial para o gerenciamento e controle de multidões. A avaliação foi realizada sobre um conjunto de aproximadamente 300 imagens, capturadas na estação de trens de Liverpool Street, em Londres, Reino Unido, utilizando-se 4 medidas de texturas extraídas das imagens por meio dos quatro seguintes métodos: matrizes de dependência espacial dos tons de cinza, segmentos de retas (proposto nesta tese), análise de Fourier e dimensões fractais. As estimações das densidades de congestionamentos são dados em termos das classificações das imagens de entrada em cinco classes de densidades (baixíssima densidade, baixa densidade, densidade moderada, alta densidade e altíssima densidade). Durante os experimentos, utilizaram-se três tipos de classificadores: neural (implementado segundo o modelo de Kohonen), Bayesiano, e uma abordagem baseada em funções de ajuste. Os resultados obtidos pelos classificadores, utilizando as quatro medidas de texturas, permitiu concluir-se que, para o problema da estimação de densidades de congestionamentos de pessoas: (i) as texturas são medidas bastante efetivas; (ii) orientação não é essencial; (iii) os descritores para texturas orientadas baseados em segmentos de retas, propostos nesta tese, possibilitam um desempenho equivalente àqueles produzidos por descritores tradicionais que também foram utilizados nos experimentos.

Abstract

One of the goals of this thesis is to present a new method for oriented texture characterization and classification, inspired in biological vision systems, including the human vision system, which are particularly sensitive to rectilinear stimuli. The features proposed, which are based on straight line segments extracted from the image by using the Hough transform, were used to classify real oriented textures, providing good results. Another goal of this thesis was to assess the efficacy of texture measures and the relevance of orientation as means for estimating levels of crowd densities in images. This estimation is crucial for the problem of crowd monitoring and control. The assessment was carried out on a set of almost 300 real images captured from Liverpool Street Train Station, London, UK, using texture measures extracted from the images through the following four different methods: gray level dependence matrices, straight line segments (proposed in this thesis), Fourier analysis, and fractal dimensions. The estimations of crowd densities are given in terms of the classification of the input images in five classes of densities (very low, low, moderate, high and very high). During the experiments, three types of classifiers were used: neural (implemented according to the Kohonen model), Bayesian, and an approach based on fitting functions. The results obtained by these three classifiers, using the four texture measures, allowed the conclusion that, for the problem of crowd density estimation: (i) texture is a very effective measure; (ii) orientation is not essential; (iii) the proposed texture analysis, based on straight line segments, yields performance similar to the other traditional texture methods considered in the experiments.

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos	5
1.3	Estrutura da Tese	6
2	Análise de Textura	8
2.1	Processamento de Textura no Sistema de Visão Humana	11
2.1.1	Teoria dos Textons	14
2.1.2	Teoria do “Primal Sketch”	17
2.1.3	Teoria da Integração de Características	18
2.1.4	Aspectos Neurais da Percepção de Textura	20
2.2	Processamento de Textura nos Sistemas de Visão Computacional	25
2.2.1	Abordagens Estatísticas	27
2.2.1.1	Matrizes de Dependência Espacial dos Tons de Cinza	27
2.2.1.2	Matrizes de Dependência Espacial Generalizada .	33
2.2.1.3	Diferença dos Tons de Cinza	34

2.2.1.4	Matrizes de Co-ocorrência de Comprimentos de Cadeias de Tons de Cinza	35
2.2.1.5	Análise de Fourier	37
2.2.1.6	Autocorrelação	38
2.2.1.7	Filtros de Laws	39
2.2.1.8	Filtros de Gabor	41
2.2.1.9	Wavelets	44
2.2.1.10	Morfologia Matemática	45
2.2.2	Abordagens Estruturais	46
2.2.2.1	Gramáticas de Formas	46
2.2.2.2	Densidade de Bordas	47
2.2.2.3	Densidade de Máximos e Mínimos	47
2.2.3	Abordagens Baseadas em Modelos	48
2.2.3.1	Modelos de Auto-regressão	48
2.2.3.2	Modelos dos Campos de Markov	49
2.2.3.3	Modelo dos Caminhos Aleatórios	50
2.2.3.4	Modelo de Fractais	50
3	Análise de Texturas Orientadas Baseada em Segmentos de Retas	56
3.1	Texturas Orientadas	56
3.2	Extração de Descritores de Texturas Orientadas	62
3.3	Transformada de Hough	64
3.3.1	Descritores de Texturas Extraídas de Segmentos de Retas	71
3.3.1.1	Técnica Baseada em Ângulos	72

3.3.1.2	Técnica Baseada em Intensidades Direcionais . . .	79
3.3.2	Descritores de Texturas Extraídas do Espaço de Hough	79
3.3.2.1	Amostragem Não-Uniforme do Espaço de Hough	80
3.3.2.2	Medidas da Distribuição dos Picos do Espaço de Hough	81
3.3.3	Determinação dos Picos do Espaço de Hough	84
3.3.4	Análise de Conectividade	89
4	Reconhecimento de Padrões	91
4.1	Classificação	91
4.1.1	Classificador Estatístico Ótimo	94
4.1.2	Redes Neurais	98
4.1.2.1	Mapas de Auto-Organização	100
4.2	Funções de Ajuste	103
4.2.1	Estimação dos Coeficientes de Polinômios Cúbicos . . .	104
5	Gerenciamento e Controle de Multidões	107
5.1	Estimação dos Níveis de Congestionamento Baseada nos Pix- els das Imagens	110
5.2	Estimação dos Níveis de Congestionamento Baseada nas Tex- turas das Imagens	113
5.2.1	Medidas Locais de Textura	118
5.2.2	Medidas Globais de Textura	121
6	Aplicação das Técnicas Baseadas em Textura para a Es- timação de Densidades de Multidões	125

6.1	Medidas Locais de Texturas	131
6.1.1	Matrizes de Co-ocorrência com Classificador Neural . .	132
6.2	Medidas Globais de Texturas	134
6.2.1	Matrizes de Co-ocorrência	135
6.2.1.1	Classificador Bayesiano	135
6.2.1.2	Redes Neurais	137
6.2.1.3	Funções de Ajuste	141
6.2.2	Segmentos de Retas	145
6.2.2.1	Classificador Bayesiano	146
6.2.2.2	Redes Neurais	148
6.2.2.3	Funções de Ajuste	150
6.2.3	Análise de Fourier	156
6.2.3.1	Classificador Bayesiano	156
6.2.3.2	Redes Neurais	158
6.2.3.3	Funções de Ajuste	160
6.2.4	Fractais	165
6.2.4.1	Classificador Bayesiano	165
6.2.4.2	Redes Neurais	166
6.2.4.3	Funções de Ajuste	168
6.3	Comparação dos Resultados e Conclusões	170
7	Conclusões	180
8	Trabalhos Futuros	185

A Estimativas Manuais das Densidades de Congestionamentos de Pessoas nas Imagens Utilizadas nos Experimentos	187
---	-----

Lista de Figuras

- 1.1 Esquema do relacionamento entre a Neurofisiologia e a Psicofísica. 3
- 1.2 Paradigma de sistemas de visão computacional. 4

- 2.1 Microelementos numa escala constituem macroelementos numa escala maior (imagem extraída do álbum de texturas de Brodatz). 9
- 2.2 Ilustração da percepção de textura de forma atencional e pré-atencional. 12
- 2.3 Imagens constituídas de pares de texturas que possuem estatísticas de segunda ordem idênticas. Os pares de texturas são indistinguíveis pelo mecanismo de visão pré-atencional. . . 15
- 2.4 Imagens constituídas de pares de texturas que possuem estatísticas de segunda ordem idênticas. Os pares de texturas são distinguíveis pelo mecanismo de visão pré-atencional. . . . 15
- 2.5 Esquema do modelo de integração de características. 19
- 2.6 Respostas de uma “grating cell” para as variações nas frequências espaciais e nas orientações dos estímulos. 22

2.7	Campo receptivo das células simples.	23
2.8	Campo receptivo das células complexas.	24
2.9	Campo receptivo das células hiper-complexas.	25
2.10	Exemplos de texturas determinísticas (imagens extraídas do álbum de texturas de Brodatz).	26
2.11	Exemplos de texturas estocásticas (imagens extraídas do álbum de texturas de Brodatz).	26
2.12	Exemplo de matriz de dependência dos tons de cinza. (a) I- magem com tons de cinza variando 0 a 3; (b) Matriz de de- pendência dos tons de cinza calculados fixando-se os parâmetros $d = 1$ pixel e $\theta = 45^\circ$	29
2.13	Banco de filtros de Laws.	40
2.14	Sequência de imagens obtidas com as dilatações de uma ima- gem binária original, com elementos estruturantes quadrados com tamanhos variando de 3x3 até 22x22 pixels.	54
2.15	(a) Relação entre as medidas dos tamanhos dos elementos es- tructurantes (T) e os números de pixels pretos (N) nas imagens dilatadas; (b) Reta determinada pela regressão linear do con- junto de pontos ($\log(T), \log(N)$).	55
3.1	Exemplos de texturas orientadas (imagens extraídas do álbum de texturas de Brodatz); (a) Texturas orientadas fortemente acopladas; (b) Texturas orientadas fracamente acopladas. . . .	57

3.2	Exemplos de padrões de direcionalidade. (a) Objetos excêntricos alongados numa orientação preferencial; (b) Bordas alinhadas numa orientação preferencial; (c) Objetos concêntricos dispostos em orientações preferenciais.	60
3.3	Componentes que influenciam na medida de intensidade direcional. (a) Intensidade dos tons cinza; (b) Dispersão angular; (c) Comprimento; (d) Densidade espacial.	61
3.4	Parametrização normal da reta.	65
3.5	(a) Imagem constituída de 17 retas; (b) Espaço de Hough da imagem apresentada em (a).	68
3.6	(a) Imagem sintética de entrada; (b) Segmentos detectados via TH; (c) Histograma de orientações; (d) Região segmentada caracterizada pela orientação próxima a 30^0 ; (e) Região segmentada caracterizada pela orientação próxima a 120^0 ; (f) Texturas classificadas.	74
3.7	(a) Imagem de entrada; (b) Bordas detectadas; (c) Segmentos de retas detectados via TH; (d) Texturas classificadas; (e) Histograma de orientações; (f) Espaço de Hough da imagem das bordas apresentada em (b).	78
3.8	(a) Imagem D17 do álbum de texturas de Brodatz; (b) Classificação baseada nos filtros de Gabor; (c) Classificação baseada na amostragem do espaço de Hough; (d) Classificação baseada nas distribuições dos picos do espaço de Hough.	82

3.9	(a) Imagem com bordas detectadas; (b) Segmentos de reta detectados utilizando-se a técnica simples de limiarização; (c) Segmentos de reta detectados utilizando-se a técnica de “back-mapping” com análise de conectividade; (d) Segmentos de reta detectados utilizando-se a técnica de remoção de pixels da imagem original.	85
4.1	Esquema do reconhecimento de padrões supervisionado.	92
4.2	Esquema típico de classificação de padrões.	94
4.3	Funções unidimensionais de densidades de probabilidades para duas classes de padrões, ω_1 e ω_2	98
4.4	Função de ajuste polinomial cúbica unidimensional.	106
5.1	Imagens de congestionamentos de pessoas de diferentes densidades (linha superior) com suas bordas detectadas (linha inferior). Da esquerda para a direita, as rugosidades dos padrões de textura aumentam, conforme aumentam as densidades dos congestionamentos de pessoas. Pode-se perceber que as quantidades de pixels de bordas não aumentam linearmente com os aumentos dos números de pessoas na cena. Na imagem mais a direita da linha inferior é possível notar uma discreta redução do número de pixels de borda detectados, comparando-se com a imagem anterior da sequência.	112
5.2	Diagrama da técnica para estimação de densidade de multitudes baseada em medidas locais de texturas.	120

5.3	Relacionamento entre a densidade de multidões e o formato dos histogramas de classes de texturas. Na coluna da esquerda estão imagens com as áreas sob monitoramento definidas por janelas quadradas centralizadas na imagem. Na coluna central estão as imagens classificadas. Na coluna da direita estão os histogramas de classes das imagens classificadas, onde os níveis baixos de tons de cinza representam baixos níveis de congestionamentos de pessoas, enquanto que os altos níveis de tons de cinza representam altos níveis de congestionamentos de pessoas. Os histogramas estão apresentados normalizados com relação à maior moda.	122
5.4	Diagrama da técnica para estimação de densidade de multidões baseada em medidas globais de texturas.	123
6.1	Conjunto de imagens de congestionamentos de pessoas. (a) Baixíssimas densidades; (b) Baixas densidades; (c) Densidades moderadas; (d) Altas densidades; (e) Altíssimas densidades. . .	127
6.2	Exemplos de densidades de multidões estimadas manualmente. (a) imagem #76 (7 pessoas) - Baixíssima Densidade; (b) imagem #27 (24 pessoas) - Baixa Densidade; (c) imagem #32 (40 Pessoas) - Densidade Moderada; (d) imagem #40 (52 pessoas) Alta Densidade; (e) imagem #280 (82 pessoas) Altíssima Densidade.	128

6.3	(a) Histograma dos números de pessoas nas imagens do conjunto de treinamento; (b) Histograma dos números de pessoas nas imagens do conjunto de teste.	129
6.4	Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 2 medidas: Contraste 0^0 e Homogeneidade 0^0 (82,52% de acerto); (b) Classificador Neural com 16 medidas: Contraste (0^0 , 45^0 , 90^0 e 135^0), Homogeneidade (0^0 , 45^0 , 90^0 e 135^0), Energia (0^0 , 45^0 , 90^0 e 135^0) e Entropia (0^0 , 45^0 , 90^0 e 135^0) (80% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: Contraste 0^0 e Homogeneidade 0^0 (82,76% de acerto).	138
6.5	Relação entre as medidas Contraste 0^0 e Homogeneidade 0^0 . .	139
6.6	Relação entre a medida de Contraste 0^0 e o número de pessoas na imagem.	139
6.7	Relação entre a medida Homogeneidade 0^0 e o número de pessoas na imagem.	140
6.8	Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.	144

6.9	Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 8 medidas: NS, MC(10%), MC, DPC, EC, MA, DPA e EA (78,62% de acerto); (b) Classificador Neural com 4 medidas: MC, DPC, EC e DPA (72,41% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: DPC e DPA (72,41% de acerto).	149
6.10	Relação entre as medidas DPC e DPA.	153
6.11	Relação entre a medida DPC e o número de pessoas na imagem.154	
6.12	Relação entre a medida DPA e o número de pessoas na imagem.154	
6.13	Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.	155
6.14	Regiões do espectro de Fourier onde foram tomadas medidas de textura. (a) 16 regiões radiais; (b) 8 anéis concêntricos. . .	157
6.15	Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 8 medidas radiais (1,3,5, 7,9,11,13,15) e 4 circulares (17,19,21,23) (70,34% de acerto); (b) Classificador Neural com todas as 24 medidas (75,17% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: 1 radial (4) e 1 circular (17) (74,48% de acerto).	159
6.16	Relação entre as medidas 4 (Rad 4) e 17 (Circ 1).	163

6.17	Relação entre a medida 4 (Rad 4) e o número de pessoas na imagem.	163
6.18	Relação entre a medida 17 (Circ 1) e o número de pessoas na imagem.	164
6.19	Gráfico dos erros relativos e absolutos das estimativas do nú- mero de pessoas nas imagens do conjunto de teste.	165
6.20	Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com a dimensão fractal da imagem (71,03% de acerto); (b) Clas- sificador Neural com a dimensão fractal da imagem (78,62% de acerto); (c) Classificador baseado em funções de ajuste a dimensão fractal da imagem (75,86% de acerto).	167
6.21	Relação entre a dimensão fractal e o número de pessoas na imagem.	170
6.22	Gráfico dos erros relativos e absolutos das estimativas do nú- mero de pessoas nas imagens do conjunto de teste.	171
6.23	Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste (“fitting”), empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da di- mensão fractal.	174

6.24	Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste, empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da dimensão fractal. (a) classe de baixíssimas densidades; (b) classe de baixas densidades; (c) classes de densidades moderadas.	176
6.25	Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste, empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da dimensão fractal. (a) classe de altas densidades; (b) classe de altíssimas densidades.	177

Lista de Tabelas

5.1 Definição de Níveis de Serviço. 117

6.1 Relacionamento entre as classes de densidades de congestionamentos e os níveis de serviços. 130

6.2 Número de imagens dos conjuntos de treinamento e de teste nas cinco classes de densidades (VL,L,M,H e VH). 131

6.3 Resultados obtidos aplicando-se a técnica baseada em medidas locais de textura extraídas de matrizes de co-ocorrências dos tons de cinza das imagens. As linhas da tabela mostram as distribuições das estimações das densidades de multidões para cada grupo de imagens do conjunto de teste. Os valores da diagonal da tabela (impressos em negrito) indicam as taxas percentuais de acerto das classificações para cada grupo. . . . 134

6.4 Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e as medidas das matrizes de co-ocorrência. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. 136

6.5	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas, cujos pares de medidas proporcionaram taxas de acerto acima de 75%. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. Hg = Homogeneidade; Ct = Contraste; Ep = Entropia.	142
6.6	Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. Hg = Homogeneidade; Ct = Contraste; Ep = Entropia.	144
6.7	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.	147
6.8	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.	150
6.9	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. .	151

6.10	Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.	155
6.11	Resultado das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e medidas de textura extraídas do espectro de Fourier. . .	157
6.12	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e 24 medidas de texturas extraídas do espectro de Fourier. . .	160
6.13	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas e medidas de textura extraídas do espectro de Fourier. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.	161
6.14	Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.	164
6.15	Resultado das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e dimensões fractais das imagens.	166
6.16	Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e as dimensões fractais das imagens como descritores de texturas.	168

6.17	Resultado das classificações das densidades de multidões das imagens do conjunto de teste, utilizando funções de ajuste e dimensões fractais das imagens.	169
6.18	Melhores resultados obtidos com as quatro medidas de textura (co-ocorrências dos tons de cinza, segmentos de retas, espectro de Fourier e dimensões fractais) e com os três classificadores (estatístico, neural e baseado em funções de ajuste). Os valores da tabela marcados por retângulos indicam os melhores resultados obtidos.	172
6.19	Taxas de classificações corretas de cada classe de densidades de congestionamentos, para os melhores resultados obtidos com as quatro medidas de texturas (co-ocorrências dos tons de cinza, segmentos de retas - Hough, espectro de Fourier e dimensões fractais), com os classificadores estatístico (Bayes), neural (modelo de Kohonen) e baseados em funções de ajuste (polinômios cúbicos). Os valores da tabela marcados por retângulos indicam os melhores resultados obtidos.	175
A.1	Distribuição das imagens entre as classes de densidades e os conjuntos de treinamento e de teste. Os números fora dos parênteses são os identificadores das imagens e aqueles entre parênteses são as estimativas manuais de pessoas presentes em cada imagem.	188

Capítulo 1

Introdução

1.1 Motivação

Ao contrário de outras áreas da Computação, nas quais o desempenho das máquinas superou o desempenho humano, tanto em velocidade, quanto em quantidade e qualidade de processamento, na área de Visão Computacional o desenvolvimento de máquinas com capacidade de visão similar ao sistema de visão humano (SVH) ainda é uma meta a ser atingida [19, 70].

Sendo resultados de processos evolutivos biológicos, os sistemas naturais de visão, e em particular o SVH, não possuem os melhores e mais adequados métodos para serem implementados nos sistemas de visão computacional (SVC), que apresentam características bastante diferentes. No entanto, os sistemas naturais continuam inspirando o desenvolvimento dos sistemas artificiais [21]. Essa inspiração justifica-se quando se observa a notável capacidade de reconhecimento de padrões complexos dos SVH, por exemplo.

Os sistemas naturais de visão ainda são muito pouco compreendidos, o que tem contribuído para limitar o desenvolvimento dos sistemas artificiais. Outros fatores que contribuem para essa limitação estão relacionados aos grandes esforços computacionais exigidos para o processamento das enormes quantidades de informações presentes nas imagens e pela complexidade dos processos envolvidos.

As pesquisas sobre o processamento natural de visão têm sido realizadas nas áreas da *Neurofisiologia*, cujos objetos de estudo são os processos realizados por células especializadas do sistema nervoso, os neurônios, e da *Psicofísica*, cujos objetos de estudo são o pensamento, as percepções, as sensações e outros aspectos da experiência consciente dos indivíduos [121].

No âmbito da Neurofisiologia, os experimentos são normalmente realizados em sistemas naturais, exceto no SVH, uma vez que empregam técnicas invasivas para medirem os níveis de atividades dos neurônios, diante de determinados estímulos visuais. As funções neurológicas do SVH têm sido determinadas por analogia através de experimentos realizados nos sistemas visuais dos primatas, que são bastante semelhantes ao SVH.

No âmbito da Psicofísica, os experimentos baseiam-se em respostas verbais ou ações motoras específicas para determinados estímulos. Por esse motivo, a grande maioria dos experimentos psicofísicos são realizados com humanos.

A Figura 1.1 mostra um esquema do relacionamento entre a Neurofisiologia e a Psicofísica. Tal esquema foi apresentado por Levine [70], que também apresentou um paradigma para os sistemas de visão computacional, segundo o qual os SVCs são constituídos de dois estágios, sendo o primeiro de baixo

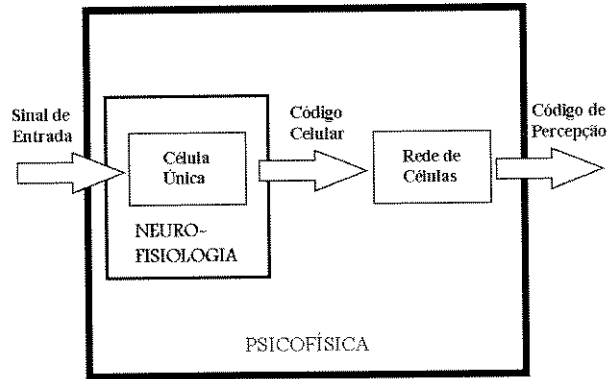


Figura 1.1: Esquema do relacionamento entre a Neurofisiologia e a Psicofísica.

nível e o segundo de alto nível. O estágio de baixo nível é responsável pelo processamento das imagens de entrada e suas representações e codificações em informações mais especializadas. O estágio de alto nível utiliza a codificação gerada pelo estágio de baixo nível na interpretação da imagem. A Figura 1.2 apresenta um esquema gráfico desse paradigma.

Segundo o paradigma de Levine, o processamento de baixo nível é dividido em três sub-fases: digitalização, realce e extração de descritores da imagem. A primeira sub-fase trata do problema de representação da imagem na máquina, ou seja, de sua digitalização, que é realizada através de uma amostragem espacial e de uma quantização das intensidades luminosas da imagem em intervalos limitados de tons de cinza ou cores. Na segunda sub-fase a imagem digitalizada é melhorada, tendo certos detalhes realçados, detalhes esses que serão utilizados na sub-fase seguinte, durante a segmentação

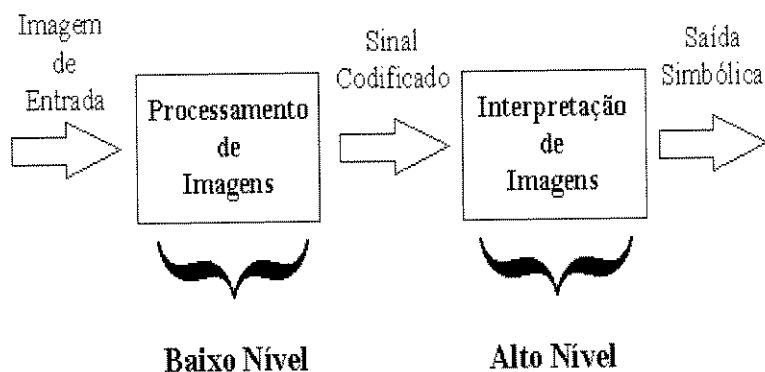


Figura 1.2: Paradigma de sistemas de visão computacional.

e a extração de descritores da imagem.

Ainda segundo esse paradigma, o estágio de alto-nível é dividido em várias sub-fases, que vão desde a classificação dos descritores extraídos da imagem, até a completa descrição da cena tridimensional presente no campo visual do SVC.

O paradigma proposto por Levine é inspirado nos sistemas biológicos de visão, nos quais foram encontrados módulos de processamento em diferentes níveis hierárquicos, sendo aqueles presentes nos níveis mais altos responsáveis por processamentos mais especializados e complexos do que aqueles dos níveis primários.

O paradigma de Levine para os SVCs é bastante simplificado e essa simplificação decorre do fato de não existir uma teoria completa de percepção. Acredita-se, por exemplo, que os níveis mais altos do processamento visual realimentem os níveis mais baixos, mas ainda não se sabe exatamente como

essa realimentação opera.

A ausência de uma teoria completa sobre a percepção visual, entretanto, não tem impedido que problemas mais modestos da área de Visão Computacional sejam resolvidos. Muitos desses problemas envolvendo classificações de padrões, por exemplo, têm sido tratados com certo êxito pelos SVCs [31].

A classificação de padrões consiste em associar um objeto físico, ou um evento, a uma determinada categoria, dentre diversas categorias pré-definidas.

O estudo extensivo de problemas de classificação tem conduzido ao desenvolvimento de modelos matemáticos abstratos que provêm as bases teóricas necessárias para o projeto de classificadores artificiais [31].

A Figura 4.2 apresenta o esquema de um sistema típico de classificação de padrões.

Esse sistema engloba completamente as três sub-fases do estágio de baixo nível e a primeira sub-fase do estágio de alto nível, descritos no paradigma de Levine para os SVCs. O sensor capta o estímulo da entrada e o digitaliza. Em seguida, o módulo extrator de descritores executa o melhoramento e o realce da imagem para depois extrair seus atributos. Finalmente, o módulo classificador utiliza os atributos extraídos para decidir a que classe, dentre um conjunto finito de classes, pertence o objeto observado.

1.2 Objetivos

O objetivo principal desta tese reside na extração de descritores adequados para classificação de imagens baseadas em informações de texturas orientadas. Texturas orientadas são formadas por padrões alongados, geralmente

produzidos por fenômenos naturais que ocorrem frequentemente na natureza. Esses padrões podem fornecer importantes informações sobre tais fenômenos.

A proposição desta tese consiste na análise de texturas orientadas através de informações obtidas dos segmentos de retas detectados nessas texturas. A motivação para tal proposição vem dos sistemas biológicos de visão, nos quais as informações sobre estímulos retilíneos com determinados comprimentos, orientações e frequências espaciais parecem ser extremamente relevantes [50, 51, 58, 59, 60, 70, 124].

Outro objetivo desta tese consiste na avaliação do desempenho dos descritores de texturas na estimação automática de densidades de multidões. A automação dessa tarefa é importante para a solução do problema de gerenciamento e controle de multidões, principalmente em situações rotineiras, como o monitoramento de estações de trens e metrô, aeroportos, etc, nas quais esse problema é crítico.

1.3 Estrutura da Tese

Além deste capítulo introdutório, que descreve a motivação e os objetivos do trabalho realizado, esta tese contém os seguintes capítulos:

Capítulo 2: apresenta o conceito de textura e as principais técnicas que vem sendo empregadas para a sua análise nos sistemas de visão computacional;

Capítulo 3: apresenta o conceito de textura orientada e propõe novas técnicas para a sua análise baseada em informações extraídas dos seg-

mentos de retas que a compõem;

Capítulo 4: descreve soluções para o problema do reconhecimento de padrões utilizando técnicas estatísticas, de redes neurais, e baseadas em funções de ajuste;

Capítulo 5: apresenta o problema do gerenciamento e controle de multidões; apresenta uma técnica existente para estimar automaticamente o número de pessoas numa cena, baseada nas informações dos pixels¹ das imagens; propõe novas estratégias para estimar densidades de multidões utilizando descritores de texturas;

Capítulo 6: descreve os resultados obtidos com os experimentos realizados com um conjunto de imagens reais para a estimação automática de densidades de multidões, utilizando medidas de texturas.

Capítulo 7: apresenta uma síntese da tese e conclusões sobre os resultados obtidos;

Capítulo 8: sugere extensões ao trabalho realizado.

¹Pixel: abreviação para “picture element”.

Capítulo 2

Análise de Textura

A visão, o mais poderoso dos cinco sentidos, permite ao ser humano perceber o mundo que o cerca, através da observação dos objetos físicos e das suas relações espaciais, desvendando o que são os objetos e determinando suas localizações precisas.

A percepção visual depende das propriedades dos objetos físicos, tais como cores, formas e textura. A propriedade textura descreve os relacionamentos dos elementos que se encontram próximos no espaço. A definição desses elementos e de seus relacionamentos, assim como o conceito de distância, dependem diretamente da escala de observação adotada. Dessa forma, um elemento primitivo de um arranjo espacial numa determinada escala torna-se um arranjo espacial numa escala mais fina. Os elementos primitivos são denominados *microtexturas*, enquanto que os arranjos desses elementos são denominados *macrotexturas* [127]. Veja na Figura 2.1 como microelementos dispostos inclinados à esquerda e à direita formam macroelementos (colunas)

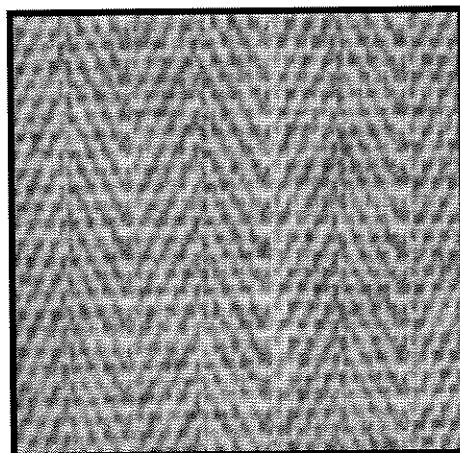


Figura 2.1: Microelementos numa escala constituem macroelementos numa escala maior (imagem extraída do álbum de texturas de Brodatz).

numa escala maior.

Quando se analisa a imagem discreta de uma textura, os menores elementos dessa imagem, os pixels, constituem as microtexturas mais simples.

Os arranjos espaciais mais complexos podem ser representados hierarquicamente através da redução recursiva das macrotexturas de uma dada escala de observação, que passam a ser consideradas microtexturas na escala seguinte. Texturas naturais costumam apresentar essa organização hierárquica, possuindo características bastante diversas nas diferentes resoluções [1].

Embora não haja definição formal para textura, algumas definições qualitativas podem ser encontradas na literatura.

Segundo Pickett [98], textura é usada para descrever arranjos bidimensionais de variações, onde os elementos e regras de posicionamento podem ser manipulados arbitrariamente, desde que se mantenha uma característica

de repetição.

Hawkins [46] descreve a noção de textura de forma mais detalhada, como sendo dependente de três ingredientes: (1) alguma “ordem” local é repetida numa região que é comparativamente maior que o tamanho dessa “ordem”; (2) essa “ordem” consiste de arranjos não aleatórios das partes elementares, e (3) tais partes são entidades quase uniformes apresentando aproximadamente as mesmas dimensões em quaisquer partes da região.

Para Gool [122], textura pode ser definida como uma estrutura constituída de um grande número de elementos ou padrões similares, mais ou menos ordenados, sem que nenhum desses se sobressaia dentre os demais, propiciando ao observador uma impressão global e unitária.

Para Haralick [44], uma textura é descrita pelo número, tipo e organização espacial de suas primitivas. A organização espacial pode ser aleatória ou pode ter dependências estruturais, probabilísticas ou funcionais mútuas entre um conjunto de primitivas numa vizinhança.

He [47] considera textura como sendo um termo utilizado para descrever a superfície de um dado fenômeno (o relacionamento espacial das intensidades entre pixels) numa imagem.

Estas descrições de textura parecem ser razoáveis do ponto de vista de percepção visual, contudo elas não conduzem imediatamente a medidas quantitativas.

A busca por medidas quantitativas efetivas de textura que traduzam suas propriedades qualitativas têm motivado diversas pesquisas sobre análise de textura na área de visão computacional.

2.1 Processamento de Textura no Sistema de Visão Humana

Há muitos anos, a visão humana tem sido estudada por psicólogos e neurofisiologistas [8, 50, 51, 56, 57, 58, 59, 60, 86, 87, 88, 94, 95, 114, 115, 116, 124]. Nesse período, muitas teorias empíricas têm sido propostas, refinadas e mesmo rejeitadas.

No campo da Psicofísica, as interações entre percepção e organização das imagens têm sido estudadas através de experimentos nos quais os parâmetros de estímulos visuais gerados artificialmente em monitores de televisão são sistematicamente manipulados.

Enquanto alguns aspectos do processamento visual humano têm gerado controvérsias, há um consenso com relação ao processamento de baixo nível, o qual acredita-se operar em dois modos, um serial e outro paralelo, através dos mecanismos de visão *atencional* (“attentive”) e *pré-atencional* (“pre-attentive”), respectivamente [61].

A Figura 2.2 exibe uma imagem que ilustra os conceitos de percepção atencional e pré-atencional.

Num relance é possível notar nessa imagem a área composta de cruces. A segregação dessa área é realizada sem esforço e quase instantaneamente, de forma pré-atencional. Por outro lado, a segregação da área composta por Ts rotacionados aleatoriamente, situado na parte direita superior da imagem, exige o esforço de uma inspeção pixel a pixel, de forma atencional.

Muitas teorias têm conseguido progressos para explicar a dicotomia visão

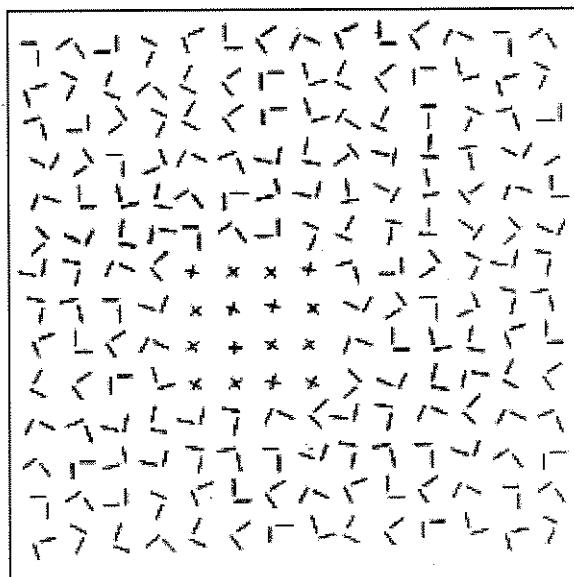


Figura 2.2: Ilustração da percepção de textura de forma atencional e pré-atencional.

atencional – visão pré-atencional. Modelos atuais propõem que a análise rudimentar pré-atencional é executada globalmente e seus resultados dirigem a subsequente análise seletiva e focalizada (atencional), marcando as posições onde existem diferenças nas estatísticas de primeira ordem de características primitivas.

Teorias de percepção podem ser divididas em dois grandes grupos: *seleção precoce* (“early selection”) e *seleção retardada* (“late selection”). Os modelos de seleção precoce sustentam que comparativamente pouco é realizado pelo estágio paralelo e que funções de resposta linear observadas para estímulos visuais mais complicados refletem o processamento sequencial atencional [114]. Nesse modelo, a percepção é associada apenas a uma pequena parte da cena a cada instante. A teoria da seleção retardada, ao contrário, argumenta que o estágio paralelo pré-atencional deveria ser denominado estágio de pré-consciência, uma vez que nele uma representação completa é construída. O gargalo serial acontece somente durante a troca de informações com o estágio consciente ou motor - é como se o sistema visual tivesse sido projetado isoladamente, com as interfaces com as demais funções sendo adicionadas posteriormente [32].

Acredita-se que o processamento pré-atencional seja dirigido a dados, sendo pouco influenciado pelos conhecimentos presentes nos níveis mais altos do processo de visão. Sua principal função seria formar uma representação da imagem nos níveis primitivos de abstração.

Algumas teorias têm sido desenvolvidas propondo formas para a representação pré-atencional, dentre as quais: a *Teoria dos Textons*, a *Teoria do “Primal Sketch”* e a *Teoria da Integração de Características*, que são

discutidas brevemente nas sub-seções seguintes.

2.1.1 Teoria dos Textons

Na década de 60, Julesz conjecturou que duas texturas com estatísticas de segunda ordem idênticas não poderiam ser distinguíveis pelo mecanismo de visão pré-atencional [56]. Estatísticas de segunda ordem dos tons de cinza de uma imagem são definidas como sendo as distribuições de probabilidades conjuntas de pares de tons de cinza ocorrerem simultaneamente em dois pixels da imagem, separados por uma distância d , numa orientação θ .

A Figura 2.3, apresenta duas imagens constituídas de pares de texturas que apresentam estatísticas de segunda ordem idênticas. Para estas duas imagens, a conjectura inicial de Julesz é verdadeira, uma vez que os pares de texturas diferentes que as compõem são indistinguíveis entre si.

Alguns anos mais tarde o próprio Julesz conseguiu gerar pares artificiais de texturas que possuíam estatísticas de segunda ordem idênticas e que eram distinguíveis pelo mecanismo de visão pré-atencional [58], contrariando, portanto, sua conjectura inicial. A Figura 2.4 apresenta imagens compostas por pares de texturas com estas propriedades.

Os pares de texturas apresentados na Figura 2.3 possuem texturas de segunda ordem idênticas, porém suas estatísticas de terceira ordem são distintas. Este fato levou a conclusão de que o SVH não se baseia em estatísticas de terceira ordem, ou de ordens superiores, para discriminar de forma pré-atencional pares de textura com estatísticas de segunda ordem idênticas. Estatísticas de terceira ordem dizem respeito às probabilidades de ocorrência

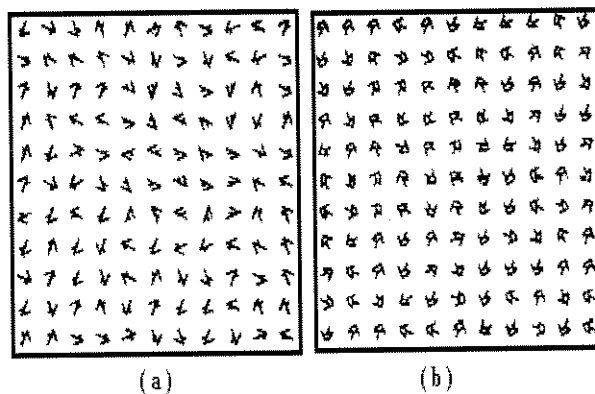


Figura 2.3: Imagens constituídas de pares de texturas que possuem estatísticas de segunda ordem idênticas. Os pares de texturas são indistinguíveis pelo mecanismo de visão pré-atencional.

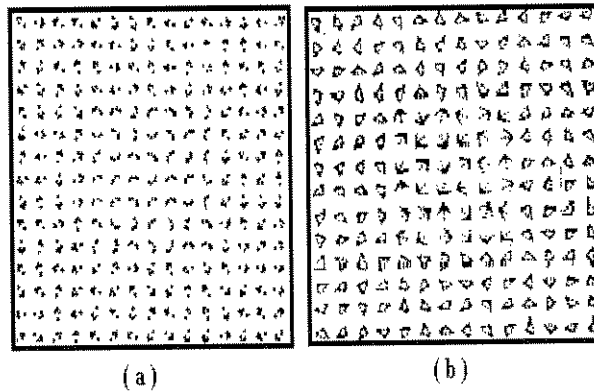


Figura 2.4: Imagens constituídas de pares de texturas que possuem estatísticas de segunda ordem idênticas. Os pares de texturas são distinguíveis pelo mecanismo de visão pré-atencional.

conjunta de um trio de tons de cinza em um trio de pixels, dispostos de forma triangular fixa sobre a imagem.

Após constatar os fatos acima, Julesz propôs a *Teoria dos Textons* [58, 59, 60], segundo a qual as informações visuais são representadas por um conjunto finito de primitivas de textura, os *textons*, que são detectados localmente, de forma pré-atencional, no campo visual por um sistema paralelo que computa as densidades locais desses textons. Diferenças nas densidades locais são registradas na representação pré-atencional, formando efetivamente uma segmentação primitiva da imagem.

A teoria dos textons pode ser resumida por duas heurísticas:

1. A visão humana opera em dois modos distintos: pré-atencional e atencional. No primeiro modo, o campo visual é processado instantaneamente, em paralelo, independentemente da complexidade da cena, sobre uma grande área. O modo atencional, ao contrário, é lento e sequencial, exigindo muito esforço e atuando numa pequena região do campo visual.
2. O sistema pré-atencional avalia a densidade espacial dos textons e dirige o modo de visão atencional para as posições do campo visual onde ocorrem diferenças nas densidades dos textons. Todas as demais funções são executadas pelo sistema de visão atencional.

Textons propostos e normalmente aceitos incluem: cor, orientação, largura, comprimento, extremidades e cruzamentos de linhas, disparidade binocular e disparidade de movimento.

Cada tipo de texton é processado independentemente por um sistema modular pré-atencional, que preserva apenas as informações de densidades dos textons. Até mesmo a informação de posição relativa dos textons do mesmo tipo são perdidas, a menos que algum alinhamento faça surgir uma nova propriedade de texton [5]. A representação pré-atencional na teoria dos textons constitui-se em um mapa de fronteiras de primitivas. Quando esse mapa fornece evidências suficientes para uma resposta, o comportamento paralelo pode ser observado. A teoria dos textons preocupa-se principalmente com as capacidades do sistema pré-atencional, não estabelecendo claramente o papel do sistema atencional.

2.1.2 Teoria do “Primal Sketch”

A teoria dos textons compartilha muitos componentes com a teoria do “primal sketch” de Marr [87], que propõe uma descrição simbólica da cena produzida por operações de filtragem de numa maneira dirigida a dados. Marr propôs primitivas parametrizadas similares àquelas aceitas como textons: linhas, bordas, barras, e “blobs”, escolhidas parcialmente por suas qualidades de invariância a projeções [86, 88].

Como na teoria dos textons, Marr propôs que a discriminação pré-atencional ocorre como resultado de diferenças de estatísticas de primeira ordem das características. A “raw primal sketch” é uma abstração primitiva da imagem e não pode prover acesso à informações de alto nível na forma em que estas são requeridas. Uma hierarquia de representações acima da raw primal sketch (cada uma num nível superior de abstração) transforma uma grande e com-

plicada quantidade de dados numa forma compacta e mais compreensível. Sucessivas abstrações são computadas a partir das representações mais primitivas sem acessar novamente a imagem de forma direta. O agrupamento de regiões adjacentes da imagem com características compatíveis é uma parte importante desse processo.

2.1.3 Teoria da Integração de Características

Segundo a teoria da integração de características proposta por Treisman [114, 116], a imagem formada na retina é processada inicialmente em paralelo por analisadores de percepção independentes, ao longo de dimensões separáveis de características. Dentre as propriedades da imagem extraídas nesse estágio encontram-se: orientação, cor, brilho, direção de movimento e contraste espacial, sendo cada atributo representado num espaço de características independente. Este primeiro estágio é automático, opera numa grande parte do campo visual e tem sido identificado como sendo pré-atencional. No estágio seguinte, as características primárias são combinadas e sintetizadas para que os objetos físicos sejam percebidos como formas unitárias. Esse segundo estágio requer atenção para que a identificação e combinação das características de um objeto sejam efetuadas. Esse processo é serial e opera local e lentamente sobre uma pequena área do campo visual.

Segundo a teoria de Treisman, a cada característica independente está associado um conjunto de mapas, um para cada possível valor dessa característica. Esses mapas preservam as informações espaciais, que são acessadas posteriormente pelo mecanismo serial.

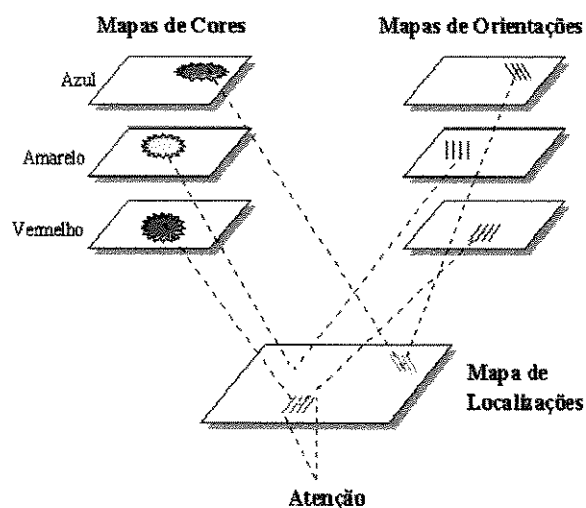


Figura 2.5: Esquema do modelo de integração de características.

Uma nova faceta do modelo de integração de características é a *pooled response*, através da qual uma medida global do nível de atividade dentro de cada espaço de características é obtida rapidamente sem a necessidade de se tratar cada posição do estímulo visual sequencialmente [115].

Veja na Figura 2.5 o esquema do modelo de integração de características relacionando os mapas de cada uma das várias dimensões independentes com o mapa geral de localizações, através do qual o mecanismo serial de atenção focalizada integra as propriedades de diferentes objetos da cena.

Em contraste com a teoria de Julesz, que inclui no conjunto dos textons conjunções de valores de dimensões diferentes, como por exemplo um “blob” alongado e vermelho num fundo azul, a teoria proposta por Treisman sugere que várias representações diferentes podem ser estabelecidas de forma pré-atencional, separadamente, dentro de cada dimensão. As conjunções

dessas características seriam realizadas num estágio posterior por meio de um mecanismo que opera com atenção focalizada.

2.1.4 Aspectos Neurais da Percepção de Textura

Assim como os experimentos psicofísicos, os experimentos neurofisiológicos geralmente utilizam estímulos visuais artificiais que são exibidos num monitor e que podem ser controlados convenientemente segundo um conjunto de parâmetros. Os experimentos neurofisiológicos realizados nos sistemas visuais de primatas têm sido geralmente extrapolados para o sistema visual humano devido às suas semelhanças.

Existem relativamente poucos estudos neurofisiológicos realizados com estímulos texturais mais complexos [42, 94, 95]. A maioria dos experimentos utiliza estímulos compostos por barras e “gratings¹” para avaliar os mecanismos básicos de codificação neural.

Um desses poucos estudos foi desenvolvido por Gallant et al [37], onde foram examinados os mecanismos neurais que suportam os fenômenos clássicos de percepção de textura, tais como segregação instantânea realizada de forma pré-atencional, e a percepção de cenas naturais e formas tridimensionais, utilizando textura como ferramenta. Os neurônios estudados foram coletados nas áreas V1 e V4 do córtex visual de macacos. A área V1 é a primeira grande região cortical de processamento de informações visuais, onde os neurônios apresentam campos receptivos² pequenos e respostas lineares aos estímulos

¹Estímulos retilíneos paralelos, periodicamente dispostos a uma mesma distância uns dos outros.

²O termo *campo receptivo* de uma célula do processamento visual refere-se aos re-

que se situam dentro do campo receptivo (alguns neurônios dessa região apresentam não-linearidades relativamente simples). A área V4 é uma região cortical de processamento visual entre a área V1 e o complexo inferotemporal. Os campos receptivos dos neurônios dessa área são muitas vezes maiores do que os da área V1. Tanto as células da área V1 quanto as da área V4 são seletivas às dimensões visuais, tais como: disparidade da retina, cor, orientação e frequência. As células da área V4 são ainda seletivas a estímulos mais complexos.

Os experimentos de Gallant et al [37] mostraram que as respostas de algumas células aumentam significativamente quando bordas de texturas encontram-se próximas aos seus campos receptivos, em contraste com as respostas apresentadas por essas mesmas células quando as texturas são homogêneas nos seus campos receptivos. A maioria dessas células parecem ter uma preferência por determinadas orientações de bordas de texturas, sendo que para uma significativa população dessas células, a orientação é uma informação irrelevante. Embora as respostas obtidas próximas às bordas de texturas sejam maiores do que as obtidas em texturas homogêneas, essas respostas são geralmente menores do que as respostas apresentadas para uma única barra situada no campo receptivo. Esse fato confirma que estímulos fora do campo receptivo têm, em geral, efeito supressivo [41, 66].

Os experimentos de Gallant et al [37] sugerem ainda que as respostas dos neurônios da área V4 a estímulos compostos por superfícies tridimensionais texturizadas parecem estar relacionadas com sua orientação e sua frequência

ceptores (cones e bastonetes) da retina que alimentam essa célula e a forma como esses receptores precisam ser estimulados para que a célula responda a tais estímulos.

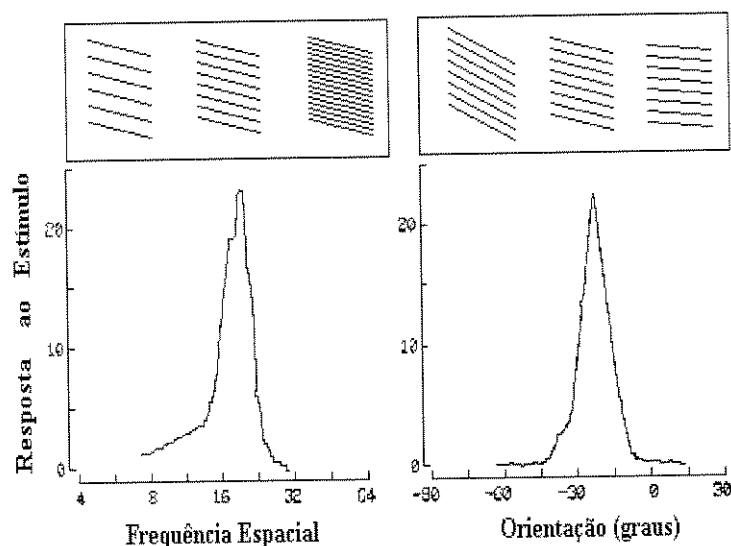


Figura 2.6: Respostas de uma “grating cell” para as variações nas frequências espaciais e nas orientações dos estímulos.

espacial.

Orientação e frequência espacial também foram considerados dois fatores importantes no estudo realizado por Heydt [124] sobre um tipo de células as quais ele denominou “grating cells”. Essas células são seletivas a estímulos periódicos formados por barras, com orientação e período específicos. A Figura 2.6 mostra as respostas de uma “grating cell” para um estímulo do tipo “grating”. Observe que os intervalos de sintonização de orientação e frequência espacial são bastante finos.

Hubel e Wiesel [50, 51, 52] pioneiramente encontraram na área visual primária (área 17 do córtex visual) e na área visual secundária (áreas 18 e 19 do córtex visual) tipos de células que são seletivas à orientação, direção

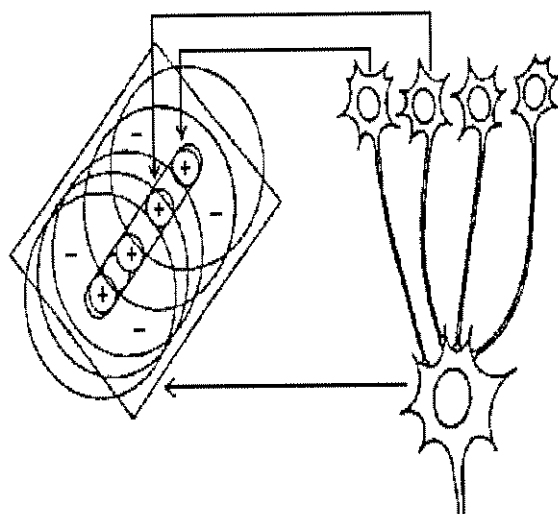


Figura 2.7: Campo receptivo das células simples.

de movimento, tamanhos e extremidades dos estímulos retilíneos.

A Figura 2.7 apresenta um possível esquema neural que produz o campo receptivo de uma célula simples (encontrada na área 17). O campo receptivo desse tipo de célula faz com que ela apresente as mais fortes respostas a estímulos retilíneos com determinada orientação [70].

A Figura 2.8 apresenta um possível esquema neural que produz o campo receptivo de uma célula complexa (encontrada nas áreas 17 e 18). O campo receptivo desse tipo de célula faz com que ela apresente as mais fortes respostas a estímulos retilíneos com determinada orientação movendo-se em determinada direção [70].

A Figura 2.9 apresenta um possível esquema neural que produz o campo receptivo de uma célula hiper-complexa (encontrada nas áreas 17, 18 e 19). O

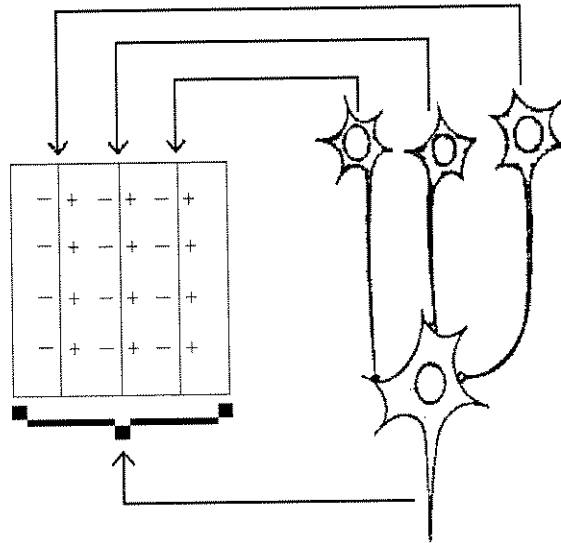


Figura 2.8: Campo receptivo das células complexas.

campo receptivo desse tipo de célula faz com que ela apresente as mais fortes respostas a estímulos retilíneos com determinada orientação, movendo-se em determinada direção e possuindo um tamanho específico [70].

A existência dessas células nos níveis primários do córtex visual sugere a importância dos estímulos retilíneos, assim como seus atributos de orientação e comprimento para o processamento visual como um todo e o processamento de textura, em particular.

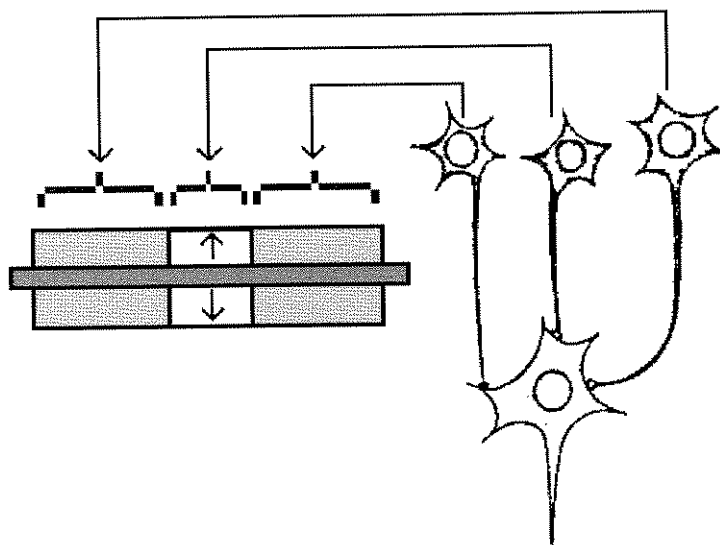


Figura 2.9: Campo receptivo das células hiper-complexas.

2.2 Processamento de Textura nos Sistemas de Visão Computacional

O fato de não existir uma definição precisa e formal para textura, dificulta a definição do conceito de *descriptor de textura*, que tem sido definido intuitivamente, assim como o próprio conceito de textura.

Várias técnicas computacionais têm sido propostas na literatura para se extrair os atributos de textura de uma imagem. Tradicionalmente, essas técnicas têm sido categorizadas como sendo *estruturais* ou *estatísticas* [44, 70, 122]. Recentemente, uma categoria adicional de texturas *baseadas em modelos* vem sendo proposta [4, 113].

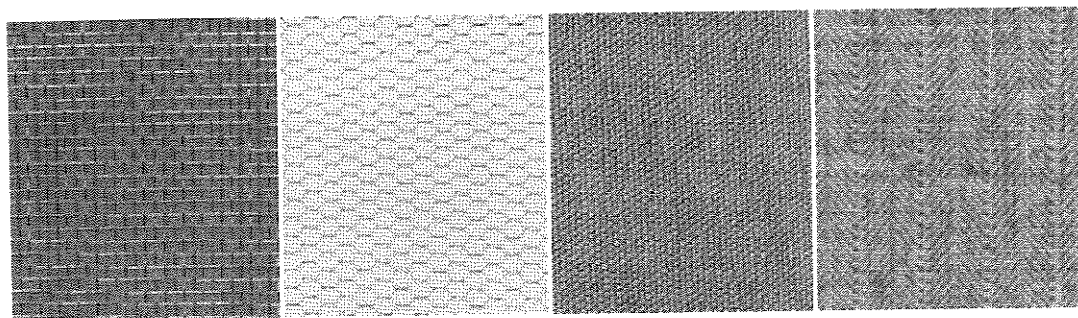


Figura 2.10: Exemplos de texturas determinísticas (imagens extraídas do álbum de texturas de Brodatz).

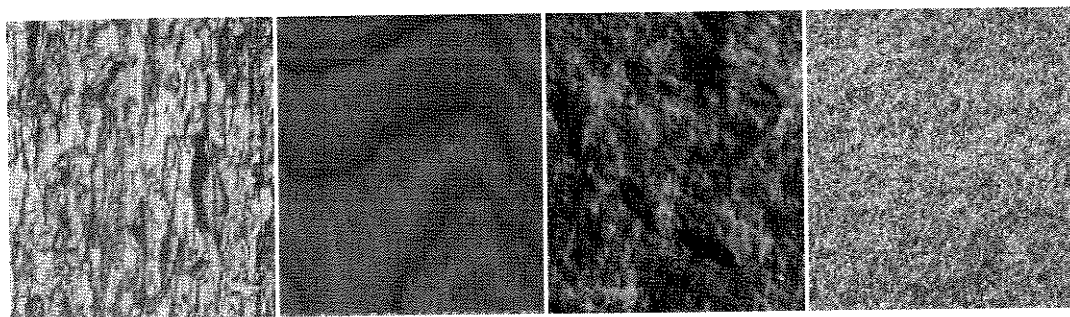


Figura 2.11: Exemplos de texturas estocásticas (imagens extraídas do álbum de texturas de Brodatz).

As técnicas *estruturais* tratam as texturas determinísticas, que podem ser consideradas como uma repetição de primitivas componentes. Essa repetição obedece à determinadas regras de posicionamento. Veja exemplos de texturas determinísticas na Figura 2.10.

As técnicas *estatísticas*, por sua vez, são mais gerais e objetivam a caracterização de propriedades estocásticas da distribuição espacial dos níveis de cinza numa imagem. Veja exemplos de texturas estocásticas na Figura 2.11.

As técnicas *baseadas em modelos* foram, em geral, desenvolvidas para a síntese de texturas e, posteriormente, adaptadas para a análise. Tais modelos assumem, em geral, certas dependências estatísticas nas vizinhanças dos pixels das imagens [4].

A categorização das técnicas de análise de textura em estruturais, estatísticas ou baseadas em modelos nem sempre pode ser feita de forma simples e clara. Entretanto, essa categorização ajuda a estruturar de forma lógica a apresentação das mesmas.

2.2.1 Abordagens Estatísticas

Texturas nessa abordagem são consideradas como sendo caracterizadas por um conjunto de estatísticas, que podem variar de um simples histograma de primeira ordem dos tons de cinza de uma imagem, até sofisticadas estatísticas de ordem superior, envolvendo elementos da imagem, como as bordas, por exemplo.

2.2.1.1 Matrizes de Dependência Espacial dos Tons de Cinza

Estudos iniciais realizados por Julesz [56, 57] sobre a percepção visual humana levaram-no a conjecturar que o mecanismo de visão humana instantânea (“pre-attentive”) dependia unicamente de estatísticas de primeira e segunda ordens. Tal conjectura motivou o desenvolvimento de técnicas baseadas nessas estatísticas, dentre elas, a técnica baseada na dependência espacial dos tons de cinza, proposta por Haralick [44].

Esta técnica diz respeito à distribuição e à dependência espacial dos níveis

de cinza da imagem e baseia-se na estimativa das funções de densidade de probabilidade condicional de segunda ordem $f(i, j, d, \theta)$. Essas funções são as probabilidades de ocorrência de um par (i, j) de tons de cinza em dois pixels da imagem separados por uma distância d , na direção definida por um ângulo θ . Os valores estimados podem ser escritos na forma matricial, as chamadas matrizes de co-ocorrência dos tons de cinza.

Por exemplo, suponha a imagem de dimensão 5x5 pixels, com tons de cinza variando entre 0 e 3, apresentada na Figura 2.12(a). Para calcular as matrizes de co-ocorrências dos tons de cinza dessa imagem, deve-se inicialmente fixar a distância d e o ângulo θ . Em seguida, com os valores d e θ fixados, calcula-se as ocorrências na imagem de cada par (t_i, t_j) separados por d pixels, na direção θ . No exemplo sendo considerado, $t_i, t_j = 0, \dots, 3$. A Figura 2.12(b) apresenta a matriz de dependência espacial dos tons de cinza da imagem da Figura 2.12(a), para os valores $d = 1$ e $\theta = 45^\circ$.

As matrizes de co-ocorrência têm dimensões dadas pelo número de tons de cinza N_g possíveis de ocorrerem numa imagem. Numa imagem que utiliza 1 byte para representar a informação de cada pixel, podem ocorrer 256 tons distintos de cinza, resultando numa matriz de co-ocorrência de dimensão 256x256. Essa dimensão pode ser reduzida para se aumentar a eficiência computacional da extração das características de textura. Além disso, a redução das dimensões da matriz aumenta a significância estatística de cada entrada.

A partir das matrizes de co-ocorrência definidas por uma distância d e um ângulo θ , pode-se computar medidas para se caracterizar texturas. Haralick [43] e Pressman [101] propuseram as medidas descritas abaixo.

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

(a)

$\frac{1}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$
$\frac{3}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$
$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$
$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{0}{16}$

(b)

Figura 2.12: Exemplo de matriz de dependência dos tons de cinza. (a) Imagem com tons de cinza variando 0 a 3; (b) Matriz de dependência dos tons de cinza calculados fixando-se os parâmetros $d = 1$ pixel e $\theta = 45^\circ$.

Considere as seguintes notações:

N_g : número de tons de cinza distintos da imagem;

$p(i, j)$: (i, j) -ésima entrada da matriz de co-ocorrência normalizada $f(i, j|d, \theta)/R$ (R é o fator de normalização);

$p_x(i)$: i -ésima entrada da matriz de probabilidade marginal obtida somando-se as linhas da matriz de co-ocorrência $P(i, j)$
 $(p_x(i) = \sum_{j=1}^{N_g} P(i, j));$

$p_y(j)$: j -ésima entrada da matriz de probabilidade marginal obtida somando-se as colunas da matriz de co-ocorrência $P(i, j)$
 $(p_y(j) = \sum_{i=1}^{N_g} P(i, j));$

$p_{x+y}(k = i + j) = \sum \sum_{i=1}^{N_g} p(i, j)$, onde $k = 2, 3, \dots, 2N_g$;

$p_{x-y}(k = |i - j|) = \sum \sum_{i=1}^{N_g} p(i, j)$, onde $k = 0, 1, \dots, N_g - 1$;

μ_x : média de p_x ;

μ_y : média de p_y ;

μ : média de p ;

σ_x : desvio-padrão de p_x ;

σ_y : desvio-padrão de p_y .

Medidas de Texturas:

Segundo Momento Angular ou Energia:

$$f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)^2 \quad (2.1)$$

Correlação:

$$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2.2)$$

Momento Diferença (Contraste):

$$f_3 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j) \quad (2.3)$$

Momento Diferença Inversa (Homogeneidade):

$$f_4 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 + (i - j)^2} \quad (2.4)$$

Entropia:

$$f_5 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j)) \quad (2.5)$$

Entropia-Soma:

$$f_6 = \sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i)) \quad (2.6)$$

Média-Soma:

$$f_7 = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad (2.7)$$

Variância-Soma:

$$f_8 = \sum_{i=2}^{2N_g} (i - f_7)^2 p_{x+y}(i) \quad (2.8)$$

Entropia-Diferença:

$$f_9 = - \sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i)) \quad (2.9)$$

Variância-Diferença:

$$f_{10} = \sum_{i=0}^{N_g-1} (i - f_9)^2 p_{x-y}(i) \quad (2.10)$$

Coefficiente de Variação:

$$f_{11} = \frac{\sigma_x \sigma_y}{\mu_x \mu_y} \quad (2.11)$$

Momento-Diagonal:

$$f_{12} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \sqrt{p(i, j) |i - j|} \quad (2.12)$$

Soma dos quadrados ou variância:

$$f_{13} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (i - \mu)^2 \quad (2.13)$$

Variância-Diagonal:

$$f_{14} = \sum_{i=1}^{N_g} p(i, j)(i - \mu)^2 \quad (2.14)$$

Probabilidade de Transição de Pico:

$$f_{15} = \text{Max}(p(i, j)) \quad (2.15)$$

Como o cálculo das matrizes de co-ocorrência para um número grande de direções θ e distâncias d é computacionalmente inviável, na prática, apenas quatro direções (0° , 45° , 90° e 135°) e uma distância (1 pixel) são comumente adotadas.

Embora Haralick [44] tenha sugerido a existência de “constraints” naturais entre as probabilidades de co-ocorrências entre diferentes distâncias espaciais, a habilidade de se discriminar algumas texturas é criticamente dependente da escolha da distância d [92]. A escolha de distâncias pequenas justifica-se pelo fato de que os pixels separados por grandes distâncias tendem a apresentar fraca correlação entre seus níveis de cinza, resultando em estatísticas ruidosas [126].

Infelizmente, as medidas de textura propostas não capturam todas as estruturas significantes [18, 92]. Além disso, elas não apresentam correlações óbvias com os aspectos de percepção de textura [110].

Apesar dessas restrições, técnicas para análise de textura baseadas em matrizes de dependência dos tons de cinza têm mostrado bons resultados na discriminação de texturas naturais [18, 126].

Resultados obtidos com a aplicação dessa técnica no problema de estimação de densidades de congestionamentos de pessoas estão descritos nas

seções 6.1.1 e 6.2.1.

2.2.1.2 Matriz de Dependência Espacial Generalizada

Uma abordagem estatística para análise de textura deveria idealmente considerar a partição de uma imagem em células e a distribuição dos tons de cinza entre essas células [1]. O método baseado nas dependências espaciais dos tons de cinza dedicam-se principalmente ao segundo aspecto. Davis [24] propôs uma generalização a esse método visando dar-lhe a habilidade de considerar aspectos espaciais da imagem.

A fim de se calcular a matriz de dependência generalizada é preciso especificar-se um protótipo P da imagem, um predicado espacial S e um atributo A do protótipo. P é a primitiva a ser utilizada (intensidade dos tons de cinza, bordas, retas, etc). O predicado espacial S é a condição imposta sobre as posições espaciais dos pixels sendo considerados. Finalmente, A significa a medida do atributo que será utilizada pelo método.

Por exemplo, P poderia ser um pixel de uma borda, S a distância máxima e A a orientação. O acumulador $(45,90)$ da matriz de dependência generalizada seria incrementado de 1 cada vez que um pixel pertencente a uma borda com orientação 45° que tivesse um pixel vizinho pertencente a uma borda com orientação de 90° com a distância sendo especificada por S . Similarmente ao método baseado nas dependências dos tons de cinza, um conjunto de características pode ser computado a partir de cada matriz de dependência generalizada e utilizado para análise de textura.

2.2.1.3 Diferença dos Tons de Cinza

Seja $g(n, m)$ uma imagem. Para um deslocamento $d = (M, N)$ (M e N inteiros), seja $g_d(m, n) = |g(m, n) - g(m + M, n + N)|$. Se a função de distribuição de probabilidades $f'(i, d) = P(g_d(m, n) = i)$ é aplicada, algumas características de texturas podem ser derivadas. Note que $f'(i, d)$ pode ser obtida a partir da matriz de co-ocorrência dos tons de cinza somando-se as entradas dessa matriz ao longo dos eixos paralelos e simétricos com relação à diagonal principal.

Para prevenir a dependência angular da matriz de co-ocorrência, pode-se utilizar o método que calcula a matriz de co-ocorrência dos tons de cinza considerando-se todos os elementos pertencentes à vizinhança simultaneamente no cálculo da matriz [109]. Esse método considera um elemento e todos os seus vizinhos, ao invés de considerar apenas uma direção por vez. Essa nova matriz é indexada por k e s . O valor s de um pixel é definido como o número de vezes que um pixel vizinho tem um tom de cinza diferença i do tom de cinza k do pixel central de uma vizinhança $d \times d$. Para uma vizinhança 3×3 , por exemplo, s pode variar entre 0 e 8. Cada entrada (k, s) da matriz representa o número de ocorrências do par (k, s) . Características de textura podem então ser definidas a partir dessa nova matriz de modo análogo àquele empregado nas matrizes de co-ocorrência dos tons de cinza.

2.2.1.4 Matrizes de Co-ocorrência de Comprimentos de Cadeias de Tons de Cinza

Uma cadeia $c(i)$ de tons de cinza é definida como sendo um conjunto consecutivo de pixels colineares da imagem contendo o mesmo tom de cinza i . O comprimento $comp(c(i))$ da cadeia $c(i)$ é o número de pixels que compõem essa cadeia.

Dada uma figura, pode-se calcular uma matriz M_θ das co-ocorrências dos comprimentos das cadeias $c(i)$ na direção definida por θ . Cada elemento $M_\theta(i, j)$ especifica o número de ocorrências de cadeias $c(i)$ com comprimento $comp(c(i)) = j$ na imagem, na direção definida por θ .

Além das medidas propostas para se obter características de textura a partir das matrizes de dependência dos níveis de cinza, as seguintes medidas, propostas por Galloway [38] e Loh [75], podem também ser aplicadas nas matrizes de co-ocorrências dos comprimentos das cadeias de tons de cinza:

Considere as seguintes notações:

$m(i, j)$: (i, j) -ésima entrada da matriz normalizada de co-ocorrência

$M_\theta(i, j)$;

N_g : número máximo de tons de cinza possíveis de ocorrer na imagem;

N_r : número máximo de comprimentos de cadeias $c(i)$ possíveis de ocorrer na imagem;

P : número de pixels na imagem;

$F = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} M_\theta(i, j)$ é o número de cadeias presentes numa imagem (utilizado como fator de normalização).

Medidas de Texturas:

Ênfase das cadeias curtas:

$$f_1 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{m(i, j)}{j^2}}{F} \quad (2.16)$$

Ênfase das cadeias longas:

$$f_2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} m(i, j) \cdot j^2}{F} \quad (2.17)$$

Ênfase à não-uniformidade dos tons de cinza das cadeias:

$$f_3 = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} m(i, j))^2}{F^2} \quad (2.18)$$

Ênfase à não-uniformidade dos comprimentos das cadeias:

$$f_4 = \frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} m(i, j))^2}{F^2} \quad (2.19)$$

Ênfase às não-linearidades:

$$f_5 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} m(i, j)}{P} \quad (2.20)$$

Como o método para calcular as matrizes de dependência dos tons de cinza, o método baseado no comprimento das cadeias com mesmos níveis de cinza mede características de textura considerando o relacionamento espacial entre as cadeias $c(i)$ [38].

2.2.1.5 Análise de Fourier

Como o grau de suavidade de uma textura é proporcional ao seu período espacial, uma região suave apresenta suas energias espectrais concentradas nas baixas frequências. Por outro lado, regiões apresentando texturas finas exibem uma concentração de energias espectrais nas altas frequências [69].

O espectro de Fourier pode ser expresso em coordenadas polares como $S(r, \theta)$, onde S é a função espectro, e r e θ são as variáveis nesse sistema de coordenadas. Para um valor fixo de θ , $S(r, \theta)$ pode ser considerado como uma função unidimensional $S_\theta(r)$, que fornece o comportamento radial do espectro visto ao longo da direção dada por θ . Analogamente, para um valor fixo de r , $S(r, \theta)$ pode ser considerado como sendo uma função unidimensional que fornece o comportamento do circular do espectro visto ao longo da circunferência de raio r centralizado na origem.

Descritores globais de textura podem ser obtidos a partir do espectro de Fourier, integrando-se tais funções unidimensionais [40].

No caso discreto, esses descritores são:

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \quad (2.21)$$

e

$$S(\theta) = \sum_{r=1}^R S_r(\theta) \quad (2.22)$$

onde R é o raio do círculo centralizado na origem. Para um espectro de dimensão $N \times N$, R é comumente escolhido como sendo $N/2$.

Resultados obtidos com a aplicação dessa técnica no problema de estimação de densidades de congestionamentos de pessoas estão descritos na seção 6.2.3.

2.2.1.6 Autocorrelação

Um aspecto importante na análise de uma textura diz respeito aos tamanhos das primitivas que a compõem. Primitivas grandes indicam texturas mais suaves, ao passo que primitivas pequenas estão associadas a texturas mais ásperas. A função de autocorrelação provê informação sobre os tamanhos das primitivas [119]. Fixado um deslocamento (m, n) , as regiões contendo primitivas grandes exibem correlações maiores do que aquelas constituídas por primitivas pequenas.

Para uma imagem discreta $f(j, k)$, a função de autocorrelação é definida como:

$$A_f(m, n) = \sum_j \sum_k f(j, k) f(j - m, k - n) \quad (2.23)$$

calculada sobre uma janela $W \times W$, com $-T \leq m, n \leq T$.

A aspereza de uma textura tende a ser proporcional ao espalhamento da função de autocorrelação. Faugeras e Pratt [35] propuseram o seguinte conjunto de medidas de espalhamento da função de autocorrelação:

$$S(u, v) = \sum_{m=0}^T \sum_{n=-T}^T (m - \eta_m)^u (n - \eta_n)^v A_f(m, n) \quad (2.24)$$

onde

$$\eta_m = \sum_{m=0}^T \sum_{n=-T}^T m A_f(m, n) \quad (2.25)$$

$$\eta_n = \sum_{m=0}^T \sum_{n=-T}^T n A_f(m, n) \quad (2.26)$$

Apenas metade da função de autocorrelação é utilizada no cálculo das medidas de espalhamento, devido à sua simetria. As medidas de interesse

incluem particularmente os perfis do espalhamento, $S(2, 0)$ e $S(0, 2)$, a relação cruzada $S(1, 1)$ e o espalhamento de segundo grau $S(2, 2)$.

2.2.1.7 Filtros de Laws

Textura é uma propriedade local das imagens e alguns tipos de operadores locais são intuitivamente úteis para a extração de medidas de textura. O passo básico para se analisar textura a partir de propriedades locais de imagens são a determinação dos tamanhos das máscaras de convolução, a escolha dos pesos dessas máscaras e a definição das medidas de textura. Laws foi um dos pioneiros dessa técnica [3, 68, 100], propondo empiricamente nove máscaras 3x3 para serem convoluídas linearmente com a imagem a fim de realçar algumas de suas estruturas intrínsecas. Essas máscaras ficaram conhecidas como filtros de Laws e estão apresentadas na Figura 2.13.

Esses filtros são convolucionados com a imagem a fim de se acentuar suas microestruturas. Dessa forma, a i -ésima microestrutura é dada por:

$$M_i(j, k) = F(j, k) * H_i(j, k) \quad (2.27)$$

onde $H_i(j, k)$ é um dos filtros de Laws.

A seguir, cada uma das nove imagens resultantes são transformadas numa imagem energia $T_i(j, k)$ através do cálculo da seguinte fórmula, que mede o desvio padrão numa vizinhança $W \times W$:

$$T_i(j, k) = \frac{1}{W^2} \left\{ \sum_{m=-W}^W \sum_{n=-W}^W \left(M_i(j+m, k+n) - \overline{M}(j+m, k+n) \right)^2 \right\}^{1/2} \quad (2.28)$$

$$\frac{1}{36} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

(a)

$$\frac{1}{12} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

(b)

$$\frac{1}{12} \begin{bmatrix} -1 & 2 & -1 \\ -2 & 4 & -2 \\ -1 & 2 & -1 \end{bmatrix}$$

(c)

$$\frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

(d)

$$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

(e)

$$\frac{1}{4} \begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

(f)

$$\frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ 2 & 4 & 2 \\ -1 & -2 & -1 \end{bmatrix}$$

(g)

$$\frac{1}{4} \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{bmatrix}$$

(h)

$$\frac{1}{4} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

(i)

Figura 2.13: Banco de filtros de Laws.

onde W é a dimensão da janela sendo considerada para definir a vizinhança dos pixels $F(j, k)$ e $\overline{M}(j, k)$ é a amplitude da imagem média.

2.2.1.8 Filtros de Gabor

Técnicas para análise de textura que utilizam filtros de Gabor pertencem à classe das recentes e bastante promissoras técnicas baseadas em filtragens espaciais multicanais. Os principais tópicos envolvidos na abordagem de filtragem multicanais para análise de textura são: a caracterização funcional dos canais e o número de canais, a extração de características de texturas das imagens filtradas, o relacionamento entre os canais e a integração das diferentes características de textura obtidas para produzir uma segmentação.

Diferentes técnicas de filtragem multicanal têm sido propostas, dentre elas as baseadas nos filtros de Gabor [9, 33, 55, 97, 111, 117].

Filtros de Gabor são originados da Teoria de Comunicação proposta por Gabor [36]. Um filtro ou função bidimensional de Gabor tem a seguinte forma geral:

$$h(x, y) = g(x, y)e^{-2\pi j(u_0x + v_0y)} \quad (2.29)$$

onde $g(x, y)$ é uma função Gaussiana bidimensional definida como:

$$g(x, y) = e^{-\frac{1}{2} \left[\left[\frac{x}{\sigma_x} \right]^2 + \left[\frac{y}{\sigma_y} \right]^2 \right]} \quad (2.30)$$

O filtro de Gabor $h(x, y)$ especificado pela equação 2.29 é uma grade senoidal complexa modulada por uma Gaussiana bidimensional definida em 2.30. As constantes espaciais σ_x e σ_y determinam a escala nas direções x e y , respectivamente. A razão $\lambda = \frac{\sigma_x}{\sigma_y}$ pode ser utilizada para impor o grau de

sensitividade à orientação do filtro. A frequência total de modulação ω_0 do filtro $h(x, y)$ é $\omega_0 = (u_0^2 + v_0^2)^{1/2}$. O eixo de modulação é orientado segundo o ângulo $\theta_0 = \tan^{-1} \left[\frac{v_0}{u_0} \right]$ com relação ao eixo x.

Se ω_0 e θ_0 forem utilizados como parâmetros do filtro, ao invés de u_0 e v_0 , então o filtro de Gabor $h(x, y)$ definido em 2.29 pode ser expresso de forma equivalente:

$$h(x, y) = g(x, y)e^{-2\pi\omega_0 j[x\cos\theta_0 + y\sin\theta_0]} \quad (2.31)$$

onde $g(x, y)$ é definido em 2.30.

No domínio da frequência, o filtro de Gabor definido em 2.29 é dado por:

$$H(u, v) = 2\pi\sigma_x\sigma_y e^{-2\pi^2[\sigma_x^2(u+u_0)^2 + \sigma_y^2(v+v_0)^2]} = G(u + u_0, v + v_0) \quad (2.32)$$

onde $G(u, v)$ é a transformada de Fourier de $g(x, y)$, dada por:

$$G(u, v) = 2\pi\sigma_x\sigma_y e^{-2\pi^2(\sigma_x^2 u^2 + \sigma_y^2 v^2)} \quad (2.33)$$

que também é uma Gaussiana bidimensional localizada na origem do sistema de coordenadas do espaço de frequências.

$H(u, v)$ também é uma Gaussiana centralizada nas coordenadas de frequência $(-u_0, -v_0)$, cuja razão $\lambda' = \frac{\frac{1}{\sigma_x}}{\frac{1}{\sigma_y}} = \frac{1}{\lambda}$, ou seja, a razão inversa do filtro no domínio da imagem. A distância do centro de $H(u, v)$ até a origem é $(u_0^2 + v_0^2)^{1/2} = \omega_0$, ou seja, a frequência de modulação total de $h(x, y)$ no domínio espacial da imagem. A orientação de $H(u, v)$ é dada em relação ao eixo x pelo mesmo ângulo do eixo de modulação do domínio espacial da imagem. Os parâmetros independentes σ_x , σ_y , u_0 e v_0 definem o filtro de Gabor dado pela equação 2.29. Eles descrevem unicamente o filtro de Gabor

e determinam as larguras das bandas de frequência espacial e orientação do filtro.

O filtro dado em (2.29) é uma função complexa. Para aplicações práticas, contudo, o filtro de Gabor pode ser definido como uma função real da seguinte forma:

$$h(x, y) = g(x, y) \operatorname{sen}[2\pi(u_0x + v_0y) + \rho] \quad (2.34)$$

onde $g(x, y)$ é a Gaussiana bidimensional definida em (2.30) e ρ é a fase da senóide de modulação.

Geralmente, atribui-se a ρ os valores $\frac{\pi}{2}$ ou 0 , resultando os respectivos filtros:

$$h_c(x, y) = g(x, y) \cos[2\pi(u_0x + v_0y)] \quad (\rho = \frac{\pi}{2}) \quad (2.35)$$

ou

$$h_s(x, y) = g(x, y) \operatorname{sen}[2\pi(u_0x + v_0y)] \quad (\rho = 0) \quad (2.36)$$

Como $h_c(x, y) = h_c(-x, -y)$ e $h_s(x, y) = -h_s(-x, -y)$, $h_c(x, y)$ é denominada de filtro de Gabor simétrico e $h_s(x, y)$ como filtro de Gabor anti-simétrico. Pode-se mostrar que $h(x, y) = h_c(x, y) - jh_s(x, y)$.

Vários estudos têm mostrado que as funções de Gabor descrevem as funções resposta dos campos receptivos das células simples do córtex visual de alguns mamíferos [9, 97, 111, 117].

A técnica para segmentação não-supervisionada de imagens baseadas na segmentação de texturas proposta por Jain [55] utiliza apenas os filtros simétricos de Gabor. Malik e Perona justificam, baseados em resultados da Psicofísica, a utilização apenas dos filtros simétricos de Gabor [77].

Assim como a técnica baseada nos filtros de Laws, técnicas baseadas em filtros de Gabor consideram os aspectos locais de textura. A extração das características de textura das imagens filtradas podem ser realizadas, calculando-se, assim como na técnica de Laws, imagens energia $T(j, k)$.

2.2.1.9 Wavelets

A teoria das *wavelets* possibilita uma análise precisa e unificada de sinais em diferentes escalas [17, 25, 48, 76]. Carter [12] foi o primeiro a relatar resultados de classificação de textura utilizando wavelets. No entanto, as wavelets por ele utilizadas não eram ortonormais e nem eram sensíveis a orientações (uma das wavelets utilizadas foi o “chapéu mexicano”). O conjunto de filtros de Gabor formam uma base aproximada para a transformação de wavelets, com a função de Gabor sendo a wavelet. As funções de Gabor (veja seção 2.2.1) permitem a sintonia de orientações e formam uma base completa, porém não ortogonal, podendo resultar em características de textura correlacionadas. Além disso, decomposições de imagens através de funções de Gabor são computacionalmente caras, especialmente para a avaliação de componentes de baixa frequência [120]. Tais problemas podem ser evitados utilizando-se transformações de wavelets.

A utilização da transformada de wavelets para a análise de textura, apresentada por Mallat [76], propõe a utilização de pirâmides de wavelets para a decomposição das imagens. Essa proposta inicial vem sendo seguida por diversos estudos sobre classificação de textura, com atenção especial sendo dada ao uso de pacotes de wavelets [67], que constituem extensões multi-

banda das pirâmides das transformadas de wavelets.

2.2.1.10 Morfologia Matemática

Granulometrias foram introduzidas por Matheron como um meio de caracterizar imagens binárias [89, 123]. Para se calcular as granulometrias, as imagens são tratadas como coleções de grãos que são “peneirados” com peneiras de furos de tamanhos variados.

A operação morfológica chave desse processo denomina-se *abertura* e é definida como:

$$S \circ E = \cup \{B + x : B + x \leq S\} \quad (2.37)$$

onde S é a imagem binária, E é o elemento estruturante e $E + x = \{y + x : y \in E\}$ é o conjunto translação de E pelo elemento x .

Se $F \circ E = F$, então, para qualquer conjunto S , $S \circ F \leq S \circ E$. Quando $F \circ E = F$, diz-se que F é um conjunto E -aberto. Portanto, se um conjunto F é E -aberto, então $S \circ F \leq S \circ E$, qualquer que seja o conjunto S . Uma consequência imediata desse resultado é que se $E_1, E_2, E_3, \dots$ é uma sequência crescente de elementos estruturantes, tais que E_{k+1} é E_k -aberto, então a sequência das imagens filtradas por essa sequência de elementos estruturantes é decrescente, isto é,

$$S \circ E_1 \geq S \circ E_2 \geq S \circ E_3 \geq \dots \quad (2.38)$$

A sequência de imagens $\{S \circ E_k\}$ é dita uma granulometria.

Assumindo-se que S seja um conjunto limitado, contando-se o número de pixels que sobra após cada operação de abertura de uma granulometria

obtem-se uma função decrescente $\psi(k)$, tal que, para algum K , $\psi(k) = 0$, para $k \geq K$. Informações de textura são reveladas estudando-se a função $\psi(k)$ [28].

2.2.2 Abordagens Estruturais

As abordagens para análise estrutural das texturas pressupõem que texturas são constituídas de primitivas que aparecem em arranjos espaciais repetitivos e regulares na imagem.

Haralick [44] subdividiu a categoria estrutural em dois grupos: fracamente acopladas (texturas nas quais as primitivas são fracamente ligadas ou dispostas aleatoriamente na imagem) e fortemente acopladas (texturas nas quais não existe interação não aleatória entre as primitivas componentes). Outro grupo, porém, pode ser descrito: o sintático, que utiliza gramáticas para descrever (e gerar) texturas. Uma gramática é uma maneira poderosa de se descrever as regras que governam as estruturas regulares das texturas [3].

2.2.2.1 Gramáticas de Formas

Uma gramática de formas [108] é definida como $G_s = (V_t, V_m, R, S)$, onde:

1. V_t é um conjunto de formas terminais (em analogia aos símbolos terminais);
2. V_m é um conjunto finito de formas não-terminais (em analogia aos símbolos não terminais), tal que $V_t \cap V_m = \emptyset$;

3. R é um conjunto finito de regras na forma de um par ordenado (u, v) tal que u é uma forma composta de elementos de V_t^+ e v é uma forma composta de um elemento de V_t^* combinado com um elemento de V_m^* ($V_t^* = V_t^+ \cup \Lambda$, onde Λ é uma forma vazia);
4. S é o símbolo inicial da gramática ($S \in V_m$)

Uma textura é gerada a partir da forma inicial S , seguindo-se as regras definidas em R . A geração termina quando todas as regras referentes às formas não terminais tenham sido aplicadas, ou seja, quando existirem apenas formas terminais na sequência gerada.

2.2.2.2 Densidade de Bordas

A quantidade de pixels pertencentes a bordas por unidade de área foi proposta por Rosenfeld [105] como uma medida de textura. A primitiva de textura nesse caso é o pixel. Dada uma janela centralizada num certo pixel, a distribuição das magnitudes dos gradientes na vizinhança desse pixel, definida pela janela, é determinada. A média dessa distribuição é o número de bordas por unidade de área associado com o dado pixel.

2.2.2.3 Densidade de Máximos e Mínimos

Rosenfeld também sugere o número de pontos de máximo e mínimo locais por unidade de área como sendo uma medida de textura. Os pontos de máximos e mínimos locais são encontrados em cada linha da imagem. Um pixel i de uma dada linha g é considerado um ponto de mínimo local se: $g(i) \leq g(i+1)$ e

$g(i) \leq g(i-1)$. Analogamente, um pixel i de uma dada linha g é considerado um ponto de máximo local se: $g(i) \geq g(i+1)$ e $g(i) \geq g(i-1)$.

O algoritmo utilizado por Rosenfeld percorre as linhas das imagens marcando todos os pixels que satisfazem as condições de máximos e mínimos locais. Em seguida, uma janela quadrada é centralizada em cada pixel e o número de pixels marcados na sua vizinhança definida pela janela são computados, resultando numa imagem textura.

2.2.3 Abordagens Baseadas em Modelos

2.2.3.1 Modelos de Auto-regressão

Modelos de auto-regressão [45] provêm um modo de se utilizar estimativas lineares dos tons de cinza dos pixels a fim de se caracterizar texturas em imagens (dados os tons de cinza numa vizinhança contendo o pixel). Os coeficientes são similares quando as texturas são suaves. Para as texturas mais finas (rugosas), os coeficientes variam bastante.

A dependência linear que um pixel de uma imagem tem sobre outra pode ser representada pelo modelo de auto-regressão para texturas e foi utilizado pela primeira vez por McCormick [90]. No estudo realizado por McCormick, um método de análise de séries temporais foi utilizado para se estimar os parâmetros de uma dada textura. Então, utilizando-se os parâmetros estimados e um conjunto de valores iniciais, uma imagem foi sintetizada. A textura exibida por essa imagem sintetizada era bastante similar à textura dada.

2.2.3.2 Modelos dos Campos de Markov

O modelo de campos aleatórios de Markov [22, 72, 80, 127] considera a influência do contexto espacial. Texturas são caracterizadas pela densidade de probabilidade marginal de cada pixel condicionada a sua vizinhança na imagem, modificando o comportamento do pixel, de acordo com o contexto.

O modelo dos campos aleatórios de Markov para textura assume que um campo de textura é estocástico e estacionário, e deveria satisfazer a propriedade de Markovianidade. Seja $R \times C$ o domínio espacial de uma imagem e para qualquer par $(r, c) \in R \times C$ seja $N(r, c)$ uma vizinhança de (r, c) , tal que $N(r, c)$ não inclua (r, c) . Como o campo é estacionário, $(a, b) \in N(r, c)$ se, e somente se, $(a + i, b + j) \in N(r + i, c + j)$ para todo (i, j) . Assim, a configuração de vizinhança espacial é a mesma sobre toda a imagem.

A condição de Markovianidade é que a probabilidade condicional de um pixel, dados todos os demais pixels da imagem, é igual à probabilidade condicional de um pixel, dados apenas uma vizinhança desse pixel. Ou seja,

$$P[I(r, c) | I(i, j) : (i, j) \in R \times C, (i, j) \neq (r, c)] = P[I(r, c) | i(i, j) : (i, j) \in N(r, c)] \quad (2.39)$$

Pode-se assumir que a distribuição dos tons de cinza de uma imagem possam ser modelados através dos campos aleatórios de Markov. Nesse caso, o problema de caracterização de uma textura resume-se em estimar os parâmetros que caracterizam aquele campo aleatório.

2.2.3.3 Modelo dos Caminhos Aleatórios

Um modelo baseado em caminhos aleatórios procura descrever a distribuição dos tons de cinza e a variação em uma imagem através de um caminho aleatório [125], que é caracterizado no plano por uma partícula movendo-se em passos unitários em direções pré-definidas. Considere um pixel (i, j) numa textura $f(i, j)$. Defina um conjunto de quatro movimentos: $0^\circ[(i, j) \rightarrow (i, j + 1)]$, $90^\circ[(i, j) \rightarrow (i - 1, j)]$, $180^\circ[(i, j) \rightarrow (i, j - 1)]$, e $270^\circ[(i, j) \rightarrow (i + 1, j)]$. Um caminho aleatório é completamente especificado se as probabilidades de se mover uma partícula de um pixel (i, j) para um dos seus pixels vizinhos (k, l) estão definidas. Pixels pertencentes à região nas quais os caminhos aleatórios são executados podem ser separados em dois conjuntos: os pixels pertencentes à borda e os pixels que não pertencem à borda. Os pixels de borda são aqueles que têm pelo menos um vizinho fora da região sendo considerada.

A maioria dos algoritmos para análise de textura baseados em modelos de caminhos aleatórios utilizam as distribuições de probabilidades de se deixar a região interna e chegar até a borda da região.

2.2.3.4 Modelo de Fractais

O universo é bastante complexo e contém inúmeras formas que não podem ser descritas por modelos matemáticos convencionais. Mandelbrot desenvolveu uma nova classe de modelos, denominada Geometria Fractal [78, 79], que permite superar essa dificuldade. Um fractal é definido como um conjunto para o qual a dimensão de Hausdorff-Besicovich é estritamente maior que

a dimensão topológica, ou seja, um conjunto para o qual a única descrição consistente de suas propriedades métricas requer uma dimensão maior do que a nossa definição intuitiva e padrão de dimensão de conjunto. Essa dimensão consistente denomina-se dimensão fractal, frequentemente denotada por D . A dimensão fractal D fornece informação de rugosidade de uma superfície. Para uma superfície lisa, $D = 2$, enquanto que para uma superfície altamente complexa e rugosa, $D = 3$.

A aplicação do modelo de superfícies fractais em análise de imagens e texturas foi feita inicialmente por Pentland [96]. Para realizar segmentação de texturas, Pentland propõe que a imagem de entrada seja percorrida por uma janela de tamanho pré-estabelecido e, para cada posição da janela, a dimensão fractal da sub-imagem definida pela janela seja determinada. O próximo passo consiste em calcular o histograma das dimensões fractais encontradas na imagem. A segmentação da imagem é realizada em seguida, utilizando-se limiares adequadamente escolhidos a partir do histograma das dimensões fractais.

Como foi mencionado anteriormente, a dimensão fractal fornece apenas informações sobre rugosidade da superfície. É possível existirem duas superfícies com texturas distintas e que possuem a mesma dimensão fractal [91]. Dubuisson também conclui que a dimensão fractal utilizada sozinha não permite segmentar todos os tipos de texturas naturais [29]. Para melhorar a aplicabilidade do modelo de superfícies fractais na análise de textura, Ait-kheddache e Rajala sugeriram o uso de dimensões fractais de ordem superior [2].

A dimensão fractal de uma imagem binária pode ser estimada, por exem-

plo, utilizando-se a técnica de contagem de caixas (“box counting”) [15, 29]. Nessa técnica, dada uma caixa de tamanho n , verifica-se qual é o número mínimo de caixas adjacentes, m , necessárias para “cobrir” totalmente o objeto da imagem (definido pelos pixels pretos, caso o fundo seja branco, ou vice-versa). Variando-se o tamanho das caixas, obtém-se uma sequência de pontos (n, m) , onde os valores de n são inversamente proporcionais aos de m . Em seguida, faz-se uma regressão linear dos pontos $(\log \frac{1}{n}, \log m)$. O valor q do coeficiente angular encontrado é então utilizado para determinar a dimensão fractal da imagem em questão. Como o método é sensível à rotação e à translação do sistema de caixas, normalmente utiliza-se média nestes parâmetros.

A técnica proposta por Coelho e Costa [16] para estimar a dimensão fractal de imagens binárias é análoga à técnica descrita acima, no entanto, ela utiliza *salsichas de Minkowski* obtidas através da operação de dilatação com elementos estruturantes de tamanhos crescentes, definida pela Morfologia Matemática [89, 123]. Nessa técnica, estabelece-se uma relação diretamente proporcional entre o tamanho do elemento estruturante (T) e o número de pixels (N) dos objetos dilatados com aquele elemento estruturante.

A Figura 2.14 mostra uma sequência de 20 imagens obtidas através da dilatação da imagem binária original, utilizando-se elementos estruturantes quadrados variando de 3x3 até 22x22 pixels.

A Figura 2.15(a) mostra a relação entre as medidas T e N, e a Figura 2.15(b) mostra a reta $y = 0,4102x + 9,8926$, determinada pela regressão linear dos pontos $(\log(T), \log(N))$.

Portanto, a dimensão fractal da imagem que originou a sequência de di-

latações apresentada na Figura 2.14 é aproximadamente $1,5898 (2 - 0,4102)$.

Resultados obtidos com a aplicação dessa técnica no problema de estimação de densidades de congestionamentos de pessoas estão descritos na seção 6.2.4.

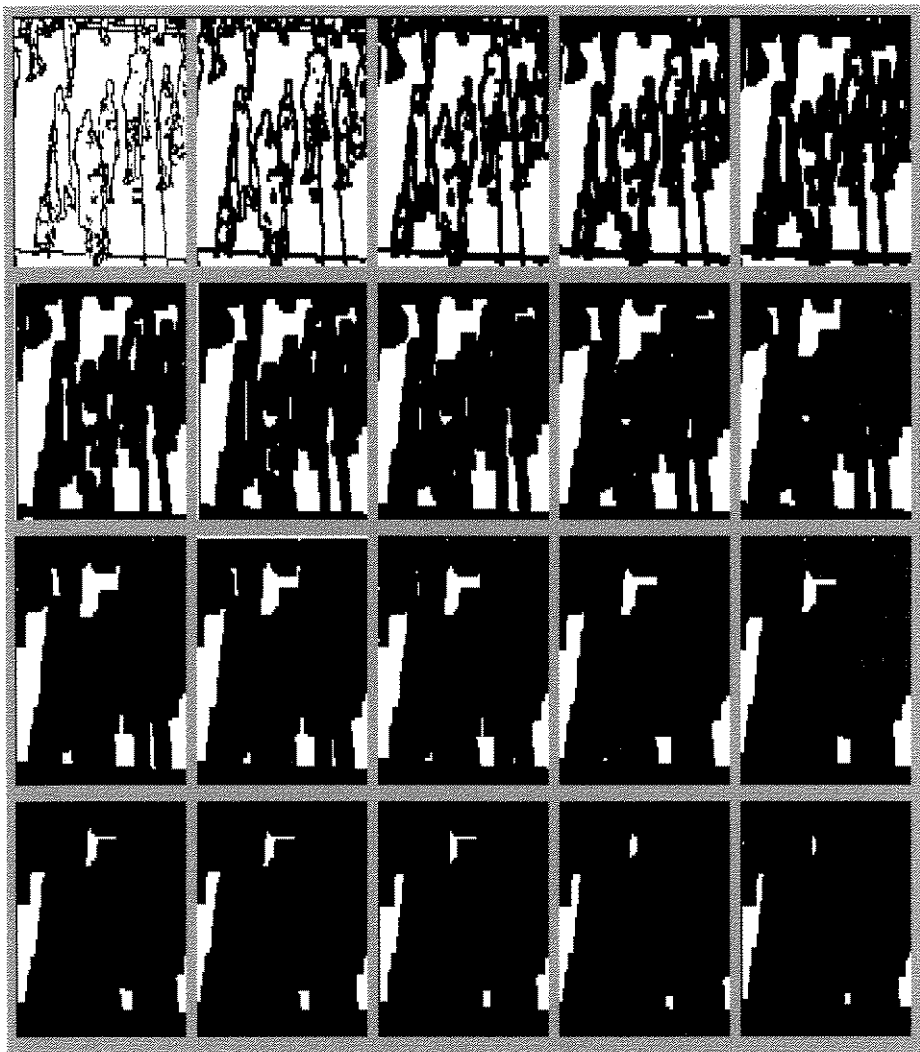


Figura 2.14: Sequência de imagens obtidas com as dilatações de uma imagem binária original, com elementos estruturantes quadrados com tamanhos variando de 3x3 até 22x22 pixels.

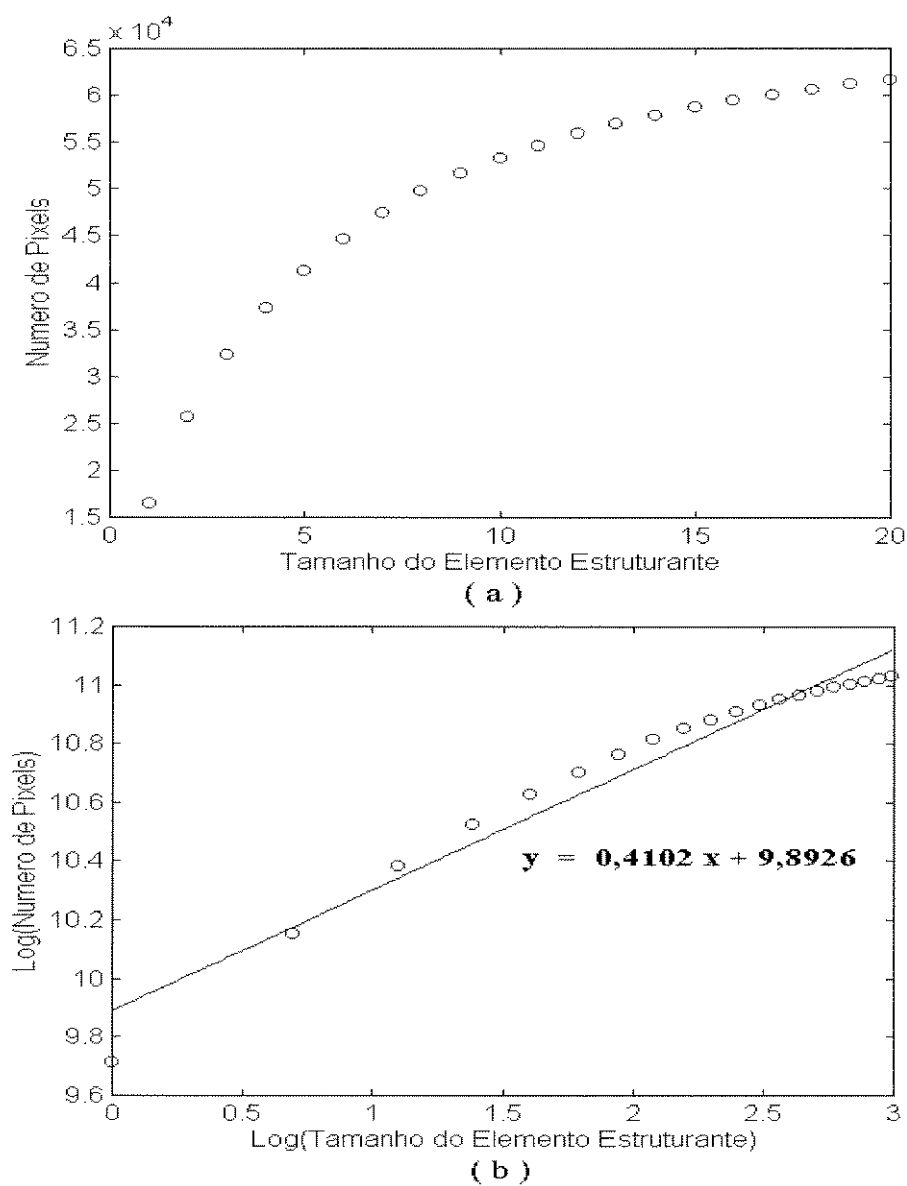


Figura 2.15: (a) Relação entre as medidas dos tamanhos dos elementos estruturantes (T) e os números de pixels pretos (N) nas imagens dilatadas; (b) Reta determinada pela regressão linear do conjunto de pontos $(\log(T), \log(N))$.

Capítulo 3

Análise de Texturas Orientadas Baseada em Segmentos de Retas

3.1 Texturas Orientadas

Texturas orientadas são compostas por padrões retilíneos e/ou alongados que ocorrem frequentemente em fenômenos naturais e constituem uma importante classe de imagens. Um dos objetivos da visão computacional é desenvolver técnicas para extrair informações de imagens compostas por texturas orientados.

Estudos sobre mecanismos seletivos a orientações no sistema visual humano têm inspirado o desenvolvimento de técnicas de visão computacional para a análise de padrões orientados [62, 73, 103].

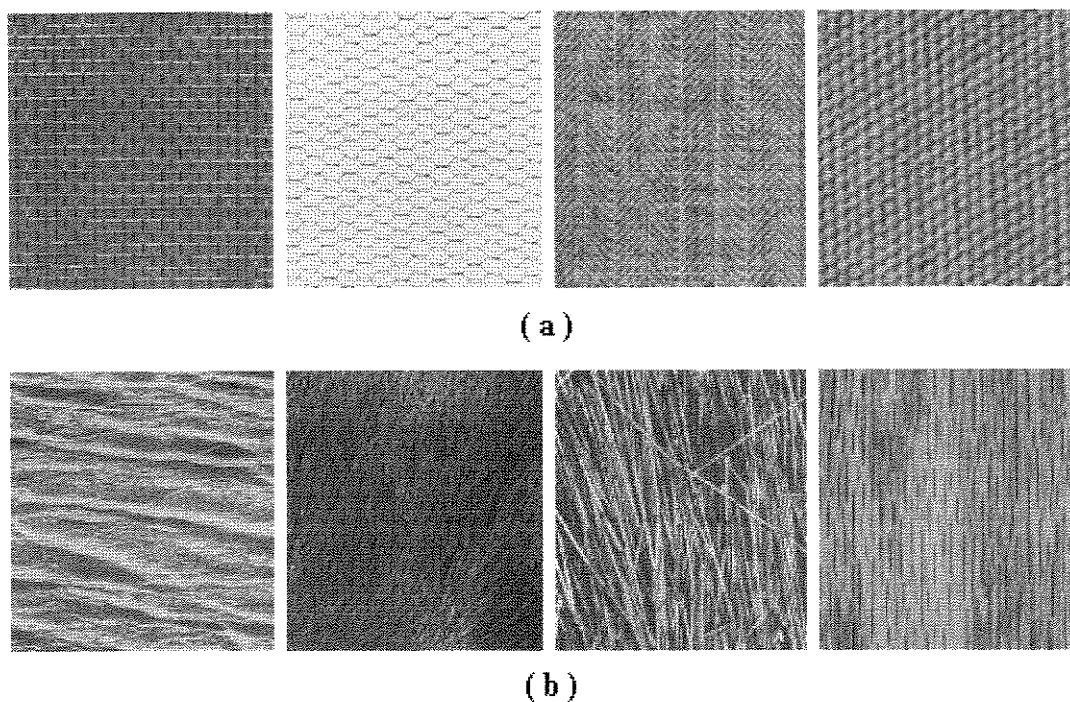


Figura 3.1: Exemplos de texturas orientadas (imagens extraídas do álbum de texturas de Brodatz); (a) Texturas orientadas fortemente acopladas; (b) Texturas orientadas fracamente acopladas.

As texturas orientadas são normalmente exemplos de texturas estruturais, podendo ser forte ou fracamente acopladas. Naquelas fortemente acopladas, os estímulos lineares (que são as primitivas que se repetem na imagem) são deterministicamente posicionados, ao contrário das fracamente acopladas, nas quais as regras de posicionamento são aleatórias.

A Figura 3.1 mostra exemplos de texturas orientadas forte e fracamente acopladas.

Rao [102] propõe uma taxonomia de textura na qual as texturas são separadas em duas grandes categorias: texturas ordenadas e texturas desordenadas. As texturas orientadas, segundo essa classificação, constituem uma sub-classe das texturas ordenadas.

Os métodos para análise de texturas orientadas possuem em geral uma fase inicial onde se realiza a detecção de bordas, através de operadores como o gradiente (primeira derivada) ou o Laplaciano (segunda derivada).

O método proposto por Rao [103] para análise de texturas orientadas baseia-se em estimativas das orientações locais de uma imagem, que são computadas em cinco passos:

1. suaviza-se a imagem através de um filtro Gaussiano;
2. calcula-se o gradiente da imagem suavizada;
3. estima-se as orientações locais dos gradientes;
4. calcula-se a média das orientações locais sobre uma vizinhança;
5. calcula-se uma medida de coerência do padrão.

Kass [62] propõe um método para análise de texturas orientadas similar ao proposto por Rao [103], porém utilizando um operador Laplaciano na fase de detecção de bordas.

Liu [73] propõe um método para tratar o problema da análise direcional no espaço de escalas, que pode ser sumarizado nos seguintes passos:

1. gera-se um conjunto de mapas de cruzamento de zeros, através da utilização de filtros baseados nos Laplacianos de Gaussianas (a variância

- que define a Gaussiana é o parâmetro que define a escala da análise);
2. a partir dos mapas dos cruzamentos de zeros, geram-se mapas de estabilidade;
 3. a partir dos mapas de estabilidade, geram-se índices de estabilidade;
 4. a partir dos índices de estabilidade, calculam-se as orientações das bordas;
 5. calculam-se as distribuições de orientações dos segmentos detectados;
 6. calculam-se medidas estatísticas para quantificar a distribuição angular dos padrões lineares.

Denslow [26] também trata da questão de texturas orientadas propondo uma caracterização estatística dos elementos que quantificam a direcionalidade nessas texturas. Denslow define direcionalidade como a anisotropia em imagens que contém pelo menos um dos três seguintes tipos de padrões:

1. objetos excêntricos dispostos na imagem em uma direção preferencial (Figura 3.2(a));
2. bordas alinhadas em alguma direção preferencial, não sendo necessário que tais bordas pertençam a objetos discretos (Figura 3.2(b));
3. arranjos de objetos concêntricos (circulares), alinhados em orientações preferenciais (Figura 3.2(c)).

Denslow [26] propõe ainda que o ângulo e a intensidade (“strength”) direcional dos padrões sejam utilizados como medidas de direcionalidade. O

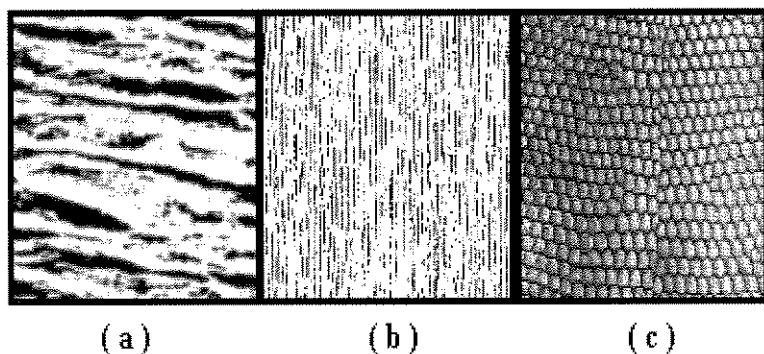


Figura 3.2: Exemplos de padrões de direcionalidade. (a) Objetos excêntricos alongados numa orientação preferencial; (b) Bordas alinhadas numa orientação preferencial; (c) Objetos concêntricos dispostos em orientações preferenciais.

conceito de intensidade direcional é definido heurísticamente combinando-se os comprimentos médios, as mudanças de intensidades de tonalidades, as extensões dos alinhamentos angulares e as densidades espaciais dos elementos de borda da imagem.

A Figura 3.3 ilustra os componentes empregados na definição do conceito de intensidade direcional proposto por Denslow [26].

Conforme pode ser observado na Figura 3.3, os quatro componentes propostos contribuem para aumentar ou diminuir a intensidade direcional dos padrões, na medida em que estão mais ou menos presentes nesses padrões. Na coluna mais a esquerda, conforme aumenta-se o nível de ruído, altera-se as intensidades dos tons de cinza de cinza da imagem, diminuindo-se o contraste e, conseqüentemente, diminuindo-se também a intensidade direcional da ima-

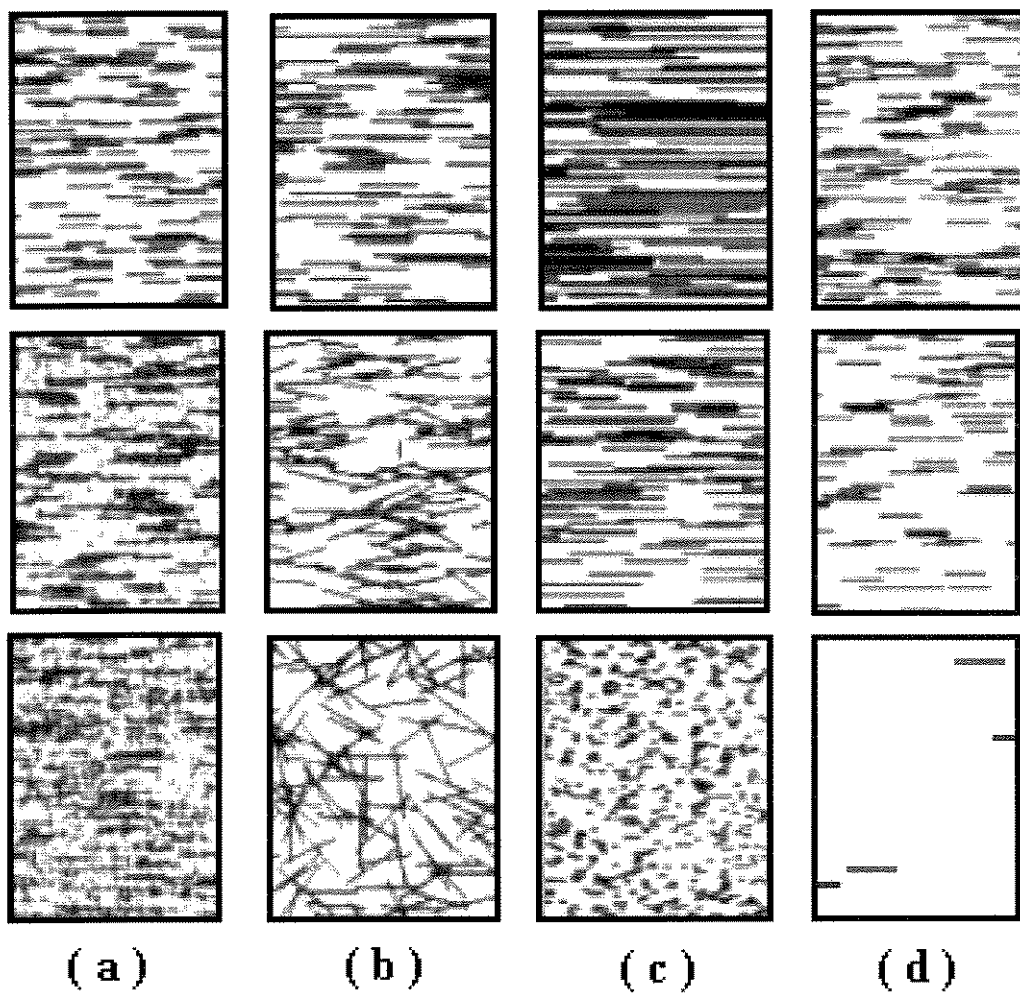


Figura 3.3: Componentes que influenciam na medida de intensidade direcional. (a) Intensidade dos tons cinza; (b) Dispersão angular; (c) Comprimento; (d) Densidade espacial.

gem. Na segunda coluna da esquerda para a direita, conforme aumenta-se a dispersão angular dos elementos de borda da imagem, diminui-se a intensidade direcional da mesma. Na coluna seguinte, pode-se observar que, fixada uma direção, as imagens constituídas de elementos de bordas mais compridos, dispostos nessa orientação, apresentam maiores intensidades direcionais. Na coluna mais à direita, pode-se observar a influência que a densidade espacial dos elementos da imagem exercem na determinação da intensidade direcional. Fixada a orientação e o comprimento, pode-se observar que a intensidade direcional aumenta, conforme aumenta a densidade dos elementos de borda. Obviamente essa afirmação vale até um certo limite, pois o aumento excessivo da densidade desses elementos levariam à formação de uma região compacta e, conseqüentemente, à perda da informação de direção.

Os conceitos de ângulo e intensidade direcional são utilizados nas técnicas apresentadas nas seções 3.3.1 e 3.3.2 para a extração de características de texturas orientadas.

3.2 Extração de Descritores de Texturas Orientadas

James Bergen [6] considera que enquanto os modelos para percepção de texturas baseados em conceitos tais como os “textons” de Julesz [59, 60] têm sido importantes para motivar as pesquisas sobre o processo de visão biológica, a formalização dos mesmos não tem sido concretizada, tornando difícil a aplicação de tais modelos numa classe ampla de texturas. A maior dificul-

dade em se utilizar os conceitos dos “textons” reside no fato deles serem descritos verbalmente, e não a partir dos valores dos tons de cinza da imagem, por exemplo.

Para que seja possível analisar classes mais gerais de textura utilizando-se o conceito de “textons”, é necessária a construção de operadores capazes de extrair das imagens as características de texturas descritas por eles.

Nessa seção, são descritos operadores para extrair segmentos de retas, que são particularizações dos “blobs” alongados, definidos como sendo textons na teoria de Julesz.

Existem muitos tipos de texturas que podem ser representadas em termos de segmentos de retas. Os experimentos psicofísicos e neurofisiológicos realizados para se investigar o processamento biológico das texturas, por exemplo, empregam comumente estímulos construídos por meio de segmentos de retas [59, 60, 115, 124]. Além disso, as retas constituem elementos importantes na representação visual, haja visto que quaisquer curvas pertencentes a espaços amostrados, tais como aqueles encontrados nos sistemas biológicos de visão, podem ser representadas exatamente em termos de segmentos de retas discretos [21].

Acredita-se que grande parte do córtex visual dos primatas é dedicado à transformação de versões das imagens com as bordas detectadas recebidas da retina, através do Núcleo Lateral Geniculado, em representações mais compactas em termos de segmentos de retas. Os estudos pioneiros de Hubel e Wiesel [50, 51] demonstraram a existência de células no córtex visual primário (área 17), que apresentam campos receptivos próprios para a detecção de estímulos retilíneos. Outras células possuem campos receptivos mais com-

plexos, o que permite-lhes ser seletivas às orientações, aos comprimentos e às direções dos movimentos dos estímulos retilíneos.

A importância dos segmentos de retas para o sistema visual dos primatas também vem sendo investigada por Blasdel [8] e Heydt [124], tendo esse último detectado um tipo de célula no córtex visual que apresenta respostas máximas para estímulos no formato de “gratings”. Heydt acredita que essas células, que são sensíveis às mudanças nas orientações e frequências espaciais dos estímulos, executam uma função importante na percepção de texturas.

Nas seções 3.3.1 e 3.3.2 são apresentadas técnicas para extração de descritores de texturas orientadas a partir das informações de orientação, frequência espacial e comprimento dos segmentos de retas (textons) presentes na imagem. Graças à sua eficácia e relativa simplicidade, a transformada de Hough, que encontra-se descrita na seção 3.3, é adotada em tais técnicas para a extração dos descritores retilíneos de textura.

3.3 Transformada de Hough

Um problema comum no processo de extração de primitivas de imagens consiste na detecção de curvas analíticas, tais como: retas, círculos, parábolas e elipses. A transformada de Hough (TH) [49, 54] fornece um meio efetivo para a detecção de tais curvas, particularmente para a detecção de retas.

Geralmente, a entrada da transformada de Hough consiste de uma imagem binária obtida por um processo de detecção de bordas aplicado na imagem original, normalmente representada em tons de cinza.

O princípio básico da TH consiste no mapeamento de todos os pontos de

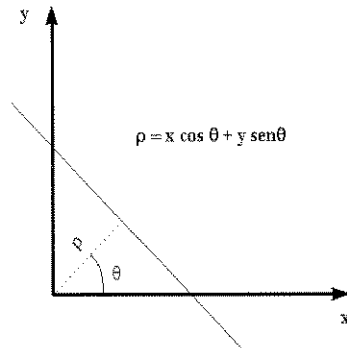


Figura 3.4: Parametrização normal da reta.

uma curva no espaço da imagem, num único ponto no espaço dos parâmetros (também denominado de espaço de Hough).

A parametrização de retas através da equação $y = ax + b$, originalmente adotada por Hough [49] para a detecção de retas, apresenta o inconveniente de ter os parâmetros de coeficiente angular, a , e de interseção ao eixo das ordenadas, b , não limitados, quando as retas possuem inclinações quase verticais.

Para se evitar esse problema, a transformada de Hough é geralmente calculada utilizando-se equação normal da reta, dada por:

$$\rho = x \cos \theta + y \sin \theta \quad (3.1)$$

onde ρ é a distância normal da reta até a origem do sistema de coordenadas do espaço da imagem, e θ é o ângulo formado entre o eixo das abscissas e a reta perpendicular à reta sendo definida. Veja a ilustração dos parâmetros da equação normal da reta na Figura 3.4.

A equação normal da reta, originalmente proposta por Duda e Hart [30]

para ser utilizada na transformada de Hough, tem os parâmetros θ e ρ limitados, ou seja, $0 \leq \theta \leq \pi$ e $-\frac{\sqrt{n^2+m^2}}{2} \leq \rho \leq +\frac{\sqrt{n^2+m^2}}{2}$, onde m é a largura e n é a altura da imagem, e o ponto $(\frac{m}{2}, \frac{n}{2})$ é a origem do sistema de coordenadas do espaço da imagem.

Utilizando-se a equação normal da reta, a TH tem as seguintes propriedades:

Propriedade 1: um ponto no espaço da imagem corresponde a uma senóide no espaço dos parâmetros (θ, ρ) ;

Propriedade 2: um ponto no espaço dos parâmetros (θ, ρ) , corresponde a uma reta no espaço da imagem;

Propriedade 3: pontos da mesma senóide no espaço dos parâmetros (θ, ρ) correspondem a retas que se interceptam num mesmo ponto no espaço da imagem;

Propriedade 4: pontos da mesma reta no espaço da imagem correspondem a senóides que se interceptam num mesmo ponto no espaço de parâmetros (θ, ρ) .

A quarta propriedade da TH é a mais importante para a detecção de segmentos de retas, pois se todas as senóides $S(x, y)$ que correspondem a pontos colineares $P(\theta, \rho)$ no espaço da imagem têm um ponto de intersecção em comum no espaço dos parâmetros, então, as coordenadas (θ, ρ) desse ponto de intersecção definem a reta que contém todos os pontos colineares $P(\theta, \rho)$ no espaço da imagem.

Para se calcular a TH utiliza-se uma matriz acumuladora definida a partir da amostragem do espaço de parâmetros, cujos elementos iniciais têm valor zero. As dimensões dessa matriz dependem dos intervalos de amostragem dos eixos θ e ρ do espaço dos parâmetros. A matriz acumuladora tem as mesmas dimensões do espaço discreto de Hough.

Para cada θ , amostrado no intervalo $[0, \pi]$, e para cada elemento de borda (x, y) da imagem binária de entrada, o valor de ρ é determinado através da equação 3.1 e aproximado para o valor discreto mais próximo. Em seguida, a célula da matriz acumuladora correspondente à posição (θ, ρ) é incrementada de 1.

De acordo com a quarta propriedade da transformada de Hough descrita acima, todos os pontos colineares do espaço da imagem contribuem para incrementar a mesma célula (θ, ρ) da matriz acumuladora, onde as senóides se interceptam no espaço dos parâmetros, resultando num grande valor acumulado (ou seja, um pico) nessa posição.

Portanto, os picos resultantes na matriz acumuladora após a transformada de Hough indicam as direções no espaço da imagem, bem como as distâncias à origem desse espaço, onde existem grandes quantidades de pontos colineares.

A Figura 3.5 mostra um exemplo de transformada de Hough utilizando-se a equação 3.1. Nessa figura, a transformada de Hough foi mapeada em tons de cinza, sendo que o maior pico foi mapeado em branco e o valor 0 do espaço de Hough foi mapeado na cor preta.

Pode-se notar no espaço de Hough (Figura 3.5(b)), a existência de 17 pontos principais de intersecções de senóides. Esses pontos, que apresentam tons bastante claros de cinza, indicam as direções θ e as distâncias da origem

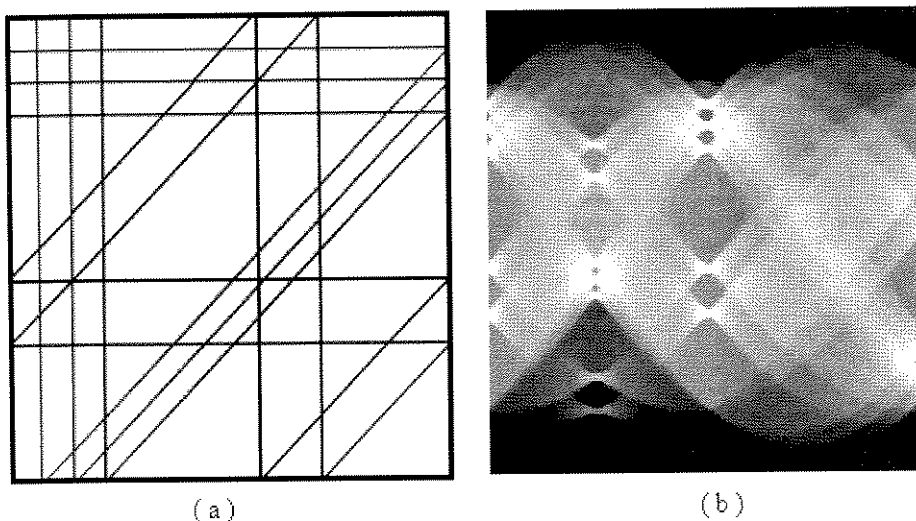


Figura 3.5: (a) Imagem constituída de 17 retas; (b) Espaço de Hough da imagem apresentada em (a).

ρ que definem as 17 retas presentes na imagem da Figura 3.5(a).

Embora a transformada de Hough possa ser generalizada para a detecção de quaisquer curvas paramétricas, as curvas de ordens maiores são definidas por um número maior de parâmetros e, conseqüentemente, requerem mais memória para armazenar o espaço de parâmetros multidimensional. Além disso, exigem maior esforço computacional para se calcular a transformada de Hough.

A literatura apresenta vários tipos de algoritmos destinados ao cálculo da TH, dentre os quais a transformada rápida de Hough [71], a transformada adaptativa de Hough [53], a transformada combinatorial de Hough [118], a transformada binária de Hough [19] e a transformada dinâmica de Hough [14]. De modo geral, esses algoritmos objetivam melhorar alguns aspectos da trans-

formada padrão de Hough, tais como o tempo de execução ou a acurácia. Illingworth e Kittler [54] apresentam uma revisão dos algoritmos para o cálculo da transformada de Hough que foram propostos até o ano de 1988.

Uma das principais dificuldades da transformada de Hough diz respeito à discretização do espaço dos parâmetros (matriz acumuladora). Se esse espaço for sub-amostrado, algumas curvas podem ser mapeadas no espaço de Hough em pontos correspondentes a parâmetros muito distantes daqueles que realmente as definem, dificultando a tarefa de detecção. Por outro lado, a super-amostragem, além de aumentar o tempo computacional necessário para calcular a transformação, faz com que pontos pertencentes à mesma curva no espaço da imagem sejam mapeados em curvas que se interceptam em pontos distintos no espaço de Hough. Dessa forma, os picos no espaço dos parâmetros, que possibilitam a detecção das curvas no espaço da imagem, são pulverizados localmente no espaço dos parâmetros, dificultando também a tarefa de detecção [20].

No caso da detecção de segmentos de retas, outra deficiência da transformada de Hough consiste no fato dela não fornecer informações sobre as extremidades (ou seja, os comprimentos) desses segmentos, uma vez que a contiguidade dos pontos colineares é uma informação não explorada por essa transformação.

Quando a questão da conectividade for importante, é necessário realizar algum tipo de pós-processamento depois de calculada a transformada de Hough. Costa [20] propõe um algoritmo para efetuar a análise de conectividade durante a detecção dos segmentos de retas (veja detalhes desse algoritmo na seção 3.3.4).

A transformada de Hough para a detecção de retas possui ainda algumas propriedades importantes relacionadas às transformações geométricas:

Propriedade 5: se duas retas se interceptam a um ângulo ϕ , onde $0^0 < \phi \leq 90^0$, então os picos correspondentes a essas retas no espaço de parâmetros localizam-se a uma distância ϕ no eixo θ ;

Propriedade 6: a translação linear da imagem tem o efeito de deslocar a TH na direção ρ . O deslocamento não é uniforme, ou seja, picos em determinados ângulos têm deslocamentos maiores do que em outros. Por exemplo, os parâmetros ρ das retas paralelas à direção do deslocamento não sofrem alterações com o deslocamento, enquanto que os parâmetros ρ das retas perpendiculares à direção do deslocamento são bastante afetadas;

Propriedade 7: translações no espaço da imagem não afetam as distâncias relativas dos picos no eixo ρ , pois, embora as posições das retas relativas à origem possam mudar, os posicionamentos relativos das mesmas permanecem inalterados com as translações;

Propriedade 8: rotações no espaço da imagem por um ângulo ϕ equivale a um deslocamento circular ϕ no eixo θ no espaço dos parâmetros;

Propriedade 9: alterações de escala iguais nas duas direções do espaço da imagem equivale a alterar a escala no eixo ρ do espaço dos parâmetros.

Essas propriedades mostram os efeitos das transformações geométricas na TH e indicam que normalizações adequadas podem compensar tais trans-

formações. Isso faz com que a TH possa ser utilizada com êxito em aplicações que necessitem ter a propriedade de invariância às transformações geométricas.

3.3.1 Descritores de Texturas Extraídas de Segmentos de Retas

Para obterem êxito, as técnicas para caracterização de texturas baseadas em informações extraídas dos segmentos de retas detectados na imagem precisam ser eficazes na detecção desses segmentos. Além da transformada de Hough, outros processamentos complementares são necessários, como por exemplo, a análise de conectividade e a determinação dos picos relevantes do espaço de Hough.

O processo de detecção dos segmentos de retas através da TH pode ser resumida nos seguintes passos:

1. a imagem de entrada é submetida a um filtro passa-alta (por exemplo, o gradiente da gaussiana) a fim de ter suas bordas detectadas. As bordas detectadas são afinadas e a imagem resultante é binarizada;
2. calcula-se a transformada de Hough (ou seja, a matriz acumuladora) da imagem binária utilizando-se a equação normal da reta (equação 3.1);
3. determina-se os picos relevantes da matriz acumuladora e armazena-se suas coordenadas (θ, ρ) numa lista de picos P ;
4. detecta-se, com o auxílio de uma técnica de análise de conectividade, os segmentos de retas da imagem nas direções θ e distâncias ρ indi-

casas pelas coordenadas dos picos armazenados na lista P . Insere-se os segmentos detectados numa lista de segmentos de retas L .

Detalhes da determinação dos picos relevantes da matriz acumuladora (passo 3) e da análise de conectividade (passo 4) são discutidos nas seções 3.3.3 e 3.3.4, respectivamente.

Detectados os segmentos de retas que constituem uma aproximação das texturas orientadas da imagem, a próxima etapa consiste em caracterizar as texturas, baseando-se em informações extraídas dos segmentos de retas. Como foi discutido na seção 3.1, o conceito de ângulo, bem como os de dispersão angular, intensidade dos níveis de cinza, comprimento e densidade dos segmentos de retas, que definem o conceito de intensidade direcional, são informações importantes para a caracterização de texturas orientadas.

A seguir são apresentadas duas técnicas para caracterização de texturas orientadas. A primeira baseia-se somente na informação de ângulo direcional, enquanto que a segunda utiliza as demais informações.

3.3.1.1 Técnica Baseada em Ângulos

Nessa técnica, a caracterização de texturas orientadas é baseada na densidade das orientações dos segmentos de retas coletados em histogramas de orientações [81].

Calculado o histograma das orientações dos segmentos de retas, a imagem é classificada de acordo com as principais modas do histograma. Para cada moda relevante, a lista de segmentos de retas L é varrida para se determinar os segmentos que possuem ângulos próximos àquela moda. As larguras dos

intervalos de orientações que definem o critério de proximidade às modas são parâmetros importante dessa técnica, pois determinam quais são os segmentos que pertencem à mesma textura orientada.

Outro parâmetro importante é a significância das orientações. Orientações com baixa densidade não são consideradas, pois não são relevantes o suficiente para caracterizar novas classes de textura.

Para retratar mais fielmente a relevância de cada orientação na imagem, o cálculo do histograma de orientações considera as densidades de pixels em cada orientação, número esse que pode ser aproximado através da soma dos comprimentos dos segmentos de retas de cada orientação dividida pela soma dos comprimentos de todos os segmentos de retas detectados.

A Figura 3.6(b) apresenta o conjunto de segmentos de retas detectados via transformada de Hough na imagem sintética apresentada na Figura 3.6(a). O histograma de orientações está apresentado na Figura 3.6(c). Na imagem da Figura 3.6(a) existem duas orientações principais, portanto o histograma de orientações apresenta apenas duas modas relevantes (próximas a 30^0 e 120^0).

O passo seguinte dessa técnica consiste em segmentar as regiões da imagem caracterizadas que possuem orientações similares. Para tanto, para cada moda significativa do histograma de orientações, traçam-se numa imagem auxiliar todos os segmentos da lista de segmentos de retas que possuem ângulos similares àquela moda. Posteriormente, aplica-se o operador morfológico de fechamento [106] na imagem auxiliar, a fim de se compactar (segmentar) as regiões da imagem caracterizadas pela orientação sendo considerada (veja as Figuras 3.6(d) e 3.6(e)). A operação de fechamento é definida como uma

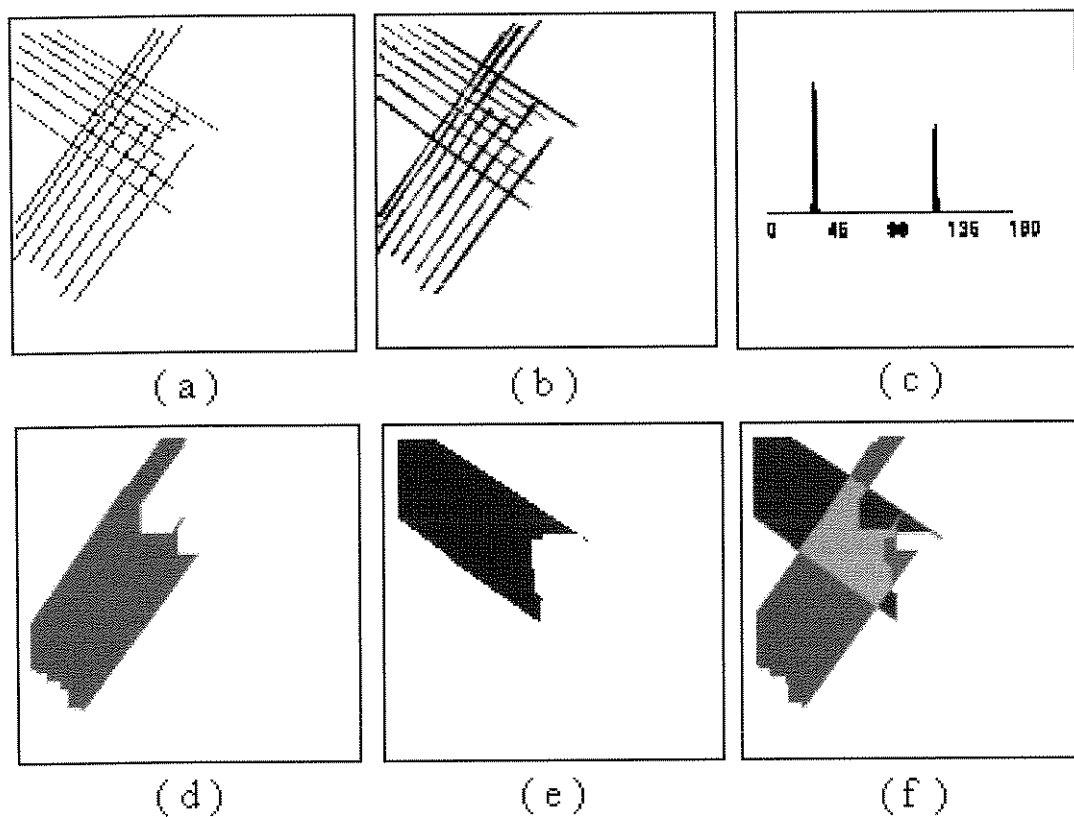


Figura 3.6: (a) Imagem sintética de entrada; (b) Segmentos detectados via TH; (c) Histograma de orientações; (d) Região segmentada caracterizada pela orientação próxima a 30^0 ; (e) Região segmentada caracterizada pela orientação próxima a 120^0 ; (f) Texturas classificadas.

dilatação seguida por uma erosão utilizando-se o mesmo elemento estruturante.

O problema nesse ponto é determinar um tamanho n adequado para o elemento estruturante a fim de promover as compactações (segmentações) das sub-regiões orientadas. Os tamanhos dos elementos estruturantes podem ser determinados em função das distâncias relativas dos segmentos de retas que possuem orientações similares àquela sendo considerada num certo instante do processamento. Essas distâncias são facilmente calculadas, uma vez que se tem disponíveis na lista de segmentos de retas os valores ρ que definem tais segmentos.

Observando-se o espaço de Hough apresentado na Figura 3.5(b), por exemplo, é possível supor a existência no espaço da imagem de cinco retas verticais, sendo que as duas mais próximas do centro estão afastadas a uma distância maior, d_1 , entre si do que as três retas que se encontram mais distantes do centro da imagem, que estão afastadas entre si a uma distância d_2 . O valor da maior distância relativa d_1 é um bom candidato para ser usado para definir o tamanho n do elemento estruturante da operação de fechamento. No caso de existirem i distâncias distintas, pode-se tomar a média das k distâncias maiores como o valor de n .

Outra possibilidade é tomar o valor da menor distância relativa como sendo o primeiro valor do parâmetro n e depois promover várias operações de fechamento, incrementando-se paulatinamente o parâmetro n até que o número de pixels das áreas compactadas se estabilize. Os valores do parâmetro n para cada orientação são aqueles que marcam o início das seções de valores constantes das funções de medida de granulometria.

Na fase de classificação das texturas encontradas na imagem, os intervalos de orientações são associados a rótulos que são utilizados para classificar os pixels das regiões da imagem caracterizadas pelas orientações daqueles intervalos. Esses rótulos são números da forma 2^i , $i = 1, 2, 3, 4, \dots, r$, onde r é o maior valor possível de intervalos de orientações (se cada intervalo tem tamanho 30, por exemplo, então $r = \frac{180}{30} = 6$). O rótulo $l = 2^0 = 1$ é associado ao intervalo de orientações que contém a maior moda detectada no histograma de orientações. Os pixels da região segmentada utilizando-se as retas desse intervalo são rotulados com esse valor. Em seguida, o rótulo $l = 2^1 = 2$ é associado ao intervalo de orientações que contém a segunda maior moda detectada no histograma de orientações e os pixels da região segmentada da imagem através dos segmentos de retas desse intervalo são rotulados com esse valor, exceto os pixels dessa região que já foram rotulados em passos anteriores. Nesse caso, esses pixels recebem o valor do rótulo anterior somado ao valor do rótulo atual. Esse processo de rotulação se repete até que todos os intervalos de orientações associados às modas relevantes do histograma de orientações sejam considerados.

Com esse esquema de classificação, se um pixel p pertence à intersecção das áreas segmentadas correspondentes às três maiores modas, ele receberá o rótulo $l = 2^0 + 2^1 + 2^2 = 7$. Por outro lado, um pixel com rótulo $l = 42$, por exemplo, indica que ele pertence às regiões segmentadas correspondentes às modas 1, 3 e 5, pois $42 = 2^1 + 2^3 + 2^5$.

Nessa esquema de classificação, as orientações e as intersecções de orientações definem classes distintas de textura. Para se determinar quantas classes distintas de texturas orientadas existem na imagem basta verificar

quantos rótulos distintos foram utilizados para classificar os pixels da imagem.

Uma característica interessante desse método de classificação de textura orientada é a invariância que ele apresenta às transformações geométricas de translação, rotação e escala. A invariância à rotação decorre do fato do método considerar como sendo da classe $l = 1$, os pixels das áreas da imagem caracterizadas pelas orientações dos intervalos relacionados à maior moda, da classe $l = 2$ os pixels relacionados à segunda maior moda, e assim sucessivamente.

A invariância com relação às alterações de escala decorre do fato do parâmetro n que define o elemento estruturante ser obtido como função das distâncias relativas dos segmentos de retas com orientações similares. Aumentando-se (diminuindo-se) a escala, aumentam-se (diminuem-se) proporcionalmente as distâncias relativas e, conseqüentemente, os valores dos parâmetros n .

A invariância com relação às translações é garantida pelo fato do método levar em consideração os posicionamentos dos segmentos de retas detectados na imagem para executar a segmentação das áreas com texturas distintas.

A Figura 3.7 apresenta os resultados da aplicação em uma imagem real da técnica estrutural para classificação de texturas orientadas descrita nessa seção.

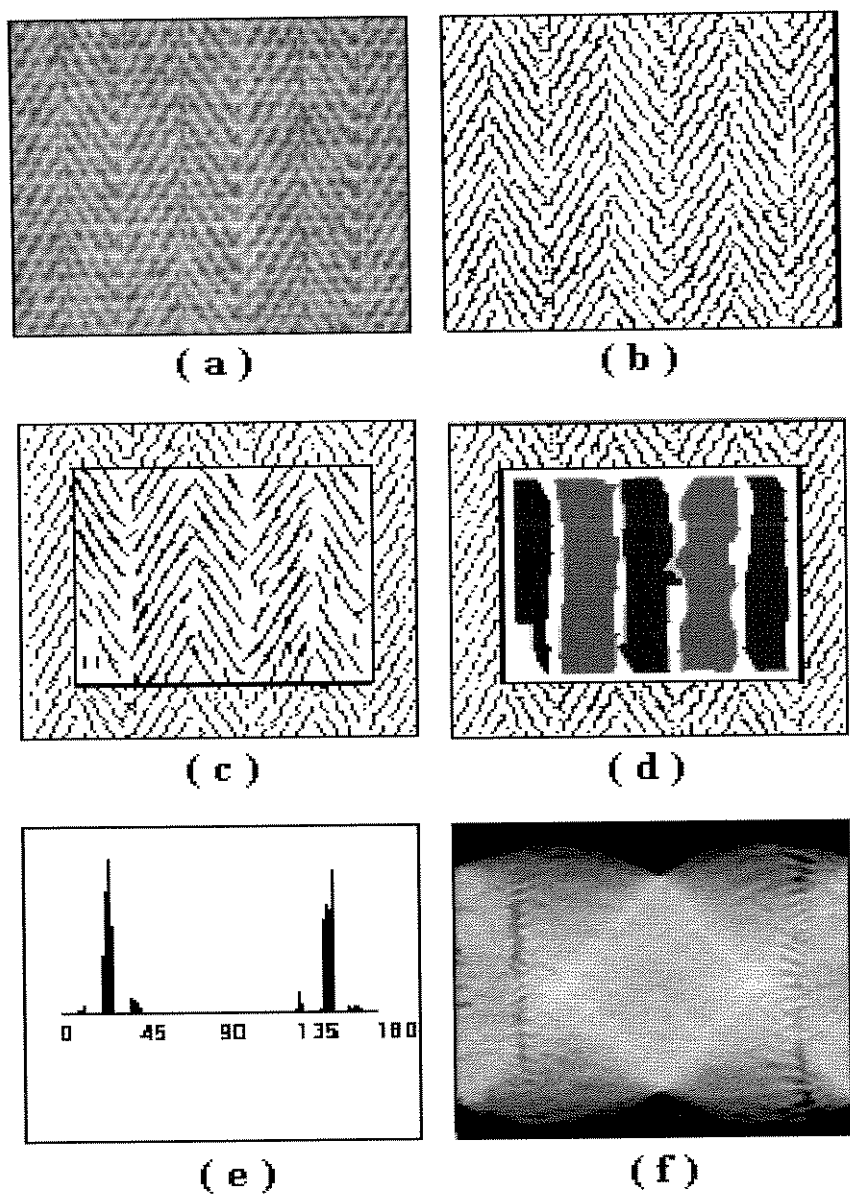


Figura 3.7: (a) Imagem de entrada; (b) Bordas detectadas; (c) Segmentos de retas detectados via TH; (d) Texturas classificadas; (e) Histograma de orientações; (f) Espaço de Hough da imagem das bordas apresentada em (b).

3.3.1.2 Técnica Baseada em Intensidades Direcionais

Além das informações dos ângulos, os segmentos de retas fornecem também informações que caracterizam as intensidades direcionais, definidas por Denslow [26].

Essas medidas referem-se aos comprimentos, às densidades espaciais, às dispersões angulares e às intensidades dos tons de cinza dos segmentos de retas.

No Capítulo 5 as medidas de média, variância e entropia dos ângulos e dos comprimentos dos segmentos de retas, assim como a quantidade e o comprimento máximo dos segmentos encontrados na imagem, são propostos para serem descritores de texturas orientadas. Dessa forma, exceto as informações das intensidades dos tons de cinza, todos os demais elementos que definem o conceito de intensidade direcional são considerados.

O desempenho desses descritores são avaliados comparando-os com descritores propostos por outras técnicas tradicionais para análise de textura, numa aplicação em que se está interessado em estimar automaticamente a densidade de multidões baseadas em informações de texturas.

3.3.2 Descritores de Texturas Extraídas do Espaço de Hough

O espaço de Hough de uma imagem, calculado por meio da equação normal da reta (equação 3.1), contém todas as informações a respeito das orientações dos elementos de bordas presentes nessa imagem. Dessa forma, o espaço de Hough pode ser utilizado para a caracterização de texturas orientadas,

podendo ser tomado como um vetor de características.

O ideal seria tomar todo o espaço como sendo tal vetor, porém isso é inviável do ponto de vista computacional devido ao tamanho desse espaço (o parâmetro θ é amostrado no intervalo $0 \leq \theta \leq \pi$ e o parâmetro ρ é amostrado no intervalo $-\frac{\sqrt{n^2+m^2}}{2} \leq \rho \leq \frac{\sqrt{n^2+m^2}}{2}$, onde m é a largura e n é a altura da imagem).

Uma forma de reduzir o tamanho do vetor de características de textura extraído diretamente do espaço de Hough é definir esse vetor como sendo uma amostra do espaço, ou defini-lo através de medidas das distribuições dos picos nesse espaço [82].

3.3.2.1 Amostragem Não-Uniforme do Espaço de Hough

Nessa abordagem, o espaço de Hough de tamanho $N_\rho \times N_\theta$ é uniformemente amostrado em m colunas no eixo θ , onde $m < N_\theta$. Para cada uma das m colunas, efetua-se uma amostragem não-uniforme dos seus elementos, considerando-se para tanto apenas aqueles que pertencem ao envelope da transformada de Hough. O envelope da transformada de Hough de uma imagem calculada utilizando-se a equação normal da reta (equação 3.1), é formado por todos os pontos (θ, ρ) do espaço de parâmetros que podem pertencer a alguma senóide definida por esta equação, sendo (x, y) as coordenadas dos pixels da imagem em questão.

Essa amostragem não-uniforme consiste em subdividir cada coluna dentro do envelope da transformada em n intervalos ($n < N_\rho$), e tomar os valores máximos dentro de cada intervalo. Os n valores máximos obtidos para cada

coluna amostrada são os elementos que constituem o vetor de características de textura. Portanto, o vetor de características tem tamanho $m \times n$.

A técnica de adotar uma amostragem do espaço de Hough como sendo o vetor de características foi utilizada por Chan [14] para fazer reconhecimento de padrões.

A Figura 3.8(c) mostra o resultado da classificação de texturas da imagem D17 do álbum de texturas de Brodatz [10], apresentada na Figura 3.8(a), utilizando-se $m = 15$ e $n = 4$, totalizando 60 características. A classificação foi supervisionada e realizada através de uma rede neural implementada segundo o modelo proposto por Kohonen [63, 64, 65], com um mapa de auto-organização de tamanho 7×7 . A seção 4.1.2.1 apresenta detalhes desse modelo de rede neural.

Na fase de treinamento, foram utilizadas 100 amostras de cada uma das duas classes distintas de textura, que foram apresentadas aleatoriamente 10000 vezes para a rede neural.

Observando-se o resultado obtido com os descritores propostos nessa seção, pode-se verificar que houve um melhor desempenho dessa técnica, se comparado ao resultado obtido utilizando-se a técnica baseada em bancos de filtros de Gabor descrita na seção 2.2.1.8. A imagem classificada utilizando-se bancos de filtros de Gabor está apresentada na Figura 3.8(b).

3.3.2.2 Medidas da Distribuição dos Picos do Espaço de Hough

Nessa abordagem, o vetor de características de texturas contém informações sobre orientação, frequência espacial e densidade em cada orientação.

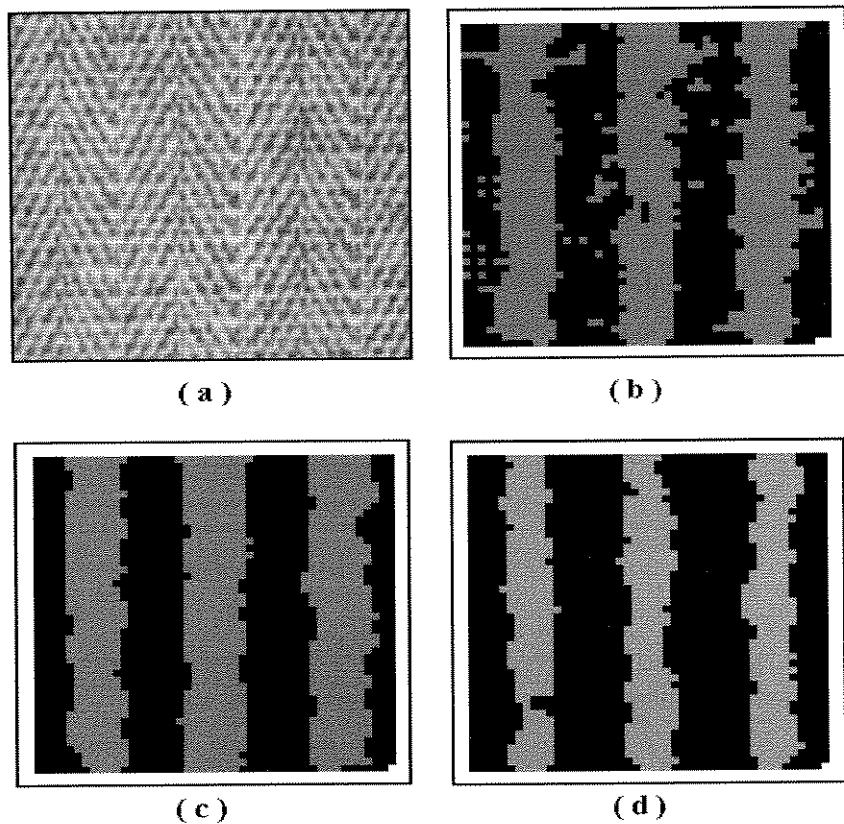


Figura 3.8: (a) Imagem D17 do álbum de texturas de Brodatz; (b) Classificação baseada nos filtros de Gabor; (c) Classificação baseada na amostragem do espaço de Hough; (d) Classificação baseada nas distribuições dos picos do espaço de Hough.

Os elementos do vetor de características são obtidos da seguinte forma:

1. consideram-se intervalos de colunas do espaço de Hough;
2. determinam-se os picos do espaço de Hough (veja a seção 3.3.3);
3. para cada intervalo de colunas (ou seja, para cada intervalo de orientações) calcula-se a distância média dos picos nessas colunas; armazenam-se os valores calculados num vetor auxiliar v_1 ;
4. para cada intervalo de colunas somam-se os valores dos picos; armazenam-se os valores calculados num vetor auxiliar v_2 ;
5. o vetor de características é obtido concatenando-se os vetores v_1 e v_2 .

A Figura 3.8(d) mostra o resultado da classificação da imagem apresentada na Figura 3.8(a), utilizando-se descritores de textura obtidos da forma exposta acima.

Assim como na abordagem anterior, a classificação foi supervisionada e realizada por uma rede neural implementada segundo o modelo de auto-organização proposto por Kohonen [63, 64, 65], com mapa de auto-organização de tamanho 7×7 .

Na fase de treinamento, 100 amostras das duas áreas distintas de texturas foram apresentadas aleatoriamente 10000 vezes para a rede neural. Os vetores de características eram constituídos de 60 elementos, o que significa que foram considerados intervalos de orientações de 6 graus.

Mesmo tendo utilizados apenas duas medidas para cada intervalo de orientação, os resultados da classificação foram bastante satisfatórios. Outras

medidas, como as dispersões e as entropias para cada intervalo de orientações também poderiam ter sido consideradas.

3.3.3 Determinação dos Picos do Espaço de Hough

A determinação dos picos relevantes do espaço de Hough é fundamental para se fazer uma detecção precisa dos segmentos de retas, pois eles indicam os posicionamentos e as orientações prováveis desses segmentos na imagem.

A técnica mais simples consiste em adotar um valor de limiar e considerar como sendo picos todos as células (θ, ρ) da matriz acumuladora que possuem valores maiores do que esse limiar. Essa técnica apresenta vários inconvenientes, pois faz com que segmentos curtos em orientações com baixa densidade de pixels sejam impossibilitados de serem detectados. Além disso, como a análise de conectividade considera uma faixa de largura 3 pixels (veja a seção 3.3.4), muitos falsos segmentos são detectados, comprometendo as técnicas que utilizam informações extraídas dos segmentos de retas detectados na imagem (veja a seção 3.3.1).

A Figura 3.9(b) apresenta os segmentos de retas detectados na imagem apresentada na Figura 3.9(a), adotando-se essa técnica para determinação dos picos. Note que, embora a detecção de retas tenha alcançado algum êxito, pois conseguiu depreender da imagem uma aproximação das bordas através de segmentos de retas, ela gerou uma imagem ruidosa (com muitos falsos segmentos tendo sido detectados).

Os falsos segmentos detectados através da técnica simples de limiarização são decorrência da utilização de falsos picos do espaço de Hough. A técnica de

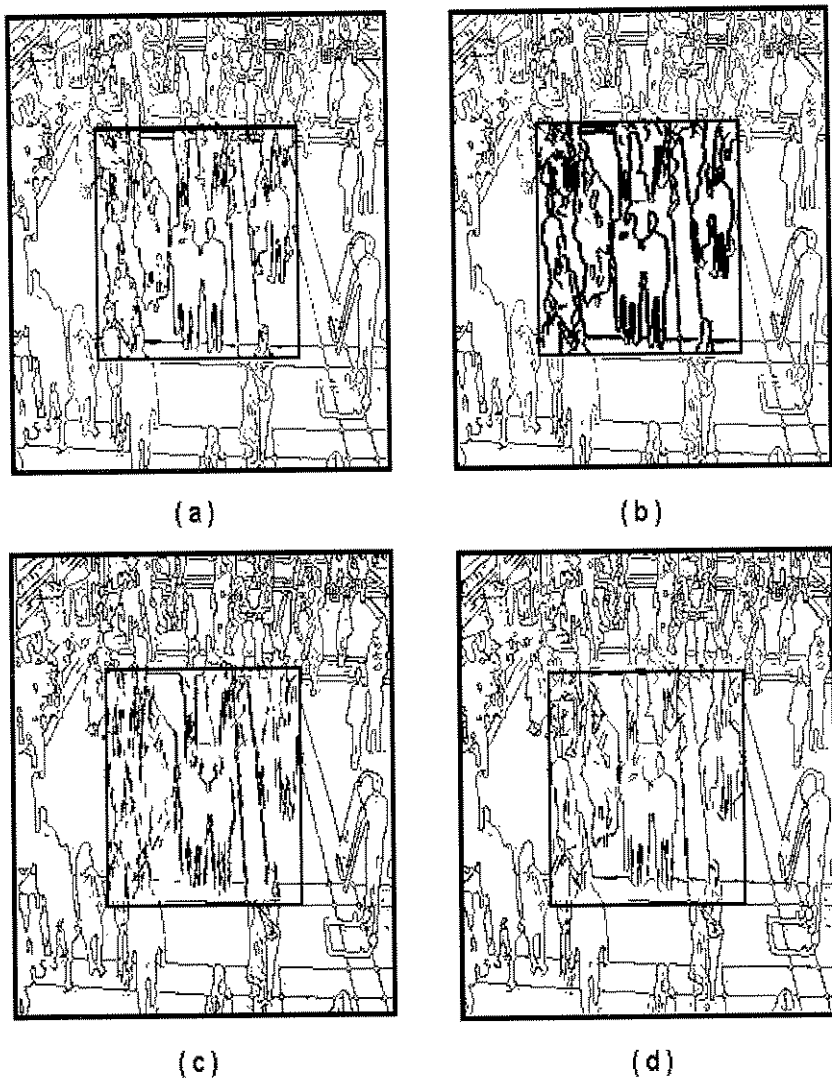


Figura 3.9: (a) Imagem com bordas detectadas; (b) Segmentos de reta detectados utilizando-se a técnica simples de limiarização; (c) Segmentos de reta detectados utilizando-se a técnica de “backmapping” com análise de conectividade; (d) Segmentos de reta detectados utilizando-se a técnica de remoção de pixels da imagem original.

“backmapping” [39] pode melhorar o desempenho do detector de segmentos de retas, uma vez que filtra o espaço de Hough, eliminando muitos falsos picos.

A técnica de “backmapping” pode ser resumida nos seguintes passos:

1. calcula-se a transformada de Hough da imagem;
2. para cada pixel de borda da imagem, percorre-se a senóide determinada por esse pixel no espaço de Hough, calculado no passo anterior, e determina-se as coordenadas (θ, ρ) do valor de máximo dessa senóide;
3. incrementa-se de 1 a posição (θ, ρ) , determinada no passo anterior, da matriz acumuladora auxiliar, que é inicializada com zeros e possui as mesmas dimensões da matriz acumuladora principal;

A matriz acumuladora auxiliar, que corresponde ao espaço de Hough filtrado, é limiarizada de forma análoga à técnica descrita anteriormente. As coordenadas (θ, ρ) que possuem valores maiores do que o limiar são tomados como sendo os picos do espaço de Hough e utilizados na detecção dos segmentos de retas.

Como por um ponto (x, y) “passam” infinitas retas, na técnica padrão da transformada de Hough, um pixel de borda contribui para incrementar os acumuladores (θ, ρ) de todas as retas que podem “passar” por esse pixel. Dessa forma, as orientações definidas por θ , com distâncias definidas por ρ , que apresentam uma grande densidade de pixels de borda, acumularão os valores de máximos das senóides do espaço de Hough.

Na técnica de “backmapping”, um pixel de borda contribui para incrementar apenas o acumulador correspondente à reta que “passa” por ele e cuja orientação θ e distância ρ possui a maior densidade de pixels de borda. Essa técnica tem a capacidade de reduzir sensivelmente o número de falsos picos. No entanto, ela apresenta um efeito colateral de eliminar os picos correspondentes a pequenos segmentos de retas em orientações pouco densas na imagem, o que pode ser indesejável em muitas aplicações.

Para amenizar essa falha, pode-se considerar uma análise de conectividade. Essa análise verifica se o pixel de borda em questão está conectado a seus pixels vizinhos na direção dada por θ , na qual encontrou-se o valor de máximo quando percorreu-se a senóide no espaço de Hough. Em caso afirmativo, incrementa-se o acumulador (θ, ρ) da matriz acumuladora auxiliar. Caso contrário, verifica-se a conectividade desse pixel na orientação referente ao segundo valor de máximo encontrado na senóide, e assim sucessivamente.

A Figura 3.9(c) apresenta os segmentos de retas detectados utilizando-se a técnica de “backmapping” com análise de conectividade. Pode-se notar uma diminuição considerável de falsos segmentos na boa aproximação obtida. Pode-se também notar que, embora tendo sido empregada a análise de conectividade, muitos pequenos segmentos de retas ainda foram descartados.

Eichmann [34] sugere que os picos locais do espaço de Hough sejam capturados da seguinte forma:

1. aplica-se um filtro de passa baixa na matriz acumuladora que armazena a transformada de Hough e armazena-se o resultado dessa operação numa matriz auxiliar;

2. subtrai-se da matriz acumuladora a matriz auxiliar calculada no passo anterior;
3. aplica-se a técnica de limiarização na matriz resultante da operação de subtração;

Utilizando-se um filtro de passa-baixa Gaussiano garante-se que não sejam gerados na imagem filtrada picos (ou vales) que não existem na imagem original [74]. A subtração permite dessa forma detectar as regiões onde encontram-se os pontos de máximos e mínimos locais. Quanto maior for a variância que define a Gaussiana, maiores serão essas regiões, gerando um número maior de falsos segmentos de retas detectados. Variâncias baixas tendem a eliminar picos locais pequenos e consequentemente tendem a não detectar alguns segmentos de retas.

Outra técnica para reduzir a detecção de falsos picos e consequentemente de falsos segmentos de retas é fazer uma limiarização em diversas escalas. Essa técnica consiste dos seguintes passos:

1. calcula-se a transformada padrão de Hough da imagem;
2. limiariza-se o espaço de Hough considerando-se um valor de limiar bastante alto;
3. detectam-se os segmentos de retas da imagem;
4. removem-se da imagem original os pixels de bordas correspondentes aos segmentos de retas detectados;

5. calcula-se a transformada padrão de Hough para a imagem obtida no passo anterior;
6. repetem-se os passos 2, 3, 4 e 5 até que o maior pico do espaço de Hough seja menor do que o menor comprimento de segmento de reta que se espera detectar na imagem;

A Figura 3.9(d) apresenta os segmentos de retas detectados utilizando-se a técnica de limiarização em diferentes escalas.

A técnica para detecção de picos relevantes em diversas escalas apresentou os melhores resultados, tanto do ponto de vista de eficácia na detecção dos pequenos segmentos de retas, quanto do ponto de vista de eliminação de falsos segmentos. A desvantagem dessa técnica com relação as outra apresentadas é o tempo de processamento, que é significativamente maior do que as demais, pois ela exige que a transformada de Hough seja calculada várias vezes.

3.3.4 Análise de Conectividade

A análise de conectividade objetiva verificar quais são os pontos da imagem que se encontram conectados, constituindo segmentos de retas. A fim de evitar que segmentos de retas desconectados por algum tipo de ruído não sejam detectados, a análise de conectividade considera conectados os pontos que estejam separados por uma distância menor do que um valor previamente estabelecido.

O algoritmo para análise de conectividade proposto por Costa [20], e que vem sendo utilizado nas técnicas apresentadas nessa tese para a extração de segmentos de retas, utiliza um vetor de “flags”. Durante a análise de

conectividade dos pixels na orientação dada por θ e distância dada por ρ , esse vetor sinaliza quais são as coordenadas que pertencem a algum segmento de reta.

Se os coeficientes angulares dos segmentos de retas sendo procurados são maiores do que um pixel, a busca é realizada na imagem variando-se os valores da ordenada y de 0 até o valor correspondente à altura da imagem. Para cada valor de y , o valor da abscissa x é calculado, utilizando-se a equação normal da reta (equação 3.1). Verifica-se então se o pixel (x, y) pertence a alguma borda. Caso ele não pertença, os seus pixels vizinhos $(x + 1, y)$ e $(x - 1, y)$ também são verificados. Se pelo menos um dos pixels pertencer a alguma borda, então a posição y do vetor de “flags” é marcada para sinalizar esse fato. Analogamente, se os coeficientes angulares dos segmentos de reta sendo procurados são menores do que 1, a busca é realizada variando-se o valor da abscissa x de 0 até o valor da largura da imagem. Para cada valor de x , o valor da ordenada y é calculado, utilizando-se a equação normal da reta (equação 3.1). Caso o pixel (x, y) não pertença a nenhuma borda, os seus pixels vizinhos $(x, y - 1)$ e $(x, y + 1)$ são verificados. Se pelo menos um dos três pixels verificados pertencer a alguma borda, então a posição x do vetor de “flags” é marcada para sinalizar esse fato.

Terminada a análise de conectividade referente aos parâmetros (θ, ρ) , o vetor de “flags” é varrido para se determinar as extremidades dos segmentos de retas.

Capítulo 4

Reconhecimento de Padrões

4.1 Classificação

O problema de tomada de decisões ocorre em inúmeras aplicações científicas e de engenharia. Normalmente, as informações disponíveis para essa tomada de decisão não são completamente precisas, fazendo com que os procedimentos de decisão sejam estatísticos por natureza. Assim, a área de reconhecimento digital de padrões tem estudado métodos de decisão baseados em estatísticas [31, 112].

O objetivo do reconhecimento de padrões é classificar objetos de interesse em classes ou categorias distintas. Os objetos de interesse são genericamente denominados de padrões.

Quando existe um conjunto de padrões para os quais as classes são conhecidas à priori, tem-se um problema de classificação supervisionada. O esquema básico da classificação supervisionada está ilustrado na Figura 4.1.

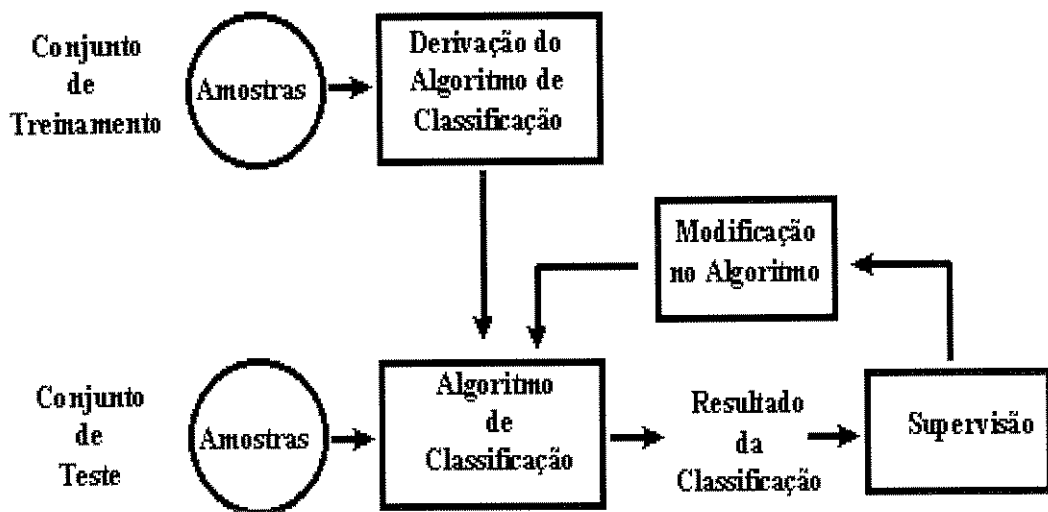


Figura 4.1: Esquema do reconhecimento de padrões supervisionado.

Uma parte dos padrões cujas classificações são previamente conhecidas são tomados como sendo o conjunto de treinamento, o restante constitui o conjunto de teste.

O conjunto de treinamento é utilizado para derivar o algoritmo de classificação, e o conjunto de teste é utilizado para avaliar o desempenho desse algoritmo. Daí decorre o termo de classificação supervisionada. Dependendo dos resultados obtidos, o algoritmo pode ser ajustado de tal modo a atingir um nível mínimo de erro na classificação.

Depois da fase de determinação de um algoritmo classificador que atinge níveis aceitáveis de erros na classificação, o algoritmo pode ser utilizado na classificação de padrões cujas classificações são desconhecidas.

Na classificação não-supervisionada não existem disponíveis padrões classificados a priori que possam auxiliar na proposição de um algoritmo para

classificação. Existem situações nas quais nem mesmo se sabe o número de classes existentes. Essa classificação requer uma análise de agrupamentos para se determinar as classes de padrões que possuem propriedades similares. O problema da classificação não supervisionada é, portanto, mais complexo e de tratamento mais difícil do que a supervisionada.

A Figura 4.2 ilustra um sistema típico de classificação de padrões. O vetor de observação x , obtido através de algum dispositivo de medida, é adequadamente transformado no vetor y , que é denominado *vetor de descritores* (“feature vector”) do padrão. Idealmente, os padrões observados devem ser reduzidos em um conjunto mínimo de descritores que contém as informações relevantes para as suas classificações.

O vetor y pode ser representado no espaço de descritores Y da mesma forma que o vetor x é representado no espaço das observações X . Entretanto, a dimensão do espaço de descritores é bem menor do que a dimensão do espaço de observações. No espaço dos descritores, pode-se projetar regras de decisão mais confiáveis para a classificação correta dos padrões.

No processo de reconhecimento de padrões o vetor de descritores y é passado para o classificador cujo propósito é tomar uma decisão sobre o padrão que ele descreve. O classificador essencialmente induz a partição do espaço de descritores em um número de regiões disjuntas. Se um dado vetor de descritores pertence a uma região R_i do espaço dos descritores, então o padrão correspondente àquele vetor é classificado como sendo da classe ω_i .

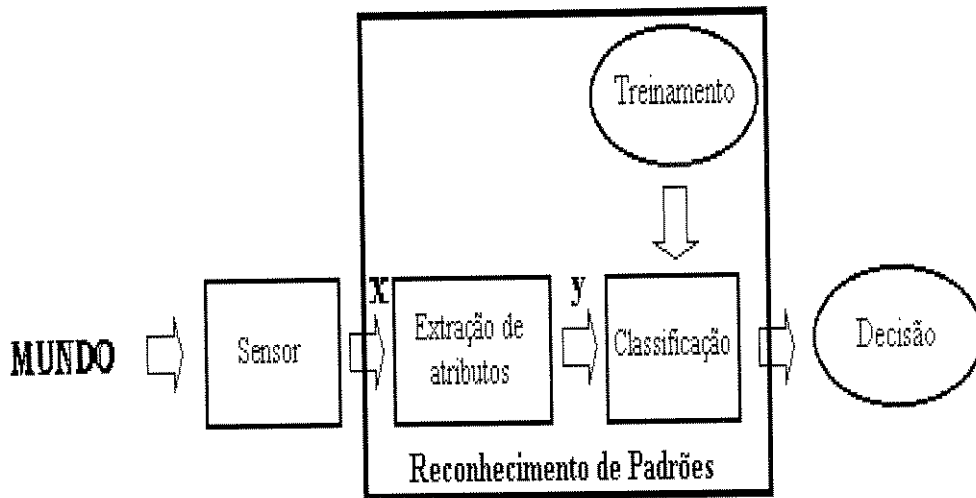


Figura 4.2: Esquema típico de classificação de padrões.

4.1.1 Classificador Estatístico Ótimo

É possível derivar uma estratégia de classificação ótima de padrões de tal modo que, na média, a probabilidade de se cometer erros seja mínima [31].

A probabilidade de um determinado padrão x pertencer à classe ω_i é denotada por $p(\omega_i|x)$. Se o classificador de padrões decide que x pertence à classe ω_j quando na verdade ele pertence à classe ω_i , ocorre um erro ponderado por L_{ij} . Como o padrão x pode pertencer a M classes, o erro médio cometido ao se classificar x como sendo da classe ω_j é dado por:

$$r_j(x) = \sum_{k=1}^M L_{kj} p(\omega_k|x). \quad (4.1)$$

A equação 4.1 é denominada *erro médio condicional* na teoria da decisão

e pode ser reescrita da seguinte forma:

$$r_j(x) = \frac{1}{p(x)} \sum_{k=1}^M L_{kj} p(x|\omega_k) P(\omega_k) \quad (4.2)$$

onde $p(x|\omega_k)$ é a função de densidade de probabilidade (FDP) dos padrões da classe ω_k . Como $\frac{1}{p(x)}$ é positivo e comum a todas as equações $r_j(x)$, para $j = 1, 2, \dots, M$, a equação 4.2 pode ser simplificada para:

$$r_j x = \sum_{k=1}^M L_{kj} p(x|\omega_k) P(\omega_k). \quad (4.3)$$

O classificador pode escolher, para cada padrão x com classificação desconhecida, M possíveis classes. Se ele calcula o erro médio $r_1(x), r_2(x), \dots, r_M(x)$, para cada padrão x e o classifica como sendo pertencente à classe com menor erro, então o erro médio cometido durante a classificação de todos os padrões desconhecidos será mínimo.

O classificador que utiliza essa estratégia é conhecido como *Classificador Bayesiano*.

Em muitos problemas de reconhecimento, $L_{ij} = 1$ se $i \neq j$ e $L_{ij} = 0$, se $i = j$. Nesses casos, a equação 4.3 pode ser reescrita da seguinte forma:

$$r_j(x) = p(x) - p(x|\omega_j)P(\omega_j) \quad (4.4)$$

Portanto, o classificador Bayesiano classifica o padrão x como pertencente à classe ω_i se:

$$p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j), j = 1, 2, \dots, M; i \neq j. \quad (4.5)$$

A função de decisão $d_j(x) = p(x|\omega_j)P(\omega_j)$ do classificador de Bayes depende do conhecimento prévio das FDPs $p(x|\omega_j)$ de cada classe ω_j , bem como das probabilidades de cada uma dessas classes ocorrer, $P(\omega_j)$.

A determinação das probabilidades de ocorrência das classes ω_j não constitui, em geral, um problema. Por exemplo, se todas as classes forem igualmente prováveis, então $P(\omega_j) = \frac{1}{M}$.

Um problema mais difícil consiste em determinar as FDPs de n variáveis das classe de padrões n -dimensionais. Na maioria das vezes, a única informação que se tem disponível sobre os padrões que se deseja classificar é um conjunto de amostras desses padrões. Nesses casos, podem ser adotadas dois tipos de técnicas para derivar um classssificado baseado no conjunto de amostras: técnicas paramétricas e técnicas não paramétricas [31].

Nas técnicas paramétricas supõe-se que os padrões a serem classificados apresentam FDPs conhecidas (por exemplo, supõe-se que elas sejam Gaussianas), e estima-se os parâmetros que definem essas funções, utilizando-se os conjuntos de amostras.

As técnicas não-paramétricas não fazem suposições sobre as FDPs e tentam estimá-las diretamente a partir dos conjuntos de amostras dos padrões, ou tentam estimar as probabilidades à posteriori, $P(\omega_i|x)$, dos padrões para cada classe ω_i . Algumas técnicas de classificações não-paramétricas procuram transformar o espaço de atributos em outros espaços, nos quais podem-se utilizar técnicas paramétricas de classificação.

Na prática, é comum utilizar as técnicas paramétricas e supor que as FDPs $p(x|\omega_j)$ são funções Gaussianas. Assumida essa hipótese, basta estimar os parâmetros que definem a FDP de cada classe, utilizando para tanto conjuntos de amostras dessas classes. O resultado da classificação depende nesses casos de quão próxima da realidade está a hipótese assumida.

No caso n -dimensional, a densidade Gaussiana dos vetores pertencentes

à j -ésima classe tem a forma:

$$p(x|\omega_j) = \frac{1}{(2\pi)^{n/2} |C_j|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_j)^T C_j^{-1} (x - \mu_j) \right] \quad (4.6)$$

onde μ_j é o vetor médio e C_j é a matriz de covariância, definidos como:

$$\mu_j = E_j\{x\} \quad (4.7)$$

e

$$C_j = E_j\{(x - \mu_j)(x - \mu_j)^T\} \quad (4.8)$$

onde $E\{.\}$ denota o valor esperado, n é a dimensão do vetor, e $|C_j|$ é o determinante da matriz C_j .

Os parâmetros μ_j e C_j que definem a FDP Gaussiana da classe ω_j podem ser estimados tomando-se um conjunto de amostras de padrões x pertencentes a essa classe, através da seguinte aproximação:

$$\mu_j = \frac{1}{N_j} \sum_{x \in \omega_j} x \quad (4.9)$$

e

$$C_j = \frac{1}{N_j} \sum_{x \in \omega_j} (xx^T - \mu_j \mu_j^T) \quad (4.10)$$

onde N_j é o número de amostras da classe ω_j .

A Figura 4.3 mostra um exemplo onde existem duas classes de padrões unidimensionais, ω_1 e ω_2 , cujas funções de densidade de probabilidades são dadas por duas funções Gaussianas. O ponto x_0 é o limite de decisão utilizado pelo classificador, caso as duas classes sejam igualmente prováveis de ocorrer, ou seja, $P(\omega_1) = P(\omega_2)$.

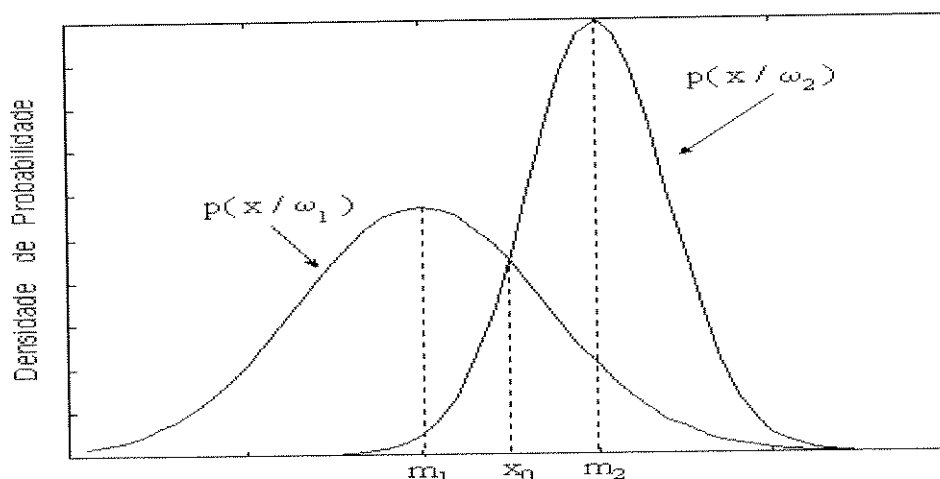


Figura 4.3: Funções unidimensionais de densidades de probabilidades para duas classes de padrões, ω_1 e ω_2 .

4.1.2 Redes Neurais

Na técnica de classificação estatística ótima, os parâmetros que definem as funções de densidade de probabilidades são estimados utilizando-se um conjunto de amostras de cada classe, para as quais já se conhece a classificação. Esses conjuntos são denominados *conjuntos de treinamento*.

Entretanto, as propriedades estatísticas das classes de padrões nem sempre são conhecidas e nem sempre podem ser adequadamente estimadas. Na prática, as técnicas não-paramétricas, que conseguem gerar as funções de decisões para a classificação diretamente dos conjuntos de treinamento sem precisar das FDPs, são mais úteis, uma vez que não estão sujeitas a determinadas suposições. Exemplos dessas técnicas são aquelas baseadas em redes

neurais.

Em geral, a utilização de redes neurais, comparando-se com outras técnicas, possui as seguintes vantagens [104]:

1. os relacionamentos funcionais entre os padrões de entrada e saída podem ser capturados por uma rede neural, sendo que tais relacionamentos não precisam ser conhecidos ou descritos explicitamente;
2. nenhuma hipótese precisa ser feita sobre as distribuições estatísticas dos padrões de entrada;
3. as redes neurais são tolerantes à falhas no sentido que, mesmo quando alguns elementos ou conexões do processamento apresentam-se degradados, a performance da rede pode ser pouco afetada.

As arquiteturas de redes neurais artificiais utilizadas para modelar os sistemas neurais biológicos podem ser divididas em três categorias distintas. A primeira categoria de redes neurais, “feedforward” transforma conjuntos de sinais de entrada em conjuntos de sinais de saída, sendo que os parâmetros dessa transformação são ajustados de forma supervisionada, de acordo com os resultados desejados. Na segunda categoria de redes neurais, “feedback”, as informações de entrada definem o estado inicial de atividade do sistema de realimentação. Depois de algumas transições de estado, atinge-se um estado que é considerado o resultado final da computação. Na terceira categoria, as células vizinhas da rede neural interagem mutuamente, a fim de transformarem-se adaptativamente em detectores especializados de diferentes padrões. Nessa última categoria, o aprendizado da rede neural é realizado de

forma não supervisionada, através de mapas de auto-organização [64].

4.1.2.1 Mapas de Auto-Organização

No modelo de redes neurais baseado em mapas de auto-organização (*Self-Organizing Map* - *SOM*), proposto por Kohonen [63, 64, 65], a segregação espacial de diferentes respostas e suas organizações em sub-conjuntos relacionados topologicamente resultam em um alto grau de eficiência nas operações típicas de uma rede neural.

O modelo SOM implementa uma projeção não-linear de um espaço multidimensional X em um espaço bidimensional M , denominado mapa de auto-organização. Essa projeção, de modo análogo à maioria das projeções encontradas no cérebro, implementam mapeamentos topológicos, ou seja, elementos vizinhos em X são geralmente mapeados em elementos vizinhos em M .

O mapeamento topológico constitui uma importante característica do modelo de redes neurais proposta por Kohonen, pois permite encontrar características e outras abstrações do espaço multidimensional através de identificação de agrupamentos (“clusters”) no mapa bidimensional.

O mapeamento do espaço multidimensional no bidimensional é realizado por uma função $f : X \rightarrow M$, onde $X \subset R^n$ e $M \subset R^2$, que associa a cada elemento $x \in X$ a um par $(i, j) \in M$. Os elementos m_{ij} do mapa M , assim como os dados de entrada, são vetores de tamanho n , que mantém os pesos das ligações sinápticas da rede neural.

Apresentado um estímulo específico $x \in X$ à rede neural, o “nó vencedor”

do mapa M para esse estímulo é determinado verificando-se qual dos elementos do mapa possui a menor distância ao estímulo, ou seja, o nó vencedor tem as coordenadas $(i, j) \in M$, tal que:

$$\|m_{ij} - x\| = \min\{\|m_{kl} - x\|, \forall i, j, k, l = 0 \dots D\} \quad (4.11)$$

onde D é o tamanho de M .

Portanto, um estímulo $x \in X$ apresentado à camada de entrada da rede neural é mapeado no elemento (i, j) da camada de saída da rede neural, ou seja do mapa de auto-organização M , para o qual a distância entre esse estímulo e o vetor dos pesos das ligações sinápticas m_{ij} da rede neural seja mínima, comparando-se com os demais vetores de pesos das ligações sinápticas m_{kl} .

Nesse modelo de rede neural, os neurônios da camada de saída interagem lateralmente com seus vizinhos. Essas interações se dão na fase de aprendizado, quando os pesos da célula “vencedora” e dos seus vizinhos são ajustados, de tal modo que a célula vencedora tenha pesos idênticos aos valores do estímulo sendo apresentado à rede num determinado instante do treinamento, e as células vizinhas tenham pesos semelhantes aos da célula vencedora, sendo que o grau de semelhança é ponderado de acordo com a distância entre a célula vizinha e a célula vencedora.

A interação da célula vencedora (i, j) com uma célula vizinha (k, l) , num determinado instante t do treinamento, é definida por uma função $h_{kl}(t)$ da distância entre elas. Normalmente, a função $h_{kl}(t)$ é uma Gaussiana.

O aprendizado nesse modelo é dito não supervisionado porque o mapeamento do espaço multidimensional para o bidimensional é feito a despeito

dos resultados esperados na saída do processamento, ao contrário dos outros modelos, onde os valores esperados e os valores obtidos pela rede, para um dado estímulo de entrada, influenciam no reajuste dos pesos das ligações sinápticas.

No modelo de redes neurais baseadas em mapas de auto-organização, os pesos das ligações sinápticas são ajustadas da seguinte forma:

$$m_{kl}(t+1) = m_{kl}(t) + (x - m_{kl}(t))h_{kl}(t) \quad (4.12)$$

onde $h_{kl}(t) = \alpha(t) \exp \left[-\frac{\|r_{kl} - r_{ij}\|^2}{\delta(t)^2} \right]$, (i, j) são as coordenadas do nó vencedor para o estímulo x , $\|r_{kl} - r_{ij}\|$ é a distância espacial entre o nó vencedor e o nó (k, l) , e $\delta(t)$ é uma função do tempo de treinamento t , que determina a variância da Gaussiana h_{kl} .

O tamanho da vizinhança $N_{ij}(t)$ decresce de acordo com o número de iterações T da fase de treinamento. No início do treinamento, isto é, quando $t = 0$, $N_{ij}(t)$ inclui toda a rede, e no final, quando $t = T$, $N_{ij}(t)$ contém somente o nó vencedor (i, j) . A seguinte função linear pode ser adotada para determinar os ajustes nas extensões da vizinhança: $N_{ij}(t) = D \left[1 - \frac{t}{T} \right]$, para $t = 0, \dots, T$.

De forma análoga, as funções $\delta(t)$ e $\alpha(t)$, podem ser reajustadas em cada iteração através das funções lineares $\delta(t) = 2 - \frac{t}{T}$ e $\alpha(t) = 1 - \frac{t}{T}$, para $t = 1, \dots, T$, respectivamente.

Após a fase de treinamento, os nós da rede são rotulados de acordo com a classe de estímulos para os quais eles apresentam respostas mais intensas. Se essa fase for supervisionada, então deve-se apresentar para a rede conjuntos de estímulos para os quais se conhece a classificação. Para cada classe de

estímulos conhecidos verifica-se quais são os nós vencedores no mapa de auto-organização. Esses nós são rotulados com o rótulo associado àquela classe de estímulos. Se essa fase não for supervisionada, é necessário fazer uma análise de agrupamentos no mapa de auto-organização.

4.2 Funções de Ajuste

O problema de reconhecimento de padrões normalmente resume-se em classificar um padrão desconhecido em uma classe dentre um conjunto finito de classes pré-definidas. Contudo, existem aplicações nas quais não existem classes pré-definidas e o que se deseja é estimar, para cada padrão, uma medida dentro de um intervalo de valores.

Se existe um conjunto de padrões para os quais se conhecem exatamente as medidas desejadas (conjuntos de treinamento), é possível utilizar esse conjunto e estimar os coeficientes de uma função de ajuste dos padrões conhecidos. Essa função pode ser utilizada posteriormente para estimar as medidas de padrões desconhecidos.

Funções polinomiais podem ser utilizadas como sendo as funções de ajuste. Quanto maior a ordem dos polinômios, melhores ajustes tendem a ser obtidos, entretanto, os polinômios de ordem superior possuem mais coeficientes, exigindo um conjunto maior de treinamento para serem determinados, e implicam em maiores oscilações na função ajustada.

O polinômio de ordem três para padrões bidimensionais, $x = [x_1 \ x_2]$, por

exemplo, é definido como:

$$\begin{aligned}
p(x) = p([x_1 \ x_2]) = & c_1x_1^3x_2^3 + c_2x_1^3x_2^2 + c_3x_1^3x_2 + c_4x_1^3 + \\
& c_5x_1^2x_2^3 + c_6x_1^2x_2^2 + c_7x_1^2x_2 + c_8x_1^2 + \\
& c_9x_1x_2^3 + c_{10}x_1x_2^2 + c_{11}x_1x_2 + c_{12}x_1 + \\
& c_{13}x_2^3 + c_{14}x_2^2 + c_{15}x_2 + c_{16}
\end{aligned} \tag{4.13}$$

Como pode ser visto na equação 4.13, os polinômios de ordem 3 de duas variáveis possuem 16 coeficientes $c_i, i = 1, 2, \dots, 16$, que precisam ser estimados a partir do conjunto de treinamento.

4.2.1 Estimação dos Coeficientes de Polinômios Cúbicos

Dado o polinômio da equação 4.13, suponha que sejam conhecidos os valores desse polinômio para um conjunto de N padrões bidimensionais $x_i = (x_{i1}, x_{i2})$, ou seja, são conhecidos os valores de $p(x_i), i = 1, \dots, N$.

Tem-se então o seguinte:

$$\begin{aligned}
p(x_1) = & c_1x_{1_1}^3x_{1_2}^3 + c_2x_{1_1}^3x_{1_2}^2 + c_3x_{1_1}^3x_{1_2} + c_4x_{1_1}^3 + \\
& c_5x_{1_1}^2x_{1_2}^3 + c_6x_{1_1}^2x_{1_2}^2 + c_7x_{1_1}^2x_{1_2} + c_8x_{1_1}^2 + \\
& c_9x_{1_1}x_{1_2}^3 + c_{10}x_{1_1}x_{1_2}^2 + c_{11}x_{1_1}x_{1_2} + c_{12}x_{1_1} + \\
& c_{13}x_{1_2}^3 + c_{14}x_{1_2}^2 + c_{15}x_{1_2} + c_{16} \\
p(x_2) = & c_1x_{2_1}^3x_{2_2}^3 + c_2x_{2_1}^3x_{2_2}^2 + c_3x_{2_1}^3x_{2_2} + c_4x_{2_1}^3 + \\
& c_5x_{2_1}^2x_{2_2}^3 + c_6x_{2_1}^2x_{2_2}^2 + c_7x_{2_1}^2x_{2_2} + c_8x_{2_1}^2 + \\
& c_9x_{2_1}x_{2_2}^3 + c_{10}x_{2_1}x_{2_2}^2 + c_{11}x_{2_1}x_{2_2} + c_{12}x_{2_1} + \\
& c_{13}x_{2_2}^3 + c_{14}x_{2_2}^2 + c_{15}x_{2_2} + c_{16}
\end{aligned}$$

$$\begin{aligned}
& \vdots \\
p(x_N) &= c_1 x_{N_1}^3 x_{N_2}^3 + c_2 x_{N_1}^3 x_{N_2}^2 + c_3 x_{N_1}^3 x_{N_2} + c_4 x_{N_1}^3 + \\
& c_5 x_{N_1}^2 x_{N_2}^3 + c_6 x_{N_1}^2 x_{N_2}^2 + c_7 x_{N_1}^2 x_{N_2} + c_8 x_{N_1}^2 + \\
& c_9 x_{N_1} x_{N_2}^3 + c_{10} x_{N_1} x_{N_2}^2 + c_{11} x_{N_1} x_{N_2} + c_{12} x_{N_1} + \\
& c_{13} x_{N_2}^3 + c_{14} x_{N_2}^2 + c_{15} x_{N_2} + c_{16}
\end{aligned} \tag{4.14}$$

O sistema apresentado em 4.14 possui 16 incógnitas (os coeficientes do polinômio cúbico) e N equações. Tal sistema pode ser representado na forma matricial:

$$\underbrace{\begin{bmatrix} p(x_1) \\ p(x_2) \\ \vdots \\ p(x_N) \end{bmatrix}}_P = \underbrace{\begin{bmatrix} x_{1_1}^3 x_{1_2}^3 & x_{1_1}^3 x_{1_2}^2 & \dots & x_{1_2} & 1 \\ x_{2_1}^3 x_{2_2}^3 & x_{2_1}^3 x_{2_2}^2 & \dots & x_{2_2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N_1}^3 x_{N_2}^3 & x_{N_1}^3 x_{N_2}^2 & \dots & x_{N_2} & 1 \end{bmatrix}}_X \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{16} \end{bmatrix}}_C \tag{4.15}$$

A matriz C dos coeficientes do polinômio pode ser determinada a partir da equação 4.15 da seguinte forma:

$$\begin{aligned}
P &= XC \implies \\
X^T P &= X^T X C \implies \\
[X^T X]^{-1} X^T P &= [X^T X]^{-1} X^T X C \implies \\
C &= [X^T X]^{-1} X^T P
\end{aligned} \tag{4.16}$$

A equação 4.16 consiste na formulação do método dos mínimos quadrados.

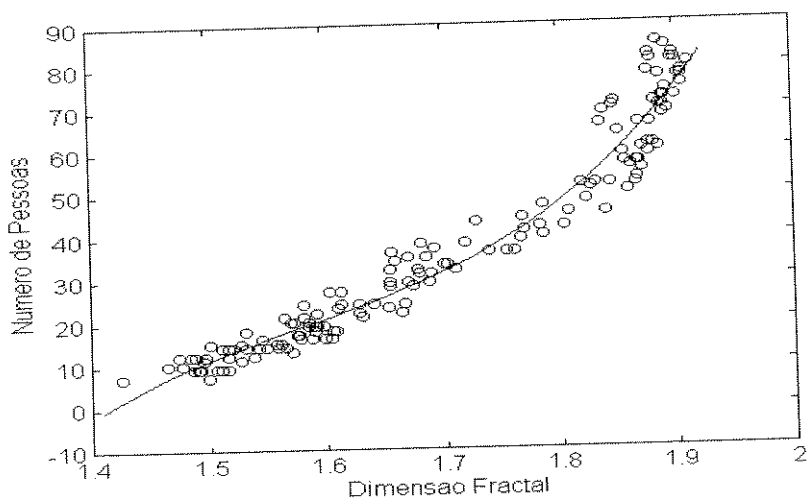


Figura 4.4: Função de ajuste polinomial cúbica unidimensional.

A Figura 4.4 apresenta um exemplo de função polinomial cúbica ajustada a um conjunto de pontos (x_{i1}, x_{i2}) , onde x_{i1} é o valor da amostra e x_{i2} é o valor de uma medida associada à amostra x_{i1} .

Após serem determinados os coeficientes que definem o polinômio $p(x)$, que fornece o valor de uma medida específica para os padrões, basta aplicá-lo nos padrões para os quais se deseja conhecer os valores da medida em questão.

Capítulo 5

Gerenciamento e Controle de Multidões

O gerenciamento e o controle de multidões é um problema crucial para a segurança dos seres humanos, uma vez que inúmeros acidentes podem ocorrer em áreas onde existem congestionamentos de pessoas, principalmente naquelas em que o congestionamento atinge níveis próximos ao limite máximo suportado [27, 93]. Muitos desses problemas podem ser prevenidos e rapidamente solucionados quando todos os aspectos envolvidos no gerenciamento e controle de multidões são bem planejados e executados.

O gerenciamento de multidões inclui medidas tomadas rotineiramente para facilitar a presença e a movimentação segura das pessoas em áreas onde ocorrem eventos de curta duração, como estádios de futebol e ginásios de esportes, ou em áreas onde grandes concentrações de pessoas são esperadas diariamente, como estações de ônibus, metrô, trens, etc. As medidas de

gerenciamento incluem procedimentos para a venda de ingressos, reserva de assentos, estacionamento de veículos, ingresso e egresso de pessoas, controle de ruído, comunicação, dentre outros.

O controle de multidões, por sua vez, diz respeito àquelas situações onde a massa de pessoas inicia uma reação a alguma ação ocorrida inesperadamente, buscando fugir de algum perigo eminente. As medidas de controle de multidões incluem procedimentos para restaurar a ordem, conduzindo a massa humana para uma nova situação de segurança.

Segundo Berlonghi [7], o gerenciamento de multidões é uma tarefa preventiva (“proactive”), enquanto que o controle é uma tarefa curativa (“reactive”).

Dois aspectos importantes para o gerenciamento e o controle adequados de multidões são o projeto de ambientes seguros, onde se espera a ocorrência de congestionamentos de pessoas, e o monitoramento em tempo real das multidões dentro das estruturas já construídas e em funcionamento [23].

No que diz respeito ao primeiro aspecto, o desenvolvimento de modelos do comportamento de multidões pode fornecer o embasamento necessário para auxiliar os arquitetos, bem como os demais profissionais ligados a essa área, a projetar construções mais seguras. Ao rever a psicologia do comportamento de multidões em termos de seu relacionamento com as áreas de Engenharia e Segurança, Sime [107] enfatiza a necessidade da validação de simulações computadorizadas da movimentação e do comportamento das multidões fugindo de situações de risco, baseadas em critérios da Engenharia e da Psicologia.

Com relação ao segundo aspecto, a monitoração de multidões em tempo real vem sendo realizada na prática com o uso extensivo de circuitos fechados

de televisão. Os monitores de televisão são constantemente observados por supervisores humanos a procura de situações anormais que podem acarretar problemas para a segurança das pessoas. Detectada uma situação de perigo, uma prática normal é controlar o acesso de novas pessoas às áreas de risco, fechando-se os portões de entrada, e conduzir em segurança as pessoas que se encontram nessas áreas para áreas mais seguras. Entretanto, o fechamento de acesso a determinadas áreas pode levar ao congestionamento de novas áreas que também precisarão por sua vez serem controladas, dificultando assim a tarefa de controle.

A monitoração de áreas interligadas (como as que caracterizam as estações de metrô, por exemplo) requerem um número enorme de câmeras de vídeo e, conseqüentemente, um número também enorme de pessoas para exercerem a tarefa de supervisão. O grande número de câmeras de vídeo apresenta inconvenientes, como por exemplo a armazenagem das imagens gravadas. Por razões de custo operacional, normalmente um único supervisor é responsável pelo monitoramento simultâneo de várias áreas, através de um painel de monitores de televisão.

No caso de eventos de curta duração, pode-se esperar um alto grau de concentração dos supervisores humanos na tarefa de monitoramento. Todavia, nas situações cotidianas, como o monitoramento de aeroportos, estações de metrô, ônibus ou trens, por exemplo, é bastante provável que supervisores humanos percam a concentração, devido a monotonia dessa tarefa. A solução para tornar mais confiável e reduzir os custos operacionais da tarefa de monitoramento em tempo real, principalmente em situações rotineiras, é indubitavelmente a automação dessa tarefa.

Um dos problemas envolvidos na tarefa de monitoramento automático de multidões é a estimação automática do número de indivíduos presentes num congestionamento de pessoas. Esse problema é crucial, uma vez que quando a densidade de pessoas em determinada área excede o limite para o qual essa área foi projetada, a chance de ocorrer acidentes torna-se muito alta [93], sendo às vezes inevitável.

5.1 Estimação dos Níveis de Congestionamento Baseada nos Pixels das Imagens

Davies et al [23] propõem uma técnica para estimar automaticamente o número de pessoas através do uso de imagens captadas por câmeras de vídeo nas áreas sob monitoramento. Nessa técnica, duas medidas são extraídas das imagens. A primeira diz respeito ao número de pixels que compõem os objetos da cena (“foreground”). A segunda medida diz respeito ao número de pixels pertencentes às bordas dos objetos da cena.

Nos casos onde existem poucas pessoas (em torno de 30, no máximo), foram verificadas as existências de relacionamentos lineares entre o número de pessoas na cena e o número de pixels do “foreground” e das bordas dos objetos. Ou seja, aumentando-se o número de pessoas na cena, aumenta-se linearmente o número desses pixels. Portanto, foi possível combinar essas duas medidas num estimador ótimo de densidade de pessoas através de um filtro linear de Kalman [11].

Testada em imagens capturadas em intervalos de 10 segundos, a técnica de

estimação automática de pessoas descrita acima resultou em um erro relativo de menos de 8%, comparando-se com medidas realizadas manualmente [23].

Infelizmente, o sucesso da técnica baseada no número de pixels de “foreground” e de bordas restringe-se às situações onde um número pequeno de pessoas são esperadas na cena. O relacionamento linear observado nesses casos, dificilmente se mantém quando existem congestionamentos de pessoas, pois as imagens desses congestionamentos apresentam sobreposições das imagens das pessoas, inviabilizando as medidas propostas. O efeito das sobreposições das imagens das pessoas poderia ser reduzido caso as câmeras fossem instaladas no teto das áreas sob monitoração, perpendiculares ao chão. No entanto, isso acarretaria no aumento do número de câmeras necessárias para o monitoramento da mesma área, além de ser muitas vezes inviável, devido às alturas inadequadas de muitos ambientes já construídos. A Figura 5.1 ilustra um caso para o qual a técnica de Davies não obteria êxito na estimação do número de pessoas. Comparando-se as duas imagens apresentadas mais à direita da Figura 5.1, é possível notar que houve um aumento no número de pessoas da quarta para a quinta imagem. Por outro lado, devido à limitações das técnicas de detecção de bordas, ocorreu uma sensível diminuição no número de pixels de bordas detectados na quinta imagem, com relação à quarta imagem.

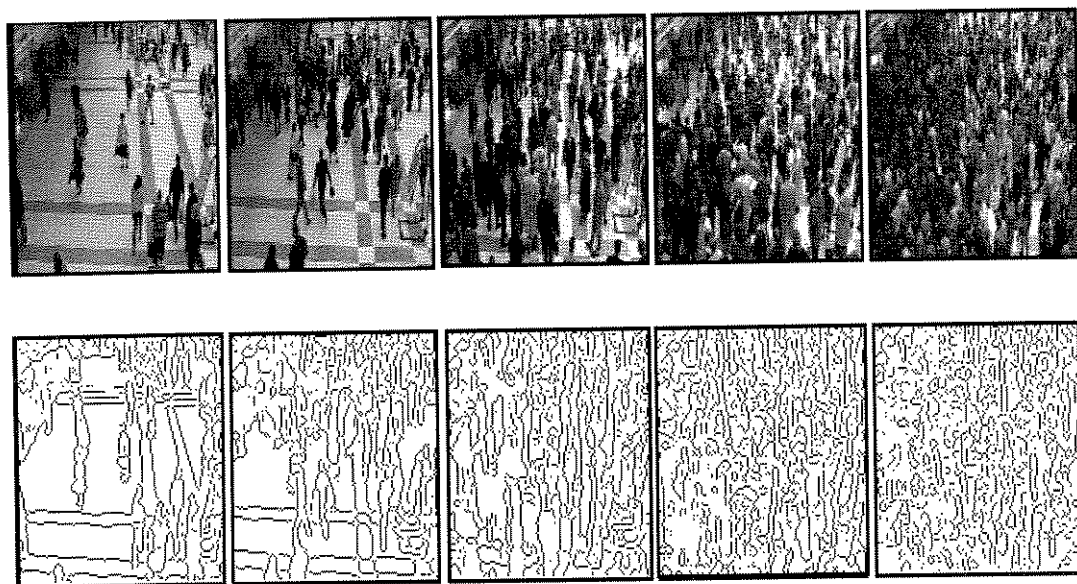


Figura 5.1: Imagens de congestionamentos de pessoas de diferentes densidades (linha superior) com suas bordas detectadas (linha inferior). Da esquerda para a direita, as rugosidades dos padrões de textura aumentam, conforme aumentam as densidades dos congestionamentos de pessoas. Pode-se perceber que as quantidades de pixels de bordas não aumentam linearmente com os aumentos dos números de pessoas na cena. Na imagem mais a direita da linha inferior é possível notar uma discreta redução do número de pixels de borda detectados, comparando-se com a imagem anterior da sequência.

5.2 Estimação dos Níveis de Congestionamento Baseada nas Texturas das Imagens

A tarefa de estimar o número exato de pessoas em uma multidão não é, em geral, possível, nem mesmo para observadores humanos. Quando se deseja gerenciar e controlar multidões, geralmente é suficiente estimar um número aproximado de pessoas, não sendo necessário conhecer o número exato.

Observando-se imagens de multidões, como as mostradas na Figura 5.1, pode-se notar padrões de texturas que variam de acordo com as densidades dessas multidões. As imagens de multidões de altíssimas densidades tendem a apresentar texturas mais finas, com os tons de cinza variando rapidamente em porções relativamente pequenas da imagem. Por outro lado, as imagens de multidões de baixas densidades tendem a apresentar padrões de texturas mais grossas e uniformes, com poucas variações dos tons de cinza em áreas relativamente grandes da imagem. A Figura 5.1 (linha superior) exhibe uma sequência de cinco imagens de aglomerações de pessoas com diferentes níveis de congestionamento. Da esquerda para a direita, as densidades dos congestionamentos aumentam, modificando os padrões de textura das imagens, tornando-os cada vez mais finos.

Pode-se também verificar na Figura 5.1 que a intensidade direcional (seção 3.1) é maior na imagem central da linha inferior, uma vez que ela é constituída por várias bordas compridas dispostas verticalmente. Nas duas imagens mais à esquerda, as intensidades direcionais são menores do que

a intensidade direcional da imagem central, pois, embora as bordas que as constituem também sejam compridas, várias delas estão dispostas horizontalmente, aumentando a dispersão angular das bordas. Nas duas imagens mais à direita, além das dispersões angulares das bordas também serem consideráveis, as bordas são curtas, contribuindo para diminuir suas intensidades direcionais.

As observações acima motivaram as duas proposições desta tese para o problema de gerenciamento automático de multidões, que são:

1. avaliar a eficácia de medidas de texturas para a estimação automática das densidades de congestionamentos de pessoas;
2. avaliar a relevância das informações de direcionalidade nas texturas das imagens de congestionamentos de pessoas.

Para atingir os objetivos propostos, foram estabelecidas duas estratégias distintas para estimar as densidades de congestionamentos de pessoas utilizando-se medidas de texturas extraídas das imagens.

A primeira estratégia classifica os pixels da imagem, utilizando medidas de texturas, extraídas nas vizinhanças desses pixels [83]. Em seguida, medidas estatísticas de primeira ordem dos pixels classificados (histogramas de classes) são utilizadas para determinar os níveis de congestionamentos de pessoas na imagem. Esta estratégia está descrita com mais detalhes na seção 5.2.1.

A segunda estratégia extrai medidas globais de texturas, considerando toda a imagem de uma única vez [84, 85]. Essas medidas são utilizadas diretamente por um classificador para estimar os níveis de congestionamentos

de pessoas na imagem. Esta estratégia está descrita com mais detalhes na seção 5.2.2.

Pode-se encontrar na literatura inúmeras técnicas computacionais para extrair medidas de texturas de uma imagem (seção 2.2). Desse conjunto de técnicas, foram escolhidas quatro para serem avaliadas nesta tese:

Matriz de Co-ocorrência dos Tons de Cinza: esta técnica pertence à categoria das técnicas estatísticas (seção 2.2.1), e encontra-se descrita na seção 2.2.1.1. Teoricamente, ela não se destina a análise de texturas orientadas, porém, como na prática é necessário fixar um conjunto finito de orientações para as quais são calculadas as matrizes de dependências, as medidas obtidas a partir dessas matrizes estão relacionadas com as direcionalidades dos padrões que formam as texturas. Esta técnica é frequentemente adotada para a análise de texturas nas mais diversas aplicações, mesmo apresentando o inconveniente de ser baseada em estatísticas de segunda ordem, as quais são consideradas insuficientes para o processamento de visão no SVH.

Transformada de Hough: esta técnica, que baseia-se nos segmentos de retas extraídos das imagens através da transformada de Hough, pertence à categoria das técnicas estruturais de análise de texturas (seção 2.2.2). Ela foi originalmente proposta nesta tese para a análise de texturas orientadas baseada no conceito de intensidade direcional dos padrões que compõem as texturas, e encontra-se descrita na seção 3.3.1.2.

Espectro de Fourier: esta técnica pertence à categoria das técnicas es-

tatísticas para análise de texturas (seção 2.2.1) e encontra-se descrita na seção 2.2.1.5. Alguns autores consideram as técnicas para análise de texturas baseadas no espectro de Fourier como constituindo uma categoria independente de técnicas [40]. Além de fornecer informações sobre as direcionalidades dos elementos que compõem as imagens, o espectro de Fourier informa sobre as frequências espaciais e os tamanhos desses elementos, sendo que os grandes, característicos das texturas grossas, estão relacionados às baixas frequências, ao passo que os pequenos, característicos das texturas finas, estão relacionados com as altas frequências.

Fractais: esta técnica pertence à categoria das técnicas para análise de texturas baseadas em modelos (seção 2.2.3). As dimensões fractais são estimadas utilizando-se as salsichas de Minkowski apresentadas na seção 2.2.3.4. Esta técnica não leva em consideração as informações sobre direcionalidade dos elementos que compõem as imagens.

Das técnicas propostas para as análises de texturas, três são consagradas na literatura (matrizes de co-ocorrências, Fourier e fractais) e uma foi proposta originalmente nesta tese (Hough). Três destas técnicas consideram relevantes as informações de direcionalidade dos componentes das imagens (matrizes de co-ocorrências, Hough e Fourier), enquanto que para uma delas essa informação é irrelevante (fractais). Todas as abordagens para a análise de texturas (estatística, estrutural e baseada em modelos) são representadas, uma vez que as técnicas propostas para a análise de texturas cobrem todas as abordagens.

Nível de Serviço	Densidade de Pedestres ($peds/m^2$)
A: Fluxo Livre (normal)	≤ 0.6
B: Fluxo Restrito	$0.6 - 0.75$
C1: Fluxo Denso	$0.75 - 1.25$
C2: Fluxo Muito Denso	$1.25 - 2.0$
D: Fluxo Extremamente Denso	> 2.0

Tabela 5.1: Definição de Níveis de Serviço.

As duas estratégias para a estimação de densidades de multidões adotadas nesta tese (seções 5.2.1 e 5.2.2), utilizam classificadores para atribuir as densidades de congestionamentos de pessoas às classes previamente definidas, que são cinco: baixíssima densidade (VL), baixa densidade (L), densidade moderada (M), alta densidade (H) e altíssima densidade (VH).

A classificação de imagens em classes de densidades é adotada na literatura. Polus [99], por exemplo, utiliza as seguintes classes, as quais ele define como sendo *níveis de serviço*: fluxo livre (A), fluxo restrito (B), fluxo denso (C1), fluxo muito denso (C2) e fluxo extremamente denso (D). A Tabela 5.1 apresenta os índices de congestionamentos ($pedestres/metro^2$) estabelecidos por Polus para definir os cinco níveis de serviços por ele propostos.

Nas estratégias para estimação de densidades de multidões propostas nesta tese, as imagens são atribuídas a uma das classes de densidades (VL,L,M,H e VH) utilizando-se os classificadores descritos no Capítulo 4: o classificador estatístico ótimo, o classificador neural e o classificador baseado

em funções de ajuste.

O classificador neural, implementado de acordo com o modelo de auto-organização proposto por Kohonen (seção 4.1.2), foi adotado pois, além do suporte biológico desse modelo, ele propiciou bons resultados nos experimentos realizados no *Vision and Robotics Laboratory*, do *King's College London*, Londres, Reino Unido, para a classificação de padrões (dígitos) utilizando-se a transformada de Hough, em comparação com resultados obtidos com redes neurais implementadas de acordo com o modelo de “backpropagation” [13].

Objetivando comparar os resultados obtidos com o classificador neural, o classificador Bayesiano e o classificador baseado em funções de ajuste também são utilizados. O Classificador Bayesiano é o mais tradicional dos classificadores e fornece solução ótima, desde que as funções de densidade sejam perfeitamente conhecidas.

O Capítulo 6 apresenta os resultados obtidos quando as estratégias de estimação de densidades de pessoas descritas neste capítulo foram aplicadas numa situação real, com imagens capturadas na estação de trens da Liverpool Street, em Londres, Reino Unido.

As próximas sub-seções apresentam mais detalhes das estratégias propostas nesta tese para a estimação de densidades de congestionamentos de pessoas utilizando-se medidas de texturas.

5.2.1 Medidas Locais de Textura

A Figura 5.2 mostra um diagrama de blocos da técnica que utiliza medidas locais de texturas da imagem (no caso, extraídas de quatro matrizes de

dependência de tons de cinza).

O primeiro passo dessa técnica consiste em detectar as bordas e binarizar as imagens, caso estejam sendo empregadas as técnicas de análise de textura baseadas na transformada de Hough ou nas dimensões fractais, pois estas exigem imagens binárias de entrada.

No passo seguinte, todos os pixels da imagem de entrada são classificados em uma das classes de textura previamente identificadas. Essa classificação é realizada com o auxílio de uma janela quadrada móvel sobre a imagem, que define as vizinhanças dos pixels onde são extraídas as medidas locais de texturas. Essas medidas locais de texturas constituem os vetores de características de entrada do classificador responsável pela classificação dos pixel da imagem.

Em seguida, são calculados histogramas das ocorrências das classes de texturas na imagem classificada. Tais histogramas são utilizados na fase seguinte como vetores de características de entrada do classificador que determina, finalmente, uma das classes de densidades de multidões para a imagem que está sendo analisada.

A Figura 5.3 apresenta exemplos de histogramas de classes de texturas de algumas imagens de multidões e a relação existente entre os perfis desses histogramas com os níveis de congestionamento de pessoas. Nesse exemplo, foram utilizadas seis classes distintas de texturas, associadas aos tons de cinza 36, 72, 108, 144, 216 e 255, que são proporcionais aos graus de congestionamentos indicados pelas texturas a eles associados. A ocorrência da classe rotulada pelo tom de cinza 36 indica um baixíssimo grau de congestionamento, enquanto que a ocorrência da classe rotulada pelo tom 216

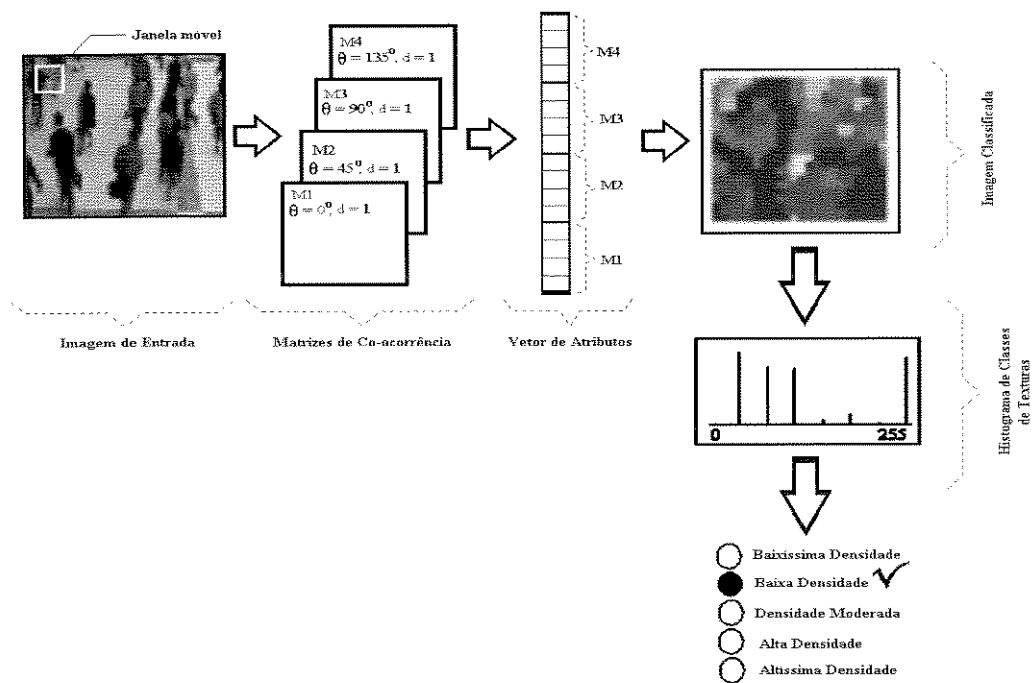


Figura 5.2: Diagrama da técnica para estimação de densidade de multidões baseada em medidas locais de texturas.

indica um altíssimo grau de congestionamento (no exemplo, o tom de cinza 0 indica os pixels que não foram classificados como pertencentes a nenhuma das classes de textura e o tom 255 indica os pixels da borda da área classificada). Dessa forma, as imagens classificadas associadas àquelas com altos níveis de congestionamentos de pessoas tendem a ser claras, produzindo modas mais à direita dos histogramas, ao contrário das imagens classificadas associadas às imagens com poucas pessoas, cujas modas tendem a situar-se mais à esquerda dos histogramas.

Esta estratégia, que utiliza medidas locais de textura, necessita, portanto, de dois classificadores. O primeiro classifica os pixels da imagem de entrada e o segundo classifica os histogramas das classes de texturas da imagem classificada em classes de densidades de congestionamentos de pessoas previamente definidas.

5.2.2 Medidas Globais de Textura

Ao contrário da técnica baseada em medidas locais de texturas, que são calculadas com o auxílio de uma janela quadrada móvel numa vizinhança da imagem, a técnica apresentada nesta seção calcula medidas globais de textura sobre toda a imagem.

A Figura 5.4 apresenta um diagrama da técnica que emprega medidas globais de textura sobre toda a imagem. O diagrama exemplifica a utilização da técnica de matrizes de co-ocorrência descrita na seção 2.2.1.1. No exemplo, são extraídas quatro medidas de textura de quatro matrizes de co-ocorrências, totalizando 16 medidas, que são utilizadas diretamente pelo

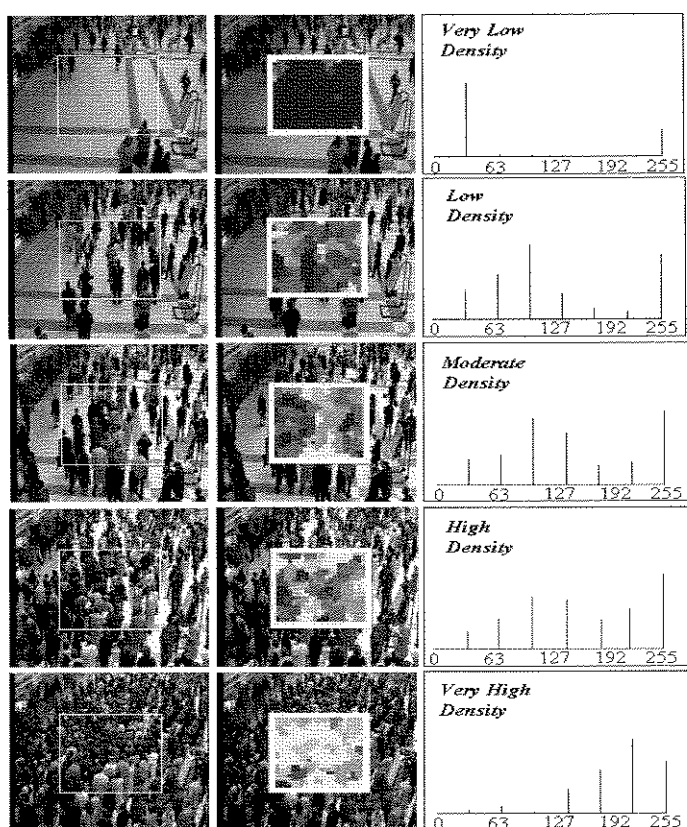


Figura 5.3: Relacionamento entre a densidade de multidões e o formato dos histogramas de classes de texturas. Na coluna da esquerda estão imagens com as áreas sob monitoramento definidas por janelas quadradas centralizadas na imagem. Na coluna central estão as imagens classificadas. Na coluna da direita estão os histogramas de classes das imagens classificadas, onde os níveis baixos de tons de cinza representam baixos níveis de congestionamentos de pessoas, enquanto que os altos níveis de tons de cinza representam altos níveis de congestionamentos de pessoas. Os histogramas estão apresentados normalizados com relação à maior moda.

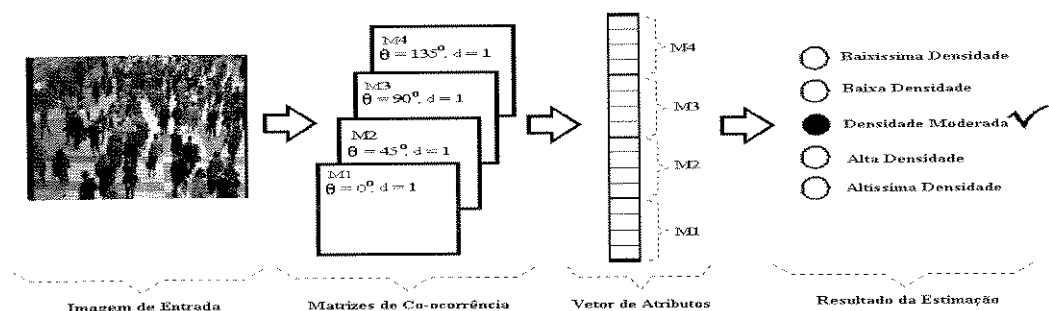


Figura 5.4: Diagrama da técnica para estimação de densidade de multidões baseada em medidas globais de texturas.

classificador como o vetor de atributos para classificar diretamente a imagem de entrada em uma das classes de multidões previamente definidas.

A grande vantagem dessa técnica, comparada com a anterior, é o seu melhor desempenho com relação ao tempo de processamento, que lhe permite ser aplicada em problemas que exigem soluções em tempo-real.

Nesta técnica, primeiramente são calculadas as medidas globais de textura das imagens do conjunto de treinamento. Essas medidas são tomadas como sendo os vetores de atributos que são utilizados para treinar a rede neural dos classificadores neurais, ou determinar os parâmetros das funções de densidades de probabilidades dos classificadores Bayesianos ou determinar os coeficientes dos polinômios cúbicos das funções de ajuste.

Em seguida, determinados os parâmetros do classificador Bayesiano, os pesos sinápticos do classificador neural, ou os coeficientes da função de ajuste, pode-se determinar as densidades de congestionamentos de pessoas em ima-

gens desconhecidas, através de suas classificações em uma das classes de densidades previamente estabelecidas.

Caso estejam sendo utilizadas as medidas de texturas baseadas na transformada de Hough ou nas dimensões fractais, o primeiro passo desta estratégia global, assim como na estratégia local, consiste em detectar as bordas e binarizar a imagem de entrada.

Capítulo 6

Aplicação das Técnicas Baseadas em Textura para a Estimação de Densidades de Multidões

Neste capítulo são apresentados os resultados obtidos com a aplicação das técnicas baseadas em medidas de textura para a estimação dos níveis de congestionamentos de pessoas. Esses experimentos foram realizados sobre um conjunto de 291 imagens capturadas na Liverpool Street Train Station, em Londres, Reino Unido.

O objetivo é classificar as densidades das multidões registradas pelas imagens em cinco classes distintas de densidades de multidões: baixíssima densidade (VL), baixa densidade (L), densidade moderada (M), alta densi-

dade (H) e altíssima densidade (VH).

A Figura 6.1 apresenta algumas das imagens utilizadas nos experimentos, pertencentes às classes distintas de densidades de congestionamentos de pessoas (Figura 6.1(a) Baixíssimas densidades; Figura 6.1(b) Baixas densidades; Figura 6.1(c) Densidades moderadas; Figura 6.1(d) Altas densidades; Figura 6.1(e) Altíssimas densidades).

Através de contagens manuais, foram calculados os números aproximados de pessoas presentes em cada uma das imagens utilizadas nos experimentos. Estas estimativas estão apresentadas na Tabela A.1, do Apêndice A. A Figura 6.2 apresenta exemplos de estimativas manuais cada uma das classes de densidades de congestionamentos de pessoas.

As Figuras 6.3(a) e 6.3(b) apresentam os histograms dos números de pessoas nas imagens dos grupos de treinamento e teste, respectivamente.

As áreas centrais das imagens utilizadas nos experimentos, onde foram calculadas as medidas de texturas, cobrem uma área de aproximadamente $18m^2$. Portanto, os níveis de serviços propostos por Polus [99], apresentados na Tabela 5.1, para essas áreas são os seguintes:

Fluxo livre: de 0 a 10,8 pessoas;

Fluxo restrito: de 10,8 a 13,5 pessoas;

Fluxo denso: de 13,5 a 22,5 pessoas;

Fluxo muito denso: de 22,5 a 36 pessoas;

Fluxo extremamente denso: acima de 36 pessoas.



Figura 6.1: Conjunto de imagens de congestionamentos de pessoas. (a) Baixíssimas densidades; (b) Baixas densidades; (c) Densidades moderadas; (d) Altas densidades; (e) Altíssimas densidades.

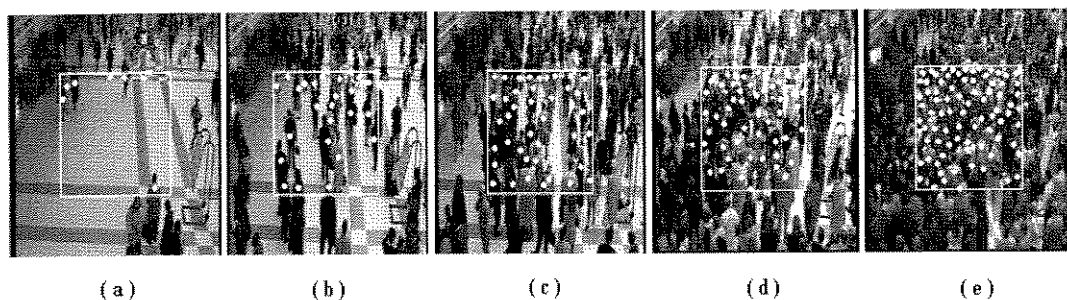


Figura 6.2: Exemplos de densidades de multidões estimadas manualmente. (a) imagem #76 (7 pessoas) - Baixíssima Densidade; (b) imagem #27 (24 pessoas) - Baixa Densidade; (c) imagem #32 (40 Pessoas) - Densidade Moderada; (d) imagem #40 (52 pessoas) Alta Densidade; (e) imagem #280 (82 pessoas) Altíssima Densidade.

Como as contagens manuais dos números de pessoas nas imagens disponíveis para realização dos experimentos mostraram a existência de várias imagens com mais de 36 pessoas, e havia interesse na diferenciação de sub-níveis de congestionamentos acima desse patamar, decidiu-se definir intervalos de densidades de congestionamentos distintos daqueles definidos pelos níveis de serviço propostos por Polus. Foram, então, definidos os seguintes intervalos de densidades de multidões:

Baixíssima Densidade (Very Low - VL): 0 a 15 pessoas;

Baixa Densidade (Low - L): 16 a 30 pessoas;

Densidade Moderada (Moderate - M): 31 a 45 pessoas;

Alta Densidade (High - H): 46 a 60 pessoas;

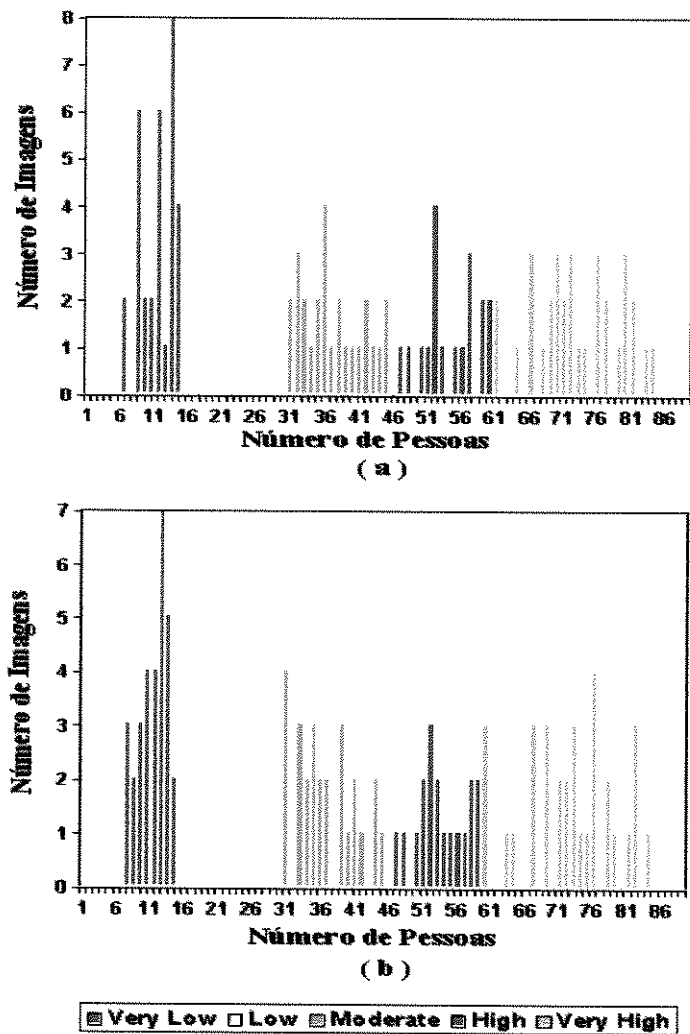


Figura 6.3: (a) Histograma dos números de pessoas nas imagens do conjunto de treinamento; (b) Histograma dos números de pessoas nas imagens do conjunto de teste.

Classes	número de Pessoas	Densidade	Nível de Serviço
Very Low	0 – 15	0 – 0.833	A,B
Low	16 – 30	0.89 – 1.67	C1
Moderate	31 – 45	1.72 – 2.5	C2,D
High	46 – 60	2.56 – 3.33	D
Very High	> 60	> 3.39	D

Tabela 6.1: Relacionamento entre as classes de densidades de congestionamentos e os níveis de serviços.

Altíssima Densidade (Very High - VH): 61 ou mais pessoas.

Com esses novos intervalos pode-se distinguir situações distintas dos fluxos congestionados definidos por Polus. Por exemplo, pode-se distinguir entre congestionamentos contendo até 45 pessoas, daqueles contendo de 45 até 60 pessoas e daqueles contendo mais de 60 pessoas.

A Tabela 6.1 mostra o relacionamento entre as cinco classes de densidades de multidões utilizadas nos experimentos (VL, L, M, H e VH) e os cinco níveis de serviços definidos por Polus (A, B, C1, C2 and D).

As imagens foram divididas em dois grupos: conjunto de treinamento e conjunto de teste. As imagens do conjunto de treinamento foram utilizadas para estimar os parâmetros das FDPs dos classificadores Bayesianos, treinar as redes neurais e determinar os coeficientes das funções de ajuste. As imagens do conjunto de teste serviram para avaliar a eficácia das técnicas de classificação empregadas. O número de imagens dos grupos de treinamento

Conjunto	VL	L	M	H	VH	Total
Treinamento	31	39	26	18	32	146
Teste	30	39	26	18	32	145

Tabela 6.2: Número de imagens dos conjuntos de treinamento e de teste nas cinco classes de densidades (VL,L,M,H e VH).

e de teste estão apresentadas na Tabela 6.2.

6.1 Medidas Locais de Texturas

Foi verificado, durante os experimentos realizados, que o processo de dupla classificação (classificação de texturas, seguida da classificação dos histogramas de classes de texturas) requerido pela técnica baseada em medidas locais de texturas (seção 5.2.1) para a estimação automática de densidades de multidões, inviabiliza sua aplicação em tempo-real, principalmente pelo fato da classificação local das texturas ser um processo bastante lento. Portanto, essa técnica deve ser aplicada preferencialmente quando são fundamentais as análises locais que não precisam ser efetuadas em tempo real, ou que possam ser executadas por sistemas distribuídos.

Os experimentos realizados com essa técnica restringiram-se ao uso de matrizes de co-ocorrência (seção 2.2.1), com um classificador neural implementado segundo o modelo de auto-organização (seção 4.1.2).

6.1.1 Matrizes de Co-ocorrência com Classificador Neural

A técnica de análise de texturas baseada em medidas das matrizes de co-ocorrências, também conhecidas como matrizes de dependência espacial dos tons de cinza da imagem, exige a fixação dos parâmetros, θ (orientação) e d (distância), com os quais as matrizes são calculadas (seção 2.2.1). Nos experimentos realizados foram fixados quatro pares (θ, d) , a saber: $(0^\circ, 1)$, $(45^\circ, 1)$, $(90^\circ, 1)$ e $(135^\circ, 1)$, sendo, portanto, calculadas quatro matrizes de co-ocorrências.

Para cada uma das quatro matrizes, foram computadas quatro medidas de textura, a saber: energia (equação 2.1), contraste (equação 2.3), homogeneidade (equação 2.4) e entropia (equação 2.5), totalizando 16 medidas para cada imagem.

Essas 16 medidas foram tomadas com sendo o vetor de atributos do classificador neural, implementado de acordo com o modelo de auto-organização (SOM), proposto por Kohonen (seção 4.1.2).

Foram identificadas seis classes de texturas distintas nas 146 imagens do conjunto de treinamento. Essas classes, que variavam de texturas muito grossas até texturas muito finas, foram associadas aos tons de cinza 36, 72, 108, 144, 180 e 216, respectivamente. Para cada uma dessas classes foram extraídas 100 amostras das imagens do conjunto de treinamento, totalizando 600 amostras. Cada uma dessas amostras eram constituídas das 16 medidas obtidas das matrizes de co-ocorrência, calculadas localmente sobre a imagem utilizando-se janelas de dimensão 20x20 pixels.

As 600 amostras foram utilizadas, então, para treinar a rede neural responsável pela classificação das texturas. Elas foram apresentadas aleatoriamente 10000 vezes à rede neural durante a fase de treinamento, gerando mapas auto-organizados de dimensões 7x7.

Após as classificações das texturas, os histogramas das classes de texturas foram calculados para as imagens pertencentes ao conjunto de treinamento. Tais histogramas foram utilizados para treinar a segunda rede neural, responsável pela classificação dos histogramas em uma das cinco classes de densidades de multidões previamente identificadas (baixíssima densidade (VL), baixa densidade (L), densidade moderada (M), alta densidade (H) e altíssima densidade (VH)). O treinamento da segunda rede neural também foi efetuada através da apresentação aleatória dos histogramas das imagens de treinamento classificadas, durante 10000 vezes.

A Figura 5.3 mostra exemplos de histogramas de classes de texturas e suas relações com os níveis de congestionamentos de pessoas. Essas relações são aprendidas pela segunda rede neural, que também gera mapas auto-organizados de dimensões 7x7.

Os resultados das classificações dos histogramas das imagens classificadas do conjunto de teste, em classes de densidades de multidões VL, L, M, H e VH estão descritos na Tabela 6.3. Nessa tabela pode-se notar, por exemplo, que a taxa de acerto do classificador foi de 100% para a classe de imagens de baixíssima densidade. Pode-se notar ainda que o pior desempenho (58,98% de acerto) foi obtido para a classe de imagens com baixa densidade de congestionamentos. Para essa classe de imagens, 25,64% de imagens foram classificadas como sendo de baixíssima densidade.

Conjunto Teste	VL	L	M	H	VH
VL	100,00%	-	-	-	-
L	25,64%	58,98%	15,38%	-	-
M	-	7,69%	76,93%	15,38%	-
H	-	-	5,56%	83,33%	11,11%
VH	-	-	6,25%	25,00%	68,75%

Tabela 6.3: Resultados obtidos aplicando-se a técnica baseada em medidas locais de textura extraídas de matrizes de co-ocorrências dos tons de cinza das imagens. As linhas da tabela mostram as distribuições das estimações das densidades de multidões para cada grupo de imagens do conjunto de teste. Os valores da diagonal da tabela (impressos em negrito) indicam as taxas percentuais de acerto das classificações para cada grupo.

Observando-se o número de imagens do conjunto de teste classificadas corretamente, observamos uma taxa global de 76,51% de classificação correta.

6.2 Medidas Globais de Texturas

A técnica que utiliza medidas globais de texturas foi avaliada utilizando-se medidas extraídas das matrizes de co-ocorrências, dos segmentos de retas, do espectro de Fourier e das dimensões fractais das imagens. As próximas seções descrevem os resultados obtidos durante os experimentos realizados com cada uma destas medidas, para a estimação de densidades de congestionamentos de pessoas.

6.2.1 Matrizes de Co-ocorrência

Uma das técnicas de análise de textura utilizadas nos nossos experimentos com medidas globais foi a baseada em matrizes de co-ocorrências. Assim como nos experimentos utilizando medidas locais de textura (seção 6.1.1), nesses experimentos foram fixados quatro pares (θ, d) , a saber: $(0^\circ, 1)$, $(45^\circ, 1)$, $(90^\circ, 1)$ e $(135^\circ, 1)$, sendo, portanto, calculadas quatro matrizes de co-ocorrências para cada imagem.

Para cada uma das quatro matrizes, foram computadas quatro medidas de textura, a saber: energia (equação 2.1), contraste (equação 2.3), homogeneidade (equação 2.4) e entropia (equação 2.5), totalizando 16 medidas para cada imagem.

Os resultados das classificações das imagens do conjunto teste em classes de densidades de multidões VL, L, M, H e VH, utilizando descritores globais de textura extraídos das quatro matrizes de co-ocorrência estão descritos nas próximas sub-seções.

6.2.1.1 Classificador Bayesiano

Na técnica de classificação Bayesiana (seção 4.1.1) foram empregadas várias combinações das 16 medidas de textura extraídas das matrizes de co-ocorrência. Os resultados das classificações obtidos com algumas dessas combinações estão apresentados na Tabela 6.4.

Tais resultados mostram que o melhor desempenho do classificador Bayesiano (85,52% de acerto) foi obtido quando apenas as duas medidas Contraste 0° e Homogeneidade 0° foram utilizadas.

número de Medidas	Medidas	Taxa de Classificação Correta
16	Energia ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) Contraste ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) Homogeneidade ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) Entropia ($0^\circ, 45^\circ, 90^\circ, 135^\circ$)	69,65%
4	Contraste (0°) Homogeneidade ($0^\circ, 45^\circ$) Entropia (45°)	84,83%
4	Contraste (0°) Homogeneidade ($0^\circ, 90^\circ$) Entropia (45°)	83,44%
3	Contraste (0°) Homogeneidade (0°) Entropia (45°)	83,45%
2	Contraste (0°) Homogeneidade (0°)	85,52%

Tabela 6.4: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e as medidas das matrizes de co-ocorrência. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

A Figura 6.4(a) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) utilizando essas medidas. O gráfico mostra que as classes VL, M e VH alcançaram um índice de acerto superior a 90% e que a classe L apresenta o pior desempenho (em torno de 67%).

A Figura 6.5 mostra o relacionamento entre as medidas Contraste 0° e Homogeneidade 0° . Pode-se observar que os agrupamentos das classes de densidades são quase disjuntos para essas medidas. A classe de baixa densidade (amarela) é a que mais se mistura com as classes vizinhas. Esse fato explica o seu pior desempenho durante a classificação. Também é possível observar que essas medidas são inversamente proporcionais entre si.

As Figuras 6.6 e 6.7 mostram os relacionamentos entre medidas de contraste 0° e homogeneidade 0° , respectivamente, com o número de pessoas na imagem. Conforme aumentam os níveis de congestionamentos, as medidas de contraste tendem a aumentar, ao passo que as medidas de homogeneidade tendem a diminuir, e vice-versa.

6.2.1.2 Redes Neurais

Na técnica baseada em redes neurais (seção 4.1.2) foram utilizadas todas as 16 medidas de textura extraídas das matrizes de co-ocorrência, como sendo o vetor de atributos para a estimação de densidades de multidões. Utilizou-se, nos experimentos, mapas de dimensão 7x7, com 39 amostras em cada classe, durante a fase de treinamento da rede neural. As classes que não tinham 39 amostras, foram complementadas com amostras escolhidas aleatoriamente

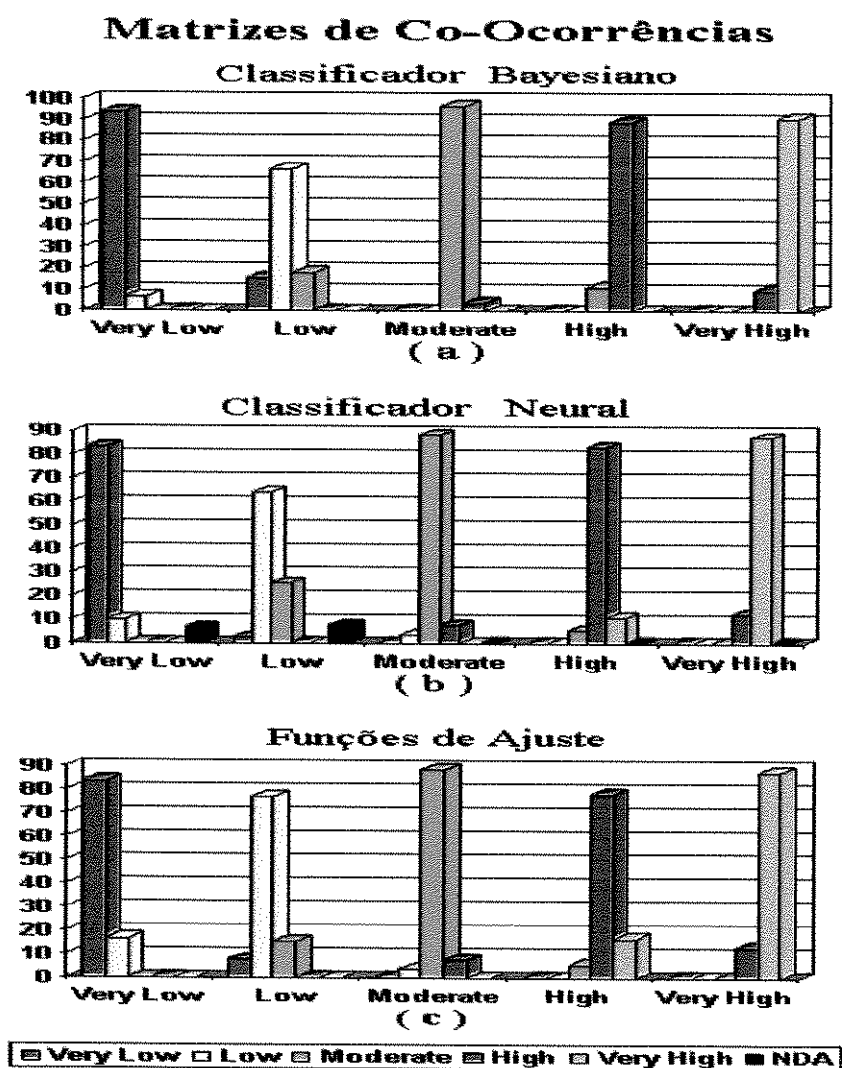


Figura 6.4: Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 2 medidas: Contraste 0^0 e Homogeneidade 0^0 (82,52% de acerto); (b) Classificador Neural com 16 medidas: Contraste (0^0 , 45^0 , 90^0 e 135^0), Homogeneidade (0^0 , 45^0 , 90^0 e 135^0), Energia (0^0 , 45^0 , 90^0 e 135^0) e Entropia (0^0 , 45^0 , 90^0 e 135^0) (80% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: Contraste 0^0 e Homogeneidade 0^0 (82,76% de acerto).

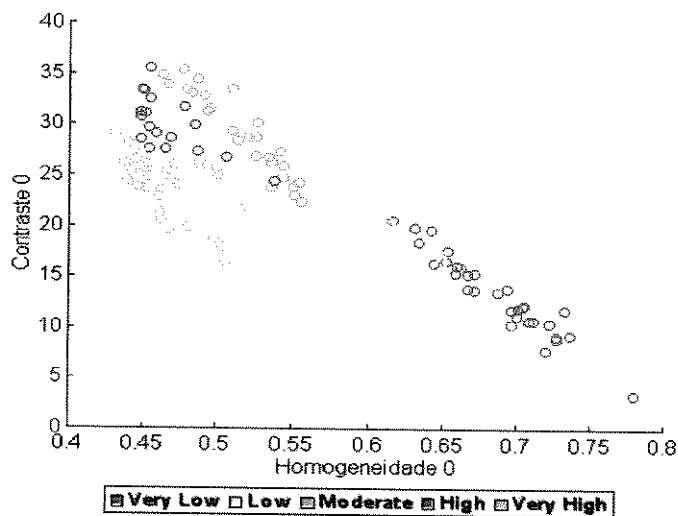


Figura 6.5: Relação entre as medidas Contraste 0° e Homogeneidade 0°.

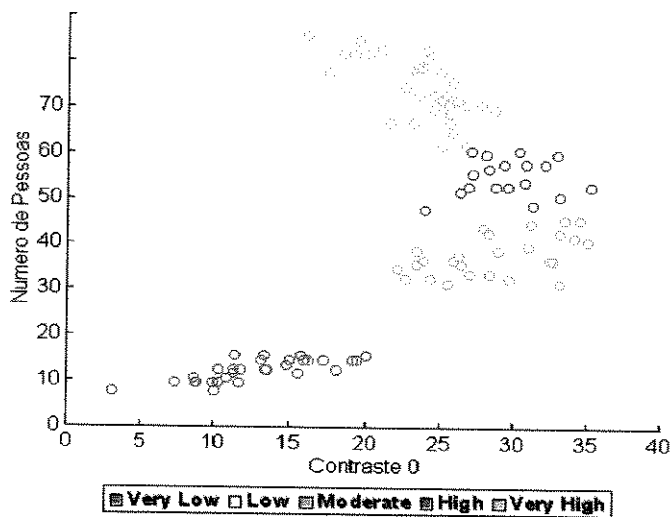


Figura 6.6: Relação entre a medida de Contraste 0° e o número de pessoas na imagem.

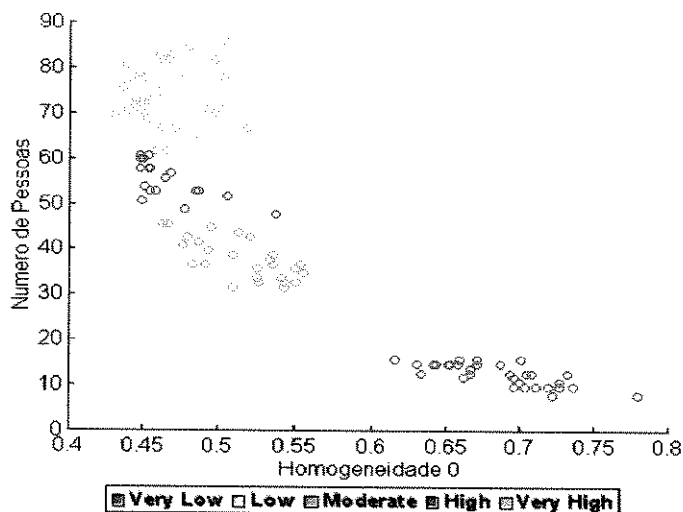


Figura 6.7: Relação entre a medida Homogeneidade O^o e o número de pessoas na imagem.

do seu próprio conjunto de treinamento. As 195 amostras (5 classes x 39 amostras) foram submetidas aleatoriamente 30000 vezes durante o treinamento da rede neural.

A Figura 6.4(b) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) obtidas pelo classificador neural com as medidas extraídas das matrizes de co-ocorrências.

O gráfico mostra que as classes M e VH alcançaram as melhores taxas de acerto (88,46% e 87,5%, respectivamente), seguidas pelas classes VL e H (ambas com 83,33% de acerto). Assim como na técnica baseada no classificador Bayesiano, a classe L apresentou o pior desempenho (64,10%).

De modo geral, 116 imagens, de um conjunto de 145, foram classificadas corretamente, implicando em uma taxa de acerto de 80%.

6.2.1.3 Funções de Ajuste

Outra técnica para a estimação de densidades de congestionamentos de pessoas utiliza funções de ajuste (seção 4.2) da seguinte forma: o número aproximado de pessoas, estimado para cada imagem do conjunto de teste via função de ajuste, é mapeado em uma das cinco classes de densidades pré-estabelecidas. Desse modo, para cada imagem associa-se uma classe de densidade.

O classificador baseado em funções de ajuste foi avaliado através de experimentos que utilizaram pares de medidas, do conjunto total de 16 medidas extraídas das matrizes de co-ocorrência, para determinar os 16 coeficientes dos polinômios cúbicos de duas variáveis (seção 4.2).

Os resultados da classificação obtidos com algumas dessas combinações, cujas taxas de acerto foram superiores a 75%, estão apresentados na Tabela 6.5.

Os resultados apresentados pela Tabela 6.5 mostram, assim como na classificação Bayesiana, que o par de medidas Homogeneidade 0º e Contraste 0º propiciaram o melhor desempenho, 82,76% de acerto.

Embora o melhor resultado obtido através dessa técnica (82,76%) seja sensivelmente menor do que aquele obtido utilizando-se o classificador Bayesiano (85,54%), a dispersão das classificações entre as classes é menor quando se utiliza as funções de ajuste, cujo pior desempenho foi obtido para a classe L, com 76,92% de acerto.

A Figura 6.4(c) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) obtidas pelo classificador que utiliza funções de ajuste e medidas das matrizes de co-ocorrências.

	Hg. 0	Hg. 45	Hg. 90	Hg. 135	Ct. 45	Ct. 135	Ep. 135
Hg. 0			77,24%		80,69%		
Hg. 45						75,17%	
Hg. 90							79,31%
Hg. 135							76,55%
Ct. 0	82,76%	78,62%		77,93%			
Ct. 45		79,31%	77,24%	78,62%			
Ct. 90			75,86%				
Ct. 135				76,55%			
Ep. 0			77,93%	75,17%			
Ep. 45			80,00%	79,31%			

Tabela 6.5: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas, cujos pares de medidas proporcionaram taxas de acerto acima de 75%. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. Hg = Homogeneidade; Ct = Contraste; Ep = Entropia.

O gráfico mostra que as classes M e VH alcançaram as melhores taxas de acerto (88,46% e 87,5%, respectivamente), seguidas pelas classes VL (com 83,33% de acerto). Assim como na técnica baseada no classificador Bayesiano, a classe L apresentou o pior desempenho (76,92% de acerto).

No melhor resultado obtido, 120 imagens, de um conjunto de 145, foram classificadas corretamente, implicando em uma taxa de acerto de 82,76%.

Quando se utiliza as funções de ajuste, estima-se o número de pessoas em cada imagem. Ao invés de mapear esse valor em uma das classes de densidades, pode-se utilizar esse valor diretamente para avaliar o desempenho do estimador, calculando-se os erros relativos e absolutos entre os valores estimados e os valores observados, sendo o erro absoluto dado por $v_{obs} - v_{est}$, e o erro relativo dado por $\frac{v_{obs} - v_{est}}{v_{obs}}$, onde v_{obs} é o valor observado e v_{est} é o valor estimado.

A Tabela 6.6 mostra os erros relativos cometidos pelo estimador quando são utilizados os pares de medidas apresentados na Tabela 6.5. Pode-se observar nesta tabela que o par de medidas Homogeneidade 0° e Contraste 0° produz também o melhor resultado na estimação do número de pessoas, segundo o critério do erro relativo, apresentando um erro de apenas 1,73.

A Figura 6.8 apresenta os gráficos dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

O desempenho do estimador baseado em funções de ajuste também foi verificado utilizando-se apenas uma medida de cada vez, com funções de ajuste polinomiais cúbicas de uma variável (quatro coeficientes). Nesses experimentos, o melhor resultado foi obtido utilizando-se a medida Homogeneidade 90°, que proporcionou uma taxa de acerto de 67,59%, taxa esta

	Hg. 0	Hg. 45	Hg. 90	Hg. 135	Ct. 45	Ct. 135	Ep. 135
Hg. 0			2,65		2,16		
Hg. 45						3,23	
Hg. 90							2,57
Hg. 135							2,11
Ct. 0	1,73	2,52		2,43			
Ct. 45		2,87	2,22	2,77			
Ct. 90			2,51				
Ct. 135				3,70			
Ep. 0			2,70	2,43			
Ep. 45			2,47	1,98			

Tabela 6.6: Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido. Hg = Homogeneidade; Ct = Contraste; Ep = Entropia.

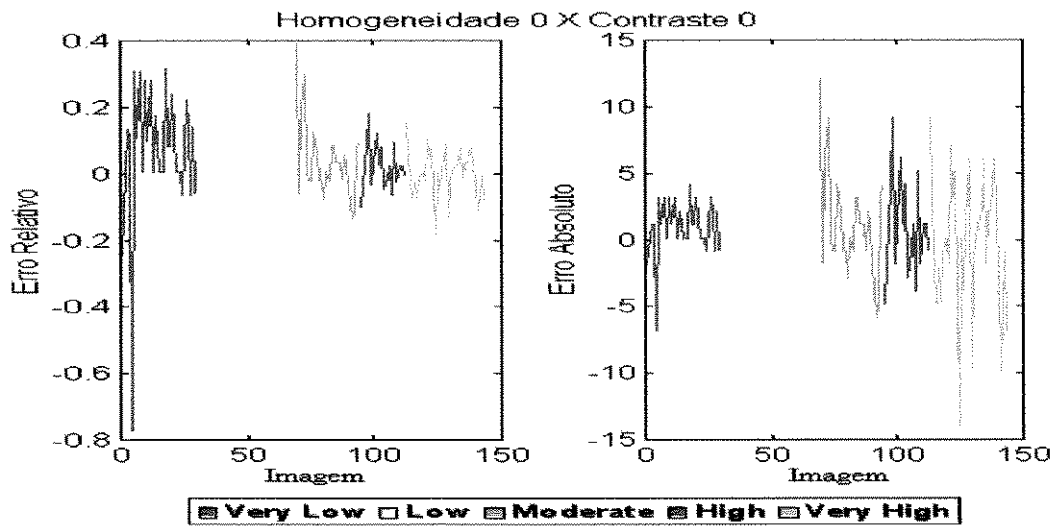


Figura 6.8: Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

significativamente menor do que aquela obtida utilizando-se duas medidas simultaneamente.

6.2.2 Segmentos de Retas

No Capítulo 3.1, foi apresentado o conceito de *intensidade direcional* definido por Denslow [26]. A Figura 3.3 ilustra os componentes de intensidade dos níveis de cinza, dispersão angular, comprimento e densidade, que influenciam as medidas de intensidade direcional.

Observando-se as bordas detectadas das imagens de multidões apresentadas na linha inferior da Figura 5.1, por exemplo, pode-se verificar que os componentes que contribuem para as medidas de intensidade direcional definidas por Denslow estão presentes, com exceção das intensidades dos tons de cinza, uma vez que estas imagens são binárias.

Na imagem mais à esquerda (baixíssima densidade), a dispersão angular das bordas é baixa (supondo-se que as faixas horizontais do piso não sejam consideradas), assim como a densidade espacial dessas bordas, ao passo que as medidas de média e dispersão dos comprimentos das bordas são altas.

À medida em que as densidades de congestionamentos aumentam também tendem a aumentar as dispersões angulares e as densidades espaciais das bordas, enquanto que as medidas de média e de dispersão dos comprimentos tendem a diminuir.

Motivados pelas observações acima, realizamos experimentos a fim de estimar as densidades de congestionamentos de pessoas em imagens, utilizando as seguintes 8 medidas sobre o conjunto de segmentos de retas detectados

nas imagens:

NS: número total de segmentos de retas;

MC(10%): média dos comprimentos do conjunto de segmentos de retas constituído por 10% de todos os segmentos de retas apresentando os maiores comprimentos;

MC: média dos comprimentos de todos os segmentos de retas;

DPC: desvio-padrão dos comprimentos de todos os segmentos de retas;

EC: entropia dos comprimentos de todos os segmentos de retas;

MA: média dos ângulos de todos os segmentos de retas;

DPA: desvio-padrão dos ângulos de todos os segmentos de retas;

EA: entropia dos ângulos de todos os segmentos de retas;

Uma aproximação do valor da média angular é obtida tomando-se o ângulo da direção que maximiza a soma dos comprimentos das projeções de todos os segmentos de retas detectados na imagem [103].

Os segmentos de retas foram detectados através da técnica que utiliza a transformada de Hough, descrita na seção 3.3.

6.2.2.1 Classificador Bayesiano

Na técnica de classificação Bayesiana (seção 4.1.1) foram avaliadas três combinações das 8 medidas de textura extraídas dos segmentos de retas detec-

Número de Medidas	Medidas	Taxa de Classificação Correta
8	NS, MC(10%), MC, DPC, EC, MA, DPA, EA	78,62%
4	MC(10%), DPC, EC, DPA	74,48%
2	DPC, DPA	71,72%

Tabela 6.7: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

tados nas imagens. Os resultados das classificações obtidos com essas combinações estão apresentados na Tabela 6.7.

Os resultados apresentados pela Tabela 6.7 mostram que o melhor desempenho do classificador Bayesiano (78,62%) foi obtido quando foram utilizadas todas as medidas referentes aos segmentos de retas detectados nas imagens.

A Figura 6.9(a) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) utilizando todas essas medidas. A classe L obteve o melhor desempenho (90% de acerto) durante a classificação, seguida pelas classes VH (87,5% de acerto) e M (84,62% de acerto). O classificador Bayesiano obteve um desempenho apenas regular (55,56%) ao classificar as imagens com altos níveis de congestionamento de pessoas (classe H). Pode-se observar no gráfico que aproximadamente 27%

das imagens dessa classe foram consideradas de altíssima densidade, e que aproximadamente 16% delas foram consideradas de densidade moderada. O desempenho regular dessa classe deve-se principalmente às falhas do algoritmo de detecção de bordas, que é fundamental no processo de detecção dos segmentos de retas (seção 3.3).

A Figura 5.1 apresenta exemplos de imagens de congestionamentos de pessoas com suas respectivas bordas detectadas. Pode-se notar que há falhas na detecção das bordas, principalmente na imagem de altíssima densidade (imagem mais à direita da Figura), fazendo com que as imagens dessa classe fiquem similares às imagens da classe de alta densidade.

6.2.2.2 Redes Neurais

Na técnica de classificação neural (seção 4.1.2) também foram avaliadas três combinações das 8 medidas de textura extraídas dos segmentos de retas detectados nas imagens. Os resultados das classificações obtidos com essas combinações estão apresentados na Tabela 6.8.

Os resultados apresentados pela Tabela 6.8 mostram que o melhor desempenho do classificador neural (72,41%) foi obtido quando foram utilizadas apenas 4 medidas dos segmentos de retas detectados nas imagens: MC(10%), DPC, EC e DPA.

A Figura 6.9(b) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) utilizando as quatro medidas. A classe L obteve o melhor desempenho (86,67% de acerto) durante a classificação, seguida pela classe H (76,92% de acerto). As demais classes

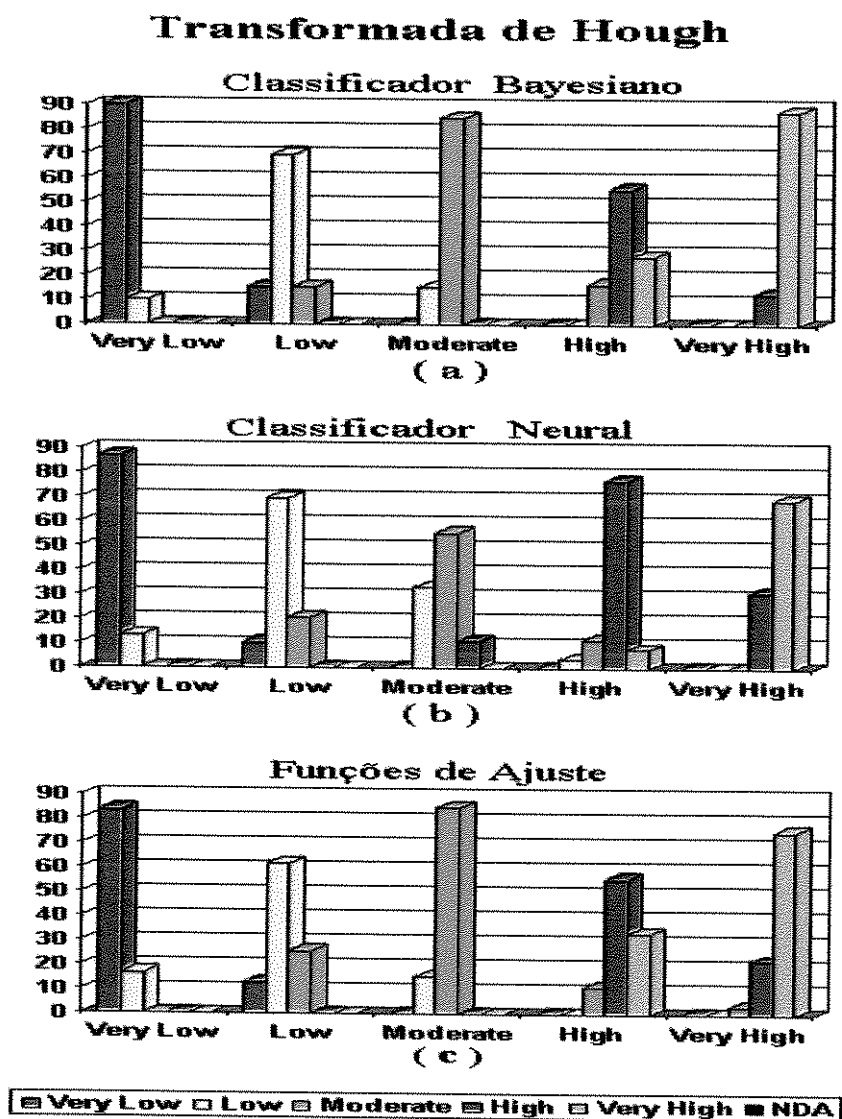


Figura 6.9: Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 8 medidas: NS, MC(10%), MC, DPC, EC, MA, DPA e EA (78,62% de acerto); (b) Classificador Neural com 4 medidas: MC, DPC, EC e DPA (72,41% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: DPC e DPA (72,41% de acerto).

Número de Medidas	Medidas	Taxa de Classificação Correta
8	NS, MC(10%), MC, DPC, EC, MA, DPA, EA	68,97%
4	MC(10%), DPC, EC, DPA	72,41%
2	DPC, DPA	71,03%

Tabela 6.8: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

apresentaram taxas de acerto inferior a 70%, sendo que a classe M apresentou um resultado regular, com apenas 55,56% de acerto. Cerca de 32% das imagens da classe M foram classificadas como pertencentes à classe L.

6.2.2.3 Funções de Ajuste

O classificador baseado em funções de ajuste foi avaliado através de experimentos que utilizaram pares de medidas, do conjunto total de 8 medidas extraídas dos segmentos de retas detectados nas imagens, a fim de determinar os 16 coeficientes dos polinômios cúbicos de duas variáveis (seção 4.2).

Os resultados da classificação obtidos com as combinações de duas entre as 8 medidas (NS, MC(10%),MC,DPC,EC,MA,DPA e EA) estão apresentados na Tabela 6.9.

	MC(10%)	MC	DPC	EC	MA	DPA	EA
NS	64,83%	48,97%	71,03%	57,24%	40%	49,66%	55,86%
MC(10%)		71,03%	71,03%	64,83%	61,38%	68,97%	60%
MC			71,03%	60%	45,52%	42,76	49,66%
DPC				68,28%	69,66%	72,41%	69,66%
EC					64,14%	66,21%	60,69%
MA						22,07%	48,97%
DPA							60%

Tabela 6.9: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas e medidas dos segmentos de retas detectados nas imagens. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

Os resultados apresentados pela Tabela 6.9 mostram que o par de medidas DPC e DPA, que determinam as dispersões dos comprimentos e dos ângulos, respectivamente, propiciam o melhor desempenho, 72,41% de acerto, quando se utiliza funções de ajuste e medidas dos segmentos de retas para se estimar as densidades de congestionamentos de pessoas.

A Figura 6.9(c) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) obtidas pelo classificador que utiliza funções de ajuste e as medidas DPC e DPA dos segmentos de retas. O gráfico mostra que as classes M e VL alcançaram as melhores taxas de acerto (84,62% e 83,33%, respectivamente), seguidas pela classe VH (75%). As demais classes tiveram um desempenho inferior a 70%, sendo que o pior deles foi o obtido pela classe H (55,56%).

No melhor resultado obtido, 105 imagens, de um conjunto de 145, foram classificadas corretamente, implicando em uma taxa de acerto de 72,41%.

A Figura 6.10 mostra o relacionamento entre as medidas DPC e DPA. Pode-se observar nessa figura que o agrupamento da classe de densidades H (cor-de-rosa) encontra-se diluído entre as classes M (verde) e VH (azul-claro), o que justifica o desempenho apenas regular das classificações das imagens pertencentes a essa classe.

As Figuras 6.11 e 6.12 mostram os relacionamentos das medidas DPC e DPA, respectivamente, com o número de pessoas na imagem. Pode-se observar na Figura 6.11 que, conforme aumentam os níveis de congestionamentos, as medidas DPC tendem a diminuir. Pode-se observar na Figura 6.12 que o DPA é baixo na classe M, assumindo valores maiores quando parte-se dessa classe em direção às classes VH e VL. Isto acontece porque na classe VH pre-

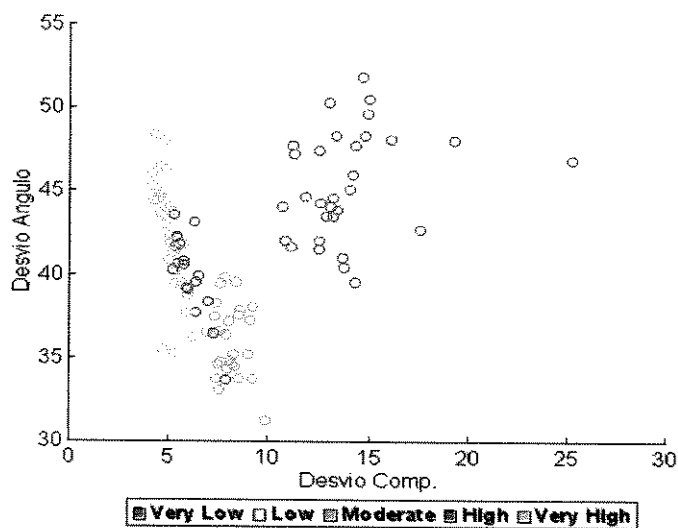


Figura 6.10: Relação entre as medidas DPC e DPA.

dominam os segmentos de retas curtos nas direções verticais e horizontais, e porque na classe VL predominam os segmentos longos nas direções verticais (bordas das pessoas) e horizontais (referentes às faixas horizontais do piso da área sob monitoramento, que não foram desconsideradas nos experimentos realizados).

A Tabela 6.10 mostra os erros relativos cometidos pelo estimador quando são utilizados os pares de medidas apresentados na Tabela 6.9. Pode-se observar nesta tabela que o par de medidas NS e DPC produz um erro relativo (3,01) menor do que o produzido pelo par DPC e DPA (3,42).

A Figura 6.13 apresenta os gráficos dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

O desempenho do estimador baseado em funções de ajuste também foi verificado utilizando-se apenas uma medida de cada vez, com funções de

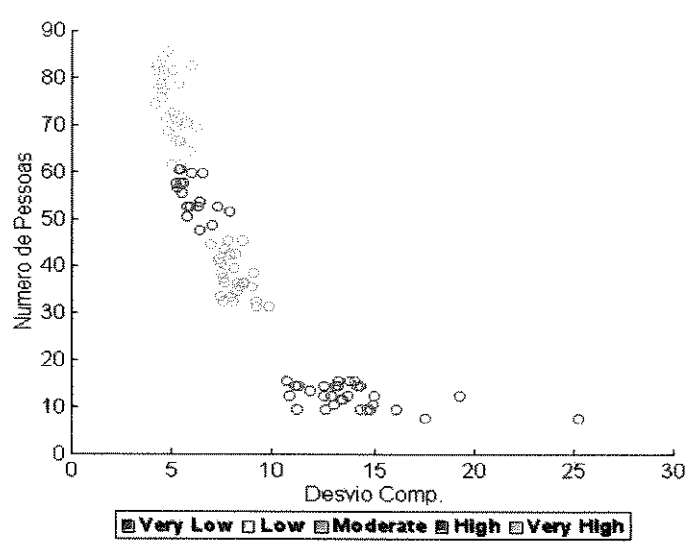


Figura 6.11: Relação entre a medida DPC e o número de pessoas na imagem.

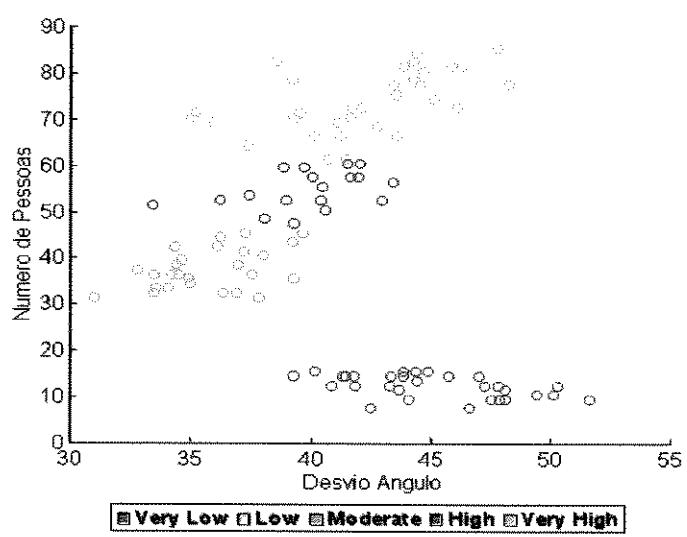


Figura 6.12: Relação entre a medida DPA e o número de pessoas na imagem.

	MC(10%)	MC	DPC	EC	MA	DPA	EA
NS	3,44	4,61	3,01	3,41	5,95	5,94	5,07
MC(10%)		3,5	4,47	3,45	5,24	5,86	4,31
MC			3,42	4,63	11,08	7,74	11,41
DPC				3,19	3,39	3,42	3,09
EC					3,87	7,57	4,69
MA						14,57	9,64
DPA							8,05

Tabela 6.10: Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

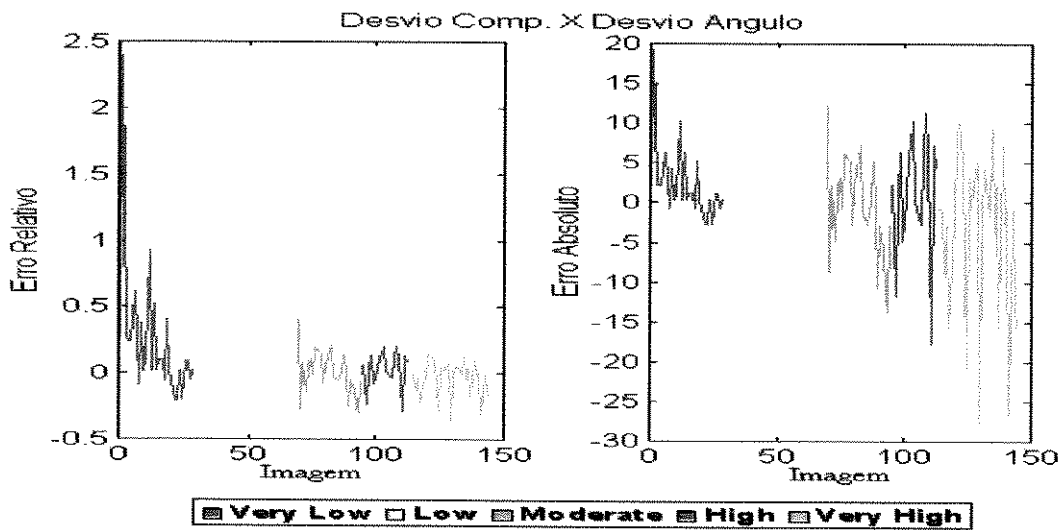


Figura 6.13: Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

ajuste polinomiais cúbicas de uma variável (quatro coeficientes). Nesses experimentos, o melhor resultado foi obtido utilizando-se a medida DPC, que proporcionou uma taxa de acerto de 70,34%, taxa esta comparável à melhor taxa (72,41%) obtida utilizando-se duas medidas simultaneamente.

6.2.3 Análise de Fourier

Medidas de textura extraídas do espectro de Fourier (seção 2.2.1) também foram avaliadas nos experimentos para determinação dos níveis de congestionamentos de pessoas. As medidas utilizadas consistem nas somas dos valores de 24 regiões do espectro localizadas na sua região central. Essas medidas foram tomadas considerando-se apenas $\frac{1}{3}$ do espectro de Fourier, centralizado no par de frequências (0,0).

Das 24 regiões, 16 são radiais e 8 são anéis concêntricos em torno da origem do espectro. A Figura 6.14 ilustra as regiões do espectro de Fourier onde foram tomadas as 24 medidas de textura.

6.2.3.1 Classificador Bayesiano

Na técnica de classificação Bayesiana (seção 4.1.1) foi avaliada uma combinação das 24 medidas de textura do espectro de Fourier, constituída por 12 valores das regiões rotuladas pelos números ímpares (Figura 6.14).

O resultado da classificação obtido com essa combinação está apresentado na Tabela 6.11.

Com essas 12 medidas, 102 imagens de um total de 145 do conjunto de teste foram classificadas corretamente, ou seja, 70,34% de acerto. A

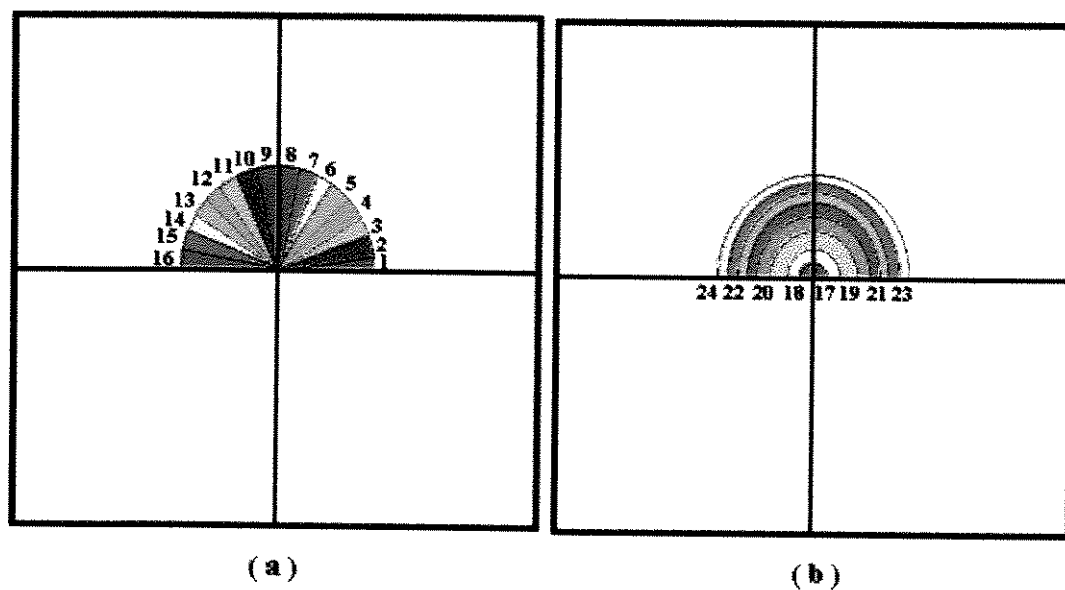


Figura 6.14: Regiões do espectro de Fourier onde foram tomadas medidas de textura. (a) 16 regiões radiais; (b) 8 anéis concêntricos.

Número de Medidas	Medidas	Taxa de Classificação Correta
12	1,3,5,7,9,11,13, 15,17,19,21,23	70,34%

Tabela 6.11: Resultado das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador Bayesiano e medidas de textura extraídas do espectro de Fourier.

Figura 6.15(a) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) utilizando essas 12 medidas.

A classe VH obteve o melhor desempenho (93,75% de acerto) durante a classificação, seguida pelas classes L (76,92% de acerto) e VL (70% de acerto). O classificador Bayesiano obteve um desempenho ruim (44,45%) ao classificar as imagens com altos níveis de congestionamento de pessoas (classe H). Pode-se observar no gráfico que quase 35% das imagens dessa classe foram consideradas de altíssima densidade, e que mais de 20% delas foram consideradas de densidade moderada. A classe M teve um desempenho regular com 50% de acerto. Quase 40% das imagens dessa classe foram classificadas como sendo de densidade baixa, L.

6.2.3.2 Redes Neurais

Na técnica de classificação neural (seção 4.1.2) foram empregadas todas as 24 medidas de textura extraídas do espectro de Fourier (Figura 6.14). O resultado das classificações obtidos com todas essas medidas estão apresentados na Tabela 6.12.

Com as 24 medidas de textura do espectro de Fourier, o classificador neural obteve uma taxa de acerto de 75,17%, o que corresponde a classificação correta de 109 imagens do conjunto teste de 145 imagens. A Figura 6.15(b) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) utilizando as 24 medidas. A classe VH obteve o melhor desempenho (93,75% de acerto) durante a classificação, seguida pelas classes VL (80% de acerto) e H (77,78% de acerto). As demais classes

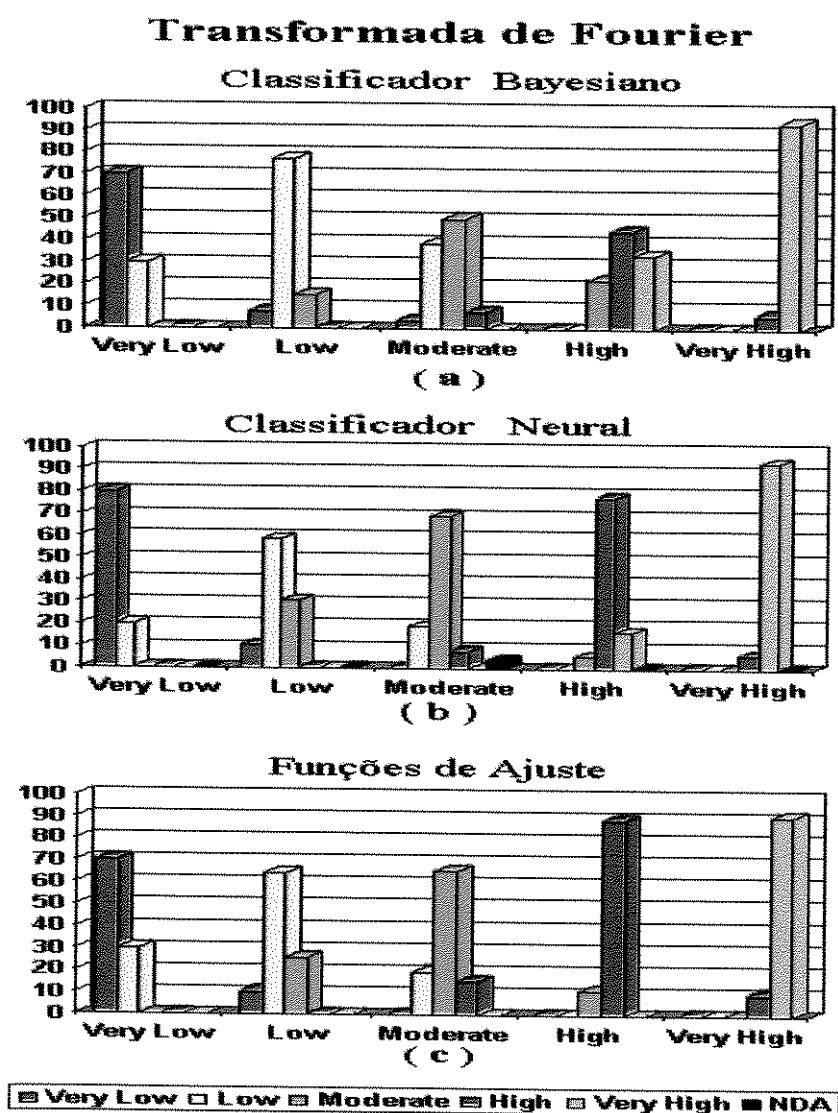


Figura 6.15: Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com 8 medidas radiais (1,3,5, 7,9,11,13,15) e 4 circulares (17,19,21,23) (70,34% de acerto); (b) Classificador Neural com todas as 24 medidas (75,17% de acerto); (c) Classificador baseado em funções de ajuste com 2 medidas: 1 radial (4) e 1 circular (17) (74,48% de acerto).

Número de Medidas	Medidas	Taxa de Classificação Correta
24	1,2,3,4,5,6,7,8, 9,10,11,12,13,14, 15,16,17,18,19,20, 21,22,23,24	75,17%

Tabela 6.12: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e 24 medidas de texturas extraídas do espectro de Fourier.

apresentaram taxas de acerto inferior a 70%. A classe com pior desempenho foi a constituída de imagens de baixas densidades de congestionamentos (L), que atingiu 58,97% de acerto. Nessa classe, cerca de 30% das imagens foram consideradas de densidade moderada (M).

6.2.3.3 Funções de Ajuste

O classificador baseado em funções de ajuste foi avaliado através de experimentos que utilizaram pares de medidas, do conjunto total de 24 medidas extraídas do espectro de Fourier, a fim de determinar os 16 coeficientes dos polinômios cúbicos de duas variáveis (seção 4.2).

Os resultados das classificações obtidos com as combinações de duas entre as 24 medidas, que obtiveram um índice de acerto acima de 65%, estão apresentados na Tabela 6.13.

	3	4	6	7	13	14	16	17
16		66,90%				66,21%		
17	66,90%	74,48%	68,28%	69,66%	67,59%	72,41%		
18								67,59%
19								66,21%
20		66,90%					70,34%	69,66%
21								71,72%

Tabela 6.13: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador baseado em funções de ajuste polinomiais cúbicas e medidas de textura extraídas do espectro de Fourier. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

Os resultados apresentados pela Tabela 6.13 mostram que o par de medidas 4 (radial) e 17 (circular) propiciaram o melhor desempenho, 74,48% de acerto, quando se utiliza funções de ajuste e medidas de texturas extraídas do espectro de Fourier para se estimar as densidades de congestionamentos de pessoas.

A Figura 6.15(c) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) obtidas pelo classificador que utiliza funções de ajuste e as medidas 4 e 17 do espectro de Fourier.

O gráfico mostra que as classes VH e H alcançaram as melhores taxas de acerto (90,63% e 88,89%, respectivamente). As demais classes apresentaram taxas de acerto entre 60% e 70%, sendo que o pior desempenho foi apresentado pela classe L (64,10% de acerto).

No melhor resultado obtido, 108 imagens, de um conjunto de 145, foram classificadas corretamente, implicando em uma taxa de acerto geral de 74,48%.

A Figura 6.16 mostra o relacionamento entre as medidas 4 e 17.

As Figuras 6.17 e 6.18 mostram os relacionamentos das medidas 4 e 17, respectivamente, com o número de pessoas nas imagens.

A Tabela 6.14 mostra os erros relativos cometidos pelo estimador quando são utilizados os pares de medidas apresentados na Tabela 6.13. Pode-se observar nesta tabela que o par de medidas 4 e 20 produz um erro relativo (3,36) menor do que o produzido pelo par 4 e 17 (3,54).

A Figura 6.19 apresenta os gráficos dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

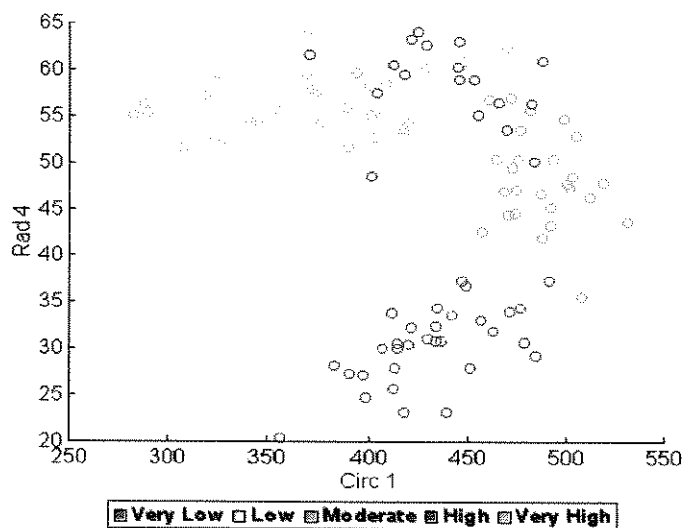


Figura 6.16: Relação entre as medidas 4 (Rad 4) e 17 (Circ 1).

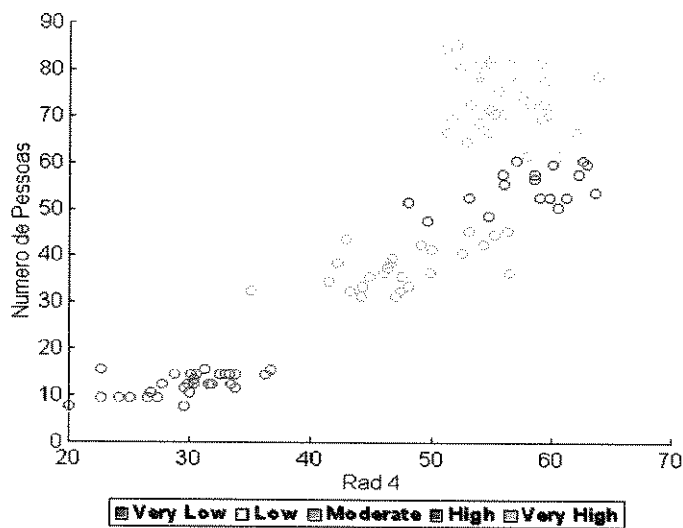


Figura 6.17: Relação entre a medida 4 (Rad 4) e o número de pessoas na imagem.

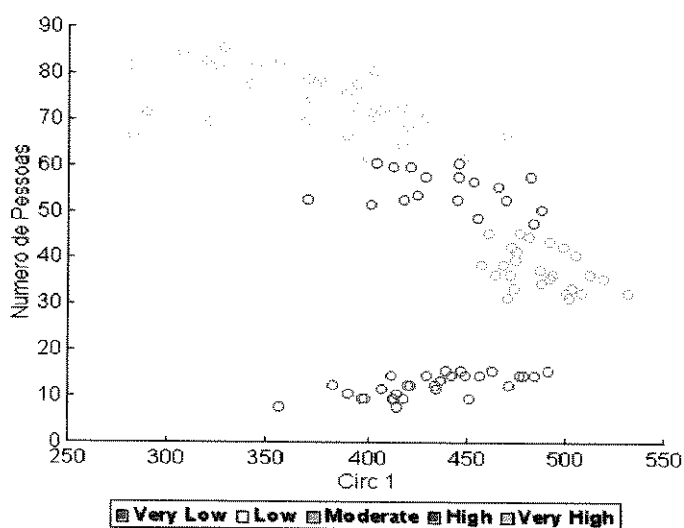


Figura 6.18: Relação entre a medida 17 (Circ 1) e o número de pessoas na imagem.

	3	4	6	7	13	14	16	17
16		5, 47				5, 12		
17	5, 47	3, 54	3, 77	4, 18	4, 66	4, 48		
18								5, 01
19								4, 60
20		3, 36					4, 04	4, 33
21								4, 38

Tabela 6.14: Erros relativos das estimativas das densidades de multidões das imagens do conjunto de teste. O valor da tabela marcado com um retângulo indica o melhor resultado obtido.

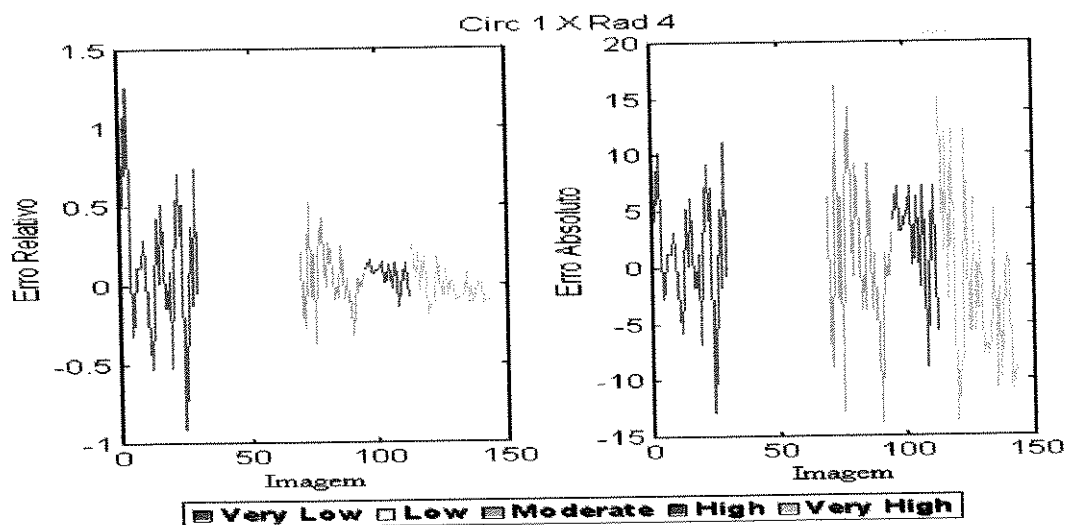


Figura 6.19: Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

6.2.4 Fractais

Medidas de textura determinadas pelas dimensões fractais das imagens também foram empregadas nos experimentos para determinação dos níveis de congestionamentos de pessoas. As dimensões fractais das imagens foram determinadas através da técnica de Minkowski descrita na seção 2.2.3, utilizando-se 15 elementos estruturantes (quadrados de dimensões 3x3, 4x4, ..., 17x17).

6.2.4.1 Classificador Bayesiano

O resultado da classificação obtido com essa combinação está apresentado na Tabela 6.15.

Utilizando-se apenas a dimensão fractal das imagens, 103 imagens de um

Número de Medidas	Medidas	Taxa de Classificação Correta
1	Dimensão Fractal	71,03%

Tabela 6.15: Resultado das classificações das densidades de multidões das i-
magens do conjunto de teste, utilizando o classificador Bayesiano e dimensões
fractais das imagens.

total de 145 do conjunto de teste foram classificadas corretamente, ou seja,
71,03%.

A Figura 6.20(a) apresenta um gráfico da distribuição das classificações
entre as cinco classes de densidades (VL,L,M,H e VH) utilizando as dimensões
fractais.

A classe VH obteve o melhor desempenho (100% de acerto) durante a
classificação, seguida pela classe VL (86,67% de acerto). As demais classes
apresentaram taxas de acerto menores do que 70%. O classificador Bayesiano
não conseguiu distinguir as imagens de altas densidades (H), classificando-as
quase que totalmente na classe VH.

6.2.4.2 Redes Neurais

A técnica de classificação neural (seção 4.1.2) também foi avaliada utilizando-
se a dimensão fractal como descritor de texturas. O resultado das classi-
ficações obtido com esse descritor de textura está apresentado na Tabela 6.16.

A Figura 6.20(b) apresenta um gráfico da distribuição das classificações

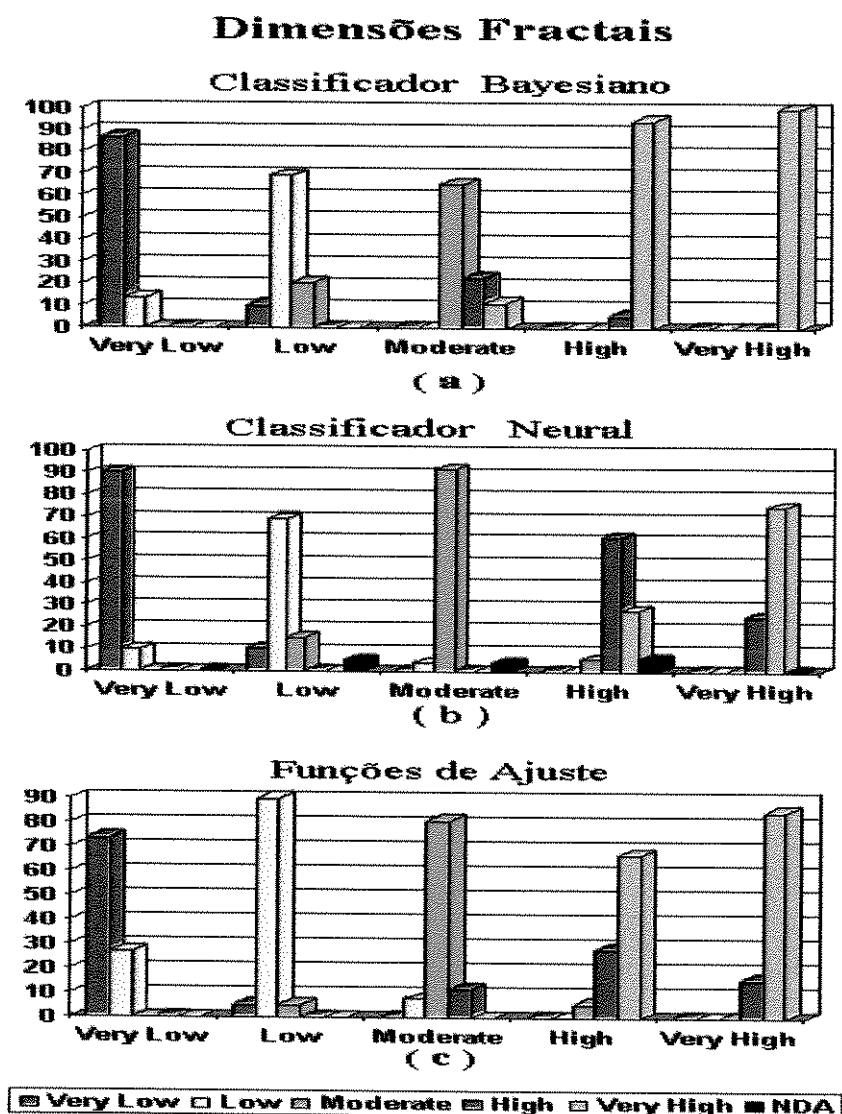


Figura 6.20: Gráficos das distribuições das classificações das densidades de multidões das imagens do conjunto de teste entre as cinco classes VL, L, M, H e VH. (a) Classificador Bayesiano com a dimensão fractal da imagem (71,03% de acerto); (b) Classificador Neural com a dimensão fractal da imagem (78,62% de acerto); (c) Classificador baseado em funções de ajuste a dimensão fractal da imagem (75,86% de acerto).

Número de Medidas	Medidas	Taxa de Classificação Correta
1	Dimensão Fractal	78,62%

Tabela 6.16: Resultados das classificações das densidades de multidões das imagens do conjunto de teste, utilizando o classificador neural e as dimensões fractais das imagens como descritores de texturas.

entre as cinco classes de densidades (VL,L,M,H e VH) utilizando-se as dimensões fractais. A classe M obteve o melhor desempenho (92,30% de acerto) durante a classificação, seguida pelas classes VL (90% de acerto) e VH (75% de acerto). As demais classes apresentaram taxas de acerto inferiores a 70%. O pior desempenho foi obtido pela classe H (61,10% de acerto). Cerca de 28% das imagens dessa classe foram classificadas como sendo de altíssima densidade. Além disso, 25% das imagens da classe VH foram consideradas como sendo de alta densidade, demonstrando que o classificador neural também teve dificuldades para distinguir essas duas classes.

6.2.4.3 Funções de Ajuste

O classificador baseado em funções de ajuste utilizou um polinômio cúbico de uma variável (4 coeficientes), uma vez que na técnica baseada em fractais apenas uma medida, a dimensão fractal, é tomada como descritor de textura.

O resultado das classificações obtido com as funções de ajuste e as dimensões fractais das imagens estão apresentados na Tabela 6.17.

Número de Medidas	Medidas	Taxa de Classificação Correta
1	Dimensão Fractal	75,86%

Tabela 6.17: Resultado das classificações das densidades de multidões das i-magens do conjunto de teste, utilizando funções de ajuste e dimensões fractais das imagens.

A Figura 6.20(c) apresenta um gráfico da distribuição das classificações entre as cinco classes de densidades (VL,L,M,H e VH) obtidas pelo classificador que utiliza funções de ajuste e as dimensões fractais das imagens como descritores de textura.

O gráfico mostra que a classe L alcançou a melhor taxa de acerto (89,74%), seguida pelas classes VH (84,38%) e M (80,77%). A classe VL obteve um índice superior a 70%. O desempenho global, com 75,86% de acerto, ou seja, 110 imagens corretamente classificadas, de um conjunto de 145 imagens, ficou prejudicado por causa do desempenho do classificador com as imagens da classe H, 66,67% das quais foram classificadas erroneamente como sendo da classe VH. O classificador baseado nas funções de ajuste também não conseguiu distinguir as imagens das classes H e VH.

A Figuras 6.21 mostra a relação entre a dimensão fractal de uma imagem e o número de pessoas presentes nessa imagem.

A Figura 6.22 apresenta os gráficos dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste. O erro

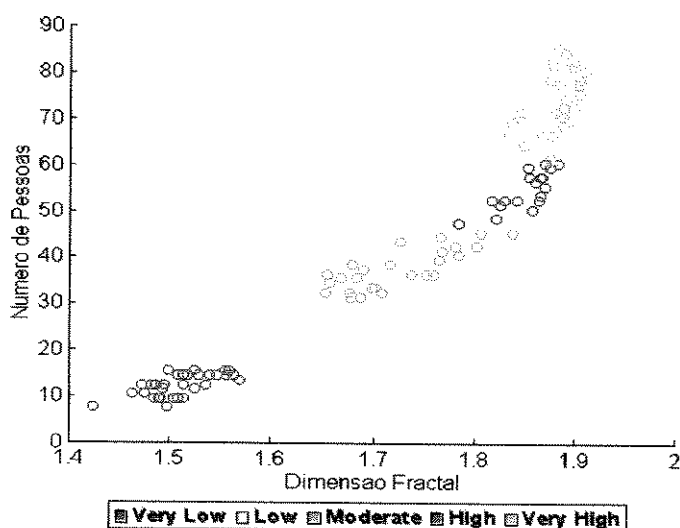


Figura 6.21: Relação entre a dimensão fractal e o número de pessoas na imagem.

relativo cometido pelo estimador foi de apenas 1,91.

6.3 Comparação dos Resultados e Conclusões

Os resultados obtidos durante os experimentos realizados com as 145 imagens do conjunto de teste, com base nas informações colhidas no conjunto de 146 imagens de treinamento, permitem concluir que, de modo geral, os descritores de texturas são medidas bastante efetivas para a determinação de estimativas de congestionamentos de pessoas, visto que foi possível atingir taxas de acerto global em torno de 85% e taxas de acerto por classes de densidades no patamar de 90%.

Também foi possível verificar que a estratégia que utiliza medidas globais

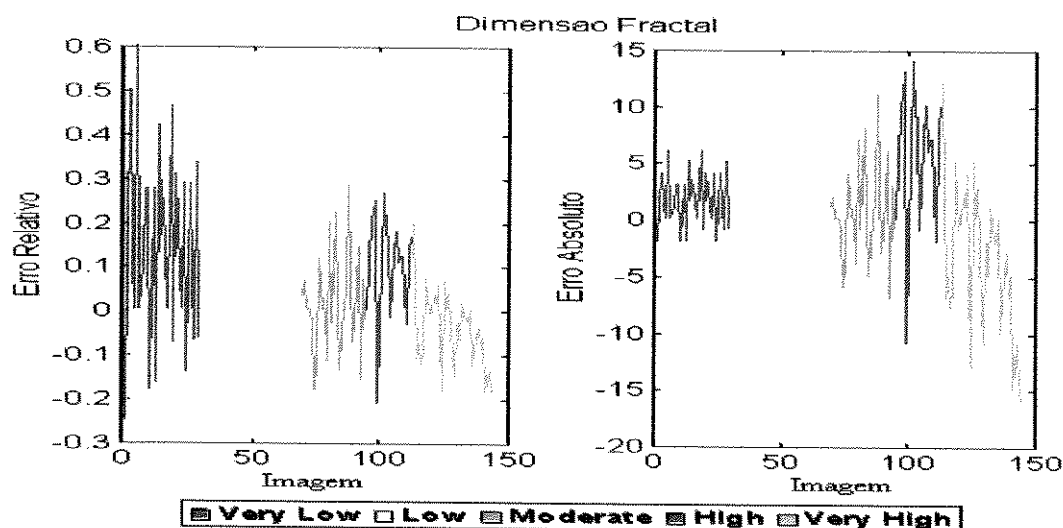


Figura 6.22: Gráfico dos erros relativos e absolutos das estimativas do número de pessoas nas imagens do conjunto de teste.

de texturas é mais apropriada caso sejam requeridas respostas em tempo-real, além de apresentar um índice de acerto (80%) superior àquele apresentado pela estratégia baseada em medidas locais de texturas (76,51%), quando são utilizadas as medidas de matrizes de co-ocorrências e o classificador neural implementado segundo o modelo de Kohonen.

A Tabela 6.18 mostra os melhores resultados obtidos com cada uma das quatro medidas avaliadas nos experimentos com medidas globais de texturas (matrizes de co-ocorrências, segmentos de retas - transformada de Hough, transformada de Fourier e dimensões fractais), com os três classificadores (estatístico - Bayes, neural - modelo de Kohonen, e baseado em funções de ajuste - polinômios cúbicos). Parte dos dados da Tabela 6.18 estão apresentados graficamente na Figura 6.23.

	Bayes	Neural	Função de Ajuste	Média das Medidas
Co-ocorrências	85,52%	80%	82,76%	82,76%
Hough	78,62%	72,41%	72,41%	74,48%
Fourier	70,34%	75,17%	74,48%	73,33%
Fractais	71,03%	78,62%	75,86%	75,17%
Média dos Classificadores	76,38%	76,55%	76,38%	

Tabela 6.18: Melhores resultados obtidos com as quatro medidas de textura (co-ocorrências dos tons de cinza, segmentos de retas, espectro de Fourier e dimensões fractais) e com os três classificadores (estatístico, neural e baseado em funções de ajuste). Os valores da tabela marcados por retângulos indicam os melhores resultados obtidos.

É possível observar na Tabela 6.18 e na Figura 6.23, que as medidas baseadas nas matrizes de co-ocorrências dos tons de cinza produziram os melhores resultados com os três classificadores empregados, sendo que o seu melhor desempenho foi obtido com o classificador estatístico (85,52% de acerto). A taxa média de acerto desta medida foi de 82,76%.

As demais medidas produziram resultados um pouco inferiores, com uma pequena vantagem para as medidas das dimensões fractais (75,17% de acerto médio), seguida pelas medidas dos segmentos de retas (74,48% de acerto médio) e pelas medidas extraídas do espectro de Fourier (73,33% de acerto médio).

Com relação aos classificadores, todos apresentaram taxas médias de acerto em torno de 76%, tendo o classificador Bayesiano apresentado o melhor desempenho quando foram utilizadas medidas das matrizes de co-ocorrências (85,52% de acerto).

A Tabela 6.19 mostra as taxas de classificações corretas para cada uma das classes individualmente, quando foram obtidos os melhores resultados globais, apresentados na Tabela 6.18. Parte dos dados mostrados por esta tabela estão apresentados graficamente nas Figuras 6.24 e 6.25.

Pode-se observar nos gráficos apresentados na Figura 6.25 que as classificações das classes de alta e altíssima densidades alcançaram taxas de acerto em torno de 90%. A classe de alta densidade, por exemplo, atingiu em torno de 90% de classificações corretas quando as medidas das matrizes de co-ocorrências foram classificadas com o classificador Bayesiano e quando as medidas do espectro de Fourier foram classificadas com o classificador baseado nas funções de ajuste. O desempenho da classe de imagens de

Comparação dos Métodos

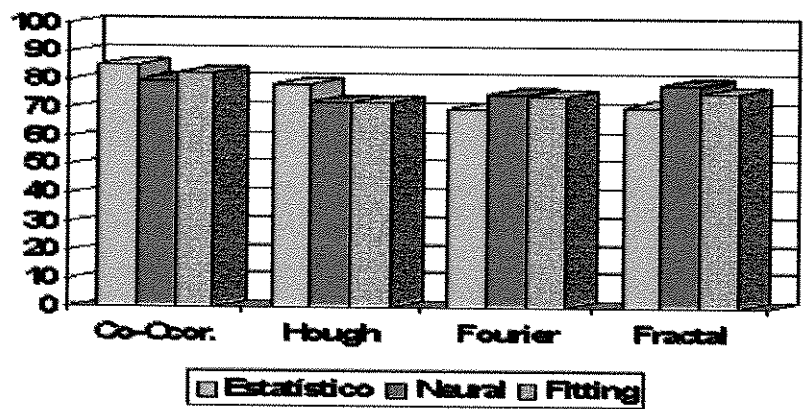


Figura 6.23: Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste (“fitting”), empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da dimensão fractal.

Classe		Co-ocor.	Hough	Fourier	Fractais	Média
VL	Bayes	93,33%	90%	70%	86,67%	85%
	Neural	83,33%	86,67%	80%	90%	85%
	Ajuste	83,33%	83,33%	70%	73,33%	77,50%
	Média	86,66%	86,67%	73,33%	83,36%	
L	Bayes	66,67%	69,24%	76,92%	69,23%	70,52%
	Neural	64,10%	69,23%	58,97%	69,23%	65,38%
	Ajuste	76,92%	61,54%	64,10%	89,74%	73,08%
	Média	69,23%	66,67%	66,66%	76,07%	
M	Bayes	96,15%	84,62%	50%	65,38%	74,04%
	Neural	88,46%	55,56%	69,23%	92,30%	76,39%
	Ajuste	88,46%	84,62%	65,38%	80,77%	79,81%
	Média	91,02%	74,93%	61,54%	79,48%	
H	Bayes	88,89%	55,56%	44,45%	5,56%	48,62%
	Neural	83,33%	76,92%	77,78%	61,10%	74,78%
	Ajuste	77,78%	55,56%	88,89%	27,78%	62,50%
	Média	83,33%	62,68%	70,37%	31,48%	
VH	Bayes	90,62%	87,50%	93,75%	100%	92,97%
	Neural	87,50%	68,75%	93,75%	75%	81,25%
	Ajuste	87,50%	75%	90,63%	84,38%	84,37%
	Média	88,54%	77,08%	92,71%	86,46%	

Tabela 6.19: Taxas de classificações corretas de cada classe de densidades de congestionamentos, para os melhores resultados obtidos com as quatro medidas de texturas (co-ocorrências dos tons de cinza, segmentos de retas - Hough, espectro de Fourier e dimensões fractais), com os classificadores estatístico (Bayes), neural (modelo de Kohonen) e baseados em funções de ajuste (polinômios cúbicos). Os valores da tabela marcados por retângulos indicam os melhores resultados obtidos.

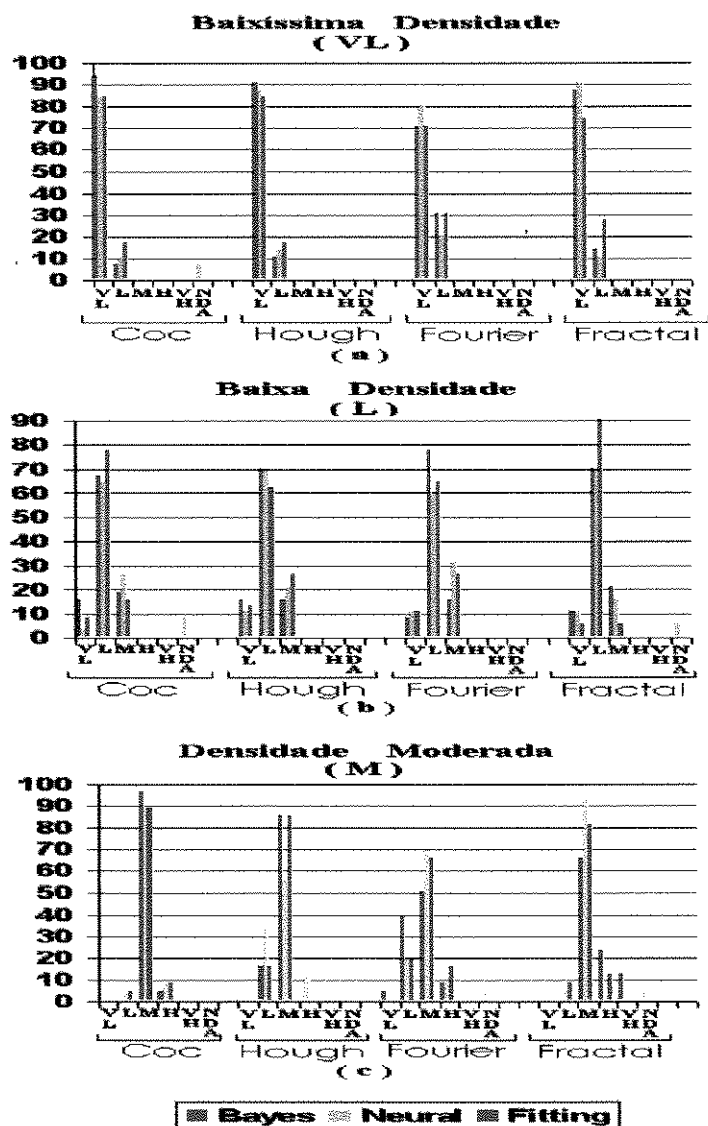


Figura 6.24: Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste, empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da dimensão fractal. (a) classe de baixíssimas densidades; (b) classe de baixas densidades; (c) classes de densidades moderadas.

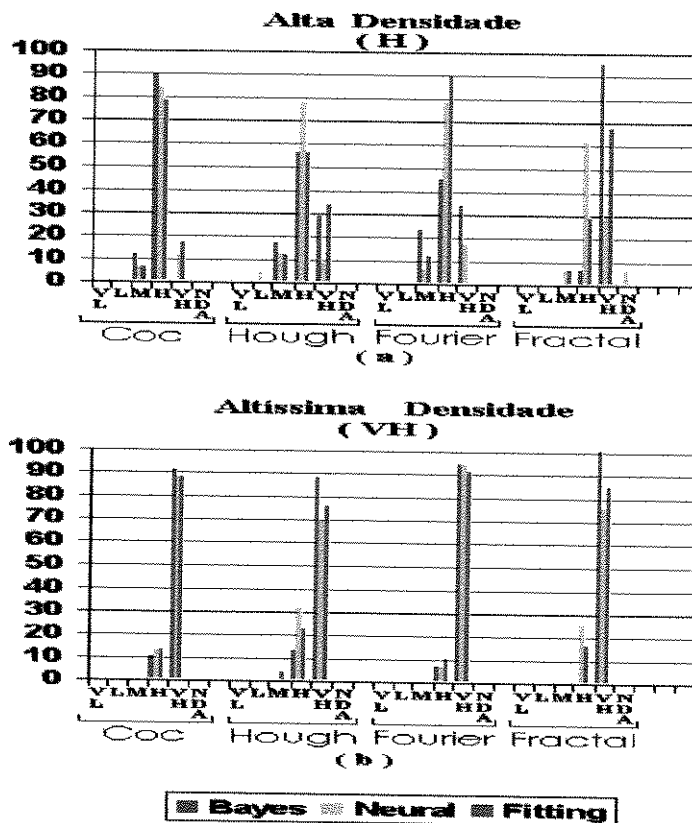


Figura 6.25: Comparação dos desempenhos dos classificadores Bayesiano, neural e baseado em funções de ajuste, empregando medidas de texturas extraídas das matrizes de co-ocorrência, dos segmentos de retas, da transformada de Fourier e da dimensão fractal. (a) classe de altas densidades; (b) classe de altíssimas densidades.

altíssimas densidades foi ainda melhor, tendo atingido 100% de classificação correta quando as medidas das dimensões fractais foram classificadas utilizando o classificador Bayesiano. As medidas de co-ocorrências e do espectro de Fourier obtiveram em torno de 90% de classificação correta para todos os classificadores avaliados. As medidas baseadas nos segmentos de retas também apresentaram um bom desempenho (quase 90% de acerto) quando foi utilizado o classificador Bayesiano. A obtenção de resultados um pouco piores para essas medidas deve-se ao fato dessa técnica requerer a detecção de bordas que, como pode ser observado nas imagens da Figura 5.1, apresenta falhas quando existem altíssimas densidades de congestionamentos de pessoas.

Se tivessem sido adotados os intervalos obedecendo-se as definições de níveis de serviços propostos por Polus, os resultados obtidos com as medidas de texturas para essas duas classes de imagens teriam sido ainda melhores. Até mesmo as medidas das dimensões fractais, que apresentaram péssimos resultados durante as classificações das imagens da classe de altas densidades produziram excelentes resultados, uma vez que grande parte dessas imagens foram classificadas como sendo de altíssimas densidades quando utilizou-se medidas das dimensões fractais.

Outro fato positivo que pode ser observado nos gráficos das Figuras 6.24 e 6.25 são as distribuições das classificações pelas classes de densidades. Essas distribuições podem, na maioria das vezes, ser modeladas com Gaussianas (ou seja, são normais), com baixíssimas dispersões e médias iguais aos valores esperados para cada uma das classes. Disso decorre que os erros cometidos nas classificações podem até ser aceitáveis na prática, pois as classes obtidas

erroneamente pelos classificadores são geralmente vizinhas adjacentes das classes corretas.

Outra característica positiva da utilização de medidas de textura para estimar as densidades de congestionamentos de pessoas é o fato dos classificadores nunca terem confundido as classes de baixíssimas e altíssimas densidades, dessa forma, em uma aplicação onde o objetivo é apenas distinguir situações de super-lotação de uma plataforma de trens, por exemplo, da situação de normalidade, esta técnica pode propiciar quase 100% de acerto.

Capítulo 7

Conclusões

Um dos objetivos desta tese foi propor novos descritores para a caracterização e classificação de texturas orientadas, inspirados nos sistemas biológicos de visão, dentre os quais o sistema de visão humana (SVH), dada a notável capacidade de reconhecimento de padrões complexos observada nesses sistemas.

Para atingir este objetivo, buscou-se inicialmente conhecer os mecanismos gerais dos sistemas naturais de visão, dando ênfase ao estudo do processamento de textura. Nessa fase, tomou-se conhecimento da importância dos estímulos retilíneos e da representação das informações visuais por meio de aproximações de contornos por segmentos de retas, através dos resultados obtidos por pesquisas nas áreas da Neurofisiologia e da Psicofísica, especialmente aquelas realizadas por Hubel e Wiesel [50, 51] e Julesz [56, 57, 58, 59, 60].

Propôs-se, então, novos descritores para texturas orientadas baseados em

informações dos segmentos de retas, que são extraídos das imagens através de técnicas que utilizam a transformada de Hough (Capítulo 3). A aplicação dos novos descritores propostos para as segmentações e classificações de imagens reais de texturas orientadas (algumas delas extraídas do álbum de texturas de Brodatz [10]), foi realizada com êxito [81, 82].

Na fase seguinte do desenvolvimento dos trabalhos, deparou-se com um problema real, cuja solução vem sendo investigada por um grupo de pesquisadores do King's College e da University College London, de Londres, e de outras Universidades do Reino Unido, reunidos no projeto *Cromatic*, para automatizar as tarefas de monitoração de multidões. Uma sub-tarefa desse problema consiste na automação da estimação de densidades de congestionamentos de pessoas, para a qual havia sido proposta uma solução, que, no entanto, não pode ser aplicada em congestionamentos muito densos.

Esse problema motivou as fases seguintes deste trabalho, que consistiram em propor, desenvolver e avaliar novas técnicas baseadas em medidas de texturas para a estimação automática de densidades de multidões. Decidiu-se avaliar também a eficácia das medidas de texturas orientadas para a solução do problema, uma vez que as imagens alvo deste trabalho são constituídas, na maioria das vezes, por silhuetas de pessoas que encontram-se na postura vertical. Tais silhuetas podem ser consideradas os elementos básicos que se repetem na imagem, formando texturas orientadas fracamente acopladas, uma vez que os posicionamentos das pessoas na imagem podem ser considerados aleatórios.

Propôs-se, então, duas novas estratégias para a estimação de densidades de congestionamentos de pessoas, uma baseada em medidas locais [83] e

outra baseada em medidas globais de texturas [84, 85]. Adotaram-se, nestas estratégias, medidas de texturas já consagradas na literatura (matrizes de co-ocorrências, análise de Fourier e fractais) e aquela medida proposta nesta tese para a análise de texturas orientadas baseada em segmentos de retas, de tal forma a obter as melhores estimativas para as densidades de multidões, explorando e avaliando a questão da direcionalidade dos padrões de textura das imagens de multidões.

Como não se está, a princípio, interessado em obter o número exato de pessoas nos congestionamentos, ambas as estratégias baseiam-se nas classificações de padrões, ou seja, dada uma imagem de multidão, obtém-se como resposta, uma classe de densidade de congestionamento para aquela imagem. Foram avaliados os desempenhos de três classificadores de abordagens distintas: um classificador neural implementado segundo o modelo de Kohonen, um classificador Bayesiano e um classificador baseado em funções de ajuste (polinômios cúbicos).

Ambas as estratégias foram avaliadas utilizando-se um conjunto de aproximadamente 300 imagens, capturadas na estação de trens da Liverpool Street, Londres, Reino Unido, para as quais foram determinadas manualmente as densidades de congestionamentos de pessoas. As estimativas manuais serviram como parâmetros de comparação para as estimativas automáticas obtidas com as duas estratégias propostas. Metade dessas imagens foi atribuída ao conjunto de treinamento (as classificações são supervisionadas) e a outra metade ao conjunto de teste.

Embora a estratégia que utiliza medidas locais de textura seja, a princípio, mais adequada, pois ela considera o fato de poder existir diferentes densi-

dades de congestionamentos em sub-áreas da imagem, ela foi abandonada, pois exige a classificação de todos os pixels (ou de uma grande parte deles) da imagem, consumindo muito tempo do processamento, tornando esta estratégia inviável para ser aplicada na prática, onde se espera obter respostas em tempo-real. Esta estratégia foi avaliada utilizando-se apenas as medidas de texturas extraídas das matrizes de co-ocorrência dos tons de cinza, com o classificador neural, atingindo um índice de acerto de 76,51%, na classificação das imagens do conjunto de teste.

Os resultados dos experimentos obtidos com a estratégia que utiliza medidas globais de textura para a estimação de densidades de multidões permitiram concluir que as medidas de textura são bastante efetivas para tratar o problema em questão. Além desta conclusão principal, pode-se também notar a partir dos resultados obtidos que:

1. não houve diferenças significativas nos índices de acerto obtidos pelos três classificadores, que ficaram, na média, em torno de 76%. Entretanto, o classificador Bayesiano produziu o melhor resultado dos testes (85,52% de acerto), quando foi utilizado em conjunto com as medidas de texturas das matrizes de co-ocorrências;
2. as medidas de texturas baseadas nas matrizes de co-ocorrências produziram, na média dos classificadores, os melhores resultados, 82,76% de acerto, contra aproximadamente 74% de acerto das demais medidas avaliadas;
3. não houve diferenças significativas entre os resultados obtidos pelas medidas baseadas nos segmentos de retas, no espectro de Fourier e

nas dimensões fractais. Este fato indica que as informações de direcionalidade não são fundamentais para essa aplicação, uma vez que as medidas das dimensões fractais não dependem dessa informação, ao contrário das outras duas medidas, para as quais as informações sobre orientação são importantes.

Capítulo 8

Trabalhos Futuros

Ao final deste trabalho, propõem-se algumas extensões que são consideradas interessantes e importantes para complementar o trabalho realizado:

1. considerar sequências temporais de imagens para melhorar os resultados das estimações de densidades de congestionamentos de pessoas;
2. investigar técnicas para melhorar o desempenho da estratégia que utiliza medidas locais de textura para segmentação das imagens, para que ela possa apresentar resultados em tempo-real, uma vez que essa estratégia é bastante relevante para o problema tratado nesta tese;
3. reavaliar o conjunto de medidas de texturas utilizadas, incluindo novas medidas (utilizando campos markovianos, ou wavelets, por exemplo) e melhorando as já empregadas (é possível, por exemplo, melhorar as estimativas das dimensões fractais utilizando apenas as escalas espaciais para as quais os objetos nas imagens apresentam natureza fractal);

4. investigar mais profundamente a influência dos parâmetros envolvidos na captura das imagens na robustez das estratégias e métodos utilizados;
5. utilizar duas ou mais medidas distintas de texturas nos vetores de características de entrada dos classificadores;
6. incorporar conhecimento de alto nível na estimação de densidades de multidões, por exemplo, através da detecção das cabeças das pessoas na imagem.

Apêndice A

Estimativas Manuais das Densidades de Congestionamentos de Pessoas nas Imagens Utilizadas nos Experimentos

Classes	Conjunto de Treinamento	Conjunto de Teste
VL	1(12), 6(12), 17(13), 61(15), 64(10), 66(14), 68(9), 70(9), 76(7), 78(14), 86(14), 102(11), 105(12), 108(12), 123(14), 125(15), 134(14), 136(11), 138(12), 141(9), 144(15), 147(15), 207(14), 209(14), 211(12), 214(14), 216(7), 218(9), 220(9), 222(9), 223(10)	2(15), 15(15), 22(9), 59(11), 63(12) 65(8), 67(8), 69(12), 73(14), 77(13) 84(13), 101(11), 103(13), 107(14), 122(12), 124(13), 126(11), 135(10) 137(10), 140(13), 142(10), 145(14) 202(13), 208(12), 210(14), 212(14), 215(8), 217(11), 219(13), 221(9)
L	3(19), 5(18), 8(22), 10(23), 12(27), 14(24) 18(16), 20(16), 23(22), 25(27), 27(24), 56(29), 60(19), 71(17), 74(16), 79(20), 81(23), 83(22), 87(24), 89(28), 104(19), 112(21), 114(18), 116(21), 118(18), 120(24), 127(19), 129(20), 131(24), 133(18), 146(19), 167(28), 172(17), 177(29), 190(29), 201(20), 204(21), 206(16), 305(16)	4(21), 7(18), 9(22), 11(26), 13(20) 16(18), 19(26), 21(21), 24(27), 26(19) 28(28), 58(30), 62(16), 72(17), 75(19), 80(24), 82(27), 85(24), 88(26), 90(26), 106(19), 113(23), 115(23), 117(17), 119(20) 121(21), 128(19), 130(18), 132(26) 139(17), 163(27), 169(27), 175(30) 185(30), 197(28), 203(22), 205(23), 213(20), 306(18)
M	29(36), 31(44), 33(43), 35(42), 54(42), 57(37), 92(36), 94(45), 164(32), 166(31), 170(38), 174(33), 178(32), 184(35), 187(36), 189(34), 198(32), 224(31), 226(41), 228(38), 230(39), 232(40), 237(45), 292(33), 294(35), 300(36)	30(39), 32(40), 34(44), 37(44), 55(35) 91(34), 93(39), 162(42), 165(31) 168(36), 171(31), 176(33), 186(31) 188(37), 196(35), 200(36), 225(34) 227(41), 229(41), 231(39), 234(45) 291(35), 295(33), 301(37), 303(31), 304(33)
H	36(52), 40(52), 42(56), 95(59), 97(57), 99(57), 161(52), 195(47), 235(52), 5 238(48), 243(60), 245(53), 247(57), 249(50), 251(55), 254(59), 258(60), 290(51)	38(53), 41(54), 53(54), 96(59), 98(50) 143(60), 194(59), 233(47), 236(48) 239(53), 240(57), 244(56), 246(53) 248(52), 250(55), 252(58), 255(60) 289(52)
VH	43(66), 45(72), 47(77), 49(66), 51(66), 100(61), 149(75), 151(84), 153(81), 155(78), 157(69), 159(69), 173(64), 192(82), 253(61), 257(70), 261(72), 263(72), 265(68), 267(77), 269(77), 271(80), 273(78), 275(74), 277(82), 279(85), 281(81), 283(81), 285(70), 287(71), 296(71), 298(70)	44(69), 46(73), 48(61), 50(78), 52(64) 148(69), 150(76), 152(76), 154(72) 156(73), 158(82), 160(61), 191(67) 193(71), 256(67), 259(67), 262(74) 264(73), 266(72), 268(71), 270(75) 272(76), 274(78), 276(81), 278(76) 280(82), 282(79), 284(69), 286(82) 288(64), 297(84), 299(63)

Tabela A.1: Distribuição das imagens entre as classes de densidades e os conjuntos de treinamento e de teste. Os números fora dos parênteses são os identificadores das imagens e aqueles entre parênteses são as estimativas manuais de pessoas presentes em cada imagem.

Referências Bibliográficas

- [1] Ahuja, N. and Rosenfeld, A. “Mosaic models for textures”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-3**, pp. 1-11, 1981.
- [2] Ait-Kheddache, A. and Rajala, A. “Texture classification based on higher-order Fractals”, *IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 1112-1115, New York, USA, 1988.
- [3] Ballard, D. H. and Brown, C. M. *Computer Vision*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1982.
- [4] Batista, J. E. S. “Techniques for Computer-Based Anatomical Segmentation of the Brain Using MRI”, PhD. Thesis, University of London (Imperial College), London, UK, 1996.
- [5] Beck, J., Prazdny, K. and Rosenfeld, A. “A theory of textural segmentation”, *Human and Machine Vision*, pp. 1-38, Academic Press, London, 1983.

- [6] Bergen, J. and Adelson, E. "Early vision and texture perception", *Nature*, **vol. 333**, pp. 363-365, 1988.
- [7] Berlonghi, A. E. "Understanding and planning for different spectator crowds", *Safety Science*, **vol. 18**, pp. 239-247, 1995.
- [8] Blasdel, G. G. and Salama, G. "Voltage-sensitive dyes reveals a modular organization in monkey striate cortex", *Nature*, **vol. 321**, pp. 579-585, 1986.
- [9] Bovik, A. C., Clark, M. and Geisler, W. S. "Multichannel texture analysis using localized spatial filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-12**, pp. 55-73, 1990.
- [10] Brodatz, P. "Textures: a photographic album for artists and designers", Dover Publications, Inc., New York, 1966.
- [11] Brown, R. G. and Hwang, P. Y. C. *Introduction to Random Signal Analysis and Kalman Filtering*, Wiley, 1992.
- [12] Carter, P. H. "Texture Discrimination Using Wavelets", *SPIE Applications of Digital Image Processing XIV*, **vol. 1567**, pp. 432-438, 1991.
- [13] Castellano, G. "Hand Written Digits Recognition Using Hough Transform and Neural Networks", *King's College London - Report*, London, UK, 1995.
- [14] Chan, C. K. "Hough Transform techniques for pattern recognition", PhD. Thesis, University of London (King's College), London, UK, 1994.

- [15] Coelho, R. C. and Costa, L. da F. “The Box-Counting Dimension: Does it Provide an Accurate Subsidy for Experimental Shape Characterization? If So, How to Use It?”, *VIII Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens (SIBGRAPI)*, São Carlos, 1995.
- [16] Coelho, R. C. and Costa, L. da F. “On the Application of the Bouligand-Minkowski Fractal Dimension for Shape Characterisation”, *Applied Signal Processing*, **vol. 3**, pp. 163-176, 1996.
- [17] Cohen, A. and Kovacević, J. “Wavelets: the mathematical background”, *Proceedings of the IEEE*, **vol. 84**, pp. 514-522, 1996.
- [18] Connors, R. W. and Harlow, C. A. “A theoretical comparison of texture algorithms”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2**, pp. 204-222, 1980.
- [19] Costa, L. da F. “Effective detection of line segments with Hough Transform”, PhD. Thesis, University of London (King’s College), London, UK, 1992.
- [20] Costa, L. da F. “Effective detection of digital bar segments with Hough Transform”, *Computer Vision, Graphics, and Image Processing*, **vol. 55**, pp. 180-191, 1993.
- [21] Costa, L. da F., Roda, V. O. and Köberle, R. “A Biologically-Inspired System for Visual Pattern Recognition”, *Proceedings of IEEE International Symposium on Industrial Electronics*, Santiago, Chile, pp. 25-30, May 1994.

- [22] Cross, G. and Jain, A. K. "Markov Random Field Texture Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-5**, pp. 25-39, 1983.
- [23] Davies, A. C., Yin, J. H. and Velastin, S. A. "Crowd Monitoring Using Image Processing", *Electronics & Communication Engineering Journal*, pp. 37-47, February 1995.
- [24] Davis, L. S. et al. "An Empirical Evaluation of Generalized Cooccurrence Matrices", *IEEE Transactions on Pattern Analysis Machine Intelligence.*, **PAMI-3**, pp. 214-221, 1981.
- [25] Daubechies, I. "The wavelet transform, time-frequency localization and signal analysis", *IEEE Transactions on Information Theory*, **vol. 36**, pp. 961-1005, 1990.
- [26] Denslow, S., Zhang, Z., Thompson, R. P. and Lam, C. F. "Statistically characterized features for directionality quantitation in patterns and textures", *Pattern Recognition*, **vol. 26**, pp. 1193-1205, 1993.
- [27] Dickie, J. F. "Major Crowd Catastrophes", *Safety Science*, **vol. 18**, pp. 309-320, 1995.
- [28] Dougherty, E. R. , Newell, J. T. and Pelz, J. "Morphological Texture-Based Maximum-Likelihood Pixel Classification Based on Local Granulometric Moments", *Pattern Recognition*, **vol. 25(10)**, pp. 1181-1198, 1992.

- [29] Dubuisson, M. P. and Dubes, R. C. "Efficacy of fractal features in segmenting images of natural textures", *Pattern Recognition Letters*, **vol. 15**, pp. 419-431, 1994.
- [30] Duda, R. O. and Hart, P. E. "Use of the Hough Transform to detect lines and curves in pictures", *Communications of the ACM*, **vol. 15**, pp. 11-15, 1972.
- [31] Duda, R. O. and Hart, P. E. "Pattern Classification and Scene Analysis", *John Wiley & Sons, Inc.*, 1973.
- [32] Duncan, J. and Humphreys, G. W. "Visual search and stimulus similarity", *Psychological Review*, **vol. 96**, pp. 433-458, 1989.
- [33] Dunn, D., Higgins, W. E. and Wakeley, J. "Texture segmentation using 2-D Gabor elementary functions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-16**, pp. 130-149, 1994.
- [34] G. Eichmann and T. Kasparis, "Topologically Invariant Texture Description", *Computer Vision, Graphics, and Image Processing*, **vol. 41**, pp. 267-281, 1988.
- [35] Faugeras, O. D. and Pratt, W. K. "Decorrelation Methods of Texture Feature Extration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2**, pp. 323-332, July 1980.
- [36] Gabor, D. "Theory of Communication", *J. Inst. Elec. Eng.*, **vol. 93**, part III, pp. 429-459, 1946.

- [37] Gallant, J. L., Van Essen, D. C. and Nothdurft, H. C. "Two-Dimensional and Three-Dimensional Texture Processing in Visual Cortex of the Macaque Monkey", *Early Vision and Beyond* edited by Thomas Papathomas, A Bradford Book, pp. 89-98, 1994.
- [38] Galloway, M. M. "Texture Analysis Using Gray Level Run Lengths", *Computer Graphics and Image Processing*, **vol. 4**, pp. 172-179, 1975.
- [39] G. Gerig and F. Klein, "Fast contour identification through efficient Hough transform and simplified interpretation strategy", *In 8th International Joint Conference on Pattern Recognition*, Paris, France, pp. 498-500, 1986.
- [40] Gonzales, R. C. and Woods, R. E. *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.
- [41] Gulyas, B., Orban, G. A., Duysens, J. and Maes, H. "The suppressive influence of moving textured backgrounds on responses of cat striate neurons to moving bars", *J. Neurophysiology*, **vol. 57**, pp. 1767-1791, 1987.
- [42] Hammond, P. and MacKay, D. M. "Differential responsiveness of simple and complex cells in cat striate cortex to visual texture", *Exp. Brain Res.*, **vol. 30**, pp. 275-296, 1977.
- [43] Haralick, R. M. "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-3(6)**, pp. 610-621, 1973.

- [44] Haralick, R. M. "Statistical and Structural Approaches to Texture". *Proceedings of the IEEE*, **vol. 67(5)**, pp. 786-804, 1979.
- [45] Haralick, R. M. *Computer and Robot Vision*, **vol. 1**, Addison-Wesley Publishing Company, 1992.
- [46] Hawkins, J. K. "Textural Properties for Pattern Recognition", *Picture Processing and Psychopictorics*, B. C. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, pp. 347-370, 1970.
- [47] He, D. C. and Wang, L. "Detecting Texture Edges from Images", *Pattern Recognition*, **vol. 25(6)**, pp. 595-600, 1992.
- [48] Hess-Nielsen, N. and Wickerhauser, M. V. "Wavelets and time-frequency analysis", *Proceedings of IEEE*, **vol. 84**, pp. 523-540, 1996.
- [49] Hough, P. V. C. "Method and means for recognizing complex patterns", U. S. Patent 3,069,654, 1962.
- [50] Hubel, D. H. and Wiesel, T. N. "Receptive Fields and Functional Architecture in Monkey Striate Cortex", *The Journal of Physiology (London)*, **vol. 195**, pp. 215-244, 1968.
- [51] Hubel, D. H. and Wiesel, T. N. "Brain mechanisms of vision", *Scientific American*, pp. 130-144, Sept. 1979.
- [52] Hubel, H. D. "Eye, Brain, and Vision", Scientific American Library, 1988.

- [53] Illingworth, J. and Kittler, J. "The adaptative Hough Transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-9**, pp. 690-698, 1987.
- [54] Illingworth, J. and Kittler, J. "A survey of the Hough Transform", *Computer Vision, Graphics, and Image Processing*, **vol. 44**, pp. 87-116, 1988.
- [55] Jain, A. K. and Farrokhnia, F. (1991) "Unsupervised Texture Segmentation Using Gabor Filters", *Pattern Recognition*, **vol. 24**, **No. 12**, pp. 1167-1186.
- [56] Julesz, B. "Visual pattern discrimination", *IRE Transactions on Information Theory*, **IT-8**, pp. 84-92, 1962.
- [57] Julesz, B. "Experiments in the visual perception of texture", *Scientific American*, **vol. 232(4)**, pp. 34-43, 1975.
- [58] Julesz, B. "Spatial nonlinearities in the instantaneous perception of textures with identical power spectra", *Philosophical Transactions of the Royal Society of London B*, **vol. 290**, pp. 83-94, 1980.
- [59] Julesz, B. "Textons, the elements of texture perception, and their interactions", *Nature*, **vol. 290**, pp. 91-97, 1981.
- [60] Julesz, B. "A brief outline of the texton theory of human vision", *Trends in Neurosciences*, **vol. 7**, pp. 41-45, 1984.
- [61] Julesz, B. "Texton Gradients: The Texton Theory Revisited", *Biological Cybernetics*, **vol. 54**, pp. 245-251, 1986.

- [62] Kass, M. and Witkin, A. "Analyzing Oriented Patterns", *Computer Vision, Graphics and Image Processing*, **vol. 37**, pp. 362-385, 1987.
- [63] Kohonen, T. *Self-Organizing and Associative Memory*, Springer-Verlag, Berlin, 1989.
- [64] Kohonen, T. "The Self-Organizing Map", *Proceedings of the IEEE*, **vol. 78**, pp. 1464-1480, 1990.
- [65] Kohonen, T. "Things You Haven't Heard About Self-Organizing Map", *IEEE International Conference on Neural Networks*, **vol. III**, pp. 1147-1156, San Francisco, California, 1993.
- [66] Knierim, J. J. and Van Essen, D. C. "Neuronal responses to static texture patterns in area V1 of the alert macaque monkeys", *J. Neurophysiology*, **vol. 67**, pp. 961-980, 1992.
- [67] Laine, A. and Fan, J. "Texture classification by wavelets packet signatures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-15**, pp. 1186-1191, 1993.
- [68] Laws, K. I. "Texture Image Segmentation", *USC/PI Report 940*, University of Southern California, Image Processing Institute, January 1980.
- [69] Lendaris, G. C. and Stanley, G. L. "Diffraction Pattern Sampling for Automatic Pattern Recognition", *Proceedings of the IEEE*, **vol. 58(2)**, pp. 198-216, 1970.

- [70] Levine, M. D. "Vision in Man and Machine", *McGraw-Hill Publishing Co.*, 1985.
- [71] Li, H., Lavin, M. A. and Le Master, R. "Fast Hough Transform: a hierarchical approach", *Computer Vision, Graphics, and Image Processing*, **vol. 36**, pp. 139-161, 1986.
- [72] Li, S. Z. (1995) "Markov Random Field Modelling in Computer Vision", *Springer-Verlag*.
- [73] Liu, Z. Q., Rangayyan, R. M. and Frank, C. B. "Directional Analysis of Images in Scale Space", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-13**, pp. 1185-1192, 1991.
- [74] Lindeberg, T. "Scale Space Theory in Computer Vision", Kluwer Academic Publishers, 1994.
- [75] Loh, H. and Leu, J. "The Analysis of Natural Textures Using Run Length Features", *IEEE Transactions on Industrial Electronics*, **vol. 35(2)**, pp. 323-328, 1988.
- [76] Mallat, S. G. "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-11**, pp. 674-693, 1989.
- [77] Malik, J. and Perona, P. "Preattentive texture discrimination with early vision mechanisms", *J. Opt. Soc. Am. A*, **vol. 7**, pp. 923-932, 1990.

- [78] Mandelbrot, B. B. “Fractals: form, chance and dimension”, *W. H. Freeman*, San Francisco, California, 1977.
- [79] Mandelbrot, B. B. “The Fractal Geometry of Nature”, *W. H. Freeman*, San Francisco, California, 1982.
- [80] Manjunath, B. S. and Chellappa, R. “Unsupervised Texture Segmentation Using Markov Random Field Models”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-13**, pp. 478-482, 1991.
- [81] Marana, A. N., Costa, L. da F. e Lotufo, R. A. “Análise de textura utilizando transformada de Hough e Morfologia Matemática”, *VIII Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens (SIBGRAPI)*, pp. 135-142, São Carlos, 1995.
- [82] Marana, A. N., Costa, L. da F., Velastin, S. A. and Lotufo, R. A. “Oriented texture classification based on a self-organizing neural network and Hough transform”, *IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, **vol. 4**, pp. 2773-2776, Munich, Germany, April 1997.
- [83] Marana, A. N., Velastin, S. A. , Costa, L. da F. and Lotufo, R. A. “Automatic Estimation of Crowd Density Using Texture”, *Safety Science*, (no prelo).
- [84] Marana, A. N., Velastin, S. A. , Costa, L. da F. and Lotufo, R. A. “Estimation of Crowd Density Using Image Processing”, *IEE Colloquium*

- on Image Processing for Security Applications*, pp. 11/1-11/8, London, UK, March 1997.
- [85] Marana, A. N., Velastin, S. A. , Costa, L. da F. and Lotufo, R. A. “Automatic Estimation of Crowd Density Using Texture”, *4th International Workshop on Systems, Signals, and Image Processing*, Eds. M. Domanski and R. Stasinski, pp. 177-180, Poznan, Poland, May 1997.
 - [86] Marr, D. C. “Early processing of visual information”, *Philosophical Transactions of the Royal Society of London B*, **vol. 275**, pp. 483-519, 1976.
 - [87] Marr, D. C. and Hildreth, E. C. “Theory of edge detection”, *Proceedings of the Royal Society of London B*, **vol. 207**, pp. 187-217, 1980.
 - [88] Marr, D. C. “Vision: a computational investigation into the human representation and processing of visual information”, *W. H. Freeman*, San Francisco, California, 1982.
 - [89] Matheron, G. *Random Sets and Integral Geometry*, Wiley, New York, 1975.
 - [90] McCormick, B. H. and Jayaramamurthy, S. N. “Time Series Models for Texture Synthesis”, *International Journal of Computer and Information Sciences*, **vol. 3**, pp. 329-343, 1974.

- [91] Medioni, G. G. and Yasumoto, Y. "A Note on Using the Fractal Dimension for Segmentation", *Proc. IEEE Workshop on Computer Vision*, Annapolis, MD, pp. 25-30, 1984.
- [92] Modestino, J. W., Fries, R. W. and Vickers, A. L. "Texture discrimination based upon an assumed stochastic texture model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-3**, pp. 557-580, 1981.
- [93] Nicholson, C. E. and Roebuck, B. "The Investigation of the Hillsborough disaster by the Health and Safety Executive", *Safety Science*, **vol. 18**, pp. 249-259, 1995.
- [94] Nothdurft, H. C. and Li, C. Y. "Representation of spatial details in textured patterns by cells of the cat striate cortex", *Experimental Brain Research*, **vol. 57**, pp. 9-21, 1984.
- [95] Nothdurft, H. C. "Texture discrimination by cells in the cat lateral geniculate nucleus", *Experimental Brain Research*, **vol. 82**, pp. 48-66, 1990.
- [96] Pentland, A. P. "Fractal-Based Description of Natural Scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, pp. 661-674, 1984.
- [97] Perry, A. and Lowe, D. G. "Segmentation of textured images", *Proceedings of IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition*, San Diego, California, pp. 326-332, 1989.

- [98] Pickett, R. M., "Visual Analysis of Texture in the Detection and Recognition of Objects", *Picture Processing and Psychopictorics*, B. C. Lipkin and A. Rosenfeld, Eds., Academic Press, New York, pp. 289-308, 1970.
- [99] Polus, A., Schofer, J. L., and Ushpiz, A. (1983) "Pedestrian Flow and Level of Service", *Journal of Transportation Engineering*, Vol. 109, No. 1, pp. 46-56.
- [100] Pratt, W. "Digital Image Processing", *John Wiley & Sons*, 1991.
- [101] Pressman, N. J. "Markovian Analysis of Cervical Cell Images", *The Journal of Histochemistry and Cytochemistry*, vol. 24(1), pp. 138-144, 1976.
- [102] Rao, A. R. "A Taxonomy for Texture Description and Identification", PhD. Thesis, Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, USA, 1989; Springer-Verlag, Berlin/New York, 1990.
- [103] Rao, A. R. and Schunck, B. G. "Computing Oriented Texture Fields", *CVGIP: Graphical Models and Image Processing*, vol. 53(2), pp. 157-185, 1991.
- [104] Raghu, P. P., Poongodi, R. and Yegnanarayana, B. "A Combined Neural Network Approach for Texture Classification", *Neural Networks*, vol. 8, pp. 975-987, 1995.

- [105] Rosenfeld, A. and Troy, E. "Visual Texture Analysis", *TR-116*, University of Maryland, Computer Science Center, June 1970.
- [106] Serra, J. "Image Analysis and Mathematical Morphology", Academic Press, New York/London, 1982.
- [107] Sime, J. D. "Crowd Psychology and Engineering", *Safety Science*, **vol. 21**, pp. 1-14, 1995.
- [108] Stiny, G. and Gips, J. "Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts", *University of California Press*, Berkeley, California, 1972.
- [109] Sun, C. and Wee, W. G. "Neighboring Gray Level Dependence Matrix for Texture Classification", *Computer Vision, Graphics and Image Processing*, **vol. 23**, pp. 341-352, 1983.
- [110] Tamura, H., Mori, S. and Yamawaki, T. "Textural Features Corresponding to Visual Perception", *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-8**, pp. 460-473, 1978.
- [111] Tan, T. N. and Constantinides, G. "Texture analysis based on a human visual model", *Proceedings of IEEE Int. Conf. Acoust., Speech and Signal Processing*, Albuquerque, New Mexico, pp. 2091-2110, April 1990.
- [112] Therrien, C. W. "Decision, Estimation and Classification: An Introduction to Pattern Recognition and Related Topics", John Wiley & Sons, Inc., 1989.

- [113] Tieniu, T. "Image Texture Analysis: Classification and Segmentation", PhD. Thesis, University of London (Imperial College), London, UK, 1989.
- [114] Treisman, A. M. and Gelade, G. "A Feature-integration theory of attention", *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.
- [115] Treisman, A. N. "Preattentive processing in vision", *Computer Vision, Graphics and Image Processing*, vol. 31, pp. 156-177, 1985.
- [116] Treisman, A. M. and Gormican, S. "Feature analysis in early vision: evidence from search asymmetries", *Psychological Review*, vol. 95, pp. 15-48, 1988.
- [117] Turner, M. R. "Texture discrimination by Gabor functions", *Biological Cybernetics*, vol. 55, pp. 71-82, 1986.
- [118] Tzvi, D. B. "New technique for the Hough Transform for shape representation", PhD. Thesis, University of London (King's College), London, UK, 1990.
- [119] Unser, M. and Coulon, F. "Detection of Defects by Texture Monitoring in Automatic Visual Inspection", *Proc. of the Second Int. Conf. on Robot Vision and Sensory Control*, pp. 27-38, Stuttgart, Germany, November 1982.
- [120] Unser, M. "Texture classification and segmentation using Wavelet Frames", *IEEE Transactions on Image Processing*, vol. 4, pp. 1549-1560, 1995.

- [121] Uttal, W. R. "A Taxonomy of Visual Processes", Lawrence Erlbaum Associates, Hillsdale, N. J., 1981.
- [122] Van Gool, L. "Texture Analysis Anno 1983", *Computer Vision, Graphics, and Image Processing*, **vol. 29**, pp. 336-357, 1985.
- [123] M. Vanrell and J. Vitrià, "Mathematical Morphology, Granulometries and Texture Perception", *SPIE*, **vol. 2030**, pp. 152-161, 1993.
- [124] Von der Heydt, R. "Periodic-Pattern-Selective Cells in Monkey Visual Cortex", *The Journal of Neuroscience*, **vol. 12**, pp. 1416-1434, 1992.
- [125] Wechsler, A. and Kidode, M. "A Random Walk Procedure for Texture Discrimination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-1**, pp. 272-280, 1979.
- [126] Weszka, J. S. , Dyer, C. R. and Rosenfeld, A. "A comparative study of texture measures for terrain classification", *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-6**, pp. 269-285, 1976.
- [127] Winder, C. P. "Markovian Analysis of Texture: Serial and Parallel Paradigms in Low-Level Vision", PhD. Thesis, University of Oxford (St. Hugh's College), Oxford, UK, 1991.