UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA

"TÉCNICAS DE MINERAÇÃO DE DADOS PARA ANÁLISE DE IMAGENS"

LUÍS AUGUSTO CONSULARO

Tese apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para obtenção do título de Doutor em Ciências "Física Aplicada – Opção Física Computacional"

USP/IFSC/SBI
8-2-001364

Orientador:
Prof. Dr. Luciano da Fontoura Costa

Comissão Julgadora:
Prof. Dr. Luciano da Fontoura Costa (IFSC-USP)
Prof. Dr. Jan Frans Willem Slaets (IFSC-USP)
Prof. Dr. Nelson Delfino D'Ávila Mascarenhas (UFSCar)
Prof. Dr. Roberto Marcondes Cesar Júnior (IME-USP)
Profa. Dra. Olga Regina Pereira Bellon (UFPr)

São Carlos – São Paulo

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

Consularo, Luís Augusto
Técnicas de Mineração de Dados para Análise de Imagens/Luís Augusto Consularo – São Carlos, 2000.
236 pp.

Tese (Doutorado) – Instituto de Física de São Carlos, 2000.

Orientador: Prof. Dr. Luciano da Fontoura Costa
1. Mineração de Dados. 2. Reconhecimento de Padrões.
3. Visão Computacional
I. Título.

UNIVERSIDADE DE SÃO PAULO
Instituto de Física de São Carlos

Caixa Postal 369 - CEP 13560-970 - São Carlos - SP
Fone: (0xx16) 273-9777 - Fax: (0xx16) 273-9777
E-mail: wladerez@if.sc.usp.br

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE LUIS AUGUSTO CONSULARO APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, DA UNIVERSIDADE DE SÃO PAULO, EM 26 DE SETEMBRO DE 2000.

COMISSÃO JULGADORA:

Prof. Dr. Luciano da Fontoura Costa/IFSC-FFI-USP

Prof. Dr. Jan Frans Willem Slaets/IFSC-FFI-USP

Nelson Delfino d'Ávila Mascarenhas

Prof. Dr. Nelson Delfino D'Ávila Mascarenhas/UFSCar

Prof. Dr. Roberto Marcondes Cesar Júnior/IME-USP

Profa. Dra. Olga Regina Pereira Bellon/UFPr

AGRADECIMENTOS

Aos meus pais, pelo eterno estímulo.

Aos meus amigos conquistados no quotidiano do laboratório.

A minha amiga Regina Célia Coelho, pelo companheirismo ímpar.

A minha amiga Júlia Sawaki Tanaka, pelos inúmeros apoios.

A minha amiga Andréa Gomes Campos, pela força de sua presença.

Ao meu orientador, Prof. Dr. Luciano da Fontoura Costa, pela imensa paciência e pelas oportunidades oferecidas.

SUMÁRIO

1 INTRODUÇÃO **1**
2 PERCEPÇÃO VISUAL E DESCOBERTA DE CONHECIMENTO **9**
3 MINERAÇÃO DE DADOS **13**
3.1 INFRA-ESTRUTURA 16
3.2 ABORDAGEM CONVENCIONAL 16
3.2.1 Data Warehousing *17*
3.2.2 OLAP (On-Line Analytic Processing) *20*
3.2.3 Mineração de Dados – DM (Data Mining) *22*
3.3 TÉCNICAS DE MINERAÇÃO DE DADOS 37
3.3.1 Teoria de Decisão de Bayes *39*
3.3.2 Vizinho mais próximo *41*
3.3.3 Agrupamentos Particionais (Clustering) *44*
3.3.4 Agrupamento Hierárquico *52*
3.3.5 Mapeamento Auto-organizável *61*
3.3.6 Algoritmo Genético *63*
3.3.7 Aproximação *71*
4 PROSPECÇÃO DE ATRIBUTOS **77**
4.1 HISTOGRAMA DE LINHAS DE CAMPO ELÉTRICO INCIDENTES 78
4.1.1 Campos conservativos usando a Transformada de Fourier *79*
4.1.2 Aplicação para interação entre neurônios *81*
4.1.3 Obtendo atributos *86*
4.2 ESQUELETONIZAÇÃO MULTI-ESCALAS COM MÉTRICAS DE MINKOWSKI VARIÁVEIS 90
4.2.1 Métricas Alternativas *91*
4.2.2 Esqueletonização multi-escalas sem deslocamento de borda *100*
4.2.3 A Propagação da Distância Euclideana *101*
4.2.4 Esqueletonização multi-escalas com métrica "chessboard" *108*
5 ANÁLISE DE FORMAS **115**
5.1 SALIÊNCIA 116
5.1.1 Por que a saliência é importante? *119*
5.1.2 Principais técnicas *119*
5.2 APRENDIZADO DE PONTOS DE SALIÊNCIA POR GENERALIZAÇÃO DE PARÂMETROS 122
5.2.1 Aplicação: Forma de Neurônios *123*
5.2.2 Aprendizado *131*
5.3 PERSEGUIÇÃO DE PONTOS SIGNIFICATIVOS 135
5.3.1 Perseguição à saliência *137*
5.3.2 Resultados e desempenho *140*
5.3.3 Conclusões *147*

6 ANÁLISE PICTÓRICA 149
6.1 MEDIDAS 155
6.1.1 texturais 155
6.1.2 contraste 161
6.1.3 complexidade multi-escalas 165
6.2 A BUSCA PELO MODELO 178
6.3 A GENERALIZAÇÃO 183
6.4 A BUSCA 186
6.5 APLICANDO O MODELO 187
6.6 COMENTÁRIOS E CONCLUSÕES 200
7 CONCLUSÕES 202
8 BIBLIOGRAFIA 206
9 APÊNDICE 219
9.1 RELAÇÕES 219
9.2 ENTROPIA E VARIÂNCIA MULTI-ESCALAS 221
9.3 COMPARAÇÃO DE DIMENSÃO FRACIONÁRIA VIA SALSICHAS DE MINKOWSKI 228
9.4 IMAGENS DE *HOME-PAGES* UTILIZADAS NO EXPERIMENTO DE ANÁLISE PICTÓRICA 232
9.5 NOTAS MÉDIAS PARA CADA IMAGEM DA SEÇÃO ANTERIOR 237

LISTA DE FIGURAS

Figura 1 – Imagem obtida por satélite de uma região de cultivo. 2

Figura 2 – A análise de uma imagem a partir de um processo de segmentação com parâmetros e critérios conhecidos. 3

Figura 3 – Etapa na qual se obtem os modelos de treinamento. 5

Figura 4 – Etapa na qual se obtém medidas e os modelos do objeto de teste. As respostas de teste são comparadas às respostas do modelo acarretando um erro. 5

Figura 5 – Etapa na qual o melhor conjunto de medidas é selecionado, baseado no menor erro. 6

Figura 6 – Arquitetura para uma abordagem representacional de percepção. 12

Figura 7 – O processo de mineração de dados segundo [FAYYAD et al., 1996]. 14

Figura 8 – Exemplo de dados transacionais: (*a*) caixa eletrônico; (*b*) operações bancárias com celulares ou computadores portáteis; (*c*) operações de comércio eletrônico via internet. 17

Figura 9 – Exemplo de dados operacionais: fluxo de caixa, folha de pagamento, controle de estoque, faturamento. 18

Figura 10 – Exemplo de dados informacionais em bolsa de valores. 18

Figura 11 – Ilustrações de relacionamentos típicos em dados operacionais e transacionais (*a*) e para dados informacionais (*b*) 19

Figura 12 – Visão multidimensional de dados fornecida por um OLAP. 21

Figura 13 – Taxonomia proposta para as abordagens sobre mineração de dados. 28

Figura 14 – Uma árvore de classificação (*a*) e uma árvore de regressão (*b*). (adapatado de [DHAR & STEIN, 1997]) 30

Figura 15 – Modelo para reconhecimento estatístico de padrões [JAIN et al., 2000]. 31

Figura 16 – Abordagens de mineração de dados baseadas em aprendizado. 32

Figura 17 - Abordagens de mineração de dados baseadas em aprendizado não supervisionado. 33

Figura 18 - Tipos de aprendizados: (*a*) supervisionado; (*b*) não-supervisionado; (*c*) com reforço 35

Figura 19 - O critério para que um vizinho seja agrupado. 42

Figura 20 – Comparação dos limites superiores de erro para diferentes *k* [DUDA et al., 1998]. 44

Figura 21 – Agrupamentos pelo algoritmo CLARA considerando agrupamentos aleatórios. 49

Figura 22 – Exemplo que será submetido a agrupamentos. São amostras identificadas por números em um espaço formado por duas medidas normalizadas entre 0 e 1. 54

Figura 23 – Resultado de um agrupamento hierárquico dos objetos ilustrados na Figura 22 com ligação simples visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn 55

Figura 24 - Resultado de um agrupamento hierárquico dos objetos ilustrados na Figura 22 com ligação completa visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn 56

Figura 25 – Comparações entre aplicações de agrupamentos hierárquicos com diferentes ligações (*a*) simples; (*b*) completa; (*c*) média. 57

Figura 26 – Diferença das abordagens para distância por (*a*) média aritmética; e (*b*) centróide. 58

Figura 27 - Resultado de um agrupamento hierárquico pelo método Ward dos objetos ilustrados na Figura 22 e visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn 60

Figura 28 – Estrutura de dados para um SOM 61

Figura 29 – Estrutura de dados para um algoritmo genético com genes binários. 64

Figura 30 – Exemplo de espaço de busca e funções objetivo (*a*) máximos (*b*) e (*c*) cruzamentos no zero. 65

Figura 31 – Explicação sinótica sobre algoritmo genético, suas operações básicas e estruturas de dados. 69

Figura 32 – Exemplos de (a) regressão linear; (b) regressão exponencial; (c) interpolação. 72

Figura 33 – Ilustração de interpolação com FBR por uma superfície. (*a*) Os pontos a serem interpolados são representados por bolas vermelhas e suas respectivas posições por cruzes azuis. (*b*) A superfície interpolando os pontos e a generalização de um ponto (bola azul). 73

Figura 34 – Exemplos de graus de liberdade de FBRs com suporte elíptico: (*a*) rotação; (*b*) translação. 75

Figura 35 – Uma imagem colorida com frutas identificáveis no espaço de cores (*a*); e uma versão em níveis de cinza (*b*). 77

Figura 36 – Funções de campo escalar usadas para calcular as componentes x (*a*) e y (*b*). 81

Figura 37 – Exemplo da função campo escalar (*b*) em uma imagem com dois neurônios (*a*) simulando uma interação. 82

Figura 38 – Dois neurônios e o campo vetorial formado por sua influência para uma lei de campo $a = 4$. 82

Figura 39 - Histograma de linhas de campo incidentes em dois neurônios 83

Figura 40 – Histograma de incidências de linhas de campo no par de formas geométricas. 83

Figura 41 – Figuras geométricas (a) tomadas em diferentes deslocamentos com as respectivas funções pontenciais (b) e campos vetoriais (*c*) f sobrepostos a $U(x,y)$. 84

Figura 42 – Figuras geométricas em um campo vetorial variando a lei (coeficiente a) 85

Figura 43 – Influência de um par de neurônios a partir de uma circunferência a qual é avaliado em cada ponto do contorno a densidade de caminhos percorridos por uma partícula. (a) mostra os caminhos para lei de campo $a = 2$ e (b) para lei de campo $a = 4$. 86

Figura 44 – Contorno de um neurônio. 86

Figura 45 – Componentes x (*a*) e y (*b*) do campo vetorial calculados aplicando-se (48) e (49) respectivamente. 87

Figura 46 – Campo escalar (*a*) e vetorial (*b*) para o contorno da Figura 44. 87

Figura 47 – Caminhos a partir de pontos de uma circunferência em torno da forma (*a*) e a partir de todos os pixels externos ao contorno (*b*). 88

Figura 48 – Gráfico mostrando para cada ponto do contorno a sua influência pela chegada de linhas de força. A lei de campo corresponde ao índice *a=1*. 88

Figura 49 - Linhas de força calculadas para leis de campo *a = 2* (*a*) e *a = 4* (b)..... 89

Figura 50 - Gráficos mostrando para cada ponto do contorno a sua influência pela chegada de linhas de força. As leis de campo correspondem aos índices *a=2* (*a*) e *a=4* (*b*). 89

Figura 51 – Comparação entre a curvatura do contorno da Figura 44 e seu histograma de linhas de forças para *a = 2*. 90

Figura 52 – Família de distâncias de Minkowski. As linhas mais fortes definem casos especiais. 92

Figura 53 – Figura contendo diversas características. 94

Figura 54 – Distâncias Minkowski aplicadas à Figura 53 com índices *n = 1*(*a*), *1.5*(*b*), *2*(*c*) e *10*(*d*). 95

Figura 55 – Distâncias *chamfer* aplicadas à Figura 53 correspondentes aos índices (1, ∞) *city block* (*a*); octagonal (*b*); (3, 4) (*c*); (1,1) *chessboard* (*d*). 95

Figura 56 – Silhuetas de mão humana (*a*) rotacionadas em (*b*) 30° e (*c*) 45°. 96

Figura 57 – Mapas de distâncias de uma mão humana rotacionada em diferentes métricas: (*a*) cityblock; (*b*) chessboard; (*c*) octagonal; (*d*) chamfer(3,4); (*e*) euclideana. 97

Figura 58 – Efeito da rotação sobre a distância de uma linha reta vertical e outra rotacionada de 45° 98

Figura 59 - Estrutura SD para propagar a distância euclideana com dilatação exata 101

Figura 60 - Figura original *(a)*; Imagem da rotulação de cada elemento do contorno *(b)*; Imagem da propagação dos rótulos do contorno pela dilatação exata. 102

Figura 61 - Cálculo da matriz de diferenças *(a)*; Resultado da aplicação do algoritmo ilustrado ao lado *(b)* mostrando os esqueletos interno e externo 103

Figura 62 – Fatias aplicadas sobre a imagem de diferenças da Figura 61b. As escalas são apresentadas em ordem crescente: (*a*) *T* = 6; (*b*) *T* = 20; (*c*) *T* = 50; (*d*) *T* = 75; (*e*) *T* = 150; (*f*) *T* = 265. As linhas mais escuras (vermelhas) indicam a fatia e as mais claras (verdes), a Figura original. 104

Figura 63 – Imagem com os discos de reconstrução da forma da Figura 60*a* em multiplas escalas. Os níveis de cinza indicam cada escala. 105

Figura 64 – Reconstrução para os respectivos esqueletos da Figura 62. 106

Figura 65 – Variação do número de distâncias com relação ao raio de uma círculo digital. Bolas vazadas correspondem à distância euclideana; bolas cheias à *Chamfer*(3,4); losangos, quadrados e triângulos (sobrepostos) correspondem respectivamente à *city block*, *chessboard* e octagonal. 107

Figura 66 – Esqueletos para a forma da Figura 60*a* extraídos (*T = 9*) usando a métrica de Minkowski com coeficientes (*a*) *n = 1*; (*b*) *n = 1.5*; (*c*) *n = 2*; (*d*) *n = 10*. 107

Figura 67 – Esqueletos para a forma da Figura 60*a* extraídos (*T = 9*) usando métodos *chamfer* para (*a*) *city block*; (*b*) octagonal; (*c*) *chamfer*(3,4); (*d*) *chessboard*. 108

Figura 68 – Pré-processamento de uma forma para a esqueletonização multiescalas com métrica chessboard. (*a*) Forma original predominantemente ramificada; (*b*) contorno extraído; (*c*) transformada distância chessboard de *b*; (*d*) rótulos propagados para o contorno *b*. 110

Figura 69 – Esqueleto interno multi-escalas da forma na Figura 68*a*. 112

Figura 70 – Reconstrução (em preto) dos esqueletos (linha em branco) da forma da Figura 68*a*. 113

Figura 71 – Desenho de um gato dormindo elaborado utilizando 38 saliências [ATTNEAVE, 1954]. 118

Figura 72 – Esquema ilustrando a aplicação de calibração de parâmetros de segmentação utilizando a curvatura de formas neurais. 124

Figura 74 - Exemplo de forma neural ganglionar (a) e de marcação de ramificações (o) e extremidades (x) 126

Figura 75 – Parâmetros utilizados para calibrar a curvatura: T^+ para terminações ou extremidades e T^- para ramificações. (*a*) ilustra uma escala $a = 28$ e (*b*) em uma escala $a = 3$. 127

Figura 77 – Efeito da aplicação da função sigmóide na curvatura. (*a*) contorno segmentado; (*b*) curvatura sem aplicação da sigmóide; (*c*) a função sigmóide com k=0.5; (*d*) curvatura com aplicação da sigmóide. 129

Figura 79 - Generalização dos parâmetros de curvatura por interpolações abrangendo os parâmetros globais (energia de dobramento – *B* versus coeficiente de variação – *CV*) e os parâmetros de segmentação: (a) escala para terminações (a^+); (b) escala para ramificações (a^-); (c) limiar para terminações (T^+); (d) limiar para ramificações (T^-); (e) sigmóide (*S*). 132

Figura 81 - Resultados conseguidos para os neurônios treinados (em preto) e neurônios de teste (em cinza). As escalas vão até 1, ou seja, o maior número de acertos e o menor número de erros. Os gráficos de barras em (*a*) são para ambos: terminações e ramificações; (*b*) apenas para terminações; e (*c*) apenas para ramificações. 135

Figura 82 - Resultado da propagação de rótulos com a restrição $d < r$ *(a)*; Visão tridimensional do resultado da aplicação do algoritmo de diferenças *(b)*. 138

Figura 84 - Resultado da limiarização do esqueleto 138

Figura 86 - O caminho vermelho ilustra o processo de perseguição à extremidade 139

Figura 88 - Imagem do esqueleto multi-escalas limiarizada em diferentes escalas. 140

Figura 89 - Saliências detectadas para algumas escalas. 140

Figura 90 - Desempenho do método proposto conforme varia o raio. 142

Figura 91 - Desempenho com relação ao número de elementos no contorno. 143

Figura 92 - Losangos com ângulos internos variando e suas respectivas detecções de saliência. 143

Figura 93 - Submetendo o algoritmo proposto à forma acima o número embaixo de cada Figura são respectivamente *r* e o *thresholding*. 145

Figura 94 - Submetendo o algoritmo a imagem rotacionada. 146

Figura 95 - Exemplo de extremidades na escala detectada. 147

Figura 96 - Aplicação do método proposto para identificar estruturas salientes em uma forma. 147

Figura 97 – Esquema para a construção do modelo de avaliação estética utilizando medidas texturais, de complexidade e de contraste. 151

Figura 99 – Experimento psicofísico 153

Figura 101 – Formulário para respostas a que os sujeitos foram submetidos no experimento psicofísico 154

Figura 103 - O espaço do espectro de potência é dividido em setores, dentro dos quais seus elementos são somados para formar descritores das variações espaciais 156

Figura 105 – Exemplos de descritores espectrais para texturas com detalhes (*a*) pequenos; (*b*) médios; (*c*) grandes. As variações do descritor espectral (*d*) espacial e (*e*)

direcional das respectivas imagens são ilustradas com os tons da barra abaixo de cada Figura correspondente. 157

Figura 107 - O espaço do espectro de potência é dividido em trilhas concêntricas, dentro dos quais seus elementos são somados para formar descritores das variações direcionais. 158

Figura 109 – Exemplos de descritores espectrais direcionais para imagens com retas de diferentes orientações predominantes (*a*) 90°; (*b*) 135°; (*c*) 0°; (*d*) 45°. No gráfico (*e*) os picos em cada tom indicam a ordem das disposições das imagens. 159

Figura 111 – Exemplo de textura real com orientação predominante. A imagem com varas de cana-de-açúcar (*a*) estão orientadas predominantemente na direção de $\pi/2$ (45°). O gráfico em (*b*) mostra essa predominância com um pico na direção ortogonal. 160

Figura 113 – Exemplo de imagem com mais de uma orientação (a) e o gráfico (b) de seu respectivo descritor espectral direcional com picos nas duas direções ortogonais. 160

Figura 115 – Exemplo de imagem apresentando uma orientação predominante, mas degradada com granularidade (a) e o gráfico (b) de seu descritor espectral direcional refletindo essa degradação. 161

Figura 117 – Ilustração do processo de extração do histograma de contraste em níveis de cinza. A imagem original é varrida por uma máscara de vizinhança 4. A diferença média entre o pixel corrente e os vizinhos para toda a imagem resulta na imagem de diferenças da qual é calculado o histograma. 162

Figura 118 – Exemplos de histogramas de contrastes em níveis de cinza para imagens de (*a*) um desenho; (*b*) um desenho gráfico; (*c*) uma fotografia. Percebe-se que os histogramas de cada imagem são distintos e podem caracterizar o processo de confecção da imagem. 163

Figura 120 – Ilustração do processo de extração do histograma de contraste em cores de uma imagem. Cada pixel é um vetor no espaço RGB e o contraste é caracterizado pelo produto escalar entre o pixel corrente e seus 4 vizinhos. O histograma de médias dos produtos escalares é o conjunto de medidas que caracteriza esse contraste. 164

Figura 121 – Exemplo de histogramas de contrastes para uma imagem colorida com grandes contrastes (a) e que em seguida é sucessivamente borrada para diluir as bordas (b); (c) e (d). Percebe-se, conforme esperado, uma uniformização do histograma de contraste de cores conforme o borramento aumenta. 165

Figura 123 – Conjunto de gaussianas usadas para os borramentos sucessivos, com o intuito de obter as várias escalas nas medidas de entropia e variância multi-escalas. (a) no espaço de origem; (b) noespaço de freqüência. 167

Figura 125 – Três tipos de imagens (a) desenho com linhas; (b) desenho gráfico com uso de sombreamentos; (c) fotografia de um olho humano. Ao lado, o gráfico de entropia multi-escalas de cada imagem. Percebe-se um comportamento diferente no desenho de linhas (a) e pouca diferença entre (b) e (c). A coluna mais à direita mostra a entropia média e o desvio padrão. 170

Figura 126 - (a) desenho com linhas; (b) desenho gráfico com uso de sombreamentos; (c) fotografia de um olho humano. Ao lado, o gráfico de variância multi-escalas de cada imagem. A maior discrepância está no desenho de linhas. Na coluna mais à direita aparecem as medidas resumo: variância média, desvio padrão da variância multi-escalas e a escala de meia-vida. 172

Figura 128 – Curva de von Koch em cada escala. Ao lado são mostrados o número de segmentos ($N(s)$) e o tamanho de cada segmento s. 173

Figura 130 – Floco de neve de von Koch, iniciado a partir de um triângulo e suas respectivas salsichas de Minkowski. Abaixo o gráfico da esquerda mostra a evolução da área da salsicha em um log-log. À direita o gráfico mostrando para cada raio de

dilatação a dimensão fracionária estimada. Percebe-se que a dimensão teórica está mais próxima quando o r=6. 174

Figura 132 – Dilatações crescentes da salsicha de Minkowski da imagem da Figura 20 para diferentes raios r. A coluna (a) corresponde à parte de cima; a (b) à parte de baixo; e (c) todo o volume. 177

Figura 134 – Volumes da salsicha de Minkowski em um gráfico (a) log-log para a imagem (b) 178

Figura 136 – Esquema de comunicação estética proposto por Bense em [BENSE, 1975]. 179

Figura 138 – Modelo "caixa preta" para o ser humano. 180

Figura 140 – Esquema ilustrando a avaliação do desempenho no treinamento da estimativa de critério estético de páginas web, baseado em notas atribuídas por sujeitos humanos. 180

Figura 141 – Gráfico que mostra o resultado de uma estimativa de critério estético para 40 imagens. Os quadrados preenchidos representam as notas médias das imagens para treinamento; os quadrados vazados, as notas das imagens de teste e o ponto a estimativa para a imagem de teste. A linha indica o erro da estimativa. 183

Figura 143 – Janela do aplicativo para testar a TPS. 188

LISTA DE TABELAS

Tabela 1 – Índices para a fórmula genérica para a regra de decisão em (18) [JAIN & DUBES, 1988]. 59

Tabela 2 - Histogramas de erros para originalidade, usando o método de força bruta para combinar duas medidas 189

Tabela 3 - Histogramas de erros para agradabilidade usando o método de força bruta para combinar três medidas. 190

Tabela 4 - As medidas selecionadas com o método "força bruta" e os erros mínimos para as várias FBRs consideradas como funções interpolantes (considerando duas medidas). 191

Tabela 5 - As medidas selecionadas com o método "força bruta" e os erros mínimos para as várias FBRs consideradas como funções interpolantes (considerando duas medidas). 191

Tabela 6 - Erros para o treinamento de agradabilidade, cores, originalidade, qualidade artística e motivação. 192

Tabela 7 – Demonstrativo de três execuções para os critérios em *C*. AGR: agradabilidade; COR: cores; ORG: originalidade; QLD: qualidade; MTV: motivação. Os números nos círculos indicam cada evolução do algoritmo genético. 193

Tabela 8 – Resultado de duas e três evoluções do GA para o atributo entropia. 194

Tabela 9 - Resultado de duas e três evoluções do GA para o atributo variância. 195

Tabela 10 - Resultado de duas e três evoluções do GA para as salsichas de Minkowski. 196

Tabela 12 - Resultado de duas e três evoluções do GA para os descritores espectrais espaciais. 197

Tabela 13 - Resultado de duas e três evoluções do GA para os descritores espectrais direcionais. 198

Tabela 14 - Resultado de duas e três evoluções do GA para contraste de níveis de cinza. 199

Tabela 15 - Resultado de duas e três evoluções do GA para contraste de cores. 200

LISTA DE ALGORITMOS

Algoritmo 1 – *k-médias* 44
Algoritmo 2 - ISODATA 46
Algoritmo 3 - Leader 47
Algoritmo 4 - CLARA 48
Algoritmo 5 - CLARANS 50
Algoritmo 6 – Algoritmo de Johnson 53
Algoritmo 7 – Algoritmo Genético 68
Algoritmo 8 - Propagação de rótulos 102
Algoritmo 9 - Propagação de rótulos e DT *chessboard.* 109

RESUMO

Imagens codificadas por matrizes de intensidade são tipicamente representadas por grande quantidade de dados. Embora existam inúmeras abordagens para análise de imagens, o conhecimento sobre problemas específicos é raramente considerado. Este trabalho trata sobre problemas de análises de imagens cujas soluções dependem do conhecimento sobre os dados envolvidos na aplicação específica. Para isso, utiliza técnicas de mineração de dados para modelar as respostas humanas obtidas de experimentos psicofísicos. Dois problemas de análise de imagens são apresentados: (1) a análise de formas e (2) a análise pictórica. No primeiro problema (1), formas de neurônios da retina (neurônios ganglionares de gato) são segmentadas e seus contornos submetidos a uma calibração dos parâmetros de curvatura considerando a segmentação manual de um especialista. Outros descritores, tais como esqueletos multi-escalas são explorados para eventual uso e avaliação da abordagem. No segundo problema (2), a análise pictórica de imagens de *home-pages* serve para avaliar critérios estéticos a partir de medidas de complexidade, contraste e textura. O sistema generaliza as respostas por um experimento psicofísico realizados com humanos. Os resultados obtivos com as duas abordagens revelaram-se promissores, surpreendentes e com ampla aplicabilidade.

ABSTRACT

Images coded by intensity matrices tipically involve large amount of data. Although image analysis approaches are diverse, knowledge about specific problems is rarely considered. This work is about image analysis problems whose solutions depend on the knowledge about the involved data. In order to do so data mining techniques are applied to model human response to psychophysical experiments. Two image analysis problems are addressed: (1) shape analysis; and (2) pictorial analysis. In the former, neuronal images (ganglion retinal cells of cat) are segmented and curvature parameters are calibrated to identify extremities and branches on the shape considering human segmentation as a reference. Descriptors such as multiscale skeletons are also explored for potential application or evaluations. In the second problem, a pictorial analysis of home-pages images feed an artificial aesthetics criteria evaluator based on complexity, contrast and texture features. The system models and generalizes the obtained human responses to psychophysical experiment. The results for these two approaches are promising, surprising and widely applicable.

1 INTRODUÇÃO

Várias teorias vêm tentando verificar se existe a construção de uma representação do meio-ambiente no sistema nervoso central [GORDON, 1989; BRUCE & GREEN, 1990]. Experimentos neurofisiológicos indicam que essas representações existem e que há um mapeamento da superfície sensorial visual (a retina) em regiões do tecido cortical [HUBEL & WIESEL, 1962]. Além disso, descobriu-se que esse mapeamento é topográfico, ou seja, preserva a vizinhança da entrada sensorial, e que em diferentes regiões do córtex cerebral residem diferentes tipos de representação. Estas evidências incentivaram as teorias de percepção que tratam o processamento visual como um sistema computacional hierárquico. Zeki [ZEKI & SHIPP, 1988], por exemplo, propôs uma arquitetura hierárquica para o processamento visual no cérebro. Marr [MARR, 1982], com um enfoque mais computacional, propôs uma abordagem representacional da forma para a percepção visual. Contudo, ainda falta muito para se conhecer os processos que fazem uma determinada representação da forma a entrada para um sistema cognitivo.

Os objetivos nas abordagens computacionais sobre o cérebro variam desde a concepção de novos sistemas de informação até a tentativa de imitar o poder computacional do cérebro. Este poder é invejado por conseguir tratar grande quantidade de informação com alto grau de paralelismo e com pequeno tempo de resposta (muitas vezes apenas o suficiente). Além disso, a capacidade de utilizar o conhecimento adquirido e descobrir novas maneiras de resolver problemas o faz uma entidade computacional ainda longe de ser reproduzida.

A descoberta de conhecimento (KD – *Knowledge Discovery*) é um paradigma computacional no qual uma máquina pode aprender a tomar decisões a partir do conhecimento adquirido. Em sistemas robóticos, esse paradigma leva o

nome de aprendizado de máquina (*Machine Learning*). O sucesso obtido com sistemas mais dedicados e específicos levou sua aplicação para sistemas administrativos e financeiros que envolvessem grande quantidade de dados armazenados em bancos de dados. Daí seu nome para essa nova aplicação: descoberta de conhecimento para bancos de dados (*Knowledge Discovery in Databases*) ou simplesmente KDD [FAYYAD et al., 1996].

A análise de imagens é o primeiro passo na direção da representação da imagem, e portanto, no desenvolvimento de um sistema de visão computacional. Esta análise pode ou não considerar o conhecimento sobre a imagem. Para termos uma idéia sobre a importância do conhecimento, consideremos a imagem de uma região de cultivo obtida por satélite (Figura 1). Suponhamos que a um indivíduo humano seja atribuída a tarefa de segmentá-la, ou seja, separar a imagem em regiões de interesse ou então distinguir um ou mais objetos de um fundo. Este indivíduo pode tentar desenhar regiões a partir de algum critério (por exemplo, intensidade de níveis de cinza) ou ainda utilizar métodos de segmentação já existentes [CAMPOS et al., 2000]. No entanto, se a ele não for dada informação alguma ou ainda se não possuir um conhecimento anterior sobre a imagem, será muito difícil ou impossível identificar as estruturas e relacioná-las na imagem.

Figura 1 – Imagem obtida por satélite de uma região de cultivo.

Portanto, o conhecimento determina os critérios sobre os quais uma imagem deve ser analisada antes de ser representada. Mais ainda, será possível avaliar se uma segmentação determinou corretamente os limites de um objeto a partir de uma referência conhecida, ou seja, o "*ground-truth*".

A análise da imagem inicia-se com sua segmentação, separando-a em regiões de interesse cujos limites definem (ou não) a forma de um objeto (Figura 2). Parâmetros como limiares de intensidade ou textura são freqüentemente usados como critérios de segmentação [SCHALKOFF, 1989]. Seus desempenhos são conhecidos e são adequados para condições controladas.

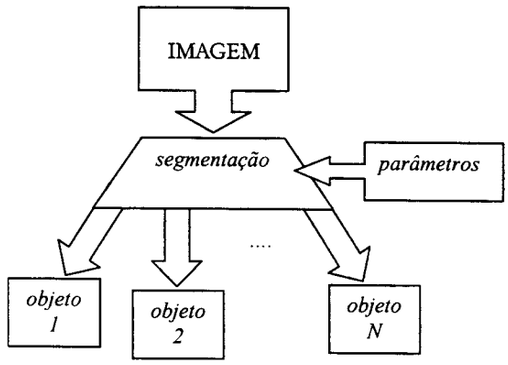


Figura 2 – A análise de uma imagem a partir de um processo de segmentação com parâmetros e critérios conhecidos.

Um outro lado da questão emerge quando o modelo não é determinado por uma característica inerente à imagem. Neste caso, o objetivo é determinar um modelo que, partindo de medidas globais ou de critérios subjetivos, influencie a escolha de um parâmetro de segmentação ou uma avaliação subjetiva. Quando este modelo é construído baseado em critério determinístico, ele é dado por uma fórmula e a escolha do parâmetro é resultante da aplicação dessa fórmula. Um exemplo é o critério entrópico para segmentação de uma imagem [BRINK & PENDOCK, 1996; WONG & SAHOO, 1989]. Nesse caso, o modelo da imagem é o histograma de intensidade e o critério é a diferença entre entropias. O parâmetro é o limiar da intensidade escolhido pelo modelo. Se a escolha for boa, os objetos serão facilmente identificados. A qualidade da segmentação, portanto, é um reflexo da qualidade do modelo.

Todavia, o critério pode ser a opinião expressada por um grupo de pessoas, ou de apenas um indivíduo, ou ainda o comportamento de animais. Neste caso não há uma fórmula que determine a escolha de um parâmetro. Por exemplo, supõe-se que a avaliação estética de imagens por pessoas seja determinada por inúmeros fatores. A partir de uma hipótese, são escolhidas medidas da imagem que possam influenciar esta avaliação estética. Neste caso, o critério corresponde à opinião estética das

pessoas e o modelo para esse critério é construído baseado em medidas da imagem [CONSULARO et al., 2000]. Para a segmentação de imagens, o critério pode também ser baseado em opiniões subjetivas [CAMPOS et al., 2000]. Assim, baseado em medidas globais o modelo construído escolhe um limiar de intensidade ou escalas adequadas.

Uma vez segmentada a imagem e as regiões delimitadas, o problema ainda não estará resolvido, pois é preciso descrever a forma de cada região para representá-la. Descrever a forma não é tarefa simples, principalmente quando se trata de formas extraídas de imagens naturais. Há uma grande quantidade de abordagens para estabelecer o repertório de símbolos com o qual uma forma deve ser descrita. Estes repertórios dependem fortemente dos resultados da aplicação de métodos de segmentação de formas ou da extração de descritores. A segmentação de formas consiste em determinar pontos de importância em um contorno. Tais pontos são os que indicarão os limites de um segmento, que por sua vez definirão um símbolo do repertório usado para descrever a forma [SONKA et al., 1993].

Este trabalho aborda dois problemas de análise de imagens. O primeiro procura identificar modelos que generalizem os critérios utilizados por especialistas para segmentar formas. O segundo, busca modelos para avaliar critérios estéticos subjetivos de uma população. Estes modelos determinam também os atributos mais significativos de um conjunto mais amplo de medidas.

A construção de modelos para descoberta de conhecimento é chamada de mineração de dados ou "*data mining*". O foco deste trabalho está no uso de técnicas de mineração de dados para análise de imagens. Três fases podem ser identificadas para extrair um modelo para análise de imagens.

Na primeira etapa, um conjunto de objetos é dividido em dois conjuntos: um conjunto de treinamento e outro de testes. Os objetos de treinamento, que no caso são imagens ou formas, são medidos gerando assim um conjunto de atributos $\{f_1, f_2, \ldots, f_n\}$. O que se quer modelar nesse trabalho são seres humanos especialistas ou então uma audiência. Deste modo, um conjunto destes sujeitos é submetido a experimentos psicofísicos e instados a responder sobre os objetos de treinamento. Estas respostas resultarão um conjunto de respostas de treinamento $R_{treinamento}$ que, junto com as medidas e uma generalização comporão um modelo. As respostas

desses objetos oferecerão um conjunto de respostas do modelo R_{modelo} (veja Figura 3).

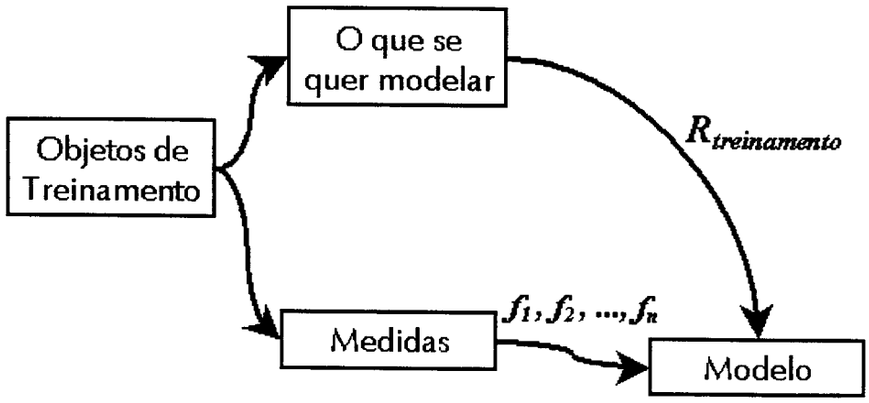


Figura 3 – Etapa na qual se obtem os modelos de treinamento.

Na segunda fase, é momento de utilizar o conjunto de objetos de teste. Este conjunto deverá testar o modelo. Deste modo, as mesmas medidas $\{f_1, f_2, \ldots, f_n\}$ são extraídas, agora para o conjunto de testes. O conjunto de respostas de teste, obtido do sujeito pelo experimento psicofísico, serve como parâmetro de comparação com a respostas do modelo. Portanto, um erro mede essa discrepância de respostas (veja Figura 4).

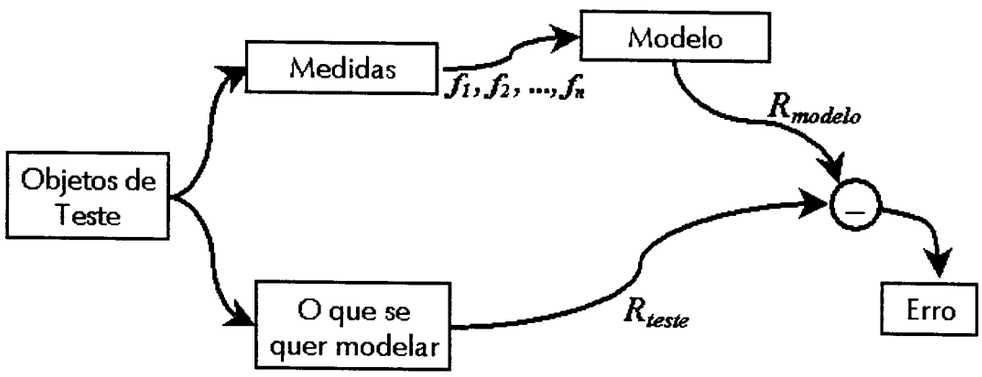


Figura 4 – Etapa na qual se obtém medidas e os modelos do objeto de teste. As respostas de teste são comparadas às respostas do modelo acarretando um erro.

A terceira etapa consiste na busca do modelo cujo erro seja o menor, se comparado aos dos objetos de treinamento. Esta busca pode encontrar o conjunto de medidas cuja resposta apresente o menor erro. Uma vez atingido um erro satisfatório

ou adequado, o modelo pode ser validado e então utilizado em uma aplicação adequada (veja Figura 5).

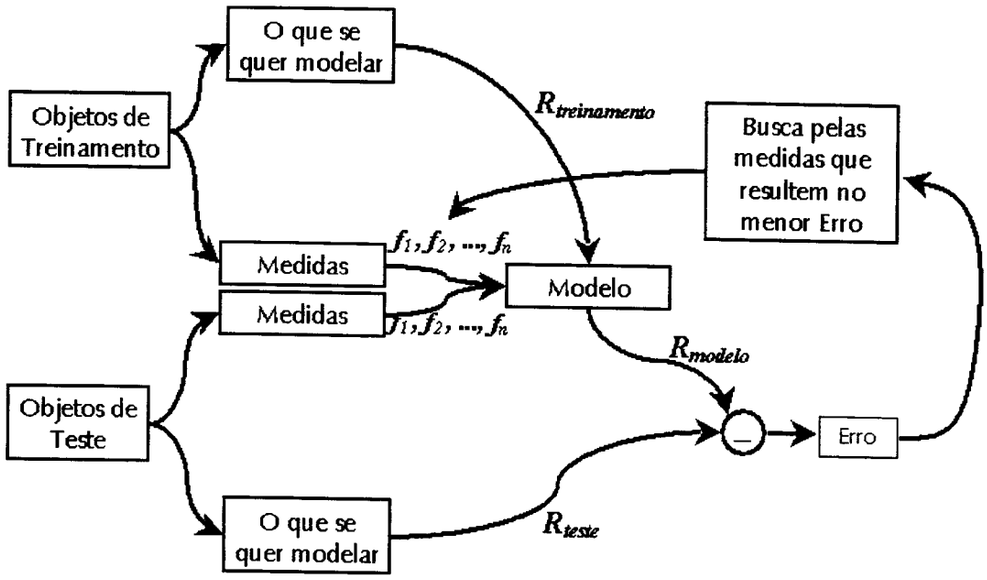


Figura 5 – Etapa na qual o melhor conjunto de medidas é selecionado, baseado no menor erro.

Este trabalho apresenta duas aplicações: uma para análise de formas e outra para análise pictórica. Na primeira aplicação, um conjunto de formas de neurônios é separado em dois conjuntos: treinamento e teste. As medidas extraídas são medidas estatísticas da curvatura de cada forma. O objetivo da aplicação é utilizar apenas estas duas medidas globais para identificar cinco parâmetros que, aplicados à curvatura, possam identificar extremidades e ramificações da forma. Na segunda aplicação, imagens de home-pages são também separadas em dois conjuntos: um de treinamento e outro de teste. O objetivo é buscar um modelo que permita a avaliação estética de home-pages relacionando medidas pictóricas destas imagens com avaliações subjetivas de uma audiência.

Estas aplicações utilizam o esquema genérico da Figura acima para buscar os modelos de cada uma delas. Além destas aplicações, métodos para descrever formas e novos atributos foram estudados e são apresentados também. Esta tese apresenta-se conforme cada um dos capítulos descritos a seguir.

Um capítulo introdutório (capítulo 2) sobre percepção visual e descoberta de conhecimento relaciona algumas das principais teorias de percepção visual e também

identifica as abordagens computacionais nas quais é viável a construção de modelos permitindo a aplicação de técnicas de mineração de dados. No capítulo 3, estas técnicas são abordadas e as etapas e desafios que devem ser vencidos para a construção de um sistema de conhecimento são apresentados. Além disso, este capítulo descreve um conjunto de técnicas que foram utilizadas para os experimentos desse trabalho como algoritmos genéticos e interpolação por "*thin-plate spline*". Outras técnicas que foram úteis no decorrer dos experimentos e que poderão ser úteis em desenvolvimentos futuros, tais como classificadores e agrupadores, são também abordados.

Um capítulo sobre prospecção de atributos e descritores de forma foi incluído (capítulo 4), descrevendo medidas que foram desenvolvidas na tentativa de obter desempenho ou como alternativas a descritores existentes. Assim, uma abordagem baseada na teoria de campos eletrostáticos e outra de esqueletonização multi-escalas sem deslocamento de borda são apresentadas. A seguir, o capítulo 5 trata dos experimentos elaborados para análise de formas. Duas abordagens ilustram diferentes enfoques: (1) um modelo para a calibração de medidas globais da curvatura, que são utilizadas para compor um modelo de detecção de extremidades e ramificações em formas neuronais planares; (2) uma abordagem que aplica a esqueletonização multi-escalas para descida da estrutura esqueletal de forma a detectar saliências.

O capítulo 6 trata sobre análise pictórica ilustrando o tema com uma aplicação de avaliação estética de home-pages. Medidas que influenciam a avaliação estética de indivíduos humanos foram modeladas por interpolações usando "*thin-plate spline*". Os melhores modelos ou combinações de *features*, foram buscados por um algoritmo genético e também pela "força bruta". Os modelos de conhecimento foram construídos a partir dos dados obtidos por um experimento psicofísico no qual indivíduos humanos explicitaram sua avaliação estética sobre as imagens das *home-pages*.

Por fim, um capítulo de discussão e conclusões (capítulo 7) apresenta os desafios que ainda precisam ser superados para que abordagens de conhecimento em análise de imagens possam ser utilizadas em aplicações comerciais ou industriais. As

perspectivas promissoras para este trabalho são também apresentadas, delineando um conjunto de desenvolvimentos futuros a partir dos resultados promissores obtidos.

2 PERCEPÇÃO VISUAL E DESCOBERTA DE CONHECIMENTO

A percepção visual é o processo como são organizadas informações sensoriais com o objetivo de fornecer significado a um ambiente. Uma grande quantidade de teorias tenta decifrar a percepção em seres humanos ou em animais. Os próprios autores de trabalhos que tentam listar estas teorias admitem que explicam apenas um pequeno conjunto delas [GORDON, 1989; BRUCE & GREEN, 1990].

O enfoque de tais teorias também é diverso. Variam entre abordagens fisiológicas, psicológicas, ecológicas ou computacionais, entre outras.

Abordagens fisiológicas buscam modelos neuronais ou orgânicos que possam descrever arquiteturalmente o funcionamento do cérebro humano. Assim, para o caso da visão, neurônios sensoriais como os cones e bastonetes, as células ganglionares na retina, o núcleo geninulado lateral (LGN), as áreas visuais do córtex cerebral, além de todas as áreas cerebrais de alguma forma relacionadas à visão têm sido intensivamente explorados por pesquisas envolvendo fisiologia e modelos. Hardclastle [HARDCASTLE, 1999] classifica estes modelos em três grupos: detecção de atributos, sistemas dinâmicos e microprocessamento dendrítico. Todas estas abordagens têm o objetivo de identificar padrões para descobrir conhecimento. A detecção de atributos considera os sinais de um neurônio como a unidade básica de cognição. Modelos por sistemas dinâmicos tentam identificar padrões em uma escala maior como redes, regiões ou tecidos neurais. O microprocessamento dendrítico foca as redes dendríticas como uma unidade cognitiva. Trabalhos nesse sentido identificam inclusive gramáticas a partir de estímulos aplicados em redes baseadas em neurônios reais[COSTA et al., 1998].

As abordagens psicológicas buscam entender a percepção por aspectos externos ao indivíduo e hipotetizando sobre seu comportamento. O que se busca

portanto, é confirmar tais hipóteses e correlacioná-las com resultados de experimentos psicofísicos. O experimento psicofísico requer apenas um conjunto de estímulos sob condições controladas e um conjunto de sujeitos ou indivíduos para serem realizados. Exemplos de experimentos psicofísicos incluem aqueles cujos estímulos testam limiares da percepção buscando determinar escalas e limites sensoriais. Devido a sua simplicidade, o experimento psicofísico é talvez a maior fonte de dados para estudo da percepção [GORDON, 1989]. Novamente a busca por padrões e conhecimento está presente na abordagem. Um padrão de comportamento definido pelos resultados de um experimento psicofísico deve estar correlacionado com uma hipótese cognitiva, como por exemplo, as leis de agrupamento da teoria Gestalt.

Um outro tipo de abordagem é a ecológica. Desenvolvida por Gibson [GIBSON, 1950], esta abordagem destaca-se por considerar a informação perceptiva presente no estímulo e não no aparato biológico. Deste modo, a percepção se caracterizaria por ser um ato de captura e não resultante de um processo passivo de recepção do aparelho sensorial [BRUCE & GREEN, 1990]. Por exemplo, a percepção de distância é influeciada, segundo Gibson, pelo gradiente da textura do fundo de uma cena. Desse exemplo, percebe-se que há a necessidade de um padrão, ou seja, uma característica do fundo da cena, para que haja a percepção de uma medida. Há, portanto, um processo de descoberta de conhecimento (distância ou profundidade de um objeto em uma cena) sobre um padrão (o fundo da cena).

Embora a classificação das teorias tente compartimentá-las, são cada vez mais freqüentes estudos relacionando-as. As hipóteses que a teoria Gestalt tentaram explicar muito subjetivamente têm sido relacionadas com modelos computacionais conexionistas [PALMER, 1992], com modelos representacionais [AHUJA & TUCERYAN, 1989] ou ainda com evidências neurofisiológicas [WESTHEIMER, 1999].

As abordagens computacionais buscam descrever algoritmicamente o processo de percepção. Esta descrição algoritmica acarreta um modelo computacional do cérebro, o qual é muitas vezes criticado por apresentar grandes discrepâncias com a realidade, mas reconhecidamente útil em simulações. Os modelos representacionais baseiam-se em descrições das cenas de um ambiente.

Marr [MARR, 1982] propõe um modelo que decompõe uma cena em quatro representações hierarquicamente relacionadas: (1) a intensidade em níveis de cinza; (2) o *primal sketch* composto por primitivas sobre as mudanças de intensidade, sua distribuição e organização; (3) o 2½-D *sketch* que descreve o ponto de vista do observador; (4) a representação 3D explicitando a organização espacial baseada no objeto descrito.

A representação pressupõe não só a extração de primitivas, mas também a construção de sentenças para que ao final resulte uma descrição global da cena. A cada etapa, portanto, devem ser armazenadas representações intermediárias para que a etapa seguinte possa prosseguir. A abordagem conexionista é baseada em modelos de neurônios ou unidades cujo papel é servir de ativador ou inibidor de entradas que influirão em uma saída. Esta saída, por sua vez, serve como entrada para outras unidades. O apelo a esse tipo de modelo é que ele dispensa o armazenamento dos resultados intermediários, permitindo que as sentenças sejam definidas apenas pela arquitetura. O paralelismo inerente e a possibilidade de aprendizado fazem esta abordagem atraente, mas sua complexidade e a forte dependência de parâmetros torna-a muitas vezes inviável em aplicações reais. Por isso, tais modelos têm servido como etapas em abordagens representacionais como mapeamentos de entradas sensoriais[BRUCE & GREEN, 1990] ou em sistemas de reconhecimento de padrões [DUDA et al., 1998].

A procura por modelos de percepção está intimamente relacionada à busca de conhecimento e ao reconhecimento de padrões. Uma ferramenta que consiga extrair a representação visual e deduzir regras cognitivas utilizando uma abordagem representacional pode ser construída utilizando a arquitetura da Figura 6. Na ilustração, são identificadas três etapas das quais as informações se transformam: (1) o sensoriamento, no qual os estímulos provenientes de um ambiente são introduzidos no sistema; (2) a percepção que transforma os estímulos em atributos; (3) a cognição que extrai regras ou induz comportamentos para compor um conhecimento.

Um sistema de descoberta de conhecimento busca reproduzir as etapas do processo cognitivo. Os objetivos podem ir desde uma simples seleção de atributos até o estabelecimento das sentenças que servem para inferir regras sobre os dados que fazem, nesse caso, o papel dos estímulos.

Para modelos de percepção visual, os estímulos são restritos às imagens. Assim, um sistema de descoberta de conhecimento que utilize modelos de percepção visual deve considerar imagens como seus dados de entrada. A correspondência biológica entre o processo cognitivo e a descoberta de conhecimento é explorada logo no tratamento do estímulo imagem.

Figura 6 – Arquitetura para uma abordagem representacional de percepção.

A abordagem computacional da qual esse trabalho trata considera um sistema KDD com uma arquitetura similar ao da Figura 6. Deste modo, uma base de imagens conterá os estímulos. Estes estímulos podem ser diretamente a imagem em níveis de cinza ou colorida para análise pictórica, ou o contorno de uma forma extraída por um processo de segmentação e que é submetida à análise de formas. O conjunto de estímulos passa por processos de mensuração (*featuring*). Estes atributos representam a percepção propriamente dita. A partir de generalizações ou classificações, cada imagem poderá ser representada por sentenças ou por regras desaguando na cognição daquela imagem.

Na próxima seção, a mineração de dados será abordada como uma etapa do processo de conhecimento. Técnicas úteis serão apresentadas e muitas delas implementadas para compor um ambiente de experimentos.

3 MINERAÇÃO DE DADOS

A sociedade da informação caracteriza-se pelo que produz, ou seja, informação em profusão. Para fins de armazenamento e recuperação de dados, a abordagem tradicional e que ainda prevalece devido a limitações tecnológicas é o banco de dados. Os bancos de dados permitem armazenar e recuperar a informação mantida em meio magnético de uma maneira efetiva. Embora exista em BDs ferramentas que permitam reduzir os dados, tais como operações embutidas da álgebra relacional em linguagens de definição (DDL – *Data-Definition Language*) e de manipulação de dados (DML – *Data-Manipulation Language*); é imprescindível que se estenda ou se adicione operações que permitam análises estatísticas e de reconhecimento de padrões para um melhor entendimento dos dados. Deste modo, evoliu-se para ferramentas que viabilizam análises estatísticas "*on-line*" (OLAP – *On-Line Analytical Processing*), ou seja, ao mesmo tempo em que o banco de dados é atualizado suas análises também o são.

Entender os dados pressupõe o objetivo de se fazer hipóteses sobre o que os dados representam e como podem ser segmentados baseados em seus aspectos. Desta maneira, técnicas de segmentação, classificação, agrupamento, generalização permitem inferir regras ou deduzir comportamentos dos objetos representados pelos dados. Estas técnicas envolvem desde abordagens mais simples como as árvores de decisão até aquelas mais sofisticadas para reconhecimento de padrões e de aprendizado de máquina (*machine learning*). O conjunto destas técnicas aplicado a grande quantidade de dados, normalmente residentes em bancos de dados especializados chamados *data warehouse* é convencionalmente chamado de mineração de dados ou *data mining*.

Esta vasta quantidade de dados sobre a qual age um minerador de dados resulta normalmente de registros históricos sobre alguma atividade; ou então de ambientes que gerem inerentemente uma grande quantidade de dados; ou ainda que apresentem relacionamentos complexos que não sejam explicitamente modelados.

Dentre as aplicações típicas que geram tais informações estão operações bancárias, seqüenciamento do genoma humano, busca de informação na "*internet*", mercado de ações, pesquisa de mercado, análise de audiência, publicidade, identificação de galáxias, meteorologia, detecção de fraudes, entre muitas outras.

Para muitas destas aplicações, a mineração de dados, embora seja a mais importante, é apenas uma das etapas de uma abordagem maior conhecida como descoberta de conhecimentos e abreviada pela sigla KD (*Knowledge Discovery*). Quando aplicada sobre bancos de dados usa-se o termo KDD (*Knowledge Discovery in Databases*).

Fayyad et al. [FAYYAD et al., 1996] dividiram um KDD em cinco etapas. Esta divisão é adequada para entender as tarefas e ferramentas necessárias em um processo de KDD e é utilizada intensivamente na literatura sobre o assunto, principalmente quando relacionada a negócios.

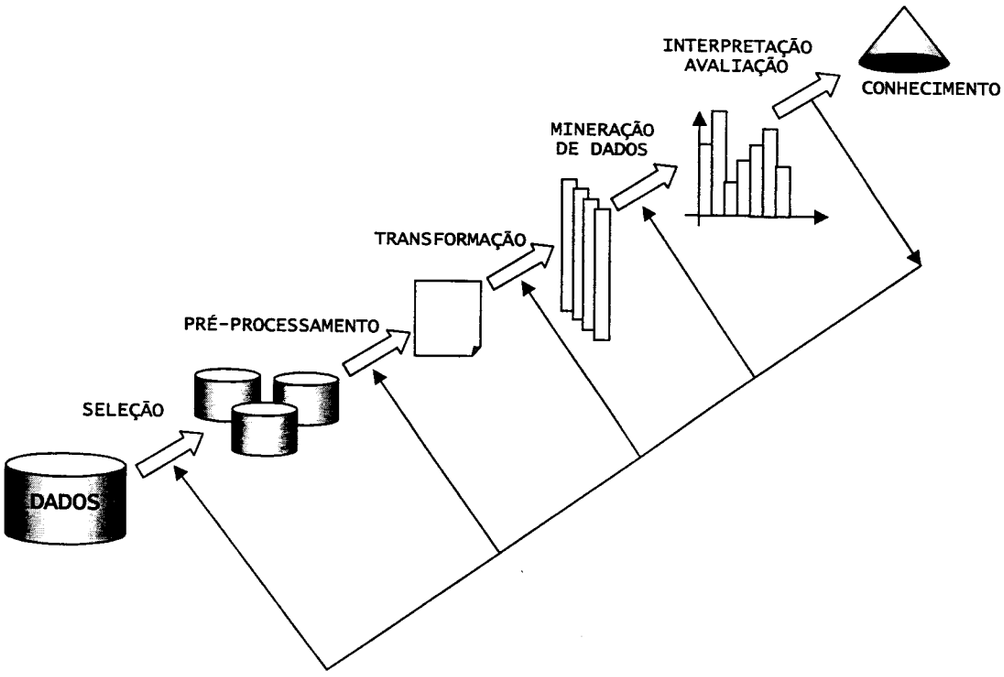


Figura 7 – O processo de mineração de dados segundo [FAYYAD et al., 1996].

A mineração de dados é a parte mais importante em um sistema de KDD porque suas técnicas é que permitirão a identificação de padrões. No entanto, para tornar um sistema KDD viável é necessária uma infra-estrutura que permita:

(1) armazenar em grande volume e com uma normalização adequada;

(2) consultar e visualizar análises estatísticas sofisticadas dos dados;

(3) gerar um banco de regras cujo papel é viabilizar inferências rápidas sobre os dados e padrões analisados;

(4) executar métodos de classificação que permitam testar hipóteses sobre algum objetivo.

Para tornar essa infra-estrutura possível, está disponível um conjunto de tecnologias. Algumas destas tecnologias têm recebido grande atenção de fabricantes e vêm permitindo constante melhoria de desempenho. Os *data warehouses* caracterizam o item (1) e as ferramentas OLAP caracterizam o item (2). Estas tecnologias permitem não só um melhor desempenho em consultas, mas também definir políticas de utilização, isto é, quais dados serão utilizados ou integrados e como serão indexados.

O uso integrado destas técnicas e tecnologias não só permite identificar padrões, mas também analisá-los com o objetivo de prever comportamentos. Os dados disponibilizados por um OLAP, por exemplo, são oferecidos de maneira descorrelacionada a um treinamento. Este treinamento gera um conjunto de medidas adequadas para que classificadores ou árvores de decisão possam identificar as relações ou regras existentes entre os dados e o objetivo. É claro que nem sempre é assim que se resolve um determinado problema. Por exemplo, se o objetivo não foi estabelecido, o sistema KDD pode "garimpar" um conjunto de objetivos. Se o problema for a existência de muitas medidas (atributos), o sistema KDD identifica quais são as mais relevantes, ou ainda indica seu grau de relevância. Percebe-se que existem trajetórias em um espaço de soluções para os problemas de KD e que em alguns casos, tais trajetórias podem estar sobrepostas ou algumas vezes não existir.

A diversidade de técnicas e ferramentas disponíveis para construção de sistemas KDD aliada à diversidade de problemas e suas peculiaridades, criou a necessidade de se estabelecer paradigmas. Esta seção preocupa-se portanto em

revisar os paradigmas que serão utilizados neste trabalho ou que tenham potencial imediato de aplicação no problema de modelos de avaliação ou segmentação de imagens ou formas. Antes de tudo, uma rápida introdução sobre as tecnologias de suporte a sistemas KDD é apresentada. Em seguida, uma estrutura taxonômica é proposta para classificar os paradigmas de KD, cujo principal objetivo é facilitar a busca por trajetórias de soluções adequadas.

3.1 Infra-estrutura

Atingir algum objetivo lidando com muitas tecnologias exige um ingrediente fundamental: sinergia. Esta sinergia deve ocorrer nos diversos níveis da aplicação, desde o modelo de dados até uma eventual decisão inferida ao sistema.

No modelo de dados esta sinergia pode ser conseguida por abordagens tradicionais ou por soluções dedicadas. As abordagens tradicionais têm a vantagem de possibilitar o uso de produtos "*off-the-shelf*" poupando tempo de desenvolvimento. No entanto, submete-se às restrições impostas pelas características de cada produto. Em uma aplicação dedicada, o grau de liberdade para teste de novas soluções é maior, porém o tempo de desenvolvimento e validação de resultados é o principal problema. Embora a solução adotada ao longo desse trabalho tenha sido predominantemente a dedicada, introduziremos alguns conceitos sobre as abordagens convencionais para oferecer suporte a sistemas KD, já que soluções tradicionais podem também serem válidas para tratar problemas de visão computacional e processamento de imagens.

3.2 Abordagem convencional

A abordagem convencional de um sistema KD consiste de cinco partes principais: acúmulo de dados, seleção de dados, pré-processamento, mineração de dados e análise ou assimilação da informação.

Para cada uma destas partes foram desenvolvidas tecnologias com características peculiares com relação ao armazenamento e acesso aos dados, e identificada com um ou mais níveis de abstrações hierárquicas. As tecnologias adquiriram acrônimos que as identificam em algumas etapas. Para o *acúmulo de dados* a tecnologia é normalmente chamada de *data warehousing*, associando-a com a Figura de um grande armazém de dados. A sigla OLAP (*On-Line Analytical*

Processing) identifica o processamento analítico que caracteriza a transformação dos dados e cujas ferramentas incluem também métodos de preparação e seleção de dados, ou seja a etapa de *pré-processamento*. A própria mineração de dados é amplamente chamada na literatura de *data mining* em uma alusão à garimpagem de informações valiosas em meio a uma grande mina de dados.

3.2.1 Data Warehousing

Uma das maiores dificuldades para pessoas que não são da área de Sistemas de Informação (SI) é identificar a necessidade de dar um novo nome a um amontoado de dados. Na verdade as tecnologias de *data warehousing* não são imprescindíveis em um sistema KD. No entanto, o tempo de resposta a uma consulta é um fator que pode inviabilizar determinadas aplicações. O que mais fortemente caracteriza um *data warehouse* é o tipo de dado que está armazenado nele.

Para entender melhor o papel de um *data warehouse*, é importante ter em mente alguns tipos de dados de um SI quanto à sua origem e utilização, ou seja, dados transacionais, operacionais ou informacionais.

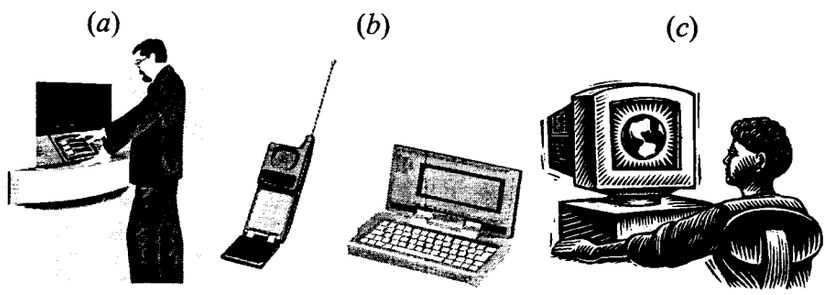


Figura 8 – Exemplo de dados transacionais: (*a*) caixa eletrônico; (*b*) operações bancárias com celulares ou computadores portáteis; (*c*) operações de comércio eletrônico via internet.

Dados transacionais são aqueles obtidos em transações, ou seja, em operações que uma vez finalizadas, seus resultados devem existir ou não (atomicidade), não havendo possibilidade de existir apenas parte do resultado. Exemplos desse tipo de dado são aqueles adquiridos em caixas eletrônicos de bancos ou quiosques sobre orientação; em aplicações realizadas em nodos remotos ou móveis de redes (locais ou não), *internet*, entre outros (Figura 8). Para esse trabalho, este tipo de dado corresponde aos dados adquiridos em um experimento psicofísico implementado em um ambiente *web*. Aplicações transacionais são implementadas por tecnologias conhecidas como OLTP (*On-Line Transaction Processing*),

dedicadas a tratar dados cujo tempo de acesso para uma alteração ou consulta a uma ou mais tabelas seja crítico.

Os dados operacionais são aqueles resultantes da operação sobre um banco de dados, acarretando a geração de informação intermediária ou relatórios operacionais. Alguns exemplos de dados operacionais são dados gerados pelo processamento de uma folha de pagamento de uma empresa; sistemas de fluxo de caixa diário, ou seja, aqueles dados envolvidos no quotidiano operacional de uma atividade (Figura 9).

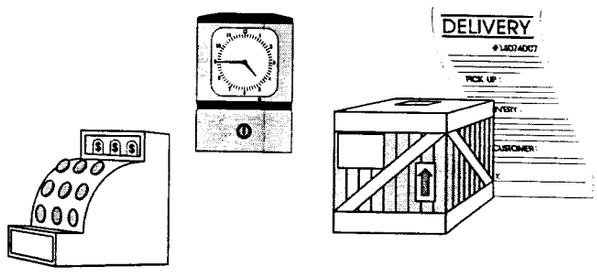


Figura 9 – Exemplo de dados operacionais: fluxo de caixa, folha de pagamento, controle de estoque, faturamento.

Os dados informacionais, por sua vez, resultam da geração de informações consolidadas e agrupadas dos dados transacionais e operacionais. Assim, é identificada a significância de um dado que será consolidado copiando-o ou transformando um conjunto de dados pela aplicação de operações matemáticas ou estatísticas em um dado informacional. Uma ilustração para esse tipo de dado é a informação freqüentemente contida em relatórios gerenciais ou a monitoração de índices em uma bolsa de valores (Figura 10).

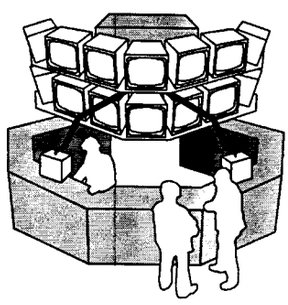


Figura 10 – Exemplo de dados informacionais em bolsa de valores.

Os *data warehouses* acumulam dados informacionais [CABENA et al., 1997].

Tipicamente, um banco de dados é modelado logicamente, privilegiando os relacionamentos conceituais. Isso permite que as consultas ou alterações sejam executadas levando-se em consideração o fluxo de informação do SI (Sistema de Informação). Um conjunto de normalizações permite ainda que se controle a redundância do modelo lógico para que o número de operações de junções e a quantidade de índices sejam minimizados reduzindo também o tempo de acesso. A Figura 11 ilustra os extremos desse processo, ou seja, a Figura 11*a* mostra um esquema de um banco de dados operacionais ou transacionais apresentando vários índices. Cada índice corresponde a uma tabela dedicada a indicar a localização de uma informação de uma tabela em outra tabela. Se o modelo lógico estiver atrelado ao modelo conceitual, os índices podem melhorar sensivelmente o desempenho, seja quanto ao tempo de acesso, seja quanto à redundância de informação. Já a Figura 11*b* mostra as tabelas de dados operacionais e transacionais consolidadas sobre uma tabela informacional ou até mesmo de um *data warehouse*. Nesta caso, abre-se mão do modelo conceitual, permitindo certa redundância para que haja uma maior flexibilidade em consultas mais genéricas.

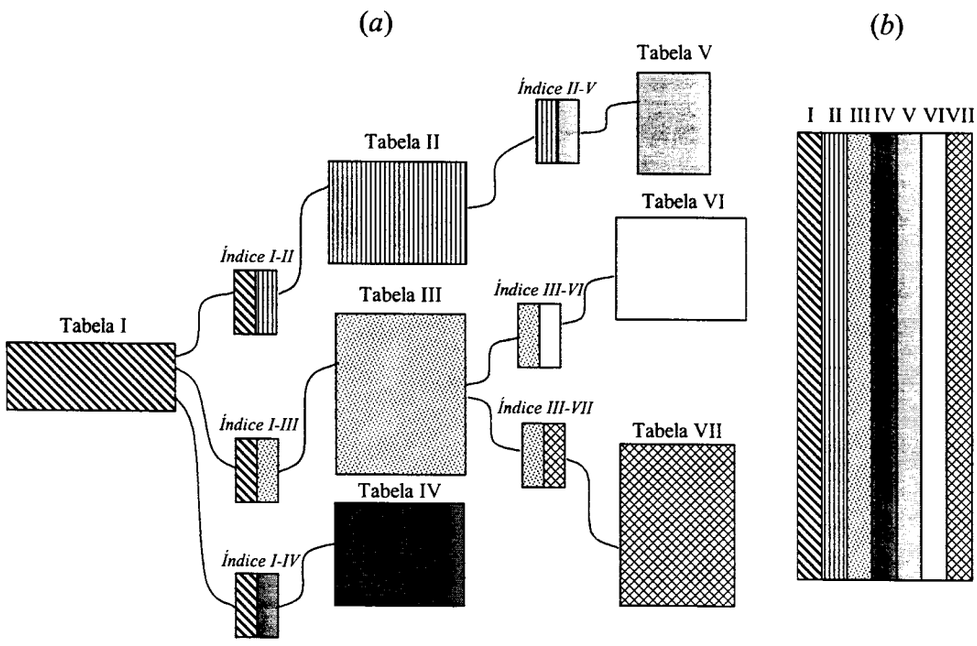


Figura 11 – Ilustrações de relacionamentos típicos em dados operacionais e transacionais (*a*) e para dados informacionais (*b*)

Além dos obstáculos sobre o tempo de acesso, os bancos de dados transacionais e operacionais oferecem pouco sobre a história dos eventos ocorridos

IESC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

em um SI. Essa limitação é muitas vezes imposta pelos modelos conceituais, para que não armazenem uma grande quantidade de dados que servirão apenas como mais um fator de aumento do tempo de acesso e no tratamento da integridade.

No entanto, o *data warehouse* foi proposto para que as tabelas residentes nele possam servir às análises históricas. Além disso, as tabelas de um *data warehouse* devem conter uma quantidade muito pequena de dados ruidosos. Para isso uma fase de limpeza dos dados deve ser realizada. Esta fase é de extrema importância, já que irá determinar quais informações influirão ou serão ignoradas seja por redundância, seja por irrelevância.

É comum que dados transacionais e operacionais estejam fisicamente localizados em diferentes lugares. Assim, é necessária uma integração destes dados no *data warehouse*. Algumas abordagens centralizam o *data warehouse* e outras distribuem seu conteúdo em unidades departamentalizadas chamadas *data marts*. Contudo, mesmo distribuídos eles devem permanecer conceitualmente integrados. Modelos conceituais para *data warehouses* têm sido concebidos [GOLFARELLI et al., 1998] e regras de normalização que considerem as redundâncias inerentes a ela são considerados em vários trabalhos.

Assim, os dados em um *data warehouse* ficam dispostos em tabelas com grande quantidade de atributos (colunas). Isto é conseqüência de quatro possíveis processamentos: (1) criação de novos atributos resultantes de consolidações (estatísticas por exemplo); (2) redução dos dados assumindo um critério de significância; (3) desnormalização para minimizar a sobrecarga de índices definidos por um modelo lógico; e (4) limpeza dos dados para eliminar problemas de integridade devido a fontes heterogêneas de dados.

3.2.2 OLAP (On-Line Analytic Processing)

O OLAP é caracterizado pelo seu principal objetivo: dispor os dados de forma que facilite seu acesso à análise e visualização. A origem dos dados pode ser transacional, operacional ou ainda um *data warehouse*. O resultado de um OLAP pode ter múltiplos destinos: visualização, relatórios, mineração de dados ou consultas especializadas (hipotetizações).

Codd et al. [CODD et al., 1993] propôs o OLAP para dar suporte aos sistemas de auxílio à decisão (DSS – *Decision-Support Systems*) em um contraponto aos modelos conceituais e lógicos que sustentam os dados transacionais (OLTP) e operacionais cujos representantes mais conhecidos são Entidade-Relacionamento [CHEN, 1976], Orientado a Objetos e Relacionais [CODD, 1970].

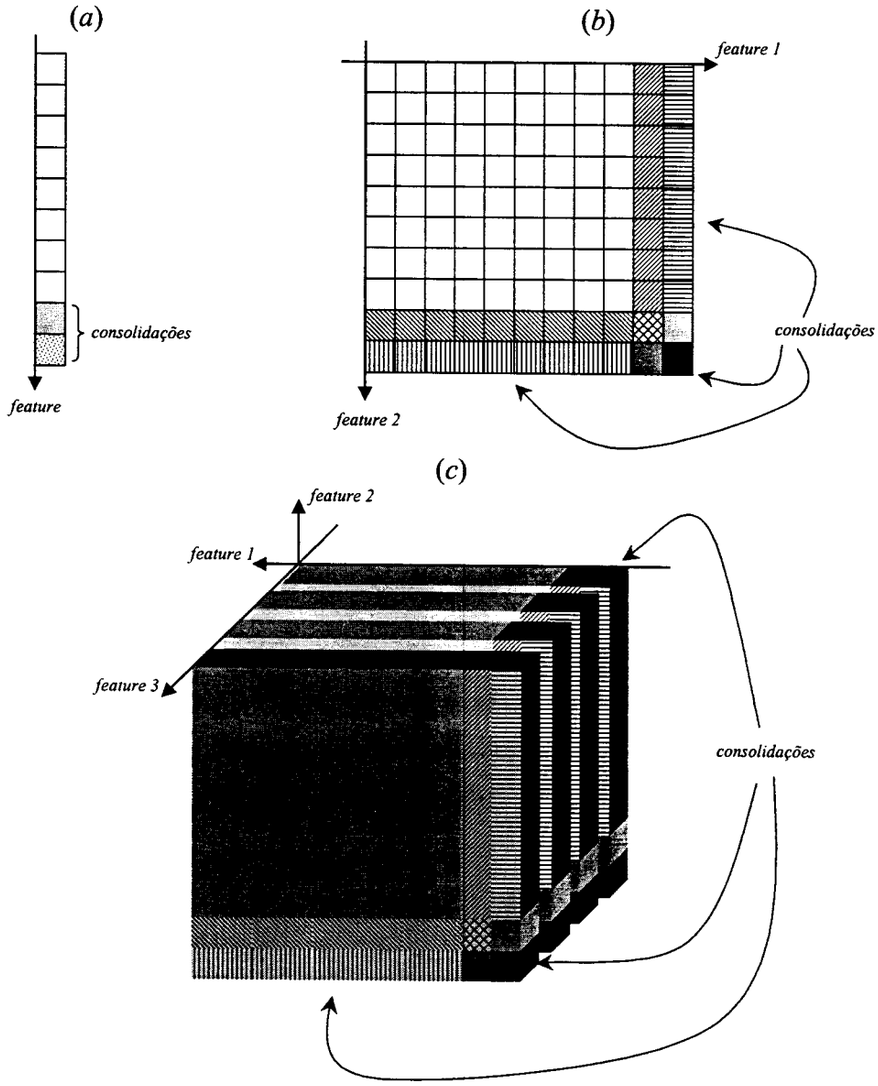


Figura 12 – Visão multidimensional de dados fornecida por um OLAP.

A motivação para a proposição de um novo modelo conceitual foi a constante necessidade de se considerar atributos temporais nas análises. Assim, sentiu-se a necessidade de um modelo que tratasse o tempo como uma dimensão e não como um atributo meramente informativo. Um caminho de consolidação define os atributos e os cálculos usados para sintetizar uma informação. Por exemplo, Na

Figura 12*a*, as consolidações podem ocorrer apenas em uma dimensão. Por exemplo, a média de todos os pontos. Já na Figura 12*b*, as consolidações podem ocorrer em qualquer uma das duas dimensões definidas por cada atributo e também para cada intervalo. Na Figura 12*c* os caminhos são mais numerosos ainda, e conforme evolui o número de atributos, maior este espaço se torna. Justificando a existência de uma Visão Conceitual Multidimensional e de ferramentas apropriadas para implementá-la, Codd et al. tratam como dimensão qualquer caminho de consolidação de dados introduzindo assim o conceito de OLAP. Vários modelos conceituais multidimensionais foram propostos e são avaliados em [PEDERSEN & JENSEN, 1998].

Basicamente, uma ferramenta OLAP deve permitir consolidação e extração de informações dos dados com um grau de liberdade oferecido pelo espaço de atributos. Além disso, sua implementação deve permitir acesso rápido às informações consolidadas ou relacionadas.

Diversas ferramentas implementam OLAP [DHAR & STEIN, 1997], porém alguns autores tratam-no meramente como um paradigma de mineração de dados [FREITAS & LAVINGTON, 1998]. Outros consideram-no uma etapa precedente à mineração, isto é, apenas um mecanismo de acesso e consolidação de dados [TANLER, 1997]. O que na verdade têm surgido são produtos que inicialmente implementavam métodos de consolidação e análise de dados, mas que posteriormente agregaram a capacidade de desempenhar tarefas de qualquer fase de um KDD. Portanto, tornaram-se ambientes ou ferramentas facilitadoras na construção e manutenção de sistemas KDD.

3.2.3 Mineração de Dados – DM (Data Mining)

Mineração de Dados (DM) é o componente determinante de um processo de descoberta de conhecimento KD. A diversidade de algoritmos e abordagens para DM é imensa e na tentativa de situar o leitor nesse intrincado espaço propusemos uma taxonomia das técnicas disponíveis dentro de suas respectivas abordagens em DM, já que uma boa quantidade destas técnicas é utilizada em mais que uma abordagem.

Vários autores [FAYYAD et al., 1996; REINARTZ, 1999; WEISS & INDURKHYA, 1997] definem fases para a descoberta de conhecimento. Em

[TWOCROWS, 1998] o processo de descoberta de conhecimento é dividido nas seguintes fases e subfases:

1 Identificação dos Objetivos;
2 Construção do Banco de Dados para a Mineração
   a.) Coleta de Dados;
   b.) Descrição dos Dados;
   c.) Seleção;
   d.) Avaliação da qualidade dos dados e Limpeza;
   e.) Consolidação e Integração;
   f.) Construção de Metadados;
   g.) Carregar os dados no BD de Mineração;
   h.) Manutenção do BD de Mineração
3 Exploração dos Dados
4 Preparação dos Dados
   a.) Seleção de atributos;
   b.) Amostragem;
   c.) Construção de novos atributos;
   d.) Transformação dos atributos
5 Construção do Modelo
6 Avaliação e Interpretação
   a.) Validação do Modelo;
   b.) Validação Externa
7 Desenvolvimento do Modelo e Resultados
8 Acompanhamento

3.2.3.1 Identificação dos Objetivos

Embora sendo a fase inicial, a identificação dos objetivos é crucial no processo de KD, pois ela define o resultado esperado e conseqüentemente influi fortemente no conjunto de métodos e ferramentas que serão utilizados. Se a meta for entender a natureza dos dados (modelo de entendimento) e obter respostas a questionamentos sobre tal natureza, então ferramentas de segmentação (classificadores e agrupadores) e tomadores de decisão serão muito provavelmente utilizados. Já se o objetivo for a previsão de comportamentos (modelo preditivo), será mais recomendável utilizar reconhecedores de padrões que consigam generalizar ou aprender a partir de um sistema mais complexo ou de ocorrências anteriores.

Os objetivos podem ainda evoluir de modelos de entendimento para modelos preditivos. Nakhaeizadeh et al. [NAKHAEIZADEH et al., 1997][1] mostram as etapas que fazem parte desta evolução. A *descrição e a sumarização dos dados* é

---

[1] apud [REINARTZ, 1999]

tipicamente uma parte do objetivo e auxilia a formulação das hipóteses para encontrar informações ocultas nos dados. A *segmentação*, como o próprio nome diz, separa os dados em classes ou subgrupos significantes e interessantes. Técnicas de agrupamentos (*clustering*) são comumente aplicadas para esse objetivo. A *descrição conceitual* consiste da atribuição de um conceito ou um significado a um grupo segmentado. Esse pode ser um objetivo ou apenas uma etapa para obter novas idéias. Algumas técnicas segmentam e atribuem conceitos concomitantemente (agrupamento conceitual). A *detecção de desvios* procura por ocorrências que sejam diferentes em alguma medida. Objetivos desse tipo podem ser um resultado final (detecção de fraudes) ou uma etapa de limpeza de dados. A *análise de dependência* associa conceitos que dependam de outros conceitos. É uma etapa fundamental para modelos preditivos e têm como entrada os resultados da descrição conceitual. A *classificação* já supõe a existência de diferentes classes e então o problema passa a ser definir o modelo que melhor classifique as ocorrências. Este é um objetivo final freqüente em diversas aplicações. A *previsão* consiste em encontrar um atributo numérico para ocorrências desconhecidas. A classificação distingue-se da previsão pela atribuição de um atributo simbólico ao resultado.

As fases seguintes dependerão diretamente dos objetivos definidos, pois decisões serão tomadas em função deles, inclusive medidas de conhecimento são definidas tendo em mente as metas a se alcançar.

3.2.3.2 Construção do Banco de Dados para Mineração

Para que um banco de dados seja submetido à mineração de dados, um conjunto de tarefas deve ser realizado para que se assegure a origem, a qualidade, a clareza e a transferência dos dados para o BD de mineração. A construção se inicia com a ***coleta dos dados.*** Nesta fase a origem dos dados é identificada. Muitas vezes, é necessário o uso de fontes externas ou então elaborar processos de coleta. No primeiro caso, as fontes externas podem ser obtidas em bancos de dados institucionais (censo, órgãos governamentais, laboratórios, hospitais) e neste caso há a suspeita de que tais dados possam gerar informação sobre o objetivo definido na fase anterior. Na elaboração de um processo de coleta, os dados podem ser mais bem definidos pelos projetistas, no entanto, corre-se o risco de se influenciar o processo

de KD, pois dados fundamentais podem ser excluídos da composição do conhecimento. É possível que todo um modelo possa ser invalidado pela incapacidade de se fazer uma coleta isenta de influências. O administrador do banco de dados (DBA – *Database Administrator*), é a pessoa que tem a responsabilidade de definir, baseado em restrições do projeto do BD. Assim, questões como custo, organização e mídias de armazenamento, requisitos de segurança e privacidade além das restrições de uso são muitas vezes determinantes no sucesso de um projeto de descoberta de conhecimento (KD).

Os dados coletados passam então por um processo de ***seleção*** para definir quais dados deverão ser submetidos ao processo de descoberta. Ressalva-se aqui que este tipo de seleção não corresponde à seleção de atributos. Nesta etapa, dados que não serão relevantes não deverão ser incluídos no BD de mineração. Por exemplo, se o interesse for descobrir modelos sobre uma determinada região geográfica ou país e na origem dos dados constam dados mundiais, deverão ser retirados os dados que não correspondem àquela região ou país.

Como os dados podem vir de orígens diferentes, é natural a ocorrência de redundâncias, problemas de integridade de domínio ou nomenclatura, ou ainda inconsistências. Uma fase de ***consolidação e integração*** dos dados deve ocorrer para que dados redundantes possam ser eliminados adotando um dos valores. Problemas de integridade como campos contendo valores com mesmo significado mas de naturezas diferentes (contínuos, discretos ou enumerados) devem ser uniformizados. Dados inconsistentes, por sua vez, são aqueles cujas ocorrências são impossíveis ou que não pertençam ao domínio. Por exemplo, homens grávidos, valores inválidos ou quantidades negativas. Estratégias para resolver este problema são similares àquelas adotadas para resolver valores inexistentes e que são descritos logo a seguir.

Os dados coletados devem passar ainda por processos de ***avaliação*** que permitam assegurar que as medidas coletadas sejam compatíveis com o que se quer medir, ou seja, é preciso avaliar a qualidade dos dados. Reinartz [REINARTZ, 1999] destaca duas alternativas: (1) a mais imediata é deixar por conta do algoritmo de mineração de dados que possa trabalhar com dados imperfeitos; ou então (2) o analista dos dados (humano) elimina os ruídos utilizando suportes como visualização, análise de dados ou agrupadores. Dentre os objetivos dessa etapa estão

(*a*) a remoção de ruídos e valores muito fora do domínio (*outliers*) e (*b*) definir um tratamento adequado para valores inexistentes. Métodos de agrupamento (*clustering*) que permitem desconsiderar *outliers* ou filtros são utilizados na etapa (*a*). Para (*b*) algumas estratégias podem ser úteis no tratamento de valores inexistentes. (*i*) Gerar um atributo com o valor inexistente permitiria considerá-lo como um estado relevante do objeto. (*ii*) Ignorar objetos que apresentem valores inexistentes pode ser útil quando tais objetos forem raros, caso contrário, há o risco de perder muitas amostras; (*iii*) Substituir por outros valores requer uma imposição baseada na amostragem. Por exemplo, valores nominais podem ser trocados pelo valor modal. Valores ordinais pela mediana e os contínuos pela média. Há ainda a possibilidade de fazer o valor acompanhar a distribuição do atributo correspondente.

A ***construção de metadados*** implica o estabelecimento de descrições sobre os dados. Estas descrições, ou seja, os metadados, devem permitir o conhecimento sobre as consolidações. Por exemplo, se um campo é resultante do cálculo da média sobre dados em outra tabela, devem ficar explícitos quais os atributos e como foi calculado o valor do campo. Isto pode ser realizado por um banco de fórmulas ou mesmo permitindo que se armazene fórmulas que serão submetidas a um *parsing* para análise sintática e semântica de sua expressão.

É momento de ***carregar os dados no BD de mineração***. Os dados devem ser armazenados em uma tabela tipicamente com grandes tuplas, ou seja, com um objeto ou amostra sendo descrito por um grande número de atributos. A ferramenta de descoberta de conhecimento deve assegurar que a partir do momento em que os dados constem no BD de mineração eles farão parte do processo da descoberta de conhecimento. A partir de cada descoberta, os dados vão sendo refinados e novos dados poderão chegar ao BD de mineração. Assim, é necessário um esquema de ***manutenção do BD de mineração*** que viabilize este refinamento permitindo a adição de novos dados passando pelas fases descritas acima.

3.2.3.3 Preparação dos Dados

Esta fase é geralmente a mais demorada no processo de KD. Ela pode ainda ser dividida em outras quatro fases: (a) Seleção de atributos, (b) Amostragem, (c) Construção de novos atributos e (d) Transformação.

a) Seleção de atributos. Com os dados limpos, é realizada a seleção das variáveis independentes que serão utilizadas no processo. É importante nessa fase que se defina medidas de independência. Dentre tais medidas está a taxa de correlação definida por (1)[MULLEN, 1973]. Um valor nulo de r indica uma independência entre os dois atributos x e y. Os valores extremos 1 e –1 indicam respectivamente uma dependência linear direta e indireta.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \qquad (1)$$

Em (1), x_i e y_i denotam um conjuntode dados com N elementos e cujos respectivos valores médios correspondem a $\bar{x}$ e $\bar{y}$.

Pode-se ainda lançar mão de ferramentas de visualização ou de algoritmos agrupadores que permitem avaliar um índice de qualidade geral dos agrupamentos [JAIN & DUBES, 1988].

b) Amostragem. A amostragem é um recurso que permite a redução dos dados para uma quantidade que viabilize processamentos mais sofisticados envolvendo grande quantidade de dados. Cabe aqui destacar que as estratégias de amostragem devem garantir que os dados sejam aleatoriamente amostrados, caso contrário estará sendo inserida uma tendência no resultado.

c) Construção de novos atributos. Quando os modelos não forem bem sucedidos, novos atributos devem ser construídos ou prospectados. Estes atributos devem ser desenvolvidos de uma maneira independente dos existentes.

d) Transformação: Feita a seleção, e dependendo de seus resultados, é possível que se queira transformar o conjunto de dados para adequá-lo a um processo de treinamento. Esta transformação pode ser realizada, discretizando-se intervalos (Alto, Médio e Baixo) em função de limiares sobre um intervalo contínuo. Pode ainda ser tomada sobre estados (ligado, desligado, desconhecido) ou ainda submetidos a funções de equalização (sigmóide ou arco-tangente) para considerar apenas determinados intervalos. É absolutamente necessário que se entenda que tais modificações podem alterar a qualidade do resultado.

3.2.3.4 Construção do Modelo

A construção do modelo é um processo iterativo. A prospecção por alternativas de modelos é realizada até que se encontre aquele adequado ao problema. É necessária, portanto, uma medida do desempenho que aplicada às ocorrências que não foram usadas no treinamento permita a aprovação ou não do modelo (veja a subseção *Avaliação*).

O modelo é construído levando-se em conta os objetivos definidos na fase inicial (veja seção 3.2.3.1). Várias abordagens são úteis para atingir cada um dos possíveis objetivos. No entanto, a inúmera quantidade de métodos e técnicas utilizados para isso torna conveniente o estabelecimento de uma taxonomia destas técnicas e métodos. Weiss et al [WEISS & INDURKHYA, 1997] propõem uma taxonomia bastante simplificada para construtores de modelos. Em um *white paper* da Information Discovery Inc.[INFORMATION DISCOVERY, 1997], toma-se o cuidado de adicionar uma classificação desses construtores de modelos em função da retenção dos dados, isto é, se os dados originais são utilizados durante o processo de descoberta.

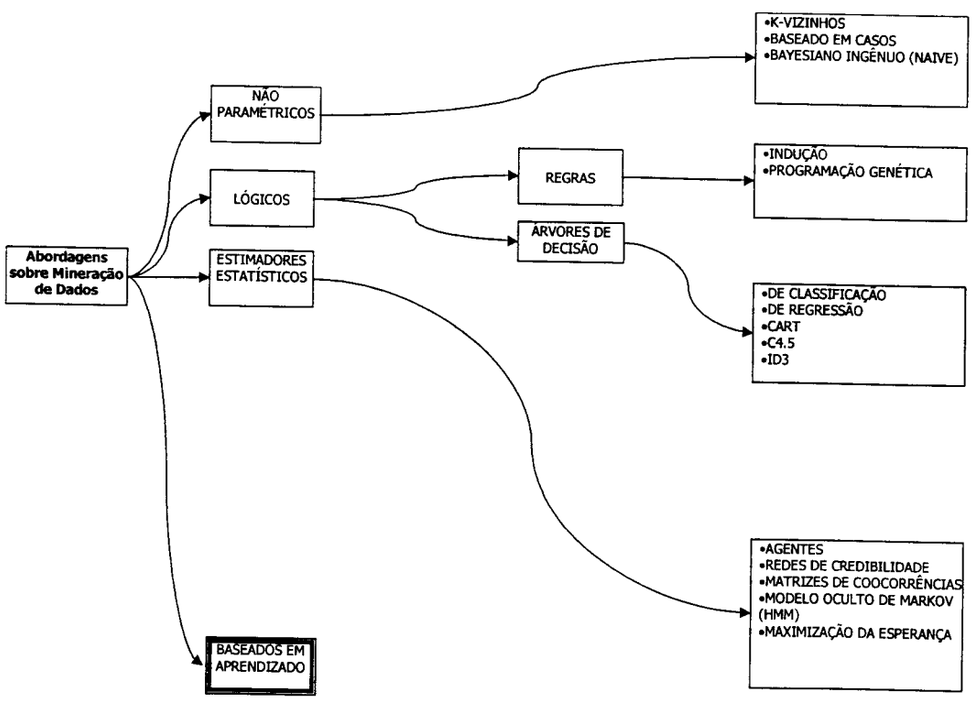


Figura 13 – Taxonomia proposta para as abordagens sobre mineração de dados.

Inspirado nos trabalhos de Weiss et al [WEISS & INDURKHYA, 1997; INFORMATION DISCOVERY, 1997], propõe-se uma taxonomia mais abrangente para as abordagens de mineração de dados baseada no paradigma de análise a que os dados serão submetidos. Quatro grupos principais são identificados: (1) Não paramétricos; (2) Lógicos; (3) Estimadores Estatísticos; e (4) Baseados em Aprendizado (veja Figura 13).

A *análise não-paramétrica* é o processo mais intuitivo de classificação [JAIN et al., 2000]. Para separar ou agrupar um elemento assume-se uma regra baseada em uma medida de distância ou de similaridade. Uma regra bastante conhecida é a regra dos vizinhos mais próximos (*k-Nearest Neighbor*), no entanto não se garante que tais regras separem as classes conforme o objetivo de uma mineração de dados ou seleção, pois muitas vezes os critérios de separação ou agrupamentos exigem mais que uma simples métrica.

Os sistemas baseados em casos ou CBR (*Case-based Reasoning*) assemelham-se a analogias. Um caso é uma relação entre um conjunto de atributos. Relações similares a um determinado caso devem ser buscadas no BD de mineração, gerando um cenário, ou seja, um conjunto de casos possíveis. Este tipo de abordagem é utilizado em ferramentas de visualização para percorrer um BD ou então para gerar regras a partir de casos emblemáticos e seus respectivos cenários.

O método Bayesiano ingênuo consiste em aplicar a teoria de Bayes escolhendo a classe de uma instância que apresente a maior probabilidade *a posteriori*. O adjetivo "ingênuo" é atribuído ao método por considerar os atributos como independentes, possibilitando o cálculo da probabilidade *a posteriori* baseado na freqüência de ocorrências dos atributos no conjunto de dados. É utilizado na análise de conteúdo de textos ou em *wrappers*, ou seja, identificadores de subconjuntos de atributos para reduzir a dimensão de um espaço de busca, por exemplo.

A classe de *abordagens lógicas* baseia-se em regras ou decisões. As regras podem ser definidas a partir de sua indução ou escolhidas por algum algoritmo que busque as regras em um universo determinado pelos objetivos inicialmente. Um exemplo deste tipo de abordagem é a programação genética introduzida por Koza [KOZA, 1992] que seleciona as melhores regras por um algoritmo evolucionário. As

árvores de classificação e as de regressão armazenam regras intermediárias. Um conjunto de atributos referente a um novo objeto pode agora ser submetido às regras da árvore para que este seja classificado e se for o caso de uma árvore de regressão, sua inclusão poderá provocar uma mudança na estrutura da árvore. A Figura 14*a* ilustra um exemplo de árvore de classificação para uma avaliação de consumidores de um produto. Os diâmetros dos círculos são proporcionais ao número de objetos para cada decisão. São avaliadas a origem (nacional ou internacional) e a idade (acima ou abaixo de 30 anos). Um consumidor preferencial pode resultar desta árvore e que no caso é um consumidor nacional com menos de 30 anos. Já na Figura 14*b* os atributos são os mesmos, mas está sendo usado um outro atributo para avaliar o consumidor: a renda média e o seu desvio padrão. Se um novo cliente é adicionado ao cadastro da empresa, a renda média daquele caminho provavelmente irá se alterar. Assim, a regra ***SE*** *cliente reside em território nacional e tem menos de 30 anos* ***ENTÃO*** *sua renda média é R$40.000,00 com desvio padrão de R$3.400,00* pode ser alterada quando um novo cliente é inserido no BD.

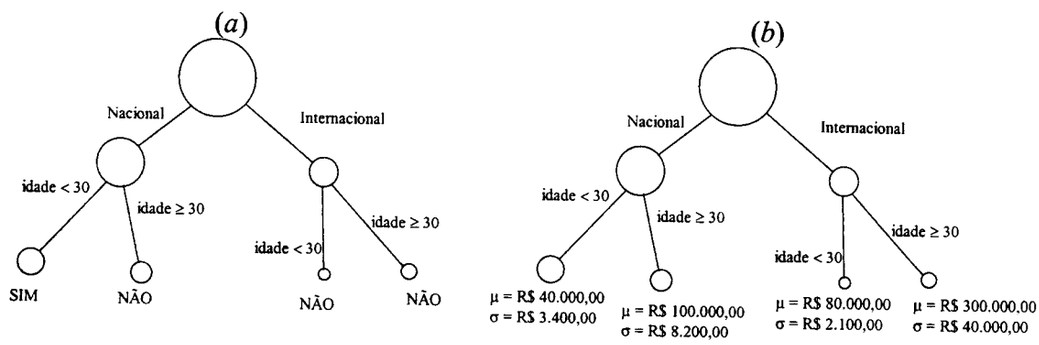


Figura 14 – Uma árvore de classificação (*a*) e uma árvore de regressão (*b*). (adapatado de [DHAR & STEIN, 1997])

As árvores de classificação e de regressão podem ser usadas para resultar regras de generalização muito utilizadas em aplicações de gerenciamento de relacionamentos com consumidores ou CRM (*Customer Relationship Management*) e em taxonomias [SRIKANT & AGRAWAL, 1997]. Bibliotecas de aprendizado de máquina (*machine learning*) incluem facilidades para implementar tais árvores. Dentre as mais conhecidas estão a C4.5 [QUINLAN, 1993] e a CART (Classification and Regression Trees) [BREIMAN et al., 1984; APTÉ & WEISS, 1997]. Os métodos implementados nesse tipo de biblioteca incluem buscas heurísticas que minimizam a desordem (entropia) em um conjunto de dados (C4.5) ou extraem os grafos (árvores

de classificação e regressão) usando algoritmos de análise de fatores ou agrupadores (*clustering*) (CART).

Os métodos que utilizam *abordagens estatísticas* são aqueles utilizados em reconhecimento estatístico de padrões. Jain et al. [JAIN et al., 2000] estabelece a abordagem estatística do reconhecimento de padrões como aquele problema em que cada padrão é representado por *d* medidas, ou seja, como um ponto em um espaço *d* dimensional. O objetivo é então delimitar este espaço vetorial em regiões compactas para que se possa conhecer sua classe. Jain et al. distinguem em sua taxonomia as abordagens estatísticas daquelas que demandam aprendizado ou treinamento com base no conhecimento de suas densidades de classe condicionais. O esquema da Figura 15 mostra um modelo para reconhecimento estatístico de padrões que se assemelha muito aos mineradores de dados para conhecimento do problema.

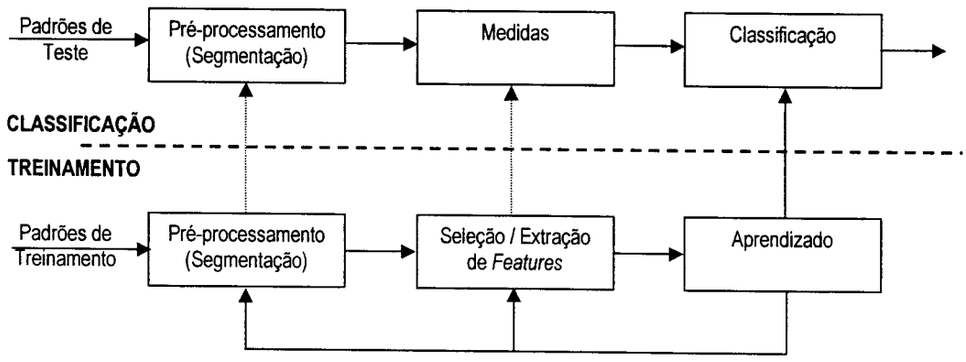


Figura 15 – Modelo para reconhecimento estatístico de padrões [JAIN et al., 2000].

Os estimadores estatísticos envolvem duas abordagens principais: o estimador da máxima verossimilhança e o estimador bayesiano. Estes estimadores diferem na forma como um vetor de atributos é determinado. No estimador de máxima verossimilhança, o vetor de atributos que melhor representa um conjunto de dados é fixo, ou seja, não é uma variável aleatória cujo treinamento define uma distribuição, como no treinamento bayesiano. A partir destas duas abordagens foram propostas soluções para problemas como classificar pontos de teste quando atributos de treinamento podem não existir. Neste caso, a maximização da esperança (EM) é um método baseado na busca de uma amostra representativa. Esta busca acontece no espaço de atributos. Por outro lado, pode ser definida uma estrutura revelando as interdependências entre os diversos atributos e as respectivas probabilidades

condicionais. Esta estrutura é chamada rede de confiabilidade e pode ser útil na definição de um estado mais provável de um vetor de atributos incompleto. Um pouco mais complexos que os dois anteriores, os modelos ocultos de Markov ou HMM (*Hidden Markov Model*) consiste de um grafo que considera a variação de seus estados no tempo, no qual também é permitida a existência de transições invisíveis entre os nodos ou estados. Estes estados invisíveis transformam uma máquina de estados determinística em uma ferramenta de aprendizado e treinamento de soluções [DUDA et al., 1998].

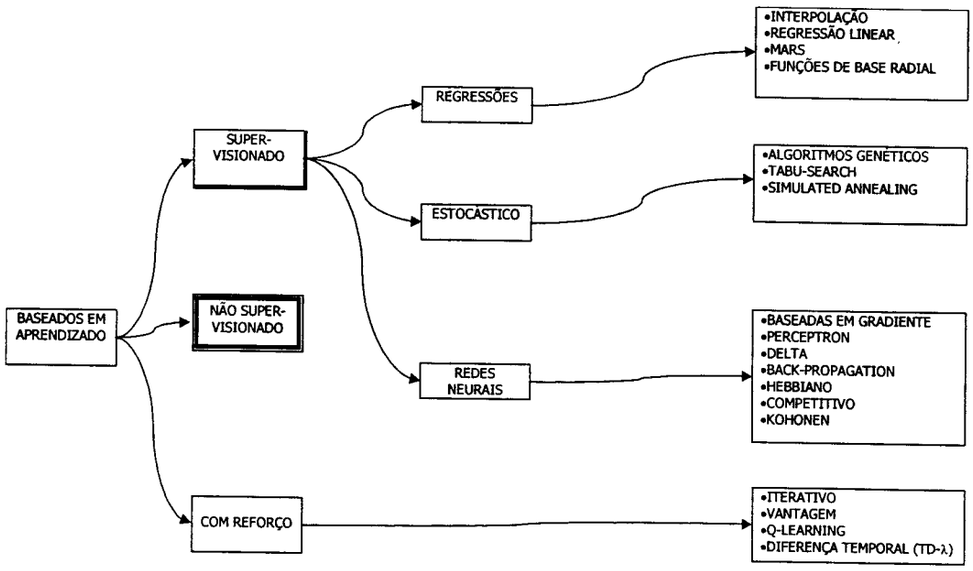


Figura 16 – Abordagens de mineração de dados baseadas em aprendizado.

As *abordagens baseadas em aprendizado* contam com um grande número de abordagens. São subdivididos de acordo com a existência ou não de referências para comparação (supervisionado ou não supervisionado) ou ainda na utilização de reforço no aprendizado (*reinforcement learning*) (veja Figura 16).

Para o aprendizado supervisionado, existe uma referência e portanto um resultado almejado conhecido. O treinamento compara os resultados obtidos por uma solução proposta pelo aprendizado com o resultado desejado. É parecido com o processo de aprendizagem em uma sala de aula, na qual o aluno é submetido a testes e se não obtiver avaliação satisfatória voltará a aprender. As alternativas que existem para o elemento de aprendizagem são abundantes e normalmente as respostas são resultados de dois tipos de abordagem: classificação ou regressão. O que entra nesses

algoritmos são medidas sobre as ocorrências de treinamento, e portanto as respostas são um conjunto de classes e distribuições (no caso da classificação) ou um valor escalar (no caso da regressão).

Na classificação a resposta é um conjunto de classes que pode estar ou não correto. Uma função de similaridade serve como avaliação de desempenho. Se um determinado valor de similaridade for alcançado, considera-se a aprendizagem suficiente e o modelo é, na verdade, o conjunto de parâmetros para o algoritmo de classificação (número de classes e distância entre elas).

Na regressão, as ocorrências são aproximadas por uma função, ou ainda por combinações lineares de funções. O valor escalar de um conjunto de aplicações na função é comparado com um valor de referência e um erro serve para determinar se o modelo está suficientemente bom ou não.

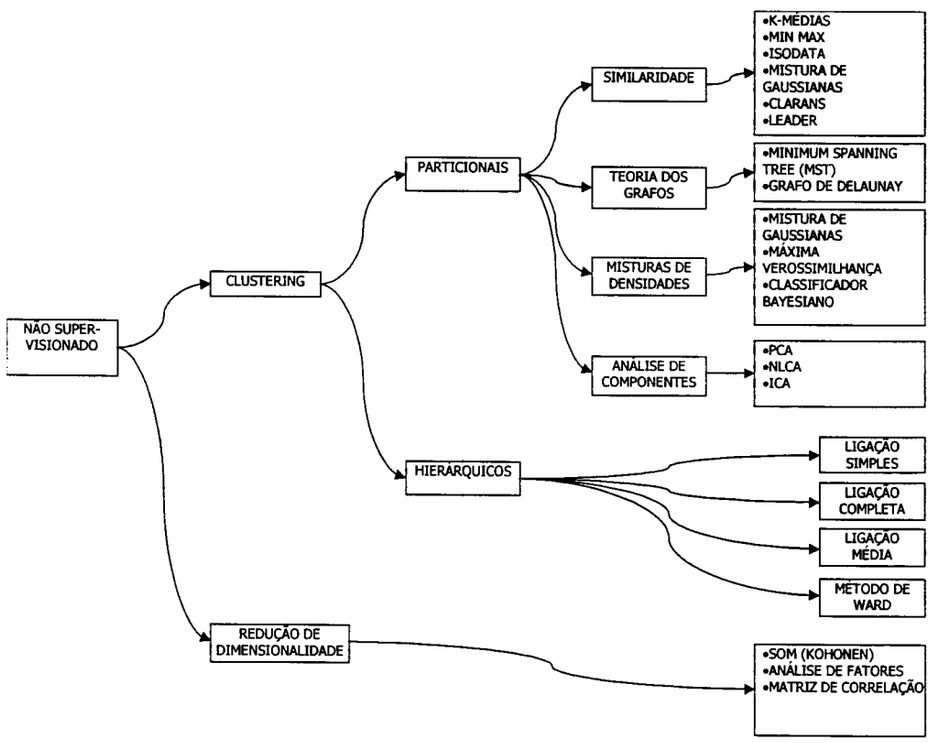


Figura 17 - Abordagens de mineração de dados baseadas em aprendizado não supervisionado.

O aprendizado não supervisionado não utiliza uma referência para avaliar o resultado de cada iteração. O desempenho é avaliado por um critério inerente aos dados de entrada e resulta, portanto, de uma saída do aprendizado que indique o momento de parar. Uma outra maneira de parar é determinar um limiar *ad hoc* do

número de iterações. Muitas vezes as duas condições de parada são usadas, isto é, determina-se um número de iterações para atingir o desempenho desejado, e se isso não acontecer assume-se como critério de parada o número de iterações atingindo um limiar.

Abordagens típicas para aprendizado não supervisionados são os agrupadores (*clustering*) [ALDENDERFER & BLASHFIELD, 1984; JAIN et al., 1999] ou os mapeamentos auto-organizáveis (SOM) [KOHONEN, 1989].

Um terceiro tipo de treinamento é o aprendizado com reforço (*reinforcement learning*). Esse tipo de aprendizado acontece com a interação entre duas partes: o agente e o ambiente. O agente sente o ambiente e baseado em sua entrada sensorial toma uma decisão que resultará em uma ação. Esta ação altera o ambiente de alguma maneira e essa mudança é comunicada ao agente por meio de um sinal escalar de reforço. O agente realiza o aprendizado usando duas funções: a função de reforço e a função de valor. A função de reforço mapeia as circunstâncias encontradas pelo agente em ações e portanto reflete o objetivo do aprendizado. A função de valor é uma maneira de medir a utilidade da ação. Para tanto, mapeia-se os estados e as ações por meio de uma política e busca-se maximizar a soma de reforços que deve ser aplicada a partir de cada estado, como se fosse tomado arbitrariamente. Para isso uma função de aproximação pode ser usada (*perceptron* multi-camadas, funções de base radial, entre outras). A função de valor pode ser modelada por uma cadeia de Markov representando um processo Markoviano de decisões, definindo os valores dos reforços como nodos e a política como arcos direcionados entre os nodos. O aprendizado ocorre conforme a trajetória da interação entre o agente e o ambiente [BALLARD, 1997].

Dentre as abordagens propostas para modelar a função de valor, a programação dinâmica tem sido amplamente usada. Ela se baseia em dois princípios: (1) se uma ação causa algo ruim de imediato, então ela não deve mais ser usada; (2) se todas as ações acarretam uma situação ruim, então a situação deve ser evitada. A grande vantagem da programação dinâmica é que o aprendizado nunca resulta em um acidente. Entre as abordagens de programação dinâmica para aprendizado reforçado (*reinforcement learning*) estão: a iteração do valor (a mais simples);

Q-*learning*; *Advantage learning*, TD(λ) [KAELBLING et al., 2000; TESAURO, 1995; BALLARD, 1997].

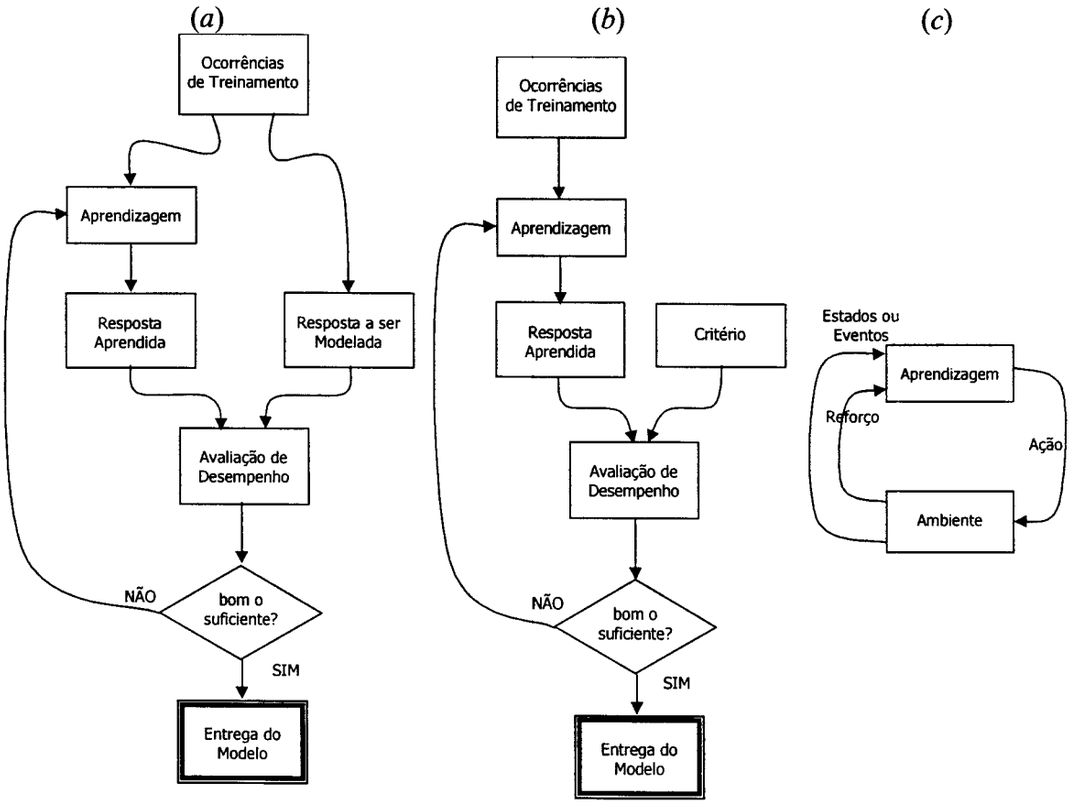


Figura 18 - Tipos de aprendizados: (*a*) supervisionado; (*b*) não-supervisionado; (*c*) com reforço

Outros critérios podem ser usados para definir as abordagens de reconhecimento de padrões utilizadas em mineração de dados. Jain et al. [JAIN et al., 1999] selecionou alguns critérios para classificar algoritmos de agrupamento. Alguns destes critérios podem ser generalizados para abordagens de reconhecimento de padrões em geral. A grande quantidade de dados, por exemplo, torna desejável ao algoritmo a possibilidade de particionar os dados de uma forma incremental, ou seja, tornar desnecessário que todos os dados estejam disponíveis para que o processo de mineração dos dados ocorra. Técnicas que buscam minimizar uma função de erro (por exemplo, erro quadrático) podem realizar uma busca utilizando métodos estocásticos ou quando possível por abordagens determinísticas. Os limites de decisão são também um critério para classificar uma abordagem. Limites nebulosos de decisão são úteis quando se quer definir apenas contribuições e uma medida na

classificação de um objeto, definindo assim um grau de pertinência do objeto a uma classe. Limites rígidos são interessantes quando o problema não admite ambiguidades.

É claro que o modelo obtido não funciona seguramente para qualquer conjunto de dados no qual for aplicado, pois o modelo é uma generalização sobre os dados disponíveis. Portanto, um resultado exato para um conjunto de treinamento ou testes não é desejável, já que o modelo estaria refletindo exatamente tais ocorrências e não uma ocorrência geral. Para determinar a acurácia de um determinado modelo é definida uma medida de distância entre o desejado e o obtido. Tal distância pode ser obtida por um erro quadrático médio, por exemplo.

Para modelos com pouca quantidade de ocorrências, é aconselhado por vários autores [TWOCROWS, 1998; WEISS & INDURKHYA, 1997], o uso de validação cruzada. A validação cruzada consiste em separar as ocorrências em mais do que dois grupos, n por exemplo. Um dos grupos é determinado como conjunto de testes e o modelo é construído treinando-se os outros $n-1$ grupos. Este processo ocorre em n iterações, sendo que em cada iteração um dos $n-1$ grupos é escolhido como conjunto de testes e o que estava sendo usado como conjunto de testes é agregado aos $n-1$ para treinamento. A medida média de acurácia de todas as iterações é utilizada para medir a acurácia geral do modelo utilizando todos os dados.

Um método alternativo é o *bootstraping* que se baseia no conjunto todo de ocorrências. A diferença é que os dados são reamostrados e para melhorar a acurácia, as ocorrências são substituídas, sendo permitida a repetição de ocorrências. A acurácia é calculada sobre o erro médio entre todas as substituições ocorridas. Com tudo isso, não é ainda garantido que funcione [TWOCROWS, 1998]. Não existe, enfim, ferramenta perfeita que dará, com certeza resultados desejados. É bastante comum, em implementações, a tentativa de várias técnicas na construção de modelos antes de encontrar uma satisfatória. Além disso, cada nova técnica alternativa testada pode exigir novas preparações de dados ou diferentes formatos de arquivos. Muitas técnicas de treinamento podem ser implementadas valendo-se de paralelismo para reduzir o tempo de execução da modelagem e permitindo que mais alternativas possam ser testadas [FREITAS & LAVINGTON, 1998].

O aprendizado não supervisionado destina-se a problemas cuja referência seja desconhecida. Algoritmos que tipicamente são usados para esse tipo de modelagem são *clustering*, detecção de seqüências e regras de associação, mapeamento auto-organizável, entre outros. Como a referência não existe, a medida de acurácia é dada por algum parâmetro inerente ao algoritmo. Essa dependência do algoritmo pode levar aos modelos as restrições a que tais métodos estejam impostos. Tais restrições podem partir desde a cobertura espacial do modelo, até algum parâmetro interno do modelo.

3.2.3.5 Avaliação e Interpretação

Validação do Modelo. Depois de construído o modelo, deve-se avaliar os resultados e interpretar sua significância. A acurácia medida nas fases de treinamento, teste e validação são parâmetros que dependem do conjunto de dados. Deve-se portanto, buscar técnicas que permitam ampliar os resultados obtidos. Uma dessas técnicas é a Matriz de Confusão[CABENA et al., 1997].

Validação externa. Como já foi dito antes, não há garantia de que o modelo obtido funcionará quando submetido ao mundo real. Assim, um modelo válido não implica em um modelo correto. Uma das principais razões para isso é que sempre haverá hipóteses implícitas no modelo. Portanto, é muito importante testar o modelo no mundo real [WEISS & INDURKHYA, 1997].

3.2.3.6 Acompanhamento

Medidas devem ser desenvolvidas para acompanhar a aplicação do modelo. O modelo pode ser avaliado durante o uso e a medida é comumente realizada por ferramentas de visualização [TWOCROWS, 1998].

3.3 Técnicas de Mineração de Dados

Esta seção descreve algumas abordagens significativas para mineração de dados e que podem ser úteis para entender e continuar a obter resultados deste trabalho.

A classificação fornecida pela teoria de decisão de Bayes é introduzida na seção 3.3.1, pelo fato de servir como parâmetro para outros classificadores. Duas técnicas de classificação não-paramétricas (seção 3.3.2) são apresentadas e logo em

seguida algoritmos de agrupamentos particionais (seção 3.3.3) e hierárquicos (seção 3.3.4) são revisados. Alguns destes algoritmos são adequados para grandes volumes de dados, como é típico para problemas de mineração de dados. Algoritmos classificadores e de agrupamento são úteis em qualquer fase de um sistema de descoberta de conhecimento, principalmente em bancos de dados (KDD). Esta aplicabilidade é adequada para a fase de construção do banco de mineração de dados, com técnicas que são utilizadas para indexar os dados para que informações possam ser obtidas mais rapidamente. Na limpeza de dados os agrupadores são utilizados na remoção de *outliers*. Pode ser ainda usado para definir classes desconhecidas ou ainda como suporte à visualização da informação. Na fase de extração de conhecimento os agrupadores ou classificadores podem segmentar o conjunto de regras e auxiliar na identificação daquelas que sejam compatíveis com os objetivos definidos pelo KDD.

A redução de dimensão de um espaço de atributos é também tarefa freqüentemente enfrentada por mineradores de dados. Técnicas como o mapa auto-organizável (SOM) são particularmente úteis para esse tipo de problema e portanto são introduzidos nessa pequena revisão (seção 3.3.5).

A busca em um espaço de atributos por aqueles mais significativos pode se tornar uma árdua tarefa. Este problema ocorre quando o espaço tem uma dimensão muito alta ou a função que define o objetivo for muito complexa. Quando se quer também determinar a influência dos atributos, este problema passa a ser chamado de construção do modelo de mineração de dados. O conjunto de técnicas para resolver a identificação dos atributos significativos freqüentemente inclui otimizadores ou ferramentas de busca de uma solução adequada. Uma técnica bastante popular devido a sua simplicidade e facilidade de implementação é o algoritmo genético (GA), introduzido na 3.3.6. Por sua vez, para encontrar o modelo, ou seja, determinar a influência dos atributos, é necessário generalizá-los a partir de um conjunto de treinamento para os atributos de um conjunto de testes ou então de novas amostras. Deste modo, técnicas de aproximação de soluções como a interpolação podem ser utilizadas para esse fim. Portanto, é descrita uma técnica de interpolação na seção 3.3.7 cujas características são desejáveis para um minerador de dados: a "*thin-plate*

spline" (TPS). Esta técnica foi utilizada intensivamente ao longo desse trabalho com bons resultados para desempenho, flexibilidade e tempo de execução.

3.3.1 *Teoria de Decisão de Bayes*

A teoria de decisão de Bayes é a principal abordagem probabilística para um classificador. Essa abordagem teórica sobre a classificação permitiu traçar os limites teóricos dos classificadores. É comum encontrar análises de desempenho de classificadores baseados na Regra de Bayes, pois ela garante uma taxa média de probabilidade de erro mínima (ótima). A decisão é tomada a partir de probabilidades conhecidas sobre a natureza de um conjunto de objetos. Cada elemento desse conjunto é caracterizado por um vetor de atributos $\vec{x} = (x_1, \cdots, x_d)$, no qual d é o número de atributos ou dimensão do espaço de possíveis atributos. Um conjunto de classes as quais se quer rotular cada objeto é dado por $w = \{w_1, \cdots, w_c\}$, (c é o número de classes). Podemos ilustrar a regra de decisão Bayseana usando o exemplo a seguir.

Um produtor rural quer criar um classificador de gado leiteiro quanto à produtividade. Ele consulta a unidade de pesquisa agropecuária de sua região e obtém um conjunto de dados dos quais ele destaca a altura e peso de cada exemplar e sua respectiva produtividade em litros por mês. Seu vetor de atributos é então dado por intervalos de altura e peso (baixo, médio e alto por exemplo) e pode ser representado por $\vec{x} = (x_{altura}, x_{peso})$. O conjunto de classes depende de uma escolha do produtor. Se ele quiser que cada indivíduo seja identificado como produtivo ou improdutivo, então $w = \{w_{produtivo}, w_{improdutivo}\}$. Como o produtor dispõe dos dados sobre o gado, as probabilidades $P(w_{produtivo})$ (a probabilidade do indivíduo ser produtivo) ou $P(w_{improdutivo})$ (a probabilidade de ser improdutivo), pode ser calculada. Além disso, pode ser conhecido o conjunto de pares de intervalos de altura e peso (x_{altura}, x_{peso}) para os quais o indivíduo é produtivo ou não. A probabilidade do indivíduo pertencer a tal conjunto é dada por $P(\vec{x} \mid w_{produtivo})$ ou então $P(\vec{x} \mid w_{improdutivo})$. A probabilidade conjunta do indivíduo ter peso e altura em intervalos nos pares (x_{altura}, x_{peso}) e ser produtivo ou improdutivo pode ser calculada de duas maneiras:

$$P(w_k, \vec{x}) = P(w_k \mid \vec{x})P(\vec{x}) = P(\vec{x} \mid w_k)P(w_k) \qquad (2)$$

k é um identificador da classe. Assim

$$P(w_k \mid \vec{x}) = \frac{P(\vec{x} \mid w_k)P(w_k)}{P(\vec{x})} \tag{3}$$

Esta é a fórmula de Bayes que permite a classificação de um objeto (conhecimento *a posteriori* ou $P(w_k \mid \vec{x})$) a partir de objetos já classificados (conhecimento *a priori* ou $P(w_k)$). A *semelhança* entre as classes já existentes e os objetos a serem classificados são determinados por $P(\vec{x} \mid w_k)$. A *evidência* ou $P(\vec{x})$ é determinada pela proporção destas características no conjunto de objetos já classificados.

$$conhecimento\ a\ posteriori = \frac{semelhança \times conhecimento\ a\ priori}{evidência} \tag{4}$$

Voltando ao exemplo, cada indivíduo que nascer ou for adquirido poderá ser classificado como produtivo ou improdutivo a partir de suas medidas de altura e peso. Esta classificação é resultado de uma decisão: se $P(w_{produtivo} \mid \vec{x}) > P(w_{improdutivo} \mid \vec{x})$ classifica-o como produtivo, caso contrário como improdutivo. Esta decisão pode estar errada, e a probabilidade de ocorrer um erro pode ser avaliada a partir da decisão tomada.

$$P(erro \mid \vec{x}) = \begin{cases} P(w_{produtivo} \mid \vec{x}) & se\ a\ decisão\ foi\ para\ indivíduo\ improdutivo \\ P(w_{improdutivo} \mid \vec{x}) & se\ a\ decisão\ foi\ para\ indivíduo\ produtivo \end{cases} \tag{5}$$

Como a decisão tomada para cada objeto é a que implica no menor erro possível, ou seja, $P(erro \mid \vec{x}) = \min\{P(w_{improdutivo} \mid \vec{x}), P(w_{produtivo} \mid \vec{x})\}$, garante-se que ao longo de sucessivas classificações a probabilidade média seja mínima também.

$$P(erro) = \int_{-\infty}^{\infty} P(erro, \vec{x})d\vec{x} = \int_{-\infty}^{\infty} P(erro \mid \vec{x})P(\vec{x})d\vec{x} \tag{6}$$

É essa propriedade que faz do classificador Bayeseano um parâmetro para o desempenho de outros algoritmos classificadores. O exemplo apresentado é bastante simples e mostra uma decisão entre duas classes e vetores de atributos dados por intervalos. A aplicação para diversas classes é direta, contudo para cada classe pode ser atribuída uma função de perda que pondera a probabilidade de erro para cada classe. Quando a função de perda assume apenas dois estados (0 ou 1), o classificador é identificado como Bayseano com função de perda 0/1. Os valores dos

atributos podem também ser contínuos. Nesse caso, a decisão é tomada a partir de uma função de custo contínua sobre o espaço de atributos que implicará em um risco de se tomar uma decisão. Estima-se uma distribuição *a priori* para a ocorrência dos atributos nos objetos. Supor que os atributos ocorram seguindo uma distribuição normal implicará em separadores lineares, quadráticos ou hiperbólicos perfeitamente estensíveis para dimensões maiores. É bom deixar claro que tais curvas, superfícies ou hipersuperfícies de decisão são tomadas a partir de um conhecimento *a priori* da distribuição das probabilidades de ocorrências dos objetos (atributos). Em aplicações com grande quantidade de objetos e cujas ocorrências são naturais e aleatórias é possível adotar esse tipo de premissa (Teorema Central do Limite).

3.3.2 Vizinho mais próximo

Para aplicar a regra de decisão Bayesiana em um espaço de atributos contínuo é necessário que haja uma estimativa de distribuição das ocorrências dos objetos nesse espaço. No entanto, em muitas aplicações esse conhecimento não existe ou não seria adequada essa estimativa. São casos, por exemplo, em que o número de objetos não é tão grande ou que não haja uma aleatoriedade na sua geração. As decisões devem ser tomadas, portanto, em função de um critério que em muitos algoritmos é assumido por uma métrica no espaço de atributos. Entre estes métodos estão os que consideram a proximidade a um vizinho como esse critério.

3.3.2.1 1-NN

"Vizinhos mais próximos" ou 1-NN é uma regra de tomada de decisão para classificar uma amostra. A regra é bastante simples: o objeto assume a classe do vizinho mais próximo. Assuma uma representação de um objeto a ser classificado como um vetor de atributos

$\vec{x} = (x_1, \cdots, x_d)$, no qual d é o número de atributos

Considere também um conjunto de classes

$w = \{w_1, \cdots, w_c\}$, no qual c é o número de classes.

Existe também um conjunto de amostras já conhecidas dadas por

$\vec{s}_i \in \{\vec{s}_1, \cdots, \vec{s}_n\}$, no qual n é o número de amostras e cada elemento $\vec{s}_i$ também é representado como um vetor de atributos.

A decisão é qual classe de w $\vec{x}$ deverá assumir. Tal decisão é tomada verificando-se a classe do vizinho mais próximo, isto é, escolhendo-se a amostra do conjunto cuja distância seja $D(\vec{s}_i, \vec{x}) = \min_k \{D(\vec{s}_k, \vec{x})\}$. Deste modo a classe atribuída a $\vec{x}$ é a de $\vec{s}_i$ (veja Figura 19).

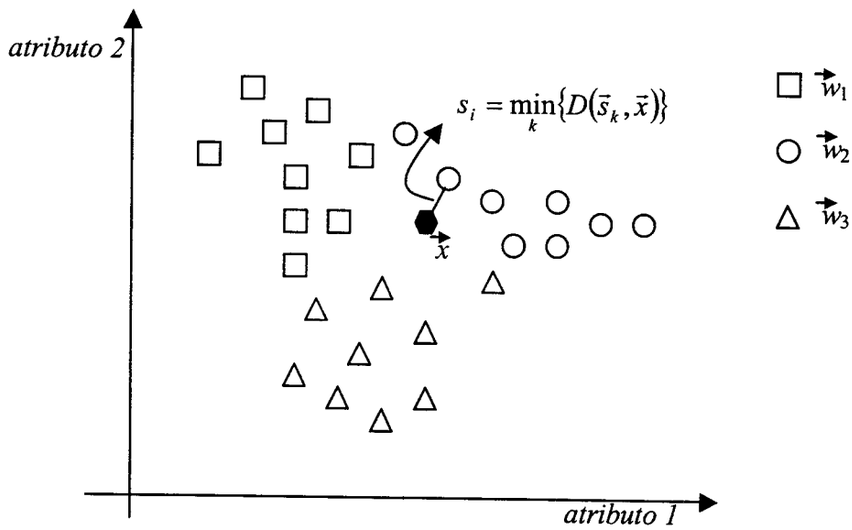


Figura 19 - O critério para que um vizinho seja agrupado.

Os limites inferior e superior da taxa de erro do 1-*NN* (probabilidade média de erro) são conhecidos e baseados na taxa de erro ótima conseguida pela aplicação da regra de Bayes (P^*). Tais limites são dados por (7) considerando c o número de classes.

$$P^* \leq P \leq P^*\left(2 - \frac{c}{c-1}P^*\right) \tag{7}$$

Conhecer tais limites é uma característica interessante para um classificador, pois não é necessário formular hipóteses sobre os dados para conhecer o comportamento da aplicação do algoritmo. Contudo, sua convergência é lenta, sua taxa de erro não converge necessariamente monotonicamente conforme aumenta o número das amostras e muitas vezes é preciso formular hipóteses sobre os dados para melhorar esse desempenho.

As vantagens em aplicar esta regra para classificação é que sua aplicação não depende de treinamento. A métrica adotada para encontrar a mínima distância é o único parâmetro que deve ser fornecido para aplicar este algoritmo. Contudo a métrica euclideana é a mais utilizada e seu resultado corresponde à tesselação de

Voronoi na qual cada célula é identificada por um rótulo de uma classe[DUDA et al., 1998]. Essa dependência de uma métrica faz com que o desempenho computacional do algoritmo 1-*NN* esteja ligado aos tipos de operações necessárias para calcular a distância.

3.3.2.2 k-NN

O algoritmo "*k*-vizinhos mais próximos" é uma extensão do 1-NN. No *k*-NN, contudo, a decisão é tomada em função da classe a que pertence os k-vizinhos do objeto a ser classificado. Assim, se uma maioria dos k-vizinhos pertence a uma classe *w*, o objeto caracterizado pelo vetor *x* pertencerá também à classe *w*. É possível que aconteça situações de empate e para que evitar ou minimizar tais situações adota-se *k* ímpar para o número de classes *c* pares e vice-versa.

Os limite superior de taxas médias das probabilidades de erros (*P*) com relação à realizada por uma regra de Bayes (*P**) é calculado em função da combinação entre os *k* vizinhos e o número de objetos possíveis com erro de classificação. Eles têm limites superiores balizados pela combinação entre os possíveis *k* e os *i* possíveis erros de classificação (8).

$$C_k(P^*) = \binom{k}{i}\left[(P^*)^{i+1}(1-P^*)^{k+i}\right] + \left[(P^*)^{k-i}(1-P^*)^{1+i}\right] \tag{8}$$

Assim, os limites são dados por (9).

$$P^* \leq P \leq C_k(P^*) \tag{9}$$

A Figura 20 ilustra uma comparação entre a variação da taxa *P* de um algoritmo *k*-NN e a taxa de erro *P** resultante da aplicação da regra de Bayes. Percebe-se que conforme *k* aumenta, a taxa de erro *P* se aproxima da taxa *P** e portanto a convergência para classificações corretas torna-se mais rápida.

O problema que resta é a complexidade computacional do algoritmo. Os métodos existentes baseiam-se em três abordagens: (1) tornar o cálculo da distância mais rápido selecionando as dimensões que podem fornecer uma resposta mais rápida, pressupondo um conhecimento do espaço em que os objetos estão sendo classificados; (2) dispor os objetos em uma estrutura de dados (árvores por exemplo) que possa facilitar a busca dos vizinhos mais próximos; (3) editar o conjunto de objetos de modo a eliminar aqueles que não forem significativos, ou seja, aqueles que

apresentem vizinhanças de mesma classe [BHATTACHARYA et al., 1981]. Isso exige um algoritmo de *pruning*, e para isso abordagens da geometria computacional como o diagrama de Voronoi são muito utilizadas.

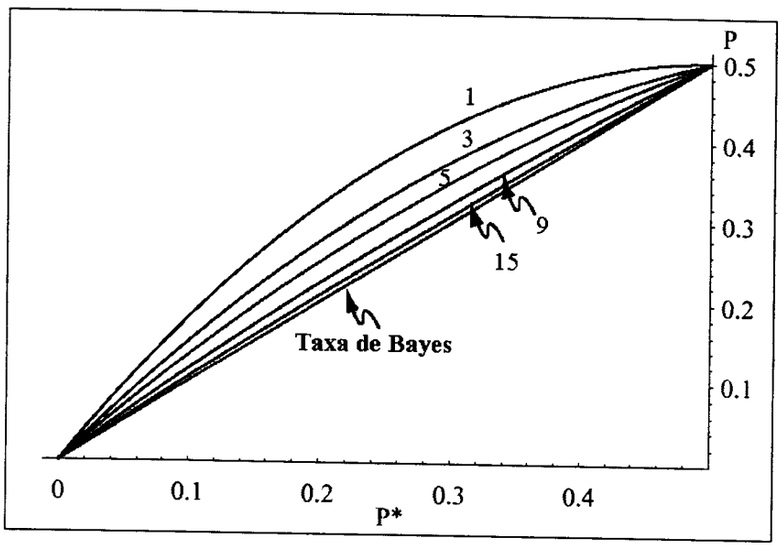


Figura 20 – Comparação dos limites superiores de erro para diferentes *k* [DUDA et al., 1998].

3.3.3 Agrupamentos Particionais (Clustering)

3.3.3.1 K-médias

O algoritmo de *k*-médias é talvez o mais intuitivo para realizar agrupamentos. O objetivo é minimizar a soma das distâncias quadráticas dos elementos de um grupo ao seu centro. Para realizá-lo, são necessários quatro passos principais:

Algoritmo 1 – *k-médias*

Passo 1	Escolher k centros iniciais de grupos que podem ser arbitrariamente escolhidos do conjunto de objetos $s = \{\vec{s}_1, \cdots, \vec{s}_n\}$ a serem classificados (n é o número de objetos). Deste modo, determina-se um conjunto de centros cuja quantidade k é fornecida como parâmetro. Seja então o conjunto inicial de centros, vetores $z(0) = \{\vec{z}_1^0, \cdots, \vec{z}_k^0\}$ (os elementos superescritos indicam a iteração);

Passo 2	Distribuir os objetos de s entre os k grupos C_m^i (m=1..k; i é a iteração corrente) assumindo o critério $\vec{s} \in C_m^i$ se $\| \vec{s} - \vec{z}_m^i \| < \| \vec{s} - \vec{z}_n^i \|$ para m,n =1..k;
Passo 3	Como os centros foram arbitrariamente determinados, calcula-se os novos centros como as médias de cada grupo. Assim, o conjunto $z(i)$ assumirá novos valores dados por $$\vec{z}_m^i = \frac{1}{N} \sum_{\vec{s} \in C_m^i} \vec{s}$$
Passo 4	Testar a convergência da nova classificação com classificação anterior, isto é, se $z(i) = z(i-1)$. Se a igualdade não ocorrer, o passo 2 deverá ser repetido até que ocorra ou então, até que um determinado número máximo de iterações seja alcançado.

Embora a convergência seja garantida (mas não para extremos absolutos) [SELIM & ISMAIL, 1984], ela pode ser lenta ou convergir apenas para um mínimo local. Esta velocidade depende fortemente da escolha inicial dos centros e da ordem na qual são apresentados os objetos ao classificador. O algoritmo acima descrito é interpretado geometricamente por [DUDA et al., 1998] como uma busca estocástica pelos centros das células de um diagrama de Voronoi.

3.3.3.2 K-médias nebuloso

O algoritmo de k-médias resulta uma classificação com limites que estabelecem se um determinado elemento pertence ou não a uma determinada classe. No entanto, para algumas aplicações é desejável que tais limites sejam substituídos por graus de pertinência. O grau de pertinência do objeto r a um grupo j é dado por (10).

$$P(C_j \mid \vec{s}_r) = \frac{(1/d_{rj})^{1/(b-1)}}{\sum_{a=1}^{k} (1/d_{ra})^{1/(b-1)}} \text{ no qual } d_{rj} = \left\| \vec{s}_r - \vec{z}_j \right\| \qquad (10)$$

O parâmetro b ajusta a mistura entre os grupos. Se b for 0, por exemplo, não haverá misturas e o algoritmo é um k-médias convencional. No entanto, conforme b aumenta a fronteira entre os grupos fica mais nebulosa.

O grau de pertinência pode ser representado por uma probabilidade condicional. Sendo assim, ela deve ser normalizada indicando que ela deve pertencer ao conjunto de classes estipulados, ou seja (11).

$$\sum_{j=1}^{k} P(C_j \mid \vec{s}_r) = 1 \quad para \ r = 1, \cdots, n \tag{11}$$

Assim, na inicialização, o grau de pertinência de cada objeto a cada "*cluster*" deve ser determinado. Esta função de pertinência indicará uma probabilidade de um objeto pertencer a um determinado grupo. A cada iteração do algoritmo *k*-médias nebuloso, os valores da função se alteram, pois os centros *z* também se alterarão conforme (12).

$$z_j^i = \frac{\sum_{r=1}^{n} \left[P(C_j \mid \vec{s}_r)\right]^b \vec{s}_r}{\sum_{r=1}^{n} \left[P(C_j \mid \vec{s}_r)\right]^b} \tag{12}$$

3.3.3.3 ISODATA

ISODATA é um acrônimo para *Iterative Self-Organizing Data Analysis Techniques*. É baseado no *k*-médias, no entanto, dispõe de parâmetros para um ajuste heurístico com o objetivo de acelerar a velocidade de convergência do *k*-médias e para determinar qual o número de classes desejado. Além do número de grupos desejados *k*, mais cinco parâmetros são necessários:

Algoritmo 2 - ISODATA

T_N	é um limiar para o número de amostras em um grupo: se o grupo tiver menos que T_N amostras ele é descartado e o número de grupos decrementado;
T_σ	é um limiar de desvio padrão. O grupo com maior desvio padrão é dividido em dois se o desvio padrão for maior que T_σ;
T_C	é um limiar de distâncias entre centros para ajuntamento. As distâncias dos centros dos grupos são calculadas. Se um limiar dessas distâncias for menor do que T_C, os grupos serão agrupados dois a dois segundo a ordem crescente de distância e até que T_L grupos tenham sido agrupados. A cada junção o número *k* é decrementado;
T_L	é um limiar para o número de grupos que podem ser agrupados.
I	é o número máximo de iterações. Serve como um limite para o caso da convergência não ocorrer;

É fácil perceber que ajustar os parâmetros acima descritos é um processo empírico tornando a intervenção de um operador imprescindível. Sua convergência não é garantida [ANDERBERG, 1973]. Seu desempenho computacional é o pior dos métodos particionais (nos quais se inclui o *k*-médias). Seu uso pode ser adequado quando um problema de classificação apresenta peculiaridades ou não-linearidades que a solução deve refletir.

3.3.3.4 LEADER

Quando o objetivo é agrupar objetos cujo espaço de atributos não é contínuo ($\mathbf{R}^n$ por exemplo), o algoritmo de *k*-médias não faz mais sentido já que nele as médias não poderão ser discretas. Para isso, foram desenvolvidos algoritmos que tentam imitar o comportamento do *k*-médias, para adaptá-lo a espaços de atributos discretos usando um objeto que represente um grupo. Um destes algoritmos é o Leader, desenvolvido por [HARTIGAN, 1975]. O Leader exige apenas um parâmetro: um limiar de similaridade δ e uma função de similaridade f_S normalizada entre os objetos. O algoritmo é dado abaixo:

Algoritmo 3 - Leader

Passo 1	Separar os objetos em grupos e selecionar um objeto líder, tal que cada objeto em um grupo esteja a uma distância menor que 1-δ do líder. Denotemos z o conjunto de líderes de cada grupo;
Passo 2	Para cada novo objeto a ser classificado, é verificado se sua similaridade com cada líder é maior que o limiar δ, isto é, $f_S(o, z_j) > \delta$ (j=1..número de grupos) . Se a similaridade for maior que δ, o objeto é agregado ao grupo, senão é criado um novo grupo cujo líder é o novo objeto.

O algoritmo acima tem a vantagem de reter pouca informação (apenas os líderes) para compará-los com os objetos. Essa característica propicia o uso do método em grandes bancos de dados. No entanto, o limiar de similaridade δ influi sobre a quantidade de grupos formados, pois quando δ cresce o número de novos líderes tende a crescer e vice-versa, influenciando a representatividade de cada líder. Outro aspecto negativo do método é a extrema dependência da ordem com que os líderes são criados, já que eles não se alteram conforme o número de objetos

classificados aumenta. Portanto, conforme cresce a quantidade de objetos os líderes podem se tornar menos representativos [REINARTZ, 1999].

3.3.3.5 PAM, CLARA e CLARANS

Algoritmos como o Leader têm problemas com a representatividade do objeto que representa um grupo. Alguns algoritmos foram desenvolvidos para contornar esse problema, tais como PAM (Partitioning Around Medoid) e CLARA [KAUFMAN & ROUSSEEUW, 1990]. O objetivo destes algoritmos é agrupar os objetos em torno de um protótipo ou como chamado pelos autores dos métodos, em torno de um medóide. O algoritmo PAM inicia-se com uma seleção arbitrária de k objetos como medóides. É calculado o custo T de se trocar cada medóide pelo novo objeto. Se esse custo for negativo, o medóide é trocado pelo objeto, senão o novo objeto é agregado ao grupo cujo medóide é o mais similar. A chave desse algoritmo está no custo T, pois para calculá-lo deve-se encontrar a distância (ou dissimilaridade) entre o novo objeto e cada objeto já agrupado além de analisar quatro situações diferentes. Segundo os próprios autores do método, a complexidade de tempo do algoritmo é $O(k(n-k)^2)$, no qual k é o número de grupos iniciais e n é o número de objetos a serem classificados. Essa complexidade de tempo inviabiliza a aplicação do algoritmo PAM (e de outros baseados em medóides) para grandes volumes de dados, apesar de sua robustez quanto a objetos muito distantes (ou dissimilares). Além disso, aspectos como invariância à translações, rotações e ordem de apresentação dos objetos são propriedades deste tipo de método e bastante desejáveis em algoritmos de agrupamentos.

Motivados pelo bom desempenho do algoritmo PAM, mas buscando uma maneira de torná-lo mais rápido, foi desenvolvido o algoritmo CLARA (Clustering LARge Applications). O algoritmo CLARA é descrito a seguir.

Algoritmo 4 - CLARA

Passo 1	Amostrar $40+2k$ objetos aleatoriamente e aplicar o algoritmo PAM para estes objetos amostrados, encontrando assim k medóides. O número de objetos amostrados foi empiricamente determinado para maximizar o desempenho do algoritmo PAM;
Passo 2	Determinar qual dos k medóides é mais similiar a cada elemento de todo o conjunto de objetos, formando k grupos;

Passo 3	Calcular a dissimilaridade média de cada grupo e encontrar a mínima dissimilaridade média. Se esse valor mínimo for menor que o da atual iteração, usa-o como o corrente; mantém os k medóides atuais para a próxima iteração e vá para o *Passo 4*. Senão os k medóides serão determinados novamente pelo algoritmo PAM, repetindo, portanto o *Passo 1* e iniciando uma nova iteração;
Passo 4	Amostrar $40+k$ objetos aleatoriamente e juntar a esses os k medóides mantidos no *Passo 3*. Para iterar novamente, ir para o *Passo 2*.

O algoritmo CLARA melhora significativamente o desempenho de tempo em relação ao PAM, e a acurácia pode ser controlada por uma medida de qualidade: a mínima dissimilaridade média. Assim o número de iterações permitiria obter mais ou menos acurácia, conforme o desejado. No entanto, busca-se um algoritmo que desempenhe tão bem ou melhor que o PAM, mas que reduza a complexidade de tempo do PAM e CLARA.

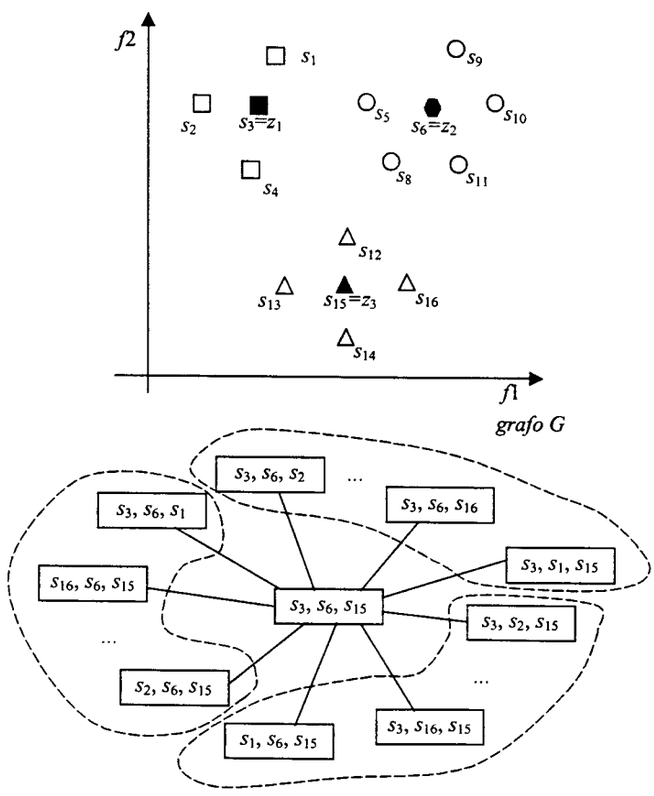


Figura 21 – Agrupamentos pelo algoritmo CLARA considerando agrupamentos aleatórios.

CLARANS foi proposto por [NG & HAN, 1994] com esse objetivo. Um conjunto de k medóides candidatos é visto como um nodo de um grafo G. Os arcos desse grafo representam uma troca de um objeto por outro no conjunto de medóides candidatos. O algoritmo PAM percorre todo o grafo com arcos rotulados com os custos de cada troca para minimizar estes custos. Sendo assim, o problema pode ser visto como uma busca pelo caminho de menor custo. CLARANS é uma alternativa para que essa busca se dê em um número menor de candidatos a medóides. CLARA já faz isso aleatorizando os próximos medóides candidatos. No entanto, essa escolha aleatória se dá sobre um conjunto de nodos sem se preocupar com a vizinhança do grafo G ilustrado na Figura 21.

Já o algoritmo CLARANS se preocupa com o custo da troca para definir os próximos candidatos escolhendo apenas os conjuntos vizinhos dos candidatos atuais para a próxima iteração. O algoritmo é dado pelos seguintes passos:

Algoritmo 5 - CLARANS

Passo 1	Inicializar parâmetros: número de mínimos locais nl e o número máximo de vizinhos nmv. Inicializar índice de mínimos locais i=1 e o custo mínimo cm com um número grande;
Passo 2	Selecione um nodo arbitrário r do grafo G (na verdade um conjunto de k medóides candidatos);
Passo 3	Inicialize o índice de vizinhanças j=1;
Passo 4	Considere um nodo s vizinho de r tomado aleatoriamente de todos os possíveis vizinhos, isto é, um conjunto s tal que $\{s \cap r\}$ possua k–1 elementos. Calcular o custo da troca do medóide candidato;
Passo 5	Se s for o vizinho de custo mais baixo, vá para o *Passo 3*. Senão incrementar j de 1. Se $j \leq nmv$, vá para o *Passo 4*.
Passo 6	Se $j > nmv$, então comparar o custo de r com o parâmetro cm. Se o custo de r for menor que cm, então selecionar r como o melhor nodo e adotar cm como o custo de se adotar r.
Passo 7	Incrementar i de 1; e se $i > nl$, selecionar r como o melhor nodo, ou seja, o melhor conjunto de medóides. Senão voltar para o *Passo 2*.

Dois parâmetros são necessários para iniciar o algoritmo: o número de mínimos locais e o número máximo de vizinhos. O primeiro é um limite para as buscas cujos custos mínimos não sejam menores que o custo mínimo na iteração

atual, dentro do número de vizinhos máximos a serem testados, estes mínimos são chamados "locais" e o parâmetro indica o número de vezes que se irá tentar localizar novos mínimos "locais". O número máximo de vizinhos é o limite da quantidade de vizinhos a qual se permitirá ao algoritmo testar. Conforme esse número aumenta mais o CLARANS se aproxima do algoritmo PAM quanto às trocas testadas. A qualidade dos agrupamentos será maior, e então os mínimos locais surgirão em menor número.

3.3.3.6 BIRCH e CURE

Embora destinados a grande quantidade de dados, os métodos Leader, PAM, CLARA e CLARANS não refletem uma preocupação de como esses dados estão armazenados. Bancos de dados exigem um gerenciamento do fluxo de dados entre a memória secundária e a memória principal [SILBERSCHATZ et al., 1999]. Como um banco de dados inteiro tipicamente não cabe totalmente na memória principal, é necessário que haja um esquema para que as regiões de memória (páginas) mantenham os blocos de dados e suas respectivas localizações. Para isso, aplicam-se diversas estruturas de dados como tabelas *hash*, árvores-B ou B+. Tais estruturas carregam informações em diversos níveis, caracterizando uma hierarquia. O algoritmo BIRCH (*Balanced Interactive Reducing and Clustering using Hierarchy*) [ZHANG et al., 1996] realiza uma fase de pré-agrupamento. Nesta fase, regiões densas são representadas por triplas (N, $\vec{LS}$, SS) chamadas *Clustering Feature* (CF) (Agrupamento de Atributos). N é o número de objetos naquele grupo, $\vec{LS}$ é a soma vetorial de todos os vetores de atributos do grupo e SS é a soma dos quadrados de todos os vetores de atributos do grupo. Esta tripla, segundo os autores, podem ser somadas, ou seja, os CFs de dois grupos podem ser somados vetorialmente e resultado é o CF dos dois grupos agregados. Durante a fase de pré-agrupamento, todo o conjunto de objetos é percorrido e estas representações CFs são armazenadas em uma estrutura dinâmica de árvore chamada *árvore* CF. Esta estrutura se assemelha à árvore-B+ [ELMASRI & NAVATHE, 1994], contudo os nodos folha, ao invés de dados contém descrições CFs de subgrupos. Para cada objeto sucessivo, a árvore-CF é percorrida para encontrar o grupo mais próximo (ou similar). Se um limiar de distância (ou dissimilaridade) for atingido, o objeto inicia um novo grupo na árvore-

CF. Em uma fase final de rotulação, os centróides dos grupos são utilizados como sementes e cada objeto é atribuído a um grupo cuja semente estiver mais próxima (ou for mais similar).

O papel que a árvore-CF desempenha é na verdade um arquivo de índices, para os objetos armazenados em um meio cujo acesso é mais lento. Este índice permite este acesso seja mais rápido quando for usado o critério de agrupamento.

3.3.4 Agrupamento Hierárquico

Para algumas aplicações é conveniente que a análise permita uma visualização hierárquica dos objetos. Duas estratégias são aplicadas para obter uma visão hierárquica: a aglomerativa e a divisiva.

Na estratégia aglomerativa, os objetos são isolados e a partir da distância ou similaridade entre eles, vão sendo agrupados. É conveniente esclarecer que tanto medidas do tipo distância como medidas do tipo similaridade podem ser utilizadas. No entanto, como distância e similaridade crescem em direções opostas, é necessário definir qual tipo será utilizado antes de definir o método. Para os exemplos utilizados nesse trabalho, serão sempre tratadas medidas do tipo distância.

De uma maneira bastante simples, uma matriz de distâncias de n linhas por n colunas contendo as distâncias (ou dissimilaridades) par a par é construída. A partir dela é derivado um diagrama (diagrama de Venn, dendrograma) na qual a visualização hierárquica é possível. Os dendrogramas são as estruturas mais utilizadas para essa visualização pois permitem uma rápida avaliação da diferença de distâncias entre os grupos.

Na estratégia divisiva, todos os objetos são inicialmente considerados como pertencentes a um único grupo. As decisões de divisões são tomadas partindo-se de um grafo dos objetos totalmente conectado. Os nodos correspondem aos objetos e os arcos às distâncias. A medida de distância juntamente com uma decisão (um limiar, por exemplo) de divisão devem ser definidas para que haja a separação dos grupos.

3.3.4.1 Agrupamentos hierárquicos aglomerativos

Os agrupamentos hierárquicos aglomerativos são realizados a partir das medidas de distância entre os objetos. Os algoritmos para realizar os agrupamentos partem de uma matriz de distâncias $D_{n\times n}=d[i, j]$, na qual cada linha e cada coluna

corresponde a um objeto. Nem todos os métodos partem de uma matriz de distânica ou similaridade.

Hubert [HUBERT, 1974] utiliza uma função de distância que é calculada a cada iteração do algoritmo. O aspecto positivo nessa abordagem é que não há necessidade da construção de uma matriz, requerendo pouca memória para elaborar a estrutura. contudo o cálculo de distância entre os grupos deve ser realizado a cada iteração, acarretando uma sobrecarga computacional talvez indesejável.

O algoritmo de Johnson [JOHNSON, 1967] realiza alterações na matriz D para manter atualizações das distâncias entre os grupos ao longo das iterações. A aplicação de diferentes critérios (ligação simples e ligação completa) torna estas matrizes modificadas também diferentes.

A matriz de distância $D_{(n-1)\times(n-1)}=d[i,j]$ é construída de tal maneira que cada uma das linhas correspondam aos n-1 primeiros grupos e as colunas aos n-1 últimos grupos. Note que essa nova matriz D elimina a diagonal principal nula. O algoritmo de Johnson para agrupamento hierárquico pressupõe a existência da matriz D e é descrito a seguir:

Algoritmo 6 – Algoritmo de Johnson

Passo 1	Iniciar com um conjunto de grupos cujo conteúdo é dado por um único objeto. Portanto, $C_i = \{o_i\}$, com i=1..n. Iniciar o índice de agrupamentos $k = 0$ e um indicador dos níveis de cada agrupamento $L[0] = 0$.
Passo 2	Encontre pelo menos um par $\{C_r, C_s\}$ no conjunto de grupos existentes tal que a distância entre eles seja a mínima, isto é, $$d[r,s] = \min_{i=1;j=1}^{i=n-1;j=n-1} \{d[i,j]\}$$
Passo 3	Incrementar o número de grupos k=k+1. Juntar os grupos C_r e C_s em um único grupo e indicar seu nível atual, i.e., $L[k]$ = d[r, s];
Passo 4	Atualizar a matriz D removendo as linhas e colunas correspondentes aos grupos C_r e C_s e adicionando as linhas e colunas correspondentes ao novo grupo formado. O critério de ligação (simples, completa, média) deve ser definido por uma regra cujas características e propriedades são descritas adiante.
Passo 5	Se todos os objetos estiverem em um único grupo, então pare. Senão, continue a execução a partir do *Passo 2*.

As aglomerações ou ligações são decididas segundo regras que influem na qualidade do agrupamento. A qualidade de um agrupamento está intimamente ligada aos objetivos de uma aplicação. Tais objetivos podem exigir uma regra mais ou menos rígida, cujas principais representantes são descritas nos próximos ítens. Um exemplo ilustra cada um destes principais métodos, mostrando os dendrogramas obtidos de objetos aleatórios medidos por dois atributos. A representação dos objetos no espaço de atributos é mostrada na Figura 22.

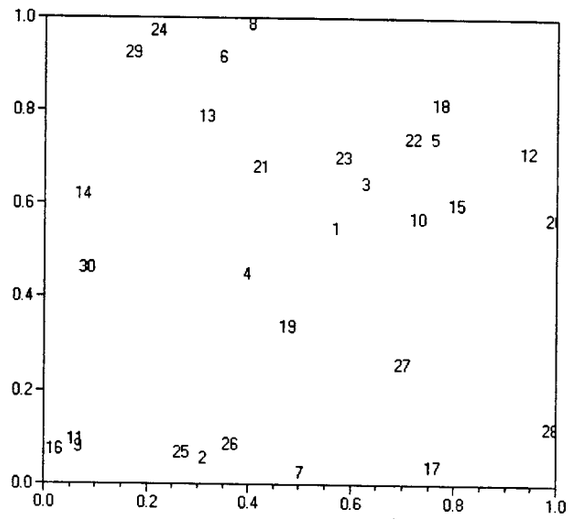


Figura 22 – Exemplo que será submetido a agrupamentos. São amostras identificadas por números em um espaço formado por duas medidas normalizadas entre 0 e 1.

Ligação simples. A ligação simples forma os grupos a partir da seguinte regra no momento da junção:

> "Um grupo é agregado a um outro grupo se PELO MENOS UM dos membros do grupo existente for similar ao do grupo a ser agregado".

No Algoritmo 6, esta regra se traduz no critério de que a distância adotada entre o novo grupo (rs) e os grupos existentes (c) será a mínima, isto é, a distância entre dois grupos é dada pela distância entre os elementos mais próximos (13).

$$d[c,(rs)] = \min\{d[c,r], d[c,s]\} \tag{13}$$

O resultado da aplicação dessa regra é uma variedade de grupos formados por um encadeamento desequilibrado. Os agrupamentos têm o objetivo de identificar os tipos de estruturas segmentadas segundo dois atributos. Percebe-se na Figura 23*a*

a tendência ao desequilíbrio na ligação simples. No entanto, um aspecto vantajoso desse método é a invariância às transformações monotônicas sobre as medidas de distância. Esta propriedade permite variar as medidas (desde que preservada a ordem) sem afetar o resultado do agrupamento. Das principais regras, a única que possui essa propriedade é a ligação simples. Um outro aspecto importante é que os empates não acarretam decisões por outros critérios, ou seja, o método suporta distâncias iguais. No dendrograma essa propriedade é representada pela possibilidade de ocorrer um agrupamento envolvendo mais que dois grupos. A qualidade dos agrupamentos é bastante comprometida pelo desequilíbrio, desde os níveis mais altos até os mais baixos. Na Figura 23*b*, este problema pode ser ilustrado por casos (entre outros) como o agrupamento dos objetos 14 e 30, cuja hierarquia é a mesma do grupo contendo todos os outros objetos.

(*a*) (*b*)

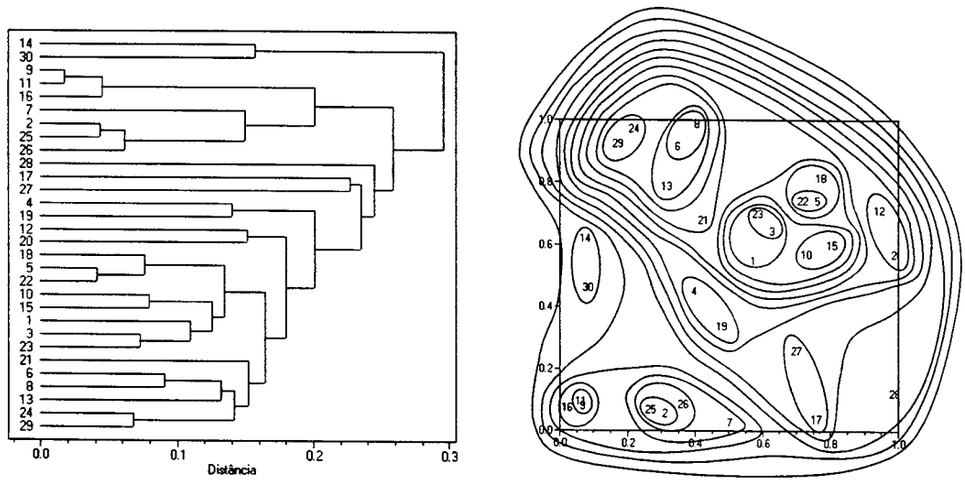


Figura 23 – Resultado de um agrupamento hierárquico dos objetos ilustrados na Figura 22 com ligação simples visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn

Ligação Completa. Esta regra é mais rígida que a regra da ligação simples. A regra para inclusão é a seguinte:

> "Um grupo é agregado a um outro grupo se TODOS os elementos do grupo existente forem similares aos do grupo a ser agregado".

A nova distância (14) na matriz D será dada pela maior distância entre os grupos.

$$d[c,(rs)] = \max\{d[c,r], d[c,s]\} \tag{14}$$

Uma vez aplicada a ligação completa, os grupos tornam-se mais bem definidos em relação à ligação simples. Contudo as diferenças de similaridades entre os grupos não são fiéis à realidade. Isto dificultaria uma aplicação do método na qual o grau de pertinência de um membro a um determinado grupo fosse uma informação relevante. A Figura 24*b* mostra o resultado da aplicação do método de ligação completa ao exemplo com dois atributos.

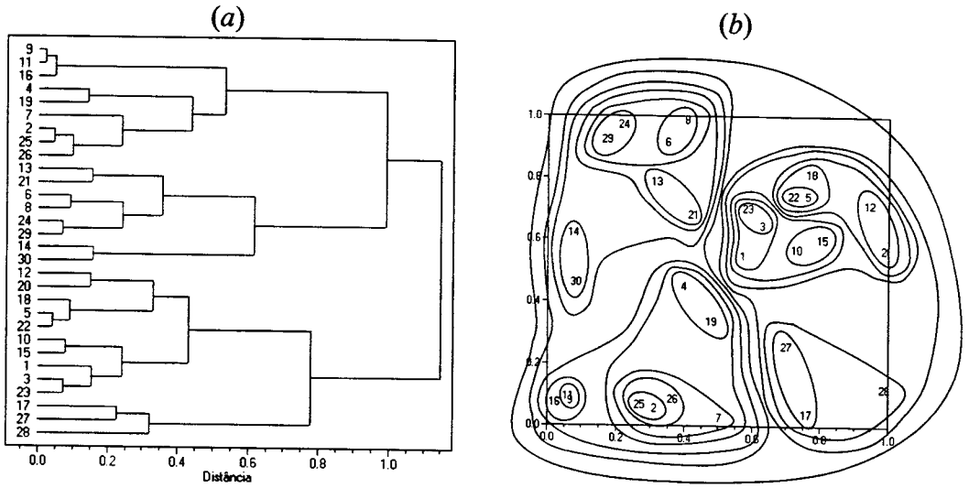


Figura 24 - Resultado de um agrupamento hierárquico dos objetos ilustrados na Figura 22 com ligação completa visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn

Ligações Generalizadas. Com o objetivo de buscar um caminho intermediário entre a ligação simples, que gera um agrupamento desequilibrado, mas com propriedades interessantes; e a ligação completa que produz um agrupamento que pode não corresponder à realidade, Sokal e Michener [SOKAL & MICHENER, 1958] desenvolveram o agrupamento por ligação média. Desde então, diversas variantes foram propostas, sendo a mais utilizada a que utiliza a média aritmética das similaridades [ALDENDERFER & BLASHFIELD, 1984].

Na ligação média as distâncias entre dois grupos r e s com n_r e n_s objetos respectivamente são calculadas tomando a média das distâncias entre cada par de objetos o_r e o_s, cada um de um grupo (15).

$$d[r,s] = \frac{\sum_r \sum_s d[o_r, o_s]}{n_r n_s} \qquad (15)$$

Os grupos são agregados de acordo com a menor distância média entre todos. Para realizar o algoritmo, no entanto, os elementos da matriz D são substituídos segundo a regra em (16).

$$d[c,(rs)] = \frac{n_r}{n_r + n_s} d[c,r] + \frac{n_s}{n_r + n_s} d[c,s] \qquad (16)$$

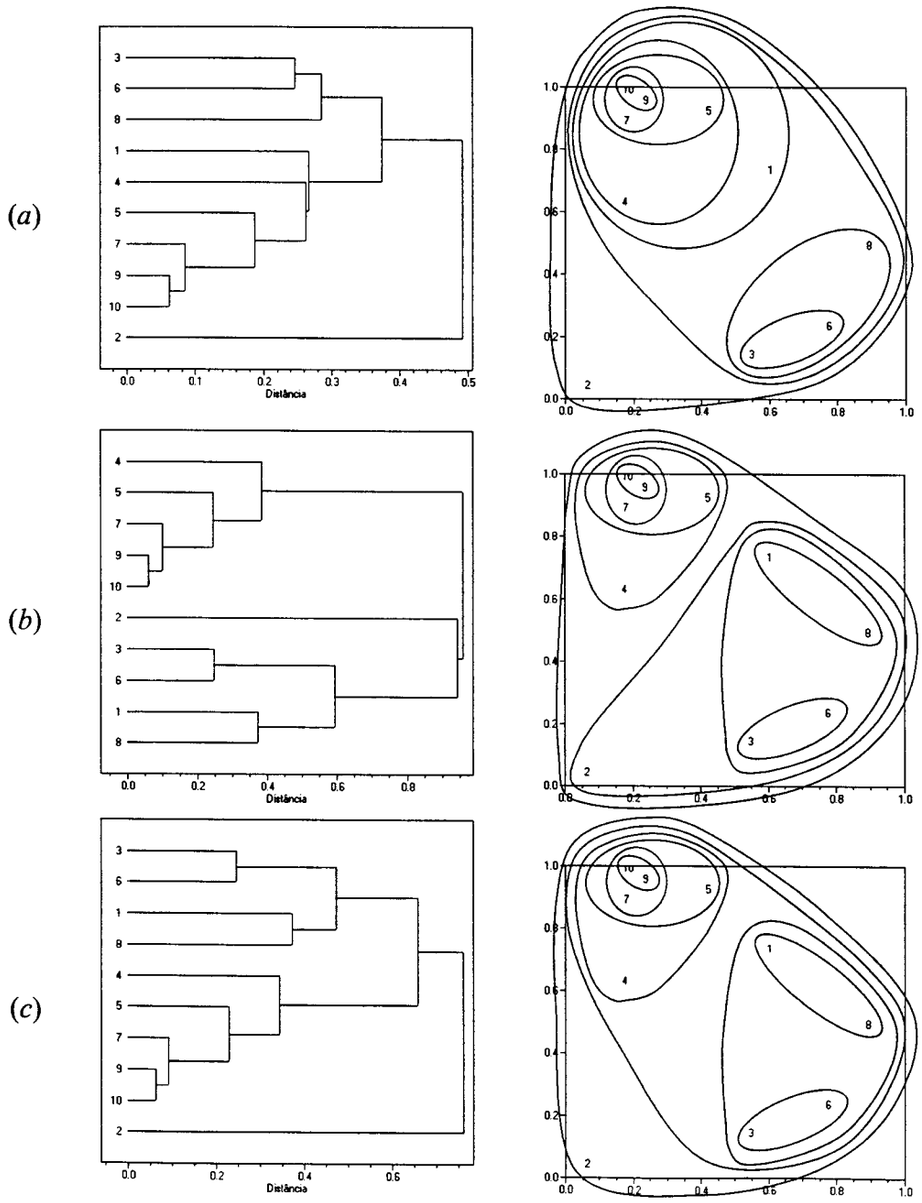


Figura 25 – Comparações entre aplicações de agrupamentos hierárquicos com diferentes ligações (*a*) simples; (*b*) completa; (*c*) média.

A aplicação da regra da ligação média pode alterar ou não as hierarquias das ligações simples ou complexas. Na verdade, a ligação média define uma estrutura

hierárquica intermediária entre as ligações simples e completa. A Figura 25 ilustra este fato. Um conjunto de dez objetos aleatoriamente distribuídos sobre um espaço de dois atributos é submetido as três aplicações: simples, completa e média. Percebe-se no exemplo, principalmente no diagrama de Venn, a transitoriedade entre os resultados da aplicação dos métodos. Nas ligações simples e média, o objeto 2 é agrupado na distância que define a maior hierarquia. Na ligação completa isso não acontece, contudo os agrupamentos de menor hierarquia são mais similares aos da ligação média.

A regra da ligação média acima é definida pela média aritmética das distâncias. Contudo, a média aritmética não é a única opção. Geometricamente o centróide representa melhor o grupo. Para o centróide a regra de substituição é dada por (17).

$$d[c,(rs)] = \frac{n_r}{n_r + n_s} d[c,r] + \frac{n_s}{n_r + n_s} d[c,s] - \frac{n_r n_s}{(n_r + n_s)^2} d[r,s] \qquad (17)$$

Na Figura 26 observa-se a diferença entre métodos que utilizam a média aritmética (*a*) e o centróide (*b*).

Os métodos baseados em centróides apresentam a propriedade de não monotonicidade, ou seja, as junções podem fazer com que as distâncias entre grupos não permaneçam sempre crescentes ao longo das iterações. Isso acarreta o que é chamado de *crossovers* (cruzamentos) nos dendrogramas. Portanto, se uma aplicação exigir monotonicidade das distâncias, o método do centróide não deve ser aplicado.

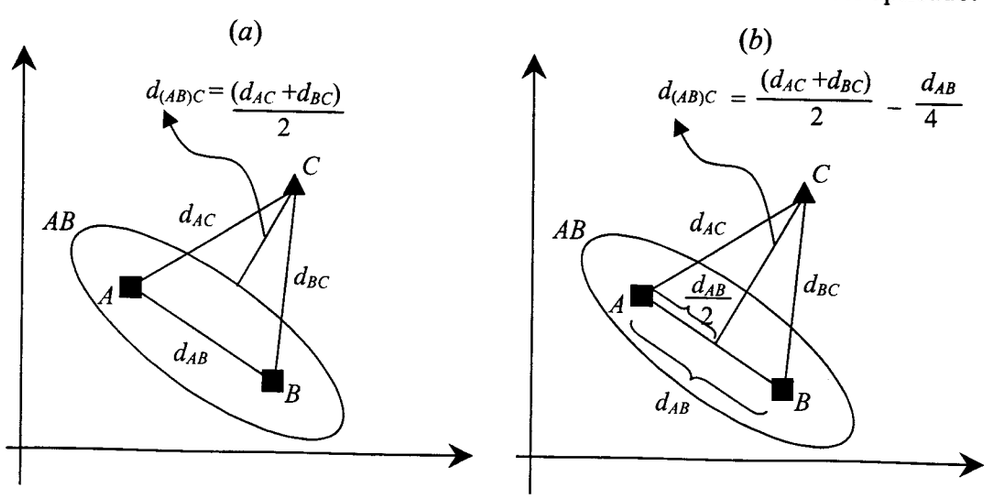


Figura 26 – Diferença das abordagens para distância por (*a*) média aritmética; e (*b*) centróide.

De qualquer maneira, tanto a média aritmética quanto o centróide fazem parte de uma generalização destas hierarquias intermediárias. Esta generalização foi proposta por Lance e Williams [LANCE & WILLIAMS, 1967] e pode ser resumida pela fórmula abaixo (18) para substituição na matriz *D*.

$$d[c,(rs)] = \alpha_r d[c,r] + \alpha_s d[c,s] + \beta d[r,s] + \gamma \left| d[c,r] - d[c,s] \right| \qquad (18)$$

Os parâmetros nessa substituição, ou seja, α_r, α_s, β e γ definem uma grande variedade de métodos. A notação *c* indica os grupos existentes; (*rs*) é o novo grupo formado pelos grupos *r* e *s*. A distância entre cada grupo existente *c* e o novo grupo *rs* é denotada por *d*[*c*, (*rs*)]. A Tabela 1 exibe este conjunto de parâmetros para os métodos mais conhecidos [JAIN & DUBES, 1988]. O parâmetro γ serve para inserir os métodos de ligação simples e completa nessa generalização. Os métodos de ligação média e centróide ponderados (WPGMA e WPGMC) não consideram o tamanho do grupo, ou seja, não incluem o número de objetos daquele grupo, para calcular a nova distância. Com isso, o método penaliza os grupos com poucos objetos. Portanto, aplicações em que os *outliers* são indesejáveis os métodos ponderados são os mais adequados.

Tabela 1 – Índices para a fórmula genérica para a regra de decisão em (18) [JAIN & DUBES, 1988].

Método de Agrupamento	α_r	α_s	β	γ
ligação simples	½	½	0	–½
ligação completa	½	½	0	½
média (UPGMA)	$\frac{n_r}{n_r + n_s}$	$\frac{n_s}{n_r + n_s}$	0	0
média ponderada (WPGMA)	½	½	0	0
centróide (UPGMC)	$\frac{n_r}{n_r + n_s}$	$\frac{n_s}{n_r + n_s}$	$\frac{-n_r n_s}{(n_r + n_s)^2}$	0
centróide ponderado (WPGMC)	½	½	–¼	0
Ward (mínima variância)	$\frac{n_r + n_c}{n_r + n_s + n_c}$	$\frac{n_s + n_c}{n_r + n_s + n_c}$	$\frac{-n_c}{n_r + n_s + n_c}$	0

Método de Ward. O método de Ward, também chamado método da mínima variância é definido sobre padrões estabelecidos. Existe um objetivo a ser perseguido, e este objetivo é a minimização do erro quadrático entre as possíveis

configurações de agrupamento para que os agrupamentos estejam ajustados aos padrões desejados.

Um objeto é definido não apenas em função de seu grupo, mas também em função de seu padrão. O erro quadrático de cada objeto é calculado tomando a diferença quadrática entre o objeto e o centróide do grupo. Assim, o erro quadrático do grupo todo define um valor dado pela soma de todos os erros quadráticos de cada objeto.

Os erros dependem somente dos centróides e em um agrupamento os erros quadráticos permanecem os mesmos para todos os grupos, exceto para os que estão sendo agrupados. Portanto, pode ser utilizada uma regra para substituição dos elementos na matriz D que permite o uso do Algoritmo de Johnson (Algoritmo 6) para realizar o método de *Ward*.

$$d[c,(rs)] = \frac{n_r + n_c}{n_r + n_s + n_c} d[c,r] + \frac{n_s + n_c}{n_r + n_s + n_c} d[c,s] + \frac{-n_c}{n_r + n_s + n_c} d[r,s] \qquad (19)$$

O método pode ainda ser generalizado por (19), que permite a substituição dos parâmetros pelos valores dados na Tabela 1.

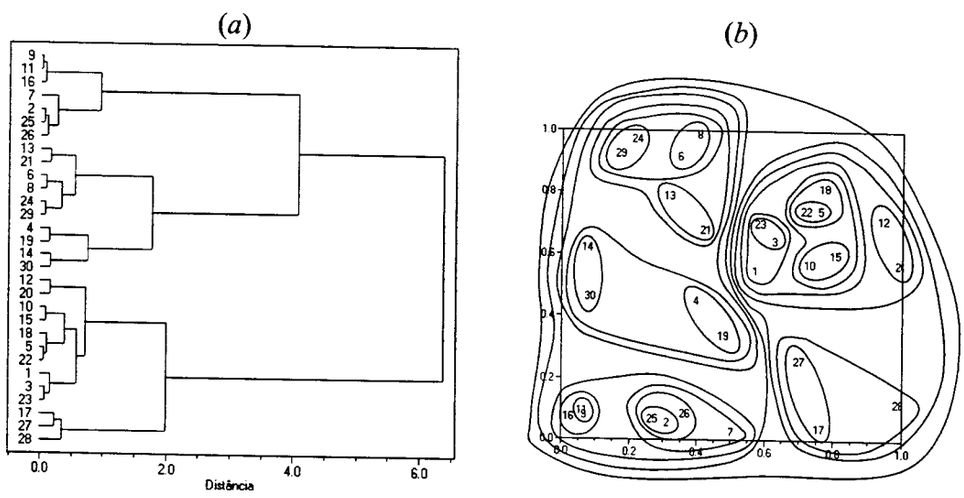


Figura 27 - Resultado de um agrupamento hierárquico pelo método Ward dos objetos ilustrados na Figura 22 e visualizados por um (*a*) dendrograma e por um (*b*) diagrama de Venn

Jain e Dubes [JAIN & DUBES, 1988] fizeram um pequeno apanhado dos trabalhos avaliando métodos de agrupamento. Dentre todos, o método *Ward* é o melhor avaliado, mas o autor ressalta que é difícil definir critérios para avaliar

algoritmos de agrupamentos. Ele ainda alerta que nem todas as aplicações caracterizam-se por ter objetos separáveis por critérios de erros quadráticos.

Para o exemplo apresentado nas ligações simples e completa, a Figura 27 ilustra uma separação bastante elegante, mas também diferente das anteriores em todos os níveis hierárquicos.

3.3.5 Mapeamento Auto-organizável

Os métodos de agrupamento hierárquicos ou particionais são baseados em regras para se tomar a decisão de junção que nem sempre consideram a forma de uma vizinhança topológica. Além disso, há a necessidade de se manter uma grande quantidade de informação, sejam os objetos originais, seja na forma de matrizes de distância. Os resultados dos agrupamentos servem também para reduzir a dimensionalidade dos dados. Muitas aplicações exigem que o resultado do agrupamento preserve a vizinhança topológica dos objetos, como a segmentação de imagens baseada em cores [MOREIRA & COSTA, 1996]. A redução de dimensionalidade é também um aspecto desejável para aplicações que lidam com percepção [OBERMAYER et al., 1992].

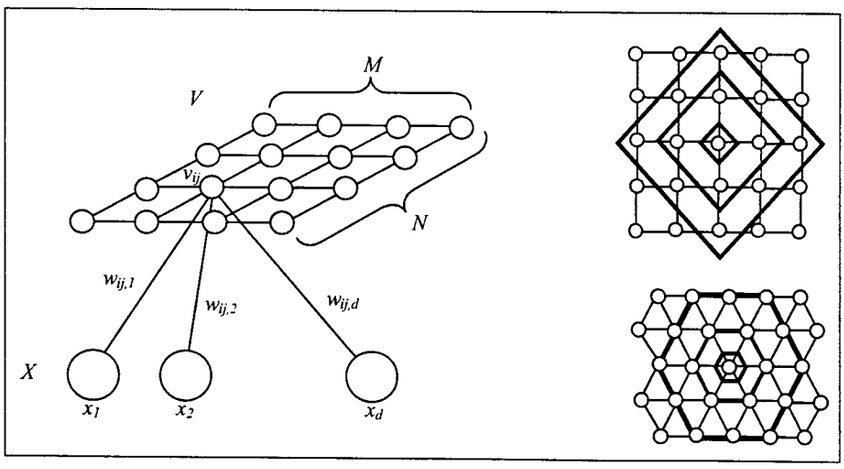


Figura 28 – Estrutura de dados para um SOM

Kohonen [KOHONEN, 1990] introduziu o mapeamento auto-organizável (SOM), cujos resultados apresentam as características desejáveis acima. A estrutura de dados do SOM consiste de um conjunto $V = \{v_{ij}\}, \quad i = 1..M; j = 1..N$, de $n=MN$ nodos dispostos em uma grade (comumente uni ou bi-dimensional) ligados a um outro conjunto de d nodos de entrada $X = \{x_1, x_2, \cdots, x_d\}$, no qual cada nodo

representa um atributo do espaço em que os objetos a serem agrupados estão representados. Cada nodo $\vec{v}_{ij}$ é ligado à entrada por um vetor protótipo $\vec{w}_{ij}$. Embora a Figura 28, ilustrando a estrutura de dados apresente uma grade retangular, ela pode ser disposta em outro tipo de topologia, como a hexagonal, por exemplo.

O SOM precisa ser treinado para generalizar um conjunto de entradas. A apresentação dos dados causará (ou não) a formação de regiões na grade *V*. Estes padrões formados na grade *V* são os grupos, e portanto a grade *V* é a saída do algoritmo. O treinamento é um processo iterativo no qual, a cada passo *t*, um objeto $\vec{x}(t)$ de um conjunto (conjunto de treinamento) é escolhido aleatoriamente. As distâncias entre $\vec{x}(t)$ e todos os vetores protótipo $\vec{w}_{ij}$ são calculadas e o vetor com menor distância $\vec{w}^*$ é encontrado.

$$\left\|\vec{x}(t) - \vec{w}^*(t)\right\| = \min_{i,j}\left\{\left\|\vec{x}(t) - \vec{w}_{ij}(t)\right\|\right\} \tag{20}$$

Este vetor com menor distância e seus vizinhos topológicos serão, no próximo passo, atualizados de acordo com uma regra. Esta regra inclui uma função de vizinhança $z^*_{ij}(t)$ e uma taxa de aprendizado $\alpha(t)$, que tornará o aprendizado mais ou menos rápido. Ambas, variando a cada passo e de acordo com a aplicação.

$$\vec{w}_{ij}(t+1) = \vec{w}_{ij}(t) + \alpha(t) z^*_{ij}(t)\left[\vec{x}(t) - \vec{w}_{ij}(t)\right] \tag{21}$$

Esta é uma fonte de críticas ao algoritmo, já que há grande liberdade para se "desenhar" soluções.

A função de vizinhança é normalmente uma função de suporte compacto, em um intervalo (0, 1] e cuja posição do valor máximo deve coincidir com a posição do melhor protótipo na grade *V*. Isso permite uma interação lateral na grade que será reduzida conforme aumenta o número de passos *t*. Vesanto [VESANTO, 2000] sugere uma gaussiana como função de vizinhança.

$$z^*_{ij}(t) = e^{-\frac{\left\|r^* - r_{ij}\right\|^2}{2\sigma^2(t)}} \tag{22}$$

Na função sugerida a redução da interação lateral se dá pela variação do desvio padrão da gaussiana $\sigma(t)$ ao longo do treinamento. As posições na grade *V* são representadas por r^* e r_{ij}.

Ao final de um número de iterações previamente determinado, cada nodo de V indicará a sua distância na grade, ao nodo protótipo. Alguns nodos estarão bastante próximos enquanto outros não. A visualização dessa grade permite distinguir as regiões (ou grupos), mas para usar o SOM como um agrupamento é necessário utilizar um algoritmo convencional de agrupamento (k-NN por exemplo) sobre o conteúdo dos nodos da grade V, para que a cada região seja identificada.

Devido a sua capacidade de redução de dimensionalidade, o SOM é aplicado em visualização de dados em espaços de muitas dimensões. Esta visualização pode ser conseguida por várias abordagens, tais como (a) matriz de distância unificada [ULTSCH & SIEMON, 1990]; (b) mapeamento de Sammon [SAMMON JR, 1969]; (c) histogramas das entradas por grupos, entre outras.

As aplicações do SOM são numerosas. Vesanto em [VESANTO, 2000] enumera diversas aplicações em cada fase de um processo de descoberta do conhecimento: determinação dos objetivos, preparação dos dados, mineração de dados, construção de modelos, acompanhamento, entre outros.

O SOM é um algoritmo bastante flexível e simples, permitindo um amplo espectro de aplicações. Sua execução dispende pouca memória, pois a apresentação dos objetos se dá por um treinamento, ou seja, o tempo é o outro lado dessa moeda. O treinamento é um processo demorado e seu desempenho depende muito de boas escolhas para a função da vizinhança e da taxa de aprendizado [KOHONEN, 1990]. Por outro lado, o SOM permite que sistemas não lineares, normalmente difíceis de se modelar, sejam aproximados por essa abordagem bastante simples. [VESANTO, 2000].

3.3.6 Algoritmo Genético

Buscar uma ou várias soluções a partir de um conjunto com grande quantidade de elementos ou cujo espaço de busca tem muitas dimensões é um desafio muito freqüente em mineração de dados. Algoritmos de busca são abundantes na literatura. Cada um deles trata alguma limitação ou peculiaridade do espaço de busca ou ainda do conjunto de dados. O algoritmo genético em particular, é bastante eficiente para espaços de busca de alta dimensionalidade e com quantidade de

elementos que tornariam inviáveis a aplicação de algoritmos como descida de gradiente ou *simulated annealing* [KIRKPATRICK et al., 1983].

A abordagem mais simples e também menos eficiente para busca é a "força bruta". A força bruta deve percorrer todo o espaço de busca para encontrar a solução do problema. É claro que, dependendo do espaço de busca, o uso da força bruta é completamente inviável. É o caso para um espaço de busca contínuo, por exemplo R^2, no qual um mínimo ou máximo para uma função $f(x, y)$ deve ser encontrada. As soluções (ou solução) poderiam ser analiticamente encontradas, contudo para uma grande parte dos problemas de busca, sequer se conhece a função, quanto mais suas propriedades.

Em outros casos, o número de alternativas é inviável para qualquer algoritmo de busca. A solução é reduzir a quantidade de consultas a estas alternativas. Porém, para isso é necessária uma estratégia.

Holland inspirou-se na teoria da evolução de Darwin para propor uma estratégia de busca. Embora seja uma aproximação grosseira do que realmente ocorre em termos macroscópicos em organismos muito simples (haplóides) [BÄCK, 1996]. Holland [HOLLAND, 1975] introduziu a noção de *schemata* para formalizar e estabelecer um limite inferior de seu desempenho [MITCHELL, 1996].

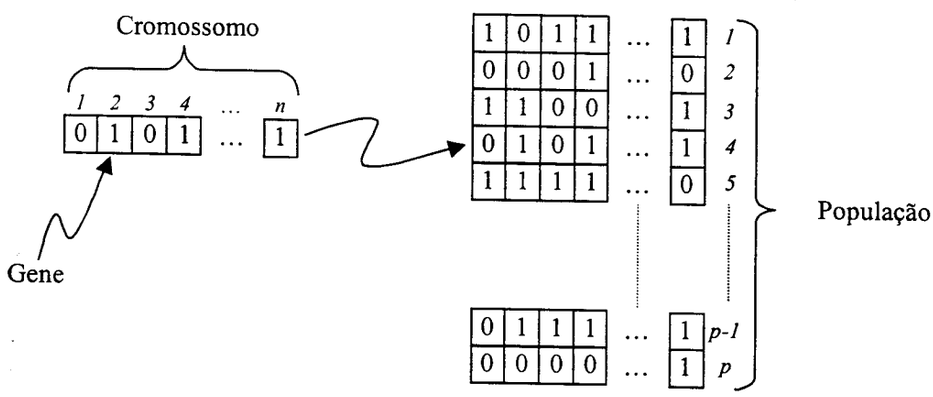


Figura 29 – Estrutura de dados para um algoritmo genético com genes binários.

O algoritmo genético ou GA deve ser implementado considerando o problema que ele irá resolver. Porém, as estruturas fundamentais e os operadores são mantidos. Estas estruturas é que são inspiradas na abordagem evolucionista. Assim, o cromossomo representa o indivíduo em uma população. O cromossomo é constituído de uma cadeia de genes que codificam um estado binário (0 ou 1) . Esta cadeia de

genes pode codificar um conjunto maior de estados. Por exemplo, um cromossomo com n genes binários codifica 2^n estados. Contudo, apenas uma população de p indivíduos será submetida ao processo evolucionário (Figura 29).

A característica do problema implica em uma função de adaptabilidade $f(i)$ aplicada a cada indivíduo, cujo resultado indica quão bem adaptado ele está. Se o problema for encontrar os máximos de uma função $h(x)$, o indivíduo deve codificar o espaço a ser buscado, ou seja, o domínio $D(f(x))$ da função. Como os genes são binários, os indivíduos apenas amostrarão o domínio D com 2^n pontos. A densidade com que o domínio é amostrado pode causar problemas, principalmente se ela for muito baixa, pois nesse exemplo, pontos de máximo poderão não ser amostrados. A função $f(i)$ é chamada de *função fitness* ou *função objetivo*, e determiná-la adequadamente é fundamental para a solução do problema. Por exemplo, na busca de máximos de $h(x)$ (Figura 30*a*), o valor de $f(i)$ deve acompanhar o valor da função. No entanto, se o que se busca for zeros de função (Figura 30*b*), $f(i)$ poderá assumir o valor $f(i) = -|h(x)|$ e assim os indivíduos mais bem adaptados serão aqueles que cruzarem o eixo (Figura 30*c*).

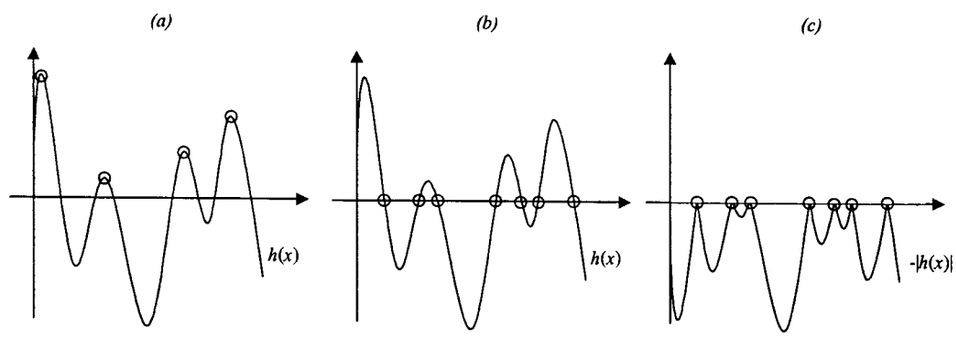


Figura 30 – Exemplo de espaço de busca e funções objetivo (*a*) máximos (*b*) e (*c*) cruzamentos no zero.

Para evoluir, a população P deve se submeter a operadores genéticos. Estes operadores são basicamente três: cruzamento, mutação e seleção.

O operador cruzamento opera sobre um par de cromossomos $C_a = [g_1^a \quad g_2^a \quad \cdots \quad g_n^a]$ e $C_b = [g_1^b \quad g_2^b \quad \cdots \quad g_n^b]$ e ocorre com uma taxa de probabilidade p_X. Um ponto de cruzamento $X \in [1, n)$ deve ser selecionado aleatoriamente. Algumas abordagens utilizam o cruzamento multipontos, ou seja, $X = \{x_1, x_2, ..., x_c\}$, é um conjunto de pontos de cruzamento, no qual $c < n$ e $x_k \leq x_{k+1}$.

Quando apenas um ponto de cruzamento existe, o resultado é um outro par de cromossomos:

$$C_a^{t+1} = [g_1^a \quad g_2^a \quad \cdots \quad g_X^a \, g_{X+1}^b \quad \cdots \quad g_n^b] \text{ e}$$

$$C_b^{t+1} = [g_1^b \quad g_2^b \quad \cdots \quad g_X^b \, g_{X+1}^a \quad \cdots \quad g_n^a].$$

Para multipontos:

$$C_a^{t+1} = [g_1^a \quad g_2^a \quad \cdots \quad g_{x_1}^a \, g_{x_1+1}^b \quad \cdots \quad g_{x_2}^b \, g_{x_2+1}^a \quad \cdots \quad g_{x_c}^a \quad g_{x_c+1}^b \cdots \quad g_n^b] \text{ e}$$

$$C_b^{t+1} = [g_1^b \quad g_2^b \quad \cdots \quad g_{x_1}^b \, g_{x_1+1}^a \quad \cdots \quad g_{x_2}^a \, g_{x_2+1}^b \quad \cdots \quad g_{x_c}^b \quad g_{x_c+1}^a \cdots \quad g_n^a]$$

O cruzamento multipontos permite um incremento na variabilidade dos cromossomos, retardando a convergência para valores que não sejam os desejados. Goldberg [GOLDBERG, 1989] ressalva que o cruzamento multipontos é uma degradação controlável do GA na direção de uma escolha totalmente aleatória.

A mutação é uma operação que age sobre um gene selecionado aleatoriamente com uma taxa de probabilidade p_M. Seu efeito é a mudança de estado do gene.

A seleção é uma operação que ocorre sobre toda a população P na iteração (ou geração) t. O resultado de sua aplicação é a identificação dos indivíduos que estão mais bem adaptados para que na próxima geração continuem evoluindo. A escolha dos cromossomos pode ser realizada de diversas maneiras. A ***seleção pela roleta*** [HOLLAND, 1975; GOLDBERG, 1989] é uma prática bastante adotada e consiste do seguinte. A função-objetivo $f(i)$ é calculada para todos os indivíduos, ou seja, para $i = 1,.., p$. A soma de todos os resultados do cálculo de $f(i)$, ou seja (23).

$$F = \sum_{i=1}^{p} f(i) \tag{23}$$

A probabilidade de seleção de cada indivíduo é determinada pela razão $p_k = f(i)/F$. A probabilidade cumulativa é calculada por (24).

$$P_i = \sum_{k=1}^{i} p_k \quad P_0 = 0 \tag{24}$$

A roleta gira e gera um número aleatório r em [0, 1] n vezes. A cada vez que a roleta gira, um cromossomo é selecionado para reproduzir. Assim, o cromossomo i irá reproduzir se $P_{i-1} < r \leq P_i$. Reproduzir significa submeter-se a operadores

genéticos como cruzamento e mutação. Observe que a reprodução deve se dar em pares, e portanto, a cada dois cromossomos selecionados é efetuada a reprodução.

Este tipo de seleção busca reproduzir os melhores indivíduos atribuindo a eles uma maior probabilidade de sobrevivência p_i. Contudo, nem sempre os melhores indivíduos de uma geração são os desejados. A roleta, em geral, causa uma convergência prematura para o melhor indivíduo.

Algumas estratégias de seleção buscam corrigir esse problema, selecionando na roleta indivíduos que estejam uniformemente espaçados. Isso permite a seleção de indivíduos com baixa probabilidade de sobrevivência, mas que pode carregar em seu código características que, se reproduzidas e combinadas com outras características, podem gerar um indivíduo melhor adaptado. Essa estratégia é conhecida como ***amostragem estocástica universal*** (SUS) [BAKER, 1987].

Em cada uma das abordagens acima, contudo, não se garante que as características do melhor indivíduo ou ele mesmo permaneça na população. Abordagens determinísticas, contudo, buscam essa preservação. Dentre estas abordagens estão a seleção por truncamento ou por blocos. Na ***seleção por truncamento*** é definido um limiar de valor da função objetivo para que apenas os indivíduos acima desse limiar reproduzirão. Já para a ***seleção por blocos***, a determinação dos indivíduos a reproduzir é feita por uma porcentagem da população (20% por exemplo), dos quais serão selecionados apenas os melhores indivíduos [THIERENS & GOLDBERG, 1994]. Estratégias de ***seleções elitistas*** [DE JONG, 1975] permitem que o melhor indivíduo sempre permaneça na população, tornando a função objetivo ao longo das iterações sempre crescente. Dentre estas estratégias está a ***seleção por disputas*** [GOLDBERG, 1989], na qual um conjunto de q indivíduos é selecionado aleatoriamente e apenas o melhor deles irá reproduzir. O processo continua até que toda uma nova população de descendentes tenha sido gerada. Este método é bastante utilizado devido a sua simplicidade computacional.

A escala da função objetivo $f(i)$ pode ser alterada considerando suas ocorrências até a iteração corrente. Dentre estas estratégias estão a escala sigma e a seleção de Boltzmann.

Para a ***escala sigma*** [GOLDBERG, 1989] o valor de $f(i)$ é alterado a cada iteração conforme sua média (μ_f^t) e desvio padrão (σ_f^t) na iteração t. A escala é resultante da aplicação de (25) [TANESE, 1989] sobre $f(i)$.

$$s(i) = \begin{cases} 1 + \dfrac{f(i) - \mu_f^t}{2\sigma_f^t} & \text{se } \sigma_f^t \neq 0 \\ 1 & \text{se } \sigma_f^t = 0 \end{cases} \tag{25}$$

Para selecionar um indivíduo, $s(i)$ é utilizada. A intenção é diminuir o que se chama *pressão seletiva*, (o número de reproduções de um único indivíduo) e conseqüentemente evitar a convergência prematura. Assim, no início, quando o desvio padrão de $f(i)$ é alto os indivíduos mais bem adaptados não serão tão privilegiados na seleção. Conforme o desvio padrão diminui, o ponto de seleção entre os melhores e os piores será mais explícito.

A ***seleção de Boltzmann*** [GOLDBERG, 1989] escala também a função objetivo (26), porém um parâmetro (T) controla a velocidade com que a distinção entre os melhores e piores. Este parâmetro é a "temperatura". Ela deve começar grande e ao longo das iterações ir decrescendo conforme um pré-estabelecimento.

$$s(i) = \frac{e^{\frac{f(i)}{T}}}{\mu(i)} \qquad \text{para } \mu(i) = \sum_{k=1}^{t} e^{\frac{f^k(i)}{T}} \tag{26}$$

O parâmetro T permite um controle da pressão seletiva, contudo é um parâmetro que deve ser estabelecido e o suas propriedades teóricas bem estabelecidas [DE LA MAZZA & TIDOR, 1991].

Os GAs evoluem por iterações. Os passos de um GA genérico pode ser visto no Algoritmo 7.

Algoritmo 7 – Algoritmo Genético

Passo 1	Inicializar a população de pais $P(t)$ e a iteração corrente com $t = 0$. Avaliar as funções objetivo $f(i)$ para cada indivíduo da população
Passo 2	Recombinar a população de pais (cruzamento e mutação) $P(t)$, gerando uma população de descendentes $D(t)$
Passo 3	Calcular as funções objetivo $f(i)$ da população $D(t)$
Passo 4	Selecionar a próxima população de pais $P(t+1)$ a partir das duas populações: $P(t)$ e $D(t)$; Incrementar t. $t = t+1$

Se a condição de término foi alcançada, então PARE, senão volte ao *Passo 2*.

A condição de término no Algoritmo 7 pode ser determinada pelo número de iterações ou então por uma condição para convergência a uma solução. A condição de parada mais comum é o número da iteração. Assim, se $t \geq T$ a evolução pára e o indivíduo ou os indivíduos com maiores $f(i)$ são os que indicam a solução. Uma outra condição de parada que pode ser utilizada em abordagens elitistas é o número de gerações em que os indivíduos mais bem adaptados permanecem inalterados.

Condições de parada que consideram a estrutura dos cromossomos (o genótipo) tentam identificar um estado de convergência da população, ou seja, uma proporção de indivíduos da população que tenham a mesma representação (ou similar). Pode ser considerada a evolução do resultado da função objetivo $f(i)$. Se ao longo da evolução $f(i)$ não se altera dentro de um intervalo ε, durante um número T de gerações, o algoritmo pára de evoluir [MICHALEWICZ, 1994].

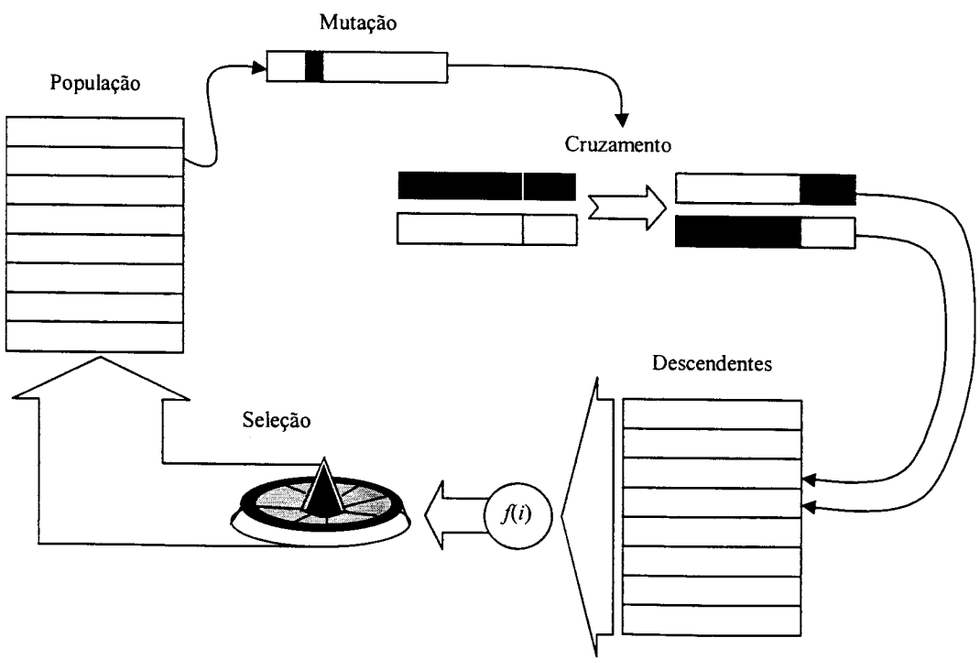


Figura 31 – Explicação sinótica sobre algoritmo genético, suas operações básicas e estruturas de dados.

Até aqui, foi comentado o algoritmo genético com codificação binária, porém há uma grande diversidade de codificações possíveis. Se a representação de

um indivíduo conta com um conjunto discreto de estados (o alfabeto), os operadores genéticos deverão ser diferentes do GA com codificação binária. Por exemplo, a mudança de estados pode não ter apenas duas opções e portanto deve ser definido um critério para uma nova operação de mutação [MITCHELL, 1996].

Uma outra maneira de representar o gene é atribuindo um valor de ponto flutuante a ele. Esse valor pode significar um peso, ou mesmo um valor de função. Nesse caso, a alteração nos operadores é muito grande. Além disso, outros tipos de operadores são adicionados aos convencionais. A ***mutação*** pode ser ***convencional uniforme***, se o gene a ser mutado for substituído por um valor aleatório dentro de um intervalo. Pode também ser ***dinâmica não uniforme***. Nesse caso, se um gene a ser mutado g_k tiver limites inferior $g_k^{\inf}$ e superior $g_k^{\sup}$, o resultado da operação terá dois valores possíveis que dependem da função (27).

$$g_k' = g_k + \Delta(t, g_k^{\sup} - g_k) \quad \text{ou} \quad g_k' = g_k + \Delta(t, g_k - g_k^{\inf}) \tag{27}$$

$$\Delta(t, y) = y \cdot r \cdot \left(1 - \frac{1}{T}\right)^b \tag{28}$$

A função em (28) retorna um valor em $[0, y)$ que aproxima-se de 0 conforme t (o número da geração) cresce. r é um valor aleatório em [0, 1], T é o número máximo de gerações e b é o grau de não-uniformidade. Gen e colaboradores [GEN et al., 1996] propuseram uma mutação baseada na direção da solução, ou seja, se o objetivo for minimizar então o valor do gene mutado deve seguir na direção do gradiente ou em direção contrária se o objetivo for maximizar. Para isso, em uma vizinhança em Δg_k, a direção do gradiente deve ser procurada.

O cruzamento também sofre alterações. Além dos operadores convencionais (um ponto e vários pontos de cruzamento), o ***cruzamento aritmético*** realiza ponderações sobre os cromossomos. Estes cromossomos são os vetores de atributos $\vec{x}$. Os descendentes são então calculados por (29) e restrições aos valores de λ_1 e λ_2 definem a convexidade da combinação: convexa quando $\lambda_1 + \lambda_2 = 1$ e $\lambda_1 > 0$ e $\lambda_2 > 0$; linear quando $\lambda_1 + \lambda_2 \leq 2$ e $\lambda_1 > 0$ e $\lambda_2 > 0$; ou afim quando $\lambda_1 = 1.5$ e $\lambda_2 = -0.5$. Tais combinações restringem a posição da solução do espaço de busca [GEN & CHENG, 1997].

$$\vec{x}_1' = \lambda_1 \vec{x}_1 + \lambda_2 \vec{x}_2$$
$$\vec{x}_2' = \lambda_1 \vec{x}_2 + \lambda_2 \vec{x}_1 \qquad (29)$$

O ***cruzamento baseado na direção*** gera apenas um indivíduo descendente que, baseado na função objetivo *f(i)* de cada pai, garante que o filho terá uma solução melhor que os pais. Isso é conseguido pela seguinte regra: $\vec{x}' = r \cdot (\vec{x}_2 - \vec{x}_1) + \vec{x}_2$ com *r* sendo um número aleatório em [0,1] e assumindo que $f(\vec{x}_2) \geq f(\vec{x}_1)$ quando se quer maximizar e o contrário se o objetivo for minimizar a solução [GEN & CHENG, 1997].

Os GAs têm um amplo potencial de aplicações. Devido a sua simplicidade de implementação os GAs vêm sendo aplicados em diversas fases para descoberta de conhecimento. Contudo seu uso exige um conhecimento do problema, para que o resultado não seja comprometido por erros. Tais erros podem ocorrer desde a codificação, ou seja, na atribuição do significado e intervalo de cada gene, até a composição da função objetivo. Dois casos extremos podem ocorrer: (1) uma função objetivo muito simples que não reflete a meta da aplicação, ou (2) superestimar a capacidade do GA incluindo na função objetivo muito complexa que transforme o espaço de busca e conseqüentemente a busca inviável.

3.3.7 Aproximação

Modelos permitem uma visão focada sobre a realidade. Para construí-los, é preciso abrir mão de diversos parâmetros ou de avaliações sobre o problema. Para dispensar parâmetros ou avaliações é preciso aprender a escolhê-los. O lavrador, quando arrisca um prognóstico sobre o tempo, exercita uma atividade que vem experimentando desde quando nasceu. Deste modo, houve um aprendizado ao longo de sua vida, para que ele pudesse ter uma taxa de acerto boa o suficiente.

O homem do campo, quando arrisca um prognóstico sobre o tempo, considera alguns fatores como a mudança de temperatura e a velocidade do vento sem sensores padronizados. Ele não dispõe de todos os recursos de um meteorologista para fazer a sua previsão e assim, a cada tentativa, ele está generalizando e muitas vezes desconsiderando algumas medidas que o profissional

utiliza. Esta generalização é conseguida por uma aproximação, que dependendo da importância dos parâmetros escolhidos pode ser suficiente para uma boa previsão.

Se o meteorologista quiser estudar essa generalização, ele poderia tomar as medidas de pressão atmosférica e o volume da precipitação ocorrida. A generalização é conseguida por uma regressão ou por uma interpolação dos pontos coletados. A regressão tenta aproximar uma curva com parâmetros conhecidos (reta, exponencial) de um conjunto de pontos buscando minimizar um critério de distância entre os pontos e a curva (veja Figura 32*a* e Figura 32*b*). Já o objetivo da interpolação é encontrar uma curva na qual todos os pontos coletados residam (veja Figura 32*c*). É claro que restrições sobre o conjunto de pontos devem ser impostas, mas o sucesso pode ter um preço alto: as generalizações podem apresentar um erro muito grande.

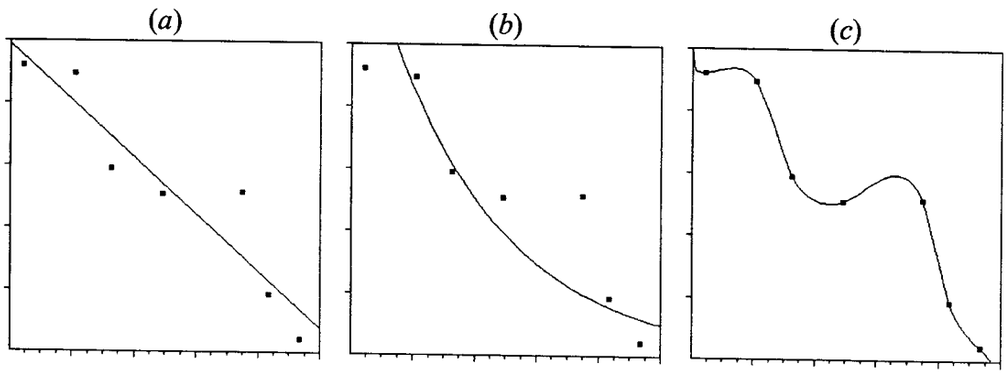


Figura 32 – Exemplos de (a) regressão linear; (b) regressão exponencial; (c) interpolação.

Métodos de regressão e interpolação são abundantes na literatura, contudo cada um deles impõem restrições, dentre as quais: pontos de origem igualmente espaçados, linearidade, grau polinomial [POWELL, 1981].

Um método de interpolação bastante utilizado para treinamentos e generalizações é aplicar funções de base radial (FBR) introduzidas por [POWELL, 1985]. Sua popularidade para aplicações de reconhecimento de padrões e aprendizado de máquina ocorre porque não é necessário um espaçamento regular entre os pontos de origem, viabilizando o uso de casos reais em treinamentos. Sua aplicação em espaços multidimensionais é facilitada por exigir poucos pontos para sua aplicação.

A interpolação utilizando FBRs é realizada considerando dois tipos de informação: a conhecida e a que se quer conhecer. A informação conhecida é

representada por um conjunto de vetores $V = \{\vec{v}_0, \vec{v}_1, \cdots, \vec{v}_N\}$ que indica uma posição no espaço $\mathbf{R}^d$ de d dimensões. A cada um dos vetores V corresponde um valor que é resultado da aplicação de uma função sobre cada um destes vetores $F_V = \{f(\vec{v}_0), f(\vec{v}_1), \cdots, f(\vec{v}_N)\}$ associado. A informação que se quer conhecer pode ser correspondente a apenas um ponto, ou então a um conjunto deles. Podemos representar as posições dos valores a serem determinados também como vetores $x = \{\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_M\}$ em $\mathbf{R}^d$. Como cada elemento de V e de X são vetores em um espaço d-dimensional, suas representações são: $\vec{v}_i = \{v_{i1}, v_{i2}, \cdots, v_{id}\}$ e $\vec{x}_k = \{x_{k1}, x_{k2}, \cdots, x_{kd}\}$. A cada vetor de V corresponde um valor de F_V e a cada vetor de X um valor escalar de F_X. Na Figura 33*a*, com $d = 2$, as cruzes correspondem aos locais V e as bolas aos valores conhecidos F_V. A superfície da Figura 33*b* corresponde aos pontos interpolados da Figura 33*a*, ou seja, à função interpolada cujos pontos estão em F_X. A cruz é um desses pontos de X e sua localização (bola indicada pela seta) na superfície é um ponto F_X.

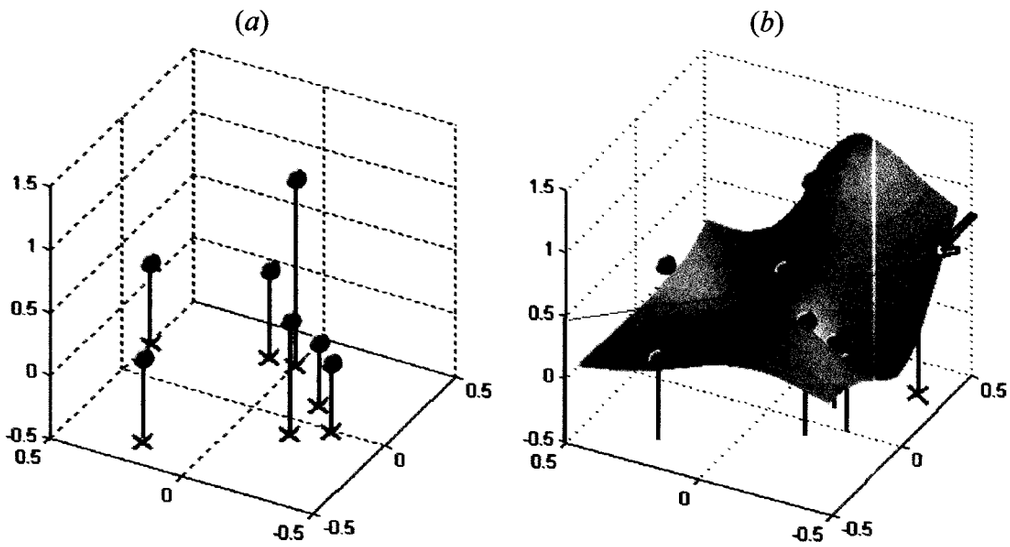


Figura 33 – Ilustração de interpolação com FBR por uma superfície. (*a*) Os pontos a serem interpolados são representados por bolas vermelhas e suas respectivas posições por cruzes azuis. (*b*) A superfície interpolando os pontos e a generalização de um ponto (bola azul).

A interpolação por FBRs aproxima ou interpola um conjunto de pontos aplicando uma base de funções escalares $\phi(r)$, no qual r é uma norma sobre o espaço dos pontos conhecidos. $\phi(r)$ é escolhida considerando um critério, como por exemplo

energia de dobramento da superfície ou linearidade. Uma vez escolhida a função base, cada valor de F_V, é aproximado por (30).

$$f(\vec{v}) = \sum_{i=1}^{N} w_i \phi(\|\vec{v} - \vec{v}_i\|) \tag{30}$$

Os coeficientes w_i indicam a intensidade com a qual a função $\phi(r)$ contribuirá sobre cada relação de distância r entre os pontos $\vec{v}$ e $\vec{v}_i$. A aproximação em (30) pode ser escrita na forma matricial como (31).

$$\begin{bmatrix} f(\vec{v}_1) \\ f(\vec{v}_2) \\ \vdots \\ f(\vec{v}_N) \end{bmatrix} = \begin{bmatrix} 0 & \phi(\|\vec{v}_1 - \vec{v}_2\|) & \cdots & \phi(\|\vec{v}_1 - \vec{v}_N\|) \\ \phi(\|\vec{v}_1 - \vec{v}_2\|) & 0 & \cdots & \phi(\|\vec{v}_2 - \vec{v}_N\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\|\vec{v}_1 - \vec{v}_N\|) & \phi(\|\vec{v}_2 - \vec{v}_N\|) & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \text{ ou} \tag{31}$$

$$F_V = \Phi \cdot w$$

A matriz Φ é simétrica, $N \times N$ e sua diagonal principal é nula. Para uma grande conjunto de $\phi(r)$, Φ é não-singular. Sendo assim, os coeficientes w_i podem ser calculados por (32).

$$w = \Phi^{-1} \cdot F_V \tag{32}$$

Obtidos os coeficientes de w, eles podem ser aplicados a um outro conjunto de pontos para generalizar seus valores de função f_X, como em (33).

$$F_X = w \cdot \Phi(\|\vec{x} - \vec{v}\|) \tag{33}$$

Na forma matricial de (34), podemos verificar a flexibilidade no número de pontos sobre o qual se pode aplicar. Assim, a matriz Φ tem $M \times N$ elementos. O desempenho computacional depende diretamente da função $\phi(r)$ e do número M de pontos que se deseja conhecer.

$$\begin{bmatrix} f(\vec{x}_1) \\ f(\vec{x}_2) \\ \vdots \\ f(\vec{x}_M) \end{bmatrix} = \begin{bmatrix} \phi(\|\vec{x}_1 - \vec{v}_1\|) & \phi(\|\vec{x}_1 - \vec{v}_2\|) & \cdots & \phi(\|\vec{x}_1 - \vec{v}_N\|) \\ \phi(\|\vec{x}_2 - \vec{v}_1\|) & \phi(\|\vec{x}_2 - \vec{v}_2\|) & \cdots & \phi(\|\vec{x}_2 - \vec{v}_N\|) \\ \vdots & \vdots & & \vdots \\ \phi(\|\vec{x}_M - \vec{v}_1\|) & \phi(\|\vec{x}_M - \vec{v}_2\|) & \cdots & \phi(\|\vec{x}_M - \vec{v}_N\|) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \tag{34}$$

Embora as FBRs definam interpolações exatas para um grande variedade de problemas e assumindo $\phi(r)$ particulares, Mees [MEES et al., 1992] afirma que funções lineares por partes são pobremente interpoladas e para contornar o problema

ele define as FBRs afins. As FBRs afins incluem graus de liberdade para rotação de $\phi(r)$ em cada dimensão do espaço em que os vetores de V e de X residem. Além da rotação, uma translação no mesmo espaço é permitida. A Figura 34, mostra o esboço de um suporte elíptico de uma função radial e os graus de liberdade para duas dimensões. Na Figura 34*a*, o suporte pode se inclinar para cada uma das duas dimensões. Já na Figura 34*b*, o suporte foi transladado. Essa liberdade permite uma superfície interpolada menos sujeita a erros.

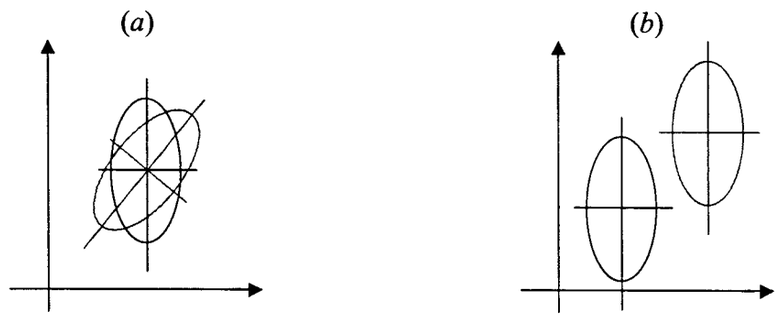


Figura 34 – Exemplos de graus de liberdade de FBRs com suporte elíptico: (*a*) rotação; (*b*) translação.

Para implementar as FBRs afins, (30) deve ser modificado de modo a adicionar os graus de liberdade descritos (35).

$$f(\vec{v}) = \sum_{i=1}^{N} w_i \phi(\|\vec{v} - \vec{v}_i\|) + \sum_{k=1}^{d} a_k v_k + b \tag{35}$$

Restrições devem ser impostas sobre os coeficientes w_i para que interpolações de funções lineares por partes sejam possíveis. Elas foram deduzidas por Powell em [POWELL, 1985] e garantem ainda a unicidade da solução. Tais condições são dadas por (36).

$$\sum_{i=1}^{N} v_i w_i = 0 \quad \text{e} \quad \sum_{i=1}^{N} w_i = 0 \tag{36}$$

Deste modo, na forma matricial, (35) toma a forma de (37).

$$\begin{bmatrix} f(\vec{v}_1) \\ f(\vec{v}_2) \\ \vdots \\ f(\vec{v}_N) \\ \hline 0 \\ 0 \\ \hdashline 0 \end{bmatrix} = \left[\begin{array}{cccc|cc:c} 0 & \phi(\|\vec{v}_1-\vec{v}_2\|) & \cdots & \phi(\|\vec{v}_1-\vec{v}_N\|) & v_{11} & v_{12} & 1 \\ \phi(\|\vec{v}_1-\vec{v}_2\|) & 0 & \cdots & \phi(\|\vec{v}_2-\vec{v}_N\|) & v_{21} & v_{22} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \phi(\|\vec{v}_1-\vec{v}_N\|) & \phi(\|\vec{v}_2-\vec{v}_N\|) & \cdots & 0 & v_{N1} & v_{N2} & 1 \\ \hline v_{11} & v_{21} & \cdots & v_{N1} & 0 & 0 & 0 \\ v_{12} & v_{22} & \cdots & v_{N2} & 0 & 0 & 0 \\ \hdashline 1 & 1 & \cdots & 1 & 0 & 0 & 0 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \hline a_1 \\ a_2 \\ \hdashline b \end{bmatrix} \tag{37}$$

E de uma maneira mais simplificada como (38).

$$F_V = \left[\begin{array}{c|c:c} \Phi & V & 1 \\ \hline V^T & 0 & 0 \\ \hdashline 1^T & 0^T & 0 \end{array}\right] \begin{bmatrix} w \\ \hline a \\ \hdashline b \end{bmatrix} \tag{38}$$

Sua aplicação para pontos não conhecidos é obtida encontrando a solução obtida em (37) aos pontos não conhecidos, como mostra (39).

$$\begin{bmatrix} f(\vec{x}_1) \\ f(\vec{x}_2) \\ \vdots \\ f(\vec{x}_M) \end{bmatrix} = \left[\begin{array}{cccc|cc:c} \phi(\|\vec{x}_1-\vec{v}_1\|) & \phi(\|\vec{x}_1-\vec{v}_2\|) & \cdots & \phi(\|\vec{x}_1-\vec{v}_N\|) & x_{11} & x_{12} & 1 \\ \phi(\|\vec{x}_2-\vec{v}_1\|) & \phi(\|\vec{x}_2-\vec{v}_2\|) & \cdots & \phi(\|\vec{x}_2-\vec{v}_N\|) & x_{21} & x_{22} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi(\|\vec{x}_M-\vec{v}_1\|) & \phi(\|\vec{x}_M-\vec{v}_2\|) & \cdots & \phi(\|\vec{x}_M-\vec{v}_N\|) & x_{M1} & x_{M2} & 1 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \hline a_1 \\ a_2 \\ \hdashline b \end{bmatrix} \tag{39}$$

Simplificando também a aplicação temos (40).

$$F_X = \left[\Phi_X \mid X \mid 1\right] \begin{bmatrix} w \\ \hline a \\ \hdashline b \end{bmatrix} \tag{40}$$

O conjunto de pontos obtidos, conforme apresentado acima, reside em $\mathbf{R}^2$. Sua extensão para d dimensões é direta e é apresentada em uma aplicação para análise pictórica na seção 6.3.

4 PROSPECÇÃO DE ATRIBUTOS

Em um sistema de mineração de dados, os objetos ou tuplas ou amostras são representadas por seus atributos. Muitas vezes, contudo, um conjunto de medidas pode não ser adequado para distinguir classes ou agrupamentos de amostras segundo algum critério desejado. Por exemplo, em uma imagem que apresente frutas coloridas (Figura 35*a*), as frutas podem ser bem identificadas utilizando medidas envolvendo cores. No entanto, se for usada a medida de intensidade dos níveis de cinza (Figura 35*b*), o trabalho pode se tornar mais difícil, e então outros atributos deverão ser adotados.

(*a*) (*b*)

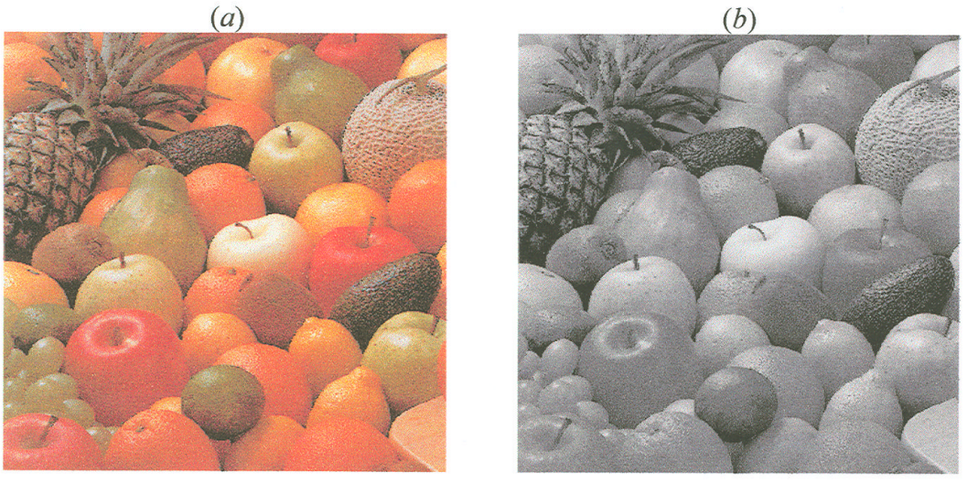


Figura 35 – Uma imagem colorida com frutas identificáveis no espaço de cores (*a*); e uma versão em níveis de cinza (*b*).

A prospecção de atributos envolve o teste e validação de novas medidas. Para que critérios de percepção possam ser considerados em uma avaliação, é desejável que além de distinguir objetos o atributo possa exprimir uma suspeita sobre sua importância na representação do objeto. Deste modo, é importante que atributos de forma considerem as saliências para caracterizá-la. Na verdade, as saliências são

pontos de segmentação de uma curva e definem segmentos que são representativos para a forma [ATTNEAVE, 1954].

Este capítulo apresenta duas novas abordagens para descrever formas. A primeira é baseada na simulação de um campo eletrostático no qual o contorno da forma reside. Cada elemento desse contorno consiste de um elemento de carga. A partir dos caminhos definidos pelas linhas de força saindo de pontos de um círculo em torno da forma é construído um atributo para cada elemento do contorno. Os elementos mais salientes apresentam mais linhas incidentes. Leis de campo alternativas e seus comportamentos são estudados. Além disso, é proposta uma aplicação para modelos de interações entre formas.

Uma outra abordagem utiliza a esqueletonização multi-escalas proposta por Costa e Estrozi [COSTA & ESTROZI, 1999]. Esta abordagem utiliza a métrica euclideana em seu desenvolvimento. O esqueleto resultante da aplicação desse método apresenta boas propriedades, contudo compromete, para vários tipos de aplicação, seu tempo de execução. A seção 4.2 apresenta a abordagem de Costa e Estrozi, e baseando-se nela faz um estudo sobre métricas alternativas que permitem reduzir o tempo de execução. O preço dessa redução é o comprometimento de algumas propriedades desejáveis com relação à métrica euclideana, mas que dependendo da aplicação podem ser ignoradas. Uma métrica em particular (*chessboard*) é apresentada junto com um algoritmo para a extração do esqueleto. Atributos a partir do esqueleto podem ser derivados e são comentados ao final.

4.1 Histograma de linhas de campo elétrico incidentes [COSTA & CONSULARO, 1999]

Como detalhado na seção 5.1, saliência é uma característica importante para a segmentação e descrição de formas. Contudo, para grande parte das abordagens, o caráter dependente da escala das saliências exige que sejam definidos parâmetros para resolver sua localização em um espaço-escala.

Uma alternativa a estas abordagens é considerar a forma residindo em um espaço vetorial regido por leis que caracterizem um tropismo, ou seja, uma força de atração. Uma lei particular imita o comportamento de um campo elétrico, com cada pixel de um contorno correspondendo a uma partícula de carga (positiva ou

negativa) unitária. Um histograma medindo a capacidade de atração de cada carga (ou pixel) nesse campo vetorial distingue as saliências em um contorno.

Para explicar melhor essa abordagem, o cálculo de campos conservativos utilizando a transformada de Fourier é descrito. A seguir, um exemplo sobre formas de neurônios será apresentado e a abordagem será então discutida para outros tipos de forma.

4.1.1 Campos conservativos usando a Transformada de Fourier

Representemos uma imagem bidimensional em termos de uma grade uniforme (x, y) com $1 \leq x, y \leq N$, e a distribuição de cargas elétricas como uma função discreta $c(x, y)$ definida sobre esta grade. Um campo vetorial genérico em qualquer ponto $\vec{p} = (x_p, y_p)$ na grade, chamado $\vec{f}(\vec{p})$, pode ser calculado usando (41), para k e a assumindo valores reais representando, respectivamente, um fator de proporcionalidade e um parâmetro de interação do campo). Uma norma genérica norma-g é definida como $\|\vec{v}\|_g = \left(v_x^g + v_y^g\right)^{1/g}$.

$$\vec{f}(\vec{p}) = k\sum_i c(\vec{v}_i)\|\vec{p} - \vec{v}_i\|_g^{-a} \frac{\vec{p} - \vec{v}_i}{\|\vec{p} - \vec{v}_i\|_g} \tag{41}$$

Sejam f_x e f_y os campos escalares correspondendo aos componentes x e y, respectivamente, do campo vetorial e assumamos que a influência da interação ponto a ponto esteja restrita a um círculo λ de raio m centrado em $\vec{p}$. Assumamos, para efeito de simplicidade, que $k = 1$ e que o campo vetorial esteja definido sobre um espaço cuja norma é a euclideana, ou seja, $g = 2$. Deste modo, os campos escalares f_x e f_y de (41) podem ser reescritos como (42) e (43), para o par (r, s) como um ponto qualquer dentro do círculo λ.

$$f_x(x,y) = \sum_{(r,s)\in\lambda} c(x+r, y+s) \cdot \frac{r}{\left(\sqrt{r^2+s^2}\right)^a} \cdot \tag{42}$$

$$f_y(x,y) = \sum_{(r,s)\in\lambda} c(x+r, y+s) \cdot \frac{s}{\left(\sqrt{r^2+s^2}\right)^a} \cdot \tag{43}$$

Observando f_x e f_y (veja Figura 36) percebe-se que o cálculo dos campos escalares corresponde a uma correlação entre a configuração de cargas $c(x, y)$ e uma função de campo h_x ou h_y definidas em (44) e (45), respectivamente.

$$h_x(x,y) = \frac{x}{\left(\sqrt{x^2+y^2}\right)^a} \tag{44}$$

$$h_y(x,y) = \frac{y}{\left(\sqrt{x^2+y^2}\right)^a} \tag{45}$$

Conseqüentemente, os campos escalares podem ser conseguidos por (46) e (47).

$$f_x = c \otimes h_x \tag{46}$$

$$f_y = c \otimes h_y \tag{47}$$

O teorema da correlação, muito utilizado em processamento de sinais [SCHALKOFF, 1989; CASTLEMAN, 1996], pode então ser utilizado para calcular (42) e (43) no domínio de Fourier. Portanto, sendo C, H_x e H_y as respectivas transformadas de Fourier para c, h_x e h_y, os campos escalares podem ser calculados tomando-se a transformada inversa de Fourier, no qual o asterisco (*) superescrito indica o complexo conjugado.

$$f_x = \Im^{-1}\left\{C \cdot H_x^*\right\} \tag{48}$$

$$f_y = \Im^{-1}\left\{C \cdot H_y^*\right\} \tag{49}$$

Devido à periodicidade intrínseca causada pela transforma de Fourier, é necessário alocar espaço maior para C e as funções de campo. Supondo assim, uma imagem quadrada de tamanho N, o acréscimo deve ser de m, pois a função de campo é definida apenas dentro do círculo λ.

A transformada de Fourier permite uma redução do tempo de execução devido a sua propriedade de reduzir a ordem do número de operações. O esforço computacional despendido para uma correlação aplicando diretamente (42) e (43) alcança a ordem $O(N^4)$. Com a aplicação de (48) e (49), o esforço se reduz a $O(N^2 \log N)$ considerando $m << N$.

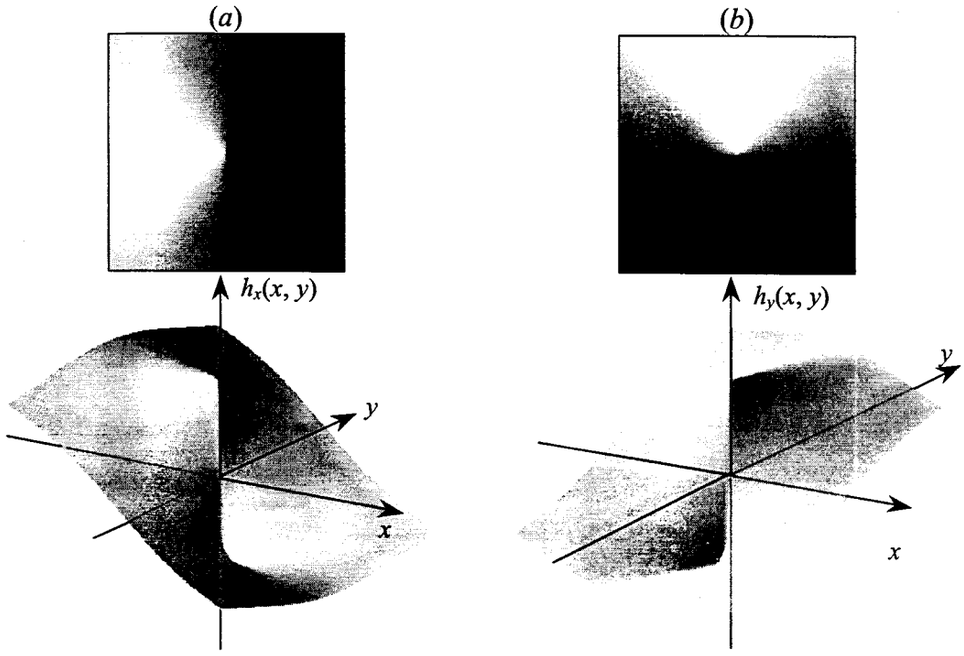


Figura 36 – Funções de campo escalar usadas para calcular as componentes x (a) e y (b).

A partir de f_x e f_y a função de campo escalar pode ser calculada por (50).

$$U = \left|\vec{f}(\vec{p})\right| = \sqrt{f_x^{\,2}(x,y) + f_y^{\,2}(x,y)} \tag{50}$$

O campo vetorial $\vec{f}$ é composto de vetores resultantes da soma vetorial expressa em (51) para $\vec{u}$ e $\vec{v}$ versores ortogonais entre si e orientados nas direções respectivas dos eixos x e y.

$$\vec{f}(\vec{p}) = f_x\vec{u} + f_y\vec{v} \tag{51}$$

4.1.2 Aplicação para interação entre neurônios

No exemplo da Figura 37*a*, duas formas de neurônios configuram cargas positivas para cada ponto. O resultado da função de campo escalar pode ser visto na Figura 37*b*, na qual extremidades de neurônios que não estão próximas apresentam valores altos (mais claros na imagem) e aquelas que estão próximas mostram valores baixos (regiões escuras). Esta abordagem pode servir como modelo de tropismo em crescimento celular, principalmente em neurônios no qual a formação de tecido é influenciada pela existência de cones de crescimento nas extremidades [GOODHILL,

1998]. Esta abordagem foi utilizada em [COSTA et al., 1998], no modelo de redes neurais morfologicamente realistas.

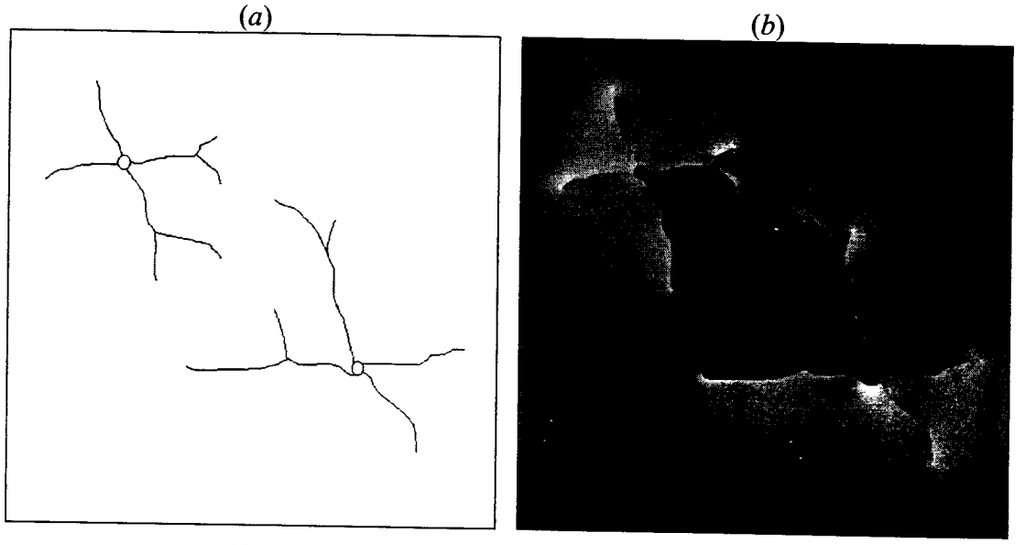


Figura 37 – Exemplo da função campo escalar (*b*) em uma imagem com dois neurônios (*a*) simulando uma interação.

Esta influência pode ser localmente definida em cada ponto do contorno do neurônio. Para isso, um conjunto de partículas ou pontos $L = \{\vec{q}_i\}$ formando uma circunferência em torno do objeto servirá como ponto de partida para um caminho entre uma partícula $\vec{q}$ e algum ponto do contorno (veja Figura 38*b*). Este caminho é orientado em cada passo pelo campo vetorial $\vec{f}$ (veja Figura 38*a*).

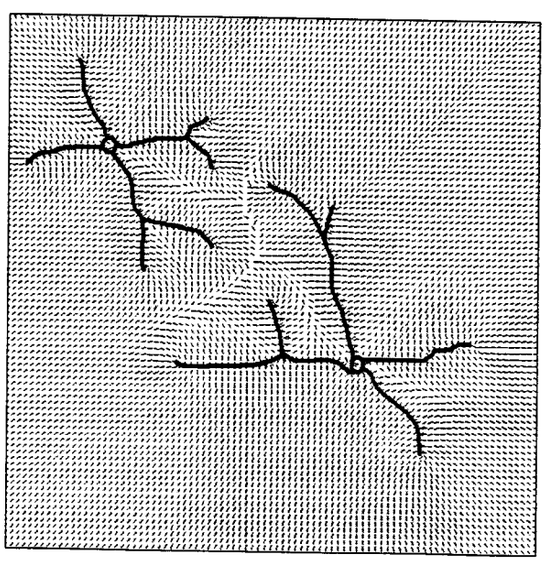


Figura 38 – Dois neurônios e o campo vetorial formado por sua influência para uma lei de campo $a = 4$.

O número de partículas $\vec{q}$ que chegam em um determinado ponto do contorno incrementa um acumulador que servirá como um histograma de partículas incidentes. Portanto, se um ponto for uma extremidade influente seu valor será mais alto. A Figura 39 mostra a visualização deste histograma para as formas da Figura 37*a*.

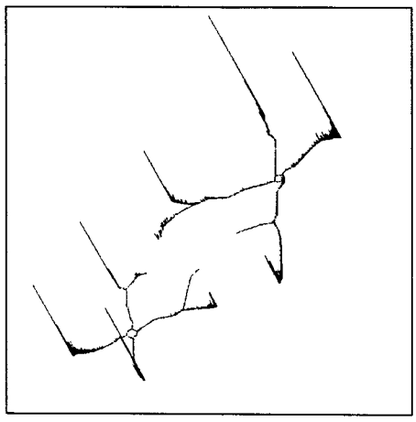


Figura 39 - Histograma de linhas de campo incidentes em dois neurônios

Esta interação pode ainda ser visualizada com contornos geométricos variando-se a distância entre eles. Na Figura 41 pode-se perceber que conforme um objeto se afasta do outro as saliências que interagiam tornam-se mais influentes que quando próximas. Na Figura 41*b*, as regiões escuras aparecem quando as formas estão próximas, que no campo vetorial aparecem como regiões nas quais as linhas de campo "evitam". A distribuição da densidade de linhas de campo ao longo do contorno pode ser visualizada na Figura 40. Ali é bastante evidente a alta densidade das saliências para uma lei de campo com valor $a=4$.

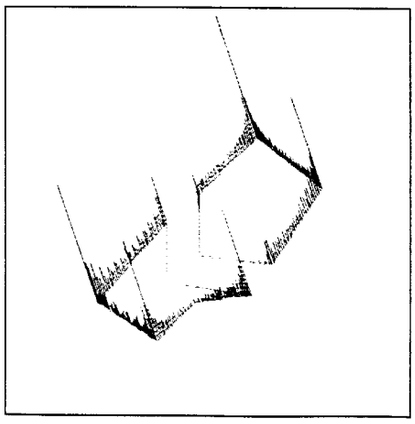


Figura 40 – Histograma de incidências de linhas de campo no par de formas geométricas.

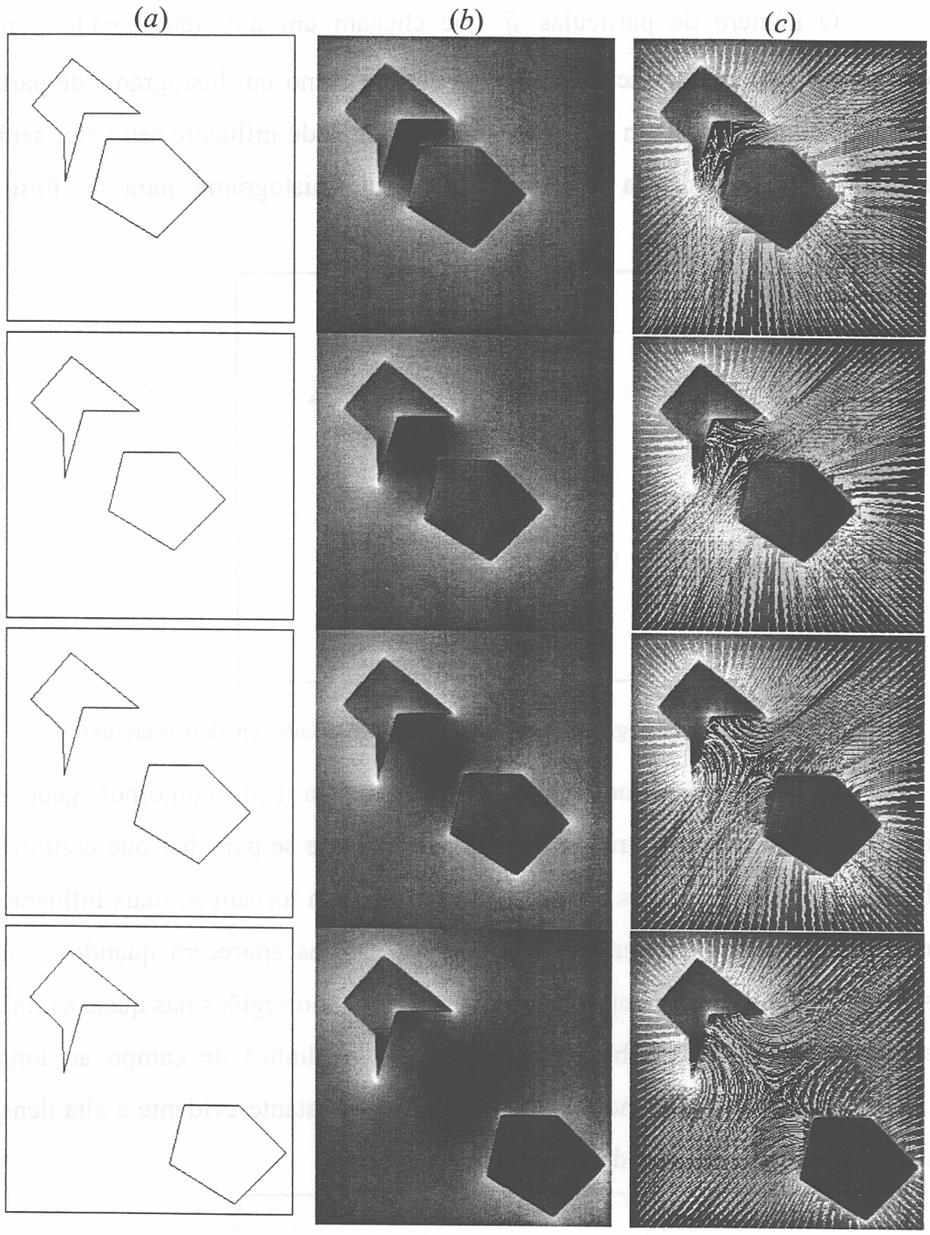


Figura 41 – Figuras geométricas (a) tomadas em diferentes deslocamentos com as respectivas funções pontenciais (b) e campos vetoriais (*c*) *f* sobrepostos a $U(x,y)$.

Uma outra análise interessante de ser realizada é sobre o comportamento do campo vetorial em diferentes leis. Isto é possível variando o valor *a* no cálculo do campo. Para os contornos geométricos, percebe-se que conforme cresce o valor de *a*, os caminhos das partículas tentam buscar as saliências das formas mostrando a analogia com o poder das pontas. A Figura 42 mostra o campo vetorial no qual as

partículas percorrerão um caminho até o contorno. Porém este campo revela as direções preferenciais.

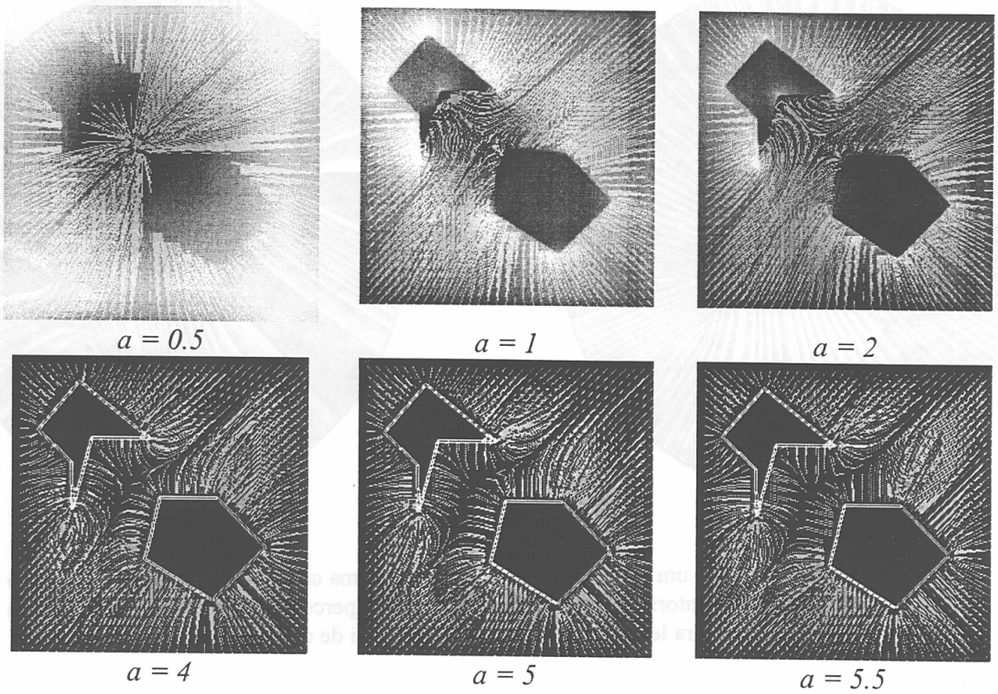


Figura 42 – Figuras geométricas em um campo vetorial variando a lei (coeficiente a) .

Para o par de neurônios da Figura 37a foram tomados os caminhos definidos para $a = 2$ (veja Figura 43a) nota-se que para $a = 4$, como ilustra a Figura 43b, uma quantidade maior de caminhos chega às saliências. Esta maior atração causa o aparecimento de regiões que praticamente não apresentam caminhos. Estas regiões têm uma incluência mais localizada, servindo como modelos para interações localmente muito fortes, mas que não produzem efeito em uma vizinhança mais abrangente. Isso não acontece para os caminhos com lei de campo $a = 2$. Embora haja regiões ainda não influenciadas no local da circunferência, os caminhos ocorrem entre os dois neurônios caracterizando assim uma influência mais ampla.

Os atributos que podem ser derivados a partir dessa abordagem são aqueles que definem, para cada elemento de contorno, a quantidade de linhas de força incidentes. As saliências são fortemente realçadas por esse atributo como mostra a próxima seção.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

(*a*) (*b*)

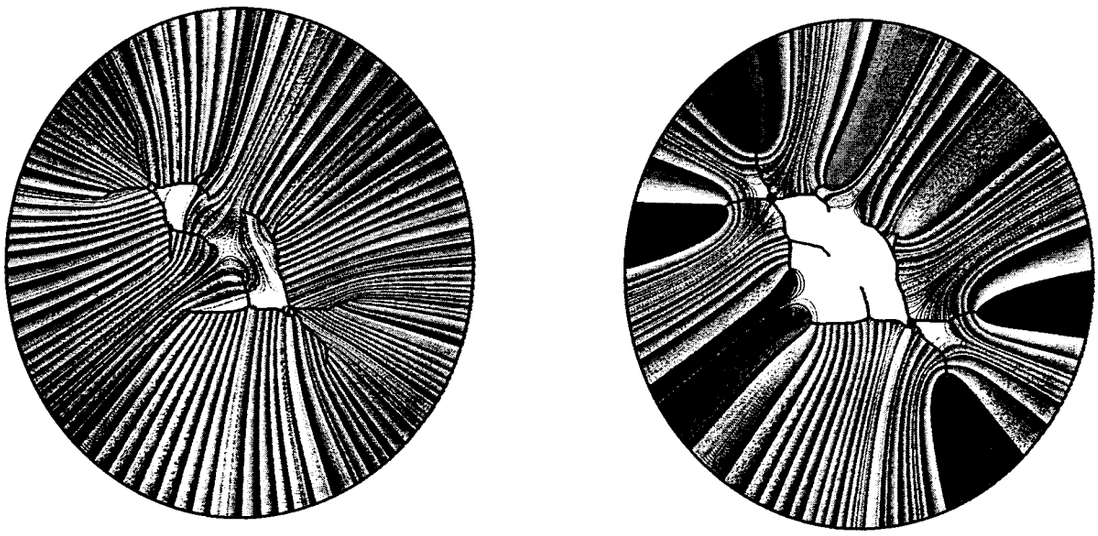


Figura 43 – Influência de um par de neurônios a partir de uma circunferência a qual é avaliado em cada ponto do contorno a densidade de caminhos percorridos por uma partícula. (a) mostra os caminhos para lei de campo $a = 2$ e (b) para lei de campo $a = 4$.

4.1.3 Obtendo atributos

A micrografia de um neurônio ganglionar da retina de galinha foi segmentada manualmente e sua forma é ilustrada pela Figura 44.

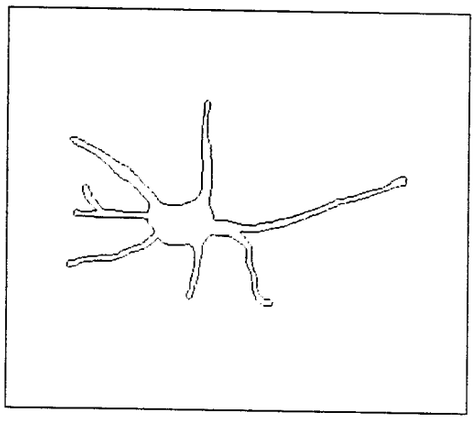


Figura 44 – Contorno de um neurônio.

Utilizando o método apresentado, calculou-se suas respectivas f_x (Figura 45*a*) e f_y (Figura 45*b*), ou seja, os respectivos componentes vetoriais ortogonais do campo eletrostático para lei de força com índice $a = 1$.

(*a*) (*b*)

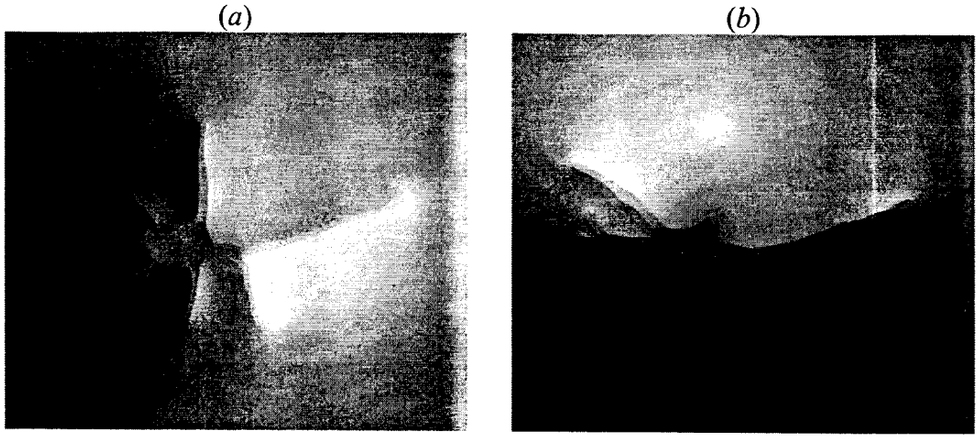


Figura 45 – Componentes x (*a*) e y (*b*) do campo vetorial calculados aplicando-se (48) e (49) respectivamente.

Sua função campo escalar $U(x, y)$ pode ser visualizada na Figura 46*a* e seu campo vetorial na Figura 46*b*.

(*a*) (*b*)

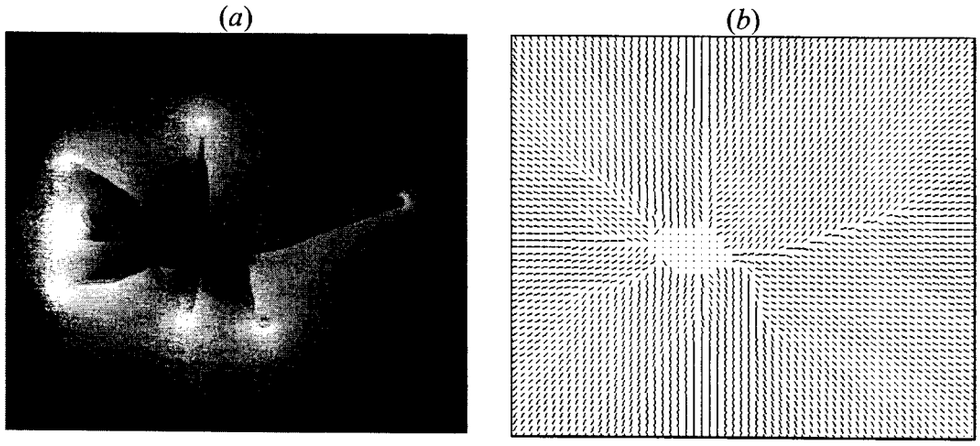


Figura 46 – Campo escalar (*a*) e vetorial (*b*) para o contorno da Figura 44.

A partir de alguns pontos de uma circunferência em torno da forma, verificou-se o elemento de contorno no qual se define o caminho no campo vetorial (Figura 46*b*). A Figura 47*a* mostra o comportamento dos caminhos. A Figura 47*b* mostra os caminhos de uma maneira mais densa, ou seja, de cada ponto da imagem externo à forma foi verificado o caminho, e cada pixel foi marcado com um nível de cinza rotulado a partir do contorno. Esta rotulação foi realizada como uma função dente-de-serra, crescendo a partir da intensidade *128* em um intervalo [0..255] para *0* correspondendo ao nível de cinza mais escuro e *255* ao mais claro.

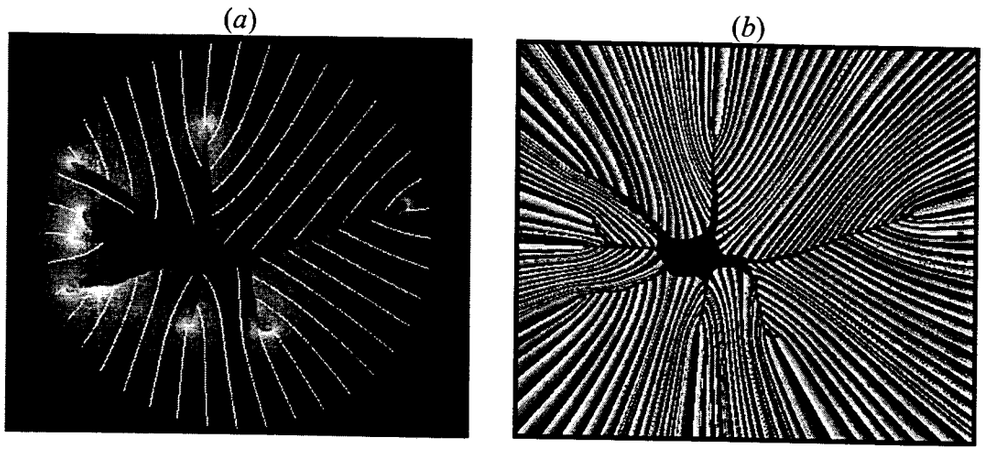


Figura 47 – Caminhos a partir de pontos de uma circunferência em torno da forma (*a*) e a partir de todos os pixels externos ao contorno (*b*).

Procedeu-se então uma contagem para cada ponto do contorno. Esta contagem é incrementada a cada vez que um caminho chega de um determinado pixel da circunferência ilustrado pela Figura 47*a*, porém agora partindo de todos os pixels da circunferência. O resultado é mostrado na Figura 48.

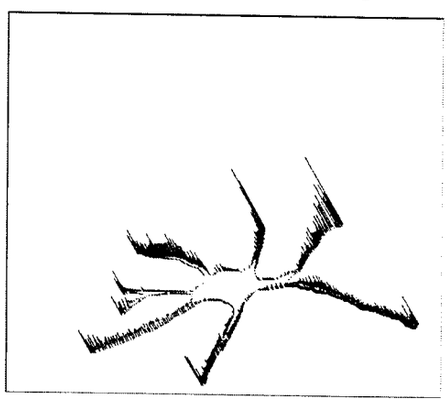


Figura 48 – Gráfico mostrando para cada ponto do contorno a sua influência pela chegada de linhas de força. A lei de campo corresponde ao índice $a=1$.

A Figura 48 ilustra uma leve preferência dos caminhos para as saliências. Na tentativa de tornar essa preferência mais forte, foram testadas outras leis de campo variando o índice a para valores $a = 2$ e $a = 4$. As respectivas figuras mostrando os caminhos chegando aos pontos do contorno (Figura 49*a* e *b*) ilustram esse aumento na preferência à saliência. As regiões contínuas com apenas um nível de intensidade revelam grandes quantidades de *pixels* chegando em apenas um ponto do contorno.

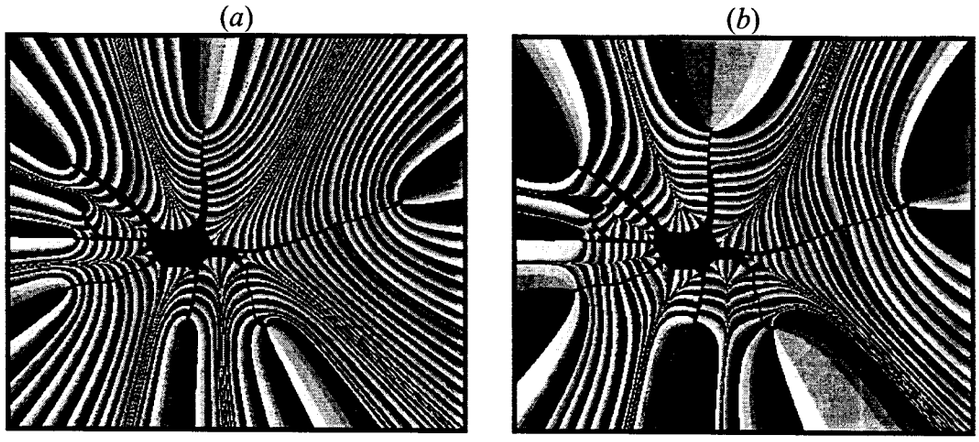


Figura 49 - Linhas de força calculadas para leis de campo $a = 2$ (*a*) e $a = 4$ (b).

A Figura 50. mostra agora uma forte preferência às saliências indicada pela altura da haste que sai do contorno do neurônio. Observa-se também que para $a = 4$ as saliências são identificadas com maior uniformidade. Para $a = 2$, algumas saliências apresentaram influências muito menores que para outras.

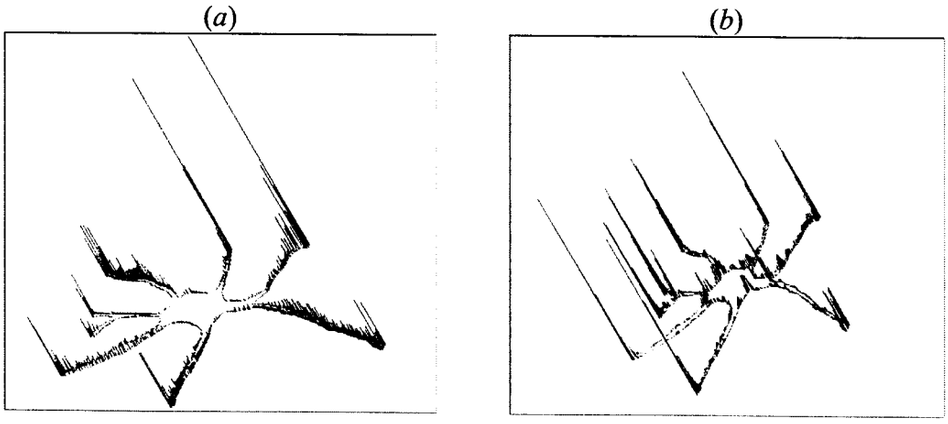


Figura 50 - Gráficos mostrando para cada ponto do contorno a sua influência pela chegada de linhas de força. As leis de campo correspondem aos índices $a=2$ (*a*) e $a=4$ (*b*).

Por fim, na Figura 51*b* o histograma de linhas de força é mostrado em um gráfico cartesiano e comparado à curvatura do contorno do neurônio (Figura 51*a*). Percebe-se uma semelhança na identificação dos pontos de maior curvatura. A vantangem do uso deste atributo ao invés da curvatura está no fato de que o cálculo de curvatura em um contorno digital depende de uma suavização, e deste modo o histograma seria menos dependente da variação de escalas. Contudo, o processo para

se obter o histograma é ainda bastante pesado computacionalmente, já que exige em vários passos o cálculo da transformada de Fourier bidimensional.

(*a*)

(*b*)

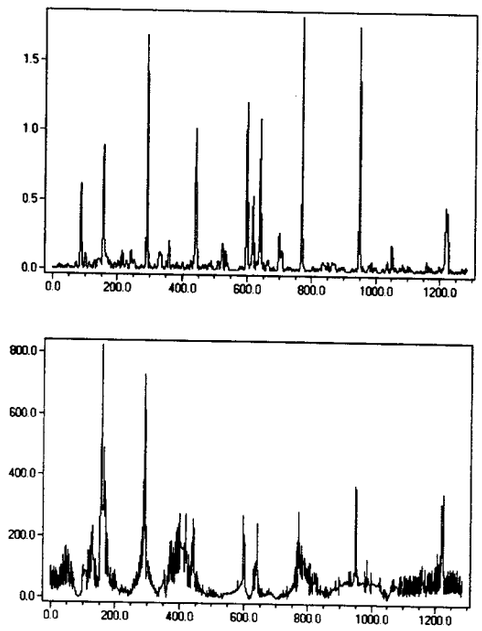


Figura 51 – Comparação entre a curvatura do contorno da **Figura 44** e seu histograma de linhas de forças para $a = 2$.

4.2 Esqueletonização multi-escalas com métricas de Minkowski variáveis

A análise e esqueletonização espaço-escala (ou multiresolução) são duas das mais promissoras e elegantes abordagens para o importante problema de análise de formas. Uma solução eficaz para estes problemas foi descrita em [COSTA & ESTROZI, 1999], baseada nos conceitos de dilatações exatas (de acordo com distâncias euclideanas) e propagação de rótulos. Esta abordagem apresentou uma série de propriedades desejáveis entre as quais:

(a) esqueletos espaço-escala de largura unitária e 8-conectados;

(b) preservação da topologia, ou seja, os esqueletos nunca se quebram durante a variação das escalas espaciais;

(c) permitem uma reconstrução espaço-escala sem deslocamento de borda, já que em escalas mais altas as formas reconstruídas não se deslocam enquanto os detalhes da escala espacial são removidos.

Apesar destas boas propriedades, o esforço computacional para realizar a esqueletonização é grande pois, depende fortemente do cálculo da distância

euclideana. Com o objetivo de minimizar a quantidade de distâncias calculadas, Costa e Estrozi [COSTA & ESTROZI, 1999] propuseram uma estrutura para que as distâncias fossem propagadas em vez de intensivamente calculadas. No entanto, um outro problema surge devido ao crescimento exponencial do tamanho da estrutura alocada para conter os deslocamentos das distâncias euclideanas. Este crescimento dificulta a esqueletonização para grandes distâncias. Embora a estrutura seja simples, ela não é computacionalmente viável de modo a permitir sua extensão à implementações 3D e em hardware paralelo.

Esta seção serve ao estudo e comparação de desempenho e de qualidade da esqueletonização multi-escalas para diversas métricas que podem tornar computacionalmente viável tal abordagem. Para isso, uma pequena revisão das métricas estudadas é apresentada. Então, é introduzido o processo de esqueletonização multi-escalas com propagação de rótulos considerando a distância euclideana exata. A seguir serão mostradas as extensões da abordagem para métricas alternativas. Por fim, uma implementação computacionalmente econômica aplicando a métrica *chessboard* é descrita.

4.2.1 Métricas Alternativas

A transformada distância de um conjunto de objetos binários A (contínuo ou discreto em $\mathbf{R}^n$), produz um mapa de distâncias D, obtido atribuindo-se a cada ponto $D(\vec{p})$ deste mapa, a menor distância entre o ponto $\vec{p}$ e algum ponto de algum objeto A. São definidas várias métricas para atribuir um valor à distância e portanto, várias transformadas distância (DT) podem ser adotadas, cada uma caracterizada pela métrica que a define.

Talvez a métrica mais conhecida seja a que define a distância euclideana definida por (52). Outras duas métricas fartamente utilizadas são as métricas *cityblock* (53) e *chessboard* (54).

$$d_E = \sqrt{\sum_{k=1}^{D} (p_k - q_k)^2} \tag{52}$$

$$d_\Diamond = \sum_{k=1}^{D} |p_k - q_k| \tag{53}$$

$$d_X = \max_{k=1}^{D}\{|p_k - q_k|\} \tag{54}$$

No entanto, tais métricas podem ser definidas como casos especiais em famílias como as métricas de Minkowski (55) ou as métricas *Chamfer* (56).

$$d = \sqrt[n]{\sum_{k=1}^{D}(p_k - q_k)^n} \tag{55}$$

$$d_{chamfer}(A,B) = A \times \max_{k=1}^{D}\{|p_k - q_k|\} + (B - A) \times \min_{k=1}^{D}\{|p_k - q_k|\} \tag{56}$$

Os casos especiais manifestam-se em (55) como valores $n = 2$ para a distância euclideana; $n = 1$ para métrica cityblock e $n = \infty$ para métrica chessboard. Para $n < 1$ não são definidas, pois não se ajustam aos quatro requisitos necessários para que uma métrica d entre um ponto p e outro q seja definida, ou seja:

1.) $d(p, q) = d(q, p)$;

2.) $d(p, q) = 0$ se e somente se $p = q$;

3.) $d(p, q) \geq 0$ para todo p, q

4.) $d(p, q) \leq d(p, w) + d(q, w)$

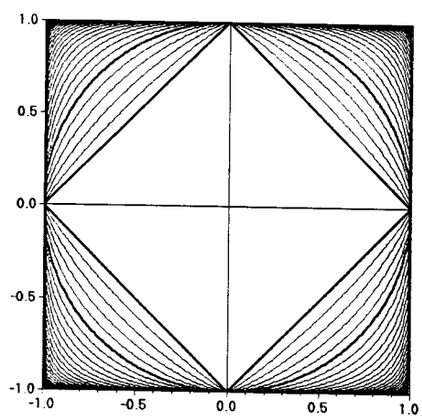


Figura 52 – Família de distâncias de Minkowski. As linhas mais fortes definem casos especiais.

A distância de Minkowski é a generalização das distâncias que aproximam uma bola unitária por formas variando entre o losango (d_1) e o quadrado (d_∞) Figura 52. A aproximação pelo losango (limite inferior da bola) resulta na bem conhecida distância city-block. Já a aproximação pelo quadrado (limite superior da

bola) é conhecida como distância *chessboard*. É fácil perceber que em direções que não são ortogonais as distâncias variam e que para a distância euclideana (o círculo) esse fato não ocorre (veja Figura 52).

A partir de (53) e (54) também é imediato concluir que as duas distâncias (*city-block* e *chessboard*) são bem menos computacionalmente custosas do que a euclideana. Esse fato tem feito com que diversas aplicações (imagens em movimento, 3D [PUDNEY, 1998], tempo-real) utilizem as aproximações para viabilizá-las.

Tentativas de aproximar a distância euclideana têm sido implementadas [MONTANARI, 1968; ARCELLI & DIBAJA, 1988]. A mais freqüente na literatura é a distância *Chamfer*. O processo de cálculo da distância Chamfer (56) leva em consideração duas direções possíveis de se caminhar em uma grade quadrada: a diagonal e a reta. Assim, atribuindo-se pesos a essas duas direções, é possível obter uma distância cujo erro pode ser formulado. A busca então, passa a ser pelos dois pesos (A e B, veja fórmula) que minimizem esse erro em relação à distância euclideana exata [BORGEFORS, 1984].

A distância octogonal é resultante de uma parametrização da distância Chamfer, (assim como *chessboard* e *city-block*), mas que permite uma menor invariância à rotação (a cada 45°) do que a *city-block* e *chessboard* (a cada 90°) [BORGEFORS, 1984].

A transformada distância é aplicada em uma série de problemas: reconhecimento de caracteres [CUISENAIRE, 1999], casamento de formas [LIU & SRINATH, 1990], esqueletonização entre outros. A propriedade de invariância à rotação é bastante desejada em reconhecimentos que envolvem variação do ângulo original da forma. É certo que reconhecimentos normalmente exigem exatidão apenas em baixas distâncias [LIU & SRINATH, 1990]. Problemas acontecem, no entanto, quando a informação extraída das cristas do mapa de distâncias (os máximos locais) é transportada para um grafo, construindo assim um descritor da forma. Se tivermos uma rotação na forma e tais máximos locais se alterarem, é possível que a descrição se altere também.

Claro que o objetivo principal dessa busca por uma aproximação à transformada distânica euclideana (EDT) é encontrar um algoritmo rápido. Existe porém um compromisso entre o tempo de execução do método e o erro com relação

a EDT. Normalmente maiores erros possibilitam uma maior rapidez. As técnicas de chamferização podem ser implementadas varrendo-se apenas duas vezes uma imagem com uma máscara. O erro pode ser reduzido aumentando-se o tamanho dessas máscaras ou também variando os seus parâmetros.

Um exemplo é ilustrado na Figura 57, para a qual a rotação do contorno de uma mão e um círculo exterior é mostrada para várias distâncias. Percebe-se que distâncias com aproximações mais grosseiras *(city block* e *chessboard*) deixam as imagens rotacionadas bastante diferentes. A distância *Chamfer* com parâmetros (3, 4) claramente aproxima bem a EDT, no entanto, existem variações em distâncias maiores que podem comprometer alguma aplicação [KAO, 1999].

Um *survey* sobre transformada distância com métricas alternativas incluindo uma visão abrangente sobre o método chamfer pode ser encontrado em [BORGEFORS, 1986]. Em [BORGEFORS, 1984], Borgefors generaliza a distância chamfer para outras dimensões (n-dimensional) e estende o conceito para grades hexagonais.

Com o intuito de estudar as propriedades destas métricas foi implementado e aplicado o algoritmo de transformada distância para um conjunto de métricas em figuras que apresentem peculiaridades que explicitem alguma deficiência da métrica e que possam apresentar comportamentos desejáveis para determinadas aplicações.

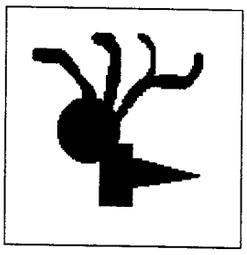


Figura 53 – Figura contendo diversas características.

Deste modo, uma Figura (veja Figura 53) reunindo um conjunto de características como: ramificações, regiões circulares, cantos vivos, regiões retangulares e triangulares foi submetido ao cálculo de um conjunto de métricas. Os resultados são ilustrados para as métricas Minkowski com valores de n variando no conjunto {1 (*city block*), 1.5, 2 (euclideana), 10 (*quasi-chessboard*)} (veja Figura 54) e para as distâncias *chamfer*$(1, \infty)$ correspondente à *city block*, $(3, 4)$: uma aproximação à distância euclideana, (1,1) cujo resultado é a métrica chessboard e um

caso especial, a métrica octagonal, resultante de quatro varreduras da imagem permitindo uma distância cujo círculo unitário é um octágono. Para facilitar a visualização, os resultados das transformada distância passaram por uma operação $D = D$ **mod** 20.

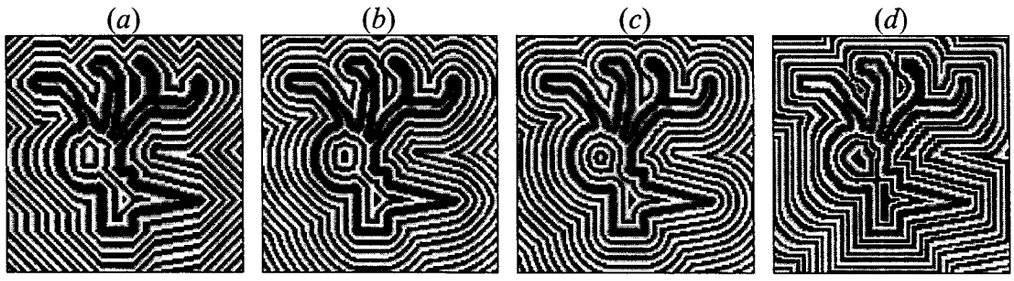


Figura 54 – Distâncias Minkowski aplicadas à **Figura 53** com índices n = *1*(*a*), *1.5*(*b*), *2*(*c*) e *10*(*d*).

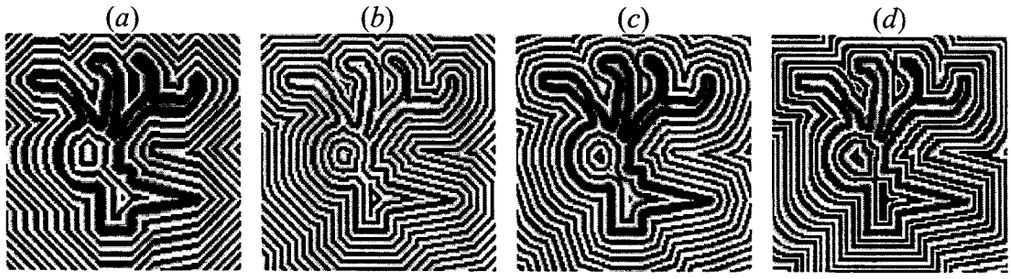


Figura 55 – Distâncias *chamfer* aplicadas à **Figura 53** correspondentes aos índices (1, ∞) *city block* (*a*); octagonal (*b*); (3, 4) (*c*); (1,1) *chessboard* (*d*).

Pode ser percebido da Figura 55 que as DTs em diferentes métricas correspondem a dilatações por diferentes elementos estruturantes. Para o caso da cityblock, um losango, para a métrica euclideana: o círculo; para a chessboard: um quadrado; para a chamfer octagonal: um octágono. Na verdade, tais formas são as aproximações a um círculo que essas métricas proporcionam. No entanto, o preço destas aproximações deve ser cobrado em perdas de algumas propriedades em relação à exata, ou seja, a euclideana. Algumas distâncias, como a chamfer 3,4 aproximam bem para distâncias pequenas e surgem erros maiores para maiores distâncias. Assim, para aplicações no qual interesse apenas uma região próxima ao objeto, com certeza será uma alternativa promissora. Em um conjunto de aplicações pode ser indesejável que a distância varie conforme o objeto rotacione. Esta propriedade, a invariância à rotação, é bastante comprometida quando distâncias com grande erro tais como cityblock e chessboard são adotadas. Como as formas destas métricas são losangos ou quadrados, ou seja, não são isométricas, as rotações do

objeto não mantém a mesma dilatação como ocorre na euclideana. A métrica octagonal apresenta um paliativo para este problema, se a aplicação permitir que o objeto rotacione apenas em ângulos proporcionais a 45°.

O efeito da rotação em mapas de distância é ilustrado na Figura 57. Ali são apresentados os mapas de distâncias para as imagens de silhuetas de uma mão humana (a) e outras duas rotacionadas em 30° (b) e 45°. Para cada conjunto de imagens rotacionadas foram calculados os mapas de distâncias para várias métricas: *city block*, *chessboard*, octagonal, chamfer(3,4) e euclideana. Para melhor visualização os mapas de distância foram submetidos a uma operação $D = D$ **mod** 40.

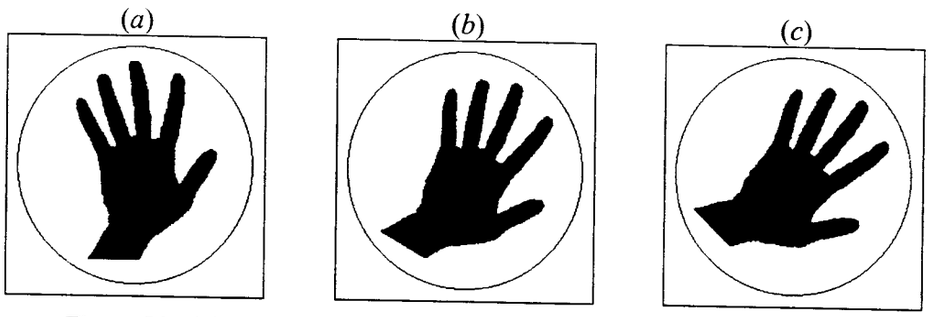


Figura 56 – Silhuetas de mão humana (*a*) rotacionadas em (*b*) 30° e (*c*) 45°.

Nas métricas com aproximações mais grosseiras ao círculo, ou seja, cityblock e chessboard, percebe-se a variação clara em relação à euclideana. Nas extremidades é mais fácil perceber que formas losangulares em uma orientação ficam retangulares em outras para estas distâncias. Para as distâncias mais aproximadas (octagonal e *chamfer*(3,4)) percebe-se uma diferença mais acentuada em distâncias maiores, ou seja, o círculo em torno da silhueta da mão aparece em distâncias maiores como dilatações retas para a octagonal e levemente poligonizada para a *chamfer*(3,4). Para a métrica euclideana, embora seja esperada a invariância à rotação, ela não acontece sem erros devido à amostragem do contorno na grade quadriculada dos pixels. Como ilustrado na Figura 58 a simples rotação de uma Figura em 45° altera a configuração de distâncias e idealmente isso deveria passar por um processo de normalização.

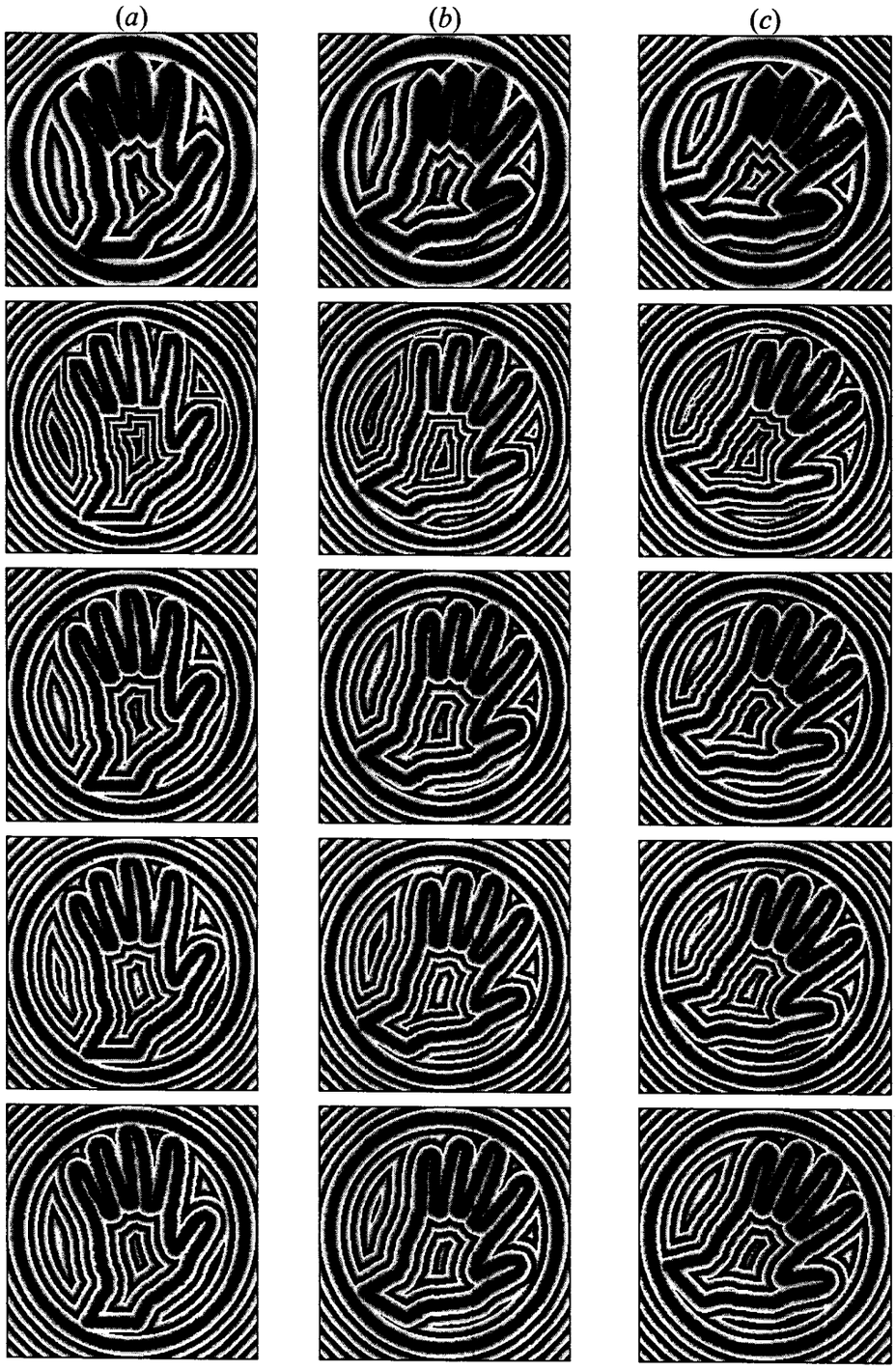


Figura 57 – Mapas de distâncias de uma mão humana rotacionada em diferentes métricas: (*a*) cityblock; (*b*) chessboard; (*c*) octagonal; (*d*) chamfer(3,4); (*e*) euclideana.

Invariância a rotação é uma propriedade desejável quando uma aplicação exige imagens em movimento rotacional. Apenas uma das métricas de Minkowski satisfaz essa propriedade em uma grade discreta: a euclideana . Porém, seu cálculo

implica em grande esforço computacional, seja por não ser possível uma propagação exata de uma frente de onda circular em uma grade discreta [THOMPSON & ROSENFELD, 1993], seja por embutir operações como raiz quadrada e multiplicação (52).

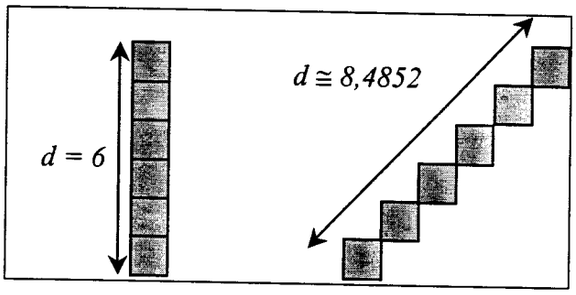


Figura 58 – Efeito da rotação sobre a distância de uma linha reta vertical e outra rotacionada de 45°.

Os métodos para se calcular a DT de uma imagem binária são implementados, ou tomando o fundo e calculando a distância de seus pontos às bordas do objeto mais próximo, ou então partindo-se da borda do objeto e calculando sua distância aos pontos de fundo. A rapidez da execução depende muito da quantidade de pontos de borda de objeto ou fundo. Algoritmos de chamfer, por exemplo, utilizam-se da varredura de uma imagem, importando-se apenas com o tamanho da imagem. Alguns algoritmos, no entanto, não só partem da borda do objeto como também se utilizam da analogia com a propagação de ondas de Huygens, para que não seja necessário revisitar pontos [CUISENAIRE & MACQ, 1999].

A propagação considera uma "onda" cuja forma é dada por uma métrica. Vimos que ela pode ser um losango (city block), um quadrado (chessboard), um octágono, etc. A propagação define, em uma grade discreta, uma vizinhança (aqui a analogia com a onda) na qual a medida mais distante dependa somente da soma dos passos percorridos no caminho. A busca por distâncias que imitam a euclideana vem desde trabalhos como o de Montanari [MONTANARI, 1968] que introduz o conceito de pseudo-distâncias, baseado na minização do erro percorrido por dois passos.

Danielsson [DANIELSSON, 1980] propõe uma propagação vetorial baseada na varredura da imagem binária em várias direções, só que com máscaras contendo as diferenças das distâncias em cada direção ortogonal. Leymarie conseguiu reduzir o

tempo de execução reparametrizando a propagação para diminuir o número de operações [LEYMARIE & LEVINE, 1992]

Ragnelmam [RAGNELMAM, 1992] introduz uma propagação em forma de buquê, i.e., cada ponto pertencente à borda objeto é o centro de uma estrutura que contêm várias máscaras (o mesmo número de distâncias que se queira) e que percorre os pontos do fundo da imagem apenas uma vez, atualizando os pontos que ainda não foram preenchidos com alguma distância.

Os máximos locais resultantes da aplicação da DT sobre uma imagem binária contendo bordas define o eixo de simetria da forma original. Esse eixo é o eixo principal ou eixo medial. Wright [WRIGHT, 1995] estende o conceito de distância euclideana para que os máximos locais definam também eixos locais de simetria (*Smoothed Local Symmetries*) [BRADY & ASADA, 1984].

Um outro conceito estendido é a distância geodésica. Tal distância é definida sobre um caminho de pontos conectados em uma vizinhança. Assim, partindo-se de um ponto desse caminho e chegando a outro também no mesmo caminho pode-se calcular a distância percorrida. Conhecer a distância percorrida em cada ponto pode depender apenas da vizinhança (métricas não euclideanas) ou então conhecer a característica analítica do caminho [CUISENAIRE, 1999].

A demanda por resultados mais precisos em DT e sua aplicação em reconhecimento e descrição de formas, pode ser atendida inserindo-se paralelismo na implementação de seus algoritmos. Yamada [YAMADA, 1984] propôs o primeiro algoritmo paralelo para uma EDT exata para inteiros, que se tratava de um particionamento da DT em algoritmos sequenciais em cada bloco (complexidade $\mathbf{O}(N^3)$).

Alguns autores paralelizam algoritmos que calculam a DT por varredura (como a chamfer). Paglieroni [PAGLIERONI, 1992] propõe um algoritmo que varre linhas e colunas reduzindo as DTs bidimensionais à problemas de DTs unidimensionais. Essa abordagem não calcula EDTs exatas e ele a chamou de UDT (Unified Distance Transform). Kolountzakis e Kutulakos [KOLOUNTZAKIS & KUTULAKOS, 2000] exploraram propriedades das métricas euclideanas para reduzir a complexidade ($\mathbf{O}(N^2 \log N)$). Takala e Viitanen [TAKALA & VIITANEN, 1999] propõem uma arquitetura SIMD para implementar algoritmos chamfer

decompondo-o apenas em operações da aritmética binária, tornando-o adequado para implementação em hardware.

Outros autores paralelizam propagações, para tentar diminuir o erro entre a DT e a EDT exata lidando com o compromisso entre o tamanho da vizinhança e o erro. Ragnelmam [RAGNELMAM, 1992] sugeriu propagações sequenciais da EDT com sinal. Huang e Mitchel [HUANG & MITCHEL, 1994] implementaram um algoritimo paralelo para propagação em métrica chessboard.

Paglieroni [PAGLIERONI, 1992] propõe uma arquitetura dedicada à UDT, algoritmo baseado em varredura de linhas e colunas calculando DTs unidimensionais e com isso gerando DTs bidimensionais. Ele a aplica em esqueletonização de circuitos impressos e também em casamentos de imagens aéreas urbanas em estéreo.

4.2.2 Esqueletonização multi-escalas sem deslocamento de borda

O método para extrair os esqueletos da forma ocorre em três etapas: (1) as dilatações exatas; (2) a propagação de rótulos; (3) e as máximas diferenças. No primeiro passo, cada ponto do limite da forma original é dilatado, simulando a propagação de ondas de choque. Para identificar o ponto do contorno do qual se originou o choque, cada um deles é seqüencialmente rotulado. A imagem resultante desse processo serve como entrada para a avaliação das máximas diferenças. O processo de máximas diferenças consiste em varrer a imagem, tomando cada elemento e subtraindo seu conteúdo do conteúdo de cada um de seus quatro vizinhos. O resultado é uma imagem cuja limiarização em cada nível de cinza, partindo do maior valor para o menor, produz esqueletos em escalas representadas pelo nível de cinza no qual ocorreu o corte. Cada esqueleto obtido pela limiarização em uma escala específica pode ser reconstruído tomando a distância de cada ponto do esqueleto e construindo um elemento estruturante com as dimensões dessa distância.

Todo este processo é computacionalmente intensivo. No entanto, para extrair saliências, apenas algumas escalas são necessárias, permitindo restringir a quantidade de distâncias euclideanas a serem avaliadas.

Esta seção segue explicando em maiores detalhes a estrutura para avaliar a distância euclideana por propagação exata, e sua aplicação para gerar a propagação

de rótulos. Além disso, as máximas diferenças e as restrições para reduzir o tempo de execução serão apresentadas.

4.2.3 A Propagação da Distância Euclideana

Conforme visto na seção 4.2.1, a Transformada Distância (DT) de uma imagem binária é dada por uma outra imagem, cujos elementos são calculados tomando a menor distância entre o elemento corrente e um elemento pertencente à borda da forma.

Visualizando a borda de uma forma como um conjunto $B = (b_1, b_2, \ldots, b_N)$, tal que $b_i = (x_{bi}, y_{bi})$. A DT de um ponto $p = (x_p, y_p)$ é dada por d_p, sendo $d_p = (x_p, y_p, z_p)$; x_p e y_p coordenadas em R^2 e z_p o valor da transformada distância $z_p = \min_{i=1}^{N}\left\{\sqrt{(x_p - x_{bi})^2 + (y_p - y_{bi})^2}\right\}$. Neste caso, assume-se uma métrica Euclideana. Buscando reduzir a sobrecarga de se tomar todos os pontos do domínio para calcular a DT [COSTA, 1999a; COSTA, 1999b], Costa e Estrozi [COSTA & ESTROZI, 1999] propõem um método de propagação de distâncias. Esse método consiste em dilatar a forma em questão, cujo critério é dado, não por um elemento estruturante estático, mas por uma estrutura de dados que contém distâncias Euclideanas exatas. Tal estrutura é uma lista contendo outras listas com as posições que em uma grade possuem a mesma distância. Tal estrutura pode ser melhor visualizada na Figura 59.

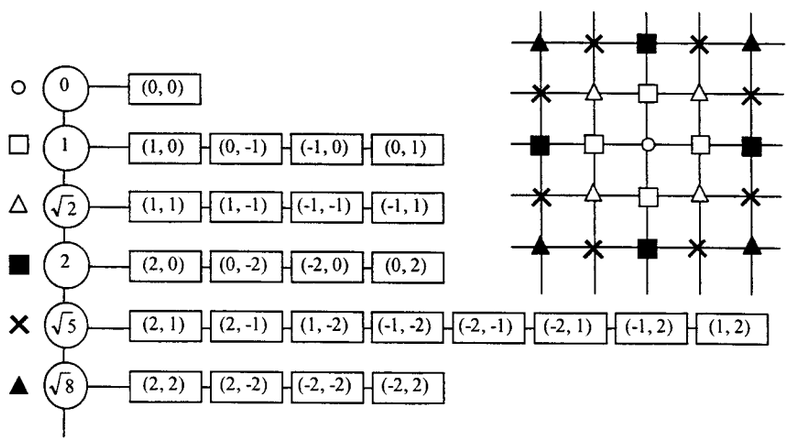


Figura 59 - Estrutura SD para propagar a distância euclideana com dilatação exata

Sendo SD a estrutura de distâncias, SD_d nada mais é do que uma lista cujos nodos contém os deslocamentos que devem ser feitos em uma grade quadrada para percorrer a distância d nessa grade a partir do ponto central. Ou seja,

$SD_d = \{(x_{dk}, y_{dk})\}$, na qual $d_k = \sqrt{x_{dk}^2 + y_{dk}^2}$ para todo k. Uma vez calculada SD, cada d ocupará a vizinhança de um ponto b_i do contorno.

Essa estratégia serve também para propagar identificadores de caminhos entre cada ponto do domínio e um elemento do contorno. Para isso, cada elemento do contorno é rotulado com um valor inteiro e sequencial (veja Figura 60b), ou seja, $L_B = \{i\}$ para $i = 1..N$.

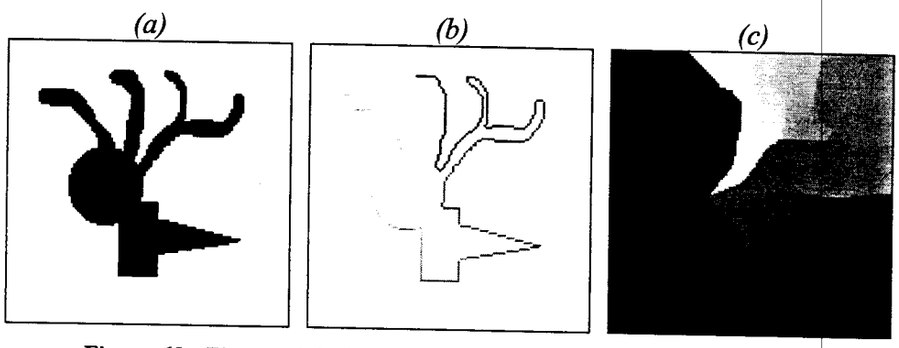


Figura 60 - Figura original *(a)*; Imagem da rotulação de cada elemento do contorno *(b)*; Imagem da propagação dos rótulos do contorno pela dilatação exata.

Então, ao invés do valor da distância, o que preenche os pontos do domínio é o rótulo. Ao final do processo, resulta uma imagem que identifica, para cada ponto fora da borda, o menor caminho percorrido na superfície da DT (veja Figura 60c).

Algoritmo 8 - Propagação de rótulos

```
para cada distância d
   para cada elemento e do contorno
      para cada nodo da estrutura SD_d
         visite cada ponto b_i + SD_d corrente;
         se o ponto ainda não contem um rótulo
         então atribua e ao ponto;
      fim_para
   fim_para
fim_para
```

Para obter os esqueletos, é necessário agora, encontrar as máximas diferenças que identificam os caminhos dos choques das ondas ao longo da sua propagação.

Isso é feito passeando-se a máscara da Figura 61a sobre cada pixel da imagem e tomando a máxima diferença entre o ponto corrente e cada um de seus

vizinhos (4-vizinhança). O resultado é uma imagem cuja limiarização define um esqueleto na escala daquele limiar (veja Figura 61b).

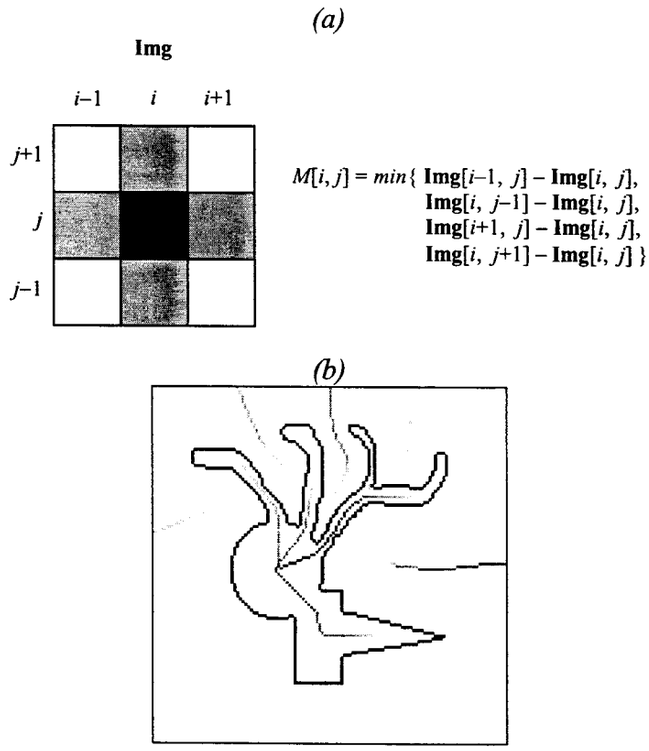


Figura 61 - Cálculo da matriz de diferenças *(a)*; Resultado da aplicação do algoritmo ilustrado ao lado *(b)* mostrando os esqueletos interno e externo

O comportamento multi-escalas do esqueleto da Figura 61*b* pode ser melhor visualizado pelas fatias extraídas. Cada fatia T contem um esqueleto E_T para um nível de detalhe.

$$E_T = \{p_k\} \qquad k = 1..N_T \tag{57}$$

As escalas mais baixas descrevem maiores detalhes (veja Figura 62*a*) e quando muito baixas refletem a grade de amostragem. Conforme aumenta a escala, os detalhes vão sendo "podados" pelo critério da distância geodésica.

De posse das respectivas imagens do esqueleto e da transformada distância euclideana d é possível reconstruir a forma para uma determinada escala. A reconstrução, para essa abordagem, apresenta uma propriedade em particular: as bordas da forma reconstruída não se deslocam com relação às bordas da forma original. A reconstrução é obtida pela união dos círculos (ou bolas) cujo centro é cada ponto do esqueleto em uma escala T. Estes discos têm raio igual à distância $d(p_k)$ entre o ponto p_k do esqueleto e a borda.

$$H = \bigcup_k C(p_k, d(p_k)) \tag{58}$$

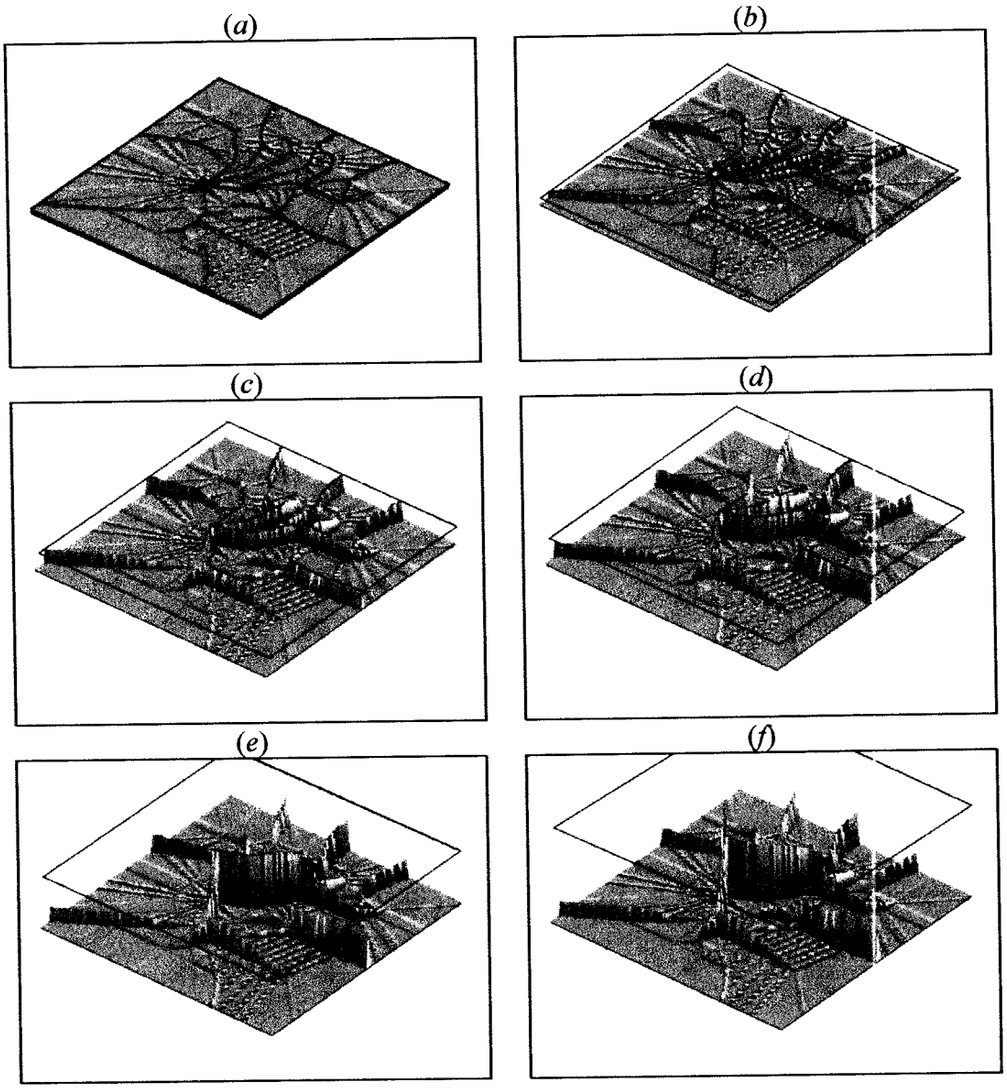


Figura 62 – Fatias aplicadas sobre a imagem de diferenças da Figura 61b. As escalas são apresentadas em ordem crescente: (*a*) $T = 6$; (*b*) $T = 20$; (*c*) $T = 50$; (*d*) $T = 75$; (*e*) $T = 150$; (*f*) $T = 265$. As linhas mais escuras (vermelhas) indicam a fatia e as mais claras (verdes), a Figura original.

Analogamente à Figura 61*b*, a reconstrução consiste de uma imagem em múltiplas escalas (veja Figura 63), as quais definem uma regra de formação da imagem. Para o caso corrente, o critério para essa formação é o perímetro entre o círculo bitangente à borda cujo centro é um ponto do esqueleto. Esta regra define

qual ponto do esqueleto ou parte da imagem será "podada" antes de outra. Outros critérios poderiam ser definidos, porém, estes teriam que ser impostos às escalas.

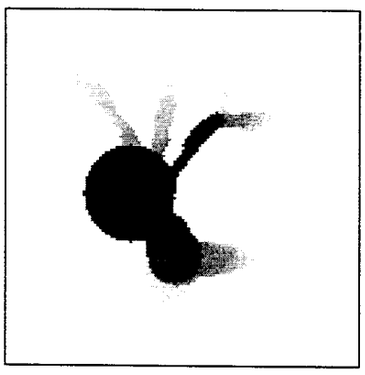


Figura 63 – Imagem com os discos de reconstrução da forma da Figura 60*a* em multiplas escalas. Os níveis de cinza indicam cada escala.

A Figura 64 mostra a reconstrução dos esqueletos internos da Figura 62. Os círculos obtidos para cada escala são destacados. O contorno original acompanha cada reconstrução para efeito de comparação.

Todos os processos envolvidos com a esqueletonização apresentada foram executados partindo-se de uma métrica euclideana. A isometria desta métrica é a propriedade mais desejada para aplicações as quais a invariância à rotação seja imprescindível. No entanto alguns problemas surgem na implementação. O primeiro destes problemas, o tempo de execução, é causado pelo grande número de distâncias euclideanas na grade ortogonal. A Figura 65 mostra o total de distâncias euclideanas, N_d, observadas em uma esfera digital de raio R para vários valores de R. Vê-se que N_d cresce substancialmente conforme o incremento de R. Como o método que usa a métrica euclideana exige uma varredura completa da forma do contorno (contendo N elementos) para cada distância, um total de $N_o = N_d.N$ operações básicas são necessárias para sua execução. A relativa complexidade daquela abordagem implica na necessidade de armazenar todas as possíveis distâncias e os respectivos vetores das suas posições.

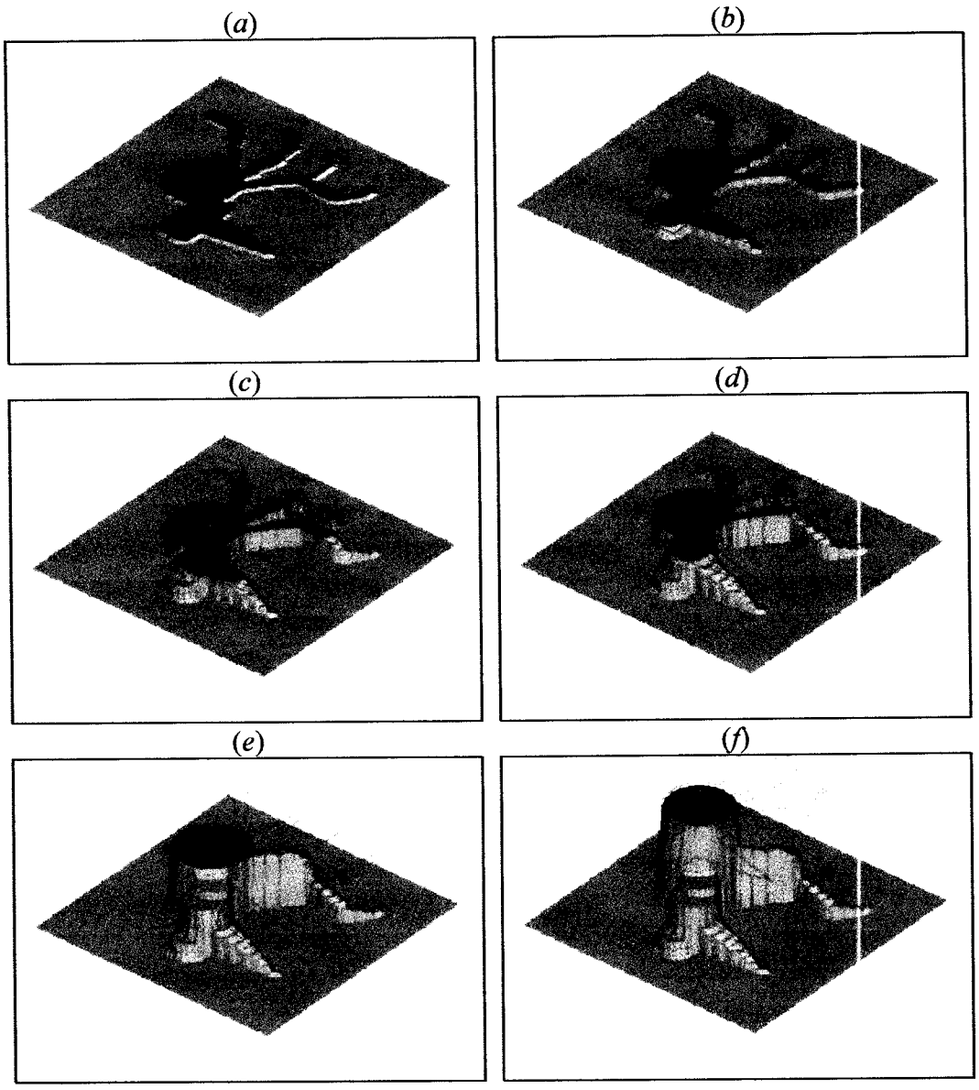


Figura 64 – Reconstrução para os respectivos esqueletos da Figura 62.

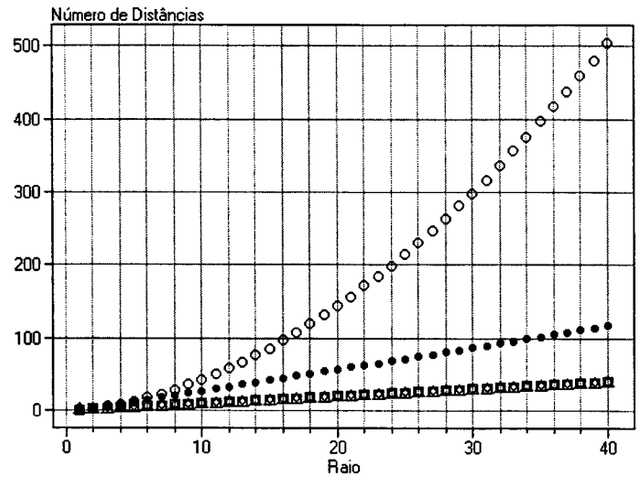


Figura 65 – Variação do número de distâncias com relação ao raio de uma círculo digital. Bolas vazadas correspondem à distância euclideana; bolas cheias à *Chamfer*(3,4); losangos, quadrados e triângulos (sobrepostos) correspondem respectivamente à *city block*, *chessboard* e octagonal.

Na busca de soluções para o problema, foram implementadas esqueletonizações e reconstruções espaço-escala a partir de outras métricas (*Minkowski* e *Chamfer*) além da euclideana. Surpreendentemente, em algumas delas (*city block* e *chessboard*) as boas propriedades dos esqueletos gerados com o método euclideano são mantidas, ou seja, largura unitária, conectividade 8 e a preservação da topologia conforme a variação da escala. A Figura 66 mostra os esqueletos para a forma da Figura 60 nas métricas Minkowski variando seu coeficiente de $n = 1$ correspondendo à distância city block, passando por um coeficiente intermediário ($n = 1.5$) e pela distância euclideana ($n = 2$) até chegar a uma aproximação da distância chessboard ($n = 10$).

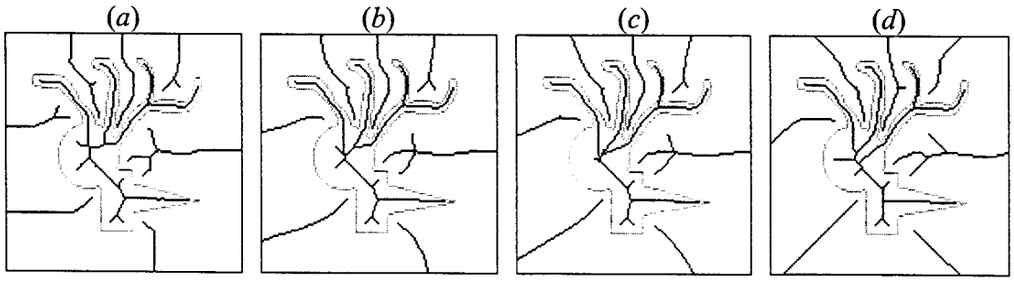


Figura 66 – Esqueletos para a forma da Figura 60*a* extraídos ($T = 9$) usando a métrica de Minkowski com coeficientes (*a*) $n = 1$; (*b*) $n = 1.5$; (*c*) $n = 2$; (*d*) $n = 10$.

Para métodos usando distâncias Chamfer, os esqueletos são ilustrados na Figura 67.

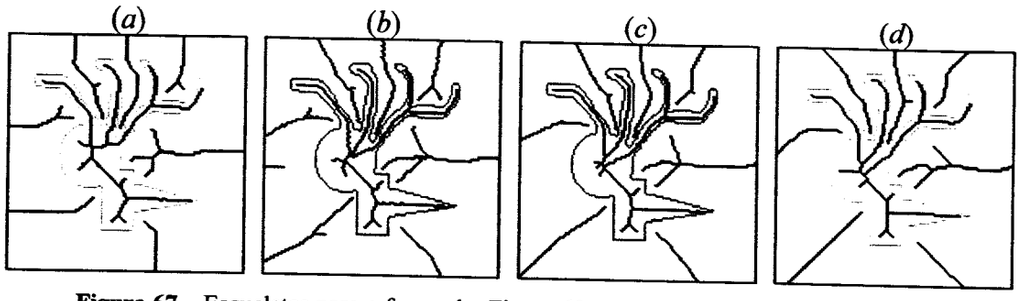


Figura 67 – Esqueletos para a forma da Figura 60*a* extraídos ($T = 9$) usando métodos *chamfer* para (*a*) *city block*; (*b*) octagonal; (*c*) *chamfer*(3,4); (*d*) *chessboard*.

Percebe-se que a adoção de diferentes métricas provoca alterações nas ramificações do esqueleto. Estas mudanças afetam o espaço-escala refletindo-se na ordem com que partes da imagem são desconsideradas ou reconstruídas. Além disso, nem todas as métricas apresentam propriedades de continuidade ao longo das escalas. Métricas simples e anisotrópicas como *chessboard* e *city block* (casos especiais da métrica de Minkowski) contudo, permitem esqueletos contínuos e algoritmos rápidos. Para a *chessboard* é proposto um algoritmo viável para a esqueletonização e reconstrução espaço-escalas.

4.2.4 Esqueletonização multi-escalas com métrica "chessboard"

Nas técnicas atuais, a distância (euclideana ou chessboard) é usada para extrair a transformada distância dos contornos da forma e propagar os rótulos atribuídos a cada um dos elementos do contorno. A nova técnica de esqueletonização espaço-escala é obtida usando a distância chessboard no lugar da euclideana. De imediato, duas vantagens podem ser identificadas: (a) a distância *chessboard* é extremamente simples e não requer uma estrutura de dados específica para conter todos os valores possíveis de distância; e (b) dado um valor máximo de distância, digamos R, há exatamente e somente R possíveis distâncias chessboard exatas em um quadrado digital de largura $2R+1$. Isto é muito menor que o total de distâncias euclideanas (veja Figura 65).

O algoritmo completo é extremamente simples e é apresentado no pseudo-código do Algoritmo 9.

Algoritmo 9 - Propagação de rótulos e DT *chessboard*.

```
para d =1 até R
  para i=1 até N
    x = lista_x[i];
    y = lista_y[i];
    p_Dir = x+d;  p_Esq = x–d;
    p_Cima = y+d;  p_Baixo = y–d;
    para ix=p_Esq até p_Dir
      se (Img_Dist[ix, p_Cima] < 0) ou
          ((i ≥ N–2d) e (Img_Rótulos[ix, p_Cima] = 1))
         Img_Dist[ix, p_Cima] = d;
         Img_Rótulo[ix, p_Cima] = i;
      fimse;
      se (Img_Dist[ix, p_Baixo] < 0) ou
          ((i ≥ N–2d) e (Img_Rótulos[ix, p_Baixo] = 1))
         Img_Dist[ix, p_Baixo] = d;
         Img_Rótulo[ix, p_Baixo] = i;
      fimse;
    fimpara;
    para iy=p_Baixo até p_Cima
      se (Img_Dist[p_Esq, iy] < 0) ou
          ((i ≥ N–2d) e (Img_Rótulos[p_Esq, iy] = 1))
         Img_Dist[p_Esq, iy] = d;
         Img_Rótulo[p_Esq, iy] = i;
      fimse;
      se (Img_Dist[p_Dir, iy] < 0) ou
          ((i ≥ N–2d) e (Img_Rótulos[p_Dir, iy] = 1))
         Img_Dist[p_Dir, iy] = d;
         Img_Rótulo[p_Dir, iy] = i;
      fimse;
    fimpara;
  fimpara;
fimpara;
```

No Algoritmo 9, N é o número de elementos de contorno da forma; R é a máxima distância chessboard considerada; *lista_x* e *lista_y* são vetores (listas) contendo as coordenadas dos elementos de contorno; *Img_Dist* é uma matriz (mesmo tamanho da imagem contendo a forma) usada para guardar a transformada distância e é iniciada com todos elementos iguais a -1; *Img_Rótulos* é uma matriz (mesmo tamanho da imagem contendo a forma) inicializada com zeros e que abrigará os rótuos i propagados. A condição "($i \geq N-2d$) **e** (Img_Rótulo[$x+ix$, *p_Cima*] = 1)" é necessária para assegurar uma propagação correta dos rótulos em uma região maior que aquela coberta inicialmente pelo primeiro elemento de contorno (como não há restrições quando o primeiro rótulo $i=1$ é propagado, ele se estende em uma região maior que o necessário para ser corrigido depois). O Algoritmo 9 não inclui a verificação se as posições atualizadas estão dentro de um espaço válido de endereçamento das matrizes usadas, já que é uma tarefa trivial.

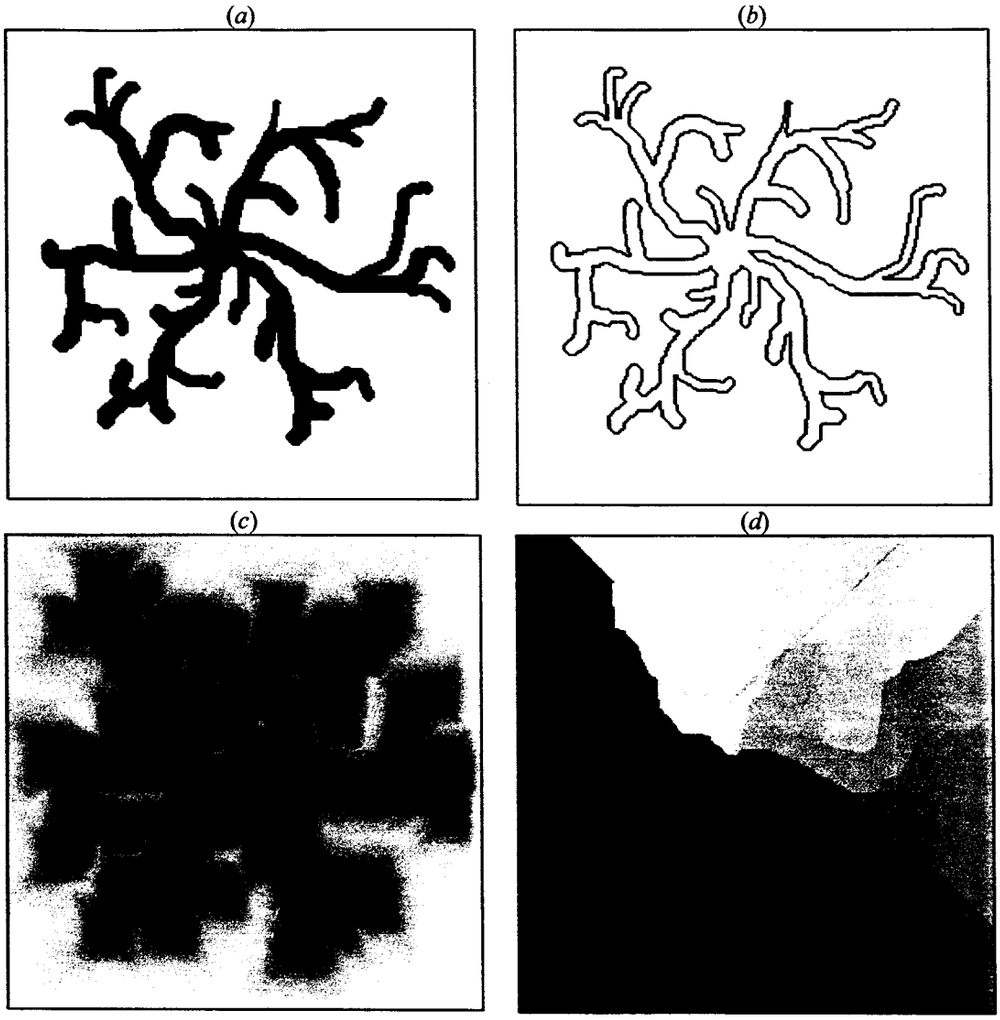


Figura 68 – Pré-processamento de uma forma para a esqueletonização multiescalas com métrica chessboard. (*a*) Forma original predominantemente ramificada; (*b*) contorno extraído; (*c*) transformada distância chessboard de *b*; (*d*) rótulos propagados para o contorno *b*.

A Figura 68 ilustra a abordagem de esqueletonização chessboard. A forma original mostrada na Figura 68*a* tem sua borda binária detectada na Figura 68*b*, que é rotulada (Figura 68*d*) perseguindo o contorno na direção horária. A imagem de diferenças apenas com o esqueleto interno é mostrada na Figura 69. Alguns dos esqueletos multi-escalas obtidos pela limiarização da imagem de diferenças para vários valores T de limiares, são exibidos na Figura 70.

Assim, uma nova e extremamente simples abordagem foi proposta para esqueletonização espaço-escalas e análise de formas foi apresentada. Esta metodologia adotou a distância chessboard substituindo a euclideana exata,

implicando um número de vantagens substanciais e algumas desvantagens. Primeiro, como a distância *chessboard* é muito simples, não há necessidade de usar uma estrutura de dados para armazenar os possíveis valores da transformada distância e os respectivos vetores posição. Como conseqüência, o algoritmo se torna bem mais simples, favorecendo extensões para implementações 3D e com paralelismo em hardware dedicado. Além disso, como o número de distâncias *chessboard* para um R específico é muito menor que o número de distâncias euclideanas, especialmente para R grande, o algoritmo é muito mais rápido. Por exemplo, o tempo de execução para obter as imagens de diferença do contorno da Figura 69 usando um código Delphi em uma máquina PC-compatível 450MHz, para R=50 pixels foi de 3.956 segundos para a distância chessboard e 11.454 segundos para a euclideana. A principal desvantagem da esqueletonização *chessboard* multi-escalas é a variação à rotações, especialmente para formas que apresentem grandes áreas interiores na forma original.

Figura 69 – Esqueleto interno multi-escalas da forma na Figura 68*a*.

A estrutura do esqueleto pode ser percorrida para representá-lo como um dendrograma. Este dendrograma permite examinar as distribuições de tamanhos de segmentos entre ramificações, de ângulos entre os respectivos segmentos. Estes

atributos são úteis na caracterização de neurônios ganglionares [COSTA & VELTE, 1999] ou para definir gramáticas a partir da forma de neurônios reais [COSTA et al., 1998]. O número de extremidades e de ramificações em determinadas escalas pode caracterizar também formas ramificadas.

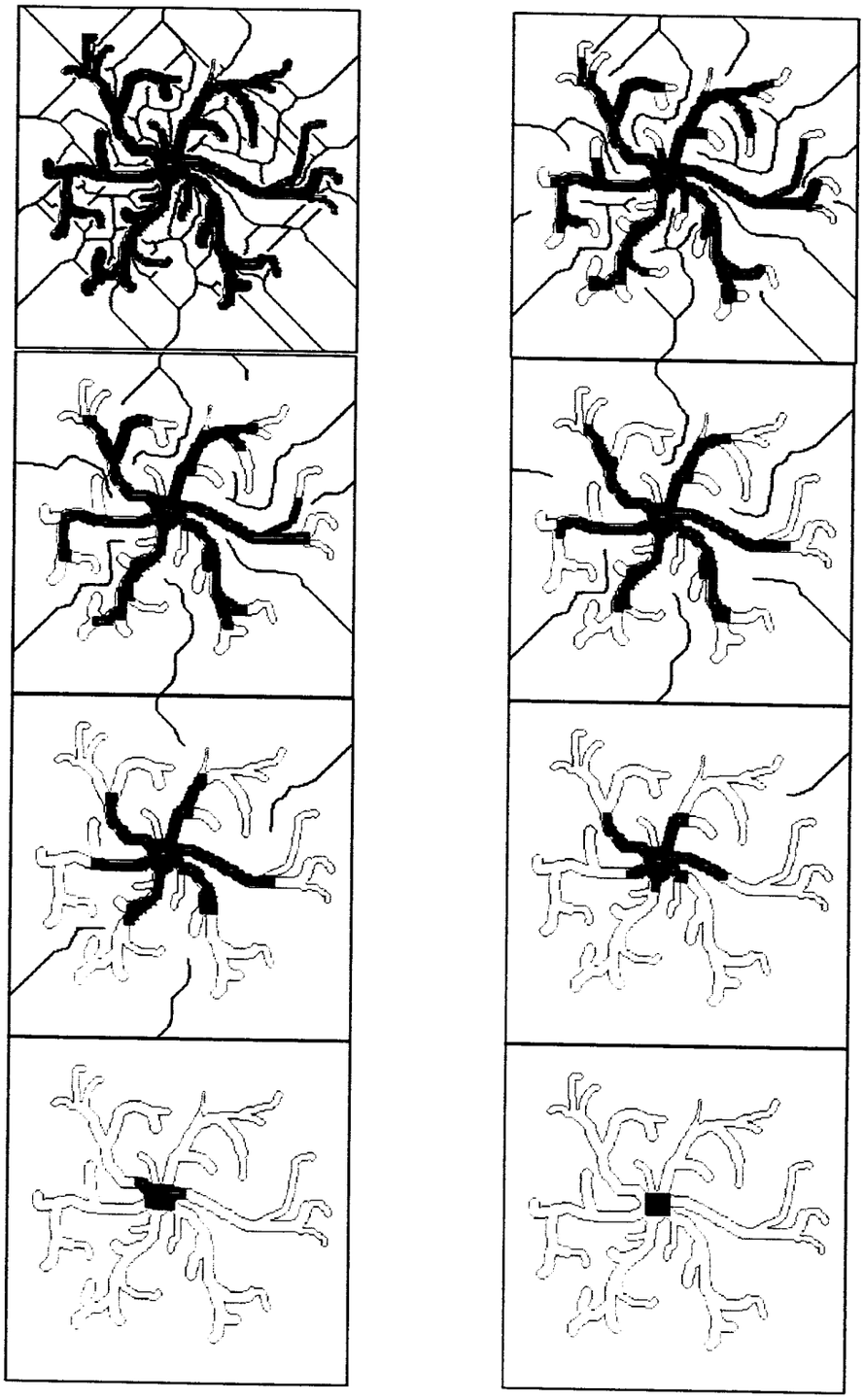


Figura 70 – Reconstrução (em preto) dos esqueletos (linha em branco) da forma da Figura 68*a*.

5 ANÁLISE DE FORMAS

Analisar algo é separá-lo em partes para posteriormente entendê-lo [PROCTER, 1995]. A análise de formas não é diferente. Sua estratégia é separar ou segmentar a forma para poder então descrevê-la analiticamente ou sintaticamente. A segmentação é, portanto, um processo fundamental que precede a cognição de um objeto. Há evidências que a visão humana utiliza esta mesma estratégia, porém com diversas descrições. O passo posterior, a descrição da forma, é bastante influenciado pela aplicação. No entanto, o compromisso entre a quantidade de símbolos em um repertório e a quantidade de informação a ser descrita está sempre presente na mente de quem projeta sistemas de visão. Para exemplificar, podemos pensar este problema como uma questão de hierarquia da representação. Em um nível hierárquico mais alto, poucos símbolos representam individualmente uma forma ou imagem complexa. A palavra face, por exemplo, resume uma grande complexidade, já que se pressupõe a existência de olhos, nariz, boca, testa, queixo, cabelo, pele, orelhas, etc. Em um nível hierárquico mais baixo, precisaremos de mais símbolos para descrever a face. Neste caso, a face pode ser composta por aquelas partes que formam-na e os locais corretos que, deste modo, definem um relacionamento sintático entre os símbolos.

Este é um problema estabelecido pelo princípio da mínima descrição (MDL). Contudo, para descrever formas é necessário estabelecer os símbolos com os quais a forma será descrita. Estes símbolos podem variar desde simples formas poligonais ou contornos lineares até aproximações polinomiais mais complexas como B-*splines*.

Uma vez estabelecidos os símbolos com o qual se descreverá a forma, esta será lida e sua descrição montada utilizando o "alfabeto" definido. Esta leitura

consiste em identificar os pontos que isolam os segmentos definidos como símbolo. Para caracterizar tais pontos é necessário um critério que estabeleça este limite. Um critério comumente utilizado é a simetria e assim, os pontos em que ocorre uma alteração na simetria definem a transição entre um símbolo e outro. Tais pontos são conhecidos como saliências, sejam elas protuberantes ou reentrantes.

A seguir será apresentada uma pequena revisão sobre saliência e sua relação com a percepção visual. Então, será dada uma visão geral sobre os métodos para detecção de saliências existentes na literatura e por fim, serão apresentados dois métodos para detecção de saliências: (1) baseado na generalização de parâmetros de curvatura por parâmetros globais; (2) baseado na perseguição de saliências em esqueletos multi-escalas.

5.1 Saliência

Quando se diz que algo é saliente, é intuitivo pensar em algo que se destaca e com importância. O conceito de saliência em formas é análogo, porém uma pergunta é imediata: Se destaca onde? Importante para quê? Percebe-se então que o conceito de saliência depende bastante do espaço em que a forma reside e ainda dos objetivos de se salientar determinados pontos. Podemos nos preocupar com a percepção da forma, ou então sermos orientados por alguma aplicação específica.

As teorias da percepção visual têm buscado elaborar modelos sobre como os estímulos vindos do mundo físico são codificados e elaborados no aparelho sensorial biológico. Tais modelos arbitram o conceito de uma boa imagem (*Gestalt*) ou inferem formas de seu ambiente (*Gibson*), ou ainda propõem estruturas hierárquicas sobre a cognição visual (*Primal Sketch*). A borda de uma forma, embora tratada de maneiras diferentes em cada modelo, sempre está presente neles [GORDON, 1989; LONCARIC, 1998].

Abordando a percepção fisiologicamente, Hubel e Wiesel [HUBEL & WIESEL, 1962] comprovaram a existência de células *end-stopped*, ou seja, que respondem bem a extremidades de linhas ou barras retangulares. Würtz and Laurens [WÜRTZ & LAURENS, 2000] modelaram tais células como operadores que detectam saliências em imagens coloridas.

Particularmente, no modelo de Marr [MARR, 1982], a borda é essencial na formação do *raw primal sketch*, que pressupõe sua análise e consequentemente sua segmentação que resultará na entrada para a representação. Tais representações devem considerar pontos críticos que separam segmentos significantes segundo paradigmas de particionamento estabelecidos conforme os objetivos da aplicação. Fischler e Bolles [FISCHLER & BOLLES, 1986] classificam tais paradigmas em: (1) detecção local (curvatura), (2) detecção global (descrições polinomiais ou splines), (3) pela confirmação de evidências (vocabulário de códons) ou (4) por simplificação recursiva (multi-escalas); além disso, apresentam experimentos psicofísicos justificando tal classificação.

Mehrotra e Nichani [MEHROTRA & NICHANI, 1990] definem saliência de uma maneira convencional: "saliência ou canto é o ponto de interseção ou junção entre duas ou mais bordas que tenham descontinuidades ao longo de uma linha reta". Percebe-se nessa definição que é necessário assumir as bordas como retas, pois isto permite que se conheça apenas a vizinhança de um dado ponto para classificá-lo como saliência ou não. Eles ainda argumentam que um ponto de saliência deve ter como atributos o ângulo e a orientação da saliência.

Outros autores se preocupam com os critérios necessários a uma boa descrição. Mokhtarian e Mackworth [MOKHTARIAN & MACKWORTH, 1986] listam alguns desses critérios: (1) que seja eficientemente computável; (2) que seja invariante à rotação, mudança de escala e translação; (3) que seja representável em vários níveis de detalhes; (4) que preserve a localidade e intensidade das alterações; (5) que seja unívoca;

Ilustrando a propriedade de mínima descrição de uma forma, Attneave [ATTNEAVE, 1954] propôs o clássico exemplo no qual 38 saliências ligadas por linhas retas descrevem uma Figura tão complexa como a de um gato dormindo (veja Figura 71). Asada e Brady [ASADA & BRADY, 1986], propuseram um vocabulário de cinco saliências em função do tipo de alteração da curvatura: cantos, junção suave, terminação, excentricidade e protuberância como um modelo para o *raw primal sketch* de Marr.

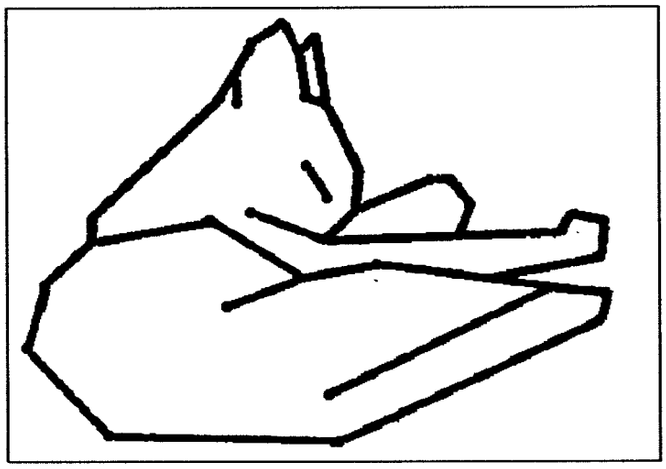


Figura 71 – Desenho de um gato dormindo elaborado utilizando 38 saliências [ATTNEAVE, 1954].

Em grande parte dos modelos de percepção, a saliência é entendida como um ponto crítico ou por uma singularidade na curvatura de uma forma [HOFFMAN & SINGH, 1997] ou então como um máximo local de uma medida de simetria [OGAWA, 1989]. A curvatura, no entanto, é um parâmetro local, de modo que um método de detecção de saliências utilizando simplesmente a curvatura teria seu desempenho comprometido para contornos constituídos de segmentos em escalas diferentes. Alguns trabalhos propõem medidas de simetria para a detecção de saliências [FREEMAN, 1978], porém tais medidas não consideram ainda uma variação nos níveis das escalas entre os segmentos.

É conhecida a dualidade entre eixos de simetria e a curvatura de uma forma [LEYTON, 1987]. Portanto não seria exótico utilizar o eixo de simetria para localizar as saliências de uma forma [OGAWA, 1989]. A vantagem do eixo de simetria está na possibilidade de usar uma abordagem multi-escalas na descrição da forma e consequentemente uma descrição não só local, mas também global da forma. Trabalhos como [OGNIEWICZ, 1994; COSTA & ESTROZI, 1999] têm proposto soluções multi-escalas, cuja evolução do eixo de simetria conforme uma escala causa um crescimento fazendo com que as extremidades evoluam conforme uma hierarquia. Esta hierarquia é dada em função de parâmetros globais como perímetro ou área.

5.1.1 Por que a saliência é importante?

Aplicações que envolvem atenção visual utilizam intensivamente a detecção de saliências, já que tais locais na imagem são reconhecidamente os que captam mais a atenção do sistema visual humano [ATTNEAVE, 1954; HOFFMAN & SINGH, 1997]. Assim, dado que a saliência é o elemento de maior interesse e como o tempo na perseguição de um alvo é um parâmetro crítico, um ganho considerável de tempo é possível quando apenas tais pontos de interesses são considerados. Esse aspecto econômico no tratamento de imagens em movimento é freqüentemente desejável em projetos de robótica, navegação veicular ou trajes computacionais. Grande parte destas aplicações exige a implementação em *hardware* dedicado, dando assim preferência a abordagens de visão que possibilitem um projeto compacto.

Muitas vezes, os objetivos de uma aplicação demandam a análise de uma forma. Essa análise isola as partes da forma que podem ser usadas como atributos ou como parte de um vocabulário para descrevê-la. O primeiro caso refere-se a medidas da forma que podem ser agrupadas ou classificadas com o intuito de caracterizá-la [BRUNO et al., 1998]. Já como vocabulário, a forma pode ser transformada em um grafo ou em uma sentença gramatical que possibilitaria, por exemplo, a elaboração de modelos funcionais [ASADA & BRADY, 1986; LEYTON, 1988].

5.1.2 Principais técnicas

Técnicas de detecção de saliências são abundantes na literatura científica [LI, 1996]. O que se percebe nesse conjunto de trabalhos é o compromisso entre a generalidade e desempenho da técnica, seja com relação ao tempo de execução, seja ao número de acertos. Tais técnicas podem ser divididas em duas classes: as que utilizam como entrada a imagem em níveis de cinza ou em cores [MEHROTRA & NICHANI, 1990; DELAGNES et al., 1995; KOHLMANN, 1996; LUO et al., 1999]; e as que utilizam o contorno de imagens já segmentadas. No primeiro caso, a vantagem é que a independência de um algoritmo de segmentação permite a economia de um passo que consome tempo de execução. Esta economia interessa muito para a implementação de sistemas de visão. Como contrapartida, muitos destes algoritmos de detecção são bastante sofisticados e computacionalmente complexos. Neste trabalho será dada atenção ao segundo caso, no qual a imagem está

IESC-USP SERVIÇO DE BIBLIOTECA

presumivelmente já segmentada e o contorno já representado. Os ruídos passam então a existir na borda, fazendo com que a imunidade a ruídos seja um requisito freqüentemente desejado e perseguido na literatura para algoritmos detectores de saliências.

Dentre as técnicas mais comuns na literatura estão aquelas que utilizam a descrição por *chain-code* [LIU & SRINATH, 1990; ARREBOLA et al., 1997; ARREBOLA et al., 1999]. Na verdade, tais técnicas se limitam a um cálculo discretizado da curvatura a partir da descrição em *chain-code*. Muitas vezes lançam mão de filtros discretizados para reduzir o ruído do contorno [ROSENFELD & JOHNSTON, 1973; RUTKOWSKI & ROSENFELD, 1978]. O maior problema nesse tipo de algoritmo, na verdade, é a definição de um limiar que resulte em saliências aceitáveis.

Koyama et al. [KOYAMA et al., 1980] propõem uma técnica que traça a corda entre um ponto e seus vizinhos. A medida que define a saliência é a taxa média entre o comprimento de arco entre cada combinação de dois pontos de um segmento e o tamanho da corda (ou distância euclideana) entre eles. O objetivo é obter uma medida de simetria local como a definida em [ASADA & BRADY, 1986].

Ainda na busca de uma medida local a partir do *chain-code*, Arrebola et. al. [ARREBOLA et al., 1997] desenvolveram um método cuja medida é dada pelo coeficiente de correlação (ρ) entre duas variáveis locais da curvatura calculadas a partir do *chain-code*. Para decidir se um ponto é uma saliência ou não, são tomados os pontos a partir de um limiar de ρ. Os autores destacam as propriedades de invariância à rotação, baixo tempo de execução e a estabilidade como os pontos fortes desse método. Mais tarde, Arrebola et.al [ARREBOLA et al., 1999] definem um novo parâmetro para escolha de saliências baseado na média circular do histograma de curvatura, alegando para isso uma redução no tempo de execução.

Medioni e Yasumoto [MEDIONI & YASUMOTO, 1987] ajustam uma B-spline cúbica e utiliza a diferença entre a curva e o contorno (em *chain-code*) resultando assim em uma medida de proeminência. Essa medida é submetida a um limiar que decide se o ponto é uma saliência ou não. Vê-se claramente uma tentativa de definir uma medida que considere as características globais do contorno.

Mokhtarian e Suomela [MOKHTARIAN & SUOMELA, 1998] propõem um método para detecção robusta de saliências. Eles utilizam o espaço-escala de curvatura em bordas extraídas utilizando o detector de borda de Canny. As saliências são perseguidas ao longo da "evolução" da escala desde as altas, até as mais baixas. A robustez é conseguida a um preço muito computacionalmente dispendioso. Além da extração da borda (80% do tempo de execução segundo os autores), as curvaturas devem ser calculadas nas escalas em que os pontos candidatos a saliências serão perseguidos.

Lee et al. [LEE et al., 1993] abordam o problema da detecção de saliências com escalas definidas por transformadas *wavelet*. Assim como Mokhtarian e Suomela [MOKHTARIAN & SUOMELA, 1998], o algoritmo busca por pontos de saliência (três tipos) estáveis ao longo do diagrama de escalas (curvograma). A vantagem desse método é a independência de parâmetros, já que as escalas são separadas em intervalos de oitavas. Os resultados ilustraram casos nos quais saliências em escalas muito pequenas são desprezadas revelando uma propriedade de imunidade a ruídos. O desempenho em termos de tempo de execução é provavelmente comprometido pela robustez, porém os autores não expuseram esse tipo de resultado.

Os algoritmos que utilizam um diagrama espaço-escala são mais robustos, porém computacionalmente custosos, pois a dimensão da escala é adicionada à busca pelas saliências. No entanto, essa busca pode ser treinada e o resultado desse treinamento aplicado a um novo conjunto de contornos de formas. Tsai [TSAI, 1997] utiliza duas redes neurais para treinar dois tipos de saliências de escalas diferentes. Uma delas treina saliências de alta curvatura, cujo conjunto de treinamento é formado por segmentos de curva com configurações angulares específicas. As entradas dessa rede são pontos da vizinhança do elemento do contorno e a saída é seu respectivo ângulo de curvatura. As entradas para a segunda rede são similares à primeira. As saídas são os pontos de tangência e de inflexão. A aplicação dos pesos como operadores em um contorno identifica seus pontos de junção.

5.2 Aprendizado de Pontos de Saliência por Generalização de Parâmetros

A curvatura de uma forma é uma propriedade local, ou seja, em cada ponto do contorno desta forma existe um valor que, para formas planares, corresponde ao valor inverso do raio do círculo tangente. Assim, uma região de contorno suave, ou seja, com um círculo tangente cujo raio de curvatura é grande reflete uma curvatura pequena. Já uma região cujo círculo tangente tem raio pequeno corresponde a uma saliência em potencial. Conforme esse raio tende a zero, mais saliente é a região. Quando o raio é nulo, uma singularidade ocorrerá na curvatura, caracterizando no contorno a ocorrência de quebras ou cantos vivos. Portanto, a curvatura é uma característica local do contorno que, a partir de limiares pode ser usada para detectar saliências.

Um contorno digital, por sua vez, é dado por um conjunto de pontos de intersecção em uma grade amostrada em um espaço contínuo. A avaliação da curvatura demanda o uso de derivadas e estas não são definidas em pontos de descontinuidade. Portanto, o cálculo da curvatura em contornos digitais é um problema mal-condicionado. Porém, a suavização desse contorno permite utilizar as propriedades da curvatura de um contorno contínuo estimada a partir de um contorno digital [CESAR JR & COSTA, 1995].

Percebe-se portanto que, para detectar saliências usando a curvatura de um contorno são necessários pelo menos dois parâmetros: um limiar e um fator que indique a suavização desse contorno. A aplicação de um simples limiar em um contorno suavizado acarretará no reconhecimento de conjuntos de pontos candidatos à saliência. É necessário selecionar em cada um destes conjuntos aquele mais adequado, ou seja, o de máxima curvatura.

Detectar saliências em um contorno digital é uma tarefa que se torna subjetiva, já que depende da seleção da escala de aproximação. Por exemplo, pixels são normalmente aproximados por quadrados. Portanto, um contorno digital observado de perto é um conjunto sucessivo de saliências. A seleção da escala é, na verdade uma escolha subjetiva de um especialista que determina as saliências que lhe interessam, ignorando as irrelevantes ao seu propósito.

Se o objetivo de uma aplicação é imitar o comportamento da seleção de saliências de um especialista, torna-se necessário encontrar uma combinação de

parâmetros que aproxime este comportamento. Isto é viável com a construção de um algoritmo que possa ser treinado para escolher estes parâmetros.

A aplicação descrita a seguir serve como uma ilustração do processo de aprendizado de seleção de saliências. O objetivo da aplicação é detectar automaticamente terminações e ramificações em formas de células neurais ganglionares [CONSULARO et al., 1999].

5.2.1 Aplicação: Forma de Neurônios

Um importante desafio comumente encontrado na análise de formas neurais é descrever a estrutura de ramificação de dentritos em neurônios. Tais descrições podem ser úteis para diversos processos envolvidos com simulações da relação entre a forma e a função neural [POZNANSKI, 1992; COSTA et al., 1998].

Muitos esquemas úteis para descrever dendritos são baseados na identificação de duas classes importantes de pontos ao longo dos dendritos: as terminações e as ramificações. Por exemplo, os dendrogramas neurais podem ser obtidos diretamente a partir destes pontos por um processo gramatical (*parsing*) [CESAR JR & COSTA, 1997a; POZNANSKI, 1992].

Um algoritmo efetivo para determinar os pontos de terminação e ramificações pode ser implementado baseado na estimativa da curvatura ao longo do contorno que delimita a célula neural. Uma vez estimada a curvatura baseada em técnicas de visão computacional, as terminações e ramificações podem ser buscadas como pontos de máximos ou mínimos locais de curvatura (veja Figura 74) [CESAR JR & COSTA, 1995].

Infelizmente, a estimativa de curvatura envolve o cálculo das primeira e segunda derivadas da representação parametrizada e discretizada do contorno, que é um probema mal-condicionado e depende da suavização da curva (no caso do contorno ser um sinal discreto). Esta necessidade de suavização do contorno enfatiza a importância da curvatura multi-escalas na segmentação e análise de formas. Neste sentido, o resultado final da análise da forma depende criticamente do grau de suavização (isto é, da escala) adotada.

O objetivo da aplicação é encontrar parâmetros que aplicados à curvatura possam identificar ramificações e extremidades com desempenho próximo de um

especialista. Para isso, um conjunto de silhuetas de neurônios foi dividio em dois grupos: um de treinamento e outro de testes. Para cada silhueta do primeiro foram calculadas suas respectivas curvaturas e duas medidas globais sobre a curvatura: coeficiente de variação (CV_k) e energia de dobramento (B_k). Foi também pedido a um especialista que marcasse os pontos de ramificação e extremidades de cada forma neural. Uma superfície interpolando o espaço de medidas globais de curvatura serviu como modelo para generalizar os parâmetros de segmentação (limiares e suavizações). A marcação do especialista atuou como referência para comparar a marcação obtida pela aplicação dos parâmetros do modelo. Foi construída então uma função objetivo cujo valor máximo é 1. Este valor é atingido quando um máximo de acertos e um mínimo de erros é alcançado. Um algoritmo genético foi utilizado para buscar as configurações de parâmetros de segmentação da curvatura, cuja função objetivo mais se aproximasse de 1. O desempenho foi medido com o conjunto de testes sendo submetido ao modelo também. Seus erros servem como uma medida da qualidade do treinamento. (veja Figura 72)

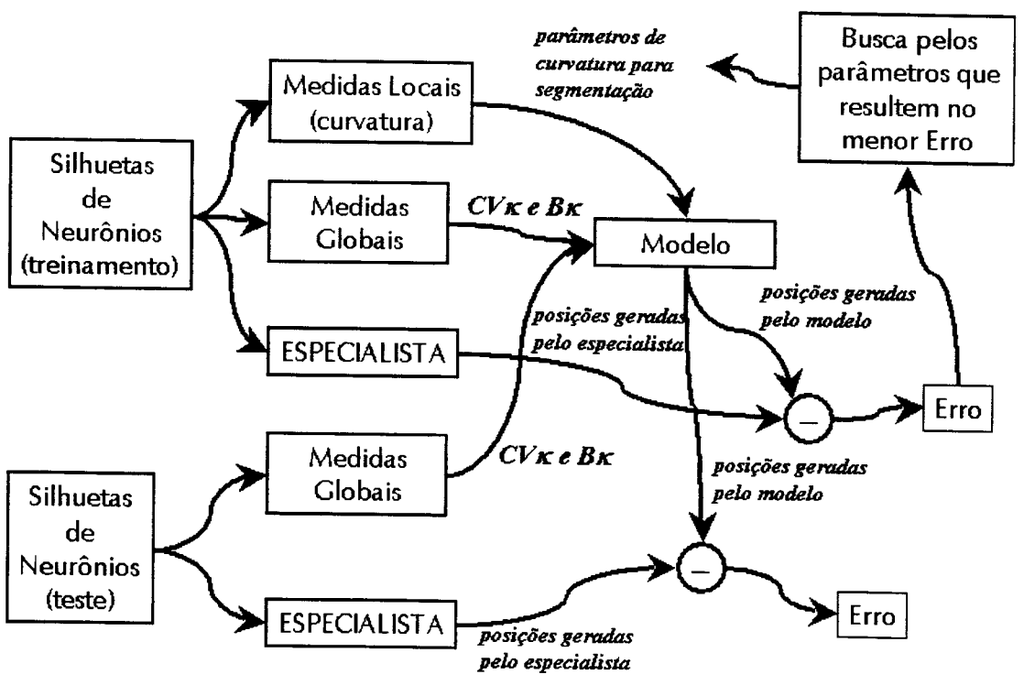


Figura 72 – Esquema ilustrando a aplicação de calibração de parâmetros de segmentação utilizando a curvatura de formas neurais.

A seguir serão detalhadas as etapas desse processo de treinamento e calibração dos parâmetros de segmentação.

1) Quais são os critérios para segmentação de contornos em imagens de células neurais?
2) A curvatura é uma abordagem eficiente para a segmentação de contornos com respeito a percepção humana?
3) Quais estratégias deveriam ser adotadas para buscar o melhor conjunto de parâmetros de curvatura com o intuito de obter resultados otimizados para a segmentação de contornos?

Para responder à primeira questão, um sujeito humano foi submetido a um experimento o qual foi inquirido a marcar pontos de ramificações e terminações em um conjunto de 39 neurônios (veja Figura 73). Estes pontos foram armazenados em uma tabela contendo as coordenadas de cada ponto e sua respectiva identificação. As coordenadas variaram dentro de uma imagem de 256 por 256 pixels.

O objetivo deste experimento é comparar uma segmentação obtida por um humano com a segmentação automática baseada nos parâmetros da forma descritos e justificados adiante. A estratégia adotada para selecionar os melhores parâmetros, conforme descrito na seção 3.3.6, foi o algoritmo genético (GA). Sua escolha foi orientada pelas características do problema:

1) o algoritmo genético não requer conhecimento anterior sobre o espaço de busca;
2) as descontinuidades no espaço de busca influem pouco no processo de busca;
3) permite uma certa robustez quanto ao problema de extremos locais;
4) apresenta, em geral, bom desempenho em grandes espaços de busca.

É importante, no entanto, ressaltar os possíveis problemas para o uso do algoritmo genético:

1) não é garantido que encontre um ótimo global;
2) o cálculo da função objetivo pode ser computacionalmente intensiva;
3) depende fortemente da configuração de parâmetros envolvidos.

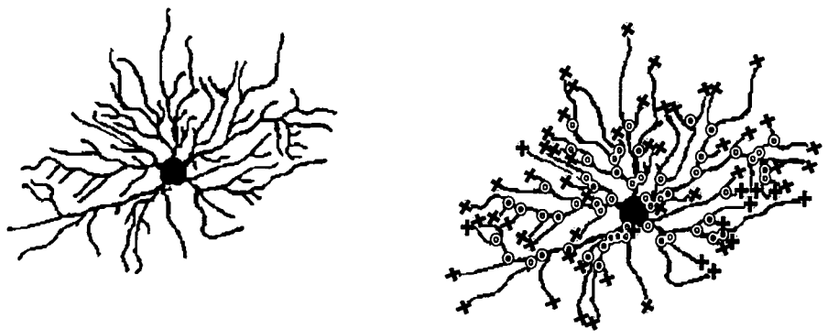


Figura 73 – Exemplo de forma neural ganglionar (a) e de marcação de ramificações (o) e extremidades (x)

Medidas Locais

Esta abordagem para análise de forma neural é baseada na representação da célula neural em termos de seu contorno. Tal representação é obtida por uma curva paramétrica (59), para a qual t corresponde à parametrização do tamanho de arco.

$$C(t) = \{x(t), y(t)\} \tag{59}$$

Uma representação complexa do contorno pode ser obtida definindo um sinal complexo $u(t)$ conforme (60).

$$u(t) = x(t) + iy(t) \tag{60}$$

As propriedades da transformada de Fourier permitem uma estimativa da descrição da curvatura multi-escalas (curvograma) de C, possibilitada pela aplicação de (61).

$$k(t,a) = \frac{-\operatorname{Im}\{\dot{u}(t,a)\,\ddot{u}^*(t,a)\}}{|\dot{u}(t,a)|^3} \tag{61}$$

Para (61), $\dot{u}(t)$ e $\ddot{u}(t)$ correspondem aos sinais diferenciados e filtrados na escala a, conforme descrito detalhadamente em [CESAR JR & COSTA, 1997b].

Uma configuração de pontos de ramificações e terminações dos dendritos é obtida pela seleção de três parâmetros: (1) suavização da curvatura na escala a; (2) um limiar para valores positivos de curvatura T^+; (3) um limiar para valores negativos de curvatura T^-. Assume-se que o parâmetro T cresce no sentido anti-horário. A Figura 74 mostra os efeitos da aplicação de cada um dos parâmetros no contorno da silhueta.

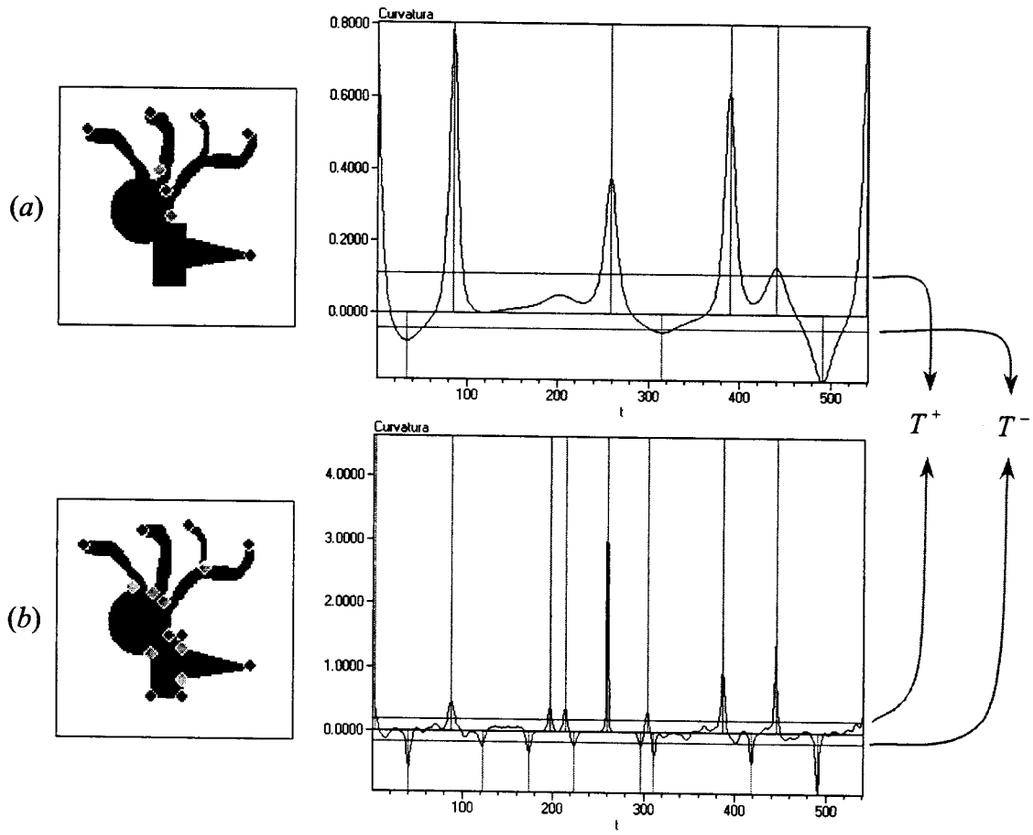


Figura 74 – Parâmetros utilizados para calibrar a curvatura: T^+ para terminações ou extremidades e T^- para ramificações. (*a*) ilustra uma escala $a = 28$ e (*b*) em uma escala $a = 3$.

Para as formas de neurônios, a segmentação de dendritos depende da escala do curvograma e dos limiares da curvatura para determinar as terminações e as ramificações. Além disso, é comum que porções negativas de curvatura apresentem magnitudes significativamente diferentes, justificando também a equalização da curvatura por uma sigmóide (63).

O objetivo deste trabalho é calibrar os parâmetros de curvatura usando os critérios de percepção humana para extremidades e ramificações. Para isso, foi desenvolvido um algoritmo para extrair a representação complexa $u(t)$ de uma silhueta. As imagens de neurônios, para efeito de normalização, foram padronizadas em 256 × 256 pixels. A curvatura foi estimada pela aplicação de (61). A abordagem multi-escalas é implementada por convoluções gaussianas com o contorno, cujo desvio padrão σ representa o parâmetro a.

Na Figura 74 as suavizações são ilustradas nas escalas $a = 28$ (a) e $a = 3$ (b). Em escalas maiores percebe-se uma maior suavização do contorno e consequentemente, da curvatura.

O grau de suavização deve ser distinto para terminações e ramificações, porque estas estruturas podem ocorrer em diferentes escalas espaciais. Por exemplo, normalmente para um contorno as terminações apresentam valores absolutos de curvatura altos, enquanto os pontos de ramificação valores absolutos baixos. Foram adotadas duas escalas: uma para terminações (a^+) e outra para ramificações (a^-).

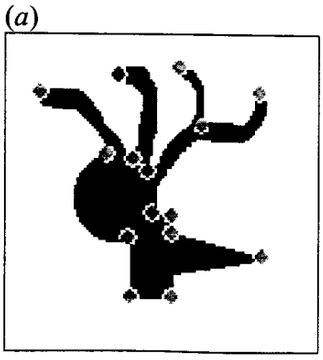
O próximo passo na detecção dos pontos de segmentação é o limiar da curvatura e a busca por pontos de máximo local para terminações e de mínimos locais para ramificações. Cada um com um limiar correspondente a extremidades (T^+) e a ramificações (T^-).

Um outro parâmetro é aplicado sobre a curvatura para equalizá-la, isto é, atenuar os valores altos de curvatura. Isto é conseguido pela aplicação de uma função sigmóide sobre a curvatura estimada.

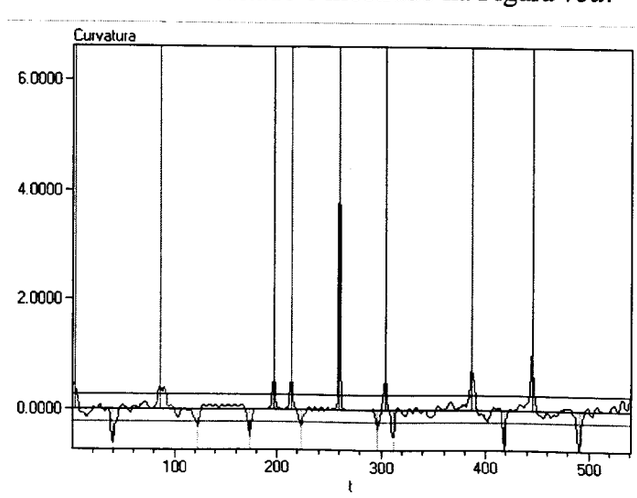
$$S(k) = \frac{2}{1+e^{-sk}} - 1 \qquad (63)$$

Os efeitos da aplicação da função sigmóide na curvatura são visualizados na Figura 75. Observa-se que a curvatura é re-escaladal, reduzindo as grandes diferenças entre os extremos de curvatura. Como pode ser visto na Figura 75*b*, as extremidades residem em uma escala maior que as ramificações. A aplicação da sigmóide serve também para deixar as escalas de extremidades e ramificações iguais. Na Figura 75*c* está a função sigmóide aplicada na curvatura e o resultado é mostrado na Figura 75*d*.

(*a*)



(*b*)



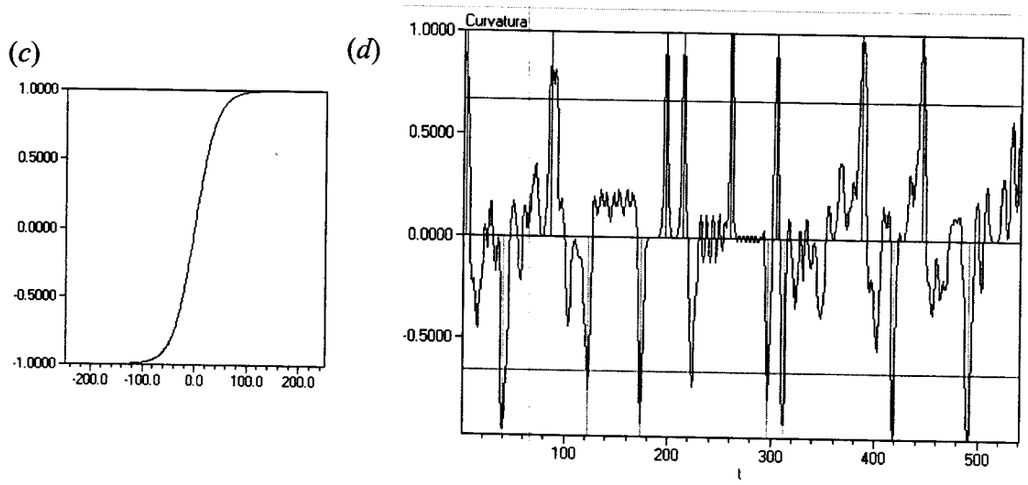


Figura 75 – Efeito da aplicação da função sigmóide na curvatura. (*a*) contorno segmentado; (*b*) curvatura sem aplicação da sigmóide; (*c*) a função sigmóide com k=0.5; (*d*) curvatura com aplicação da sigmóide.

Medidas Globais

Há vários tipos de parâmetros globais para caracterizar a forma. As recomendações para seleção de descritores globais [CASTLEMAN, 1996] implicam quatro propriedades desejáveis:

1) fracamente correlacionados para reduzir a redundância;

2) alta variância, maximizando a abrangência de formas nas classes;

3) similaridade equidistante entre os descritores;

4) redução de dimensão buscando combinação de descritores, se possível.

Três medidas globais foram consideradas: a energia de dobramento (B), a dimensão fracionária (df) e o coeficiente de variação da curvatura (CV_k).

energia de dobramento (B)

A energia de dobramento de um contorno foi introduzida em [YOUNG et al., 1974] e aplicada para descrever formas biológicas. Esta medida é definida como a integração do quadrado da curvatura κ (64) e tem sido adotada como uma medida de complexidade de forma [CASTLEMAN, 1996] e mais especificamente caracterizando formas de células neurais [CESAR JR & COSTA, 1997b]. Pode-se fazer uma associação física entre um contorno e um fio de arame ideal, por exemplo. Assim, B seria a energia que deve ser dispendida para dobrar uma forma circular (a

forma com menor energia) e o obter a forma em questão. Como nesse trabalho, as formas são apresentadas em diversas escalas a, em cada uma delas o contorno terá uma energia de dobramento diferente, isto é, B_a.

$$B_a = \frac{P_a^{\,2}}{N}\sum_{n=1}^{N}\kappa_a[n] \qquad (64)$$

O perímetro do contorno P_a é, na verdade, o perímetro da aproximação poligonal da forma em questão (65) representada por parametrizações X_a e Y_a do contorno na escala a.

$$P_a = \sum_{n=1}^{N}\sqrt{(X_a[(n \bmod N)+1]-X_a[n])^2+(Y_a[(n \bmod N)+1]-Y_a[n])^2} \qquad (65)$$

<u>coeficiente de variação</u>

O coeficiente de variação da curvatura é uma medida relativa de dispersão e teve sucesso quando combinado com a energia de dobramento B. Esta combinação apresentou uma taxa de correlação mais baixa que as combinações nas quais participavam a dimensão fracionária.

$$CV(\kappa_a) = \frac{\sigma_{\kappa_a}}{|\mu_{\kappa_a}|} \qquad (66)$$

Um alto desvio padrão de curvatura indica uma forma mais complexa. A curvatura média reflete a existência de mais convexidade (média positiva) ou mais concavidade (média negativa). Podemos concluir que o coeficiente de variação implica em um compromisso entre a existência de muitos detalhes e a convexidade da forma.

<u>dimensão fracionária</u>

A dimensão fracionária, conforme apresentada na seção 6.1.3.3, é uma medida simples de complexidade e sua quantidade é influenciada pela cobertura espacial de uma forma. Alguns autores aplicaram-na para classificar imagens de células neuronais [CASERTA et al., 1995].

5.2.2 Aprendizado

Contornos desempenham um papel importante na visão por computador e humana. A visão humana realiza tarefas como segmentação de imagens e extração de contornos de modo bastante eficaz. Contudo, tais tarefas são difíceis para visão computacional e processamento de imagens. Várias e diferentes abordagens vêm sendo desenvolvidas pela comunidade de visão por computador com resultados promissores, principalmente quando as condições de aquisição podem ser controladas. Uma vez segmentada a imagem, cada região pode ser representada pelo seu contorno externo. Muitos problemas importantes de visão podem ser resolvidos pela caracterização do contorno destas regiões limites. Neste trabalho, acreditamos que o problema de segmentação de imagens possa ser resolvido, e que o contorno de interesse pode ser extraído efetivamente. Mais especificamente, este trabalho trata da análise de contornos de neurônios para aplicações em neurociência. Para isso, uma abordagem sinergética para resolver o problema de segmentação, isto é, a detecção de terminações e ramificações em células neurais é descrita. A segmentação de contornos depende de um conjunto de parâmetros cujos valores variam entre células neurais com formas significativamente diferentes. Este trabalho apresenta uma nova abordagem para ajuste dos parâmetros de limiarização, suavização e equalização. Neste passo, os parâmetros são calibrados para um conjunto de treinamento de contornos por um algoritmo genético baseado na segmentação definida por um especialista humano. Assim, um conjunto de parâmetros otimizados é obtido para cada contorno neural do conjunto de treinamento. Então, na segunda fase, duas medidas globais da forma são calculadas para cada neurônio do conjunto de treinamento. Sejam g_1 e g_2 os dois parâmetros globais e seja p_i o i-ésimo parâmetro para um dado contorno neural. A fase de treinamento fornece um mapeamento $p_i = f(g_1, g_2)$ para todos os neurônios do conjunto de treinamento. Querendo então segmentar uma nova entrada de contorno neural, deve-se calcular as medidas globais desta nova entrada e cada parâmetro será obtido interpolando os valores de f_i para cada parâmetro i.

Generalização

O objetivo da generalização é obter os parâmetros de segmentação da curvatura a partir de parâmetros globais. Isto é possível com a aplicação de uma interpolação de superfícies. O método de interpolação aplicado [SANDWELL, 1987] é baseado na "*thin-plate spline*" introduzida por Powell [POWELL, 1994]. Cada um dos cinco melhores parâmetros de segmentação obtidos no treinamento, neste caso com 20 contornos, foi plotado versus os parâmetros globais e cinco superfícies foram interpoladas, uma para cada parâmetro de curvatura: escala e limiar para terminações, escala e limiar para ramificações e sigmóide. (veja Figura 76).

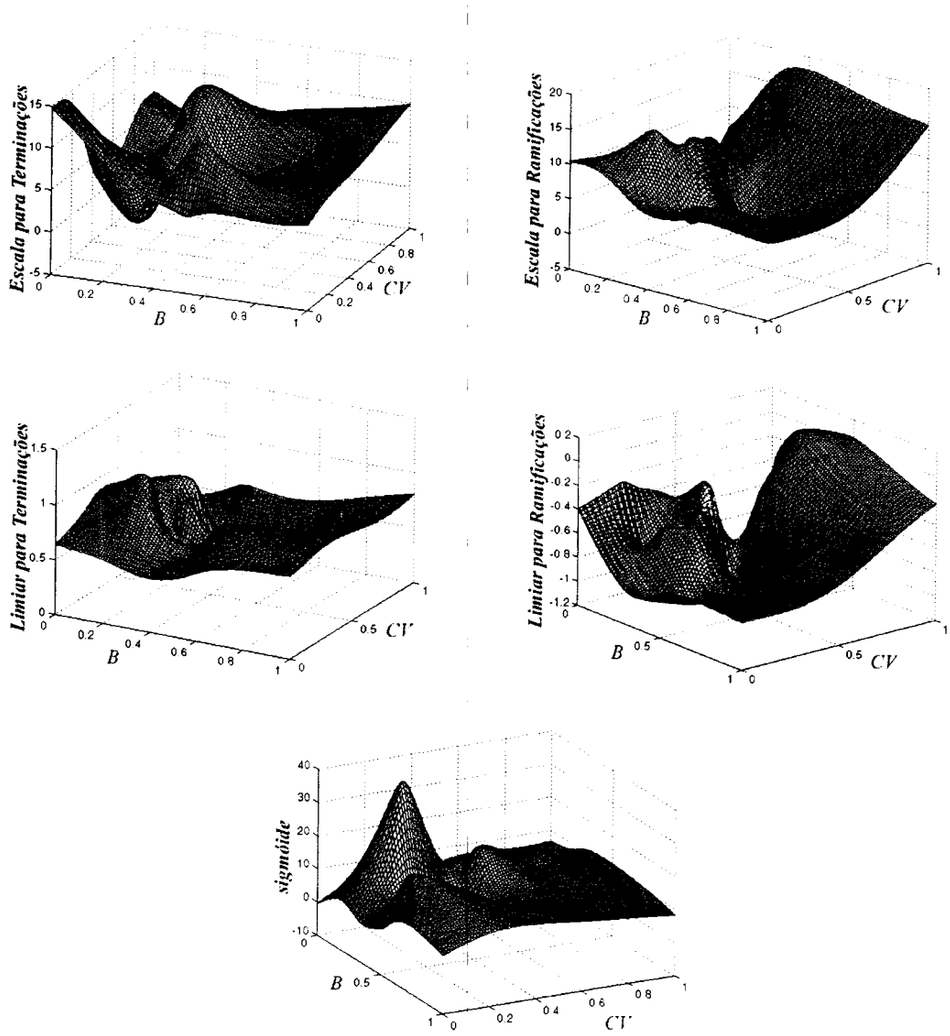


Figura 76 - Generalização dos parâmetros de curvatura por interpolações abrangendo os parâmetros globais (energia de dobramento – B versus coeficiente de variação – CV) e os parâmetros de segmentação: **(a)** escala para terminações (a^+); **(b)** escala para ramificações (a^-); **(c)** limiar para terminações (T^+); **(d)** limiar para ramificações (T^-); **(e)** sigmóide (S).

Treinamento

Algoritmos genéticos (GAs) são métodos de otimização inspirados nos mecanismos de seleção natural [GOLDBERG, 1989]. Eles foram propostos por John Holland e colaboradores [HOLLAND, 1962] e desde então tem sido aceito devido a sua robustez e aplicabilidade em diversos problemas. Embora muitas variantes têm sido propostas, apenas algumas abordagens têm sido matematicamente provada [MITCHELL, 1996].

Uma implementação simples foi desenvolvida para este trabalho para o qual cada gene foi definido como um número de ponto flutuante e associado a um parâmetro específico. Portanto, cinco genes foram agrupados em um cromossomo (ou indivíduo). Uma população de cromossomos foi criada e inicializada aleatoriamente com cada gene sendo codificado como um valor real entre 0 e 100. Embora alguns algoritmos genéticos considerem algumas tendências para a inicialização, devido ao desconhecimento do espaço de busca, este trabalho não utilizou este tipo de artifício. Cada cromossomo corresponde a uma configuração de parâmetros que permite avaliar uma função-objetivo que maximiza acertos e minimiza os erros. Deste modo, a função-objetivo para um cromossomo c em (67) foi adotada, tendo como máximo o valor um. Este valor é alcançado quando todos os pontos marcados pelo sujeito humano corresponderem aos da aplicação do conjunto de parâmetros e nenhum ponto dessa aplicação ocorrer fora do conjunto de pontos marcados pelo sujeito humano. O cálculo da função-objetivo é realizado para cada cromossomo da população que em seguida é ordenada para passar por um processo de seleção. O próximo passo é evoluir uma nova população $P(t+1)$ a partir da $P(t)$ corrente. Esta evolução é feita aplicando-se três tipos de operadores: mutação, que nessa implementação consiste da adição de um ruído gaussiano em um gene aleatoriamente selecionado; crossover, trocando os genes entre a partir de um dado gene aleatoriamente selecionado; e seleção, determinando a política de permanência ou descarte de indivíduos para a próxima geração $t+1$ com o intuito de garantir a diversidade. Novas gerações são continuamente submetidas a evoluções até que a solução ótima ou pelo menos uma razoável seja alcançada.

$$f_{objetivo}(c) = \frac{acertos - erros}{número\ de\ pontos\ marcados\ pelo\ sujeito\ humano} \qquad (67)$$

Para esta aplicação, cada gene corresponde aos parâmetros de operadores aplicados à segmentação de contorno de imagens de neurônios, ou seja, as duas escalas, uma para terminações (a^+) e outra para as ramificações (a^-); os dois limiares T^+ e T^-; e ainda o índice da função sigmóide S. Indivíduos são inicializados pela atribuição aleatória de valores entre 0 e 100. O número de indivíduos, isto é, a população utilizada foi de 40 (quarenta) indivíduos. A taxa de mutação, isto é, a probabilidade de selecionar um gene que mutará foi 0,1. A taxa de crossover é a probabilidade de ocorrer crossover em um ponto de quebra específico. O valor adotado foi 0,8 considerando todos os genes da população. Quanto à política de permanência para a seleção, o elitismo foi adotado para manter o cromossomo com melhor função-objetivo na população, garantindo deste modo que a melhor função-objetivo nunca decresça ao longo de gerações.

Resultados

Dezenove neurônios distintos foram usados para testes. Seus parâmetros globais foram avaliados e as medidas obtidas foram usadas para indexar cada superfície interpolada. Assim, cinco parâmetros de curvatura foram estimados e aplicados para extrair as terminações e as ramificações.

Esta função-objetivo foi calculada para ambos: terminações e ramificações (67), apenas para terminações (b) e apenas para ramificações (c). A função-objetivo viabilizou um modo de avaliar o desempenho das aplicações dos parâmetros.

Apesar do número de neurônios treinados e testados não ser grande o suficiente para descrever uma classe ampla de imagens, o método poderia ser aperfeiçoado não apenas para neurônios, como também um grande número de formas biológicas.

A eficiência sobre a segmentação de contorno mostrada pela abordagem descrita poderia ser conseguida aplicando outras técnicas de aprendizado, contudo a combinação de parâmetros globais e locais resulta em uma rápida segmentação de contorno. Além disso, a descoberta de conhecimento traduz as características inerentemente não lineares da percepção para a máquina, permitindo construir sistemas para classificação e descrições sintáticas orientados por critérios de percepção.

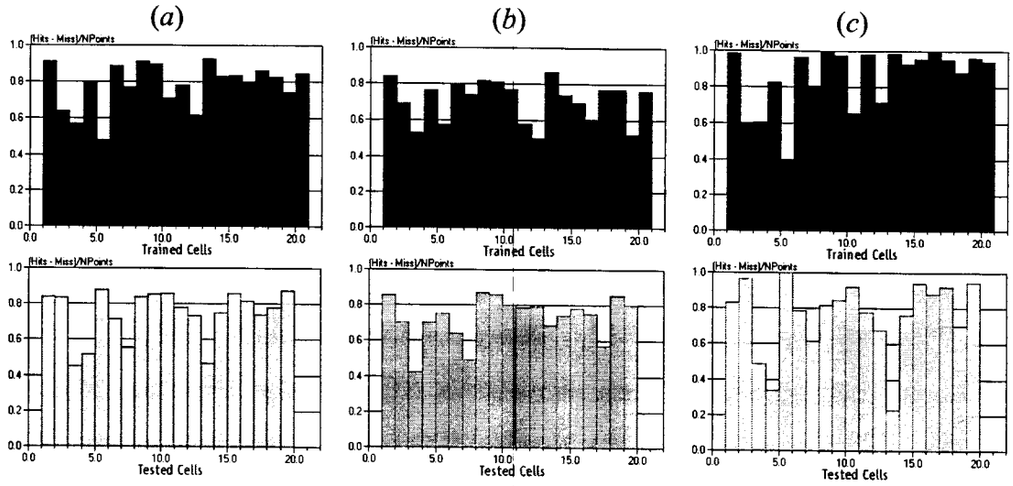


Figura 77 - Resultados conseguidos para os neurônios treinados (em preto) e neurônios de teste (em cinza). As escalas vão até 1, ou seja, o maior número de acertos e o menor número de erros. Os gráficos de barras em (*a*) são para ambos: terminações e ramificações; (*b*) apenas para terminações; e (*c*) apenas para ramificações.

Seria um desenvolvimento natural considerar um fluxo contínuo para imagens, parâmetros globais e parâmetros de treinamento adicionando seletores de escala para contornos mais detalhados e mais simples. Este fluxo descobriria conhecimentos continuamente sobre novas imagens e seus parâmetros inseridos.

Um pré-processamento de seletores de descritores poderia ser implementado permitindo uma seleção automática de atributos. Atualmente, isto pode ser conseguido pelo coeficiente de correlação entre os parâmetros globais, mas os agrupamentos (*clustering*) podem ser uma opção viável.

Apesar das limitações expostas, o método apresentado mostra potencial para bom desempenho em segmentação de contornos.

5.3 Perseguição de pontos significativos

No trabalho anterior foi explorada a curvatura de um contorno para extrair suas saliências. Para que o método conseguisse uma certa robustez, pelo menos para um grupo de formas com propriedades similares, foi necessário generalizar e treinar operadores de limiarização e suavização por medidas globais. Percebe-se que dois parâmetros de suavização, seriam úteis para formas ramificadas ou que tivessem grande parte das saliências de um mesmo tipo em uma mesma escala. Formas que apresentassem muitas escalas diferentes gerariam um problema de segmentação o qual simples limiarizações não resolveriam.

Um método robusto deve permitir selecionar escalas de detalhes em uma forma com parâmetros simples de serem ajustados.

Um método de esqueletonização multi-escalas foi desenvolvio em [COSTA & ESTROZI, 1999]. Este método permite a seleção das escalas de detalhe do esqueleto interno e externo de modo bastante elegante e robusto com relação ao nível de detalhe. Assim, explorando as propriedades do esqueleto resultante do método propõe-se um método para extrair as saliências de um contorno.

Os esqueletos obtidos em [COSTA & ESTROZI, 1999] possuem duas outras propriedades importantes para detectar saliências: (1) largura de um pixel; e (2) topologicamente conectado. Estas propriedades permitem identificar morfologicamente os pontos de terminações e de ramificações de um esqueleto; e que em qualquer escala o esqueleto não se quebrará mantendo a topologia da forma.

A extração de esqueletos multi-escalas a partir da métrica euclideana exata proposta por Costa [COSTA & ESTROZI, 1999] é computacionalmente intensiva, embora métodos rápidos possam ser considerados para a propagação dos rótulos [CUISENAIRE & MACQ, 1999]. Assim, foi proposto um modo de delimitar a região de esqueletonização para contornar este problema.

O método de esqueletonização é apresentado com mais detalhes na seção 4.2.2. Um conjunto de formas com características que dificultam os métodos tradicionais para detecção de saliências é testado e seus resultados ilustrados. Uma análise do desempenho computacional é apresentada e discutida logo após.

A representação do esqueleto em multi-escalas é de enorme importância, já que desvenda a estrutura de formação de uma dada forma. Estruturas salientes, por exemplo, ocorrem nas escalas mais baixas. No entanto, três problemas inviabilizam uma detecção de saliências utilizando o esqueleto multi-escalas: (1) quando tratamos de formas amostradas, a escala mais baixa indica a geometria da grade em que a forma está amostrada; (2) existe uma hierarquia na escala das ramificações que deve ser reconhecida ou analisada; (3) o alto custo computacional para calcular o esqueleto multi-escalas. O problema em (1) pode ser resolvido pela hierarquia reconhecida em (2), ou seja, esqueletos que surgem devido à amostragem são hierarquicamente inferiores. Ainda assim, resta o problema em (3) a ser resolvido. Para resolvê-lo, apenas uma faixa do domínio distante de um raio r da borda será propagada. Isso

permite uma redução drástica no tempo de execução da extração do esqueleto multi-escalas, primeiro porque as distâncias euclideanas menores existem em menor número de posições do que nas maiores. Segundo, porque a própria redução no número de elementos de interesse no domínio eliminará a necessidade de propagação até esses elementos.

O reconhecimento da hierarquia pode ser implementado através de uma descida a partir das escalas mais altas até as mais baixas, levantando uma espécie de dendrograma do esqueleto. Contudo, para se detectar as saliências, é necessária apenas uma perseguição que despreze ramos de hierarquia mais baixa, preserve o caminho das hierarquias mais altas e ainda bifurque quando encontrar dois ramos com hierarquias iguais. Esse algoritmo é descrito com maiores detalhes na próxima secção.

A imagem que é tomada como ponto de partida para a perseguição das saliências é obtida através do processo descrito a seguir.

5.3.1 Perseguição à saliência

O método descrito para propagação permite que os esqueletos sejam calculados apenas em uma faixa do domínio, ou seja, em uma região distante de um dado raio r da borda da forma. Assim, o tempo de execução depende diretamente do número de elementos da borda (N), do raio r, e da complexidade da forma também.

O processo se inicia na construção da estrutura de dados SD, limitada apenas às posições cujas distâncias são menores que r. Um passeio de SD sobre a borda da forma define uma região A, que compreende os pontos no domínio cuja distância à borda B é menor que r.

A seguir, é construída uma matriz de rótulos M_L, cujos valores fora de A são rotulados com ∞, e os interiores a A são preenchidos conforme o Algoritmo 8, porém restritos a $d < r$. A matriz obtida é ilustrada na Figura 4(a). Ainda utilizando SD, cada posição estritamente interior a A, isto é a região $A^* = A - B_A$, (B_A é a borda da região A) é preenchida com a máxima diferença entre cada um de seus vizinhos e o próprio rótulo, conforme descrito na seção anterior. O resultado é o esqueleto contido na região A. (Figura 78*b*).

(*a*)

(*b*)

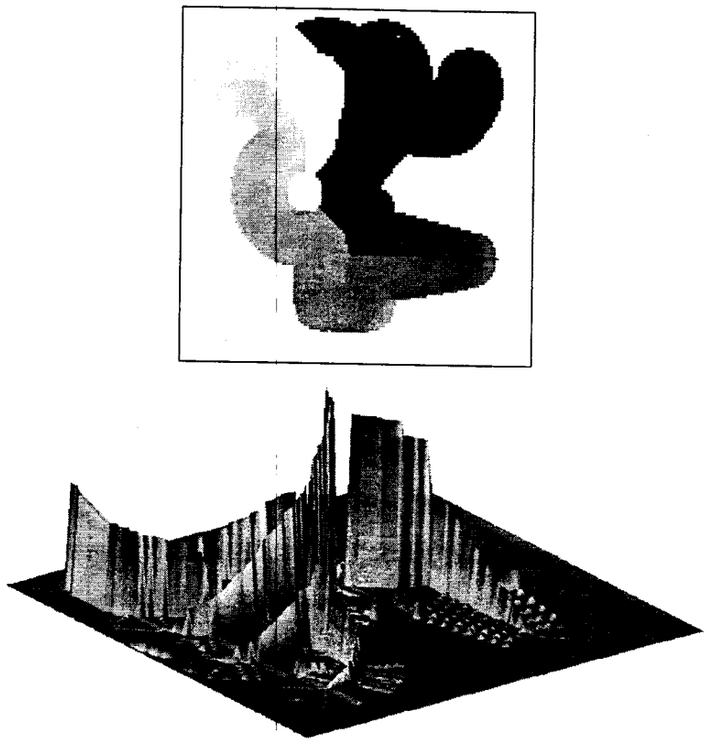


Figura 78 - Resultado da propagação de rótulos com a restrição $d < r$ *(a)*; Visão tridimensional do resultado da aplicação do algoritmo de diferenças *(b)*.

A partir do esqueleto multi-escalas calculado, seleciona-se os elementos de um conjunto $V = \{v_k\}$ candidatos a iniciar perseguições às saliências. Para isso, o esqueleto multi-escalas é limiarizado, tomando os pontos acima de $\lambda \leq r$, resultando na matriz M_C. Uma representação do resultado dessa limiarização é ilustrada na imagem da Figura 79. As extremidades dessa imagem binária são detectadas morfologicamente (apenas um vizinho em uma 8-vizinhança) e cada extremidade detectada é adicionada em V.

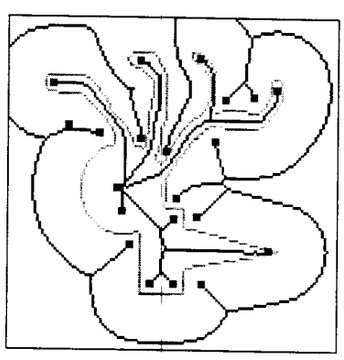


Figura 79 - Resultado da limiarização do esqueleto

A lista de pontos V é o ponto de partida para a busca das saliências. Alguns elementos de V podem já ser uma saliência. Portanto, se $d(v_k, B) < 2$, então esse elemento $v_k \in V$ é removido de V e inserido em uma lista de saliências $S = \{s_k\}$, passando a ser uma saliência e deixando de ser um candidato à perseguição.

A perseguição se inicia tomando cada elemento v_k e verificando sua vizinhança. Se um ou mais vizinhos existirem com o mesmo valor (mesma altura), então cada um desses vizinhos serão empilhados e seus vizinhos também verificados. Se novamente vizinhos com o mesmo valor forem encontrados, o mesmo procedimento acontece, isto é, são empilhados e recursivamente processados. No entanto, se houver apenas valores menores na vizinhança, então realmente vamos descer. Mas não tão rápido. Uma lista de novos pontos N conterá este ponto maior entre os menores. Essa lista N será concatenada à V, depois que os pontos v_k já verificados tiverem sido removidos de V. Uma saliência é detectada quando não houver vizinhos, nem com valores menores, nem com valores iguais. Novamente, sua distância é verificada, e se $d(v_k, B) < 2$, então v_k passa a ser indicado como um ponto de saliência e inserida na lista $S = \{s_k\}$. Ao final, S contém todas as saliências detectadas. Esta estratégia pode ser entendida como uma descida evitando descontinuidades. A Figura 80 ilustra a descida no esqueleto multi-escalas em um detalhe da Figura 78*b*.

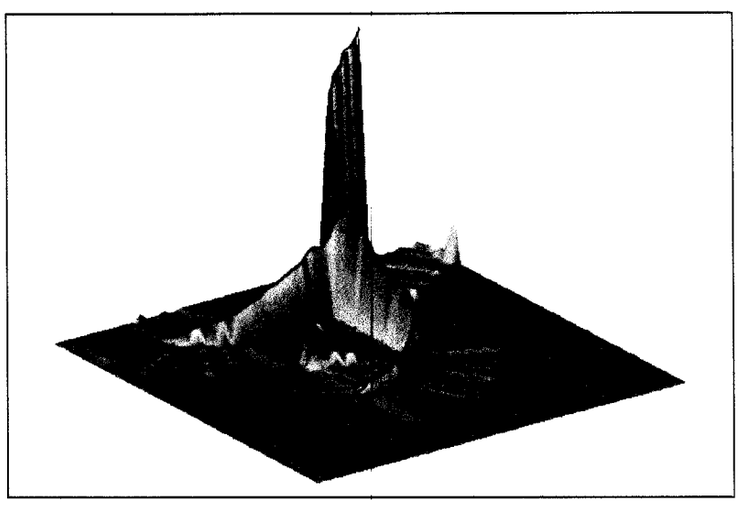
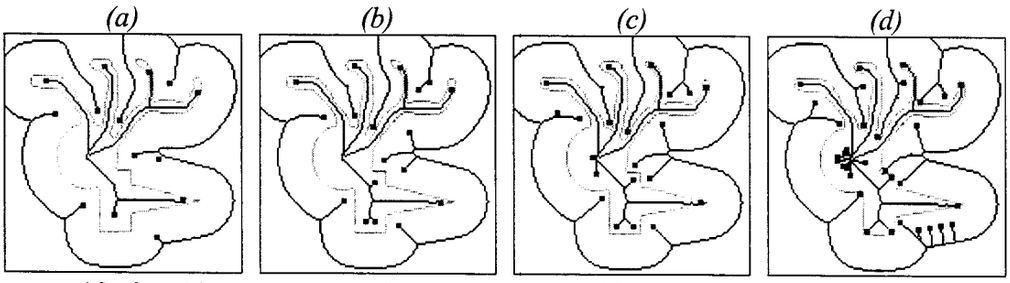


Figura 80 - O caminho vermelho ilustra o processo de perseguição à extremidade

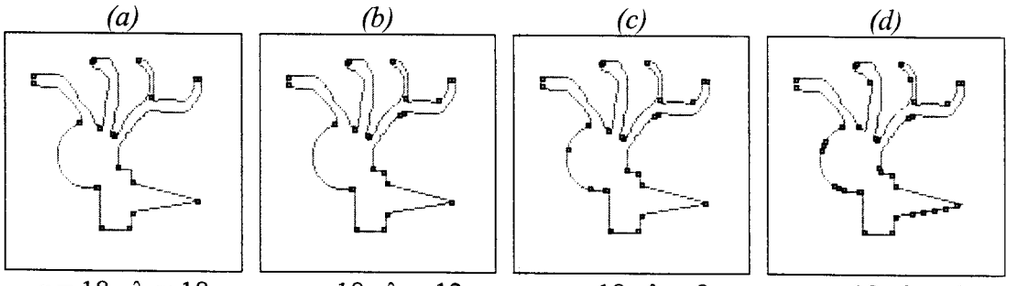
5.3.2 Resultados e desempenho

Para executar uma detecção de saliências em um contorno, é necessária a entrada de dois parâmetros (explicados acima). O parâmetro r, o raio de esqueletonização, permite reduzir ou aumentar o desempenho da detecção pois definirá o tamanho da estrutura SD e consequentemente da dilatação exata. O parâmetro λ, o limiar para início da perseguição, controla a escala na qual se deseja detectar saliências. Um limiar λ menor acarretará a detecção de saliências em escalas mais detalhadas. Já um λ maior privilegia as saliências em escalas maiores desprezando as de menor escala. A Figura 81 ilustra os pontos candidatos para quatro limiares (18, 12, 8 e 4). A Figura 82 mostra as saliências detectadas em cada escala selecionada.



$r = 18$ $\lambda = 18$ $r = 18$ $\lambda = 12$ $r = 18$ $\lambda = 8$ $r = 18$ $\lambda = 4$

Figura 81 - Imagem do esqueleto multi-escalas limiarizada em diferentes escalas.



$r = 18$ $\lambda = 18$ $r = 18$ $\lambda = 12$ $r = 18$ $\lambda = 8$ $r = 18$ $\lambda = 4$

Figura 82 - Saliências detectadas para algumas escalas.

Pode-se observar que a escala $\lambda = 4$ detecta saliências que são formadas pela amostragem do contorno sobre uma grade discreta. Isto acontece mais freqüentemente em partes que amostram retas que não estão dispostas paralelamente a algum dos eixos ou em partes circulares. Em menor grau isso acontece na Figura 82c, já que a escala é um pouco maior ($\lambda = 8$). A escala $\lambda = 12$ identifica os pontos de segmentação (saliências) de forma bastante satisfatória. No entanto, na

Figura 82a, uma saliência importante (um canto de 90 graus) não foi detectada. Pode-se observar na Figura 81a que nessa escala, o eixo de simetria não existe para aquela saliência. Portanto, uma escolha adequada da escala é necessária para que se tenha um conjunto de saliências que possa segmentar o contorno, i.e., λ influi muito no resultado da segmentação.

O raio de esqueletonização, r, deve ser escolhido de modo que contenha os pontos candidatos à perseguição. É claro que um r muito pequeno proibiria a existência dos pontos que se iniciassem em escalas maiores e fora do raio. Uma possível solução seria reduzir o valor de λ, no entanto, para formas que contenham saliências significantes em escalas muito pequenas e outras em escalas muito grandes, pode ter que se limitar às maiores, e então precisar de um r ainda maior.

O parâmetro r está diretamente relacionado com a distância máxima a qual se executará a esqueletonização (ou dilatação exata), que por sua vez é a etapa mais lenta do processo. Desse modo, r é uma escolha que influi diretamente no tempo de execução da detecção de saliências. Para ilustrar esta influência, a Figura 83 mostra um gráfico do tempo de execução (em milisegundos) pelo parâmetro r adotado em uma execução de detecção de saliências sobre o contorno da imagem da Figura 58a. Percebe-se claramente um crescimento acentuado do tempo de execução em função do crescimento de r. É interessante, portanto, escolher um r mínimo que consiga conter na região de esqueletonização todos os pontos candidatos. Esse compromisso é uma escolha do usuário. Esse gráfico mostra também as diferentes fases do processo de extração de saliências. A parte vazada (branca) identifica o processo de esqueletonização. No topo a parte sólida (preta) corresponde à perseguição da saliência propriamente dita. Percebe-se, portanto uma extrema dependência da métrica adotada para a construção da estrutura do esqueleto. Métricas alternativas (veja 4.2) podem ser aplicadas para tentar reduzir este tempo.

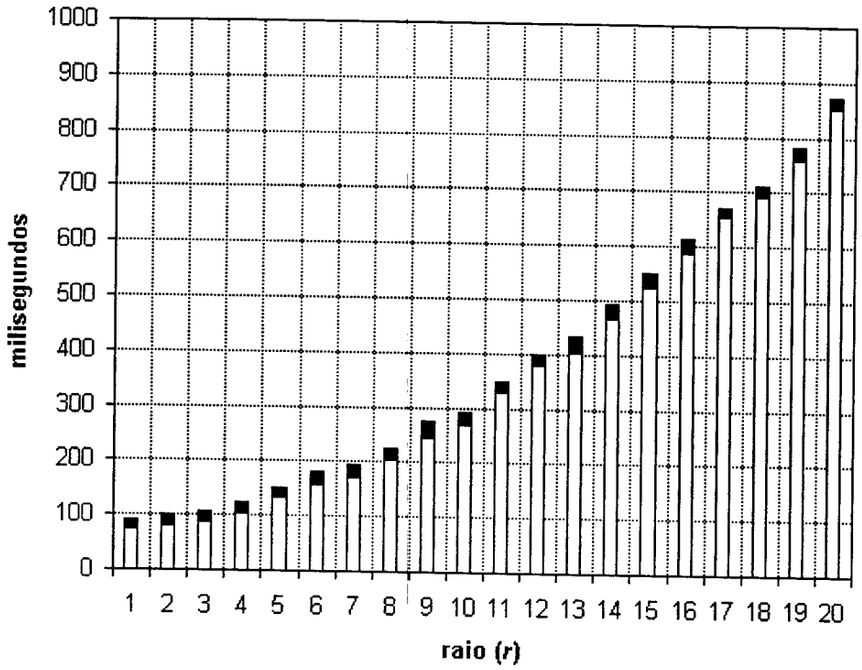


Figura 83 - Desempenho do método proposto conforme varia o raio.

Esse crescimento acentudado conforme o crescimento de *r* justifica-se pelo número de elementos na estrutura SD, dada pelo número de posições para cada distância euclideana (veja Figura 63) e que deve ser percorrida para propagar os rótulos.

Uma outra característica interessante da abordagem proposta é a linearidade do tempo de execução conforme aumenta o número de elementos no contorno. No gráfico da Figura 84, percebe-se esse comportamento na etapa mais lenta do processo, isto é, a esqueletonização. Uma vez alocada, a estrutura de distância é utilizada pelo processo de (1) esqueletonização (representada pelos quadrados vazados). Em seguida ocorre a (2) perseguição às saliências (quadrados cheios). A Figura 58a foi escalada para diferentes números de elementos de contorno, e então o tempo de execução para a detecção de saliências foi tomado em cada imagem escalada. É fácil visualizar que a etapa de perseguição às saliências (2) é bastante rápida. O tempo de construção da estrutura SD pode ser reduzido pelo armazenamento dessa estrutura em disco, ou então a calculando antes da detecção das saliências, se a aplicação consistir, por exemplo, de um grupo de imagens.

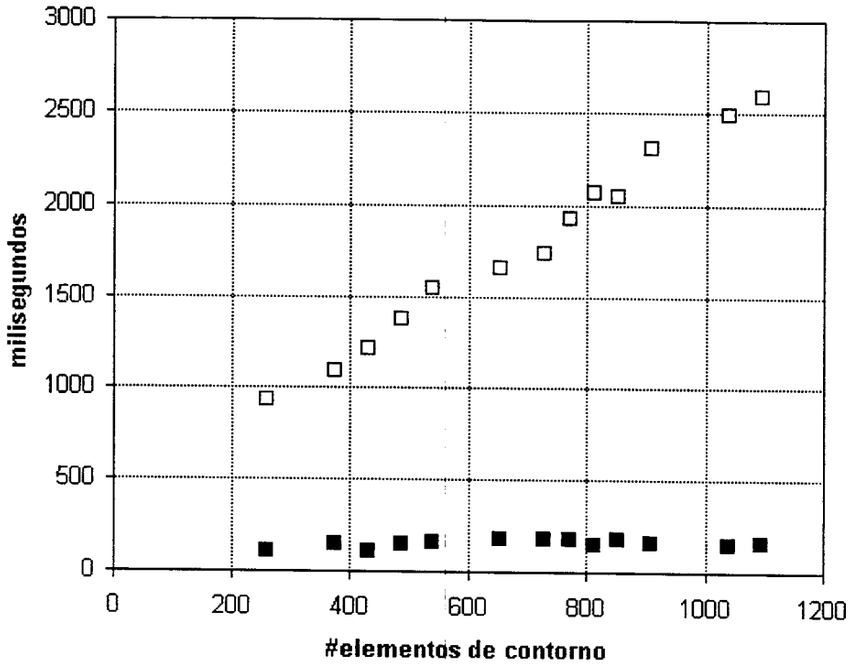


Figura 84 - Desempenho com relação ao número de elementos no contorno.

O método proposto mostra-se razoavelmente robusto na sua aplicação a imagens com variações de escala. Em um primeiro exemplo, ilustrado na Figura 85, losangos com ângulos internos opostos variando, ou seja, dois ângulos decrescendo e os outros dois crescendo, são submetidos ao algoritmo de detecção de saliências, com parâmetros $r = 8$ e $\lambda = 4$ mantidos constantes. Percebe-se que as saliências foram detectadas em um amplo intervalo de variação dos ângulos.

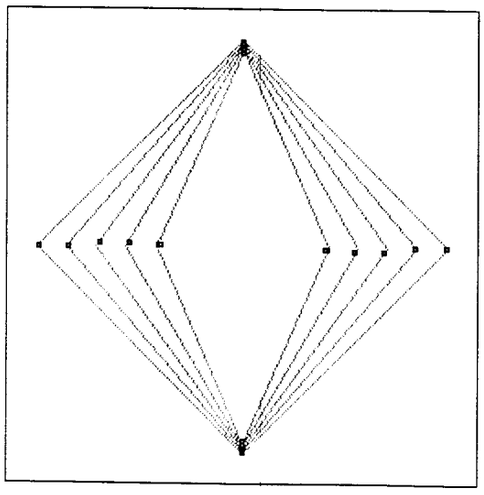


Figura 85 - Losangos com ângulos internos variando e suas respectivas detecções de saliência.

Como o método permite selecionar a escala de detalhes, é fácil utilizar esta proposta como um meio de desconsiderar ruídos na borda de uma forma. A Figura 86 ilustra esta propriedade, na qual um retângulo bastante ruidoso é submetido ao método proposto. Percebe-se que existe uma escala que identifica as quatro saliências do retângulo. É claro que a localidade não está preservada, porém a borda também foi deslocada pelo ruído.

O esqueleto multi-escalas apresenta a propriedade de invariância à rotação. Esta propriedade é útil também na detecção de saliências, conforme ilustrado pelo exemplo da Figura 87. Embora a rotação não preserve o perímetro do contorno, percebe-se que nas escalas mais importantes a localidade das saliências é preservada na rotação mesmo em uma figura ruidosa. Nas escalas menores, isso não é garantido, pois as estruturas menores sofrem muita influência da variação da distância na rotação do contorno sobre uma grade de amostragem.

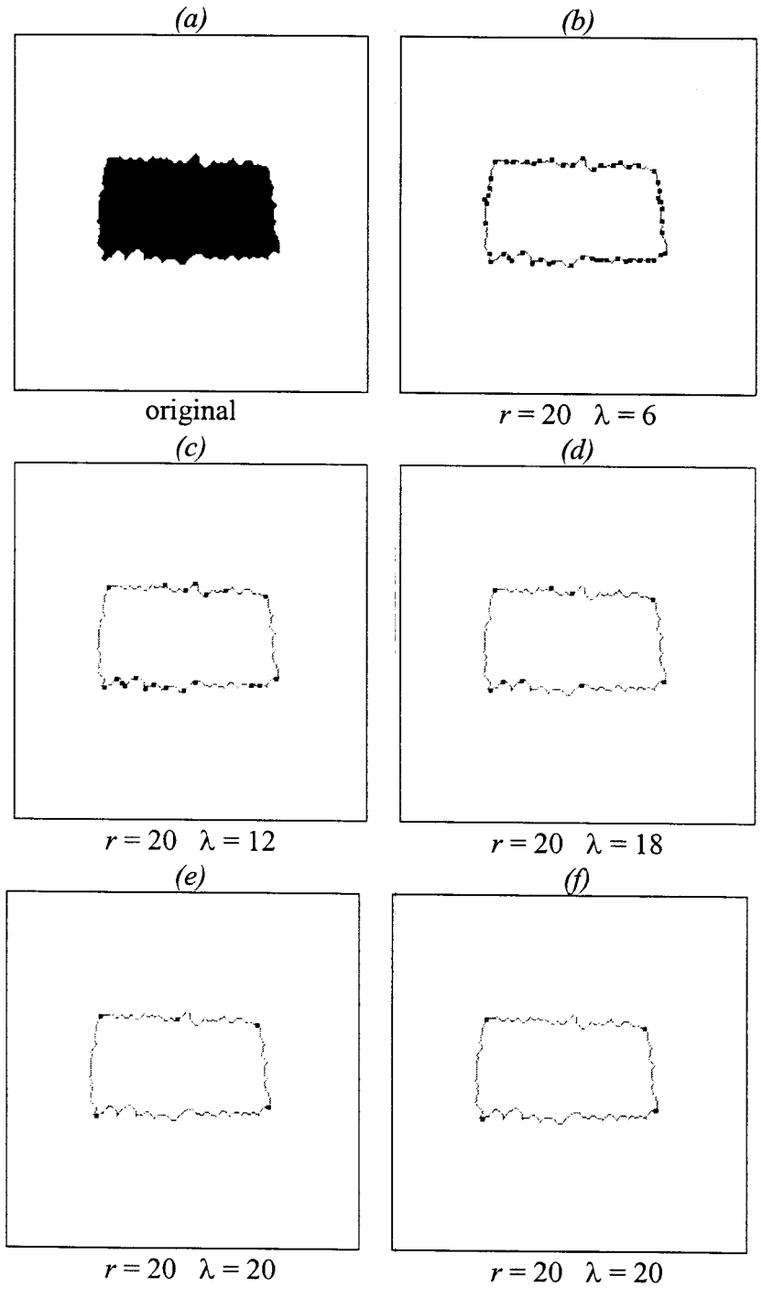


Figura 86 - Submetendo o algoritmo proposto à forma acima o número embaixo de cada Figura são respectivamente r e o *thresholding*.

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

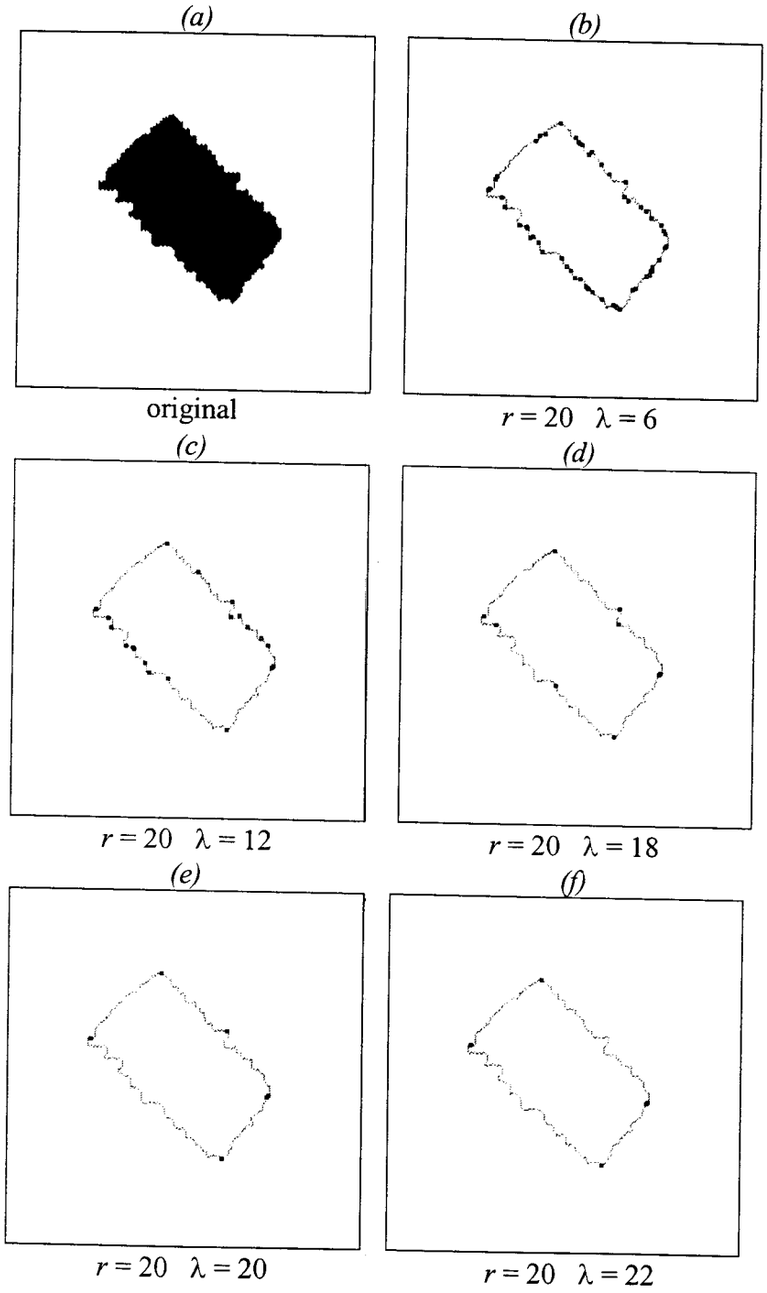


Figura 87 - Submetendo o algoritmo a imagem rotacionada.

Imagens com ramificações são particularmente beneficiadas por este método, já que possuem uma diversidade menor de escalas das saliências. O exemplo da Figura 88 ilustra bem essa propriedade do método proposto que identifica muito bem as saliências em uma forma bastante ramificada.

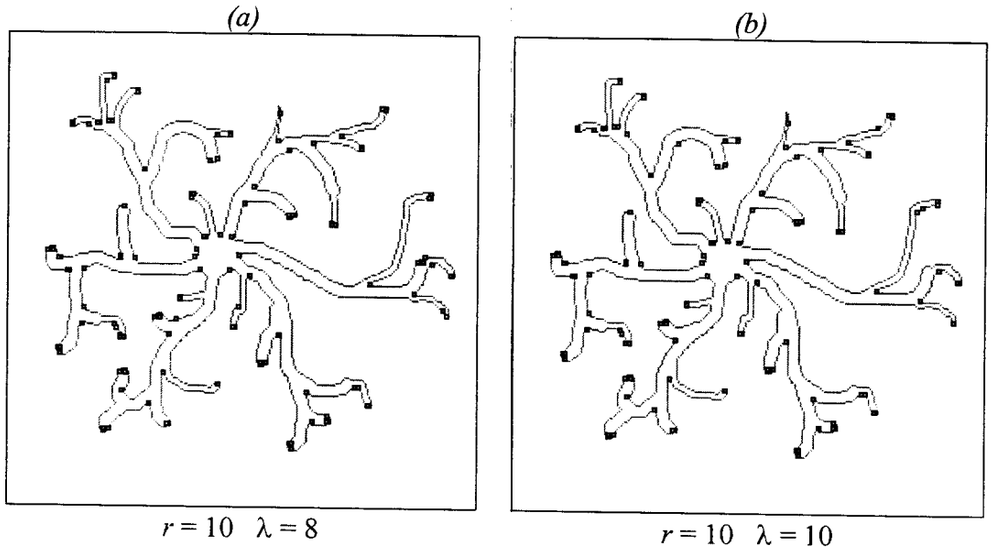


Figura 88 - Exemplo de extremidades na escala detectada.

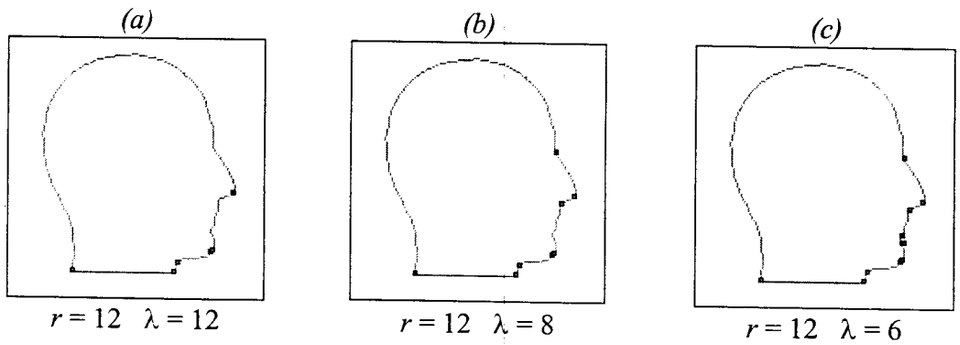


Figura 89 - Aplicação do método proposto para identificar estruturas salientes em uma forma.

Uma outra aplicação do método é a seleção de estruturas específicas em um contorno. No exemplo da Figura 89, a variação das escalas permite selecionar partes da uma face tais como olho, nariz, boca, queixo e pescoço. No caso, o nariz, queixo e pescoço são as estruturas mais salientes. Na próxima escala inferior o olho é identificado e então, na escala mais baixa, a boca.

5.3.3 Conclusões

Foi desenvolvido um método para detectar saliências em contornos planares 2D extraídos, por exemplo, de imagens binarizadas. O método baseia-se na esqueletonização multi-escalas proposta por [COSTA & ESTROZI, 1999] e apresenta um potencial bastante amplo de aplicações, conforme ilustrado ao longo desse trabalho. Dois parâmetros, um que controla o nível de detalhe e o outro que

limita o raio em que a esqueletonização multi-escalas é realizada influem, respectivamente, na significância das saliências e no tempo de execução. Os resultados mostraram as boas propriedades do método como detector de saliências, tais como: sensibilidade controlável ao ruído, invariância à rotação, seletividade a estruturas e robustez em imagens com várias escalas.

Os tempos de execução foram apresentados e confirmou-se que o processo de esqueletonização multi-escalas pode comprometer o desempenho da detecção das saliências. Porém, isto pode ser melhorado por um armazenamento da estrutura de distâncias euclideanas SD.

O método apresentado leva em consideração uma descrição estrutural da forma: o esqueleto multi-escalas. Isso permite herdar boa parte das propriedades dessa descrição tais como: ajuste do nível de detalhes (parâmetro λ), invariância à rotação, tolerância à coexistência de diferentes escalas em uma mesma forma. As abordagens revistas utilizam como característica para extrair saliências, parametrizações do contorno tais como tamanho de arco, curvatura [CHETVERIKOV & SZABÓ, 1999], histogramas de comprimentos cordais [MEHROTRA & NICHANI, 1990], histogramas de *chain-code* [ARREBOLA et al., 1997], entre outros. Muitas destas abordagens exigem parâmetros para controlar compromissos entre a localidade da saliência e a presença de ruídos. Algumas abordagens procuram calibrar parâmetros de filtragem de ruídos e limiares para que maximizem os acertos. Abordagens que consideram máscaras para comparar com gabaritos de cantos [LIN et al., 1998] não podem garantir robustez, a menos que se defina um conjunto restrito de tipos de saliências. Funcionam bem, contudo para um conjunto restrito de problemas. O método proposto mostra-se capaz de tratar uma grande diversidade de contornos, ampliando seu espectro de aplicações.

6 ANÁLISE PICTÓRICA

Modelos estéticos são procurados por pesquisadores de diversas áreas e com um amplo leque de aplicações. Trabalhos com abordagens subjetivas nesse sentido têm sido propostos [ORLAND et al., 1998] ou vêm sendo pesquisados [RIGLIS, 1998], contudo a complexidade computacional do problema dificulta a realização de propostas mais ambiciosas para resolver o problema, embora elas existam.

O meio no qual o repertório de informações trafega ou se apresenta é fundamental na viabilidade de uma proposta. Variam desde a estética do texto (poesia, por exemplo)[BENSE, 1975] até a agradabilidade de objetos sintetizados em um meio eletrônico como a beleza de ambientes virtuais [REDDY, 1996; REDDY, 1997]. A arquitetura tem explorado bastante a avaliação automática de beleza em paisagens ou em edificações [ORLAND et al., 1998; SYNEK, 2000]. A beleza estética humana também tem sido objeto de pesquisa almejando o entendimento de comportamentos ou simplesmente qualificar texturas de pele [FINK, 1998], entre outras aplicações.

O objetivo desta seção é apresentar o trabalho elaborado para avaliação pictórica de imagens obtidas de home-pages obtidas em sítios *web*.

A análise pictórica proposta procura modelos de percepção visual que sejam compatíveis com critérios estéticos subjetivos. Este trabalho realizou um ambiente para avaliação da qualidade pictórica de home-pages. O critério considerado como parâmetro de avaliação é a visão humana. O objetivo é construir um ambiente no qual a qualidade visual de home-pages possa ser avaliada apresentando-se a imagem da página e obtendo-se de um sistema, uma avaliação qualitativa baseada em medidas da imagem. A idéia é inferir a opinião sobre a qualidade visual da home-page pela escolha de um conjunto de pessoas.

Essa idéia é compatível com as definições que Bense [BENSE, 1975], apoiado por trabalhos anteriores de Birkhoff [BIRKHOFF, 1933], Maser, Moler e Gunzenhäuser[2] sobre estética.

O teorema da medida estética M_E de Birkhoff [BIRKHOFF, 1933] determinou uma relação fundamental entre a ordem (O) e a complexidade (C). Birkhoff concebeu M_E intuitivamente para tentar atribuir um índice numérico para a estética da percepção imediata. Esta relação é dada por um quociente entre ordem e complexidade (68).

$$M_E = \frac{O}{C} \tag{68}$$

Birkhoff percebeu a necessidade de definir um repertório para experimentar essa medida. Os alvos dos experimentos de Birkhoff foram polígonos, padrões reticulados e vasos. Assim, para cada um destes objetos, ele definiu um conjunto de propriedades de ordem (O) como: simetria, equilíbrio, agradabilidade entre outros. A propriedade de complexidade era determinada por uma característica atômica do objeto em questão. Estas determinações eram empírico-intuitivas, mas faziam um certo sentido para os 90 polígonos testados por Birkhoff.

Maser sofisticou um pouco mais a medida estética de Birkhoff com atributos de cores e considerando a presença ou não desses atributos em objetos e assim definiu-a como uma medida macro-estética.

Bense, por sua vez, situou o problema de avaliação estética em uma abordagem de escalas. Em uma escala maior os repertórios de atributos macro-estéticos resumem um estado estético. Na direção da menor escala um conjunto de medidas micro-estéticas configura o estado macro-estético. Traduzindo, as medidas macro-estéticas podem ser resumidas por uma opinião subjetiva sobre beleza como uma nota entre zero e dez. Já a medida micro-estética é um aspecto quantitativo de um objeto, tal como o nível de cinza médio de uma imagem.

As considerações de Bense e de Moler levaram Gunzenhäuser a formular essa concepção usando conceitos estatísticos de redundância (R) e entropia (H), resultando em (70).

---

[2] apud [BENSE, 1975]

$$M_e = \frac{R}{H} \tag{70}$$

Este trabalho propõe a construção de modelos a partir da generalização de respostas de sujeitos humanos. Estas respostas são obtidas por experimentos psicofísicos submetidos a sujeitos humanos. Seu papel é fornecer um grau para avaliação estética humana para um critério estético. As mesmas imagens utilizadas no experimento psicofísico foram então divididas em dois grupos: (1) de treinamento e outro (2) de testes (veja Figura 90). Um conjunto de medidas foi calculado para todas as imagens e o modelo foi construído pela interpolação do espaço de medidas pelos *scores* fornecidos pelos sujeitos para cada imagem. As medidas das imagens de teste e seus *scores* servem então para avaliar a qualidade do modelo. Esta qualidade é refletida pelo erro, ou seja, pela diferença entre a resposta do modelo e o *score* das imagens de teste. Para buscar o conjunto de medidas que resulta em um erro mínimo, foi construído um algoritmo genético.

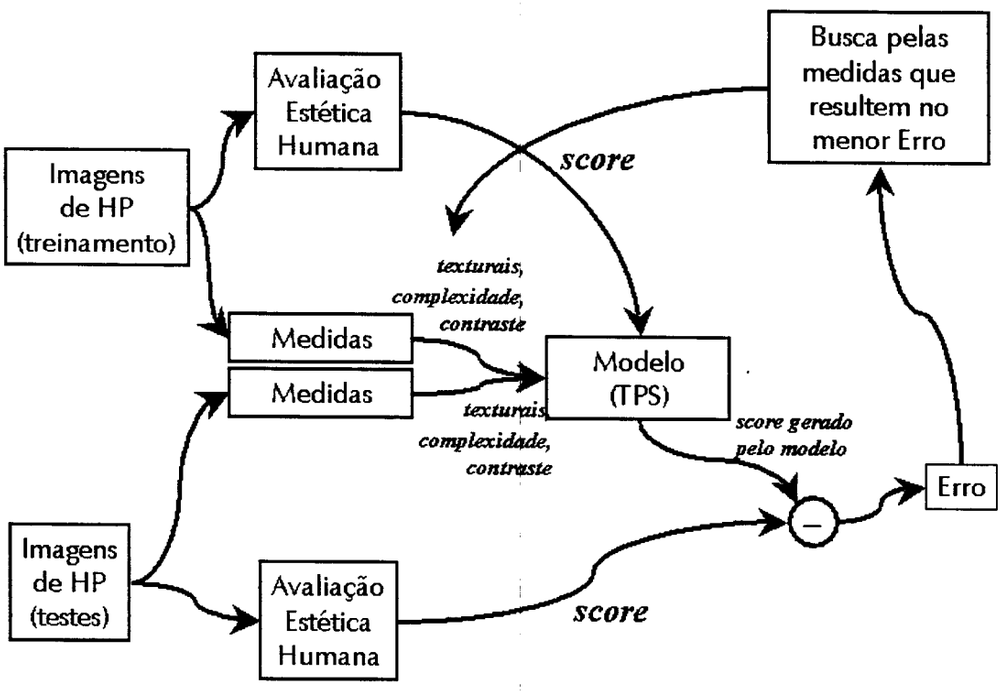


Figura 90 – Esquema para a construção do modelo de avaliação estética utilizando medidas texturais, de complexidade e de contraste.

Os objetos, ou seja, as páginas de sítios *web* possuem atributos macro e micro-estéticos mais sofisticados do que simples polígonos. Nelas foram identificados critérios estéticos dos quais alguns foram selecionados para

experimentos. Estes critérios estéticos foram obtidos por avaliações de sujeitos humanos submetidos a um experimento psicofísico implementado como um serviço web (Figura 91). O objetivo é obter um amplo espectro de sujeitos para atenuar possíveis influências culturais ou contextuais na avaliação das páginas pelos sujeitos.

Foram escolhidos critérios estéticos que procurassem definir os estados estéticos provocados pela apresentação da página web a um sujeito. São critérios subjetivos que refletem a opinião de sujeitos submetidos a um experimento psicofísico. O ***prazer visual ou agradabilidade*** serve como uma medida de aprovação visual. Suspeita-se, por exemplo, que a percepção visual de texturas estejam associadas ao prazer tátil [DONDIS, 1997]. Pentland elaborou um experimento para correlacionar o sentido tátil e o visual [PENTLAND, 1984]. As cores estão diretamente associadas a questões de sobrevivência humana no ambiente, seja como propriedades de alimentos, seja como aspectos humanos ou ainda como uma característica ambiental[BARROW, 1995]. Por exemplo, o vermelho ou amarelo de um fruto sobre uma folhagem verde de uma árvore pode implicar uma aprovação de uma determinada imagem com estas cores. Já o vermelho do sangue humano pode ter conotações de perigo ou de alimento também, já que o homem caçou por muito tempo durante sua evolução. O céu azul e a presença de núvens brancas podem ter papel determinante na avaliação estética, pela sua cosntante presença no meio-ambiente. Portanto, o critério de ***equilíbrio de cores*** tenta identificar estes elementos estéticos primitivos. A ***originalidade*** é o critério que qualifica a diferença do que já foi visto. Alguns experimentos, principalmente para publicidade, chegaram a conclusões surpreendentes sobre a importância de elementos diferentes na avaliação da motivação em chamadas comerciais televisivas [MORIARTY & SAYRE, 1992]. A ***qualidade artística*** é, sem dúvida, o critério mais subjetivo dentre os cinco propostos. Conjecturas podem relacioná-la à fidedignidade com a idéia que o artista quer transmitir. Contudo, os surrealistas buscaram a composição artística a partir de uma regressão a elementos primitivos e transportaram a idéia da obra para o seu observador. Ainda assim, não se pode questionar valores artísticos atribuídos a estas obras. A ***motivação*** é um critério associado diretamente com o meio home-pages. O alvo do modelo, neste caso, é a decisão do usuário em explorar novas páginas a partir da que está observando.

Muitas páginas apresentam um projeto gráfico adequado para meio impresso. Algumas vezes o objetivo é apenas a apresentação institucional. No entanto, como uma mídia em que o acesso é um valor e despertar a decisão de exploração aumenta esse mesmo valor, é importante responder a pergunta: "Quanto a estética visual influencia na audiência?".

Um conjunto de páginas (veja Apêndice 9.4) foi escolhido na *web*, procurando representar o universo de páginas existentes. O número de páginas foi limitado a 40, para que os sujeitos fossem poupados de cansaço e, conseqüentemente, evitando respostas displicentes ou tendenciosas. Ao sujeito foi pedida a identificação (Figura 92*a*) e foi exibida uma explicação sobre o experimento (Figura 92*b*). A partir de então, cada imagem foi exibida por 4 segundos. A seguir, ao sujeito foi mostrado um formulário para opinar sobre a página (Figura 92*c*), requerendo assim que ele desse uma nota entre 1 e 10 avaliando os critérios estéticos daquela página. O resultado da pesquisa foi armazenado em uma tabela resultante da relação **dmnAnswers** (veja Apêndice 9.1) para a qual cada sujeito é uma entidade atômica identificada pelo seu endereço eletrônico.

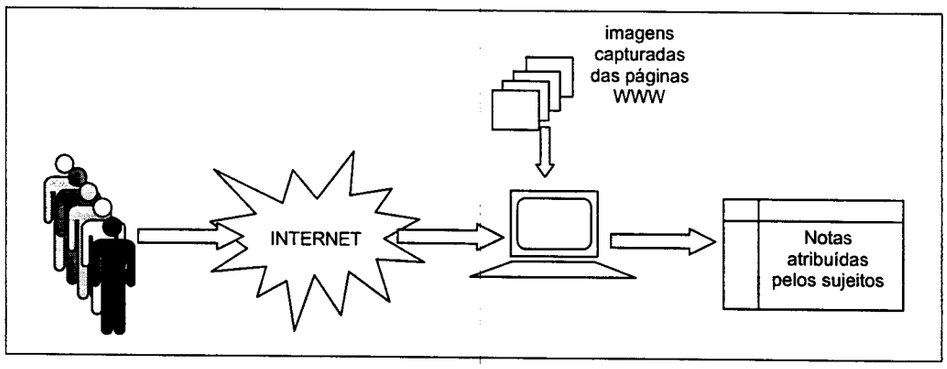


Figura 91 – Experimento psicofísico

As imagens capturadas das páginas *web* são também dispostas em uma tabela cuja relação é dada por **imgWeb**. O experimento é realizado remotamente, acessando um endereço URL que mostra a página ilustrada na Figura 92.

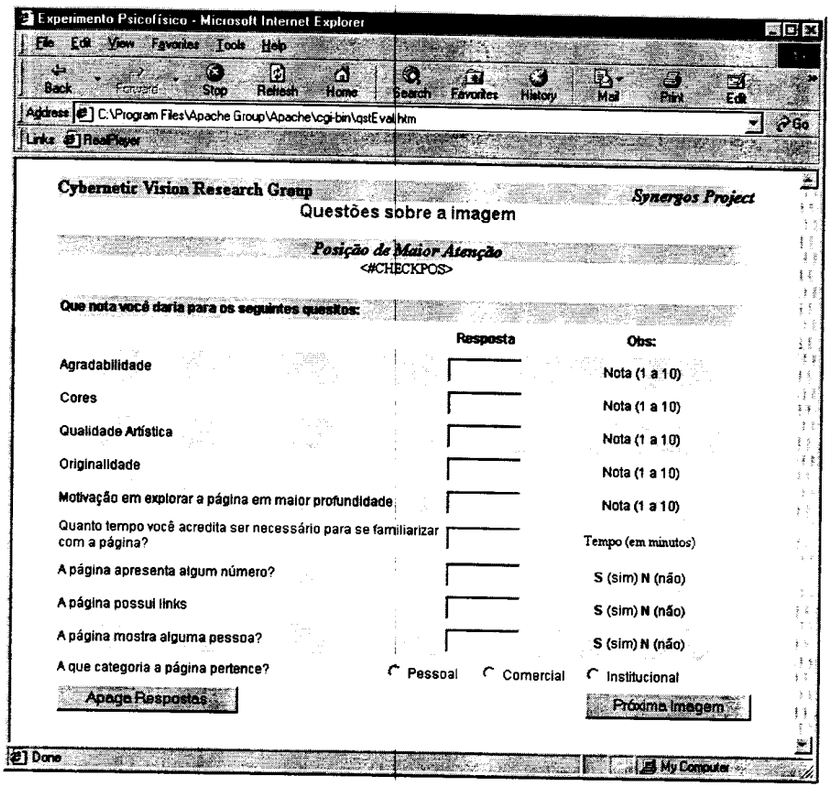


Figura 92 – Formulário para respostas a que os sujeitos foram submetidos no experimento psicofísico.

Além do experimento psicofísico, medidas micro-estéticas da imagem são geradas a partir do conjunto de imagens. Estas informações são aspectos das imagens relacionados à percepção dos quais há a suspeita de influenciarem a avaliação subjetiva de qualidade. São três tipos de medidas: texturais, de complexidade e de contraste.

As medidas texturais são histogramas do espectro de potência da imagem em níveis de cinza. São duas: descrição espectral radial e descrição espectral concêntrica. Estas medidas são calculadas para cada imagem e armazenada na tabela **dmnSpc** (veja Apêndice 9.1).

As medidas de complexidade são relacionadas à informação contida em uma imagem (entropia e variância) e foram agrupadas em escalas. Cada escala consiste de um borramento da imagem original com um filtro passa-baixas gaussiano. Os resultados de sucessivos borramentos permitem analisar a complexidade da imagem com relação a um tipo de atributo na tentativa de filtrar situações não controladas no experimento, tais como: distância do sujeito ao monitor; eventuais disfunções

oftálmicas; condições variáveis de luminosidade do ambiente. Uma outra medida de complexidade utilizada foi a salsicha de Minkowski que permite quantificar a área ou volume de influência de formas. As medidas de complexidade são armazenadas nas tabelas **dmnEnt**, **dmnVar** e **dmnFrc** para medidas de entropia, variância e salsicha de Minkowski, respectivamente (veja Apêndice 9.1).

As medidas de contraste buscam identificar e quantificar diferenças entre regiões. Para isso, duas medidas de contraste foram experimentadas: em níveis de cinza e em cores. Foram calculados os histogramas das diferenças entre pixels vizinhos sendo que cada intervalo do histograma serve como um atributo. Estas medidas são armazenadas na tabela **dmnCnt** (veja Apêndice 9.1).

Na próxima seção, cada uma destas medidas microestéticas é detalhada.

6.1 Medidas

6.1.1 texturais

O descritor espectral é um histograma da freqüência espacial de uma imagem. A freqüência espacial permite quantificar os elementos texturais de uma imagem identificando seus aspectos como posições, tamanhos, forma, orientação, número de ocorrências, entre outros. No entanto, estas características são diversas e são caracterizadas por vários tipos de medidas. Dentre estas técnicas estão as matrizes de co-ocorrências e a autocorrelação [SONKA et al., 1993] que podem ser realizadas considerando-se apenas o espaço no qual a imagem reside. Uma outra maneira de avaliar a freqüência espacial de uma imagem $f(x, y)$ é aplicando a transformada de Fourier e obtendo sua representação em freqüências. Tal representação $F(u, v)$ reside em um domínio em $\mathbf{C}^2$, cujos elementos são coeficientes de uma base de funções senoidais bidimensionais. O espectro de potência (71) da imagem no domínio de freqüência é uma representação em $\mathbf{R}^2$ que, devido à localidade de cada função da base, possibilita a identificação de características texturais.

$$E(u,v) = |F(u,v)|^2 \tag{71}$$

Duas maneiras são utilizadas para caracterizar as estruturas texturais: (i) pela distribuição ao longo de setores (variação direcional) e (ii) ao longo de faixas circulares (variação espacial).

Na variação direcional, as freqüências devem ser medidas ao longo de setores de um círculo com centro na origem. O raio r deste círculo é fixado em 0.5, supondo uma amostragem em pixels da imagem. A Figura 93 ilustra o espectro de uma imagem de um quadrado e a divisão deste espectro em setores.

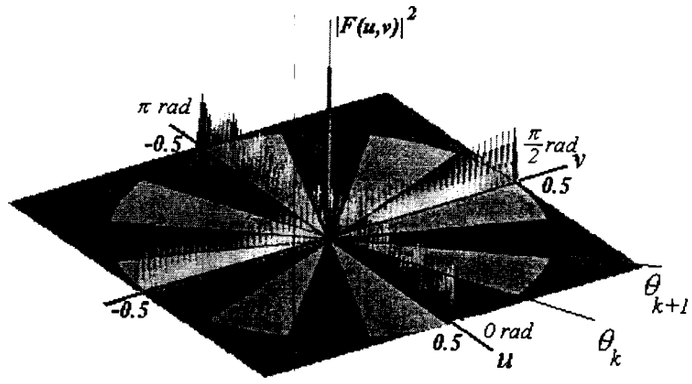


Figura 93 - O espaço do espectro de potência é dividido em setores, dentro dos quais seus elementos são somados para formar descritores das variações espaciais

O objetivo desse descritor espectral é caracterizar tamanho e freqüência das ocorrências de estruturas na imagem. Pelas propriedades da transformada de Fourier, os elementos de baixa freqüência (regiões uniformes e texturas com variações suaves) concentram-se ao redor da origem. Já os elementos de alta freqüência (bordas ou variações abruptas) distribuem-se adiante da origem ao longo do espaço, até o valor máximo de freqüência 0.5. Nas Figura 94*a*, 4*b* e 4*c*, podemos observar que conforme a imagem apresenta estruturas maiores, as respectivas variações espaciais observadas na Figura 94*d* vão se aproximando da origem.

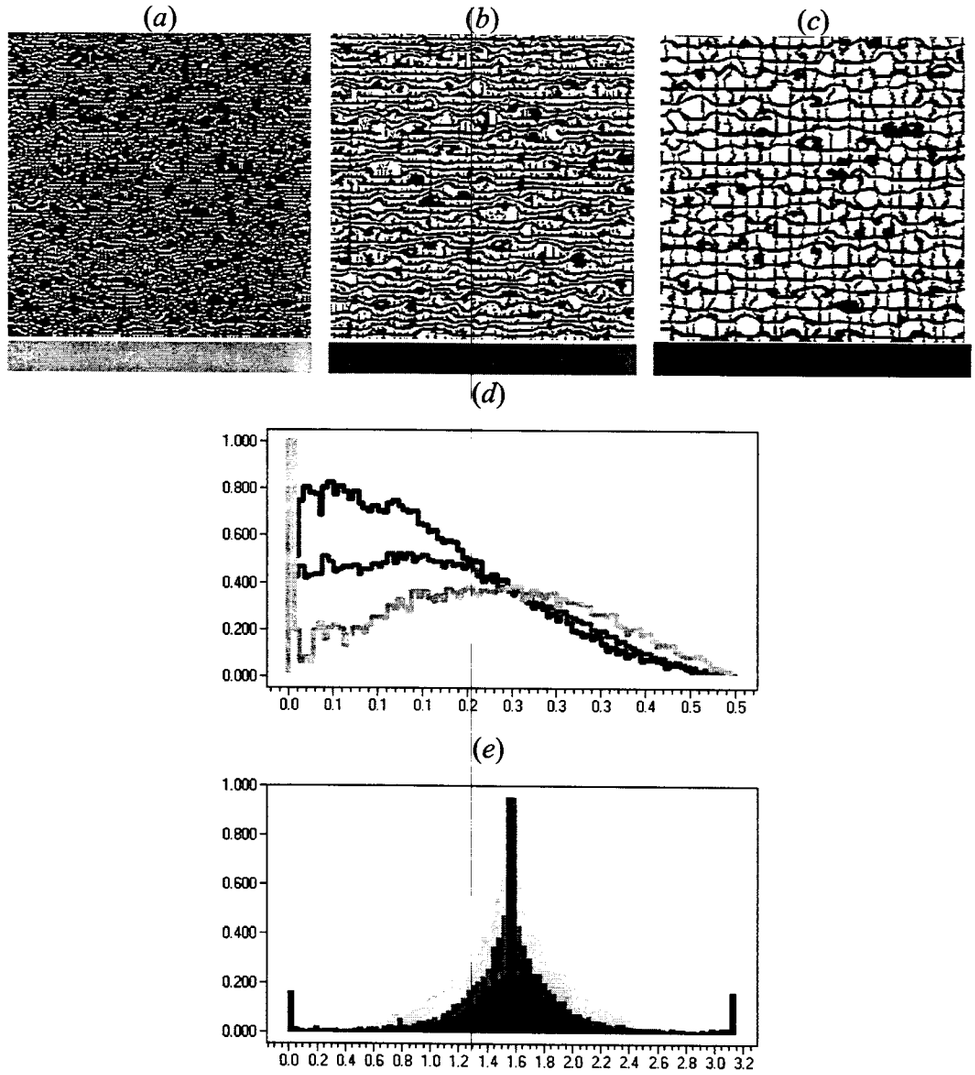


Figura 94 – Exemplos de descritores espectrais para texturas com detalhes (*a*) pequenos; (*b*) médios; (*c*) grandes. As variações do descritor espectral (*d*) espacial e (*e*) direcional das respectivas imagens são ilustradas com os tons da barra abaixo de cada Figura correspondente.

O cálculo dos descritores de variação espacial são realizados tomando a soma dos elementos do espectro de potência (71) em um setor $S_{\text{esp}}[k]$ do círculo de raio 0.5 que sobrepõe esse espectro.

$$S_{esp}[k] = \sum_{\arctan(v/u)=\theta_k}^{\arctan(v/u)=\theta_{k+1}} |F(u,v)|^2, \text{ para } \sqrt{u^2+v^2} < 0.5 \qquad (72)$$

$$k = 1..N_{\text{setores}}$$

O número de setores $N_{setores}$ deve ser introduzido. Para este trabalho, um compromisso é necessário entre a precisão desejada e o número de setores, já que é considerado um atributo.

Para melhor visualizar a variação nas escalas de freqüência mais altas, adota-se uma escala logarítmica para $|F(u, v)|^2$, aplicando-se $\log(1 + |F(u, v)|^2)$.

A descrição da variação espacial não é a única característica que se pode extrair do espectro de potência da imagem. Aspectos direcionais podem ser detectados por um descritor direcional. Esse descritor efetua a soma dos elementos do espectro de potência em regiões definidas por trilhas circulares concêntricas. A Figura 95 ilustra o espectro de uma imagem de um quadrado e a divisão deste espectro em anéis concêntricos.

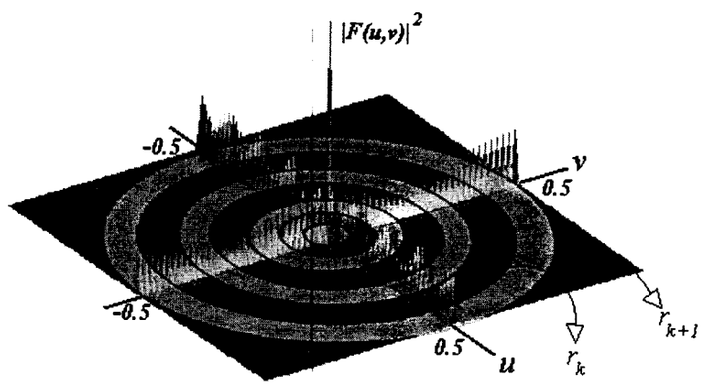


Figura 95 - O espaço do espectro de potência é dividido em trilhas concêntricas, dentro dos quais seus elementos são somados para formar descritores das variações direcionais.

O objetivo é prover uma medida que indique direções com maiores ocorrências de alguma estrutura. Nas Figura 96*a-d*, podemos observar a propriedade desse descritor em imagens apresentando linhas ao longo de uma direção. Para conhecer as direções predominantes da textura, deve-se somar $\pi/2$ a essa orientação, pois o espectro revela a orientação da função senóide bidimensional que serve como base para compor a imagem. A Figura 96*d*, mostra claramente que as orientações de cada imagem são detectadas pelos respectivos descritores direcionais.

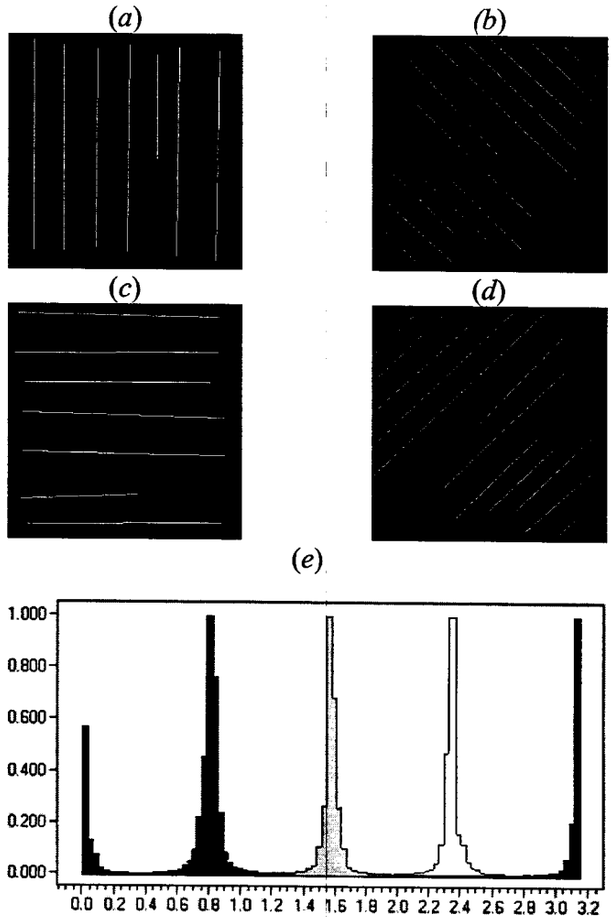


Figura 96 – Exemplos de descritores espectrais direcionais para imagens com retas de diferentes orientações predominantes (*a*) 90°; (*b*) 135°; (*c*) 0°; (*d*) 45°. No gráfico (*e*) os picos em cada tom indicam a ordem das disposições das imagens.

Para calcular o descritor direcional, o número de trilhas $N_{trilhas}$ deve ser determinado. A soma (73) dos componentes do espectro de potência para cada trilha fará parte de uma seqüência. Cada elemento dessa seqüência corresponde a um intervalo na escala entre 0 e π radianos.

$$S_{dir}[k] = \sum_{\sqrt{u^2+v^2}=r_k}^{\sqrt{u^2+v^2}=r_{k+1}} |F(u,v)|^2 \quad \text{para } \sqrt{u^2+v^2} < 0.5 \qquad (73)$$

$$k = 1..N_{\text{trilhas}}$$

Uma imagem real ou mesmo um desenho ou pintura apresentam diversas orientações. Para que seja útil na detecção de uma determinada orientação a imagem

deve ter apenas algumas orientações predominantes. Na Figura 97*a*, uma imagem tomada de um recipiente contendo varas de cana-de-açúcar dispostas em uma direção predominante ilustra o método, já que a Figura 97*b* mostra o pico que indica o máximo valor de tal orientação. Na Figura 98*a* temos uma imagem artificialmente gerada como uma grade modificada. Assim mesmo, as duas orientações predominantes podem ser visualizadas no gráfico do descritor da Figura 98*b*.

(*a*) (*b*)

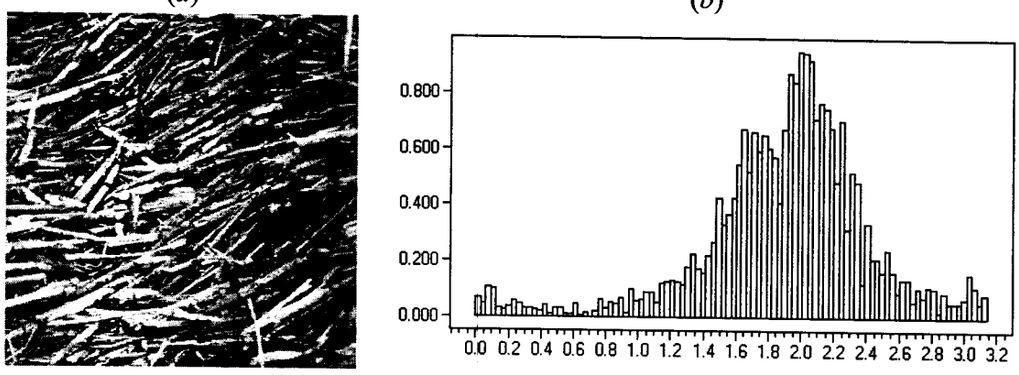


Figura 97 – Exemplo de textura real com orientação predominante. A imagem com varas de cana-de-açúcar (*a*) estão orientadas predominantemente na direção de $\pi/2$ (45°). O gráfico em (*b*) mostra essa predominância com um pico na direção ortogonal.

(*a*) (*b*)

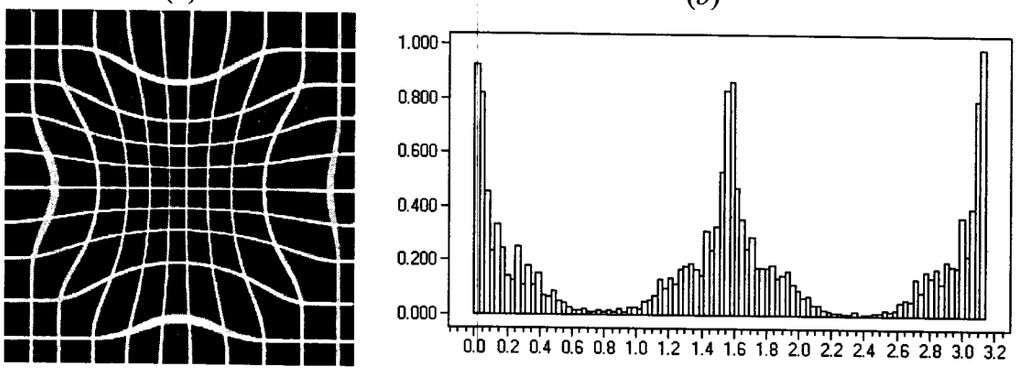


Figura 98 – Exemplo de imagem com mais de uma orientação (a) e o gráfico (b) de seu respectivo descritor espectral direcional com picos nas duas direções ortogonais.

A alteração da imagem da Figura 96*b* pela adição de componentes texturais ao longo das retas e também ao fundo resultou na imagem da Figura 99*a*. Esta alteração acarretou uma degradação do descritor espectral (veja Figura 99*b*), pois tais elementos não possuem uma orientação bem definida. Ainda assim, a orientação predominante é detectada.

O exemplo da Figura 94 mostrou os descritores espaciais de figuras texturizadas com granularidade crescente. Percebe-se naquele exemplo que as alterações da granularidade da textura influiram tanto o descritor espacial quanto o direcional. Neste último, as direções foram se definindo mais conforme aumentava a granularidade (veja Figura 94*e*).

(*a*) (*b*)

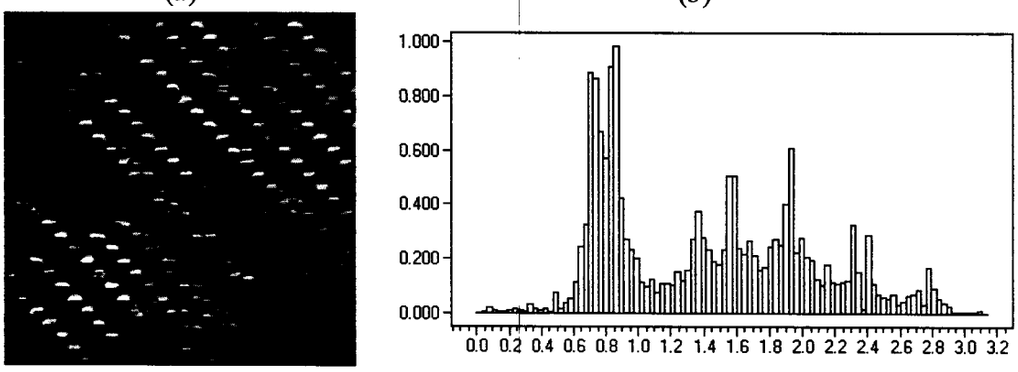


Figura 99 – Exemplo de imagem apresentando uma orientação predominante, mas degradada com granularidade (a) e o gráfico (b) de seu descritor espectral direcional refletindo essa degradação.

6.1.2 contraste

Contraste é a mudança local de intensidade. Embora sua definição esteja relacionada com a percepção, já que o olho humano distingue logaritmicamente a intensidade luminosa (lei de Weber-Fechner [GORDON, 1989]), a mudança pode ser indicada pela média das diferenças entre a intensidade de um pixel específico e a intensidade dos vizinhos.

Um conjunto de atributos pode ser definido a partir da imagem de diferenças obtida de cada pixel da imagem original. Assim, o histograma dessa imagem de diferenças revela cada quantidade de vizinhanças que contém menores ou maiores mudanças, isto é, contrastes.

O contraste pode ser definido sobre imagens em nível de cinza ou em cores. Para uma imagem em níveis de cinza $I[i,j]$, as diferenças são definidas por (74).

$$D_{cinza}[i,j] = \frac{(4\,\mathrm{Im}[i,j] - \{\mathrm{Im}[i+1,j] + \mathrm{Im}[i,j+1] + \mathrm{Im}[i-1,j] + \mathrm{Im}[i,j-1]\})}{4} \qquad (74)$$

É gerado um histograma da imagem D_{cinza}. Cada intervalo deste histograma fornece-nos um atributo de contraste de níveis de cinza.

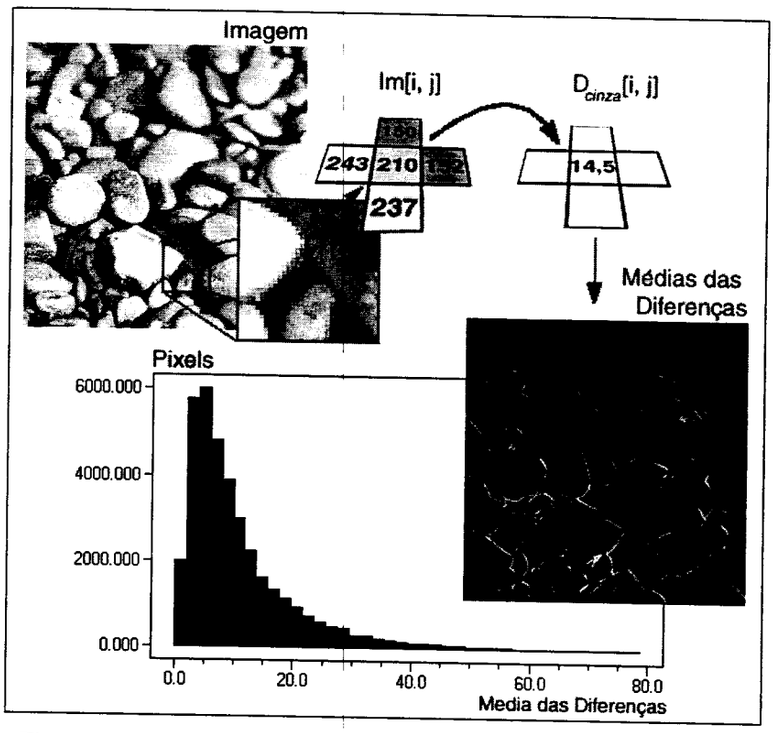


Figura 100 – Ilustração do processo de extração do histograma de contraste em níveis de cinza. A imagem original é varrida por uma máscara de vizinhança 4. A diferença média entre o pixel corrente e os vizinhos para toda a imagem resulta na imagem de diferenças da qual é calculado o histograma.

É importante observar que um grande número de ocorrências de baixos valores de diferenças revela uma imagem com regiões uniformes ou com transições suaves entre uma região e outra. Grandes diferenças indicam imagens apresentando texturas com pequenas estruturas ou contaminação de ruídos. Essa discriminação é útil para diferenciar desenhos como ilustrado nas Figura 101*a-b* de fotografias (Figura 101*c*).

As imagens em níveis de cinza são muitas vezes suficientes para caracterizar uma imagem. No entanto, cores não só aperfeiçoam essa caracterização, mas também permitem que uma segmentação ocorra considerando-as como atributos. Na verdade, a cor é produto da percepção de um aparelho visual, cujos fotosensores sinalizam faixas determinadas de freqüência da luz. O olho humano possui fotosensores sensíveis à cor, os cones, sendo que cada um deles "prefere" sinalizar um determinado intervalo de freqüência da luz. No entanto, três intervalos são os mais freqüentes nos cones, os quais identifica-se como três cores principais do espectro de percepção visual humana: o vermelho, o verde e o azul.

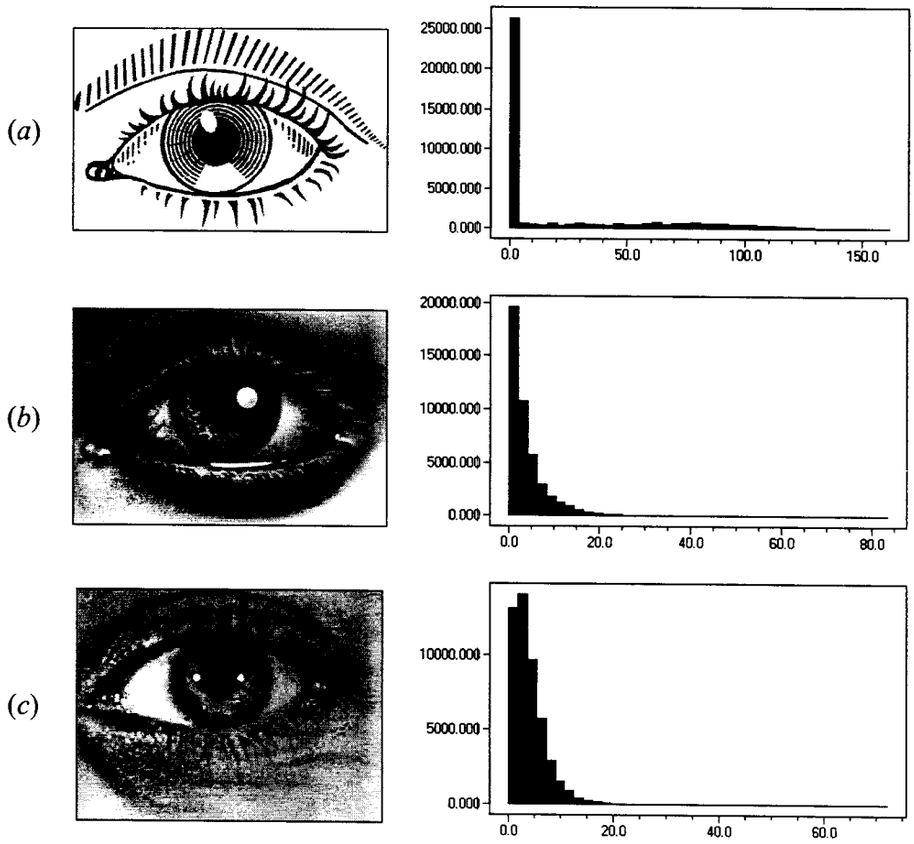


Figura 101 – Exemplos de histogramas de contrastes em níveis de cinza para imagens de (*a*) um desenho; (*b*) um desenho gráfico; (*c*) uma fotografia. Percebe-se que os histogramas de cada imagem são distintos e podem caracterizar o processo de confecção da imagem.

Uma imagem percebida pelo aparelho visual de um indivíduo em um dado momento é única e portanto não existe uma cor única para todos os indivíduos. No entanto, alguns mapeamentos de cores buscam uniformizar o vocabulário de cores dos seres humanos. Dentre estes mapeamentos está o RGB (*Red*, *Green*, *Blue*), utilizado em grande parte dos dispositivos de exibição e aquisição de imagens como monitores, scanners, câmaras. Em um mapeamento RGB, uma imagem colorida é composta por três canais, ou seja, três imagens cujas intensidades de seus respectivos pixels indicam as contribuições dos componentes (vermelho, verde ou azul) para formar a cor de cada um deles.

Para medir o contraste, as intensidades dos três canais são dispostas formando um espaço tridimensional e ortogonal RGB, no qual cada eixo corresponde a um dos componentes.

Cada pixel é um vetor $\vec{v}_{[i,j]} = (\mathrm{Im}_R[i,j], \mathrm{Im}_G[i,j], \mathrm{Im}_B[i,j])$. Assim, o contraste será maior entre dois pixels quando os dois vetores que os representam forem ortogonais. Portanto, para definir uma medida de contraste, deve-se considerar o produto escalar entre dois vetores representando pixels vizinhos. Para isso, uma imagem de contraste em cores D_{cor} é definida, cujos elementos são formados pela média entre os quatro produtos escalares entre o vetor do pixel corrente e os vetores de seus pixels vizinhos.

$$D_{cor}[i,j] = \frac{\vec{v}_{[i,j]} \cdot \vec{v}_{[i+1,j]} + \vec{v}_{[i,j]} \cdot \vec{v}_{[i,j+1]} + \vec{v}_{[i,j]} \cdot \vec{v}_{[i-1,j]} + \vec{v}_{[i,j]} \cdot \vec{v}_{[i,j-1]}}{4} \qquad (75)$$

A partir do histograma da imagem D_{cor} obtida por (75), é possível definir atributos que possibilitem caracterizar o contraste entre cores.

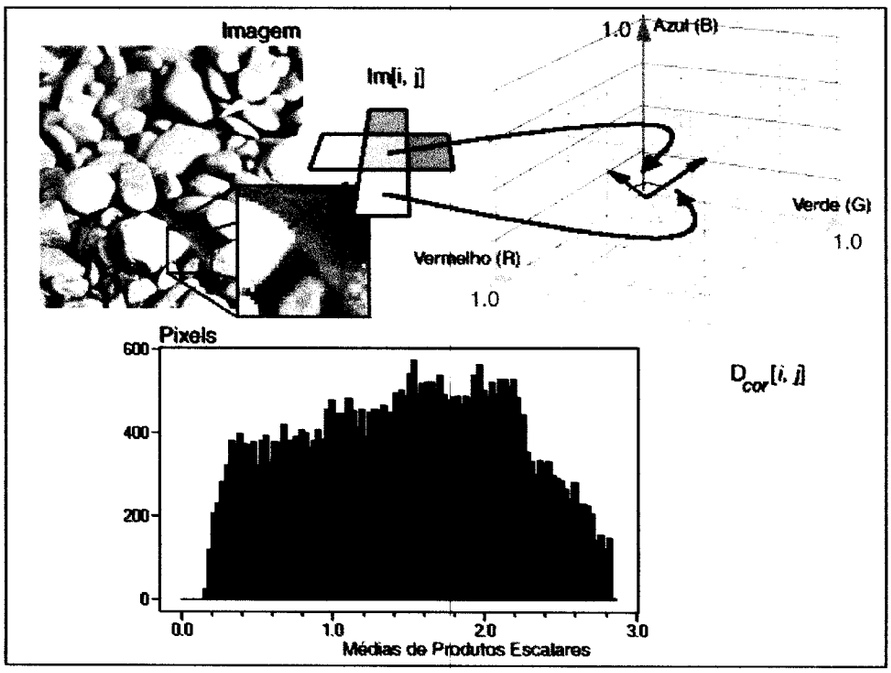


Figura 102 – Ilustração do processo de extração do histograma de contraste em cores de uma imagem. Cada pixel é um vetor no espaço RGB e o contraste é caracterizado pelo produto escalar entre o pixel corrente e seus 4 vizinhos. O histograma de médias dos produtos escalares é o conjunto de medidas que caracteriza esse contraste.

Para ilustrar essa caracterização, uma imagem com 3 objetos com as cores básicas do mapeamento RGB sobre um fundo preto e branco é submetida a sucessivos borramentos ou perdas de contraste. Podemos perceber que conforme se perde o contraste, mais uniforme se torna o histograma (Figura 103).

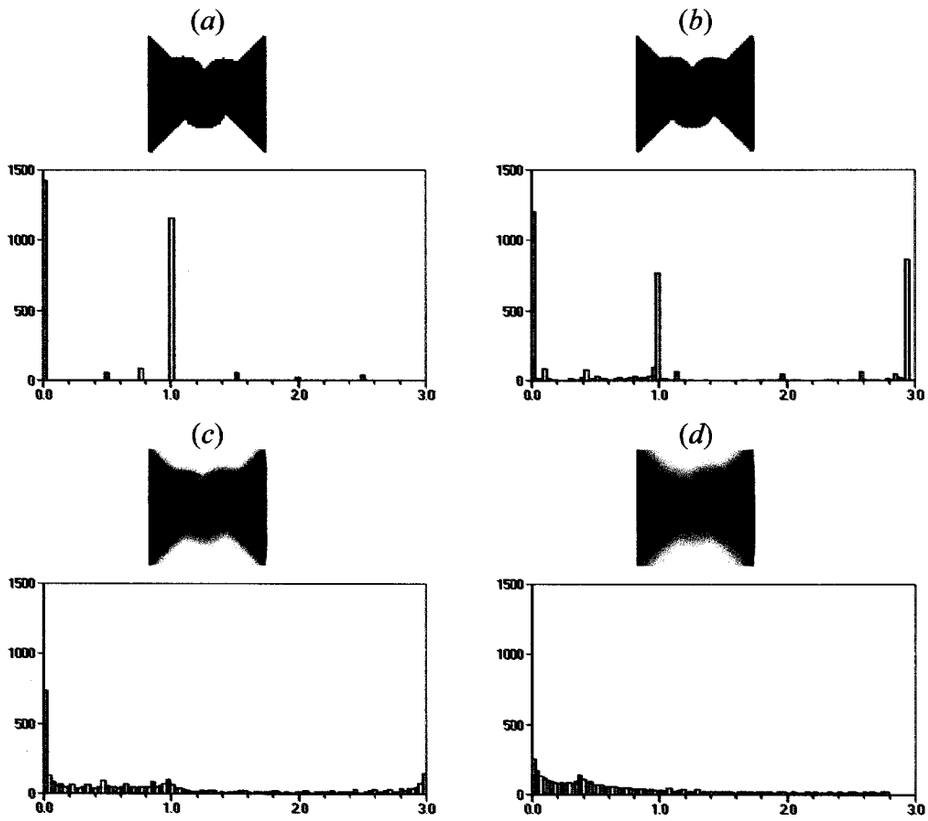


Figura 103 – Exemplo de histogramas de contrastes para uma imagem colorida com grandes contrastes (a) e que em seguida é sucessivamente borrada para diluir as bordas (b); (c) e (d). Percebe-se, conforme esperado, uma uniformização do histograma de contraste de cores conforme o borramento aumenta.

6.1.3 complexidade multi-escalas

A complexidade é um termo utilizado com grande abrangência em diversas áreas científicas. Watt [WATT, 1988] divide o conceito de complexidade em duas categorias principais: a complexidade de sistemas caóticos (previsões de tempo, mercados, sistemas biológicos) e a complexidade ontológica (complexidade algorítmica, de percepção de padrões, cognitiva e estrutural). Esta última, a ontológica, é a que está relacionada com esta seção.

A relação entre complexidade e percepção ocorre quando um sistema visual precisa reduzir as bandas de aquisição sensorial buscando otimizar a quantidade de informação capturada do ambiente.

Estimadores têm sido propostos para caracterizar a complexidade em imagens. As abordagens para construir tais estimadores variam em um vasto

conjunto de técnicas cuja meta é quantificar tal característica. Tais estimadores buscam métricas para características associadas à complexidade como simetria, similaridade, suavidade, saliências. As técnicas para calcular estas métricas são definidas em diferentes abordagens das quais as baseadas em estatística, geometria e na teoria da informação foram selecionadas para este trabalho.

Na abordagem estatística, a variância da intensidade em níveis de cinza indica quantos "tons" de cinza compõem uma determinada imagem [JÄHNE, 1997]. Já a abordagem geométrica considera a dimensão fracionária como um estimador de similaridade entre estruturas contidas na imagem e também da área de cobertura destas estruturas [KELLER et al., 1989]. Representando a teoria da informação, a entropia é considerada como medida de quantidade de informação necessária para descrever uma determinada imagem [SONKA et al., 1993].

Cada uma destas medidas consideradas em diversas escalas revela a "duração" das estruturas que compõem a imagem [LINDEBERG, 1994]. Deste modo, estruturas mais salientes resistem mais e portanto aparecem em escalas maiores influenciando estas medidas de complexidade.

As sucessivas escalas são conseguidas por sucessivos borramentos. Para efetuar um borramento, a imagem é convoluída com uma outra imagem de um núcleo gaussiano. Para reduzir o tempo de execução, contudo, a imagem gaussiana é gerada a partir do espaço de freqüência. Assim, seja uma imagem $f(x, y)$ cujas dimensões são dadas por $1 \leq x \leq 640$ e $1 \leq y \leq 480$. Seja também uma imagem gaussiana $g(x, y)$ gerada como resultado da transformada de Fourier $G(u, v)$ composta de partes real e imaginária (76).

$$G(u,v) = e^{-2\cdot(\pi\sigma\sqrt{u^2+v^2})^2} \text{ com } -0.5 \leq u \leq 0.5 \text{ e } -0.5 \leq v \leq 0.5 \qquad (76)$$

Cada borramento é realizado pela convolução de $f(x, y)$ com $g(x, y)$ (77) aproveitando-se do Teorema da Convolução [CASTLEMAN, 1996]. Assim, em uma escala σ, e tomando a transformada de Fourier para $\mathrm{F}\{f(x, y)\} = F(u,v)$ temos:

$$f * g_\sigma = \Im^{-1}\{F(u,v) \cdot G_\sigma(u,v)\} \qquad (77)$$

Foi adotada uma escala mista para σ com o objetivo de analisar também os borramentos mais extremos, utilizando a possibilidade aberta pelo aplicativo de permitir a inserção de elementos na escala ou fazê-la linear. Assim, σ iniciou-se com

0, ou seja, a imagem sem borramentos e a partir daí cresceu linearmente de 1 até 16. Prosseguiu crescendo, porém exponencialmente com σ assumindo valores 32, 64 e 128. É bom ressaltar que este valor de σ, embora introduzido como parâmetro no espaço da imagem gaussiana transformada (freqüência), indica o desvio padrão no espaço original da imagem (deslocamento). (Figura 104)

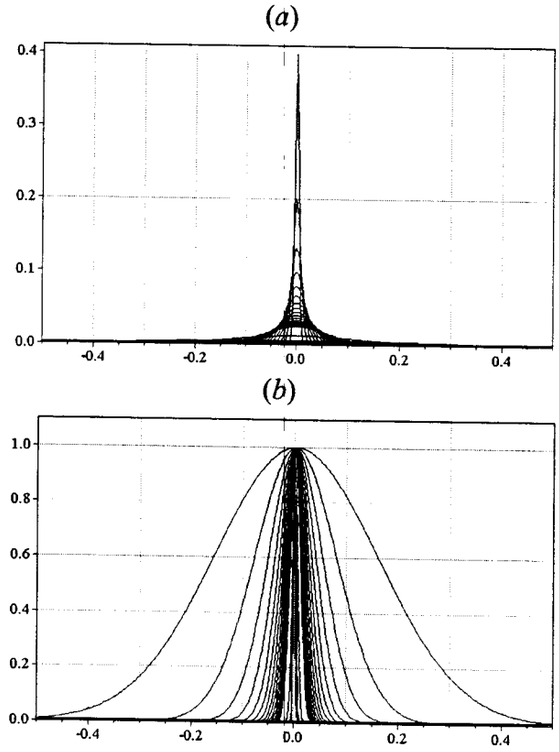


Figura 104 – Conjunto de gaussianas usadas para os borramentos sucessivos, com o intuito de obter as várias escalas nas medidas de entropia e variância multi-escalas. (a) no espaço de origem; (b) noespaço de freqüência.

Para cada conjunto de medidas multi-escalas foram calculados parâmetros estatísticos (média e variância). Outro parâmetro busca caracterizar o conjunto de medidas por apenas um fator. Por exemplo, dimensão fracionária para o conjunto de volumes da salsicha de Minkowski ou a meia-vida caracterizando o decaimento da variância multi-escalas. A seguir, cada medida será detalhada, com sua respectiva justificativa.

6.1.3.1 entropia multi-escalas

A aplicação da medida de entropia em processamento de imagens provém do conceito introduzido por Shannon [SHANNON, 1948] na formulação da teoria da

informação. Clausius em 1854, atribuiu o termo grego (evolução ou transformação), para entender a natureza do calor e a irreversibilidade em sistemas termodinâmicos, baseando-se no trabalho de Carnot (1824) e Kelvin (~1850). Esse trabalho resultou na formulação da segunda lei da termodinâmica ($dS = dq/T \geq 0$; na qual dq representa o fluxo de calor e T a temperatura), que resumidamente afirma que a entropia sempre cresce em sistemas reais [STEVENS, 1998].

A entropia revela-se como uma propriedade macroscópica de sistemas (tal como volume e tamanho). No entanto, Boltzmann procurou uma explicação estatística reduzindo o conceito de entropia para níveis moleculares e chegou à forma $S = k\ln W$ (W é o número de estados possíveis do sistema e k é a constante de Boltzmann) [STEVENS, 1998].

Shannon, aproveitando a idéia de Boltzmann, foi quem definiu uma medida de quantidade de informação dada pela ocorrência de um símbolo em uma mensagem. Assim, quanto mais raro um símbolo em uma mensagem, mais ele estaria contribuindo para aumentar a quantidade de informação passada pela mensagem [STEVENS, 1998].

Definindo W como a multiplicidade de um sistema (78), e pensando esse sistema como imagens:

$$W = \frac{N!}{\prod_i N_i!} \tag{78}$$

Seja N o número de pixels em uma imagem e conseqüentemente $N!$ o número de imagens possíveis com N pixels. $N_i!$ é o número de possíveis configurações dos pixels com nível de cinza i em uma imagem. W indica, portanto, todas as imagens possíveis com N pixels para um histograma N_i [BISHOP, 1995; HAKEN, 1977].

Para se perceber o acréscimo de informação de uma maneira aditiva e não multiplicativa, utiliza-se o logarítmo da medida W multiplicado por uma constante k[3], que mais adiante terá seu papel definido.

$$I = k\ln W \tag{79}$$

---

[3] Em sistemas termodinâmicos k é a constante de Boltzmann e vale 1.380658×10^{-23} Joules/Kelvin.

A entropia de informação, segundo Shannon, indica a quantidade de informação por símbolo. Para obtê-la, portanto substitui-se (78) em (79) e então $S = I/N$.

$$S = k\frac{1}{N}\ln\left(\frac{N!}{\prod_i N_i!}\right) = k\frac{1}{N}\left\{\ln N! - \sum_i \ln N_i!\right\} \tag{80}$$

Usando a aproximação de Stirling para $N!$ com N suficientemente grande, $\ln N! \approx N(\ln N - 1)$ em (80).

$$S = -k\left\{\sum_i \frac{N_i}{N}\ln\left(\frac{N_i}{N}\right)\right\} \tag{81}$$

Em (81) observa-se que $p_i = N_i/N$ é a probabilidade do pixel i ser encontrado na imagem. Portanto, pode-se escrever (81) como:

$$S = -k\sum_i p_i \ln p_i \tag{82}$$

A constante k, para a teoria de informação, pode ser arbitrariamente escolhida considerando alguma definição. Podemos então adotar k conforme a unidade de informação, que por conveniência, poderia ser a quantidade de "bits". Isso é possível fixando em (79) a quantidade de informação $I = n$, ou seja, uma mensagem com n bits e que contenha todos os símbolos do alfabeto tem $W=2^n$ configurações possíveis. Portanto, para se medir a quantidade de informação contida em uma imagem em bits, usa-se $k = 1/\ln 2$ e (82) pode agora ser escrito como (83).

$$S = -\sum_i p_i \log_2 p_i \tag{83}$$

A função densidade de probabilidade (fdp) de uma imagem digital é calculada a partir do histograma de intensidades de níveis de cinza da imagem. Cada valor $H_{\mathrm{Im}}[q]$ do histograma é o número de pixels cuja intensidade é $0 \le q \le 255$. Portanto, a fdp $P_{\mathrm{Im}}[q]$ da imagem é dada por (84). M e N são as dimensões das imagens e portanto, MN, o número total de pixels. A entropia da imagem é obtida por (85).

$$P_{\mathrm{Im}}[q] = \frac{H_{\mathrm{Im}}[q]}{MN} \tag{84}$$

$$Ent_{\mathrm{Im}} = -\sum_{q=0}^{255} P_{\mathrm{Im}}[q]\log_2(P_{\mathrm{Im}}[q]) \qquad (85)$$

A entropia multi-escalas é o arranjo formado pelas entropias das imagens resultantes de sucessivos borramentos de uma imagem original Im[i,j]. A escala é definida pelo parâmetro de suavização t = {0..16, 32, 64, 128}. O arranjo Ent_t (85), ou seja, a entropia sobre a imagem com borramento t, define um conjunto de atributos. A partir de Ent_t também são extraídos a média μ_{Ent} e o desvio padrão σ_{Ent}.

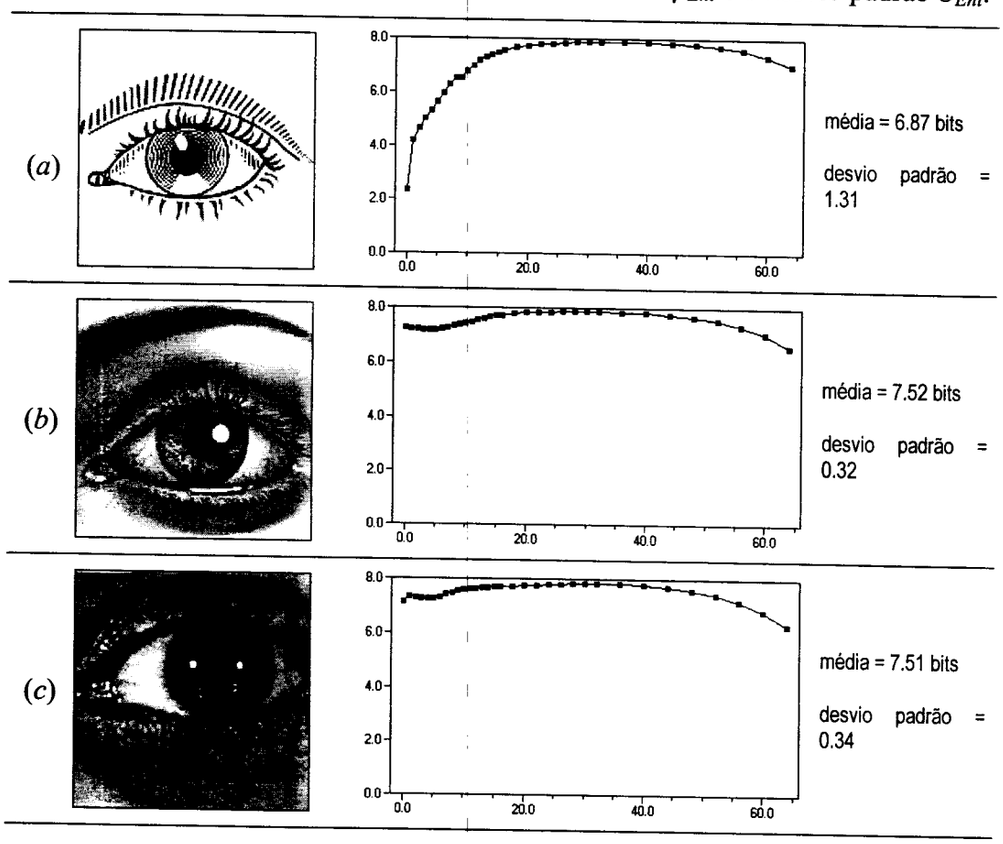


Figura 105 – Três tipos de imagens (a) desenho com linhas; (b) desenho gráfico com uso de sombreamentos; (c) fotografia de um olho humano. Ao lado, o gráfico de entropia multi-escalas de cada imagem. Percebe-se um comportamento diferente no desenho de linhas (a) e pouca diferença entre (b) e (c). A coluna mais à direita mostra a entropia média e o desvio padrão.

Na Figura 105 podemos observar um conjunto de imagens com diferentes características. Uma imagem com o desenho de um olho humano composto de curvas e preenchimentos em um fundo branco apresenta uma entropia baixa. Conforme a imagem é borrada pode-se notar a entropia aumenta até alcançar o máximo de uma imagem de 128×128 pixels e 256 tons de níveis de cinza que é 8 bits. Um outro

desenho do olho humano, agora usando recursos de desenho por computador, com regiões suaves e tentando imitar fielmente a imagem real. Deste modo, a imagem original já apresenta uma entropia alta, ou seja, há uma grande variedade de níveis de cinza por toda a imagem.

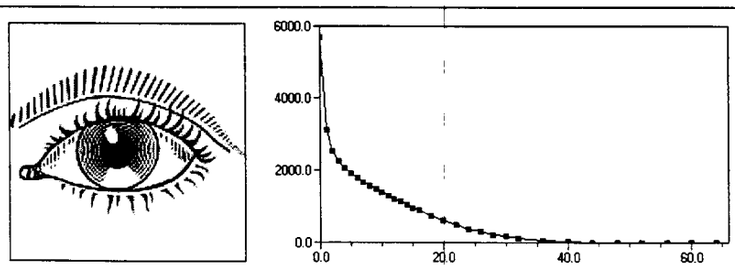
Os atributos resumo, ou seja, média e desvio padrão podem diferenciar o tipo de imagem também. Para o exemplo da Figura 105, podemos perceber que o desvio padrão para a imagem de desenho é maior que para as imagens fotográfica e do desenho mais sofisticado.

6.1.3.2 variância multi-escalas e meia-vida

A variância (86) é uma medida estatística de quanto, em uma imagem, os níveis de cinza variam em torno do nível de cinza médio. Assim como na entropia multi-escalas, a variância pode ser calculada ao longo de borramentos gaussianos da imagem resultando em um arranjo V_t, com t = {0..16, 32, 64, 128}. Além da variância média e do desvio-padrão das variâncias, um outro atributo resumo é testado nesse arranjo: a meia-vida. A meia-vida é típica de medidas que decrescem exponencialmente e indica a velocidade dessa queda. Para borramentos sucessivos em imagens, a meia-vida da variância indica a escala na qual a variância é igual a metade da variância da imagem original, isto é, $h_{\mathrm{Im}} = \arg\{V_t = V_0/2\}$. Na Figura 106, podemos observar que na imagem de desenho, a queda é bem mais rápida, isto é h_{Im} é bem menor que nas outras duas imagens.

$$Var_{\mathrm{Im}} = \sigma^2\{\mathrm{Im}[i,j]\} = E\{\mathrm{Im}^2[i,j]\} - E^2\{\mathrm{Im}[i,j]\} \qquad (86)$$

(*a*)



média = 1062.81

desvio padrão = 1189.9

meia-vida = 1.50

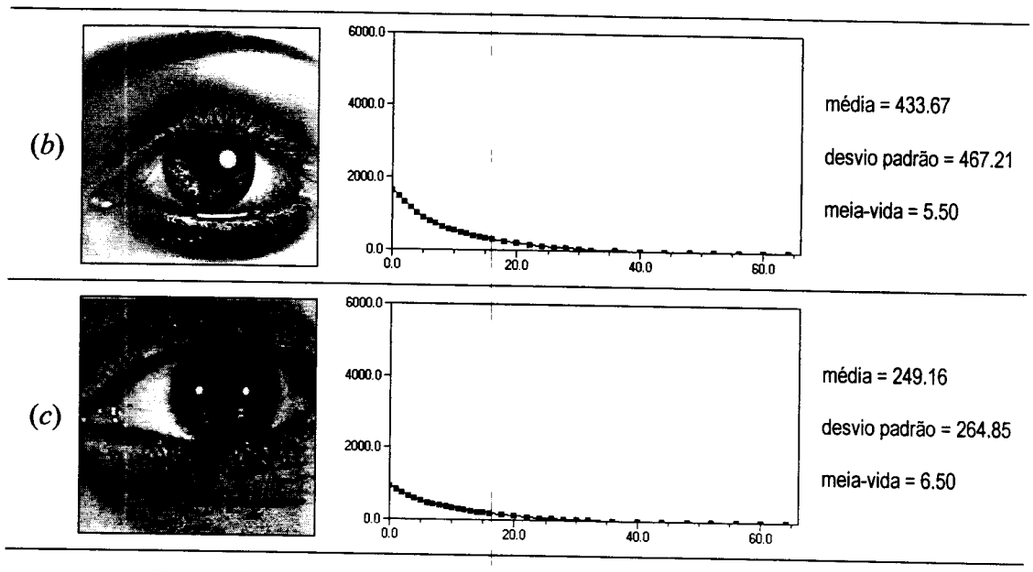


Figura 106 - (a) desenho com linhas; (b) desenho gráfico com uso de sombreamentos; (c) fotografia de um olho humano. Ao lado, o gráfico de variância multi-escalas de cada imagem. A maior discrepância está no desenho de linhas. Na coluna mais à direita aparecem as medidas resumo: variância média, desvio padrão da variância multi-escalas e a escala de meia-vida.

6.1.3.3 salsicha de Minkowski e dimensão fracionária

A geometria fracionária tem sido alvo de freqüente interesse na busca de descrição de imagens. As justificativas para esse interesse baseiam-se em um tipo de geometria que descreve o mundo natural de uma maneira elegante e econômica. Esta descrição pode manifestar-se em uma medida de auto-similaridade (usada em texturas, por exemplo), ou ainda em descrições estocásticas ou determinísticas de processos de formação de um sinal [JAIN et al., 1995].

A análise dimensional atribui uma medida, a dimensão fracionária, a uma imagem ou forma como propriedade estrutural [KRAFT & KAUER, 1995]. A dimensão topológica também define uma medida estrutural, contudo inteira. Objetos como curvas, superfícies e sólidos apresentam uma dimensão topológica 1, 2 e 3, respectivamente. Contudo curvas podem cobrir uma superfície e superfície, por sua vez, podem formar sólidos. Mandelbrot [MANDELBROT, 1977] desenvolveu então, modelos fracionários de curvas, superfícies e sólidos para descrever objetos naturais como: montanhas, árvores, lagos e rios e chamou-os *fractais*. Estes modelos são baseados na lei de potência [BAK, 1996] dada pela relação em (87) e no caso de

dimensão fracionária s é a unidade métrica de um espaço e $N(s)$ é o número dessas unidades que o objeto ocupa nesse espaço.

$$N(s) = s^{-d} \tag{87}$$

A dimensão fracionária excede a dimensão topológica euclideana e deste modo, uma curva pode ser usada para cobrir uma parte de uma superfície ($d_f \leq 2$) ou a superfície inteira ($d_f = 2$). Assim, o número d_f é usado como uma medida de rugosidade ou cobertura espacial de formas ou imagens.

A dimensão fracionária pode ser determinada teoricamente se houver conhecimento do processo de formação do objeto. Por exemplo, para uma curva de Koch, a evolução do número de segmentos $N(s)$ que compõe a curva de Koch para uma interação n é dada por uma regra, ou seja, $N(s) = 4^n$. Já o tamanho do menor segmento de linha reta é iterado conforme a regra $s = (1/3)^n$ (veja Figura 107). Estas regras de evolução permitem-nos usar a lei de potência (87) para encontrar d. O uso de tais regras de evolução nas leis de potência resulta em (88).

$$4^n = \left(\frac{1}{3^n}\right)^{-d} \tag{88}$$

Tomando-se os logarítmos de ambos os lados, resultará:

$$d = \frac{\ln(4^n)}{\ln(3^n)} = \frac{\ln 4}{\ln 3} \cong 1.2619 \tag{89}$$

Em análises de imagens e formas, tais modelos não são conhecidos. Então, para se obter a medida d_f é preciso construir um estimador da dimensão fracionária.

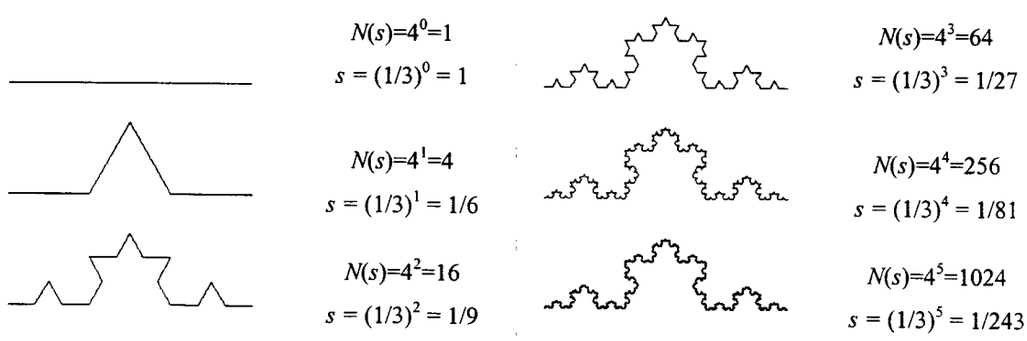


Figura 107 – Curva de von Koch em cada escala. Ao lado são mostrados o número de segmentos ($N(s)$) e o tamanho de cada segmento s.

Os métodos para elaborar tais estimadores são numerosos [ALLEN et al., 1995] e dentre eles, alguns comprovadamente reduzem o erro com relação à medida

teórica (89) . Dentre estes com pequeno erro, o método de salsicha de Minkowski ou EDM (*Euclidean Dilation Method*) é bastante utilizado e validado [ALLEN et al., 1995].

O método da salsicha de Minkowski consiste em crescer o objeto (forma ou imagem) cuja dimensão deve ser medida. O crescimento deve ser isométrico e para implementar o método, são realizadas sucessivas dilatações do objeto com elementos estruturantes circulares, para o caso de curvas, ou esféricos para o caso de superfícies. Para cada dilatação, é tomada a medida de cobertura do objeto crescido. Assim, para o caso de linhas, são tomadas as áreas de cada dilatação. Se o objeto for uma superfície, são tomados seus volumes dilatados.

$$
\begin{aligned}
A(r) &\propto kr^{2-d} & V(r) &\propto kr^{3-d} \\
\ln[A(r)] &\propto \ln\left[kr^{2-d}\right] & \ln[V(r)] &\propto \ln\left[kr^{3-d}\right] \\
\ln[A(r)] &\propto \ln k + (2-d)\ln r & \ln[V(r)] &\propto \ln k + (3-d)\ln r
\end{aligned}
\tag{90}
$$

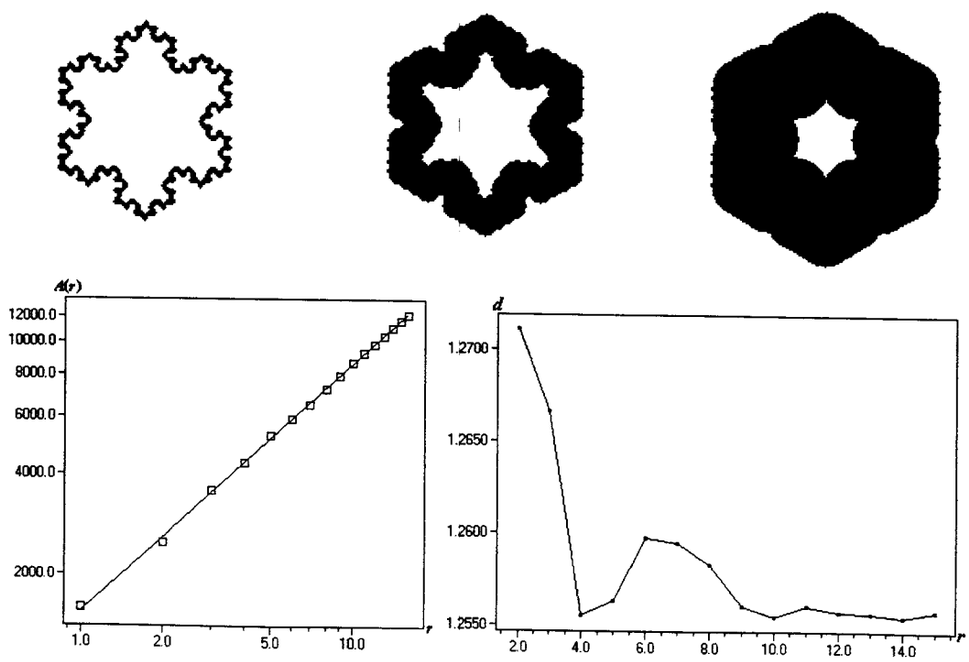


Figura 108 – Floco de neve de von Koch, iniciado a partir de um triângulo e suas respectivas salsichas de Minkowski. Abaixo o gráfico da esquerda mostra a evolução da área da salsicha em um log-log. À direita o gráfico mostrando para cada raio de dilatação a dimensão fracionária estimada. Percebe-se que a dimensão teórica está mais próxima quando o r=6.

O crescimento do raio r do elemento estruturante $e(r)$ que irá dilatar o objeto causa um aumento na área $A(r)$ do objeto se $e(r)$ for um disco, ou do volume $V(r)$ do objeto se $e(r)$ for uma esfera. Este aumento acontece segundo a relação (90) baseada na dimensão de Minkowski[4].

Para determinar o valor de d, são tomados os valores das áreas $A(r)$ ou volumes $V(r)$ para valores de r menores que um raio determinado empiricamente suficiente. Em um mapa log-log (Figura 108), o problema passa a ser determinar o coeficiente angular $m=2-d$ ou $m=3-d$ da reta ajustada. Uma vez ajustada a reta de $\log(r) \times \log(A(r))$, a dimensão fracionária estimada é d.

A dilatação do objeto é realizada convoluindo uma função S_r (91) com a imagem Im na qual o objeto está contido. Em seguida o resultado da convolução $\mathrm{Im}_r = \mathrm{Im} * S_r$ é limiarizado e a área $A(r)$ é obtida contando-se os pixels com valores não nulos. Uma outra opção é a dilatação morfológica. Nesse caso, o operador morfológico de dilatação é aplicado sobre Im, isto é, $\mathrm{Im}_r = \mathrm{Im} \oplus S_r$. A área $A(r)$ é dada pelo número de pixels não nulos. Assume-se que as imagens Im são binárias assim como os elementos estruturantes S_r. Usar um ou outro método depende do raio do elemento estruturante.

$$S_r[i,j] = \begin{cases} 255 & se\ \sqrt{i^2+j^2} \le r \\ 0 & caso\ contrário \end{cases} \qquad i,j = -r..r \tag{91}$$

A abordagem morfológica exige que cada pixel dentro do elemento estruturante seja visitado, o que torna inviável para os raios maiores. Se implementada a convolução utilizando a respectiva propriedade da transformada de Fourier e conseqüentemente o algoritmo rápido da FFT [BRIGHAM, 1988], o desempenho não dependerá do tamanho do raio, mas apenas das dimensões da imagem.

O algoritmo descrito trabalha com imagens binarizadas, ou seja, encontra a dimensão fracionária de imagens já segmentadas ou de formas. No entanto, o objetivo deste trabalho é encontrar uma medida para imagens em níveis de cinza Im.

A imagem **Im** define uma superfície $f(x, y)$, cuja dimensão fracionária pode ser estimada por diversos métodos[ALLEN et al., 1995]. Contudo, estimulado pelo

---

[4] {MINKOWSKI 1901 767 /id} *apud* {ALLEN, BROWN, et al. 1995 329 /id}

baixo erro com relação à medida teórica e também pela simplicidade do método da salsicha de Minkowski, procurou-se aplicar o mesmo algoritmo para superfície. As duas abordagens: convolução e dilatação, podem ser estendidas para a superfície. Contudo, a complexidade algorítmica de ambas inviabiliza uma extensão direta, pois a imagem **Im** em níveis de cinza deve ser mapeada em $\mathbf{R}^3$ como (92).

$$f_{\mathrm{Im}}[i,j,k] = \begin{cases} 1 & se\ k = \mathrm{Im}[i,j] \\ 0 & caso\ contrário \end{cases} \quad i=1..N;\ \ j=1..M;\ \ k=0..255 \qquad (92)$$

A proposta para resolver o problema é definir uma estrutura que permita um mapeamento da imagem em um matriz de bits. Cada bit corresponde a um nível de cinza. Portanto, cada pixel é representado por uma cadeia $B_{\mathrm{Im}}[i,j] = \{b_{ij}^0 b_{ij}^1 b_{ij}^2 \cdots b_{ij}^{255}\}$. A imagem mapeada em $B_{\mathrm{Im}}[i,j]$ exige menor quantidade de memória, pois enquanto f_{Im} ocupa $N*M*256$ valores reais, B_{Im}, por sua vez, precisa apenas de $N*M*8$ bytes. No entanto, apenas a abordagem morfológica é possível com B_{Im}, e mesmo assim, para valores grandes de r a operação de dilatação pode se tornar inviável. A solução proposta para esse problema é realizar a contagem dos pontos que não são preenchidos pela dilatação $B_{\mathrm{Im}} \oplus S_r$, com S_r definido em (93).

$$S_r[i,j,k] = \begin{cases} 1 & se\ \sqrt{i^2+j^2+k} \le r \\ 0 & caso\ contrário \end{cases} \quad i,j,k=-r..r \qquad (93)$$

Essa contagem torna desnecessário o preenchimento de todos o pixels do elemento estruturante na dilatação de cada ponto da superfície, já que importantes agora são os pontos do limite do elemento estruturante. Portanto, o elemento estruturante pode ser alterado para (94), ou seja, uma esfera oca o suficiente para não deixar "buracos" na superfície dilatada.

$$C_r[i,j,k] = \begin{cases} 1 & se\ r_{\mathrm{int}} \le \sqrt{i^2+j^2+k} \le r \\ 0 & caso\ contrário \end{cases} \quad i,j,k=-r..r \text{ e}$$

$$r_{\mathrm{int}} = \begin{cases} r\text{-}2 & se\ r > 2 \\ 0 & caso\ contrário \end{cases} \qquad (94)$$

O efeito da operação $B_{\mathrm{Im}} \oplus C_r$ é um "salsicha" (Figura 109c) cujo interior contém apenas os pontos do limite de C_r dilatado em B_{Im}. A Figura 109 ilustra esse efeito e mostra as partes de cima (Figura 109a) e de baixo (Figura 109b) da

contagem. Para cada pixel há dois contadores, um partindo do topo e indo para baixo ($cima[i, j]$) e outro partindo do nível mais baixo e indo ao topo ($baixo[i, j]$) até encontrar um bit em $B_{\mathrm{Im}}[i, j] = 1$. Quando isso ocorre com cada um dos contadores, a contribuição do pixel $[i, j]$ ao volume $Vol(r)$ é dada por $cima[i, j] - baixo[i, j]$, e portanto:

$$Vol(r) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(cima[i,j] - baixo[i,j]\right) \tag{95}$$

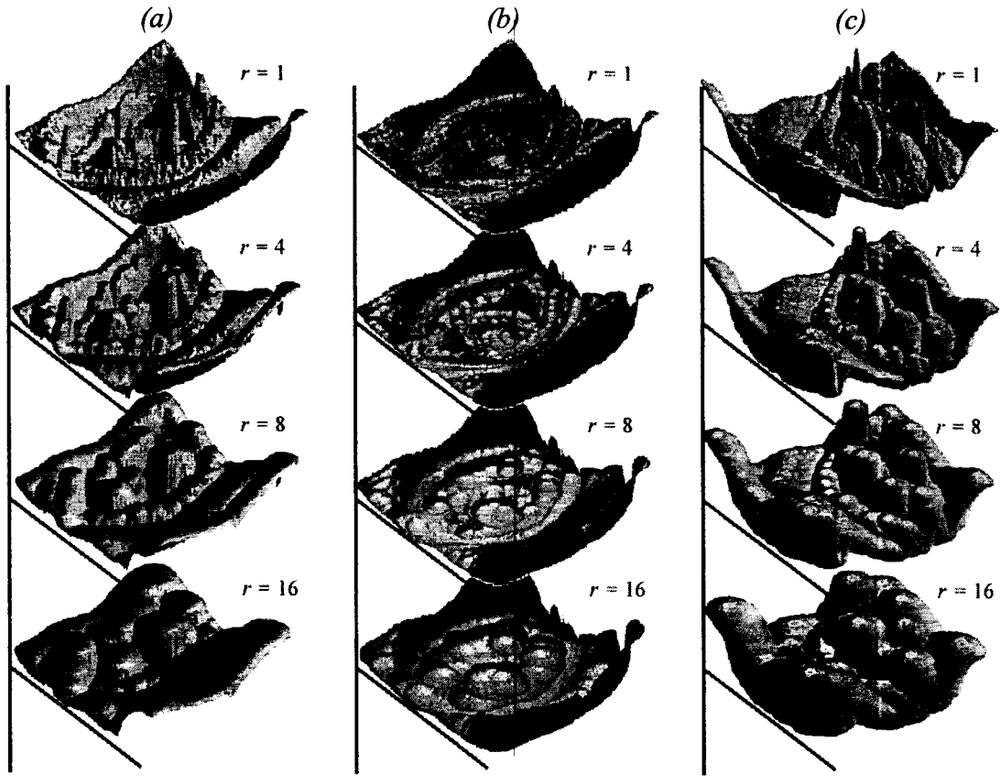


Figura 109 – Dilatações crescentes da salsicha de Minkowski da imagem da Figura 20 para diferentes raios r. A coluna (a) corresponde à parte de cima; a (b) à parte de baixo; e (c) todo o volume.

As "salsichas" acima correspondem à imagem da Figura 110b. Percebe-se que conforme o raio aumenta a "salsicha" é inflada e seu volume aumentado.

Para os pontos $\log(r) \times \log(Vol_r)$ é ajustada uma reta cujo coeficiente angular será dado por m=3 $-d$ (Figura 110a). A dimensão fracionária d define apenas o atributo resumo do conjunto de volumes $Vol(r)$, os quais são também atributos da imagem. Outros atributos resumo são o volume médio μ_{Vol} e desvio-padrão σ_{Vol}.

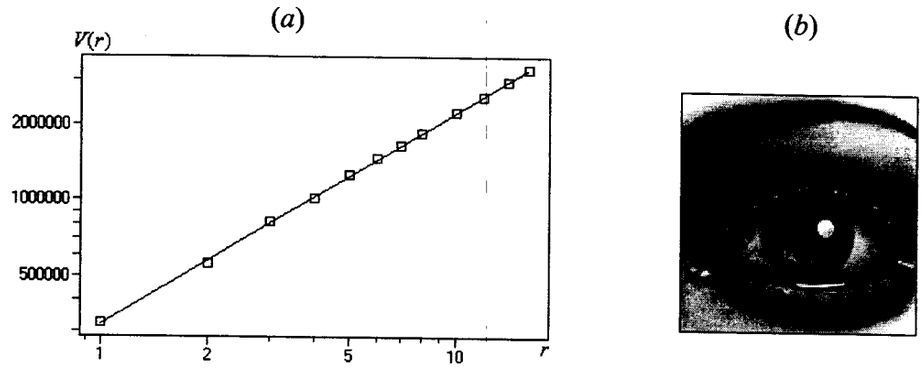


Figura 110 – Volumes da salsicha de Minkowski em um gráfico (a) log-log para a imagem (b)

Imagens podem apresentar a mesma dimensão fracionária, mas revelarem estruturas totalmente diferentes, como ilustram as imagens da Tabela (IV)*b* e (VI)*b* no Apêndice 9.3. No entanto, os respectivos volumes das salsichas indicam a diferença. A dimensão fracionária, nesse caso, não distinguiria as duas imagens, justificando o uso das medidas em cada escala para classificações ou discriminações de imagens. Observa-se ainda, que propriedades de imagens similares podem ser detectadas. No Apêndice 9.3 podemos ver pares ou grupos de imagens com suas respectivas dimensão fracionária e volumes. Nas Tabelas (I), (IV), (V), cada par apresenta uma imagem e seu detalhe para ilustrar que imagens similares podem indicar propriedades, neste caso, detalhes ou aproximação. Outros dois pares (Tabelas (III) e (VI)) mostram imagens com propriedades texturais e de escala diferentes.

Conclui-se que a dimensão fracionária como uma medida única para rugosidade e complexidade não teria sucesso em um conjunto de imagens com ampla diversidade. Contudo, o acompanhamento da evolução das escalas possibilitado pelos volumes da salsicha de Minkowski resolve esse problema.

6.2 A Busca pelo Modelo

Embora Birkhoff [BIRKHOFF, 1933] tenha definido a medida estética como uma fórmula bastante simples (68), Gunzenhäuser[5] já apontava para uma medida que relaciona a avaliação subjetiva (estado macroestético) com as medidas de

[5] apud {BENSE 1975 315 /id}

uma imagem (estado microestético). Maser[6] propõe essa relação como um somatório de medidas sobre um conjunto de informação.

Bense [BENSE, 1975] propõe a atribuição de valor estético a obras de arte, baseado em um esquema de comunicação estética ilustrado abaixo. Neste esquema, o artista é responsável pela codificação da obra usando medidas estéticas como textura, cor, forma, entre outros para descrever uma imagem. Ao observador cabe a decodificação da obra de arte e a atribuição de um valor estético. Existe portanto, uma sobreposição das representações do artista e do observador. Esta sobreposição, ilustrada no esquema (Figura 111) como uma intersecção é o conjunto de medidas estéticas determinantes para a avaliação do observador. Bense ressalva que este esquema é pertinente para avaliações estéticas nas quais o contexto não é o objetivo.

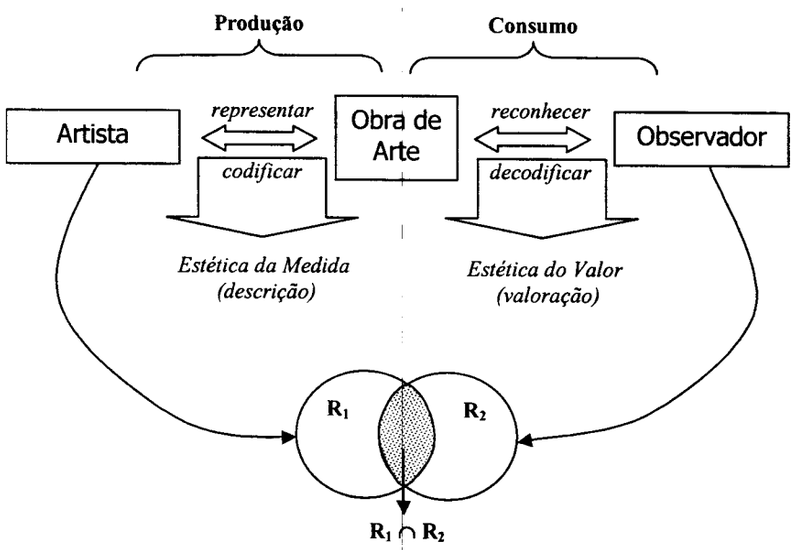


Figura 111 – Esquema de comunicação estética proposto por Bense em [BENSE, 1975].

Neste trabalho, as medidas estéticas são as medidas texturais, de complexidade e de contraste realizadas sobre as imagens e que são armazenadas em tabelas. Os valores são os critérios estéticos de agradabilidade, cores, originalidade, qualidade artística e motivação. Estes critérios também constam em tabelas.

O que se busca então, é uma representação que maximize um determinado valor estético. Para buscar essa representação foi construído um modelo

---

[6] apud {BENSE 1975 315 /id}

preditivo[EDMONDS, 1997] que generalize, com um erro suficientemente pequeno, os critérios estéticos avaliados por humanos.

A Figura 112 ilustra a abordagem deste trabalho. O sujeito humano é visto como uma "caixa preta" da qual a aplicação de estímulos (imagens) resulta uma resposta que no caso é o valor dado ao critério estético. No modelo pretendido são introduzidas as medidas destas imagens e que responderão com valores aos critérios estéticos atribuídos pelos sujeitos humanos.

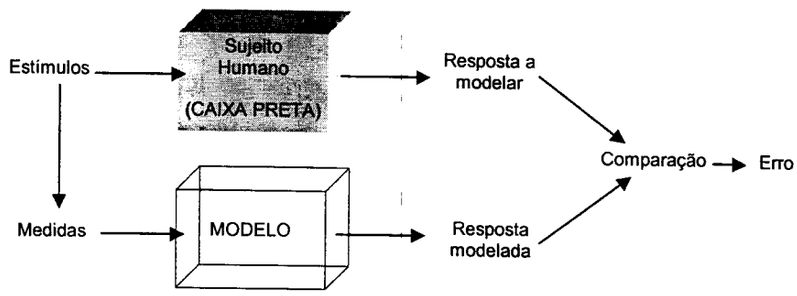


Figura 112 – Modelo "caixa preta" para o ser humano.

As respostas a serem modeladas são as médias das notas atribuídas a cada imagem pelos sujeitos humanos. Portanto, o modelo refletirá esse mesmo comportamento médio. Para determinar a configuração de medidas que fazem o modelo responder notas médias com menor erro, foi implementado um sistema utilizando as tabelas descritas no Apêndice 9.1. O esquema abaixo ilustra essa implementação.

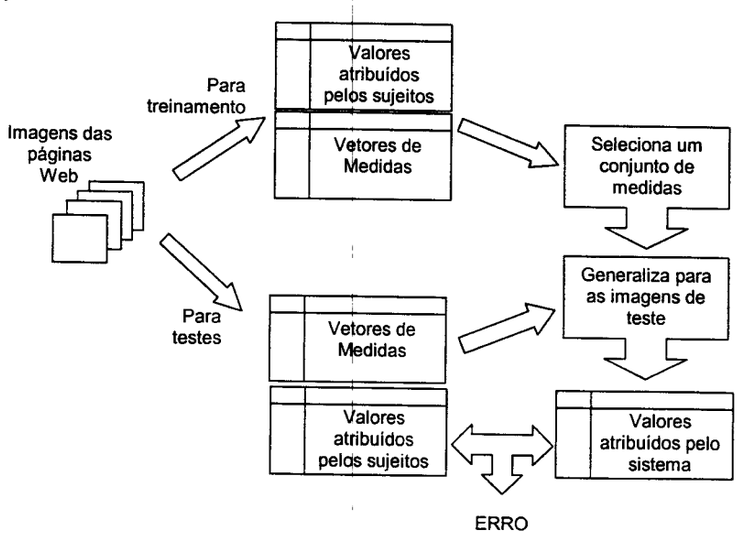


Figura 113 – Esquema ilustrando a avaliação do desempenho no treinamento da estimativa de critério estético de páginas web, baseado em notas atribuídas por sujeitos humanos.

As imagens são divididas em dois grupos: um para treinamento (20 imagens) e outro para testes (20 imagens).

Seja $\boldsymbol{C}$ o conjunto de critérios estéticos $\boldsymbol{C} = \{$*agradabilidade, cores, originalidade, qualidade artística, motivação*$\}$.

Seja $\boldsymbol{M}$ conjunto de todas as medidas calculadas das imagens. Este conjunto é composto por $N_D = 152$ medidas possíveis, sendo:

- 44 medidas texturais definidas na seção 6.1.1, ou seja, 20 faixas do histograma espectral espacial S_{esp} e 20 faixas do histograma espectral direcional S_{dir}. Para cada um deles é calculada a média (μ_{esp} e μ_{dir}) e o desvio padrão (σ_{esp} e σ_{dir}).
- 64 medidas de complexidade descritas em 6.1.3, isto é, 20 escalas de entropia, 20 escalas de variância e 16 raios de salsicha de Minkowski. Além dos momentos de cada conjunto, ou seja, médias e desvios padrões das entropias (μ_{Ent} e σ_{Ent}), das variâncias (μ_{Var} e σ_{Var}) e das salsichas (μ_{Vol} e σ_{Vol}); inclui-se também a meia-vida da variância (h) e a dimensão fracionária d_f.
- 44 medidas de contraste (seção 6.1.2), isto é, 20 faixas do histograma de contraste de cores H_{cores} e 20 faixas do histrograma de contrate em níveis de cinza H_{cinza}, além de seus respectivos momentos estatísticos média e desvio padrão: (μ_{Cores} e σ_{Cores}) e (μ_{Cinza} e σ_{Cinza}).

As imagens para treinamento, bem como as de teste possuem d medidas. O conjunto de medidas de cada imagem de treinamento é representado aqui por um vetor $\vec{u}_k = (u_k^1, u_k^2, \cdots, u_k^d)$ $(k = 1..20)$ e das imagens de teste como $\vec{v}_j = (v_j^1, v_j^2, \cdots, v_j^d)$ $(j = 1..20)$. Os dois conjuntos de imagens possuem também notas atribuídas pelos sujeitos. Sugere-se então representá-las como α_k^c as notas das imagens de treinamento para o critério estético $c \in \boldsymbol{C}$ e como β_j^c as notas das imagens de teste. Os vetores de medidas $\vec{u}_k$ e as notas α_k^c vão servir como parâmetros para o modelo, que consiste de uma hipersuperfície interpolando os

pontos $(\vec{u}_k, \alpha_k^c)$ no espaço de medidas definido em R^d em função de cada valor do critério $f_c(\vec{u}_k): R^d \mapsto R$ atribuído pelos sujeitos humanos.

Uma vez obtida a hipersuperfície f_c, a generalização é conseguida aplicando os vetores de medidas das N_{testes}=20 imagens de teste $\vec{v}_j$ (j = 1..N_{testes}) a ela, ou seja, $f_c(\vec{v}_j)$. Esse valor obtido é comparado com a nota média da respectiva imagem. A comparação é realizada tomando a distância (96) entre as respostas das imagens e suas notas.

$$Erro^c = \sqrt{\sum_{j=1}^{N_{testes}} (f_c(\vec{v}_j) - \beta_j^c)^2} \qquad (96)$$

No gráfico abaixo vemos um exemplo do resultado da aplicação $f_c(\vec{v}_j)$. Cada quadrado na Figura 114 representa uma imagem de treinamento k (preenchido) ou uma imagem de teste j (vazado). Para as imagens de treinamento os erros não existem pois são suas notas α_k^c que servem como parâmetros para a generalização . No entanto, as imagens de teste apresentam uma discrepância entre suas notas β_j^c e suas respectivas generalizações $f_c(\vec{u}_k)$. O tamanho de cada discrepância $f_c(\vec{v}_j) - \beta_j^c$ é representado pela linha vertical e o ponto em sua extremidade é o resultado da generalização para aquela imagem j específica. Observe que a escala de notas está normalizada no intervalo [-1, 1].

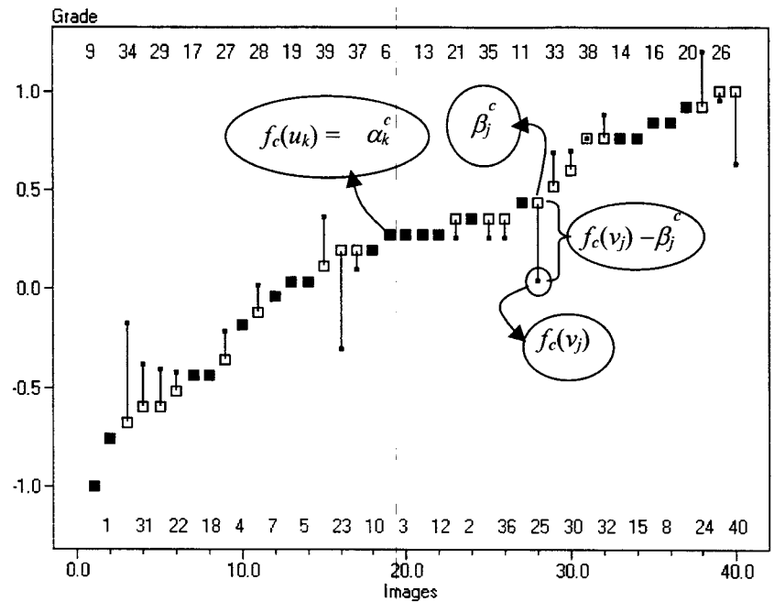


Figura 114 – Gráfico que mostra o resultado de uma estimativa de critério estético para 40 imagens. Os quadrados preenchidos representam as notas médias das imagens para treinamento; os quadrados vazados, as notas das imagens de teste e o ponto a estimativa para a imagem de teste. A linha indica o erro da estimativa.

6.3 A Generalização

A hipersuperfície é obtida interpolando-se uma "*thin-plate spline*" (TPS) multidimensional sobre os pontos $(\vec{u}_k, \alpha_k^c)$. A interpolação multidimensional é uma extensão do caso bidimensional descrito na seção 3.3.7 do capítulo 3.

Para o presente problema, o modelo almejado é um vetor de pesos $\vec{\omega}$ representando o modelo matemático, que operado sobre as medidas das imagens de teste $\vec{v}_j$, resultem em $f_c(\vec{v}_j)$. Isto pode ser conseguido de uma forma simples por um sistema linear. Contudo esta solução expressaria um relacionamento incompatível com a lei Weber-Fechner [GORDON, 1989].

Bookstein [BOOKSTEIN, 1989], resolvendo a equação diferencial parcial conhecida como equação biharmônica [POWELL, 1994] enfrentou um problema variacional cuja solução é dada por polinômios *splines* da forma (97).

$$f(\vec{u}) = \sum_{k=1}^{N_{treinamento}} w_k \phi(\|\vec{u} - \vec{u}_k\|) \qquad (97)$$

Em (97), os w_k indicam a suavização da base de funções radiais (FBR) ϕ e N o número de pontos disponíveis.

A escolha da função $\phi(\|r\|)$ pode ser realizada com uma certa liberdade, porém a função $\phi(\|r\|) = r^2\log(r)$ [POWELL, 1994] é a que minimiza a energia de dobramento de uma superfície e é chamada de "*thin-plate spline*" (TPS) devido a essa propriedade. Aqui, r é a distância considerando a norma do espaço em questão e a distância euclideana (98) é a norma adotada para este trabalho.

$$r = \|\vec{u} - \vec{v}\| = \sqrt{\sum_{i=1}^{d} (u_i - v_i)^2} \tag{98}$$

A abordagem considerada neste trabalho para TPS é a apresentada em (Mees et al., 1992), e permite graus de liberdade de torção, adicionando um polinômio linear $\vec{a} \cdot \vec{u}$ com um termo para cada dimensão de $\Re^d$. Além disso, a adição de um termo escalar b permite a translação da FBR $\phi(\|r\|)$ que servirá como base para $f(\vec{u})$. Estas adições, juntamente com imposições (100) sobre os limites de $f(\vec{u})$, viabilizam uma quantidade maior de soluções e conseguem assim hipersuperfícies afins. Esta abordagem é chamada também de TPS afins e se apresenta em (99).

$$f(\vec{u}) = \sum_{k=1}^{N_{treinamento}} w_k \phi(\|\vec{u} - \vec{u}_k\|) + \vec{a} \cdot \vec{u} + b \tag{99}$$

$$\sum_{k=1}^{N_{treinamento}} w_k = 0 \quad \text{e} \quad \vec{w} \cdot \vec{u} = \vec{0} \tag{100}$$

A equação em (99) pode ser vista como uma equação matricial que toma a forma de (101). Observe que a hipersuperfície $f_c(\vec{u}_k)$ contém o mesmo valor das notas para um determinado critério c, ou seja $f_c(\vec{u}_k) = \alpha_k^c$.

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = \left[\begin{array}{cccc|cccc:c} 0 & \phi(\|\vec{u}_1 - \vec{u}_2\|) & \cdots & \phi(\|\vec{u}_1 - \vec{u}_N\|) & u_{11} & u_{12} & \cdots & u_{1d} & 1 \\ \phi(\|\vec{u}_1 - \vec{u}_2\|) & 0 & \cdots & \phi(\|\vec{u}_2 - \vec{u}_N\|) & u_{21} & u_{22} & \cdots & u_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \phi(\|\vec{u}_1 - \vec{u}_N\|) & \phi(\|\vec{u}_2 - \vec{u}_N\|) & \cdots & 0 & u_{N1} & u_{N2} & & u_{Nd} & 1 \\ \hline u_{11} & u_{21} & \cdots & u_{N1} & 0 & 0 & \cdots & 0 & 0 \\ u_{12} & u_{22} & \cdots & u_{N2} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ u_{1d} & u_{2d} & \cdots & u_{Nd} & 0 & 0 & \cdots & 0 & 0 \\ \hdashline 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \hline a_1 \\ a_2 \\ \vdots \\ a_d \\ \hline b \end{bmatrix} \tag{101}$$

Para efeito de simplicidade em (101), os índices do critério estético foram suprimidos e $N_{treinamento}$ foi considerado apenas N. O índice d indica o número de

medidas (ou atributos) que estão sendo usados para obter os pesos. A FBR pode ser explicitada como (102).

$$\phi(\|\vec{u}_i - \vec{u}_k\|) = \|\vec{u}_i - \vec{u}_k\|^2 \log(\|\vec{u}_i - \vec{u}_k\|) \tag{102}$$

O problema é reduzido, portanto, a um sistema linear (101) cuja solução é dada por (103).

$$\begin{bmatrix} \vec{\alpha} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \Phi & U & 1 \\ U^T & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{a} \\ b \end{bmatrix} \Rightarrow \begin{bmatrix} \vec{w} \\ \vec{a} \\ b \end{bmatrix} = \begin{bmatrix} \Phi & U & 1 \\ U^T & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \vec{\alpha} \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \vec{\omega} = \mathrm{Q}^{-1}\vec{\alpha} \tag{103}$$

Uma vez obtido o vetor $\vec{\omega}^c$, ele poderá ser aplicado sobre as medidas das imagens de teste $\vec{v}$ fornecendo-nos uma estimativa $f_c(\vec{v})$ para o critério estético c (104). Novamente, para efeito de simplicidade, os índices do critério estético c foram suprimidos e N significa o número de imagens de treinamento, ou seja, $N = N_{treinamento}$ e M é o número de imagens de testes ($M = N_{teste}$).

$$\begin{bmatrix} f(\vec{v}_1) \\ f(\vec{v}_2) \\ \vdots \\ f(\vec{v}_M) \end{bmatrix} = \left[\begin{array}{cccc|cccc:c} \phi(\|\vec{v}_1 - \vec{u}_1\|) & \phi(\|\vec{v}_1 - \vec{u}_2\|) & \cdots & \phi(\|\vec{v}_1 - \vec{u}_N\|) & v_{11} & v_{12} & \cdots & v_{1d} & 1 \\ \phi(\|\vec{v}_2 - \vec{u}_1\|) & \phi(\|\vec{v}_2 - \vec{u}_2\|) & \cdots & \phi(\|\vec{v}_2 - \vec{u}_N\|) & v_{21} & v_{22} & \cdots & v_{2d} & 1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \phi(\|\vec{v}_M - \vec{u}_1\|) & \phi(\|\vec{v}_M - \vec{u}_2\|) & \cdots & \phi(\|\vec{v}_M - \vec{u}_N\|) & v_{M1} & v_{M2} & & v_{Md} & 1 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \hline a_1 \\ a_2 \\ \vdots \\ a_d \\ \hline b \end{bmatrix} \tag{104}$$

A aplicação de $\vec{\omega}^c$ para estimar $f_c(\vec{v})$ pode ser simplificada por (105).

$$[f(V)] = [\Phi \quad V \quad 1] \begin{bmatrix} \vec{w} \\ \vec{a} \\ b \end{bmatrix} \Leftrightarrow f_c(V) = \vec{\omega}\,\mathrm{P} \tag{105}$$

Temos agora uma maneira de generalizar um critério estético de imagens a partir de um modelo construído com um conjunto conhecido de imagens. Além disso, podemos avaliar o desempenho desse modelo utilizando (105).

Nem todas as N_D medidas das imagens precisam ser usadas para a generalização. Na verdade o problema que se desenha é selecionar as d medidas que mais influenciam na generalização com melhor desempenho, isto é, com o menor erro. O objetivo passa a ser determinar a combinação de medidas M_c com o melhor

desempenho na generalização do critério estético c. Apresenta-se, portanto um problema combinatório cuja solução deve responder duas questões:

(1) Quantas medidas são necessárias (dentre as N_D disponíveis) para interpolar os critérios estéticos cuja combinação resulte no melhor desempenho comparado com um outro conjunto arbitrário de imagens?

(2) Quais são estas medidas?

Para termos uma noção do tamanho do problema, o número de combinações a serem testadas é dado pela soma das combinações (binômio de Newton) possíveis em cada dimensão. Deste modo, o espaço de busca contém a quantidade de combinações dada por (106).

$$\sum_{j=2}^{N_D} \binom{N_D}{j} = 2^{N_D} - N_D \tag{106}$$

6.4 A Busca

Observando (106), vê-se que o espaço de busca para este problema é proibitivo com N_D já em torno de uma dezena. Conseqüentemente, uma estratégia de busca efetiva deve ser aplicada. Várias estratégias de busca existem na literatura, todas com vantagens e desvantagens. Métodos como "força bruta"[GEN & CHENG, 1997], algoritmos evolucionários [BÄCK, 1996], "*simulated annealing*"[KIRKPATRICK et al., 1983], ou buscas de subida/descida de montanha (uphill/downhill) [GEN & CHENG, 1997], ou ainda estratégias híbridas como "*tabu search*"[GLOVER & LAGUNA, 1997], têm sido propostos. O problema implicado pelo modelo adotado é caracterizado pelos seguintes aspectos: (1) o espaço de busca é desconhecido; (2) a solução pode não ser única o nem existir; (3) há um objetivo claro (i.e., minimizar o erro da estimativa do critério estético). Estas características fazem do algoritmo genético (GA), proposto por Holland e que simula a estratégia do processo de evolução Darwiniano [HOLLAND, 1962], uma boa escolha. Desde a abordagem original de Holland, uma grande quantidade de diferentes abordagens e aplicações tem tornado GA um algoritmo de busca muito popular. Particularmente, o GA adotado neste trabalho corresponde a uma das variantes mais simples e é descrito em [GOLDBERG, 1989] e [GEN & CHENG, 1997]. (veja seção 3.3.6).

A população é formada por cromossomos que representam individualmente um conjunto de medidas (vetor de atributos) ou combinação de medidas que deve ser testado e generalizado. Seja $X_i = \{b_1 b_2 \cdots b_{N_D}\}$ um vetor binário, cujos elementos (ou genes) podem assumir dois estados: $b_k = 0$ quando a medida deve ser desconsiderada e $b_k = 1$ quando a medida deve influir no resultado da generalização. Deste modo, N_D é o número máximo de medidas sendo consideradas simultaneamente.

A população pode ser inicializada aleatoriamente ou considerando algum conhecimento sobre a solução. Portanto, para evoluções partindo de uma inicialização aleatória verificou-se que os cromossomos selecionados eram aqueles cujo número de medidas consideradas eram próximos de 15. Com o intuito de deixar a evolução mais rápida, inicializou-se a população com cromossomos considerando 10 medidas tomadas aleatoriamente.

Conforme já observado, o objetivo é minimizar a distância entre as notas atribuídas por humanos e os critérios estéticos modelados. Esta distânica serve como um parâmetro de qualidade da generalização efetuada pela interpolação TPS conforme descrito anteriormente e formulada por (96). Ela servirá também como função objetivo $\varphi(X_i)$ para avaliar uma combinação particular de d medidas em teste de um cromossomo.

Depois de avaliados, os cromossomos são dispostos em ordem ascendente de $\varphi(X_i)$ e a primeira metade é selecionada para a recombinação. Esta recombinação é realizada por dois tipos de operadores evolucionários: o cruzamento e a mutação. Com o intuito de obter uma convergência mais rápida sem comprometer a qualidade da solução, foi escolhido o cruzamento multi-pontos [GOLDBERG, 1989]. Já a mutação é uma simples mudança no estado binário do gene. O processo continua até que se chegue a um cromossomo cuja valor da função objetivo $\varphi(X_i)$ seja adequado ou que a variância das funções objetivos de uma população seja próxima de zero.

6.5 Aplicando o modelo

O processo de modelamento foi realizado para todos os critérios estéticos em $\boldsymbol{C}$. Antes do treinamento, foi necessário testar a Base de Funções Radiais (FBR) a ser usada na interpolação. Embora a TPS apresente propriedades desejáveis assegurando uma superfície que inclua todos os pontos interpolados, em

aprendizados é também desejável uma superfície que apenas aproxime, já que o ajuste total representaria uma solução única ou muito limitada. Mees [MEES et al., 1992] adverte sobre más escolhas para a função de interpolação $\phi(\|r\|)$. Algumas dessas escolhas podem gerar mais erros que outras na solução do sistema linear em (101). Foi dedicada uma atenção especial em funções obtidas variando apenas o expoente da FBR e para isto foi implementado um pequeno ambiente para teste do comportamento da função variando esse expoente (veja Figura 115) incluindo um visualizador da superfície.

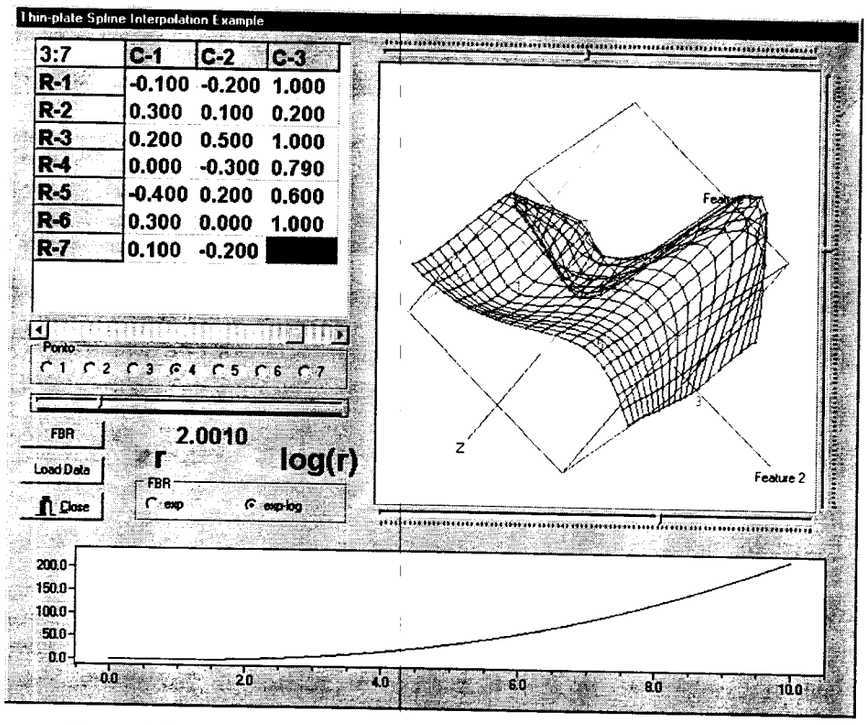


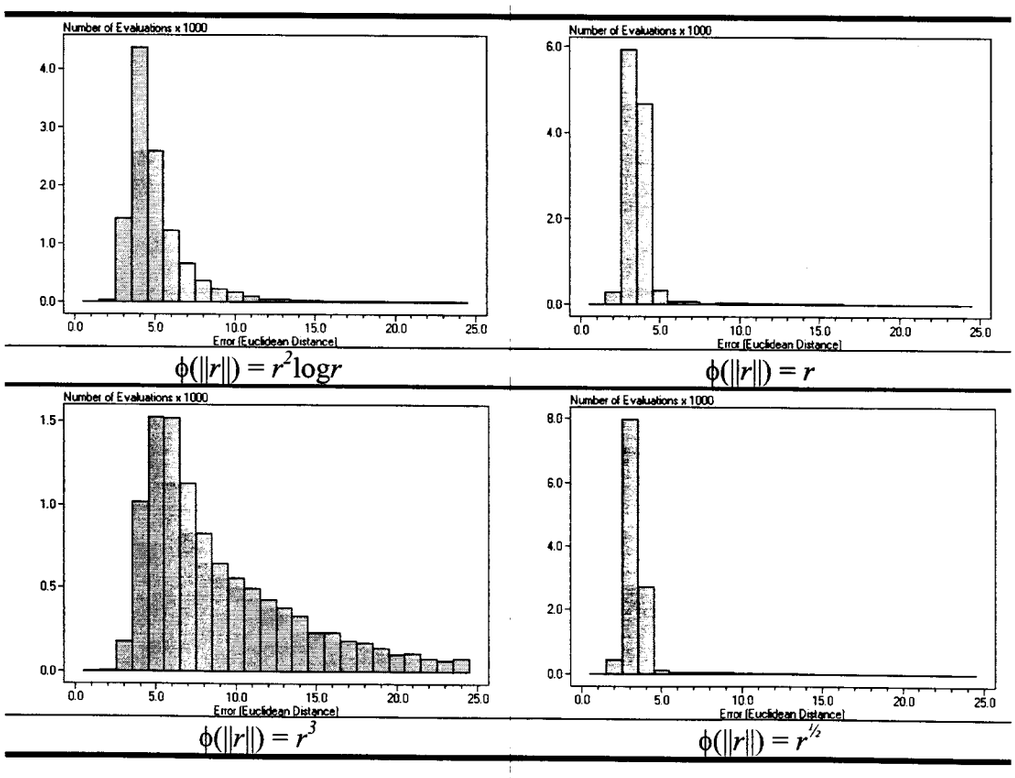
Figura 115 – Janela do aplicativo para testar a TPS.

A consistência do sistema linear que provém da interpolação (101), foi verificada durante as evoluções do GA desconsiderando as combinações que geravam casos inconsistentes. Estes acontecem quando existem pontos colineares entre os pontos a serem interpolados. Quando isso acontece, o indivíduo não prosseguirá em outras gerações, a não ser que sofra alguma recombinação.

Para avaliar o desempenho dos resultados obtidos com o algoritmo genético (GA), foi realizada uma seleção por "força bruta" das medidas, isto é, todas as combinações possíveis tomadas duas e três medidas foram consideradas nesta investigação. Somente o critério estético *agradabilidade* foi considerado para este

experimento. Os histogramas da distância euclideana entre a resposta interpolada e a resposta do sujeito humano (96), ou seja, o erro da estimativa, foram calculados para os seguintes casos de função de interpolação: (a) $\phi(\|r\|) = r$; (b) $\phi(\|r\|) = r^2\log(r)$; (c) $\phi(\|r\|) = r^3$ e (d) $\phi(\|r\|) = r^{1/2}$. Estes expoentes foram escolhidos em função de um intervalo no qual a superfície torna-se mais suave conforme aumenta o valor do expoente. Existe ainda um compromisso entre a exatidão da interpolação e a superfície de mínima distância. Ou seja, ou todos os pontos pertencem à superfície interpolada ou esta superfície é aquela de mínima distância quadrática dos pontos. Estes expoentes poderiam ser testados em um intervalo contínuo. Neste caso, os histogramas da Tabela 2 e da Tabela 3 corresponderiam a cortes em uma superfície de erro para suas respectivas quantidades de medidas.

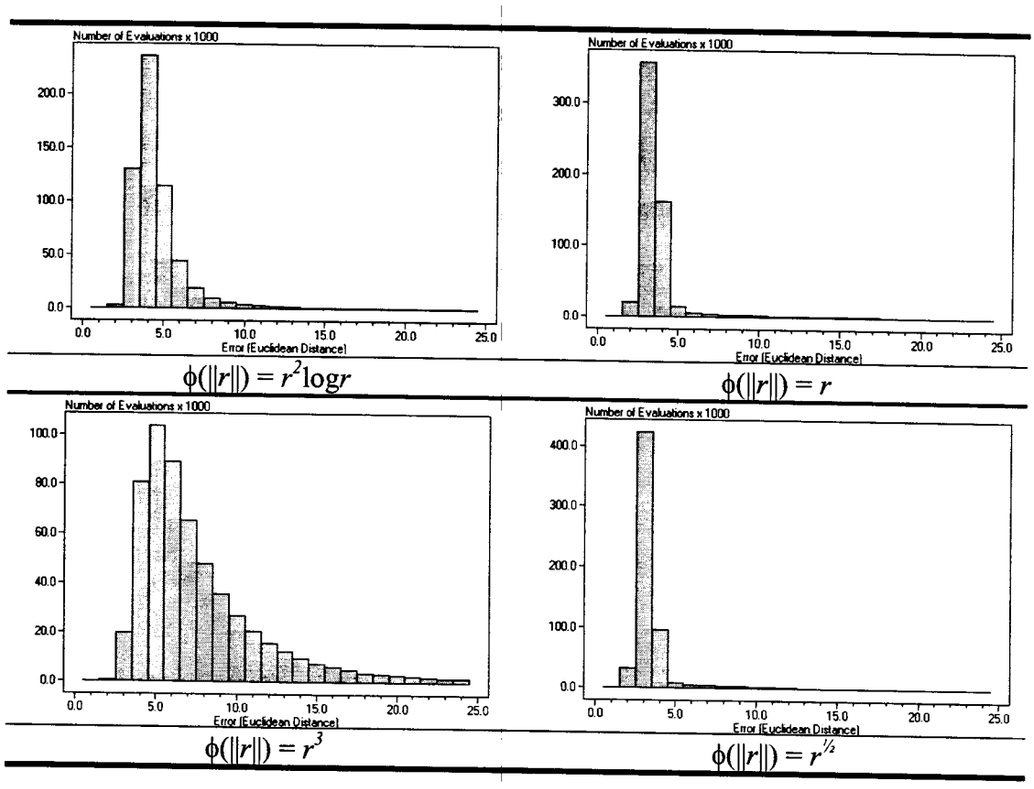
Tabela 2 - Histogramas de erros para originalidade, usando o método de força bruta para combinar duas medidas



As distribuições obtidas foram comparadas e os histogramas resultantes são mostrados na Tabela 2 e na Tabela 3. É fácil perceber que nos resultados obtidos para $\phi(\|r\|) = r$ e $\phi(\|r\|) = r^{1/2}$ há um intervalo de erro, isto é, $2.5 < \text{Erro} < 4.5$, permitindo erros menores para ambas as dimensões testadas do espaço de busca (2 e 3

dimensões). Contudo, para $\phi(\|r\|) = r^2\log(r)$ e $\phi(\|r\|) = r^3$, os valores de erros da estimativa maiores tornam-se mais freqüentes conforme advertido por Mees et al. em [MEES et al., 1992]. Para r^3, os erros mais altos são provavelmente causados pela suavidade imposta pela função.

Tabela 3 - Histogramas de erros para agradabilidade usando o método de força bruta para combinar três medidas.



Além dos histogramas de erros, a investigação sobre o método "força bruta" forneceu-nos uma abordagem sobre as medidas selecionadas e os erros mínimos para os casos de duas e três medidas mostradas na Tabela 4 e na

Tabela 5, respectivamente. Estas tabelas mostram que o contraste de cores para o intervalo do histograma $229.5 \leq H_{cores} < 242.25$ foi sempre escolhido como uma das medidas.

Conforme explicado nas seções anteriores, na abordagem adotada para modelamento é desejável reduzir o número de medidas usadas. Contudo, mesmo não considerando o número de medidas na função objetivo, poucas medidas (entre 14 e 18) foram selecionadas. Verificou-se também que o número de medidas selecionadas não varia muito ao longo de repetidas evoluções considerando condições iniciais

aleatórias. A evolução se deu em 225 gerações e foram realizadas sobre uma população aleatória, porém restringindo o número de medidas combinadas incialmente. As inicializações aleatórias implicaram uma evolução lenta, mas convergiram para números similares de medidas e valores da função objetivo (erros das estimativas). Vários números de combinações iniciais foram tentados, mas 10 combinações iniciais levaram a evoluções mais rápidas.

Tabela 4 - As medidas selecionadas com o método "força bruta" e os erros mínimos para as várias FBRs consideradas como funções interpolantes (considerando duas medidas).

FBR $\phi(\|r\|)$	Medidas Selecionadas		Erro Mínimo
$r^2 \log r$	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Espacial ($0.79 \leq s < 0.94$)	1.1562
r	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Espacial ($1.88 \leq s < 2.04$)	1.1467
r^3	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Direcional ($0.1 \leq s < 0.125$)	1.3405
$r^{1/2}$	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Espacial ($2.83 \leq s < 2.98$)	1.0750

Tabela 5 - As medidas selecionadas com o método "força bruta" e os erros mínimos para as várias FBRs consideradas como funções interpolantes (considerando duas medidas).

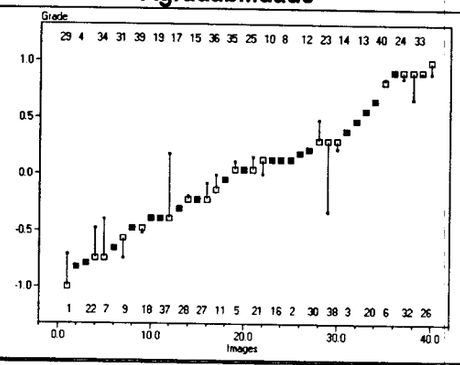
FBR $\phi(\|r\|)$	Medidas Selecionadas			Erro Mínimo
$r^2 \log r$	Histograma de Contraste RGB ($178.5 \leq d < 191.25$)	Histograma de Contraste RGB ($216.75 \leq d < 229.5$)	Histograma Espectral Direcional (em rad) ($2.36 \leq s < 2.51$)	0.9815
r	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Espacial (Variância)	Histograma Espectral Direcional (em rad) ($0.47 \leq s < 0.62$)	0.8895
r^3	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma de Contraste RGB (Variância)	Histograma Espectral Direcional (Variância)	1.0374
$r^{1/2}$	Histograma de Contraste RGB ($229.5 \leq d < 242.25$)	Histograma Espectral Direcional (em rad) ($0.47 \leq d < 0.63$)	Histograma Espectral Espacial (Variância)	0.9250

A Tabela 6 mostra as estimativas obtidas pela melhor solução com respeito a cada critério estético. A sequência de imagens no eixo horizontal é obtida pela ordenação das imagens em termos das notas médias (do experimento psicofísico). Estas notas são representadas por quadrados, os quais aqueles preenchidos representam as imagens usadas para treinamento e os quadrados vazados + pontos correspondem às imagens de teste com sua respectiva estimativa. O número da

seqüência no topo e em baixo de cada gráfico representam o código das imagens consideradas.

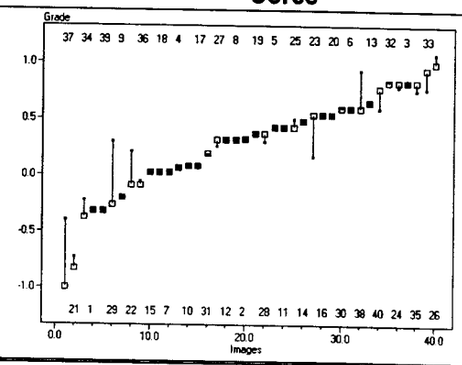
Tabela 6 - Erros para o treinamento de agradabilidade, cores, originalidade, qualidade artística e motivação.

Agradabilidade



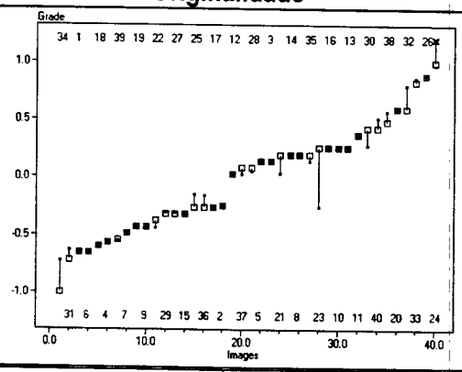
Erro	#Medidas
1.02	17

Cores



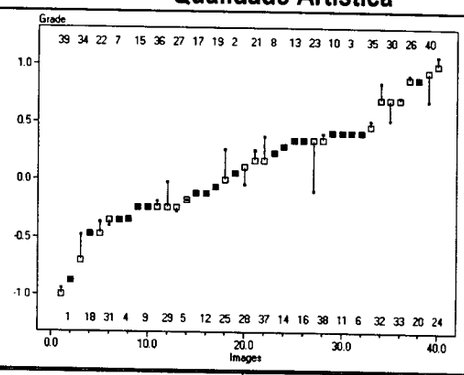
Erro	#Medidas
1.07	16

Originalidade



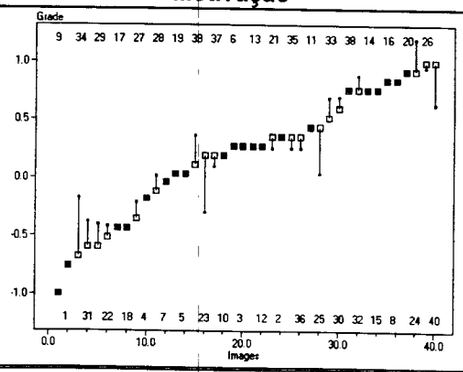
Erro	#Medidas
0.73	16

Qualidade Artística



Erro	#Medidas
0.73	16

Motivação



Erro	#Medidas
1.04	16

O algoritmo genético, usando a representação binária das medidas pelos 152 genes para cada cromossomo evoluiu com uma população de 200 indivíduos. Evoluções independentes foram realizadas para cada critério estético. A taxa de mutação foi 0.01 e a de cruzamento 0.7. O crossover multiponto quebra os cromossomos em oito pontos. Embora tenha sido usado como condição de parada o critério de convergência (variância da função-objetivo), para efeito de padronização dos resultados para os 5 critérios estéticos, adotou-se a evolução de 225 gerações.

Os resultados das evoluções têm mostrado um bom desempenho para alguns critérios estéticos e razoáveis para outros. Desempenhos similares foram obtidos pelo método de força bruta. Critérios estéticos como *originalidade* e *qualidade artística* têm sido mais receptivos ao modelo adotado, produzindo mais concordância com o experimento psicofísico. Este sucesso foi caracterizado em termos do número de medidas selecionadas e o menor erro da estimativa (função objetivo), ou seja, pequenos erros revelam uma boa concordância entre as notas originais e a estimativa para os critérios estéticos. Embora a Tabela 6 não identifique as medidas obtidas com a evolução do GA, as medidas selecionadas pelo método "força bruta" estão presentes na Tabela 4 e na

Tabela 5. O intervalo do histograma do contraste de cores (entre 229.5 e 242.25) e intervalos do descritor espectral (direcional e espacial) foram os mais freqüentemente escolhidos.

Tabela 7 – Demonstrativo de três execuções para os critérios em *C*. AGR: agradabilidade; COR: cores; ORG: originalidade; QLD: qualidade; MTV: motivação. Os números nos círculos indicam cada evolução do algoritmo genético.

		AGR	COR	ORG	QLD	MTV
fitness	①	*1,1*	*1,07*	*0,72*	*0,76*	*1,04*
	②	*1,02*	*1,34*	*0,73*	*0,73*	*1,08*
	③			*0,78*	*0,77*	
nº gerações	①	*226*	*227*	*200*	*211*	*233*
	②	*226*	*243*	*226*	*226*	*245*
	③			*226*	*226*	
nº atributos	①	*14*	*16*	*12*	*18*	*14*
	②	*17*	*14*	*16*	*14*	*16*
	③			*16*	*16*	

Foram realizadas evoluções para cada um dos critérios em *C*. Os critérios Originalidade (ORG) e Qualidade Artística (QLD) obtiveram os menores erros com

três evoluções. O número de atributos selecionados para cada um dos critérios oscilou pouco em um pequeno intervalo, entre 12 e 18 em um universo de 154. O menor deles foi 12, obtido com o critério Originalidade (ORG). Para essa mesma evolução com 200 gerações resultou o menor erro (0,72). Para obter um modelo robusto é interessante que poucos atributos sejam selecionados. É importante ressaltar que este erro é calculado em um espaço cuja dimensão é a do número de atributos selecionados.

A entropia não foi dos atributos mais influentes na composição dos critérios macro-estéticos. A motivação (MTV) e a originalidade (ORG) são os critérios menos influenciados por esse atributo. A complexidade caracterizada pela entropia parece não ser importante na definição destes critérios e de certa forma, corrobora os dados para 2 e 3 atributos da Tabela 4 e da

Tabela 5.

Tabela 8 – Resultado de duas e três evoluções do GA para o atributo entropia.

	Entropia				
	AGR	COR	ORG	QLD	MTV
μ					
σ^2				②	
0		②			
1	①	②			
2			③		
3		①			①
4	②				
5					
6	②				
7					
8					
9					
10					
11					
12	①				
13	①				
14					
15				①	
16				①	
32					
64					
128				①②	

A variância influenciou principalmente os critérios agradabilidade (AGR) e originalidade (ORG). Contudo essas influências foram concentradas em cada evolução. Para AGR, a execução ② foi bastante influenciada. Já para ORG, a execução ③ determinou um conjunto concentrado de escalas.

Tabela 9 - Resultado de duas e três evoluções do GA para o atributo variância.

	Variância				
	AGR	COR	ORG	QLD	MTV
μ	②		①		①
σ^2				②③	②
meia-vida					
0					
1			②		
2		①			①
3					
4					
5					
6					②
7					
8		②	②		
9					
10			③		
11			③		
12		②	③		
13	②		③		
14	②	②			
15	②				
16	②				
32	②		②	①	
64				③	
128					②

As salsichas de Minkowski são os atributos menos influentes, como se pode perceber na Tabela 10.

Tabela 10 - Resultado de duas e três evoluções do GA para as salsichas de Minkowski.

	Salsichas de Minkowski				
	AGR	COR	ORG	QLD	MTV
μ				①	
σ^2					
dim. frac.					
1					
2				②	
3				①	①
4					
5			②		
6					
7					
8					
9	②				
10		①		①	
11	①				
12					
13		②			
14					①
15					
16					②

O descritor espectral espacial revelou-se um dos atributos mais influentes no experimento. Além disso, na Tabela 11 percebe-se uma concentração de escolhas em determinados intervalos. Por exemplo, a origem, ou seja, as freqüências espaciais mais baixas influenciam fortemente. Regiões de freqüencias mais altas também influiram, contudo de maneira menos concentrada em um intervalo específico. Este resultado indica o papel do tamanho das estruturas texturais de uma imagem.

Tabela 11 - Resultado de duas e três evoluções do GA para os descritores espectrais espaciais.

	Descrição Espectral Espacial				
	AGR	COR	ORG	QLD	MTV
μ		①			
σ^2				①	
0,000	①②	①②	①②③	②③	①②
0,025					
0,050			②		
0,075				②	
0,100	②				②
0,125					
0,150					
0,175			③		
0,200					
0,225					
0,250	①				
0,275	①		①		①
0,300	②	②	②③	②	
0,325		①			①
0,350		②	②	①③	
0,375			①	③	
0,400				①②③	②
0,425			③	②	①
0,450					
0,475					

A Tabela 12 mostra os descritores espectrais direcionais. Percebe-se uma certa preferência para duas regiões de direção: entre 0 e 0.7854 radianos e entre 1.7279 e 2.6704 radianos.

Tabela 12 - Resultado de duas e três evoluções do GA para os descritores espectrais direcionais.

	Descrição Espectral Direcional				
	AGR	COR	ORG	QLD	MTV
μ					
σ^2					
0,000			①		
0,1571					①
0,3142	①	①	②		
0,4712		①		②	②
0,6283	②			①③	
0,7854		①			
0,9425					
1,0996					
1,2566					
1,4137					
1,5708					
1,7279			③		
1,8850				③	
2,042		①	①③		
2,1991					
2,3562		②		②	
2,5133		①	②		
2,6704	①				
2,8274					
2,9845					

Os atributos de contraste são também bastante influentes. No caso do contraste em níveis de cinza, são distribuídos pelos critérios, ou seja, são influentes, porém em diversos intervalos.

Tabela 13 - Resultado de duas e três evoluções do GA para contraste de níveis de cinza.

	Contraste em Níveis de Cinza				
	AGR	COR	ORG	QLD	MTV
μ					
σ^2		①			
00,00			①		
12,75				①	
25,50		①			
38,25			①		
51,00	①				
63,75	①	①②		①	
76,50					
89,25	②				②
102,00	②	①	③		
114,75	①	②			②
127,50					
140,25					
153,00					
165,75	②		①		
178,50	①	②	②③		
191,25	①	②		①	
204,00			②		①
216,75					
229,50					
242,25					
255,00					

O contraste de cores apresenta influência para critérios de originalidade (ORG), qualidade artística (QLD) e motivação (MTV). A agradabilidade (AGR) não foi incluenciada pelo contraste de cores, pelo menos nesse experimento. O mais interessante é que o equilíbrio de cores (COR) sofreu pouca influência desse atributo. O único intervalo que influencia é o de baixo contraste, ou seja, regiões de cores uniformes, mesmo assim em apenas uma execução.

Tabela 14 - Resultado de duas e três evoluções do GA para contraste de cores.

	Contraste em Cores				
	AGR	COR	ORG	QLD	MTV
μ					
σ^2					
00,00		①		③	
12,75					②
25,50					
38,25					①
51,00					
63,75					
76,50					②
89,25				③	②
102,00			③	②③	
114,75					
127,50				①②	
140,25			①	②③	①
153,00					①②
165,75				①	②
178,50			②③		
191,25					
204,00			①②	①②③	
216,75					
229,50					
242,25				①②③	
255,00					

6.6 Comentários e conclusões

Textura, contraste, informação visual e outras propriedades da imagem podem ser quantificadas por uma grande diversidade de medidas. Contudo, para inferir seu relacionamento com a percepção visual humana é necessária alguma atenção. O modelamento da percepção visual humana é um problema complexo que deve ser resolvido em várias frentes (neurociência, psicologia, visão computacional, teoria da informação entre outras). A abordagem apresentada representa uma tentativa de integrar várias destas possibilidades. Um dos problemas de abordagens desse tipo é que as medidas importantes não são conhecidas *a priori*. Aqui, foi considerado um conjunto de medidas que, dada sua natureza, poderiam desempenhar um papel importante no modelo da percepção. Na verdade, um dos resultados

permitido pelo modelo é inferir, dado um conjunto de medidas, quais são as mais relevantes.

Um dos resultados mais importantes desse trabalho tem sido a verificação de que modelos razoavelmente bons de percepção visual humana podem ser obtidos pela abordagem apresentada. Isto é particularmente surpreendente, já que tais critérios estéticos são intensivamente afetados por fatores subjetivos como preferências pessoais e tendências culturais. Por outro lado, as qualidades consideradas certamente confirmam, conforme sustentado pelos experimentos, alguns atributos visuais básicos e compartilhados pela maioria das pessoas como complexidade e equilíbrio. Este resultado é particularmente relevante para o objetivo inicial de usar a abordagem adotada como um modelo estético preditivo, seja em diagnósticos, seja em avaliações de *home-pages*.

7 CONCLUSÕES

Este trabalho apresentou técnicas de mineração de dados aplicadas ao problema de análise de imagens. Um conjunto de técnicas foi revisado com a proposta de uma taxonomia. Um dos combustíveis para sistemas baseados em conhecimento são os atributos. Assim, novas abordagens de extração de atributos foram propostas. Exemplos de aplicações para análise de imagens foram explorados ilustrando o potencial apresentado por abordagens baseadas em conhecimento para este tipo de enfoque.

Uma nova técnica baseada em campos eletrostáticos [GRIGORISHIN et al., 1996] para identificação de saliência em formas bidimensionais foi proposta. Embora computacionalmente inviável para aplicações que exijam pequeno tempo de execução, os resultados aplicados a formas neurais apontam para um método bastante robusto, principalmente se comparado ao uso da curvatura na detecção de saliências. A aplicação da abordagem para estudo de interações entre formas foi estudada, com potencial aplicação em modelos para crescimento de axônios. Esta aplicação pode ser também usada para análise de imagens em movimento, já que detecta a aproximação de corpos em cada ponto do contorno.

A esqueletonização multi-escalas sem deslocamento de bordas foi abordada a partir de novas métricas. Inicialmente introduzida a partir da métrica euclideana e com limitações quanto ao tempo de execução, as novas métricas permitem uma maior flexibilidade quanto às propriedades do método. Assim, pode-se abrir mão de propriedades como isometria, em função de um menor tempo de execução. Muitas propriedades esqueletais permaneceram valendo para muitas das métricas de Minkowski e *Chamfer*. Dentre estas propriedades estão: (1) a preservação da

topologia e da homotopia da forma original no esqueleto; (2) uma estrutura multi-escalas; (3) reconstrução sem deslocamento de borda.

As estruturas multi-escalas, em particular, permitem que se construa atributos e descritores (dendrogramas) que podem ser usados na mineração de dados aplicada a análise de formas. Para uma métrica específica: a *chessboard*, foi construído um algoritmo rápido para a construção da estrutura multi-escalas do esqueleto. Embora não seja invariante à rotação, o método pode ser aplicado sem problemas em formas ramificadas.

A análise de formas foi abordada nesse trabalho em duas frentes. A primeira utiliza-se de um sistema baseado no conhecimento de um operador especialista na identificação de extremidades e ramificações em neurônios ganglionares da retina de gatos. Esta abordagem utilizou técnicas de mineração de dados para calibrar parâmetros de limiarização, suavização e equalização da curvatura a partir de atributos globais da forma. Os resultados obtidos foram encorajadores no prosseguimento de aplicação da abordagem em imagens médicas e laboratoriais tais como: traçados cefalométricos e hemogramas. A generalização de uma forma usando parâmetros globais permite que as aplicações, uma vez treinadas, sejam extremamente rápidas e eficazes. Isso é possível selecionando-se parâmetros globais baseados no seu baixo custo computacional.

A segunda frente aborda o desenvolvimento de um detector de saliências a partir da estrutura do esqueleto multi-escalas da forma proposto por Costa e Estrozi [COSTA & ESTROZI, 1999]. O detector de saliências apresenta uma robustez controlada por apenas dois parâmetros. Além disso, o tempo de execução é reduzido, já que o processo de esqueletonização se dá apenas em uma vizinhança ao contorno. A viabilidade da aplicação das novas métricas estudadas na seção 4.2 é proposto como trabalho futuro e pode tornar o método rápido o suficiente para ser aplicado em sistemas dedicados. O estudo de tempos de execução mostrou essa viabilidade indicando que a maior parte do processo é tomada pelo cálculo da transformada distância.

As duas frentes apresentadas acima podem ser utilizadas em conjunto. Por exemplo, os parâmetros do detector de saliências podem ser calibrados para aplicações específicas, com o intuito de aumentar a acurácia dos pontos detectados.

A análise pictórica também foi estudada no âmbito de sistemas baseados em conhecimento. Uma aplicação de avaliação estética de *home-pages* foi desenvolvida. Esta aplicação permitiu selecionar atributos que influenciam critérios estéticos subjetivos. Não houve atuação explícita sobre o processo de seleção de atributos, por exemplo, com penalidades sobre o número de atributos selecionados. Deste modo, o sistema construído descobriu um compromisso entre os erros da aplicação de uma combinação de atributos com a quantidade de atributos selecionada. O sistema apontou preferências a determinadas medidas da imagem relacionadas ao contraste e descritores texturais. Conclui-se, portanto que é viável a construção de um sistema que implemente um fluxo contínuo entre o experimento, a avaliação e a seleção de atributos resultando em uma avaliação artificial da qualidade de imagens.

O espectro de aplicações desse tipo de abordagem é amplo. O meio de comunicação visual tem se tornado extremamente influente. A avaliação de critérios subjetivos pode ser estendida a outras mídias visuais como jornais e publicidade impressa. Estudos sobre a avaliação de qualidade de imagens degradadas por processos de compressão podem ser avaliados utilizando sujeitos representativos em uma população. Este processo pode ser útil no desenvolvimento de dispositivos de comunicação com alto tráfego de imagens ou até mesmo na avaliação e atuação no tráfego de dados provenientes de processos multimída em redes de alta velocidade. Sistemas de diagnóstico por imagens podem utilizar dessa abordagem para armazenar conhecimentos de um especialista (um patologista, por exemplo) e auxiliá-lo no diagnóstico a partir de imagens de lâminas histológicas. Sistemas de monitoramento que resultem em tomadas de decisão também podem ser auxiliados por sistemas de análise pictórica baseados em conhecimento. Um exemplo: sistemas de controle de tráfego com monitoramento por câmeras.

Novos desenvolvimentos devem ser adicionados à aplicação de avaliação estética de *home-pages*. Estes desenvolvimentos viabilizam e aperfeiçoam a abordagem apresentada. Abaixo seguem listados alguns desenvolvimentos.

- Análise dos atributos para modelos de atenção visual em *home-pages*.
- O método de Monte Carlo pode ser utilizado para generalizar a análise de atributos por "força bruta", construindo assim histogramas de erros para mais combinações de atributos.

- O experimento psicofísico deve ser aperfeiçoado para que os indivíduos sintam-se mais estimulados a participar do experimento e garantindo assim uma amostra mais representativa do conhecimento. Este aperfeiçoamento pode ainda incluir o desenvolvimento de um ambiente de experimentos que possa ser facilmente difundido.
- Desenvolver uma home-page que permita a adição sugerida de home-pages ao conjunto de testes e treinamento. Este seria um passo na direção de da construção de uma avaliação estética *on-line.*
- Investigar outras mídias ligadas à comunicação visual ou de manifestações artísticas como pinturas.
- Permitir que os modelos variem o conjunto de indivíduos. Para isso, deve-se alterar o experimento para que os indivíduos adicionem suas informações econômico-sociais ou que estas sejam capturadas por sistemas automáticos.
- Relaxar o modelo de generalização. Esta relaxação permitiria definir o grau de liberdade com o qual o modelo será construído. Isto pode ser realizado variando o expoente da função de base radial (*thin-plate spline*).
- Testar outras abordagens de generalização. Um exemplo seria a abordagem de mapas auto-organizáveis (SOM) que serviriam ao mesmo tempo para generalizar e "aprender" os atributos mais influentes.

8 BIBLIOGRAFIA

AHUJA,N.; TUCERYAN,M. Extraction of Early Perceptual Structure in Dot Patterns: Integrating Region, Boundary, and Component Gestalt. ***Computer Vision, Graphics, and Image Processing***, (48), 304-356, 1989.

ALDENDERFER,M.S.; BLASHFIELD,R.K. ***Cluster Analysis***. SAGE Publications, 87pp., 1984.

ALLEN,M.; BROWN,G.J.; MILES,N.J. Measurement of boundary fractal dimensions: review of current techniques. ***Powder Technology***, (84), 1-14, 1995.

ANDERBERG,M.R. ***Cluster Analysis for Appliations***. Academic Press, 359pp., 1973.

APTÉ,C.; WEISS,S. Data mining with decision trees and decision rules. ***Future Generation Computer Systems***, (13), 197-210, 1997.

ARCELLI,C.; DIBAJA,G.S. Finding Local Maxima in a Pseudo-Euclidean Distance Transform. ***Computer Vision, Graphics, and Image Processing***, (43), 361-367, 1988.

ARREBOLA,F.; BANDERA,A.; CAMACHO,P.; SANDOVAL,F. Corner detection by local histogram of contour chain code. ***Electronics Letters***, (33)21, 1769-1771, 1997.

ARREBOLA,F.; CAMACHO,P.; BANDERA,A.; SANDOVAL,F. Corner detection and curve representation by circular histograms of contour chain code. ***Electronics Letters***, (35)13, 1065-1067, 1999.

ASADA,H.; BRADY,M. The Curvature Primal Sketch. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (PAMI-8)1, 2-14, 1986.

ATTNEAVE,F. Some Informational Aspects of Visual Perception. ***Psychological Review***, (61)3, 183-193, 1954.

BÄCK,T. ***Evolutionary Algorithms in Theory and Practice***. Oxford University Press, 314pp., 1996.

BAK,P. ***How nature works. The science of self-organized criticality.*** Springer-Verlag , 212pp., 1996.

BAKER,J. Adaptive selection methods for genetic algorithms. *In*: ***Proceedings of the Second International Conference on Genetic Algorithms***, Hillsdale, NJ, USA, Lawrence Erlbaum Associates, 100-111, 1987.

BALLARD,D. ***An Introduction to Natural Computation***. MIT Press, 307pp., 1997.

BARROW,J.D. ***The Artful Universe***. Oxford University Press, 274pp., 1995.

BENSE,M. ***Pequena Estética***. Editora Perspectiva, 233pp., 1975.

BHATTACHARYA,B.K.; POULSEN,S.; TOUSSAINT,G.T. Application of Proximity Graphs to Editing Nearest Neighbor Decision Rule. *In*: ***Proceedings of International Symposium on Information Theory***, Santa Monica, CA, USA, 1981.

BIRKHOFF,G.D. ***Aesthetic measure***. Harvard University Press, 1933.

BISHOP,C.M. ***Neural Networks for Pattern Recognition***. Oxford University Press, 482pp., 1995.

BOOKSTEIN,F.L. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (11)6, 567-585, 1989.

BORGEFORS,G. Distance Transformations in Digital Images. ***Computer Vision, Graphics, and Image Processing***, (34), 344 -371, 1986.

BORGEFORS,G. Distance Transforms in Arbitrary Dimensions. ***Computer Vision, Graphics, and Image Processing***, (27), 321-345, 1984.

BRADY,M.; ASADA,H. Smoothed Local Symmetries and Their Implementation, *AI Memo 757*. Artificial Intelligence Laboratory, MIT-EUA, 1-43, 1984.

BREIMAN,L.; FRIEDMAN,R.; OLSHEN,R.; STONE,C. ***Classifications and Regression Trees***. Wadsworth, 1984.

BRIGHAM,E.O. ***The Fast Fourier Transform and its Applications***. Prentice Hall, 448pp., 1988.

BRINK,A.D.; PENDOCK,N.E. Minimum Cross-entropy Threshold Selection. ***Pattern Recognition***, (29)1, 179-188, 1996.

BRUCE,V.; GREEN,P.R. ***Visual Perception. Physiology, Psychology and Ecology***. Lawrence Erlbaum, 431pp., 1990.

BRUCE,V.; GREEN,P.R. ***Visual Perception. Physiology, Psychology and Ecology***. Lawrence Erlbaum, 431pp., 1990.

BRUNO,O.M.; CESAR JR,R.M.; CONSULARO,L.A.; COSTA,L.F. Automatic Feature Selection for Biological Shape Classification in Synergos. *In*: ***Proceedings of SIBGRAPI'98***, Rio de Janeiro - Brazil, IEEE Computer Society, 363- 370, 1998.

CABENA,P.; HADJINIAN,P.; STADLER,R.; VERHEES,J.; ZANASI,A. ***Discovering Data Mining. From Concept to Implementation.*** Prentice-Hall, 195pp., 1997.

CAMPOS,A.G.; CONSULARO,L.A.; ESTROZI,L.F.; RIOS,L.G.; BOSCO,A.; COSTA,L.F. Neural image segmentation: A comparative approach. ***Image and Vision Computing***, 2000. (submetido)

CASERTA,F.; ELDRED,W.D.; FERNANDEZ,E.; HAUSMAN,R.E.; STANFORD,L.R.; BULDEREV,S.V.; SCHWARZER,S.; STANLEY,H.E. Determination of fractal dimension of physiologically characterized neurons in two and three dimensions. ***Journal of Neuroscience Methods***, (56), 133 - 144, 1995.

CASTLEMAN,K. ***Digital Image Processing***. Prentice Hall, 667pp., 1996.

CESAR JR,R.M.; COSTA,L.F. Application and assessment of multiscale bending energy for morphometric characterization of neural cells. ***Review of Scientific Instrumentation***, (68)5, 2177-2186, 1997b.

CESAR JR,R.M.; COSTA,L.F. Piecewise Linear Segmentation of Digital Contours in O(NlogN) Through a Technique based on Effective Digital Curvature Estimation. ***Real-time Imaging***, (1), 409-417, 1995.

CESAR JR,R.M.; COSTA,L.F. Semi-automated Dendrogram Generation for Neural Shape Analysis. *In*: ***Proceedings of SIBGRAPI'97***, Campos do Jordão, SP - Brazil, IEEE Computer Society, 147-154, 1997a.

CHEN,P.P.-S. The Entity-Relationship Model - Toward a Unified View of Data. ***ACM Transactions on Database Systems***, (1)1, 9-36, 1976.

CHETVERIKOV,D.; SZABÓ,Z. ***Detection of High Curvature Points in Planar Curves***. 1999. [http://visual.ipan.sztaki.hu/corner/cornerweb.html]

CODD,E.F. A Relational Model for Large Shared Data Banks. ***Communications of the ACM***, (13)6, 377-387, 1970.

CODD,E.F.; CODD,S.B.; SALLEY,C.T. Providing OLAP to User-Analysts. An IT Mandate, ***E.F. Codd and Associates***, 1-24, 1993.

CONSULARO,L.A.; LOTUFO,R.A.; COSTA,L.F. Data mining based modeling of human vision perception. *In*: CIOS,K.J. (Eds.) **Medical Data Mining and Knowledge Discovery**, Physica-Verlag (Springer), 2000.

CONSULARO,L.A.; CESAR JR,R.M.; COSTA,L.F. Synergos and its Applications to Contour Segmentation. *In*: ***Proceedings of I International Seminar on Bioelectronic Interfaces and III Workshop on Cybernetic Vision***, Campinas-SP, Brazil, IFSC-USP, 77- 83, 1999.

COSTA,L.F. Multidimensional scale-space shape analysis. *In*: ***International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging***, Santorini, Greece 214-217, 1999a.

COSTA,L.F. Particle Systems Analysis by using Skeletonization and Exact Dilations. ***Particle and Particle Systems Characterization***, (16)6, 273-277, 1999.

COSTA,L.F.; CESAR JR,R.M.; COELHO,R.C.; TANAKA,J.S. Analysis and Synthesis of Morphologically Realistic Neural Networks. *In*:

POZNANSKI,R.R. (Eds.) **Modeling in the Neurosciences: From Ionic Channels to Neural Networks**, Harwood Academic Publishers, 1998.

COSTA,L.F.; CONSULARO,L.A. The Dynamics of Biological Evolution and the Importance of Spatial Relations and Shapes. *In*: CANTONI,V.; DIGESÙ,V.; SETTI,A.; TEGOLO,D. (Eds.) ***Human and Machine Perception 2*** - , Kluwer Academic/Plenum , 1-14, 1999.

COSTA,L.F.; ESTROZI,L.F. Multiresolution Shape Representation without Border Shifting. ***Electronics Letters***, (35)21, 1829-1830, 1999.

COSTA,L.F.; VELTE,T.J. Automatic characterization and classification of ganglion cells from the salamander retina. ***Journal of Computer Neurology***, (404), 33-51, 1999.

CUISENAIRE,O. ***Distance Transformations: Fast Algorithms and Applications to Medical Image Processing***. Thesis (Docteur en Science Apliquées), Université Catolique de Louvain, Belgique, 233pp, 1999.

CUISENAIRE,O.; MACQ,B. Fast Euclidean Distance Transformation by Propagation using Multiple Neighborhoods. ***Computer Vision and Image Understanding***, (76)2, 163-172, 1999.

DANIELSSON,P.E. Euclidean distance mapping. ***Computer Vision, Graphics, and Image Processing***, (14), 227-248, 1980.

DE JONG,K. ***An Analysis of the Behavior of a Class of Genetic Adaptive Systems***. 1975.

DE LA MAZZA,M.; TIDOR,B. Boltzmann Weighted Selection Improves Performance of Genetic Algorithms, ***AI Memo 1345***. MIT Artificial Intelligence Laboratory, 1991.

DELAGNES,P.; BENOIS,J.; BARBA,D. Active contours approach to object tracking in image sequences with complex background. ***Pattern Recognition Letters***, (16), 171-178, 1995.

DHAR,V.; STEIN,R. ***Seven Methods for Transforming Corporate Data into Business Intelligence***. Prentice-Hall, 269pp., 1997.

DONDIS,D.A. ***A Primer of Visual Literacy***. MIT Press, 194pp., 1997.

DUDA,R.O.; HART,P.E.; STORK,D.G. ***Pattern Classification and Scene Analysis. Part I: Pattern Classification***. John Wiley and Sons, 482pp., 1998.

EDMONDS,B. Complexity and Scientific Modelling. *In*: ***Proceedings of 20th International Wittgenstein Symposium***, Krchberg am Wechsel, Austria, 1997.

ELMASRI,R.; NAVATHE,S.B. ***Fundamentals of Database Systems***. Benjamin/Cummings, 1994.

FAYYAD,U.; PIATETSKY-SHAPIRO,G.; SMYTH,P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. *In*: ***Proceedings of Second International Conference on Knowledge Discovery and Data Mining***, Portland, OR, USA, AAAI Press, 82-88, 1996.

FINK,B. ***Skin surface and female facial beauty: The influence of skin texture on female attractiveness***. 1998. [http://evolution.humb.univie.ac.at/institutes/urbanethology/student/html/bernhard/finkpro_tex.html]

FISCHLER,M.A.; BOLLES,R.C. Perceptual Organization and Curve Partitioning. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (PAMI-8)1, 100-105, 1986.

FREEMAN,H. Shape Description Via the Use of Critical Points. ***Pattern Recognition***, (10), 159-166, 1978.

FREITAS,A.; LAVINGTON,S.H. ***Mining Very Large Databases with Parallel Processing***. Kluwer Academic Publishers, 208pp., 1998.

GEN,M.; CHENG,R. ***Genetic Algorithms and Engineering Design***. John Wiley and Sons, 411pp., 1997.

GEN,M.; LIU,B.; IDA,K. Evolution program fro deterministic and stochastic optimizations. *European Journal of Operational Research*, 1996.

GIBSON,J.J. ***The perception of the visual world***. Houghton Miffin, 1950.

GLOVER,F.; LAGUNA,M. ***Tabu Search***. Kluwer, 1997.

GOLDBERG,D. ***Genetic Algorithms in Search, Optimization, and Machine Learning***. Addison-Wesley Publishing Co., 412pp., 1989.

GOLFARELLI,M.; MAIO,D.; RIZZI,S. Conceptual Design of Data Warehouses from E/R Schemes. *In*: ***Proceedings of the Hawaii International Conference on System Sciences***, Kona, Hawaii - USA, IEEE Press, 1998.

GOODHILL,G.J. Mathematical guidance for axons. ***Trends in Neurosciences***, (21)6, 226-231, 1998.

GORDON,I.E. ***Theories of Visual Perception***. John Wiley and Sons, 273pp., 1989.

GRIGORISHIN,T.; ABDEL-HAMID,G.; YANG,Y.H. ***Skeletonization: An Electrostatic Field-based Approach***, Technical Report, Department of Computer Science, University of Saskatchewan, 1-23, 1996.

HAKEN,H. ***Synergetics. An Introduction.*** Springer-Verlag, 325 pp., 1977.

HARDCASTLE,V.G. What We Don't Know About Brains. ***Studies on Historical, Biological and Biomedical Sciences***, (30)1, 69-89, 1999.

HARTIGAN,J.A. ***Clustering Algorithms***. John Wiley and Sons, 1975.

HOFFMAN,D.D.; SINGH,M. Salience of Visual Parts. ***Cognition*** , (63), 29-78, 1997.

HOLLAND,J. ***Adaption in natural and artificial systems***. The University of Michigan Press, 1975.

HOLLAND,J. Outline for a logical theory of adaptive systems. ***Journal of ACM***, (3), 297-314 , 1962.

HUANG,C.T.; MITCHEL,O.R. A Euclidean distance transform using greyscale morphology decomposition. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (16)4, 443-448, 1994.

HUBEL,D.; WIESEL,T. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. ***Journal of Physiology***, (160), 154-160, 1962.

HUBERT,L.J. Some applications of graph theory to clustering. ***Psychometrika***, (39), 283-309, 1974.

INFORMATION DISCOVERY,I. ***A Characterization of Data Mining Technologies and Processes***. Information Discovery, 1997.

JÄHNE,B. ***Digital Image Processing***. Springer-Verlag, 555pp., 1997.

JAIN,A.K.; DUBES,R.C. ***Algorithms for Clustering Data***. Prentice-Hall, 320pp., 1988.

JAIN,A.K.; DUIN,R.P.W.; MAO,J. Statistical Pattern Recognition: A Review. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (22)1, 4-37, 2000.

JAIN,A.K.; MURTY,M.N.; FLYNN,P.J. Data Clustering: A Review . ***ACM Computing Surveys***, (31)3, 264-323, 1999.

JAIN,R.; KASTURI,R.; SCHUNCK,B.G. ***Machine Vision***. McGraw-Hill, 549pp., 1995.

JOHNSON,S.C. Hierarchical clustering schemes. ***Psychometrika***, (32), 241-254, 1967.

KAELBLING,L.P.; LITTMAN,M.L.; MOORE,A.W. Reinforcement Learning: A Survey. ***Journal of Artificial Intelligence Research***, (4), 237-285, 2000.

KAO,J.-H. Process Planning for Additive/Subtractive Solid Freeform Fabrication usin Medial Axis Transform. 1-173, 1999.

KAUFMAN,L.; ROUSSEEUW,P.J. Finding Groups in Data - An Introduction to Cluster Analysis. *In*: ***Probability and Mathematical Statistics*** - , John Wiley and Sons, 1990.

KELLER,J.M.; CHEN,S.; CROWNOVER,R.M. Texture Description and Segmentation through Fractal Geometry. ***Computer Vision, Graphics, and Image Processing***, (45), 150-166, 1989.

KIRKPATRICK,S.; GELATT,C.D.; VECCHI,M.P. Optimization by simulated annealing. ***Science***, (220), 671-680, 1983.

KOHLMANN,K. Corner detection in natural images based on the 2D Hilbert transform. ***Signal Processing***, (48), 225-234, 1996.

KOHONEN,T. ***Self-Organization and Associative Memory***. Springer-Verlag, 312pp., 1989.

KOHONEN,T. The Self-organizing Map. ***Proceedings of the IEEE***, (78)9, 1464-1480, 1990.

KOLOUNTZAKIS,M.N.; KUTULAKOS,K.N. Fast Computation of the Euclidean Distance Map for Binary Images. ***Information Processing Letters***, (43), 181-184, 2000.

KOYAMA,T.; SHIONO,M.; SANADA,H.; TEZUKA,Y. Corner detection on thinned pattern, *Technical Report PRL80-107*. Professional Group on Pattern Recognition and Learning, IECE Japan, 83-90, 1980.

KOZA,J.R. ***Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)***. MIT Press, 819pp., 1992.

KRAFT,R.; KAUER,J. ***Estimating the Fractal Dimension from Digitized Images***, 1995. [http://www.weihenstephan.de/ane/algorithms/algorithms.html]

LANCE,G.N.; WILLIAMS,W.T. A general theory of classification sorting strategies: II. Clustering Systems. ***Computer Journal***, (10), 271-277, 1967.

LEE,J.-S.; SUN,Y.-N.; CHEN,C.-H.; TSAI,C.-T. Wavelet Based Corner Detection. ***Pattern Recognition***, (26)6, 853-865, 1993.

LEYMARIE,F.; LEVINE,M.D. Simulating the Grassfire Transform using an Active Contour Model. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (14)1, 56-75, 1992.

LEYTON,M. A Process-Grammar for Shape. ***Artificial Intelligence***, (34), 213-247, 1988.

LEYTON,M. Symmetry-curvature duality. ***Computer Vision, Graphics, and Image Processing***, (38), 327-341, 1987.

LI,B. Repeatdly Smoothing, Discrete Scale-space Evolution and Dominant Point Detection. ***Pattern Recognition***, (29)6, 1049-1059, 1996.

LIN,R.-S.; CHU,C.-H.; HSUEH,Y.-C. A modified morphological corner detector. ***Pattern Recognition Letters***, (19), 279-286, 1998.

LINDEBERG,T. ***Scale-Space Theory in Computer Vision***. Kluwer, 423pp., 1994.

LIU,H.-C.; SRINATH,M.D. Corner Detection from Chain-code. ***Pattern Recognition***, (23)12, 51-68, 1990.

LIU,H.-C.; SRINATH,M.D. Partial Shape Classification using Contour Matching in Distance Transformation. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (12)11, 1072-1079, 1990.

LONCARIC,S. A Survey of Shape Analysis Techniques. ***Pattern Recognition***, (31)8, 983-1001, 1998.

LUO,B.; CROSS,A.D.J.; HANCOCK,E.R. Corner detection via topographic analysis of vector-potential. ***Pattern Recognition Letters***, (20), 635-650, 1999.

MANDELBROT,B. ***Fractals, Form, Chance, and Dimension***. W. H. Freeman, 365pp., 1977.

MARR,D. ***Vision. A Computational Investigation into the Human Representation and Processing of Visual Information***. W.H. Freeman and Co., 397 pp., 1982.

MEDIONI,G.; YASUMOTO,Y. Corner detection and curve representation using cubic B-splines . ***Computer Vision, Graphics, and Image Processing***, (39), 267-278, 1987.

MEES,A.I.; JACKSON,M.F.; CHUA,L.O. Device Modeling by Radial Basis Functions. ***IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications***, (39)1, 19-27, 1992.

MEHROTRA,R.; NICHANI,S. Corner Detection. ***Pattern Recognition***, (23)11, 1223-1233, 1990.

MICHALEWICZ,Z. ***Genetic Algorithms + Data Structures = Evolution Programs***. Springer-Verlag, 340pp., 1994.

MITCHELL,M. ***An Introduction to Genetic Algorithms***. MIT Press, 208pp., 1996.

MOKHTARIAN,F.; MACKWORTH,A. Scale-based Description and Recognition of Planar Curves and Two-Dimensional Shapes. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (PAMI-8)1, 34-43, 1986.

MOKHTARIAN,F.; SUOMELA,R. Robust Image Corner Detection Through Curvature Scale Space . ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (20)12, 1376-1381, 1998.

MONTANARI,U. A Method for Obtaining Skeletons Using a Quasi-Euclidean Distance. ***Journal of ACM***, (15)4, 600-624, 1968.

MOREIRA,J.; COSTA,L.F. Neural-based color image segmentation and classification using self-organizing maps. *In*: ***Proceedings of IX SIBGRAPI*** 47-54, 1996.

MORIARTY,S.; SAYRE,S. Visual Semiotics and the Production of Meaning in Advertising. ***Visual Communication***, (6)1992.

MULLEN,M. ***A First Course in Probability and Statistics***. Addison-Wesley, 361pp., 1973.

NAKHAEIZADEH,G.; REINARTZ,T.; WIRTH,R. Wissensentdeckung in Datenbanken und Data Mining: Ein Überblick. *In*: NAKHAEIZADEH,G. (Eds.) Data Mining: Theoretische Aspekte und Anwendungen - , Physica-Verlag, 1-33, 1997.

NG,R.T.; HAN,J. Efficient and Effective Clustering Methods for Spatial Data Mining. *In*: ***Proceedings of the 20th VLDB Conference***, Santiago, Chile, ACM, 1994.

OBERMAYER,K.; BLASDEL,G.G.; SCHULTEN,K. Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. ***Physical Review A***, (45)10, 7568-7589, 1992.

OGAWA,H. Corner Detection on Digital Curves Based on Local Symmetry of the Shape. ***Pattern Recognition***, (22)4, 351-357, 1989.

OGNIEWICZ,R.L. Skeleton-space: A Multi-scale Shape Description Combining Region and Boundary Information. *In*: ***Proceedings of CVPR'94***, Seattle - USA 746-751, 1994.

ORLAND,B.; WEIDEMANN,E.; LARSEN,L.; RADJA,P. ***Exploring the Relationship between Visual Complexity and Perceived Beauty***. 1998. [http://www.imlab.uiuc.edu/complex/index.html]

PAGLIERONI,D.W. Distance Transforms: Properties and Machine Vision Applications. ***Graphical Models and Image Processing***, (54)1, 56-74, 1992.

PALMER,S.E. Modern Theories of Gestalt Perception. *In*: HUMPHREYS,G.W. (Eds.) ***Understanding Vision - Readings in Mind & Language***, Blackwell, 39-70, 1992.

PEDERSEN,T.B.; JENSEN,C.S. ***Multidimensional Data Modeling for Complex Data***, TimeCenter, 1-27, 1998.

PENTLAND,A. Fractal-Based Description of Natural Scenes. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (PAMI-6)6, 661-674, 1984.

POWELL,M.J.D. ***Approximation theory and methods***. Cambridge University Press, 339pp., 1981.

POWELL,M.J.D. Radial basis functions for multivariate interpolation: A review, *Technical Report 1985/NA12*. Department of Applied Mathematics and Theoretical Physics, Cambridge University, 1985.

POWELL,M.J.D. The uniform convergence of thin plate spline interpolation in two dimensions. ***Numerische Mathematik***, (68), 107-128, 1994.

POZNANSKI,R.R. Modeling the Electronic Structure of Starbust Amacrine Cells in the Rabbit Retina: Functional Interpretation of Dendritic Morphology. ***Bulletin of Mathematical Biology***, (54), 905-928, 1992.

PROCTER,P. ***CAMBRIDGE International Dictionary of English***. *In*: Cambridge University Press, 1-1773, 1995.

PUDNEY,C. Distance-ordered Homotopic Thinning: A Skeletonization Algorithm for 3D Digital Images. ***Computer Vision and Image Understanding***, (72)3, 404-413, 1998.

QUINLAN,J.R. ***C4.5: Programs for Machine Learning***. Morgan Kaufmann, 302pp., 1993.

RAGNELMAM,I. Neighborhoods for distance transformations using ordered propagation. ***Computer Vision, Graphics, and Image Processing***, (56)3, 399-409, 1992.

REDDY,M. A Measure for Perceived Detail in Computer-Generated Images, *Technical Report ECS-CSG-19-96*. Department of Computer Science, University of Edinburgh, 1-23, 1996.

REDDY,M. ***Perceptually Modulated Level of Detail for Virtual Environments***. PhD Thesis, University of Edinburgh, Scotland. 224pp, 1997.

REINARTZ,T. Focusing Solutions for Data Mining. Analytical Studies and Experimental Results in Real-World Domains. *In*: CARBONELL,J.G.; SIEKMANN,J. (Eds.) ***Lecture Notes in Artificial Intelligence*** - , Springer-Verlag, 1-307, 1999.

RIGLIS,E. Modeling Complexity in Visual Images, *Technical Report of 1st year PhD*. ISEL, Heriot-Watt University, 1998.

ROSENFELD,A.; JOHNSTON,E. Angle detection on digital curves. ***IEEE Transactions on Computers***, (22), 875-878, 1973.

RUTKOWSKI,W.S.; ROSENFELD,A. A comparison of corner detection techniques for chain-coded curves, *Technical Report TR-623*. Computer Science Center, University of Maryland, 1978.

SAMMON JR,J.W. A Nonlinear Mapping for Data Structure Analysis. ***IEEE Transactions on Computers***, (C-18)5, 401-409, 1969.

SANDWELL,D.T. Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data. ***Geophysical Research Letters***, (2), 139-142, 1987.

SCHALKOFF,R.J. ***Digital Image Processing and Computer Vision***. John Wiley and Sons, 489pp., 1989.

SCHALKOFF,R.J. ***Digital Image Processing and Computer Vision***. John Wiley and Sons, 489pp., 1989.

SELIM,S.Z.; ISMAIL,M.A. K-means type algorithms: a generalized convergence theorem and characterization of local optimality. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, (PAMI-6), 81-87, 1984.

SHANNON,C.E. A Mathematical Theory of Communication. ***Bell System Technical Journal***, (27)3, 379-423, 1948.

SILBERSCHATZ,A.; KORTH,H.F.; SUDARSHAN,S. ***Sistema de Banco de Dados***. Makron Books, 778pp., 1999.

SOKAL,R.; MICHENER,C.D. A statistical method for evaluating systematic relationships, *Scientific Bulletin 38*. University of Kansas, 1409-1438, 1958.

SONKA,M.; HLAVAC,V.; BOYLE,R. ***Image Processing, Analysis and Machine Vision***. Chapman & Hall, 555pp., 1993.

SRIKANT,R.; AGRAWAL,R. Mining generalized association rules. ***Future Generation Computer Systems***, (13), 161-180, 1997.

STEVENS,C.F. ***The Six Core Theories of Modern Physics***. MIT Press, 221pp., 1998.

SYNEK,E. ***Evolutionary Aesthetics: Visual Complexity and the Development of Human Landscape.*** 2000. [http://evolution.humb.univie.ac.at/institutes/urbanethology/student/html/erich/synekpro.html]

TAKALA,J.H.; VIITANEN,J.O. Distance Transform Algorithm for Bit-Serial SIMD Architectures. ***Computer Vision and Image Understanding***, (74)2, 150-161, 1999.

TANESE,R. ***Distributed Genetic Algorithms for Function Optimization***. PhD Thesis, University of Michigan, EUA, 1989.

TANLER,R. ***The Intranet Data Warehouse***. John Wiley and Sons, 338pp., 1997.

TESAURO,G. Temporal Difference Learning and TD-Gammon. ***Communications of the ACM***, (38)3, 58 -68, 1995.

THIERENS,D.; GOLDBERG,D. Convergence models of genetic algorithm selection schemes. *In*: DAVIDOR,Y.; SCHWEFEL,H.; MÄNNER,R. (Eds.) ***Parallel Problem Solving from Nature*** - PSPN III, Springer-Verlag, 119-129, 1994.

THOMPSON,S.; ROSENFELD,A. Isotropic Growth on a Grid. ***Pattern Recognition***, (28), 241-253, 1995.

TSAI,D.-M. Boundary-based Corner Detection using Neural Networks. ***Pattern Recognition***, (30)1, 85-97, 1997.

TWOCROWS,C. ***Introduction to Data Mining and Knowledge Discovery***. Two Crows Inc., 1998.

ULTSCH,A.; SIEMON,H.P. Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. *In*: ***Proceedings of International Neural Network Conference (INNC'90)***, Dordretch, The Netherlands, Kluwer, 305-308, 1990.

VESANTO,J. ***Using SOM in Data Mining***. Licenciate Thesis, Helsink University of Technology, Finland, 1-50p, 2000.

WATT,R. ***Visual Processing, Computational, Psychophysical and Cognitive Research Essays in Cognitive Psychology***. Lawrence Erlbaum Associates Publishers, 1988.

WEISS,S.; INDURKHYA,N. ***Predictive Data Mining***. Morgan-Kaufman, 1997.

WESTHEIMER,G. Gestalt theory reconfigured: Max Wertheimer's anticipation of recent developments in visual neuroscience. ***Perception***, (28), 5-15, 1999.

WONG,A.K.C.; SAHOO,P.K. A Gray-level Threshold Selection Method Based on Maximum Entropy Principle. ***IEEE Transactions on Systems, Man, and Cybernetics***, (19)4, 866-871, 1989.

WRIGHT,M.W. ***The extended Euclidean distance transform***. PhD Thesis, University of Cambridge. 132pp, 1995.

WÜRTZ,R.P.; LAURENS,T. Corner detection in color images through a multiscale combination of end-stopped cortical cells. ***Image and Vision Computing***, (18), 534-541, 2000.

YAMADA,H. Complete Euclidean distance transformation by parallel operation. *In*: ***Proceedings of 7th International Conference on Pattern Recognition***, Montreal, Canada 336-338, 1984.

YOUNG,I.T.; WALKER,J.E.; BOWIE,J.E. An analysis technique for biological shape. ***Information and Control***, (25), 357-370, 1974.

ZEKI,S.; SHIPP,S. The Functional Logic of Cortical Connections. ***Nature***, (335), 311-317, 1988.

ZHANG,T.; RAMAKRISHNAN,R.; LIVNY,M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *In*: ***Proceedings of ACM SIGMOD'96***, Montreal, Canada, ACM, 103-114p, 1996.

9 APÊNDICE

9.1 Relações

dmnAnswers(*email, code, xaverage, yaverage, ntagrad, ntqually, ntcolor, ntorigin, ntmotiv, tmexplore, blnum, bllink, blpersons, cthomepg*)
email: endereço eletrônico do sujeito
code: código da imagem
xaverage, *yaverage*: ponto de maior atenção definido pelo sujeito
ntagrad: nota sobre prazer visual
ntqually: nota sobre qualidade artística
ntcolor: nota sobre o equilíbrio no uso das cores
ntorigin: nota sobre a originalidade da página
ntmotiv: nota sobre a motivação para continuar navegando
tmexplore: tempo estimado pelo sujeito para explorar toda a página
blnum: resposta sobre a visualização de dígitos
bllink: resposta sobre a visualização de hipertextos
blpersons: resposta sobre a visualização de imagens ou desenhos de pessoas
cthomepg: tipo de home page (comercial, institucional, pessoal)

imgWeb(*code, image, class, width, height, ncolors, map, select*)
code: código identificador da imagem;
image: um atributo BLOB contendo o mapa de bits da imagem;
width, height: dimensões largura e altura da imagem;
ncolors: profundidade de cores em bits;
map: mapa de cores utilizado;
select: indicador de imagem selecionada.

dmnSpc(*code, radscl, angscl, raddta, angdta, npoints*)
code: código identificador da imagem;
radscl: escala para o descritor espectral espacial;
angscl: escala para o descritor espectral direcional;
raddta: dados para o descritor espectral espacial;
angdta: dados para o descritor espectral direcional;
npoints: número de dados.

dmnEnt(*code, scale, data, npoints*)
code: código identificador da imagem;

scale: escala para a entropia;
data: dados para a entropia;
npoints: número de dados.

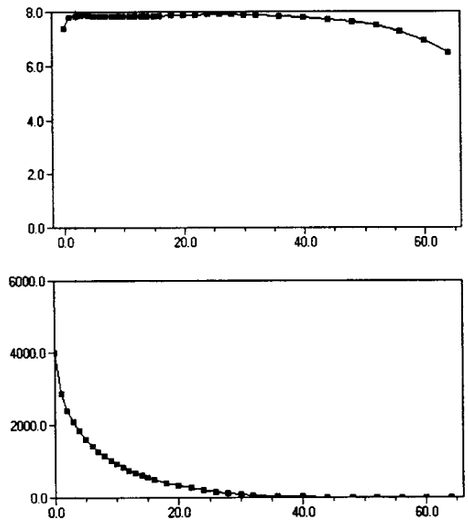
dmnVar(*code, scale, data, npoints*)
code: código identificador da imagem;
scale: escala para a variância;
data: dados para a variância;
npoints: número de dados.

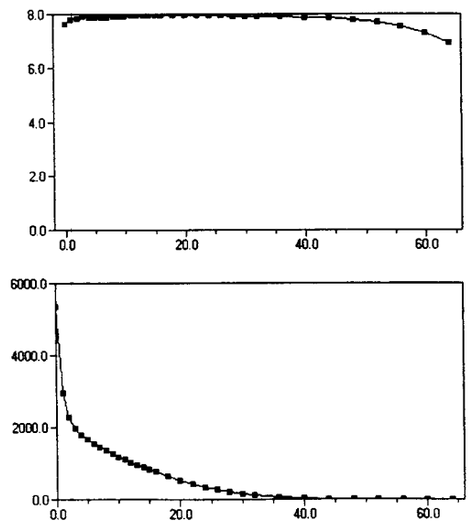
dmnFrc(*code, scale, data, frcdim, npoints*)
code: código identificador da imagem;
scale: escala para a salsicha de Minkowski;
data: dados para a salsicha de Minkowski;
frcdim: dimensão fracionária;
npoints: número de dados.

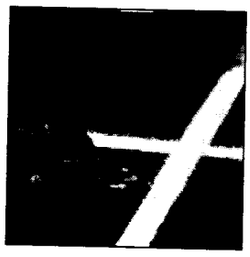
dmnCnt(*code, scale, gr_data, rgb_data, rdata, gdata, bdata, npoints*)
code: código identificador da imagem;
scale: intervalo do histograma de contraste;
gr_data: dados para o contraste em níveis de cinza;
rgb_data: dados para o contraste de cores;
rdata: dados para o contraste em níveis de vermelho;
gdata: dados para o contraste em níveis de verde;
bdata: dados para o contraste em níveis de azul;
npoints: número de dados.

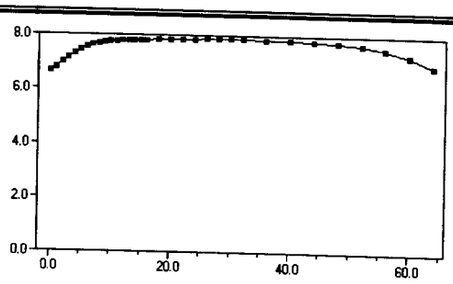
9.2 Entropia e Variância Multi-escalas

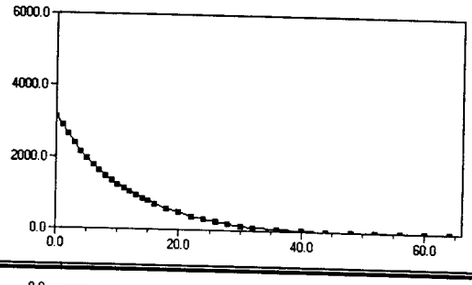
Cada uma das imagens abaixo apresenta duas medidas multi-escalas: o gráfico na parte superior corresponde à *entropia multi-escalas* e o da parte inferior à *variância multi-escalas*.

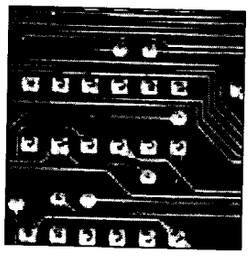


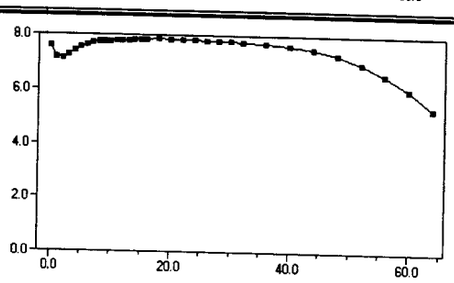


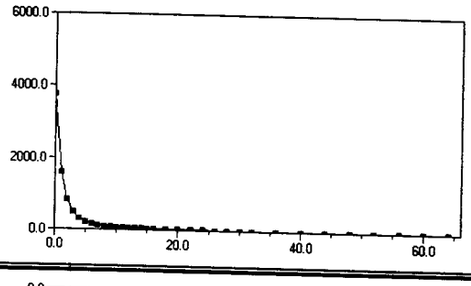


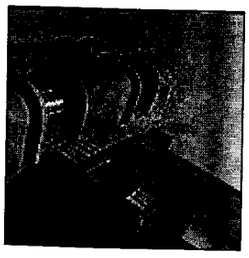

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

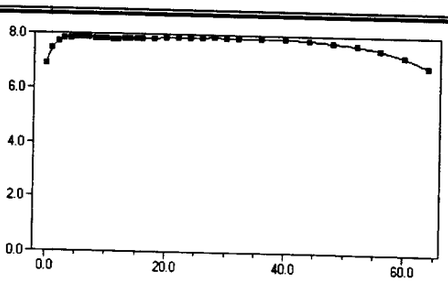

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

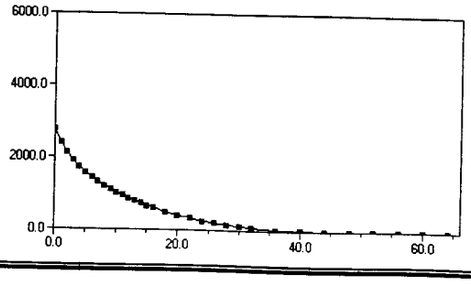


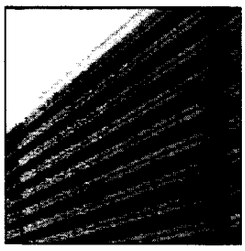

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

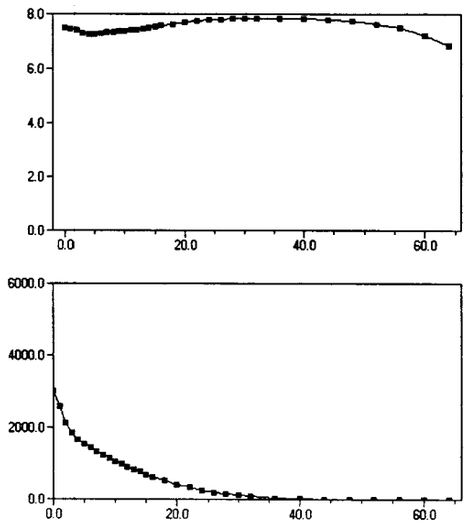

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

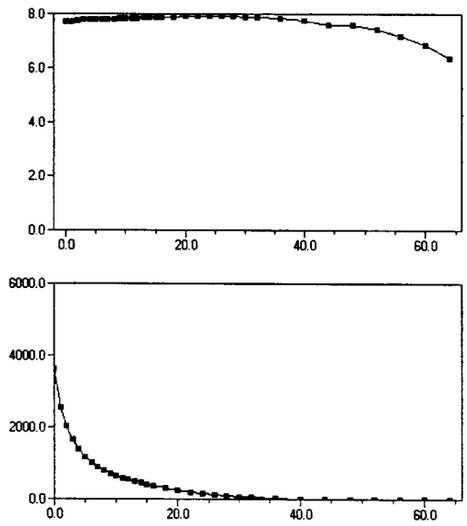


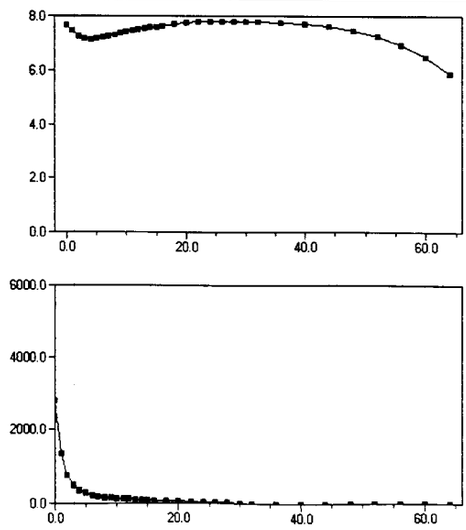

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

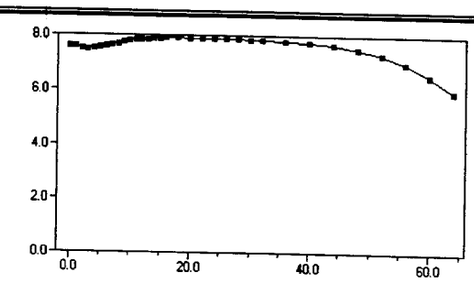

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

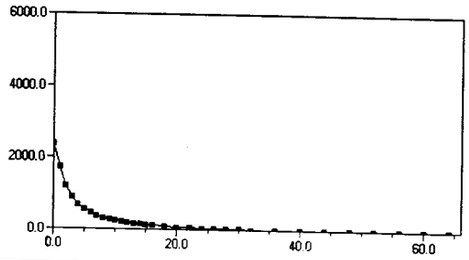


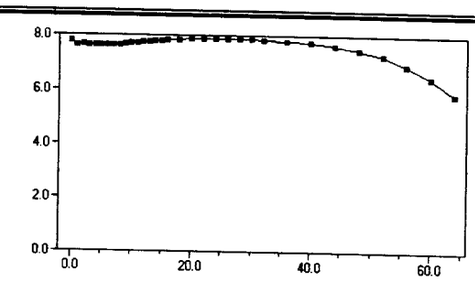

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

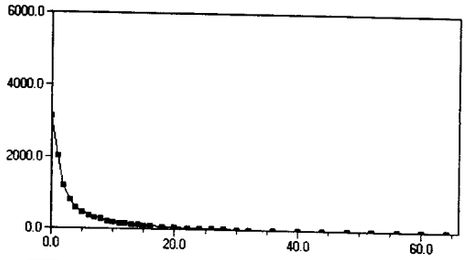

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

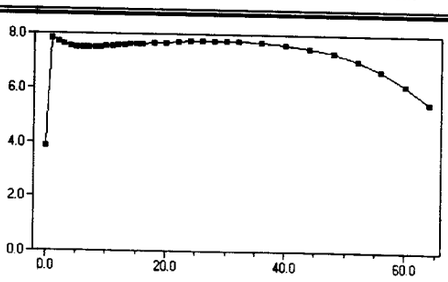

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

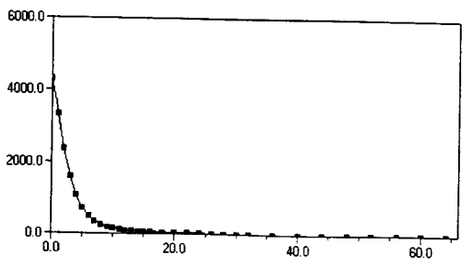

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

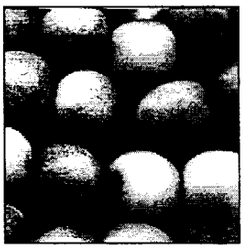

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

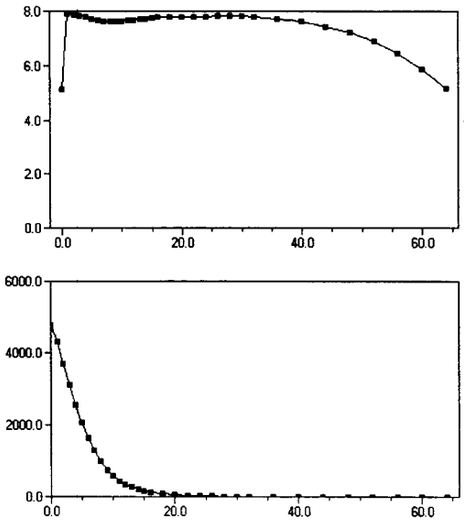

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

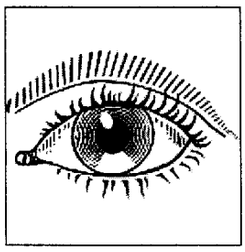

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

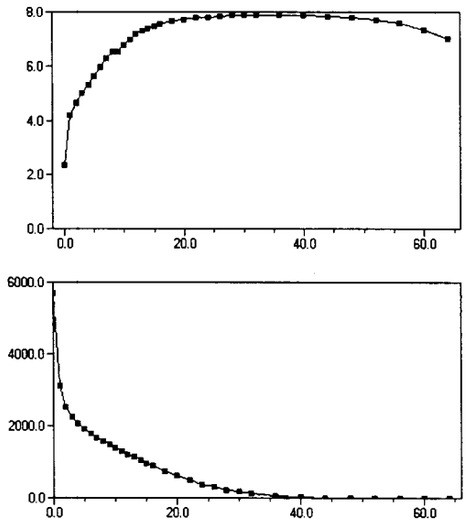

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

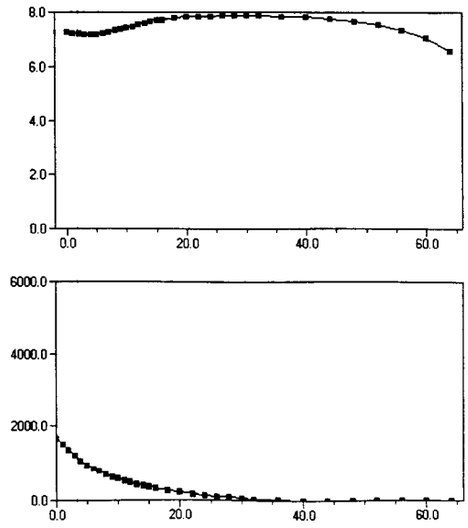

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

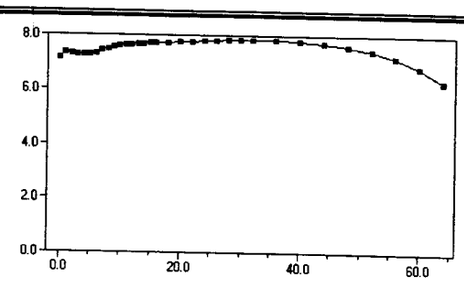


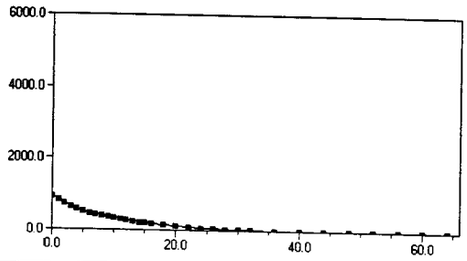

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

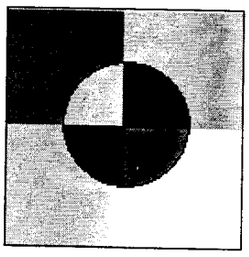


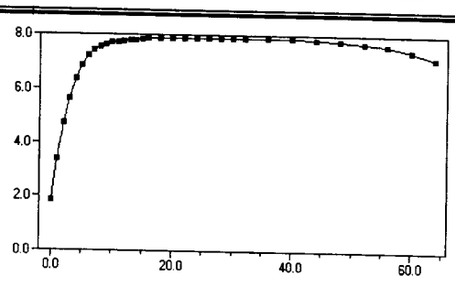

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

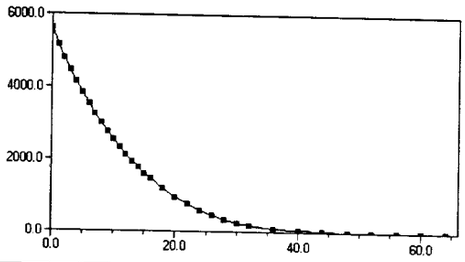

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0
6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

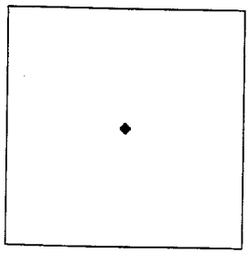

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

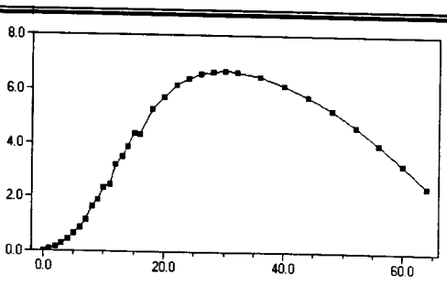

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

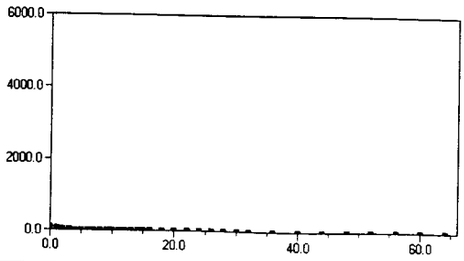


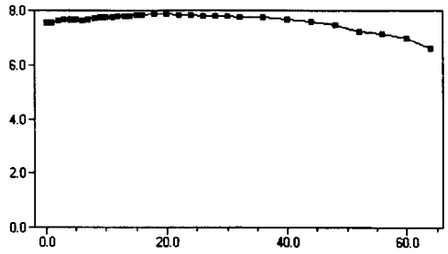

8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0

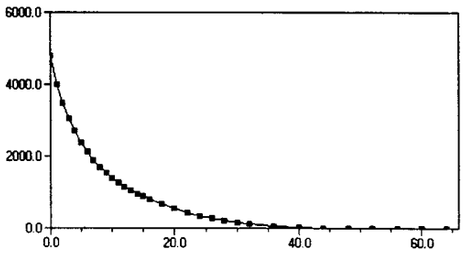

6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0




8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0


6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0


8.0
6.0
4.0
2.0
0.0
0.0
20.0
40.0
60.0


6000.0
4000.0
2000.0
0.0
0.0
20.0
40.0
60.0

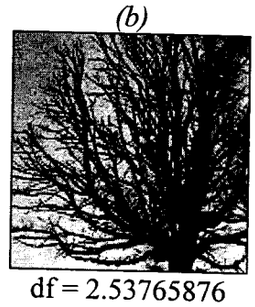
9.3 Comparação de Dimensão Fracionária via Salsichas de Minkowski

Tabela (I)

(a)

df = 2.45958614

(b)



df = 2.53765876

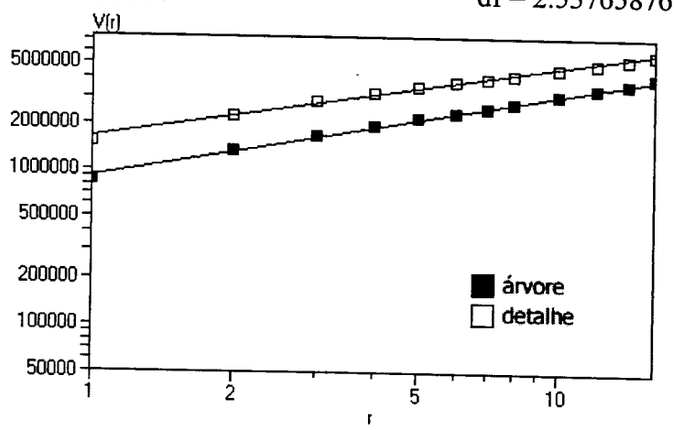


Tabela (II)

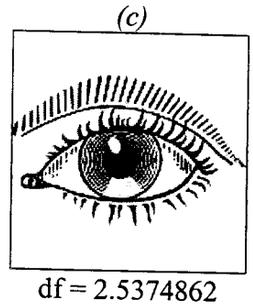
(a)

df = 2.18355212

(b)

df = 2.15504743

(c)



df = 2.5374862

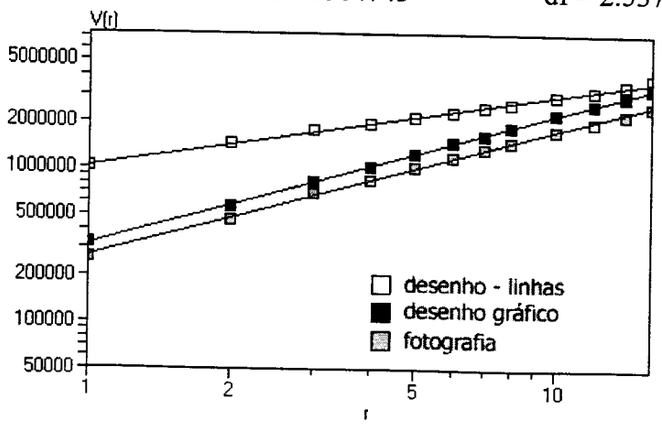
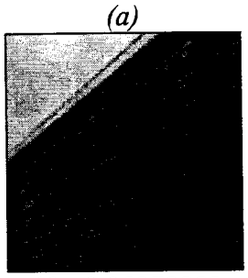


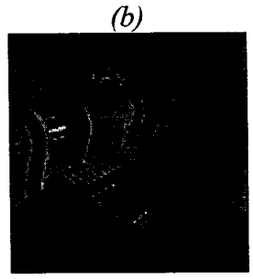
Tabela (III)

(a)



df = 2.35372302

(b)



df = 2.24132242

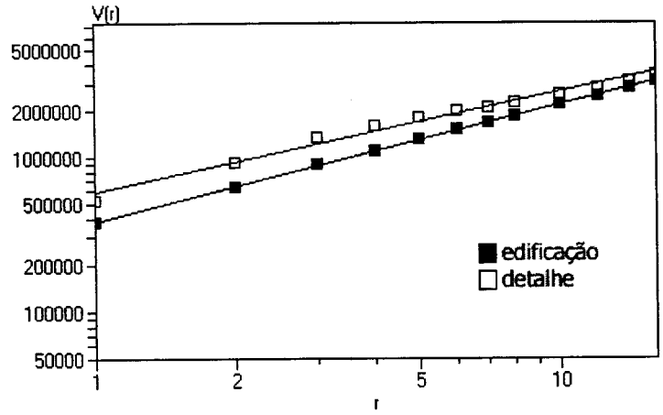
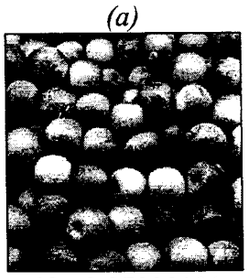


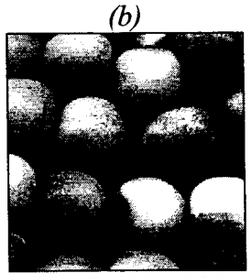
Tabela (IV)

(a)



df = 2.27000131

(b)



df = 2.10624083

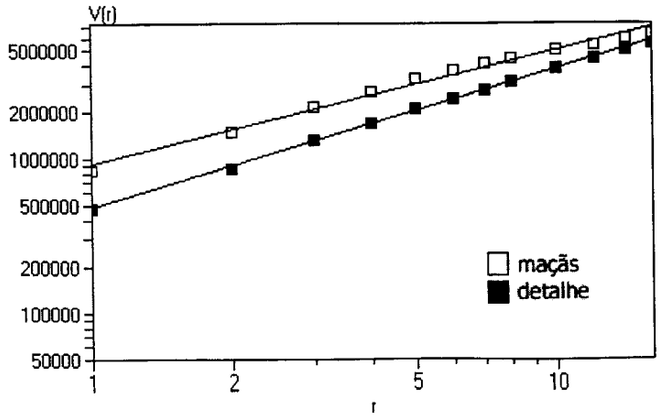


Tabela (V)

(a)

df = 2.47032940

(b)

df = 2.30011938

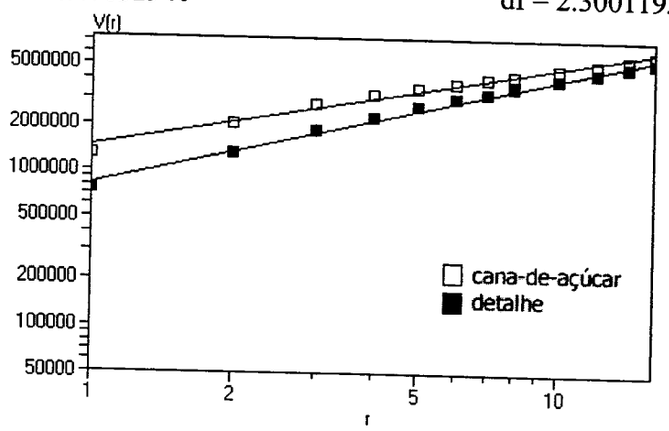
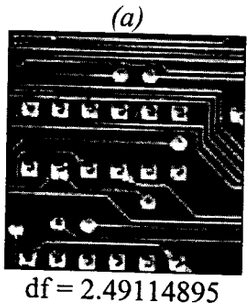


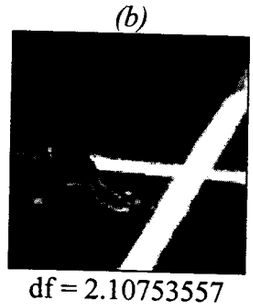
Tabela (VI)

(a)



df = 2.49114895

(b)



df = 2.10753557

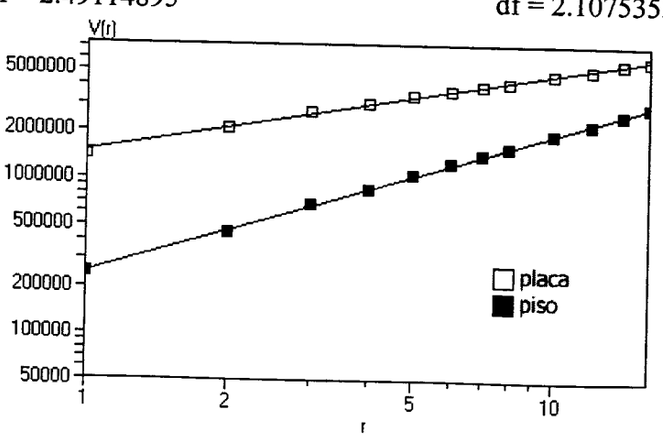
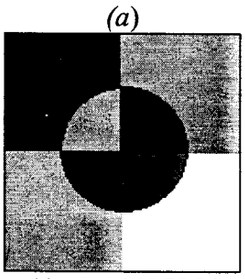


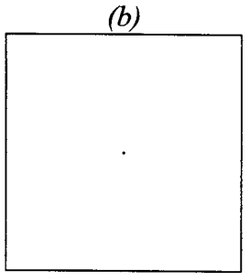
Tabela (VII)

(a)



df = 2.06435954

(b)



df = 1.92359061

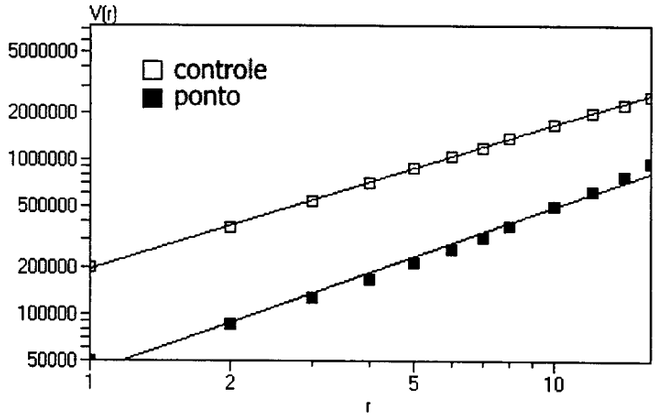
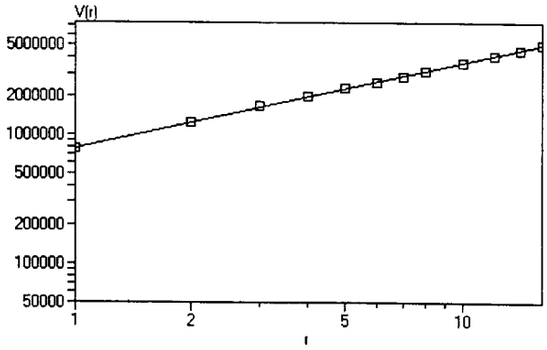


Tabela (VIII)

df = 2.33639455

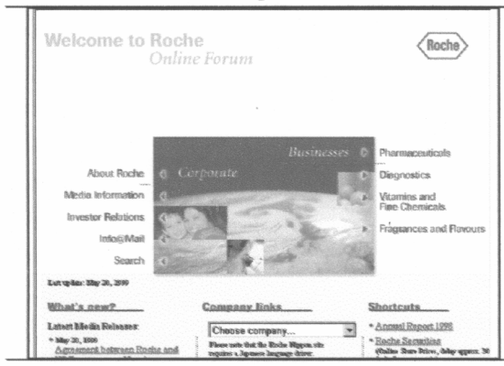


9.4 Imagens de *home-pages* utilizadas no experimento de análise pictórica

1

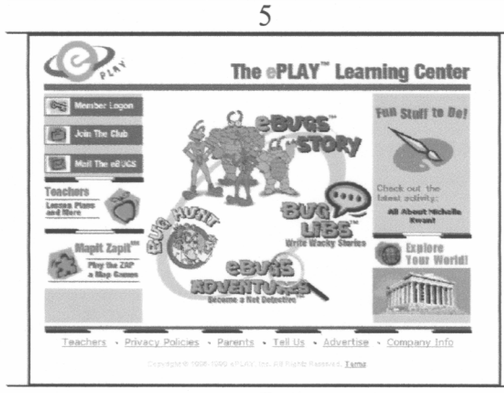
2

3



4

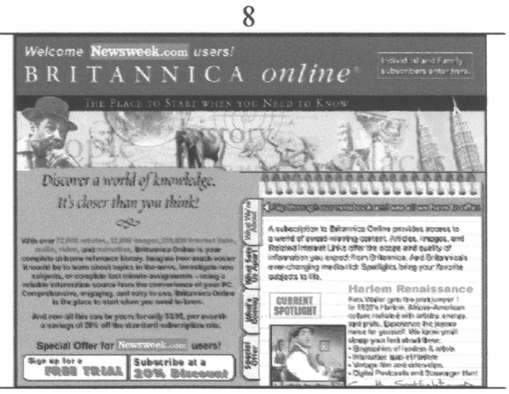
5



6

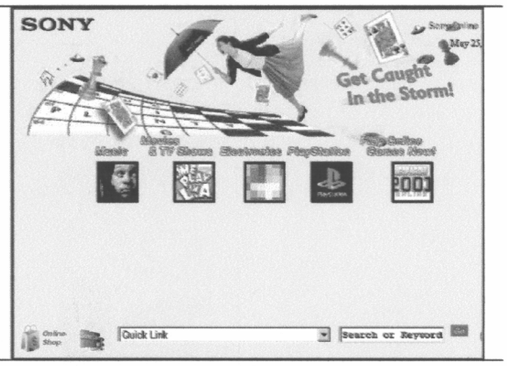
7

8

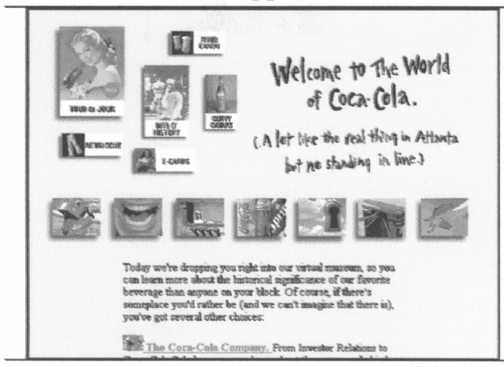


9

10



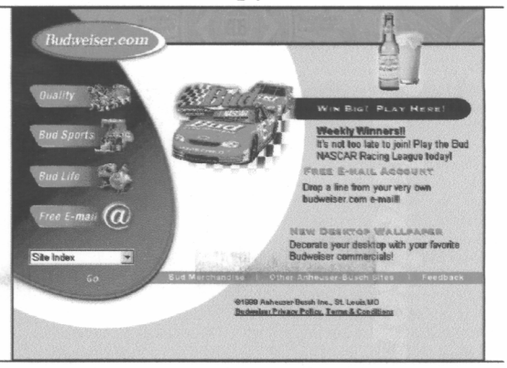
11



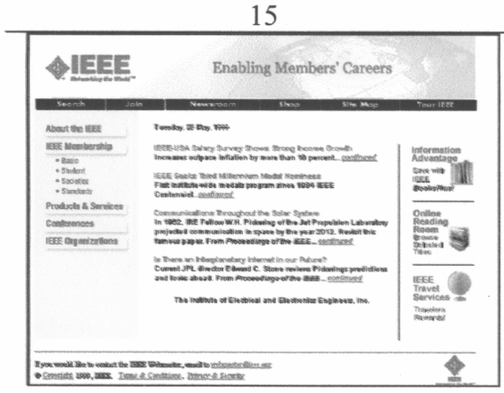
12

13

14

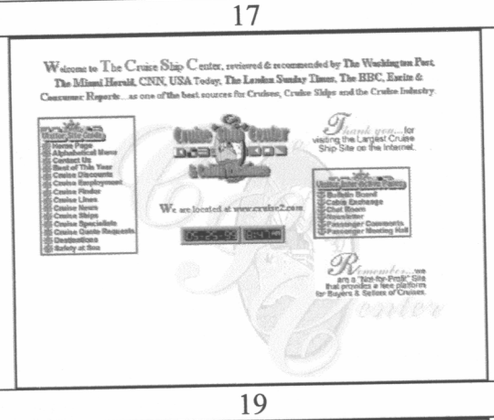


15

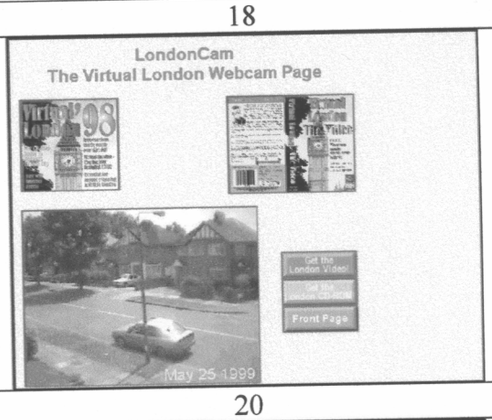


16

17

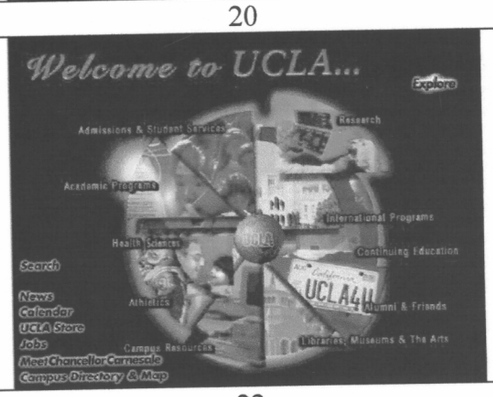


18

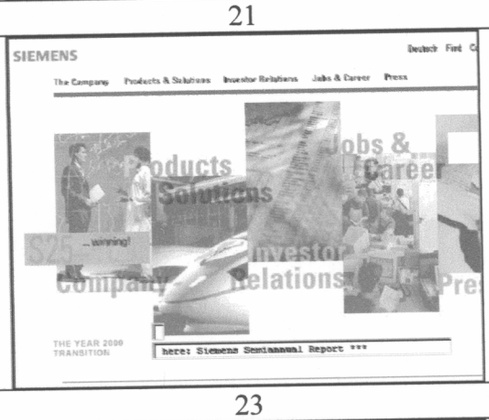


19

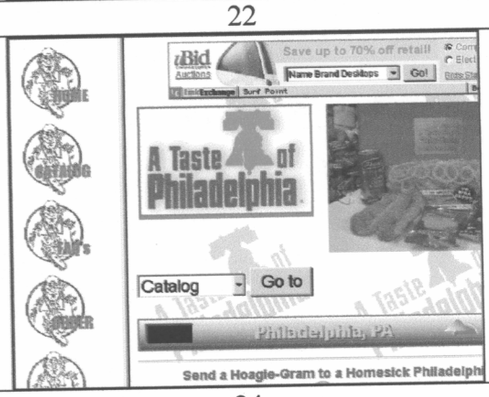
20



21

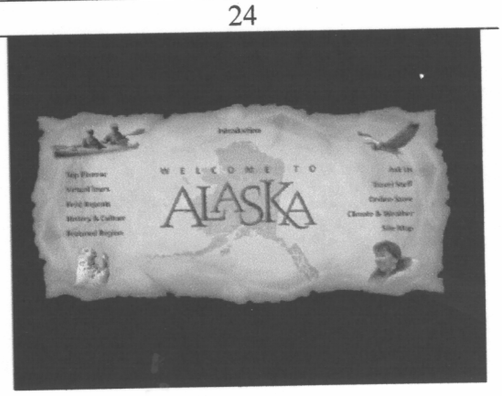


22



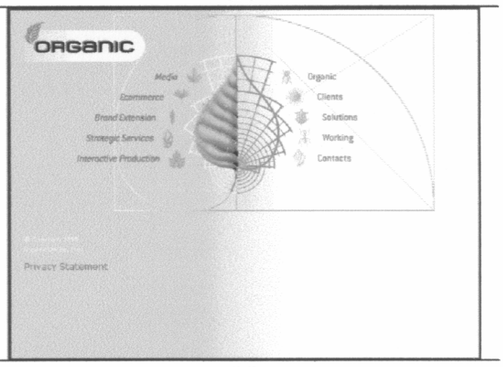
23

24

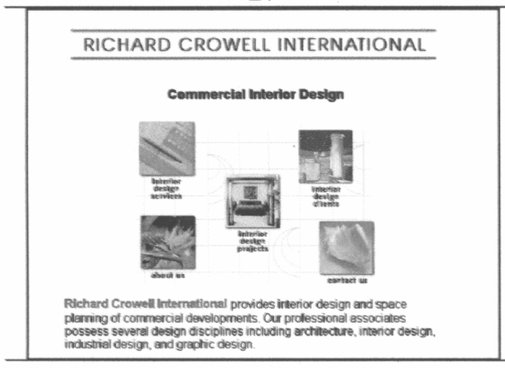


25

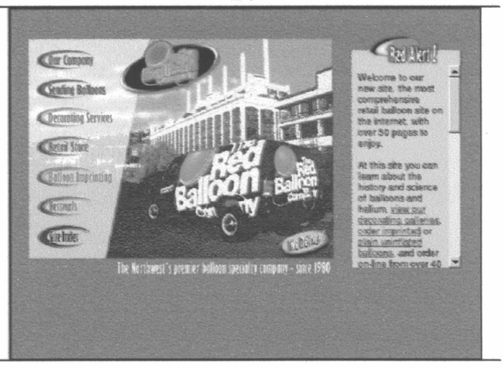
26



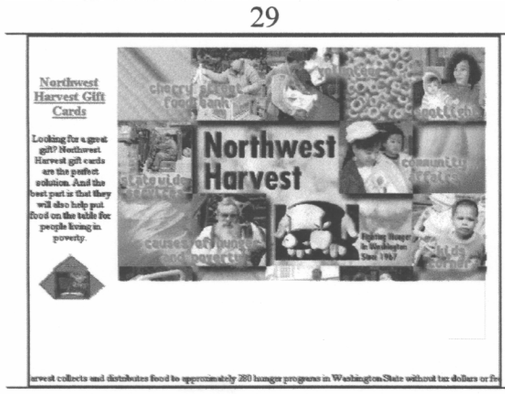
27



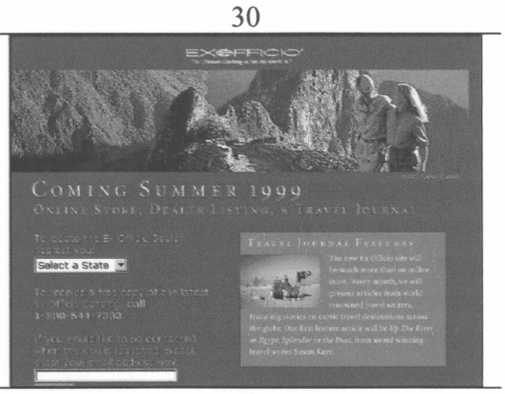
28



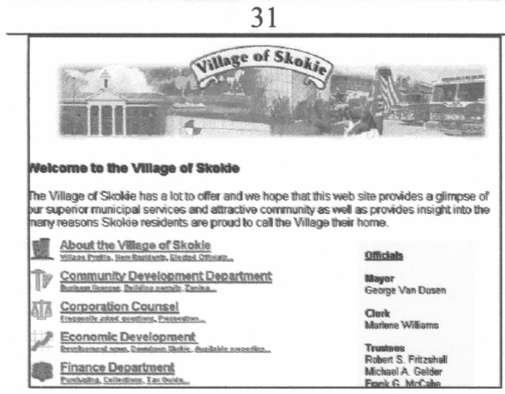
29



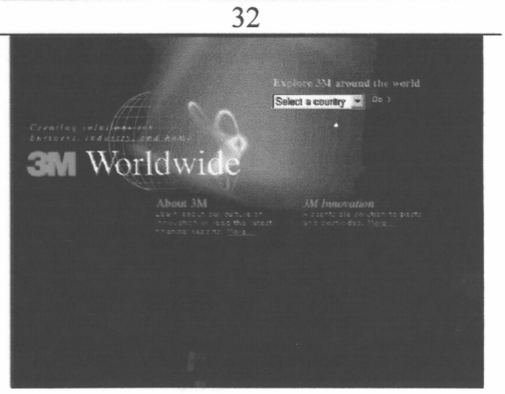
30



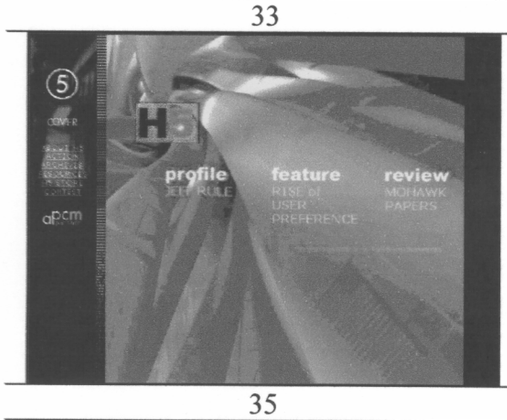
31



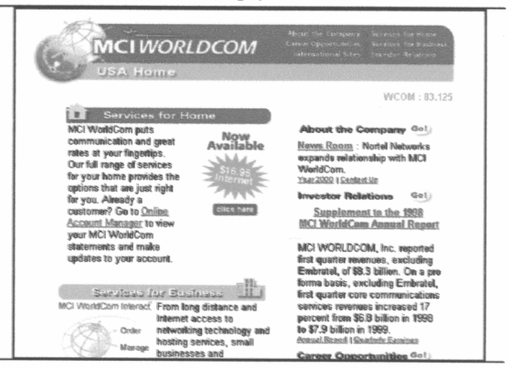
32



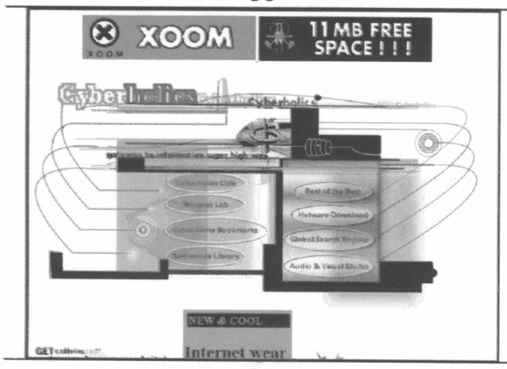
33



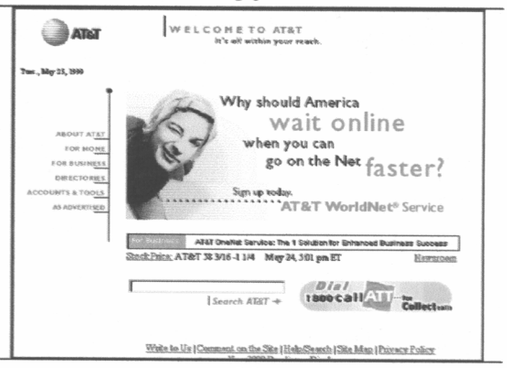
34



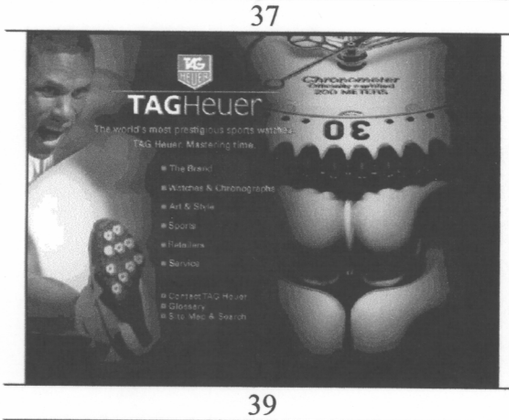
35



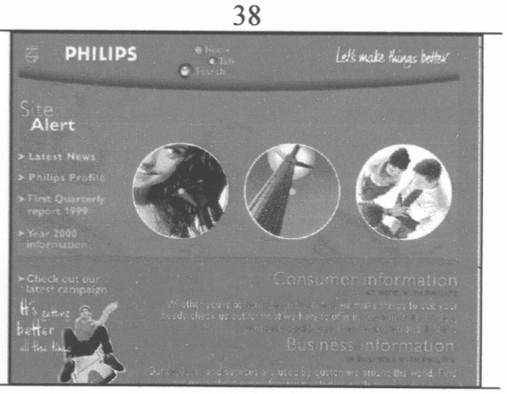
36



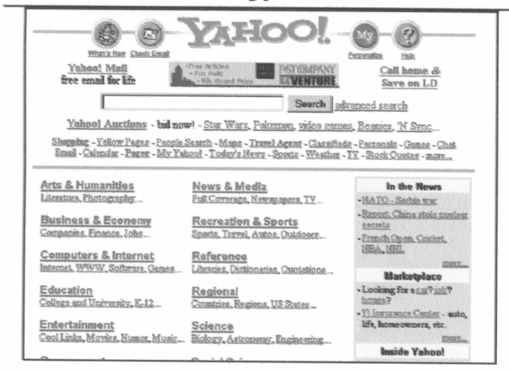
37



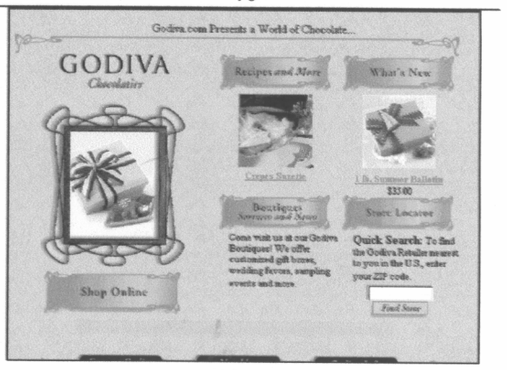
38



39



40



9.5 Notas médias para cada imagem da seção anterior

AGR = agradabilidade; COR = equilíbrio de cores;

QLI = qualidade artística; ORG = originalidade; MTV = motivação

imagem	AGR	COR	QLI	ORG	MTV
1	5,3	4,9	4,2	4,4	4,4
2	6,6	6,1	6,0	5,3	6,0
3	7,0	7,1	6,7	6,0	5,9
4	5,4	5,6	5,3	4,6	5,3
5	6,4	6,3	5,6	6,0	5,6
6	7,6	6,7	6,7	4,4	5,9
7	5,6	5,6	5,2	4,8	5,4
8	6,6	6,1	6,3	6,1	6,7
9	5,8	5,1	5,4	4,9	4,1
10	6,6	5,7	6,7	6,2	5,8
11	6,3	6,3	6,7	6,4	6,1
12	6,7	6,1	5,7	5,8	5,9
13	7,1	6,8	6,6	6,2	5,9
14	6,9	6,4	6,4	6,1	6,6
15	6,1	5,6	5,4	5,1	6,6
16	6,6	6,6	6,6	6,2	6,7
17	6,0	5,7	5,7	5,2	4,9
18	5,9	5,6	5,0	4,6	4,9
19	5,9	6,2	5,8	4,9	5,6
20	7,2	6,6	7,6	6,9	6,8
21	6,6	3,9	6,2	6,1	6,0
22	5,4	5,3	5,0	5,0	4,8
23	6,8	6,6	6,6	6,2	5,8
24	7,6	7,1	7,8	7,7	6,8
25	6,4	6,3	5,9	5,2	6,1
26	7,7	7,4	7,6	7,4	6,9
27	6,1	6,1	5,4	5,1	5,0
28	6,1	6,2	6,1	5,9	5,3
29	5,1	5,0	5,4	5,1	4,7
30	6,8	6,7	7,2	6,6	6,3
31	5,7	5,9	5,2	4,3	4,7
32	7,6	7,1	7,2	6,9	6,6
33	7,6	7,3	7,2	7,3	6,2
34	5,4	4,8	4,6	3,8	4,6
35	6,4	7,1	6,8	6,1	6,0
36	6,2	5,3	5,4	5,2	6,0
37	5,9	3,6	6,2	5,9	5,8
38	6,8	6,7	6,6	6,7	6,6
39	5,8	4,9	4,0	4,7	5,7
40	7,4	7,0	7,7	6,6	6,9