

Resumo

Os avanços tecnológicos registados nos últimos anos na área da tecnologia audiovisual levaram à explosão do uso da informação audiovisual, nomeadamente acedida via *Internet*, por um vasto número de utilizadores. No entanto, a crescente quantidade de informação disponível dificulta a procura e acesso à informação de que cada utente necessita. Para ajudar a resolver este problema, o grupo MPEG lançou um projecto de normalização, denominado MPEG-7. A norma MPEG-7 deverá especificar um conjunto de elementos normativos (descritores, esquemas de descrição, esquemas de codificação e uma linguagem de definição de descrições) de forma a satisfazer um conjunto relevante de requisitos relacionados com a descrição e procura de conteúdo audiovisual.

Uma das principais propriedades do MPEG-7 é a sua capacidade de descrever de forma independente objectos na cena audiovisual, sendo cada objecto representado através da sua informação de textura e de forma. A informação de forma aparece como a grande novidade associada à representação de objectos, ao contrário do que acontecia no passado com as tramas rectangulares. Esta tese estuda o problema da descrição de formas visuais em íntima relação com o que está a ser desenvolvido nesta área no contexto da norma MPEG-7. Assim os grandes objectivos desta tese são:

1. Revisão bibliográfica e análise comparativa dos parâmetros de descrição de forma disponíveis;
2. Revisão bibliográfica e análise comparativa dos principais mecanismos de procura disponíveis que usam a informação de forma;
3. Avaliação dos descritores de forma propostos e adoptados pelo MPEG-7 para um conjunto relevante de material de teste e em condições rigorosamente definidas;
4. Desenvolvimento de um mecanismo de descrição e procura com ênfase na característica de forma;
5. Desenvolvimentos ao descritor ângulos de curvatura do contorno.

Palavras-chave: MPEG-7, descrição e procura de conteúdo visual, descritores de forma, medidas de semelhança de forma, procuras por exemplo de conteúdo e de esboço, ângulos de curvatura do contorno.

Abstract

The technological advances registered in the last years in the audiovisual technological field, have led to a boom in the usage of audiovisual information, namely accessed through the *Internet*, by a great number of users. The deployment of audiovisual information and the increasing amount of available information have made it more difficult to retrieve the desired information. To help solve this problem, the MPEG group has launched a standardization project, named MPEG-7. The MPEG-7 standard will specify a set of normative elements (descriptors, description schemes, coding schemes and a description definition language) in order to satisfy a set of demanding requirements related to the description and retrieval of audiovisual content.

One of MPEG-7's main properties is its capacity to describe independent objects in audiovisual scenes, each being represented by its texture and shape information. Shape information appears as the novelty associated to the representation of objects, as opposed to what happened in the past with rectangular frames. This Thesis will study the problem associated to shape descriptions in close relationship to what is being developed in this area within MPEG-7. Therefore the big goals of this Thesis are:

1. Bibliographic study and comparative analysis of the available shape description parameters;
2. Bibliographic study and comparative analysis of the main retrieval mechanisms using shape information;
3. Evaluation of the proposed and adopted MPEG-7 shape descriptors using its dataset and well-defined experimental conditions;
4. Development of a description and retrieval mechanism with emphasis on the shape characteristic;
5. Improvements to the turning angles descriptor.

Keywords: MPEG-7, description and retrieval of audiovisual content, shape descriptors, shape similarity measures, retrieval by example of content and of sketch, turning angles.

Agradecimentos

Gostaria de começar por agradecer ao Professor Fernando Pereira o acompanhamento rigoroso, dedicação e paciência durante todas as fases desta tese.

Ao meu marido e colega, Luis Ducla Soares, pelo apoio incontestável, pelo interesse e disponibilidade sempre demonstrado resultando em sugestões construtivas.

Aos meus pais, irmã e avós pelo carinho, apoio, incentivo e coragem que sempre me deram permitindo ser quem sou e chegar onde cheguei.

A todos os colegas do grupo de imagem pelo bom ambiente que sempre criaram, pelas sugestões que foram dando e pela disponibilidade permanente que facilitou o trabalho desenvolvido.

Índice

Capítulo 1	Contexto e objectivos.....	1
1.1	A norma MPEG-7	4
1.1.1	MPEG-7: os elementos normativos	5
1.1.2	A relação com a norma MPEG-4.....	6
1.1.3	MPEG-7: as aplicações.....	7
1.2	Terminologia relevante.....	9
1.3	Objectivos e organização desta tese.....	10
Capítulo 2	Parâmetros de forma	13
2.1	Requisitos	13
2.2	Parâmetros de forma baseados no contorno.....	15
2.2.1	Parâmetros geométricos.....	16
2.2.2	Parâmetros baseados em transformadas.....	20
2.2.3	Parâmetros baseados em momentos.....	25
2.2.4	Parâmetros baseados em contornos normalizados	26
2.2.5	Parâmetros baseados nos ângulos de curvatura do contorno	29
2.2.6	Parâmetros baseados numa imagem <i>Curvature Scale Space</i>	31
2.3	Parâmetros de forma baseados em regiões	34
2.3.1	Parâmetros geométricos.....	35
2.3.2	Parâmetros de forma baseados em vectores próprios multi-nível.....	40
2.3.3	Parâmetros de forma baseados em transformadas	41
2.3.4	Parâmetros de forma baseados em momentos	43
2.4	Conclusões.....	45
Capítulo 3	Mecanismos de descrição e procura	49
3.1	Arquitectura de um mecanismo de descrição e procura.....	50
3.1.1	Módulo de descrição.....	52
3.1.2	Módulo da base de dados.....	52
3.1.3	Módulo de procura.....	54
3.2	Alguns mecanismos de descrição e procura de vídeo.....	60
3.2.1	NETRA.....	60
3.2.2	<i>QBIC</i>	64
3.2.3	<i>MARS</i>	69
3.2.4	<i>VideoQ</i>	71
3.3	Conclusões.....	74
Capítulo 4	Mecanismo de descrição e procura de formas	79
4.1	Arquitectura geral da aplicação	80
4.1.1	Estrutura de dados.....	81
4.1.2	Estrutura da interface gráfica	82
4.2	Menu principal.....	85
4.3	Barra de ferramentas	87
4.4	Janela da base de dados	88
4.5	Janela de descrição e procura.....	89
4.6	Módulo de descrição.....	91
4.6.1	Navegação pelas bases de dados de conteúdo.....	92
4.6.2	Geração de descrições.....	93
4.7	Módulo de procura.....	95
4.7.1	Navegação pelas bases de dados com descrições	96
4.7.2	Definição da procura.....	96
4.7.3	Apresentação dos resultados da procura	102
4.8	Comentários Finais	103
Capítulo 5	Avaliação dos descritores de forma considerados e adoptados pelo MPEG-7..	105
5.1	MPEG-7: Processo de Normalização.....	107
5.2	<i>Core Experiments</i> : Metodologias de Avaliação para a Característica de Forma.....	112
5.2.1	Core Experiment-1 (<i>CE-1</i>).....	114
5.2.2	Core Experiment-2 (<i>CE-2</i>).....	119
5.2.3	Core Experiment <i>CE-1/2</i> Optimizado (<i>CE-1/2</i>).....	123

5.2.4	Core Experiment MS-3 (CE MS-3).....	123
5.2.5	Core Experiment MS-5 (CE MS-5).....	124
5.3	Seleccção dos Descritores de Forma Propostos ao MPEG-7.....	124
5.3.1	Avaliação independente do desempenho dos descritores	125
5.3.2	Primeira fase do processo de desenvolvimento	127
5.3.3	Segunda fase do processo de desenvolvimento	142
5.3.4	Terceira fase do processo de desenvolvimento.....	146
5.3.5	Quarta fase do processo de desenvolvimento	160
5.4	Descritores de Forma Adoptados pelo MPEG-7.....	161
5.4.1	Descritor de forma baseado no contorno	162
5.4.2	Descritor de forma baseado na região.....	165
5.5	Conclusões.....	166
Capítulo 6	Desenvolvimentos ao descritor ângulos de curvatura do contorno	169
6.1	Optimização inicial da dimensão do descritor ACC.....	170
6.1.1	Eliminação das componentes não essenciais do descritor ACC	172
6.1.2	Análise da dimensão do descritor ACC.....	173
6.2	Melhoria do desempenho de procura e computacional através da inclusão de parâmetros geométricos 174	
6.2.1	Redefinição da sintaxe e semântica do descritor ACC	174
6.2.2	Definição da nova medida de semelhança	176
6.2.3	Análise do desempenho do descritor	177
6.3	Diminuição da sensibilidade à simetria do descritor ACC	178
6.3.1	Definição da nova medida de semelhança	179
6.3.2	Nova Análise do desempenho do descritor ACC.....	184
6.4	Optimização do número de ângulos de curvatura usados por descrição.....	185
6.4.1	Optimização da componente <i>Número_Ângulos</i> do descritor ACC.....	185
6.4.2	Nova análise do desempenho do descritor ACC.....	186
6.5	Estudo do descritor ACC usando a distância Euclidiana.....	189
6.5.1	Definição e implementação da medida de semelhança baseada na distância Euclidiana	190
6.5.2	Nova análise do desempenho do descritor ACC.....	191
6.6	Estudo do descritor ACC usando a distância <i>alpha-trimmed average</i>	191
6.6.1	Definição e implementação da medida de semelhança baseada na distância <i>alpha-trimmed average</i> 191	
6.6.2	Nova análise de desempenho do descritor ACC.....	193
6.7	Optimização do número de vectores de ângulos de curvatura no descritor ACC.....	195
6.7.1	Definição do critério de selecção do número de vectores de ângulos no descritor ACC.....	195
6.7.2	Redefinição da sintaxe e semântica do descritor ACC	197
6.7.3	Nova análise do desempenho do descritor ACC.....	198
6.8	Optimização adicional do número de vectores de ângulos para o descritor ACC	200
6.8.1	Definição do critério de selecção para o único ponto de começo	201
6.8.2	Nova análise do desempenho do descritor ACC.....	202
6.9	Optimização do número de <i>bits</i> para cada uma das componentes do descritor ACC	203
6.9.1	Nova análise do desempenho do descritor ACC.....	203
6.10	Proposta final para o descritor ACC	205
6.10.1	Definição final da sintaxe e semântica do descritor ACC.....	205
6.10.2	Medida de semelhança final para o descritor ACC.....	206
6.11	Conclusões.....	207
Capítulo 7	Conclusões	209
Referências	213

Lista de figuras

Figura 2.1 – <i>Bream</i> (trama1): a) Imagem com o objecto; b) contorno do objecto.....	15
Figura 2.2 – <i>Bream</i> : a) Forma do objecto simples (suporte da textura); b) Representação gráfica do perímetro do objecto em a)	17
Figura 2.3 – <i>Classic</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica da corda máxima do objecto	17
Figura 2.4 – <i>Bell</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica da circularidade do objecto ...	18
Figura 2.5 – <i>Device</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica da convexidade do objecto	19
Figura 2.6 – <i>Bird</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica dos raios máximo e mínimo do objecto	19
Figura 2.7 – Esquema representativo da sequência de operações efectuadas para o cálculo dos coeficientes da transformada baseada em <i>wavelets</i> [PrTe1994].....	24
Figura 2.8 – <i>Children</i> : a) Forma a analisar; b) Extracção do contorno, com um número de pontos variável e não equidistante, ajustado ao objecto simples ou região; c) Contorno re-amostrado a 64 pontos usando-se o método das <i>Cubic Splines</i> ; d) Contorno com descrição insensível à rotação [Taba1999]	27
Figura 2.9 – Determinação do contorno normalizado [Taba1999]	27
Figura 2.10 – Exemplo representativo de ângulos de curvatura do contorno	30
Figura 2.11 – Objecto ‘peixe’: a) textura do objecto e b) Imagem CSS do contorno do objecto [SQUID]	32
Figura 2.12 – Objecto ‘peixe’: a), c), e) e g) Representação multi-escalar através da forma contorno do objecto simples ou região; b), d), f) e h) Imagem CSS do objecto simples ou região intersectado com a linha horizontal, de valor σ , representativa do número de zeros existentes nas formas dos objectos adjacentes [SQUID]	33
Figura 2.13 – Exemplos de objectos complexos e respectivas regiões e buracos [MPEG-7XM1999]	34
Figura 2.14 – <i>KK-902</i> : a) Forma do objecto simples e <i>bounding box</i> correspondente [CE1]; <i>1001</i> : b) Forma do objecto complexo e <i>bounding box</i> correspondente [CE2S8]	35
Figura 2.15 – <i>Butterfly</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica da área do objecto; <i>1222</i> : c) Forma do objecto complexo [CE2S8]; d) Representação gráfica da área do objecto	36
Figura 2.16 – <i>Apple</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica do centróide do objecto; <i>1253</i> : c) Forma do objecto complexo [CE2S8]; d) Representação gráfica do centróide do objecto	37
Figura 2.17 – <i>AEA005D</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica das projecções do objecto; <i>1008</i> : c) Forma do objecto complexo [CE2S8]; b) Representação gráfica das projecções do objecto	37
Figura 2.18 – <i>Device2-1</i> : a) Forma do objecto simples [CE1]; b) Representação gráfica do <i>DCE</i> do objecto; <i>1007</i> : c) Forma do objecto complexo [CE2S8]; d) Representação gráfica do <i>DCE</i> do objecto	38
Figura 2.19 – <i>Device5-1</i> : a) Forma do objecto; b) Representação gráfica da forma convexa do objecto	39
Figura 2.20 – <i>Bat</i> : Exemplos de sub-divisão do objecto a) Após primeira sub-divisão, b) Após segunda sub-divisão [KiKi1999d]	40
Figura 2.21 – Parte real da função de base da transformada <i>ART</i> [KiKi2000]	43
Figura 2.22 – Funções de base de <i>Zernike</i> até à ordem 8: a) parte real b) parte imaginária [KiKi1999b].....	45
Figura 3.1 – Arquitectura genérica de um mecanismo de descrição e procura para aplicações do tipo <i>Pull</i>	51
Figura 3.2 – Exemplo de procura por estatística do sinal, nomeadamente percentagens ou outros valores numéricos [QBIC_Stamps].....	56
Figura 3.3 – Procuras baseadas em exemplos por conteúdo: interfaces para os mecanismos a) <i>MARS</i> [MARS] e b) <i>QBIC</i> [QBIC_Stamps]	57
Figura 3.4 – Exemplo de procura baseada num exemplo por esboço usando como parâmetros a relação espacial, a cor e o tamanho dos objectos ou regiões a procurar [QBIC_Stamps].....	58
Figura 3.5 – Exemplo de procura baseada num exemplo por esboço delineando o contorno sobre uma imagem de referência [MARS].....	58
Figura 3.6 – Exemplo de procura baseada num exemplo por esboço: o esboço é desenhado com figuras geométricas, cor, textura e movimento para cada um dos objectos na imagem [VisualSEEK].....	59
Figura 3.7 – Interface de procura baseada num exemplo por conteúdo para o <i>NETRA</i> : resultados obtidos usando os parâmetros a) função de coordenadas complexas; b) função da distância ao centróide [NETRA]	64
Figura 3.8 – <i>QBIC</i> – Interfaces de procura da aplicação dos logótipos: a) interface por exemplo inicial; b) interface por exemplo intermédia (escolheu-se como exemplo a <i>mask 0</i>); e c) interface que apresenta os resultados da procura [QBIC_Trade].....	68
Figura 3.9 – <i>QBIC</i> : a) interface com procura do tipo esboço como exemplo; b) interface com o resultado da procura [Finn1996]	69

Figura 3.10 – <i>MARS</i> – Interface de procura: a) procura baseada em exemplo por conteúdo; b) procura baseada em exemplo por esboço; c) resultados da procura efectuada usando a característica de forma [MARS].....	71
Figura 3.11 – <i>VideoQ</i> : a) Interface de entrada usando esboço como exemplo; b) resultados da pesquisa definida em a) [VideoQ].....	74
Figura 4.1 – Estrutura de directórios para a aplicação desenvolvida.....	81
Figura 4.2 – Interfaces da aplicação desenvolvida - a) de descrição; b) de procura baseada em conteúdo como exemplo e c) de procura baseada em esboço como exemplo.....	84
Figura 4.3 – Organização espacial do ecrã da aplicação.....	85
Figura 4.4 – Menu principal da aplicação desenvolvida.....	85
Figura 4.5 – Sub-menu <i>Database</i> da aplicação desenvolvida.....	86
Figura 4.6 – Sub-menu <i>Describing</i> da aplicação desenvolvida.....	86
Figura 4.7 – Sub-menu <i>Retrieving</i> da aplicação desenvolvida.....	87
Figura 4.8 – Barra de ferramentas da aplicação desenvolvida.....	87
Figura 4.9 – Janela da base de dados: a) para a interface de descrição e b) para a interface de procura.....	89
Figura 4.10 – Interface de descrição e de procura baseada em conteúdo como exemplo.....	90
Figura 4.11 – Interface de procura baseada em esboço como exemplo.....	91
Figura 4.12 – Caixa de diálogo com a) imagem binária original e b) informação relativa à imagem.....	92
Figura 4.13 – Menu <i>popup</i> da janela das bases de dados.....	94
Figura 4.14 – Caixa de diálogo para a selecção dos descritores a usar no processo de descrição.....	95
Figura 4.15 – Menu <i>popup</i> da janela das bases de dados.....	98
Figura 4.16 – Caixa de diálogo para selecção dos descritores a usar no processo de procura.....	98
Figura 4.17 – Interface de procura baseada em esboço livre como exemplo.....	100
Figura 4.18 – Interface de procura baseada em esboço sobre uma imagem referência como exemplo.....	101
Figura 4.19 – Janela para apresentação dos resultados da procura.....	102
Figura 5.1 – Descritor baseado no contorno: a) imagem binária que define a forma que se pretende encontrar, b), c) e d) resultados da procura; Descritor baseado na região – e) imagem binária que define a forma que se pretende encontrar, f), g) e h) resultados da procura (pouco semelhantes em termos de contorno).....	110
Figura 5.2 – Exemplo de um emparelhamento exacto: a) <i>Elephant.tif</i> : imagem dada como exemplo para a procura, b) e c) imagens resultantes da procura: imagem reduzida (0.33) e imagem rodada (150°), respectivamente.....	113
Figura 5.3 – Exemplo de um emparelhamento baseado na semelhança perceptiva: a) <i>Elephant.tif</i> : imagem dada como exemplo para a procura, b) e c) imagens resultantes da procura.....	113
Figura 5.4 – Exemplo de uma sequência de movimentos não-rígidos da sequência de teste do MPEG-7, <i>Bream</i>	114
Figura 5.5 – Classe <i>Bat</i> - a) imagem binária de referência; imagens binárias resultantes após aplicação dos factores de mudança de escala b) x2, c) x0.3, d) x0.25, e) x0.2 e f) x0.1.....	115
Figura 5.6 – Classe <i>Bat</i> - a) imagem binária de referência; imagens binárias resultantes após aplicação dos ângulos de rotação b) 9°, c) 36°, d) 45°, e) 90° e f) 150°.....	115
Figura 5.7 – Classe <i>Bat</i> – 6 das 20 imagens binárias pertencentes a uma das 70 classes de imagens binárias disponíveis.....	117
Figura 5.8 – Exemplos de imagens binárias pertencentes a), b), c) e d) à sequência <i>Bream</i> , e e), f), g) e h) à base de dados de criaturas marinhas.....	118
Figura 5.9 – Classe 368 - a) imagem binária de referência; imagens binárias resultantes após aplicação dos factores de mudança de escala b) x3, c) x1.5, d) x0.5, e) x0.33.....	120
Figura 5.10 – Classe 368 - a) imagem binária de referência; imagens binárias resultantes após aplicação dos ângulos de rotação b) 9°, c) 36°, d) 45°, e) 90°, f) 120° e g) 150°.....	120
Figura 5.11 – Classe 616 - a) imagem binária de referência; b), c) imagens binárias resultantes após mudança de escala; d) e e) imagens binárias resultantes após rotação no domínio analógico.....	121
Figura 5.12 – Classe 616 - a) imagem binária de referência; b), c) e d) imagens binárias resultantes após transformações de perspectiva no domínio analógico.....	121
Figura 5.13 – Diagrama de blocos para as 3 etapas do cálculo dos ângulos de curvatura do contorno.....	131
Figura 5.14 – a) Contornos definidos nas arestas dos <i>pixels</i> e b) Vizinhança de ordem 4 do <i>pixel</i> (<i>i,j</i>) [Ferr1995].....	132
Figura 5.15 – <i>Apple</i> - a) Imagem binária; b) contorno correspondente; c) zona ampliada do contorno (marcada em b)).....	133
Figura 5.16 – Exemplo de um objecto ambíguo relativamente ao ponto de começo a usar para o cálculo dos ângulos de curvatura.....	134
Figura 5.17 – <i>Apple</i> - a) imagem binária; e b) identificação dos quatro pontos de começo.....	135
Figura 5.18 – Exemplo da determinação dos ângulos de curvatura, $t(i)$, segundo o algoritmo descrito.....	136

Figura 5.19 – a) forma do objecto; b) menor circunferência que engloba o objecto; c) região interior para um dado raio r ; e d) região exterior para um dado raio r [KiKi1999f]	144
Figura 5.20 – Resultados obtidos para: a) desempenho do descritor <i>CSS</i> em função do número de picos e b) dimensão do descritor em função do número de picos [BoAt2000a]	148
Figura 5.21 – Desempenho do descritor <i>CSS</i> em função do nível de quantificação dos componentes: a) <i>HighestPeak</i> , b) <i>ypeak</i> , c) <i>xpeak</i> e d) <i>Curvature_Vector</i> [BoAt2000a]	150
Figura 5.22 – Desempenho em função da ordem a) momentos de <i>Zernike</i> ; e b) transformada <i>ART</i> [KiKi2000l]	156
Figura 5.23 – Desempenho em função dos níveis de quantificação a) momentos de <i>Zernike</i> ; e b) transformada <i>ART</i> [KiKi2000l]	157
Figura 6.1 – Exemplo de simetria: Cão a) forma original, b) e c) formas simétricas relativamente ao eixo vertical e horizontal, respectivamente	180
Figura 6.2 – Esquema representativo do método de compensação da simetria através do uso de uma descrição de procura analiticamente simétrica	181
Figura 6.3 – Esquema representativo do método de compensação da simetria através do uso de uma descrição de procura gerada no sentido dos ponteiros do relógio	183
Figura 6.4 – Desempenho médio total do descritor ACC para o conjunto de valores de <i>Número_Ângulos</i> definido	187
Figura 6.5 – Desempenho do descritor relativamente às reduções e ampliações existentes na base de dados correspondente à parte A1 do CE-1	188
Figura 6.6 – Exemplo representativo do cálculo da distância D_{ATA}	192
Figura 6.7 – Desempenho do descritor ACC para a distância <i>alpha-trimmed average</i> usando vários valores para o limiar M_L	194
Figura 6.8 – Exemplo ilustrativo da localização dos pontos de começo P_{11} , P_{12} , P_{21} e P_{22} , e das respectivas distâncias, D_{11} , D_{12} , D_{21} e D_{22}	196
Figura 6.9 – Desempenho médio do descritor ACC em função do número de <i>bits</i> para a componente <i>Ângulo_Curvatura</i>	204

Lista de tabelas

Tabela 2.1 – Efeito de vários tipos de transformação sobre os coeficientes calculados pela transformada de Fourier [GoWo1993].....	22
Tabela 2.2 – Resumo das propriedades dos parâmetros de forma baseados no contorno.....	46
Tabela 2.3 – Resumo das propriedades dos parâmetros de forma baseados na região.....	47
Tabela 3.1 – Parâmetros de forma escolhidos para cada um dos mecanismos analisados.....	76
Tabela 3.2 – Propriedades dos módulos de base de dados e de procura dos mecanismos analisados.....	76
Tabela 5.1 – Elementos normativos, não-normativos e metodologias de trabalho para a característica de forma apresentados no seguimento da chamada de propostas do MPEG-7 [VeHa1999].....	110
Tabela 5.2 – Avaliação das propostas a descritores de forma baseados no contorno [VeHa1999].....	111
Tabela 5.3 – Avaliação das propostas a descritores de forma baseados na região [VeHa1999].....	111
Tabela 5.4 – Imagens binárias para as várias partes do <i>core experiment CE-1</i>	119
Tabela 5.5 – Imagens binárias para as várias partes do <i>core experiment CE-2</i>	123
Tabela 5.6 – Descritores de forma que participaram nos <i>core experiments</i> MPEG-7 [MPEG7-DCE1999a].....	124
Tabela 5.7 – Bases de dados a usar para a execução dos <i>core experiments</i>	126
Tabela 5.8 – Resultados obtidos pelos proponentes do descritor CSS de acordo com o método de avaliação do <i>CE-1</i> [Bobe1999b][BoKi1999].....	129
Tabela 5.9 – Resultados obtidos pelos proponentes e outros especialistas para o descritor baseado na transformada <i>wavelet</i> de acordo com o método de avaliação do <i>CE-1</i> [BoKi1999] [MuOh1999d].....	130
Tabela 5.10 – Resultados obtidos no âmbito desta tese para os descritores CSS, baseado na transformada <i>wavelet</i> e ângulos de curvatura do contorno segundo as condições do <i>CE-1</i>	138
Tabela 5.11 – Resultados obtidos para os <i>CE-1</i> e <i>CE-2</i> com o descritor momentos de <i>Zernike</i> [BoKi1999].....	140
Tabela 5.12 – Resultados obtidos para os <i>CE-1</i> e <i>CE-2</i> com o descritor <i>MLEV</i> [BoKi1999][KiKi1999e].....	141
Tabela 5.13 – Desempenho do descritor de forma baseado no contorno, CSS, antes e após a optimização dos limiares E_L e C_L da sua medida de semelhança [Bobe1999c].....	143
Tabela 5.14 – Desempenho dos momentos de <i>Zernike</i> usando como pesos: a) desvio padrão e b) distribuição de Fischer [KiKi1999h][KiKi1999i].....	146
Tabela 5.15 – Número médio de <i>bits</i> usados antes e depois da optimização do descritor CSS [BoAt2000a][HeKa2000].....	151
Tabela 5.16 – Perdas de desempenho do descritor CSS após a optimização da sua dimensão.....	151
Tabela 5.17 – Desempenho do descritor CSS antes e depois da optimização em termos da dimensão do descritor [BoAt2000a][HeKa2000].....	152
Tabela 5.18 – Desempenho para os <i>core experiments CE-1</i> e <i>CE-2</i> : momentos de <i>Zernike</i> versus transformada ART [KiKi2000i].....	154
Tabela 5.19 – Número de momentos/coeficientes e número de <i>bits</i> necessários para a codificação de cada momento de <i>Zernike</i> ou coeficiente ART [KiKi2000i][Yama2000].....	158
Tabela 5.20 – Desempenho dos descritores momentos de <i>Zernike</i> e ART após a quantificação com os valores optimizados [KiKi2000i][Yama2000].....	159
Tabela 5.21 – Perdas de desempenho dos descritores momentos de <i>Zernike</i> e ART após optimização da sua dimensão.....	160
Tabela 5.22 – Desempenho do descritor de forma baseado no contorno, CSS, antes e depois da optimização em relação aos parâmetros globais do descritor e às transformadas efectuadas aos picos [Bobe2000e][BoAt2000b].....	161
Tabela 5.23 – Tabela de quantificação usada para os coeficientes ART [MPEG7-VWD2000b][MPEG7-CDVisual2000].....	166
Tabela 6.1 – Desempenho de procura obtido nesta tese para os descritores CSS, transformada <i>wavelet</i> e ângulos de curvatura do contorno (ACC) nas condições do <i>core experiment CE-1</i>	171
Tabela 6.2 – Número médio de <i>bits</i> usado por cada descrição usando os descritores CSS e ACC.....	172
Tabela 6.3 – Número médio de <i>bits</i> por descrição antes e após a optimização da sintaxe do descritor ACC proposta nesta secção.....	173
Tabela 6.4 – Número médio de <i>bits</i> por descrição antes e após a melhoria do descritor ACC através da inclusão dos parâmetros geométricos, excentricidade e circularidade.....	175
Tabela 6.5 – Desempenho de procura do descritor ACC antes e depois da filtragem baseada nos parâmetros geométricos sem e com influência destes na medida de semelhança final.....	177
Tabela 6.6 – Taxa percentual de formas eliminadas pelos parâmetros geométricos incluídos no descritor ACC.....	178

Tabela 6.7 – Desempenho do descritor ACC antes e depois das alterações à medida de semelhança para a compensação de transformações de simetria conforme os dois métodos anteriormente descritos	184
Tabela 6.8 – Desempenho médio total do descritor ACC para o conjunto de valores de <i>Número_Ângulos</i> definido	187
Tabela 6.9 – Desempenho do descritor ACC relativamente às reduções e ampliações existentes na base de dados correspondente à parte A1 do CE-1	188
Tabela 6.10 – Número médio de <i>bits</i> por descrição antes e após a optimização do descritor ACC relativamente ao valor da componente <i>Número_Ângulos</i>	189
Tabela 6.11: Desempenho de procura para as distâncias de <i>Minkowsky</i> e Euclidiana	191
Tabela 6.12 - Desempenho do descritor ACC para a distância <i>alpha-trimmed average</i> usando vários valores para o limiar M_L	194
Tabela 6.13: Resultados antes e após a optimização do número de vectores de ângulos enviados em cada descrição em função dos limiares de proximidade $ M_L $ e $ D_L $	199
Tabela 6.14– Número de <i>bits</i> usado por descrição ACC, antes e depois da optimização da dimensão do descritor	200
Tabela 6.15 – Desempenho médio de procura do descritor AAC usando os quatro critérios de selecção do ponto de começo propostos	202
Tabela 6.16 – Número médio de <i>bits</i> usado por descrição ACC, antes e depois da optimização da dimensão do descritor	203
Tabela 6.17 – Desempenho médio de procura do descritor ACC em função do número de <i>bits</i> para a componente <i>Ângulo_Curvatura</i>	204
Tabela 6.18 – Número de <i>bits</i> usado por descrição ACC, antes e depois da optimização da dimensão da componente <i>Ângulos_Curvatura</i>	205
Tabela 6.19: Resumo dos resultados obtidos para as alterações efectuadas ao descritor ACC ao longo deste capítulo	207

Capítulo 1

Contexto e objectivos

Os avanços tecnológicos registados nos últimos anos na área da tecnologia audiovisual, tanto em termos de algoritmos como de produtos, deram origem a uma enorme proliferação da informação audiovisual assim como ao seu mais fácil acesso por um vasto número de utilizadores, não só ao nível do consumo como também ao nível da produção. O aparecimento e crescente popularidade das diversas máquinas fotográficas e de vídeo digital, hoje em dia existentes no mercado, armazenando a informação adquirida em formatos digitais normalizados tais como o JPEG e o MPEG-1, foram factores determinantes ao incentivo da produção de informação audiovisual. Estes desenvolvimentos afectam cada um de nós, tornando-nos potenciais produtores de informação audiovisual, capazes de criar conteúdos que a seguir são facilmente distribuídos e publicados, por exemplo através da *Internet*.

Mas se a aquisição, o processamento e a distribuição de conteúdo audiovisual é cada mais fácil, também o deveria ser a sua procura e acesso para que todos pudessem ter a possibilidade de usufruir rapidamente da informação disponibilizada em qualquer parte do mundo. No entanto, a identificação e localização de conteúdos, por procura ou filtragem, tem-se tornado cada vez mais difícil já que os progressos tecnológicos têm dado origem a um crescimento galopante da quantidade de informação audiovisual disponível, sobretudo devido à rápida adaptação dos consumidores às novas tecnologias. Neste contexto, existe cada vez maior número de pessoas que se dedicam exclusivamente a actividades nas áreas dos conteúdos audiovisuais nomeadamente em termos de aquisição, processamento e descrição, de modo a tornar este tipo de conteúdo finalmente mais acessível aos seus consumidores.

Devido à crescente necessidade de encontrar informação audiovisual, já que há cada vez mais actividades onde o uso de informação audiovisual é essencial, têm surgido nos últimos anos, nomeadamente na *Internet*, inúmeros mecanismos de procura (*browsers*) baseados em

descrições textuais do conteúdo ou palavras chave, normalmente designadas por *keywords*. No entanto, a identificação da informação procurada ocorre muitas vezes quase por acaso ou com grandes dificuldade em virtude da grande quantidade de respostas não relevantes à procura em questão, fornecidas pela maior parte destes mecanismos. Ainda que, como é evidente, este facto não seja positivo em termos dos utentes, a enorme aceitação e uso destes mecanismos revela a dimensão das necessidades subjacentes. As actuais limitações dos mecanismos de procura, aceites pelos utentes com maior ou menor resignação, indicam contudo que é necessário melhorar este tipo de tecnologia, oferecendo aos utentes não só soluções mais eficazes mas também mais flexíveis e potentes, nomeadamente em termos da ‘linguagem de descrição e procura’ usada.

Embora até hoje, a descrição dos conteúdos audiovisuais tenha sido essencialmente feita através de informação textual (anotações), manualmente gerada por especialistas na matéria, é óbvio que este sistema não pode continuar a ser usado por duas razões principais:

- A crescente quantidade de informação a descrever/indexar o que implicaria grandes recursos e custos humanos e
- A subjectividade implícita a todos os processos manuais de descrição por parte de uma pessoa que tende a interpretar o conteúdo segundo um dado ponto de vista, por exemplo correspondente a um certo domínio de aplicação, o que vai limitar a gama de solicitações a que um dado conteúdo pode responder por ter sido descrito apenas segundo uma dada perspectiva.

Na verdade, se pedíssemos a 1000 pessoas diferentes para fazer uma descrição do famoso quadro de Picasso, “Guernica”, obteríamos concerteza 1000 descrições diferentes o que prova que qualquer descrição textual é normalmente subjectiva e limitada e reflecte um certo ponto de vista, por exemplo de uma pessoa, de um tipo de aplicação ou de um tipo de uso.

Por outro lado, há limitações intrínsecas associadas às linguagens usadas ou seja o texto não parece ser a forma/linguagem mais potente para descrever informação visual que necessita de outro tipo de linguagem. Por exemplo, se se estiver à procura de uma imagem com céu azul, águas esmeralda e um barco de corrida amarelado que se desloca da esquerda para a direita, o texto não será muito provavelmente o modo mais eficiente de expressar esta procura se se esperar encontrar efectivamente conteúdo com esta características, ainda que não existam hoje grandes alternativas.

Estes factos levaram à aceitação generalizada da necessidade de descrever automaticamente, e de modo tão objectivo quanto possível, a informação audiovisual o que implica o uso de parâmetros de descrição automaticamente extraíveis da informação audiovisual que possam completar os parâmetros manuais, sempre que estes existam. Estes parâmetros estão associados a características audiovisuais próprias do tipo de informação em questão; exemplos destas características são a forma, a cor, a textura e o movimento para o vídeo, e o timbre e o ritmo para o áudio. O uso deste tipo de parâmetros de descrição audiovisual apresenta as seguintes vantagens:

- **Automáticos** – Sendo a sua extracção automática, compete às máquinas e não aos especialistas resolver o problema da grande quantidade de informação audiovisual a descrever;

- **Objectivos** – Sendo automaticamente extraíveis e logo objectivos, permitem superar as limitações de subjectividade e especificidade associadas às descrições manuais;
- **Adaptados** – Usando uma linguagem mais próxima da informação que pretendem representar, permitem formular procuras de modo mais adaptado aos conteúdos em questão.

Uma vez que a descrição de conteúdo audiovisual é útil em inúmeros domínios de aplicação, e depende sempre fortemente das propriedades e requisitos da aplicação em causa, é importante que estejam à disposição um conjunto de ferramentas de descrição suficientemente flexível e potente para satisfazer as necessidades dos vários tipos de aplicação. Assim, podem distinguir-se alguns tipos principais de descrição e de características de descrição:

- **Descrição baseada no conteúdo** – Descrição baseada na informação audiovisual e que, como tal, pode ser dela extraída de forma automática, semi-automática ou manual; um exemplo pode ser a descrição de um jogo de futebol, p.e. codificado em MPEG-2, através de parâmetros textuais como tipo de conteúdo – futebol, ou através de parâmetros automáticos como a cor, a textura, a forma, o movimento e a amplitude do som;
 - ♦ **Características de descrição de baixo-nível** – Características objectivas de extracção automática, tipicamente sem qualquer carga semântica associada; exemplos: a forma, a cor, a textura e o movimento para imagens e vídeo e o ritmo para áudio;
 - ♦ **Características de descrição de alto-nível** – Características associadas a um maior nível de abstracção (semântico), cuja extracção implica normalmente a intervenção humana, pelo que são obtidas de forma manual ou semi-automática; exemplos: classificação de conteúdo, p.e. desporto, notícias, etc, detecção e reconhecimento de objectos, de pessoas, de personagens e de actividades.
- **Descrição não baseada no conteúdo** – Descrição que não pode ser extraída directamente do conteúdo, nem automaticamente, nem manualmente, mas que é útil e pode ser conhecida pelo criador das descrições através de outros meios; este tipo de descrição é normalmente textual e gerada manualmente; exemplos: título do filme ou música, data e local de produção, nome do autor e do produtor, condições de acesso, direitos de autor, apontadores para outro conteúdo de interesse e relacionado com o conteúdo principal em questão.

Assim, um mecanismo de descrição que integre ferramentas de descrição associadas aos vários tipos de descrição acima referidos, permitirá gerar descrições com propriedades muito diferentes adaptadas às necessidades da aplicação em causa, por exemplo em termos de nível de abstracção, detalhe, etc.

Embora a descrição completamente automática se encontre ainda essencialmente limitada a características com menor significado semântico, como a forma, a cor, a textura e o movimento para imagens e vídeo e o ritmo para o áudio, muito tem sido feito e continua a ser feito para que se consigam alcançar maiores níveis de abstracção semântica de forma

automática. O mapeamento automático de parâmetros de baixo-nível de abstracção como a cor, a forma e o movimento em conceitos mais abstractos é uma das áreas de investigação actualmente mais interessantes e efervescentes. Na verdade, muitos utentes procuram, por exemplo, imagens associados a conceitos como esperança, vigor ou resistência, como por exemplo no mundo da publicidade, o que, como se imagina, cria desafios enormes em termos da capacidade de mapear parâmetros de baixo-nível em parâmetros de alto-nível. Por outro lado, existem também vários projectos visando a identificação automática de certas acções em certos contextos, por exemplo um golo num jogo de futebol ou um ‘cesto’ num jogo de *basket*.

As crescentes necessidades de encontrar, de modo eficaz e eficiente, informação audiovisual, o aumento da informação disponível *on-line*, e os enormes desenvolvimentos tecnológicos na área da descrição de informação audiovisual, têm levado ao aparecimento cada vez mais frequente de mecanismos de descrição e procura com capacidades de descrição não-textuais. A maturidade das tecnologias envolvidas e a importância das aplicações em questão levou a que, recentemente, um organismo de normalização com a credibilidade da ISO (*International Standards Organization*) através do seu grupo de trabalho MPEG (*Moving Picture Experts Group*) tenha decidido considerar este problema, como se descreve de seguida.

1.1 A norma MPEG-7

Tendo em vista a solução do problema da descrição de informação audiovisual, o grupo MPEG, *Motion Picture Experts Group*, decidiu lançar, em 1996, um projecto denominado *Multimedia Content Description Interface*, mais conhecido como MPEG-7 [MPEG-7-Cont1998]. Este projecto visa especificar um conjunto de ferramentas e métodos para a descrição de vários tipos de informação audiovisual, e.g. imagens estáticas, vídeo, áudio, gráficos, modelos 3D e áudio sintético, independentemente do seu formato, e.g. digital ou analógico, ou meio de armazenamento, e.g. papel, filme ou cassete. Entre os formatos digitais de vídeo mais importantes estão, sem dúvida, os formatos MPEG-1, -2 e -4; a norma MPEG-4 tem aliás uma relação mais íntima com a norma MPEG-7 por serem as únicas que adoptaram um modelo de representação baseado em objectos, como se detalhará mais à frente.

Atendendo a que o grupo MPEG tem sido um dos grandes responsáveis pelo aumento da produção e distribuição de conteúdo audiovisual, tanto ao nível dos mercados profissionais como dos consumidores domésticos, devido aos sucessos das normas MPEG-1, -2, -4, é natural que agora também este grupo de normalização venha ajudar a resolver o problema da descrição desta informação de forma a possibilitar o seu fácil acesso. O lançamento do projecto MPEG-7 representou para o grupo MPEG um enorme desafio uma vez que muitos dos conhecimentos técnicos necessários para o sucesso deste projecto não existiam no grupo, até então mais dedicado às questões da codificação, tendo sido necessário recrutar especialistas nas áreas técnicas relevantes.

Como já ficou provado várias vezes e em vários domínios, a existência de uma norma é absolutamente essencial para a difusão em larga escala de qualquer tecnologia audiovisual e logo das aplicações e serviços a ela associados. A existência de uma norma mundial para a descrição de conteúdos audiovisuais, qualquer que seja o seu suporte, vai permitir um nível de interoperabilidade impossível de alcançar de outro modo, com a mesma informação a ser acedida por mais mecanismos de procura e cada mecanismo de procura a aceder a mais

informação e logo a satisfazer mais e de melhor forma as necessidades dos utentes [PeKo1998]. Assim ao aumentar-se o nível de interoperabilidade entre os inúmeros mecanismos de descrição e procura, vai também permitir-se a produção de conteúdos com menores custos e a possibilidade de fazer explodir novos serviços baseados na norma, como aliás já aconteceu de forma clara com as normas MPEG-1 e MPEG-2 para vídeo e áudio.

O MPEG-7, como todos os outros membros da família MPEG (MPEG-1, -2, -4), especificará métodos de representação da informação audiovisual de forma a satisfazer um conjunto relevante de requisitos. No caso do MPEG-7, estes requisitos estão relacionados com a identificação (descrição e procura) de conteúdo audiovisual [MPEG7-Req1999]. A ênfase atribuída pelo MPEG-7 às características audiovisuais não elimina a importância da descrição textual da informação audiovisual que continuará a ser usada em conjunto com outras formas de descrição. Alguns exemplos através dos quais se pode constatar a importância do texto na descrição de conteúdo audiovisual é na descrição de nomes e lugares, e mesmo na anotação individual dos conteúdos. No entanto, as descrições textuais trazem também questões delicadas, nomeadamente associadas à dependência linguística [MPEG7-Cont1999]. Por outro lado, dada a sua vocação audiovisual, a norma MPEG-7 não especificará novas ferramentas para a descrição de documentos textuais mas adoptará as soluções já existentes como, por exemplo, HTML, SGML, RDF, etc.

1.1.1 MPEG-7: os elementos normativos

Como aconteceu para as outras normas MPEG, o MPEG-7 normalizará um número mínimo possível de ferramentas de descrição para garantir interoperabilidade. A não normalização de tecnologia em áreas não essenciais à interoperabilidade é extremamente importante para o sucesso e tempo de vida de uma norma uma vez que permite [PeKo1998]:

- Integrar de modo compatível todos os desenvolvimentos tecnológicos, posteriormente obtidos, para as partes não-normativas ou seja a norma continua a evoluir tecnologicamente sem qualquer impacto negativo em termos de interoperabilidade;
- Incentivar a competição entre os vários fabricantes uma vez que os produtos compatíveis com uma dada norma não serão rigidamente semelhantes uns aos outros, havendo vários blocos não normativos e com um impacto determinante em termos do desempenho final onde os fabricantes podem competir;
- Adaptar a aplicação da norma às propriedades específicas da aplicação e conteúdos em questão, melhorando o seu desempenho.

Estas vantagens traduzem-se essencialmente em melhores e mais baratos mecanismos de geração e consumo das descrições, condições básicas para o sucesso junto dos utentes.

Atendendo a que hoje em dia, a maioria dos mecanismos de descrição e procura usa soluções específicas que impedem a sua interoperabilidade, é essencial a normalização de um conjunto básico de ferramentas de descrição audiovisual. Na verdade, mesmo quando os parâmetros de descrição usados por dois mecanismos são iguais, por exemplo para a cor, a interoperabilidade entre eles exige mais do que isso ou seja exige a adopção das mesmas

sintaxe e semântica para as descrições, para que as descrições possam ser totalmente compreensíveis para os mecanismos em questão.

É neste contexto que a norma MPEG-7 surge como a mais importante resposta ao desafio de normalização na área da descrição de informação audiovisual. Para responder a este desafio, a norma MPEG-7 normalizará cinco tipos de ferramentas de descrição, rigorosamente definidos mais à frente [MPEG7-Req1999]; mais brevemente, estas ferramentas são:

- Descritores representando as várias características de descrição, por exemplo cor, forma e movimento;
- Esquemas de descrição dando estrutura aos descritores, por exemplo estrutura no espaço e no tempo;
- Técnicas de codificação para as descrições permitindo nomeadamente a codificação eficiente e resiliente das descrições; e
- Linguagem para definir novos esquemas de descrição e (possivelmente) descritores, denominada por *Description Definition Language* (DDL), o que permitirá a extensão das capacidades da norma, mesmo depois da sua especificação;
- Ferramentas de sistema que visam, por exemplo, a sincronização e multiplexagem de descrições e conteúdos descritos.

Estes cinco tipos de ferramentas de descrição são consideradas os elementos normativos da norma, o que significa que ao serem implementados terão que estar de acordo com as especificações impostas pela norma, de modo a garantirem a interoperabilidade dos mecanismos de descrição e procura. Consequentemente, e apesar da sua importância fundamental em termos de desempenho, as ferramentas de extracção das características, as formas de formulação da procura (*querying*), as medidas de semelhança e, em geral, a organização das descrições, p.e. numa base de dados, não são normalizadas uma vez que isso não é necessário para a interoperabilidade dos mecanismos.

1.1.2 A relação com a norma MPEG-4

Para além do uso de características audiovisuais, uma das principais propriedades da norma MPEG-7 é a sua capacidade de descrever, de forma independente, elementos (objectos) na cena audiovisual e não só a cena na sua totalidade. Esta propriedade resulta da adopção de um modelo de representação da informação audiovisual baseado no conceito de objecto onde, por exemplo, uma cena visual passa a ser representada como uma composição de objectos independentes e já não como um sequência de tramas rectangulares. Este novo modelo de representação foi pela primeira vez adoptado pela norma MPEG-4 [MPEG4-1999] que permite codificar uma cena audiovisual como uma composição de objectos visuais e de áudio, oferecendo deste modo novas funcionalidades em termos de eficiência, interactividade, personalização e acesso universal.

A norma MPEG-4 é uma norma com várias partes, p.e. Áudio, Vídeo, Sistemas, que visa satisfazer as necessidades dos vários protagonistas da cadeia de produção e consumo

audiovisual em termos da codificação independente de vários tipos de objectos audiovisuais como, por exemplo, objectos naturais e sintéticos, 2D e 3D, música e voz, etc., que podem ser individualmente processados, acedidos e manipulados. Entre outras, a norma MPEG-4 permite oferecer as seguintes funcionalidades [Pere1998]:

- **Integração de conteúdos naturais e sintéticos** – Codificação de objectos com origem natural ou sintética a fim de serem compostos numa mesma cena audiovisual.
- **Interacção** – Possibilidade de interagir com os vários objectos na cena uma vez que estes são codificados de modo independente e como tal podem ser independentemente acedidos e manipulados.
- **Personalização** – Possibilidade de influenciar qual e o modo como o conteúdo audiovisual é composto, determinando a personalização do conteúdo e do *layout* do ecrã. Exemplo: modificar o conteúdo retirando, adicionando ou alterando a posição de objectos.
- **Re-utilização de conteúdos** – Reutilização de objectos audiovisuais, naturais ou sintéticos, de uma cena para outras cenas. Exemplo: reutilização de uma personagem, pertencente a uma sequência de vídeo, para a composição de uma nova cena de vídeo.
- **Hyperlinking** – Possibilidade de associar a cada objecto ligações para outro conteúdo relevante.

A normalização das técnicas que viabilizam as funcionalidades acima mencionadas, já feita pelo MPEG-4, deverá vir a ter um grande impacto na criação de novas formas de conteúdo, nomeadamente conteúdos integrando objectos de vídeo com forma arbitrária e não apenas com forma rectangular.

Atendendo à forte necessidade de oferecer rapidamente algumas capacidades em termos de descrição da informação audiovisual, a norma MPEG-4 especificou também um conjunto de ferramentas de descrição textual designadas por *Object Content Information (OCI)*. Para além disso, prevendo-se já o aparecimento da norma MPEG-7, a norma MPEG-4 permite já a associação a cada objecto codificado com o MPEG-4 de um fluxo binário correspondente à descrição em MPEG-7 do objecto em questão.

Desta forma, estabeleceu-se entre as normas MPEG-4 e MPEG-7 uma relação mais próxima, inexistente para as outras normas que contudo podem, sem qualquer problema, ver o seu conteúdo descrito de acordo com a norma MPEG-7.

1.1.3 MPEG-7: as aplicações

Como já se disse, o uso da norma MPEG-7 é independente do modo em que o conteúdo se encontra disponível, nomeadamente do seu formato ou suporte de armazenamento. Por exemplo, a informação de vídeo poderá estar disponível digitalmente codificada como MPEG-1, -2 ou -4, em qualquer outro formato digital, ou mesmo em formato analógico, em filme ou *cassete*. As imagens poderão estar codificadas em JPEG ou então existir apenas em

papel. É claro que será contudo difícil extrair descrições de baixo-nível para conteúdo que não esteja digitalmente disponível.

O MPEG-7 pretende ser uma norma de descrição genérica, considerando o maior número possível de aplicações e requisitos associados. Para isso, deverá integrar um conjunto suficientemente vasto de ferramentas de descrição, seguindo contudo sempre o princípio de “uma funcionalidade, uma ferramenta” ou seja recusando a adopção de várias ferramentas com o mesmo objectivo. Este princípio é essencial para minimizar a complexidade dos mecanismos de descrição e procura, maximizando as suas capacidades e interoperabilidade.

Assim, o MPEG-7 considera vários tipos de aplicações, nomeadamente em tempo real (quando a descrição é criada simultaneamente ao conteúdo) e tempo não real, tendo organizado as aplicações em três categorias, como se apresenta de seguida [MPEG-7App1998][PeKo1998]:

- **Aplicações do tipo *Pull*** – Aplicações em que o utilizador que pretende encontrar informação audiovisual indica de algum modo as propriedades do conteúdo que pretende encontrar ou fornece conteúdo como exemplo; essas propriedades, transformadas em parâmetros de descrição, descrições ou em conteúdo exemplo são enviadas para comparação remota com as descrições de conteúdo disponíveis, p.e. em inúmeras bases de dados na *Internet*. Os exemplos mais típicos deste tipo de aplicações são os mecanismos de descrição e procura disponíveis na *Internet* como, por exemplo, de bibliotecas musicais comerciais, de procura de bases de dados de vídeo e de registos e procuras em base de dados de logótipos;
- **Aplicações do tipo *Push*** – Aplicações em que o utilizador que pretende encontrar informação audiovisual cria um filtro de descrições que vai seleccionar entre as inúmeras descrições que lhe chegam aquelas que lhe são úteis por corresponderem a conteúdo que lhe interessa; exemplos deste tipo de aplicação são a selecção de canais de televisão em radiodifusão, via cabo ou satélite;
- **Aplicações profissionais e de controlo** – Aplicações específicas que abrangem domínios tão variados como a biomedicina, telecompras, educação e vigilância.

As aplicações que se encontram listadas de seguida são as principais aplicações consideradas pelo MPEG-7 [MPEG7-App1999]; esta lista apresenta apenas alguns exemplos já que muitas outras aplicações poderão vir a beneficiar da norma MPEG-7:

- | | |
|-------------------------------------|--|
| • Educação | • Compras |
| • Jornalismo | • Publicidade |
| • Turismo | • Vigilância |
| • Biomédicas | • Entretenimento |
| • Sistemas de Informação Geográfica | • Desenho de interiores e arquitectura |
| • Têxteis | • Bases de dados de vídeo e áudio |

Uma vez que o MPEG-7 pretende abranger todos os domínios de aplicação e não apenas algumas aplicações específicas, torna-se essencial que o MPEG-7 adopte um conjunto de ferramentas genéricas que permitam satisfazer horizontalmente algumas necessidades de descrição mas também algumas ferramentas mais específicas. Esta variedade dará ao MPEG-7 a capacidade de descrever conteúdo audiovisual para o maior número possível de aplicações, mesmo que as aplicações mais importantes tenham um maior peso no seu desenvolvimento. É de salientar que desta forma o mesmo conteúdo audiovisual poderá ser descrito de várias maneiras diferentes, usando-se parâmetros audiovisuais e esquemas de descrição diferentes, consoante o domínio de aplicação a que a descrição se destina. Consequentemente, caberá a cada criador de descrições a tarefa de escolher os melhores parâmetros audiovisuais e o nível de detalhe correspondente, para que todas as descrições geradas para um mesmo conteúdo sejam igualmente correctas e válidas. Todas estas descrições serão segundo a norma MPEG-7 e, como tal, compreensíveis por todos os mecanismos desenvolvidos de acordo com esta norma.

1.2 Terminologia relevante

Devido à sua importância e de forma a possibilitar mais facilmente a compreensão desta tese, definem-se de seguida os termos mais relevantes e mais frequentemente usados. Para evitar inconsistências, usa-se sempre que possível a terminologia definida pelo MPEG-7 em [MPEG7-Req1999]. Assim, a descrição de informação audiovisual envolve os seguintes conceitos principais (entre parêntesis indica-se o termo usado pelo MPEG-7):

- **Informação audiovisual (*data*)** – Conteúdo audiovisual (que pode também ser apenas visual ou apenas áudio) que vai ser descrito, independentemente da sua forma de armazenamento, codificação, visualização e transmissão. Exemplos: um fluxo binário MPEG-4, um filme numa cassete de vídeo, as canções num CD.
- **Característica (*feature*)** – Qualquer propriedade relevante da informação audiovisual e que pode significar algo para alguém. Exemplos: a forma de um objecto, o título de um filme, o ritmo de uma música.
- **Parâmetro** – Representação de uma dada característica usando uma dada técnica de análise; um parâmetro define apenas a semântica da representação da característica. Exemplos: a área, a circularidade, os coeficientes de Fourier para a forma de um objecto e um histograma de cor para a sua cor.
- **Descritor (*descriptor*)** – Representação completa de uma dada característica; um descritor define a sintaxe e a semântica da representação da característica, ou seja, acrescenta ao parâmetro a definição da sintaxe. Exemplos: descritores normalizados pelo MPEG-7.
- **Valor do descritor** – Instanciação de cada um dos campos de um descritor (que pode ser composto por um ou mais campos sintácticos) para a informação audiovisual em questão.
- **Esquema de descrição (*description scheme*)** – Especifica a estrutura e a semântica das relações existentes entre os seus componentes, que podem ser tanto descritores como esquemas de descrição. Exemplo: um filme temporalmente

estruturado em cenas com descritores textuais associados ao filme, p.e. nome e realizador, e descritores de cor, movimento e áudio associados a cada cena.

- **Descrição (*description*)** – Consiste num esquema de descrição (que pode ser a combinação de vários outros) e no conjunto de valores dos descritores que descrevem a informação audiovisual em questão.
- **Linguagem de Definição de Descrição (*description definition language*)** – Permite definir novos e melhorar esquemas de descrição e (possivelmente) descritores¹ existentes. Este elemento permitirá a extensão das capacidades da norma, mesmo depois da sua especificação.

Dos termos acima definidos, apenas ‘parâmetro’ não faz parte da terminologia MPEG-7. Este termo foi introduzido para melhor fazer a transição entre ‘característica’, essencialmente conceptual, e ‘descritor’ já essencialmente associado a uma dada especificação de sintaxe e semântica.

Outros conceitos essenciais no contexto desta tese que, como se verá, incidirá na descrição de objectos de vídeo com base na característica de forma são:

- **Região** – Conjunto conexo de todos os *pixels* com a mesma etiqueta de identificação.
- **Objecto** – Região ou conjunto de regiões com um significado especial, nomeadamente semântico. A norma MPEG-4 codifica de forma independente os vários objectos de uma cena que podem assim ser individualmente acedidos e processados.
- **Buraco de uma região** – Conjunto conexo de todos os *pixels* que se encontram totalmente envolvidos por uma região mas não fazem parte dela.
- **Contorno** – Conjunto de todos os *pixels* que pertencem a uma região ou objecto e que têm como vizinhos, segundo um dado tipo de vizinhança, pelo menos um pixel não pertencente a essa região ou objecto. O contorno define a forma de uma região ou objecto.
- **Shapel** – Elemento básico do suporte de um objecto ou região correspondendo aos *pixels* onde a textura do objecto tem valor não nulo.

1.3 Objectivos e organização desta tese

Como se viu atrás, algumas das tendências mais importantes de hoje em termos de tecnologia audiovisual como a interactividade, a personalização de conteúdos e o acesso universal, estão intimamente ligadas a um modelo de representação da informação audiovisual baseado no conceito de ‘objecto’. Em termos visuais, um objecto é representado através da sua informação de textura (luminância e crominância) e através da sua informação de forma. A

¹ Até Janeiro de 2001, a capacidade de definir novos descritores não era possível por não haver meios para definir a sua semântica.

informação de forma aparece pois como a grande novidade associada à representação, codificação ou descrição, de objectos uma vez que a informação de textura já era usada nos modelos de representação anteriores baseados em tramas rectangulares.

Devido à grande importância do conceito de ‘objecto’ e de ‘forma’ no mundo visual e também à sua relativa novidade, escolheu-se como tema central desta tese o estudo da descrição de formas visuais, em íntima relação com o que está a ser desenvolvido nesta área no contexto da norma MPEG-7.

Assim os grandes objectivos definidos e alcançados com esta tese são:

1. Revisão bibliográfica e análise comparativa dos parâmetros de descrição de forma disponíveis;
2. Revisão bibliográfica e análise comparativa dos principais mecanismos de procura disponíveis e que usam a informação de forma;
3. Desenvolvimento de um mecanismo de descrição e procura com ênfase na característica de forma;
4. Avaliação dos descritores de forma propostos e adoptados pelo MPEG-7 para um conjunto relevante de material de teste e em condições rigorosamente definidas; e
5. Implementação em *software* no contexto do ambiente de experimentação do MPEG-7 de alguns descritores de forma;
6. Melhoria substancial do descritor ângulos de curvatura do contorno tanto em termos de descrição como em termos de medidas de semelhança.

Para alcançar os objectivos estabelecidos, esta tese encontra-se dividida em sete capítulos, organizados do modo que se descreve de seguida.

O Capítulo 2 apresenta uma revisão bibliográfica dos parâmetros de forma mais relevantes disponíveis na literatura, nomeadamente aqueles propostos ao MPEG-7 no seguimento da chamada de propostas feita pelo MPEG em Outubro de 1998. Neste capítulo começa-se por definir os parâmetros de forma mais relevantes assim como algumas medidas de semelhança associadas e, finalmente, avaliam-se em conjunto todos os parâmetros e medidas de semelhança anteriormente apresentados.

O Capítulo 3 debruça-se sobre os principais mecanismos de descrição e procura disponíveis na *Internet* e que usam apenas características visuais de baixo-nível ou seja características automaticamente extraíveis da informação visual. Após a proposta de uma arquitectura genérica para estes mecanismos, segue-se a descrição de alguns dos mecanismos mais interessantes disponíveis na *Internet*, concluindo-se com uma análise comparativa entre a arquitectura genérica proposta e a arquitectura dos mecanismos descritos. Uma vez que este trabalho se debruça especialmente sobre a característica de forma, estes mecanismos serão mais detalhadamente analisados e comparados em tudo o que diz respeito à característica de forma.

O Capítulo 4 descreve o mecanismo de descrição e procura a desenvolver no âmbito desta tese, ou seja, um mecanismo de descrição e procura baseado principalmente na característica de forma dos objectos de vídeo. Esta aplicação permitirá a avaliação e comparação do desempenho dos descritores de forma do MPEG-7 assim como de outros descritores alternativos de descrição de forma. No entanto, a grande novidade desta aplicação relativamente a outros mecanismos existentes na *Internet* recai precisamente sobre o facto de se usarem os descritores normalizados pelo MPEG-7, garantindo-se deste modo a

interoperabilidade com outros mecanismos em termos das descrições de forma se apenas descritores MPEG-7 forem usados.

O Capítulo 5 apresenta pormenorizadamente, e de acordo com o processo de normalização que o MPEG adoptou para o desenvolvimento da norma MPEG-7, os passos dados no desenvolvimento desta norma com especial ênfase nos parâmetros de forma propostos ao MPEG-7 como candidatos a descritores. Será também analisado e comparado o desempenho dos descritores de forma considerados na fase de competição, reproduzindo todos os *core experiments* de forma definidos no contexto do MPEG-7 ou seja obtendo resultados para o desempenho de todos os descritores em questão e implementando em *software* os descritores não disponíveis no ambiente de experimentação do MPEG-7 (atendendo a que alguns dos descritores propostos não chegaram a ser integrados no *software* do MPEG-7). Por último, descrevem-se e analisam-se a sintaxe e a semântica referentes a cada descritor adoptado pelo MPEG-7.

O Capítulo 6 apresenta e descreve como consequência das conclusões tiradas no capítulo 5 sobre os parâmetros de forma propostos e/ou adoptados pelo MPEG-7, o trabalho realizado no sentido de melhorar o parâmetro de forma baseado nos ângulos de curvatura, tentando levá-lo a um desempenho semelhante aos descritores incluídos na norma MPEG-7. Este parâmetro implementado de acordo com o método de extracção sugerido pelos seus proponentes, mostrou um desempenho médio que se encontrava longe dos desempenhos médios obtidos pelos seus ‘adversários’. Consequentemente, e por este descritor aparentar ter claras ‘debilidades’ relativamente aos descritores em competição no MPEG-7, investiu-se na sua melhoria, quer em termos de descrição, quer em termos de medidas de semelhança, tendo sempre presentes as propriedades dos descritores MPEG-7, que os levaram a superar as debilidades que o descritor ângulos de curvatura do contorno ainda apresentava.

Finalmente, no Capítulo 7, tecem-se considerações finais e conclusões sobre as dificuldades inerentes à descrição e procura de vídeo baseadas na característica de forma assim como à aplicação desenvolvida que descreve e procura conteúdo visual baseada nos descritores de forma e cor do MPEG-7 e ainda no descritor ângulos de curvatura do contorno implementado no âmbito desta tese. Com base nas considerações e conclusões finais definir-se-á também o que se pensa poder propor como trabalho futuro.

Capítulo 2

Parâmetros de forma

O enorme aumento de informação audiovisual no mundo de hoje, devido à maior facilidade em adquirir, processar, armazenar e transmitir este tipo de informação, veio acentuar a necessidade e motivar a definição de ferramentas de descrição que auxiliem a descrição e a procura eficientes e rápidas de conteúdos audiovisuais. Assim, o estudo de características visuais para a descrição do conteúdo como a forma, cor, movimento e textura, intrinsecamente adequadas à informação visual merece, devido à sua riqueza e potencialidade, um cuidado particular. No contexto desta tese, a característica de forma assume, como já se explicou, o papel central pelo que é normal que se lhe dedique especial atenção em termos de revisão bibliográfica das ferramentas de descrição disponíveis.

Este capítulo dedicar-se-á à revisão bibliográfica dos parâmetros de forma mais relevantes disponíveis na literatura, nomeadamente aqueles propostos ao MPEG-7 no seguimento da chamada de propostas feita pelo MPEG no início dos trabalhos referentes a esta norma. Assim, começar-se-ão por definir os principais parâmetros de forma assim como algumas medidas de semelhança relevantes e, finalmente, avaliar-se-ão sumariamente e em conjunto todos os parâmetros e medidas de semelhança anteriormente apresentados. Ainda que ao longo deste capítulo se refiram muitos exemplos no contexto de aplicações do tipo *pull*, por exemplo procura em bases de dados, a utilidade das ferramentas de descrição apresentadas não se limita a esse tipo de aplicações.

2.1 Requisitos

Os parâmetros de forma aqui apresentados são todos parâmetros numéricos ou seja parâmetros cujo modo de instanciação é um valor numérico. Significa isto que estes

parâmetros usam uma descrição matemática das propriedades da forma, que permite assim representar automaticamente a forma de um objecto ou região. Para que estes parâmetros possam ser usados correctamente durante a descrição e a procura de objectos semelhantes em termos de forma, é necessário que satisfaçam os seguintes requisitos:

- **Eficácia na procura** – O utilizador deve encontrar aquilo que procura e não outra informação irrelevante para a procura em questão;
- **Eficiência na procura** – O utilizador deve encontrar o que procura rapidamente;
- **Procura baseada na semelhança** – As descrições geradas permitem que o resultado final de uma procura seja ordenado por grau de semelhança com a informação de definição da procura fornecida;
- **Descrições sincronizadas ou não com o conteúdo** – As descrições devem poder estar sincronizadas ou não com o conteúdo; mais especificamente devem poder ser ou não ser geradas em simultâneo com a informação audiovisual;
- **Descrições audiovisuais distribuídas** – A procura de informação audiovisual deve poder ser feita simultânea e transparentemente em qualquer base de dados que contenha descrições compatíveis;
- **Procuras interactivas** – As descrições devem poder ser utilizadas no contexto de mecanismos que permitem algum tipo de interacção por parte do utilizador;
- **Procuras por navegação** – As descrições devem permitir procuras em que o utilizador pode visualizar o conteúdo disponível, antes de efectuar a procura, de modo a ultrapassar o seu desconhecimento relativamente à estrutura da informação ou ao tipo de informação, ou mesmo para clarificar possíveis dúvidas relativamente ao que procura; e
- **Insensibilidade a transformações geométricas** – As descrições devem ser insensíveis a rotações, translações, mudanças de escala, ponto de começo, etc.

Dos requisitos acima listados, o da insensibilidade a transformações geométricas será aquele a que este capítulo dará mais ênfase por estar mais intrinsecamente relacionado com os parâmetros de forma. Por esse motivo, definem-se de seguida as quatro principais transformações geométricas a ter em conta pelos parâmetros de forma a analisar:

- **Insensibilidade à rotação** – Invariância do parâmetro de forma relativamente ao movimento de um objecto ou região que se desloca em volta de um eixo (material ou não), durante o qual cada ponto do objecto ou região tem a mesma velocidade angular;
- **Insensibilidade à translação** – Invariância do parâmetro de forma relativamente ao movimento de todos os pontos em simultâneo do objecto ou região, caracterizado por serem paralelas a posição inicial e a posição final da recta definida por dois quaisquer desses pontos;
- **Insensibilidade à mudança de escala** – Invariância do parâmetro de forma ao aumento ou redução do objecto ou região em relação às suas dimensões originais;

- **Insensibilidade ao ponto de começo** – Invariância do parâmetro de forma relativamente ao ponto de começo da análise da forma do objecto ou região.

Tendo em conta os requisitos especificados e definidos anteriormente, analisaram-se dois tipos de parâmetros de forma nas secções que seguem [MPEG-7XM1999]:

1. **Parâmetros de forma baseados no contorno** – Parâmetros que descrevem a forma de um objecto tendo em conta apenas os seus *shapels* mais exteriores, ou seja, o contorno fechado de um objecto simples ou região;
2. **Parâmetros de forma baseados em regiões** – Parâmetros que descrevem formas simples e complexas de um objecto tendo em conta todos os seus *shapels*.

Uma vez extraídos e armazenados os parâmetros de forma dando origem a descrições de forma, é necessário poder aceder-lhes para que eventualmente se possa encontrar e visualizar o conteúdo que estes representam. Deste modo, para que este conteúdo possa ser encontrado quando uma procura relevante for definida, é necessário dispor de medidas de semelhança que eficazmente meçam o significado de semelhança de forma tal como o utilizador do mecanismo de procura o entende. Como seria de esperar, as medidas de semelhança adequadas estão fortemente relacionadas com os parâmetros de descrição de forma utilizados não se comparando parâmetros mas obviamente instâncias do mesmo parâmetro. Assim, apresentar-se-ão nas sub-secções que se seguem algumas medidas de semelhança, mais comuns, como a distância Euclidiana e a distância de *Minkowsky*, entre outras.

2.2 Parâmetros de forma baseados no contorno

Os parâmetros de forma baseados no contorno descrevem a forma de um objecto tendo em conta apenas os seus *shapels* mais exteriores, ou seja, o contorno fechado de um objecto simples ou região. A Figura 2.1 mostra um exemplo de um contorno fechado a ser descrito por este tipo de parâmetros.

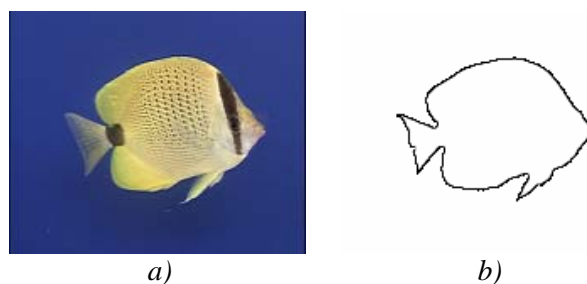


Figura 2.1 – *Bream* (trama1): a) Imagem com o objecto; b) contorno do objecto

Neste contexto, apresentar-se-ão de seguida os principais parâmetros de forma baseados no contorno disponíveis na literatura, incluindo aqueles propostos ao MPEG-7. Para melhor compreensão, os parâmetros de forma baseados no contorno serão organizados segundo as suas propriedades do seguinte modo:

- **Parâmetros geométricos** – Parâmetros que representam a forma do objecto simples usando propriedades geométricas do seu contorno. Exemplos: perímetro, corda máxima, circularidade, convexidade e excentricidade;

- **Parâmetros baseados em transformadas** – Parâmetros que representam a forma de um objecto simples ou região usando coeficientes calculados a partir de uma dada transformada; exemplos: transformada de *Fourier e wavelets*;
- **Parâmetros baseados em momentos** – Parâmetros que representam a forma de um objecto simples ou região usando um conjunto de valores estatísticos; exemplos: momentos geométricos ou invariantes;
- **Parâmetros baseados em contornos normalizados** – Parâmetros que representam a forma de um objecto simples ou região usando o seu contorno normalizado, insensível a transformações geométricas e ao número de pontos que o definem;
- **Parâmetros baseados nos ângulos de curvatura da forma** – Parâmetros que representam a forma de um objecto simples ou região através de um conjunto de ângulos de curvatura do seu contorno.
- **Parâmetros baseados numa imagem curvature scale space (CSS)** – Parâmetros que representam a forma de um objecto simples ou região com vários níveis de detalhe de acordo com os pontos de inflexão (convexos ou côncavos) do contorno fechado.

2.2.1 Parâmetros geométricos

Os parâmetros de forma geométricos inerentes ao contorno fechado de um objecto simples ou região, são parâmetros de forma que permitem representar o objecto simples ou região em relação à geometria do seu contorno fechado.

2.2.1.1 Cálculo dos parâmetros geométricos de forma

Alguns dos parâmetros de forma geométricos existentes e descritos de seguida, são o perímetro, a corda máxima, a circularidade, a convexidade e a excentricidade [Russ1995].

▪ Perímetro

O perímetro de um objecto simples ou região corresponde ao número total de *pixels* que constituem o contorno fechado do objecto simples ou região em vizinhança de ordem 8. A expressão (2.1) define o perímetro, P , como sendo:

$$P = \sum_i \sum_j I(i, j) \quad (2.1)$$

onde $I(i, j)$ é o suporte binário do contorno fechado do objecto e (i, j) são as coordenadas vertical e horizontal dos *shapels*, respectivamente. O contorno do objecto é obtido a partir da definição de vizinhos diagonais, horizontais e verticais (vizinhança de ordem 8). Portanto, todos os *pixels* constituintes do objecto que tiverem como *pixel* vizinho nas direcções

horizontal, vertical ou diagonal um *pixel* pertencente ao fundo são considerados como sendo constituintes do contorno do objecto.

Este parâmetro tem como vantagens a insensibilidade a rotações, translações e ponto de começo, apresentando como desvantagem a sua sensibilidade a mudanças de escala. O objecto *Bream* [CE1] da Figura 2.2a) tem como representação gráfica do seu perímetro a linha preta da Figura 2.2b).



Figura 2.2 – *Bream*: a) Forma do objecto simples (suporte da textura); b) Representação gráfica do perímetro do objecto em a)

▪ Corda Máxima

A corda máxima de um objecto simples ou região corresponde à distância máxima entre dois pontos do contorno fechado do objecto simples ou região. A expressão (2.2) define a corda máxima, C_{Max} , como sendo:

$$C_{Max} = \max \left\{ \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2} \right\} \quad (2.2)$$

onde (i_1, j_1) e (i_2, j_2) são quaisquer dois pontos do contorno fechado do objecto simples ou região e \max é a função maximizante da distância Euclidiana entre dois pontos do contorno fechado.

Este parâmetro, à semelhança do perímetro, tem também como vantagens a insensibilidade a rotações, translações e ponto de começo e, como desvantagem, a sensibilidade a mudanças de escala. O objecto *Classic* [CE1] da Figura 2.3a) tem como representação gráfica da sua corda máxima a linha a preto na Figura 2.3b).

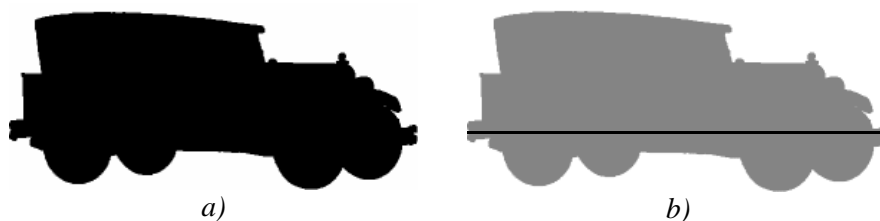


Figura 2.3 – *Classic*: a) Forma do objecto simples [CE1]; b) Representação gráfica da corda máxima do objecto

▪ Circularidade

A circularidade de um objecto simples ou de uma região traduz a sua semelhança em relação a uma circunferência. Um objecto simples ou região é tão mais circular quanto mais próximo da unidade for o valor da sua circularidade normalizada a 1. A expressão (2.3) define a circularidade, *Circ*, como:

$$Circ = \frac{4 \pi A}{P^2} \quad (2.3)$$

onde P é o perímetro em vizinhança de ordem 8 da forma a descrever e A é a área (tal como definida na secção seguinte) do objecto simples ou região.

Este parâmetro tem como vantagens a insensibilidade a rotações, translações, mudanças de escala e ponto de começo. O objecto *Bell* [CE1] da Figura 2.4a) tem como circunferência equivalente à sua circularidade (circunferência com a mesma área que o objecto) a circunferência a preto na Figura 2.4b).

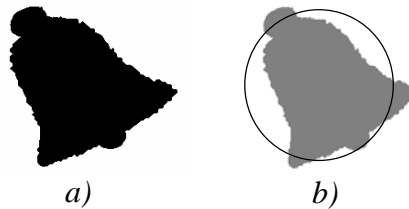


Figura 2.4 – *Bell*: a) Forma do objecto simples [CE1]; b) Representação gráfica da circularidade do objecto

▪ Convexidade

A convexidade de um objecto simples ou região traduz a semelhança entre a forma do objecto simples ou região e a sua forma convexa definida como o polígono que envolve os pontos mais extremos do objecto simples ou região em questão [Russ1995]. A expressão (2.4) define a convexidade, C_v , como:

$$C_v = \frac{P_{cv}}{P} \quad (2.4)$$

onde P é o perímetro numa vizinhança de ordem 8 e P_{cv} é o perímetro da sua forma convexa.

Este parâmetro, à semelhança da circularidade, tem como vantagens a insensibilidade a rotações, translações, mudanças de escala e ponto de começo. O objecto *Device* [CE1] na Figura 2.5a) tem como representação gráfica da sua convexidade a forma convexa a preto na Figura 2.5b).

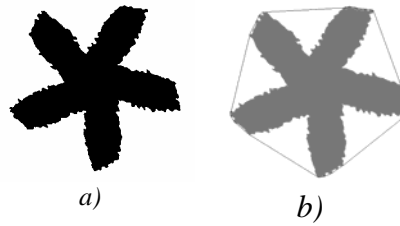


Figura 2.5 – *Device*: a) Forma do objecto simples [CE1]; b) Representação gráfica da convexidade do objecto

▪ Excentricidade

A excentricidade de um objecto simples ou região é definida como a relação existente entre o seu raio máximo, R_{max} , e o seu raio mínimo, R_{min} . Os vários raios são definidos como a distância entre o centróide do objecto e qualquer ponto da menor elipse envolvente do objecto (ou seja que engloba totalmente o objecto), centrada no centróide do objecto. A expressão (2.5) define a excentricidade, E , da forma como:

$$E = \frac{R_{max}}{R_{min}} \quad (2.5)$$

Este parâmetro tem como vantagens a insensibilidade a rotações, translações, mudanças de escala e ponto de começo. O objecto *Bird* [CE1] na Figura 2.6a) tem como raio máximo e mínimo as linhas traçadas a preto na Figura 2.6b).

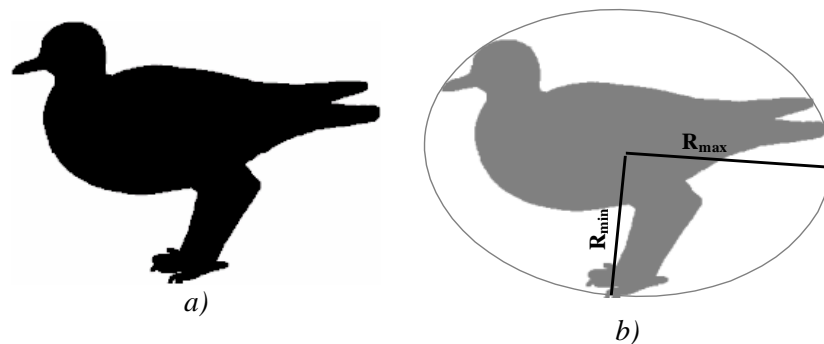


Figura 2.6 – *Bird*: a) Forma do objecto simples [CE1]; b) Representação gráfica dos raios máximo e mínimo do objecto

2.2.1.2 Medidas de semelhança dos parâmetros geométricos

As medidas de semelhança dos parâmetros geométricos pretendem medir a semelhança entre os parâmetros geométricos da forma fornecidos ao mecanismo pelo utilizador e os previamente descritos pelo mecanismo. Deste modo, as medidas de semelhança mais comumente utilizadas nestes parâmetros e definidas de seguida são a distância Euclidiana e distância de *Minkowsky* [GoWo1993].

▪ Distância Euclidiana

A distância Euclidiana, D_E , entre dois vectores de parâmetros, o correspondente à descrição previamente disponível (I) e o correspondente à descrição de procura (Q) é definida como:

$$D_E(I, Q) = \sqrt{\sum_{j=1}^N (P_j^Q - P_j^I)^2} \quad (2.6)$$

onde $P^Q = (P_1^Q, P_2^Q, \dots, P_N^Q)$ e $P^I = (P_1^I, P_2^I, \dots, P_N^I)$ são os vectores de descrição cujas componentes são os valores dos parâmetros geométricos de forma instanciados durante a procura e descrição, respectivamente. O valor N é o número de componentes do parâmetro usado para representar a característica forma.

▪ Distância de *Minkowsky*

A distância de *Minkowsky*, D_{Min} , entre dois vectores é definida como a soma das diferenças absolutas entre as várias componentes dos vectores em questão, ou seja:

$$D_{Min}(I, Q) = \sum_{j=1}^N |P_j^Q - P_j^I| \quad (2.7)$$

onde $P^Q = (P_1^Q, P_2^Q, \dots, P_N^Q)$ e $P^I = (P_1^I, P_2^I, \dots, P_N^I)$ são, novamente, os vectores de descrição cujas componentes são os valores dos parâmetros geométricos de forma instanciados durante a procura e descrição, respectivamente. Da mesma forma, N é o número de componentes do parâmetro usado para representar a característica forma.

Para qualquer uma das medidas de semelhança acabadas de definir, as formas em comparação são tão mais semelhantes quanto menor for o seu valor. Consoante a aplicação e os seus objectivos, poderá ser definido um limiar que delimita os casos onde se declara ou não estar na presença de formas semelhantes ou então as formas são pura e simplesmente ordenadas por ordem decrescente de semelhança.

2.2.2 Parâmetros baseados em transformadas

As transformadas são funções matemáticas, normalmente definidas no domínio da frequência, e utilizadas para a representação de características da forma por projectarem o sinal da forma sobre funções ortogonais, ou seja, distribuindo a energia do sinal sobre um conjunto de componentes não correlacionadas [AnGa1994]. Nesta secção, as transformadas serão aplicadas sobre o sinal de contorno do objecto simples ou região a fim de representar as suas características de forma.

2.2.2.1 Definição de transformadas como parâmetros de forma

Entre as transformadas existentes na literatura e propostas ao *MPEG-7*, foram escolhidas para apresentação nesta tese como parâmetros de forma as transformadas de *Fourier* [GoWo1993] [Jain1989] e as *wavelets* [AnGa1994] [Bimb1999] [MuOh1999a] [MuOh1999b].

▪ Transformada de *Fourier*

Este parâmetro de forma corresponde aos coeficientes obtidos pela aplicação da transformada de *Fourier* a uma função que representa o contorno do objecto simples ou região. A transformada de *Fourier* tem hoje em dia um largo domínio de aplicação, nomeadamente na análise de sinais para codificação. Por este motivo, um dos domínios de aplicação deste parâmetro, a transformada de *Fourier*, é a descrição de contornos fechados de objectos simples ou regiões.

O cálculo deste parâmetro efectua-se uma vez extraído o contorno fechado do objecto simples ou região em análise. Assim sendo, considera-se o par de sinais $x(t)$ e $y(t)$, correspondentes às funções paramétricas das coordenadas amostradas x e y dos *shapels* do objecto simples ou região, como sendo uma função complexa $c(k)$ em vez de duas funções de valores reais, $x(t)$ e $y(t)$. A notação discreta que se segue, $x(k)=x_k$ e $y(k)=y_k$, permitirá representar o contorno como uma sequência de coordenadas $c(k) = [x(k),y(k)]$, onde $k=0,1,...,N-1$. Deste modo, a representação do contorno fechado como valor complexo define-se de acordo com a expressão (2.8):

$$c(k) = x(k) + jy(k) \quad (2.8)$$

Esta função complexa apresenta a vantagem de reduzir a dimensão da transformada de *Fourier* de 2D para 1D, tendo-se por isso que calcular apenas a transformada de *Fourier* 1D através da expressão (2.9), o que vai reduzir o tempo de cálculo:

$$C(u) = \sum_{k=0}^{N-1} c(k) e^{-2\pi i k u / N} \quad (2.9)$$

onde $u=0,1,...,N-1$ e N é o número de pontos do contorno fechado.

Tendo em conta que os parâmetros de forma são tão mais eficazes na procura quanto mais insensíveis forem às transformações geométricas do objecto simples ou região, verificou-se que os coeficientes obtidos pela transformada de *Fourier* são sensíveis a essas transformações. Contudo, esta sensibilidade pode ser facilmente compensada, como se pode ver na Tabela 2.1, através de uma correcção dependente do tipo de transformação. Vejamos, por exemplo, a rotação: ainda que os coeficientes $C(u)$ não sejam os mesmos para um certo contorno e uma sua rotação, há igualdade entre os coeficientes $C_r(u)$ para o contorno rodado e os coeficientes $C(u)$ para o contorno original multiplicados por um factor que só depende do ângulo de rotação. A Tabela 2.1 indica sumariamente a relação entre os coeficientes $C(u)$ e os coeficientes $C_r(u)$, $C_t(u)$, $C_e(u)$ e $C_p(u)$ correspondentes a contornos transformados a partir do original segundo uma dada transformação.

Transformação	Contorno do objecto simples ou região	Coefficientes de <i>Fourier</i>
Identidade	$c(k)$	$C(u)$
Rotação	$c_r(k) = c(k)e^{j\theta}$	$C_r(u) = C(u)e^{j\theta}$
Translação	$c_t(k) = c(k) + \Delta_{xy}$	$C_t(u) = C(u) + \Delta_{xy}\delta(u)$
Mudança de Escala	$c_e(k) = \alpha c(k)$	$C_e(u) = \alpha C(u)$
Ponto de Começo	$c_p(k) = c(k - k_0)$	$C_p(u) = C(u)e^{-2\pi k_0 u / N}$

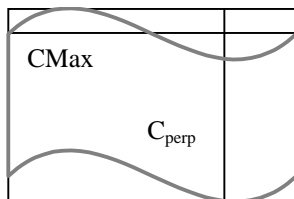
Tabela 2.1 – Efeito de vários tipos de transformação sobre os coeficientes calculados pela transformada de *Fourier* [GoWo1993]

▪ Transformada *wavelet*

A transformada *wavelet* é uma transformada que decompõe a imagem em várias componentes frequenciais. Esta decomposição é efectuada por filtragens iterativas usando-se filtros de *wavelet* passa-baixo e passa-alto. A filtragem é feita para toda a imagem e para cada banda de frequência existirá uma versão filtrada da imagem cuja resolução espacial é proporcional à frequência da filtragem. O resultado da aplicação desta transformada é uma representação hierárquica da imagem em que a cada nível está associada a informação de uma banda de frequência.

Recentemente, as *wavelets* têm sido usadas para a representação da forma de um objecto simples ou região, pela sua capacidade de descrição de propriedades locais e globais da forma e pela sua capacidade de representar a forma com vários níveis de detalhe. A transformada *wavelet* aplicada à forma tem como objectivo representar o contorno fechado, armazenando num vector os seus coeficientes, de um objecto simples ou região através do seguinte algoritmo:

1. Cálculo do *modification_ratio* (*MR*) usado para reduzir a dimensão dos contornos e consequentemente a complexidade do algoritmo, através da expressão (2.10):



$$MR = \frac{C_{perp}}{C_{Max}} \quad (2.10)$$

onde C_{Max} é a corda máxima que define o comprimento do rectângulo envolvente e C_{perp} é a corda perpendicular a C_{Max} e que define a largura do rectângulo envolvente.

2. Cálculo do valor médio das coordenadas, x e y , dos pontos pertencentes ao contorno fechado.
3. Subtração do valor médio calculado aos valores dos pontos do contorno armazenados no vector, obtendo-se valores diferença associados a cada ponto do contorno que são agora insensíveis a translações.
4. Cálculo da transformada *wavelet* dos valores diferença associados aos pontos do contorno em análise. Para o cálculo da transformada *wavelet* é necessário começar por definir um filtro, pertencente a uma classe descoberta por *Daubechies* [PrTe1994], considerando o caso particular *DAUB4* que tem apenas quatro coeficientes, C_0, \dots, C_3 , resultando numa matriz de coeficientes de *wavelet* como a apresentada na expressão (2.11) [PrTe1994]:

$$\begin{bmatrix}
 C_0 & C_1 & C_2 & C_3 & & & & & \\
 C_3 - C_2 & C_1 - C_0 & & & & & & & \\
 & & C_0 & C_1 & C_2 & C_3 & & & \\
 & & C_3 - C_2 & C_1 - C_0 & & & & & \\
 \vdots & \vdots & & & & \ddots & & & \\
 & & & & & & C_0 & C_1 & C_2 & C_3 \\
 & & & & & & C_3 - C_2 & C_1 - C_0 & & \\
 C_2 & C_3 & & & & & & & C_0 & C_1 \\
 C_1 - C_0 & & & & & & & & C_3 - C_2 &
 \end{bmatrix} \quad (2.11)$$

Uma vez definido o filtro resta definir o cálculo da transformada discreta de *wavelets*. Esta transformada consiste em aplicar a matriz de coeficientes (2.11) hierarquicamente aos vectores diferença, X e Y , de dimensão N como mostra a Figura 2.7. Inicialmente, esta matriz é aplicada aos vectores diferença com N elementos, de onde resultam $N/2$ componentes de alisamento, s_i , resultantes das linhas ímpares. De seguida estas componentes são reordenadas (como se pode ver no terceiro vector da Figura 2.7) encontrando-se as componentes s_i primeiro e as d_i a seguir. Este método repete-se para os primeiros $N/4$ elementos do vector, resultando em duas componentes “alisamento-alisamento”, S_i , que por suas vez serão novamente reordenadas. Este processo repetir-se-á até que restem apenas duas componentes de “alisamento” (também denominados por “coeficientes da função mãe”) e as restantes componentes (D_i e d_i) sejam coeficientes da transformada *wavelet*.

$$\begin{array}{c} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \end{bmatrix} \end{array} \xrightarrow{(2.11)} \begin{array}{c} \begin{bmatrix} s_1 \\ d_1 \\ s_2 \\ d_2 \\ s_3 \\ d_3 \\ s_4 \\ d_4 \end{bmatrix} \end{array} \xrightarrow{\text{permutar}} \begin{array}{c} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \frac{s_4}{d_1} \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \end{array} \xrightarrow{(2.11)} \begin{array}{c} \begin{bmatrix} S_1 \\ D_1 \\ S_2 \\ D_2 \\ \frac{D_2}{d_1} \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \end{array} \xrightarrow{\text{permutar}} \begin{array}{c} \begin{bmatrix} S_1 \\ S_2 \\ D_1 \\ D_2 \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \end{array}$$

Figura 2.7 – Esquema representativo da sequência de operações efectuadas para o cálculo dos coeficientes da transformada baseada em *wavelets* [PrTe1994]

Uma vez calculada a transformada *wavelet*, os valores do parâmetro correspondente à descrição da forma serão os coeficientes obtidos pela última iteração do procedimento anterior, ou seja, S_1 , S_2 , D_1 , D_2 , d_1 , d_2 , d_3 e d_4 .

É de salientar que este parâmetro é insensível a translações e ao ponto de começo, sendo a sensibilidade à mudança de escala dos coeficientes eliminada durante o processo de descrição.

2.2.2.2 Medidas de semelhança para os parâmetros baseados em transformadas

As medidas (ou métricas) de semelhança para as transformadas aplicadas à forma pretendem medir a semelhança entre as descrições de uma ou mais formas fornecidas pelo mecanismo de procura ou filtragem e as descrições de forma previamente disponíveis, associadas ao conteúdo descrito. As medidas de semelhança mais comumente utilizadas para os parâmetros atrás apresentados são a distância Euclidiana, a distância de *Minkowsky* e uma medida de semelhança para os coeficientes de *Fourier* que compensa as transformações geométricas definidas de seguida [GoWo1993].

As distâncias Euclidiana e de *Minkowsky* já foram anteriormente definidas na secção 2.2.1.2. Das medidas de semelhança para os parâmetros de transformadas acima mencionadas resta-nos definir apenas a distância insensível a transformações geométricas para os parâmetros de *Fourier* [Jain1989].

▪ Distância para os Coeficientes de *Fourier*

Das medidas de semelhança adequadas para as transformadas acima mencionadas, resta-nos apenas definir a distância insensível a transformações geométricas para os coeficientes de *Fourier* [Jain1989].

Esta distância mede o grau de semelhança entre dois vectores de descrição (com um ou mais coeficientes) correspondentes ao parâmetro de forma Transformada de *Fourier*, o descrito (I) e o de procura (Q), sensível a transformações geométricas. A expressão (2.12) define esta distância como:

$$d(\Delta_0, \alpha, \theta, k_0) \triangleq \min_{\Delta_0, \alpha, \theta, k_0} \left\{ \sum_{n=0}^{N-1} |C_I(u) - \alpha C_Q(u + k_0) e^{j\theta} - \Delta_0|^2 \right\} \quad (2.12)$$

onde $C_I(u)$ e $C_Q(u)$ são, respectivamente, os coeficientes de descrição da forma de um objecto simples ou região descrito e de procura, ambos obtidos através da transformada de *Fourier*. Os valores Δ_0 , α , θ e k_0 são os valores escolhidos para a compensação dos efeitos das transformações geométricas, translação, mudança de escala, rotação e ponto de começo, respectivamente, sendo N o número de coeficientes incluídos no parâmetro definido pela transformada de *Fourier*. A insensibilidade às transformações obtem-se através da consideração de múltiplas descrições da procura associadas a um conjunto relevante de transformações sobre a forma básica de procura.

2.2.3 Parâmetros baseados em momentos

Os momentos são representações estatísticas da forma do objecto simples ou região em análise. Consequentemente, os momentos têm sido usados como parâmetros de forma para representar os contornos fechados de um objecto simples ou região.

2.2.3.1 Definição de momentos como parâmetros de forma

Os momentos aqui apresentados são denominados por momentos geométricos ou invariantes, devido às suas propriedades de insensibilidade a transformações geométricas como a translação, rotação, mudança de escala e ponto de começo [MuOh1999c].

Os parâmetros de forma baseados nos momentos invariantes têm como objectivo representar o contorno fechado de um objecto simples ou região através das suas propriedades estatísticas. Normalmente, os momentos invariantes calculados são sete e baseiam-se nos momentos centrais de segunda e terceira ordem. A razão pela qual são calculados estes sete momentos invariantes, e nem mais nem menos, deve-se ao facto de a partir de certa ordem não se obter mais detalhe do que aquele que é necessário e distinguível pelo ser humano. Desta forma, os sete momentos invariantes são calculados do seguinte modo:

1. Cálculo dos sete momentos centrais, μ_{00} , μ_{11} , μ_{20} , μ_{02} , μ_{21} , μ_{12} , μ_{30} , μ_{03} , através da expressão (2.13):

$$\mu_{pq} = \sum_{n=1}^N ((x(n) - \bar{x})^p \cdot (y(n) - \bar{y})^q) \quad (2.13)$$

onde N é o número de pontos (x, y) do contorno fechado, e \bar{x} e \bar{y} são as médias das componentes x e y dos pontos de contorno, respectivamente;

2. Normalização dos momentos centrais μ_{pq} , gerando-se momentos centrais normalizados m_{pq} através das expressões que se seguem, (2.14) e consequentemente (2.15) [MuOh1999c]:

$$m_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \text{em que} \quad \gamma = \frac{1}{2}(p+q)+1 \quad e \quad (p+q) = 2, 3, \dots \quad (2.14)$$

$$\begin{aligned} m_{11} &= \mu_{11} / \mu_{00}^2; & m_{20} &= \mu_{20} / \mu_{00}^2; & m_{02} &= \mu_{02} / \mu_{00}^2; \\ m_{21} &= \mu_{21} / \mu_{00}^{3/2}; & m_{12} &= \mu_{12} / \mu_{00}^{3/2}; & m_{30} &= \mu_{30} / \mu_{00}^{3/2}; \\ m_{03} &= \mu_{03} / \mu_{00}^{3/2}; \end{aligned} \quad (2.15)$$

3. Cálculo dos momentos invariantes através das expressões (2.16), $M1...M7$, usando os valores m_{pq} acima definidos:

$$\begin{aligned} M1 &= m_{20} + m_{02}; \\ M2 &= (m_{20} - m_{02})^2 + 4m_{11}^2; \\ M3 &= (m_{30} - 3m_{12})^2 + (m_{03} - 3m_{21})^2; \\ M4 &= (m_{30} + 3m_{12})^2 + (m_{03} + 3m_{21})^2; \\ M5 &= (m_{30} - 3m_{12})(m_{03} + 3m_{21}) \left[(m_{30} + 3m_{12})^2 - 3(m_{03} + 3m_{21})^2 \right] \\ &\quad + (m_{03} - 3m_{21})(m_{30} + 3m_{12}) \left[(m_{03} + 3m_{21})^2 - 3(m_{30} + 3m_{12})^2 \right]; \\ M6 &= (m_{20} - m_{02})(m_{30} + m_{12}) \left[(m_{03} + m_{21})^2 \right] \\ &\quad + 4m_{11}(m_{30} - m_{12})(m_{03} + 3m_{21}); \\ M7 &= (3m_{12} - m_{30})(m_{30} + m_{12}) \left[(m_{30} + m_{12})^2 - 3(m_{03} + m_{21})^2 \right] \\ &\quad + (m_{03} - 3m_{21})(m_{03} + 3m_{21}) \left[(m_{03} + m_{21})^2 - 3(m_{30} + m_{12})^2 \right]; \end{aligned} \quad (2.16)$$

Estes momentos expressam o espalhamento do objecto em relação ao seu ponto médio, sendo assim úteis na caracterização da regularidade do objecto. O parâmetro de forma baseado nos momentos poderá incluir todos ou apenas alguns dos momentos acima definidos, consoante as necessidades associadas às aplicações em causa. Como já se disse, estes momentos são insensíveis às várias transformações mais relevantes.

2.2.3.2 Medidas de semelhança para os parâmetros baseados em momentos

As medidas de semelhança, para os parâmetros baseados em momentos, mais utilizadas são a distância Euclidiana e a distância de *Minkowsky*, anteriormente definidas na secção 2.2.1.

2.2.4 Parâmetros baseados em contornos normalizados

A representação da forma baseada num contorno normalizado surge como resposta à enorme variação das imagens de procura em termos do número de pontos de contorno, nomeadamente em relação às imagens existentes nas bases de dados [Taba1999].

2.2.4.1 Definição do parâmetro baseado num contorno normalizado

A normalização do contorno fechado de um objecto simples ou região é feita para compensar o número variável de pontos de contorno que surgem após a extracção do seu contorno. Os parâmetros calculados com este método serão ainda alterados a fim de ficarem insensíveis a transformações geométricas (translação, mudança de escala, rotação e ponto de começo) aplicáveis ao objecto simples ou região. Assim sendo, a representação da forma através do contorno normalizado faz-se em duas etapas, nomeadamente: extracção do contorno e normalização do contorno.

1. **Extracção do contorno** – Nesta etapa, o contorno do objecto simples ou região é extraído a partir do seu suporte binário, tendo um número variável de pontos como mostra a Figura 2.8.

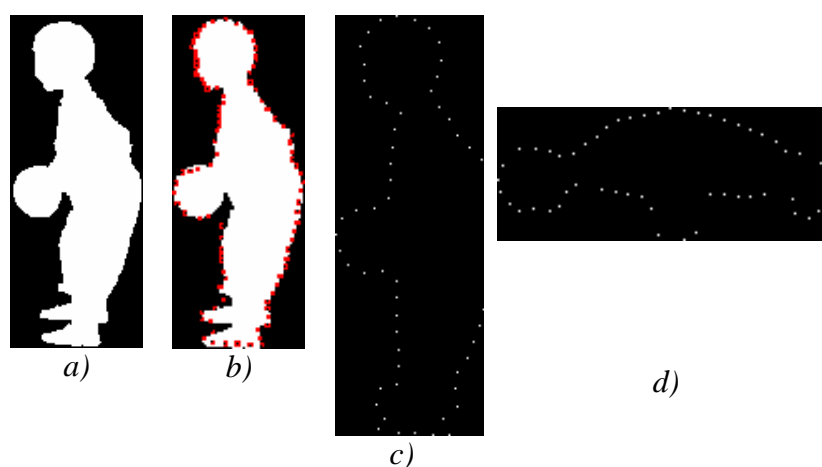


Figura 2.8 – *Children*: a) Forma a analisar; b) Extracção do contorno, com um número de pontos variável e não equidistante, ajustado ao objecto simples ou região; c) Contorno re-amostrado a 64 pontos usando-se o método das *Cubic Splines*; d) Contorno com descrição insensível à rotação [Taba1999]

2. **Normalização do contorno** – Esta etapa da normalização do contorno é feita em cinco passos, como mostra a Figura 2.9.

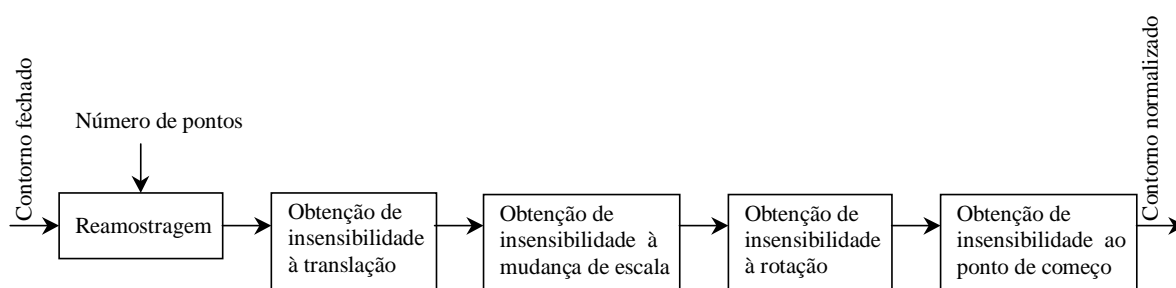


Figura 2.9 – Determinação do contorno normalizado [Taba1999]

- i. **Reamostragem** – Aplica-se para que todos os contornos normalizados tenham o mesmo número de pontos equidistantes. O número de pontos do contorno fechado

é definido pelo implementador da ferramenta ou pelo utilizador do mecanismo de descrição e procura, e fornecido ao módulo de reamostragem. A Figura 2.8c) mostra um exemplo do contorno reamostrado a 64 pontos através do método das *Cubic Splines* [Taba1999].

- ii. **Obtenção de insensibilidade à translação** – Aplica-se para que todos os contornos normalizados sejam insensíveis a translações dos objectos simples ou regiões. A insensibilidade à translação é obtida calculando-se o valor médio, (μ_x, μ_y) , das duas coordenadas para todos os pontos do contorno fechado, através da expressão (2.17):

$$(\mu_x, \mu_y) = \frac{\sum_{i=0}^{N-1} (x_i, y_i)}{N} \quad (2.17)$$

onde (x_i, y_i) são as coordenadas dos pontos do contorno reamostrado e N é o número de pontos do contorno reamostrado. De seguida, obtém-se um contorno insensível a translações, C_T , através da expressão (2.18):

$$\begin{aligned} (X_i, Y_i) &= (x_i, y_i) - (\mu_x, \mu_y) \\ C_T &= \{(X_0, Y_0), (X_0, Y_1), \dots, (X_N, Y_N)\} \end{aligned} \quad (2.18)$$

onde (X_i, Y_i) são as coordenadas dos pontos do contorno insensível a translações.

- iii. **Obtenção da insensibilidade à mudança de escala** – Aplica-se para que todos os contornos normalizados sejam insensíveis a mudanças de escala dos objectos simples ou regiões. A insensibilidade a mudanças de escala é obtida calculando-se a norma do contorno C_T :

$$\|C_T\| = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (x_i^2 + y_i^2)} \quad (2.19)$$

C_T é então escalado, obtendo-se um contorno insensível a mudanças de escala, C_{ME} , com norma igual a 1:

$$C_{ME} = \frac{C_T}{\|C_T\|} \quad (2.20)$$

- iv. **Obtenção de insensibilidade à rotação** – Aplica-se para que todos os contornos normalizados sejam insensíveis a rotações dos objectos simples ou regiões. A insensibilidade à rotação é obtida através de um método baseado nos momentos centrais e que usa o ângulo de $\pi/2$ para a normalização do contorno [Taba1999]. A Figura 2.8d) mostra o contorno insensível à rotação do objecto.
- v. **Obtenção da insensibilidade ao ponto de começo** – Aplica-se para que todos os contornos normalizados sejam insensíveis ao ponto de começo do cálculo do parâmetro. A insensibilidade ao ponto de começo é obtida começando-se por identificar os pontos do contorno cuja componente vertical (eixo-y) muda de sinal em relação aos pontos de contorno adjacentes, identificando-se, de seguida, a

coordenada da componente horizontal (eixo-x) máxima correspondente a um desses pontos de contorno. O ponto de começo será o ponto do contorno com a coordenada x máxima e o respectivo valor de y .

Após estes cinco passos, obtém-se uma descrição da forma do objecto simples ou região através de um conjunto de coordenadas (x,y) correspondente ao contorno normalizado e insensível à translação, rotação, mudança de escala e ponto de começo.

2.2.4.2 Medidas de semelhança para o parâmetro baseado em contornos normalizados

A medida de semelhança para o parâmetro de descrição de forma contorno normalizado tem como objectivo avaliar quão semelhantes são dois vectores correspondentes a dois contornos normalizados. As medidas de semelhança para os contornos normalizados mais usadas são a distância Euclidiana e a distância de *Minkowsky*.

2.2.5 Parâmetros baseados nos ângulos de curvatura do contorno

O parâmetro de forma baseado nos ângulos de curvatura do contorno é normalmente definido como o conjunto dos ângulos feitos por dois vectores: um é o que une dois pontos consecutivos representativos do contorno e o outro é definido pelo eixo principal da forma a descrever. O conjunto destes ângulos define uma aproximação poligonal do contorno.

2.2.5.1 Definição dos ângulos de curvatura do contorno

Assuma-se que O é o objecto simples ou região em análise, p o seu perímetro, C o centróide de O e A_1 e A_2 os seus eixos principais. Os eixos principais são definidos pelos vectores próprios da matriz de covariância da forma, definida pelos seus *shapels* mais exteriores ou seja pelo seu contorno, e representada através da matriz V da expressão (2.21) [IBMRe1999][NiYi1999]:

$$V = \begin{bmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{bmatrix}$$

$$V_{xx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(x_i - m_x)$$

$$V_{xy} = V_{yx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(y_i - m_y)$$

$$V_{yy} = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)(y_i - m_y)$$
(2.21)

Os eixos principais são identificados através dos respectivos valores próprios e de entre os valores próprios resultantes, o valor próprio de maior amplitude definirá o *eixo-x*, enquanto

que o segundo maior definirá o *eixo-y*. Uma vez definida a terminologia essencial ao cálculo deste parâmetro, segue-se a definição do cálculo do parâmetro:

1. Definição do número de ângulos da curvatura, N , a serem calculados;
2. Ordenação dos pontos do contorno segundo a direcção contrária à dos ponteiros do relógio;
3. Selecção dos pontos representativos do contorno para o cálculo dos ângulos da curvatura do contorno. A selecção dos pontos poderá ser feita usando-se vários algoritmos como, por exemplo, o algoritmo seguinte: sendo (x_j, y_j) com $j=0, \dots, w$, os pontos do contorno fechado do objecto simples ou região então, caso $w < N$, o y é interpolado de modo a obter-se um conjunto com N pontos. Caso contrário, se $w > N$, é calculada a média de cada conjunto de N/w pontos do contorno fechado, resultando num conjunto de N pontos, $(x'(i), y'(i))$, com $i=0, \dots, N-1$;
4. Cálculo dos ângulos de curvatura $t(i)$ a partir da expressão (2.22), considerando o algoritmo acima sugerido para a selecção dos pontos representativos, como mostra a Figura 2.10:

$$t(i) = \arctg \frac{(y'(i+1) - y'(i))}{(x'(i+1) - x'(i))} \quad (2.22)$$

que vão ser usados como os valores que definem o vector correspondente ao parâmetro de descrição.

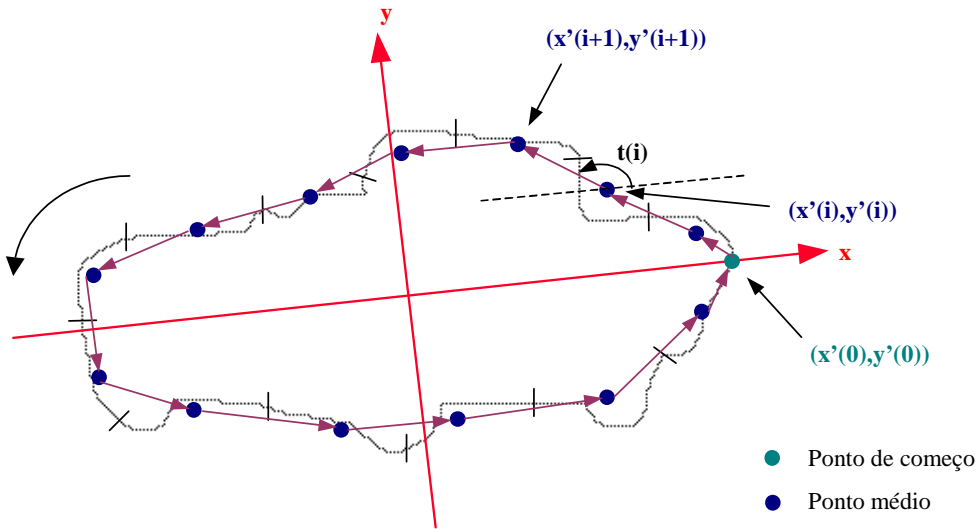


Figura 2.10 – Exemplo representativo de ângulos de curvatura do contorno

Este parâmetro de descrição da forma tem como vantagens a insensibilidade à rotação, translação e mudança de escala e apresenta problemas relativamente à sua sensibilidade ao ponto de começo. No entanto, este problema pode ser e foi solucionado por alguns especialistas [IBMRe1999][NiYi1999] considerando-se no acto da descrição quatro vectores de ângulos, cada um calculado a partir de um de quatro pontos de começo. Estes pontos de começo são definidos como os pontos de intersecção dos eixos principais com o contorno do objecto a ser descrito.

2.2.5.2 Medidas de semelhança para o parâmetro baseado nos ângulos de curvatura do contorno

A medida de semelhança apresentada de seguida tem como objectivo comparar os ângulos de curvatura das formas descritas disponíveis e daquela fornecida ao mecanismo de procura, de modo a avaliar eficazmente a semelhança entre as formas em comparação [IBMRe1999] [NiYi1999].

Esta medida de semelhança começa por considerar dois vectores, $A_B(i)$ e $A_I(i)$, com os ângulos de curvatura correspondentes a uma das formas descritas disponíveis e à forma fornecida ao mecanismo de procura. De seguida, é calculado o melhor emparelhamento entre os dois vectores, ou seja, o emparelhamento de cada elemento do vector $A_B(i)$ com um elemento do vector $A_I(i)$ resultando na menor soma de diferenças absolutas. Consequentemente, é calculada uma sequência, $S(t)$, com pares de índices dos ângulos de curvatura para as duas formas em comparação:

$$S(t) = (i_1, j_1)(i_2, j_2) \cdots (i_k, j_k), \quad (i_{t+1} = i_t \cup i_{t+1} = i_t + 1) \cap (j_{t+1} = j_t \cup j_{t+1} = j_t + 1)$$

que minimiza a soma das diferenças absolutas, D_A :

$$D_A = \sum_{t=1}^k |A_B(i_t) - A_I(j_t)| + r \quad (2.23)$$

onde $r=0$ se $i_t=i_{t-1}$ e $j_t=j_{t-1}$; caso contrário, r toma um valor penalizante para que se acentue a diferença existente entre os termos da condição. Resta dizer que o emparelhamento de vários elementos de $A_B(i)$ pode ser feito com o mesmo elemento de $A_I(i)$

Se as formas forem semelhantes, a sequência $S(t)$ terá uma correspondência do tipo 1 para 1 entre i e j , como por exemplo (1,1)(2,2)(3,3) etc. Se as formas forem diferentes, a sequência será mais extensa de modo a minimizar a distância D_A , como por exemplo a sequência (1,1)(1,2)(1,3)(2,4) etc.

Uma vez que os vectores de ângulos são sensíveis ao ponto de começo, a medida anteriormente descrita também o é, o que poderia limitar seriamente os resultados obtidos. Consequentemente, esta medida poderá ser melhorada de modo a tornar-se insensível ao ponto de começo, considerando-se quatro pontos de começo para a forma da procura e criando-se a partir de cada ponto de começo um novo vector com os ângulos de curvatura.

2.2.6 Parâmetros baseados numa imagem *Curvature Scale Space*

O parâmetro de forma baseado numa imagem *Curvature Scale Space* (CSS), tal como todos os anteriores, pretende descrever a forma de um objecto simples ou região através de propriedades do seu contorno fechado. A imagem CSS é uma representação multi-escalar, ou seja com várias resoluções (ver Figura 2.11a), dos pontos de inflexão (convexos ou côncavos) do contorno fechado [Bobe1999][MoAb1999][MPEG-7XM1999].

2.2.6.1 Definição do parâmetro baseado numa imagem CSS

O cálculo da imagem CSS baseia-se num método paramétrico para a descrição do contorno fechado de um objecto simples ou região. Este parâmetro é calculado através de 4 etapas, explicadas de seguida:

1. Reamostragem dos pontos do contorno do objecto ou região, (x,y) , para um número de pontos u pré-definido e sua normalização a 1, $(x(u),y(u))$. As coordenadas $x(u)$ e $y(u)$ representam as coordenadas paramétricas da curva Γ .
2. Definição do filtro Gaussiano 1D, $g(u,\sigma)$ de largura σ e redefinição de cada uma das componentes dos pontos do contorno em função da largura do filtro (ver expressão (2.24)), ou seja, as coordenadas $X(u,\sigma)$ e $Y(u,\sigma)$ da curva envolvente Γ_σ , respectivamente.

$$\begin{aligned} X(u,\sigma) &= x(u) * g(u,\sigma) \\ Y(u,\sigma) &= y(u) * g(u,\sigma) \end{aligned} \quad (2.24)$$

onde $*$ representa a convolução dos dois sinais.

3. Cálculo da função de curvatura $C(u,\sigma)$ do contorno envolvente do objecto ou região através da expressão (2.25):

$$C(u,\sigma) = \frac{X_u(u,\sigma)Y_{uu}(u,\sigma) - X_{uu}(u,\sigma)Y_u(u,\sigma)}{\left(X_u(u,\sigma)^2 + Y_u(u,\sigma)^2\right)^{3/2}} \quad (2.25)$$

onde $X_u(u,\sigma), Y_u(u,\sigma), X_{uu}(u,\sigma), Y_{uu}(u,\sigma)$ são as primeira e segunda derivadas em ordem a u das componentes da curva envolvente, respectivamente.

4. Após a definição da função $C(u,\sigma)$, calculam-se os zeros da função para valores crescentes de σ , podendo os resultados ser representados num plano (u,σ) como uma imagem binária CSS. Esta imagem terá como eixos horizontal e vertical, o índice do respectivo ponto, u , e a largura do filtro, σ , respectivamente.

O objecto da Figura 2.11a) tem como representação do seu contorno a imagem CSS da Figura 2.11b). A imagem CSS deste objecto é interpretada do seguinte modo:

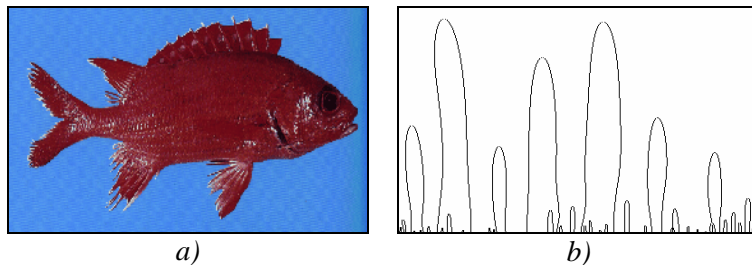


Figura 2.11 – Objecto ‘peixe’: a) textura do objecto e b) Imagem CSS do contorno do objecto [SQUID]

- ♦ Para cada σ , tem-se uma nova curva Γ_σ (ver Figura 2.12a), c) e e)), que é um contorno fechado representativo do contorno do objecto ou região, com um conjunto de zeros representativos da forma diferentes e uma imagem CSS diferente (ver Figura 2.12b), d) e f));
- ♦ A intersecção de cada linha horizontal com os contornos da imagem CSS indicam a localização dos zeros na curva Γ_σ . As Figura 2.12b), d), f) e h) apresentam a imagem CSS até à linha horizontal representativa do número de zeros existentes nas curvas das Figura 2.12a), c), e) e g).

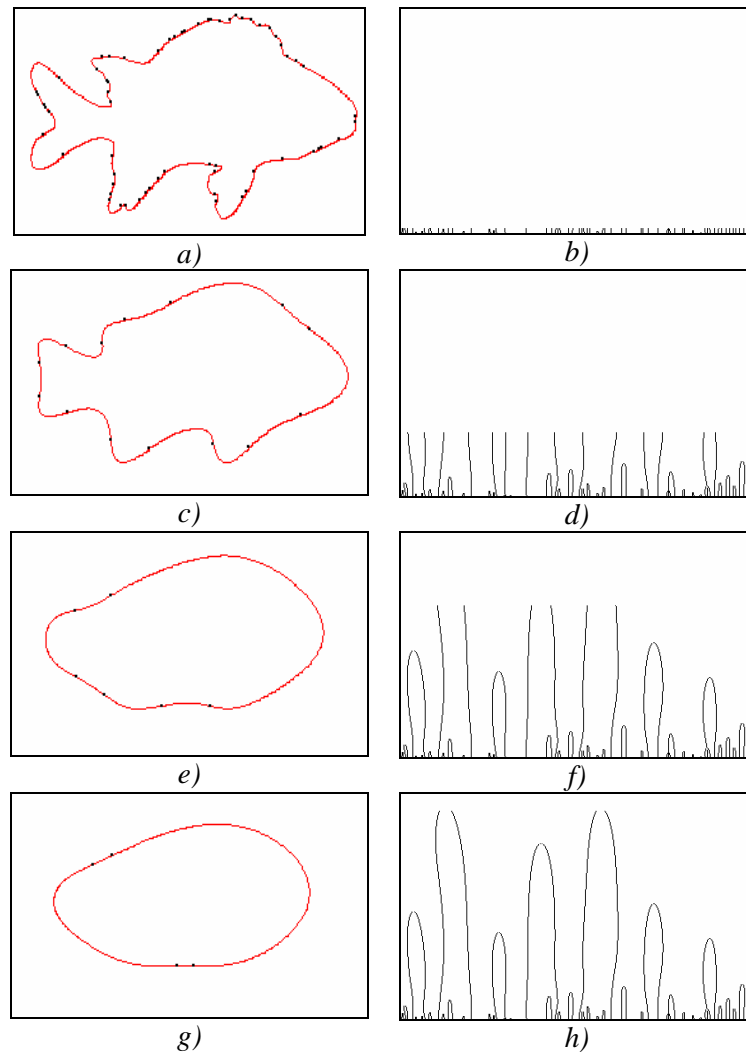


Figura 2.12 – Objecto ‘peixe’: a), c), e) e g) Representação multi-escalar através da forma contorno do objecto simples ou região; b), d), f) e h) Imagem CSS do objecto simples ou região intersectado com a linha horizontal, de valor σ , representativa do número de zeros existentes nas formas dos objectos adjacentes [SQUID]

Consequentemente, este parâmetro de descrição corresponde a um conjunto de pares (u, σ) , representativos dos picos máximos da imagem CSS. Este parâmetro de forma tem como vantagens a insensibilidade à translação, à rotação e à mudança de escala, podendo conseguir-se a insensibilidade ao ponto de começo através do uso de uma medida de semelhança adequada.

2.2.6.2 Medidas de semelhança para o parâmetro baseado na imagem CSS

A medida de semelhança definida de seguida mede a semelhança entre duas imagens CSS usadas como parâmetros de forma. A medida consiste essencialmente em medir a semelhança existente entre os picos da imagem CSS [MPEG7-XM1999].

A distância usada para medir a semelhança existente entre as formas descritas e a procurada é a M_{CSS} definida através da expressão (2.26):

$$M_{CSS} = \sum (x_i - x_j)^2 + (y_i - y_j)^2 + \sum (x_u^2 + y_u^2) \quad (2.26)$$

onde \sum_l é a soma das diferenças quadráticas das amplitudes de todos os picos, (x_i, y_i) e (x_j, y_j) são os picos da imagem de procura e da imagem descrita na base de dados, que se assemelham e \sum_u é a soma da amplitude de todos os picos, (x_u, y_u) , que não se assemelham. Para que dois picos sejam definidos como semelhantes, é necessário que a distância Euclidiana entre as coordenadas x desses picos seja inferior a um determinado limiar, por exemplo 0.1.

2.3 Parâmetros de forma baseados em regiões

Os parâmetros de forma baseados em regiões são parâmetros que descrevem formas simples e formas mais complexas tipicamente associadas a um conjunto de regiões formando um objecto complexo [MPEG-7XM1999]. A Figura 2.13 apresenta alguns objectos que poderão ser descritos por parâmetros baseados em regiões.



Figura 2.13 – Exemplos de objectos complexos e respectivas regiões e buracos [MPEG-7XM1999]

As secções seguintes apresentam alguns parâmetros deste tipo disponíveis na literatura ou submetidos ao MPEG-7 na sequência da chamada de propostas feita em Outubro de 1998. Os parâmetros serão organizados de acordo com as seguintes categorias:

- **Parâmetros geométricos** – Parâmetros que representam a forma do objecto simples ou complexo usando propriedades geométricas do seu contorno ou região. Exemplos: *bounding box*, área, centróide, projecções: altura e largura, diâmetro circular equivalente, solidez e compactação;
- **Parâmetros baseados em vectores próprios multi-nível (*Multi Layer Eigen Vectors - MLEV*)** – Parâmetros que representam a forma de um objecto complexo com vários níveis de detalhe, sub dividindo sucessivamente o objecto segundo as direcções dos vectores próprios calculados na iteração anterior;

- **Parâmetros baseados em transformadas** – Parâmetros que representam a forma de um objecto complexo usando coeficientes calculados a partir de uma dada transformada. Exemplo: transformada angular-radial (*ART*) e de *Fourier*;
- **Parâmetros baseados em momentos** – Parâmetros que representam a forma de um objecto complexo usando um conjunto de valores estatísticos associados a um dado tipo de momento. Exemplos: momentos geométricos, momentos de *Legendre*, momentos de *Zernike*, momentos rotacionais e momentos complexos [ChRo1988].

2.3.1 Parâmetros geométricos

Os parâmetros de forma geométricos são parâmetros que permitem representar um objecto simples ou complexo em relação à sua geometria de contorno ou região, respectivamente [Russ1995].

2.3.1.1 Definição de parâmetros geométricos

Alguns dos parâmetros de forma geométricos baseados na região existentes e descritos de seguida, são: a *bounding box*, a área, o centróide, as projecções: altura e largura, o diâmetro de circularidade máxima, a solidez e a compactação [GoWo1999] [MPEG-7XM1999] [Russ1995] [SoHi1993] [Tekt1999].

▪ Bounding box

A *bounding box* de um objecto complexo (ou simples) permite representar grosseiramente a forma do objecto complexo através do menor rectângulo que engloba completamente o objecto, tendo os seus lados que ser paralelos aos eixos principais do objecto complexo [MuPa1999].

Os valores do parâmetro correspondente à *bounding box* poderão ser a altura, B_A , e a largura, B_L , da mesma. Este parâmetro tem como vantagens a insensibilidade à rotação, translação e ponto de começo; no entanto, este parâmetro é sensível às mudanças de escala. A Figura 2.14 apresenta a *bounding box* (rectângulo a preto) do objecto simples *KK-902* [CE1] e do objecto complexo *1001* [CE2S8], respectivamente.

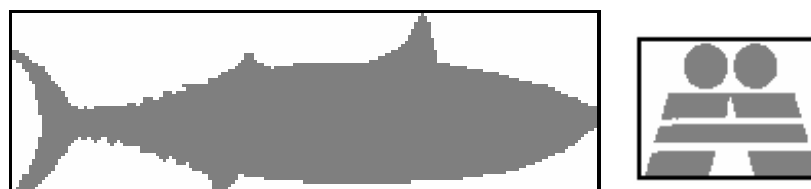


Figura 2.14 – *KK-902*: a) Forma do objecto simples e *bounding box* correspondente [CE1];
1001: b) Forma do objecto complexo e *bounding box* correspondente [CE2S8]

▪ Área

A área de um objecto complexo corresponde ao número total de *shapels* que constituem o objecto complexo. A expressão (2.27) define a área, A , como sendo:

$$A = \sum_i \sum_j I(i, j) \quad (2.27)$$

onde $I(i, j)$ é o suporte binário da forma do objecto e (i, j) são as posições dos *shapels* nesta matriz pertencentes ao objecto.

Este parâmetro tem como vantagens a sua insensibilidade a rotações, translações e ponto de começo, apresentando como desvantagem a sua sensibilidade a mudanças de escala. O objecto *Butterfly* [CE1] e *I222* [CE2S8] da Figura 2.15a) e c) tem como representação gráfica da sua área as regiões a preto da Figura 2.15b) e d).

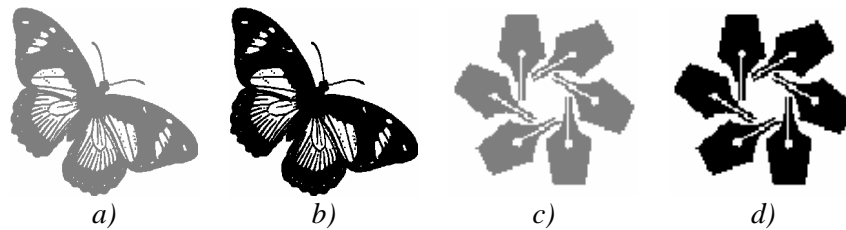


Figura 2.15 – *Butterfly*: a) Forma do objecto simples [CE1]; b) Representação gráfica da área do objecto; *I222*: c) Forma do objecto complexo [CE2S8]; d) Representação gráfica da área do objecto

▪ Centróide

O centróide CM de um objecto complexo representa o centro de massa do objecto complexo. CM_x e CM_y são as coordenadas do centróide, correspondendo aos valores médios das coordenadas de todos os *shapels*, para as coordenadas x e y , respectivamente. As expressões (2.28) e (2.29) definem CM_x e CM_y como:

$$CM_x = \frac{\sum_k \sum_i x_{k_i}}{\sum_k A_k} \quad (2.28)$$

$$CM_y = \frac{\sum_k \sum_i y_{k_i}}{\sum_k A_k} \quad (2.29)$$

onde A_k é a área de cada uma das k regiões do objecto complexo e (x_{k_i}, y_{k_i}) são os pontos pertencentes à região k do objecto complexo.

Este parâmetro tem como vantagens a sua insensibilidade a rotações, mudanças de escala e ponto de começo, apresentando como desvantagem a sensibilidade a translações. O objecto *Apple* [CE1] e *I253* [CE2S8] da Figura 2.16a) e c) tem como representação gráfica do seu centróide o ponto preto marcado sobre a Figura 2.16b) e d).

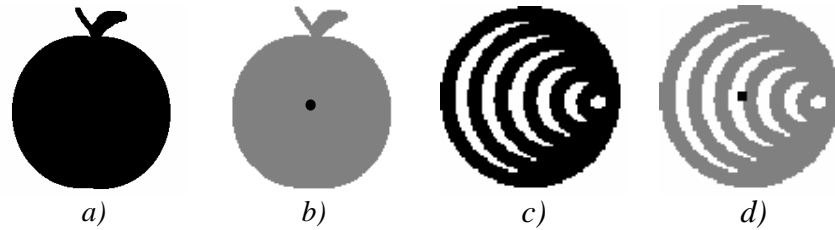


Figura 2.16 – *Apple*: a) Forma do objecto simples [CE1]; b) Representação gráfica do centróide do objecto; *1253*: c) Forma do objecto complexo [CE2S8]; d) Representação gráfica do centróide do objecto

▪ Projecções: altura e largura

As projecções de um objecto complexo representam a altura e a largura do objecto (incluindo as zonas entre as várias regiões), sendo definidas como o valor máximo da projecção vertical e horizontal, respectivamente. As expressões (2.30) e (2.31) definem a altura, AL , e a largura, L , como sendo:

$$AL = \sum_i I(i, j) \quad (2.30)$$

$$L = \sum_j I(i, j) \quad (2.31)$$

onde $I(i, j)$ é o suporte binário da forma do objecto e (i, j) são as posições dos *shapels* nesta matriz pertencentes ao objecto simples ou complexo.

Este parâmetro, à semelhança da área, tem como vantagens a sua insensibilidade a rotações, translações e ponto de começo, apresentando no entanto sensibilidade a mudanças de escala. O objecto *AEA005D* [CE1] e *1008* [CE2S8] da Figura 2.17a) e c) tem como representação gráfica das suas projecções as linhas marcadas a preto da Figura 2.17b) e d).

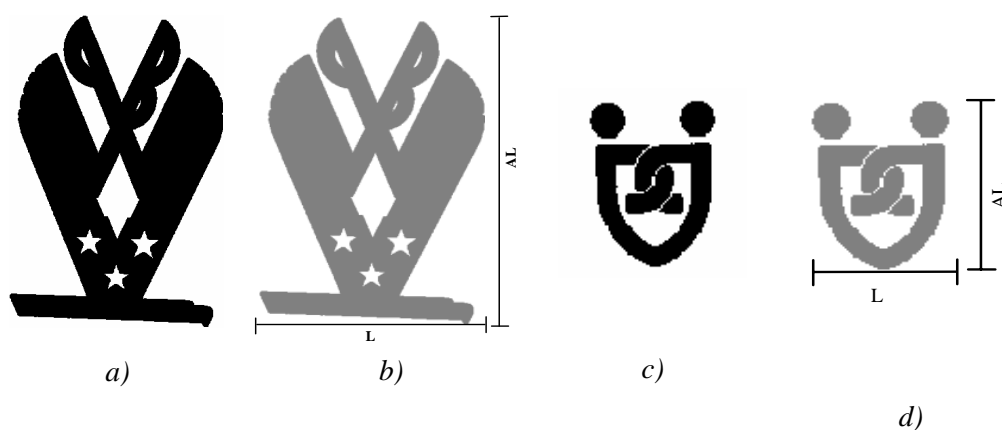


Figura 2.17 – *AEA005D*: a) Forma do objecto simples [CE1]; b) Representação gráfica das projecções do objecto; *1008*: c) Forma do objecto complexo [CE2S8]; d) Representação gráfica das projecções do objecto

▪ Diâmetro circular equivalente

O diâmetro circular equivalente de um objecto complexo representa o diâmetro da circunferência que tem a mesma área que o objecto complexo analisado. A expressão (2.32) define o diâmetro circular equivalente, DCE , como sendo:

$$DCE = \sqrt{\frac{4}{\pi} A} \quad (2.32)$$

onde A é a área do objecto complexo.

Este parâmetro tem como vantagens a sua insensibilidade a rotações, translações, mudanças de escala e ponto de começo. O objecto *Device2-1* [CE1] e *1007* [CE2S8] da Figura 2.18a) e c) tem como representação gráfica do seu diâmetro circular equivalente a linha marcada a preto da Figura 2.18b) e d).

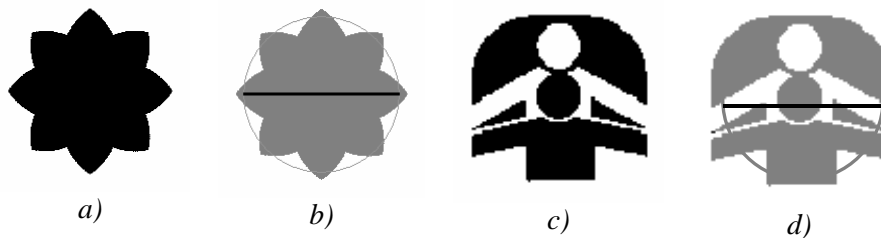


Figura 2.18 – *Device2-1*: a) Forma do objecto simples [CE1]; b) Representação gráfica do DCE do objecto; *1007*: c) Forma do objecto complexo [CE2S8]; d) Representação gráfica do DCE do objecto

▪ Solidez

A solidez de um objecto complexo representa a semelhança entre a área da forma convexa do objecto complexo (como anteriormente definida para a convexidade) e a área do mesmo objecto. A expressão (2.33) define a solidez, S , como sendo:

$$S = \frac{A}{A_{cv}} \quad (2.33)$$

onde A e A_{cv} são, respectivamente, a área e a área da forma convexa do objecto.

Este parâmetro, à semelhança do DCE , tem como vantagens a sua insensibilidade a rotações, translações, mudanças de escala e ponto de começo. O objecto *Device5-1* da Figura 2.19a) tem como representação gráfica da solidez a relação entre a sua área e a área da sua convexidade (A_{cv}), zona marcada a preto na Figura 2.19b).

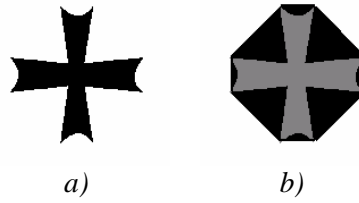


Figura 2.19 – *Device5-1*: a) Forma do objecto; b) Representação gráfica da forma convexa do objecto

▪ Compactação

A compactação de um objecto complexo representa a relação existente entre o perímetro ao quadrado e a área do objecto complexo. A expressão (2.34) define a compactação, CP , como:

$$CP = \frac{P^2}{A} \quad (2.34)$$

onde A e P são, respectivamente, a área e o perímetro do objecto complexo.

Este parâmetro tem como vantagens a sua insensibilidade a rotações, translações e ponto de começo, apresentando a desvantagem de ser sensível a mudanças de escala do objecto.

2.3.1.2 Medidas de semelhança para os parâmetros geométricos

As medidas de semelhança para os parâmetros geométricos baseados em regiões não diferem das medidas Euclidianas e de *Minkowsky* definidas anteriormente. Especificamente para a *bounding box*, existe uma outra medida de semelhança que depende do factor de forma [Tekt1999], F , da *bounding box* definido por:

$$F = \frac{B_A}{B_L} \quad (2.35)$$

onde B_A e B_L são, respectivamente, a altura e a largura da *bounding box*.

Esta medida de semelhança, M_B , é definida como o valor máximo entre a diferença absoluta dos factores de forma do objecto fornecido pelo mecanismo de procura, F_Q , e os factores de forma para os objectos descritos disponíveis, F_I , e a diferença absoluta entre o inverso de cada um dos factores de forma. A razão pela qual se calcula a medida de semelhança do parâmetro deste modo, deve-se ao facto de se ter que compensar possíveis rotações do objecto.

$$M_B = \max \left\{ \left| F_I - F_Q \right|, \left| \frac{1}{F_I} - \frac{1}{F_Q} \right| \right\} \quad (2.36)$$

2.3.2 Parâmetros de forma baseados em vectores próprios multi-nível

O parâmetro de forma baseado em vectores próprios multi-nível (*MLEV* do inglês *Multi-Layer Eigen Vectors*) é uma descrição da forma baseada numa estrutura multi-escalar, onde cada nível introduz mais detalhe relativamente à forma do objecto. Esta estrutura multi-escalar é obtida através de uma divisão recursiva de cada sector do objecto complexo segundo os seus eixos principais [KiKi1999c][KiKi1999d].

2.3.2.1 Definição dos parâmetros baseados em vectores próprios multi-nível

O parâmetro de forma *MLEV* obtém-se a partir de várias partições do objecto. Sendo $I_i = [x_i \ y_i]^T$ o vector de localização do *shapel* i e N o número total de *shapels* pertencentes ao objecto complexo. Esta sub-divisão da forma do objecto é então feita em três etapas:

1. Cálculo do centro de massa, m_L :

$$m_L = \frac{1}{N} \sum_{i=1}^N I_i \quad (2.37)$$

da matriz de covariância, C_L :

$$C_L = \frac{1}{N} \sum_{i=1}^N I_i I_i^T - m_L m_L^T \quad (2.38)$$

e dos vectores próprios e_j , assim como dos seus valores próprios λ_j :

$$C_L e_j = \lambda_j e_j \quad (2.39)$$

onde $j=1,2$.

2. Sub-divisão da forma em quatro sectores segundo a direcção dos eixos principais correspondentes aos vectores próprios e_1 e e_2 no centróide do objecto complexo. A Figura 2.20a) mostra a divisão do objecto complexo em quatro sectores R_1 , R_2 , R_3 e R_4 .
3. Para cada um dos quatro sectores anteriores, a primeira e a segunda etapa são repetidas, gerando quatro novos sectores até que o nível de detalhe requerido pela aplicação seja atingido. A Figura 2.20b) mostra o resultado dos novos sectores (R_5 - R_8 , R_9 - R_{12} , R_{13} - R_{16} e R_{17} - R_{20}) gerados pelo método.

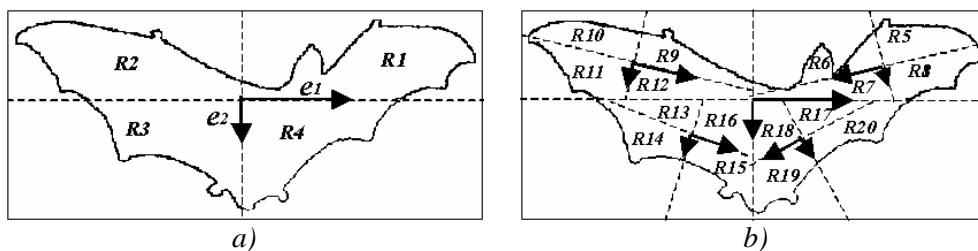


Figura 2.20 – *Bat*: Exemplos de sub-divisão do objecto a) Após primeira sub-divisão, b) Após segunda sub-divisão [KiKi1999d]

Para cada um dos sectores do objecto complexo serão calculados quatro parâmetros de forma invariantes à translação, rotação e mudança de escala, nomeadamente:

- ♦ **Relação entre os valores próprios** – Relação entre o valor próprio do eixo principal maior ($\lambda_{M,i}$) e o valor próprio do eixo principal menor ($\lambda_{m,i}$):

$$relação(i) = \frac{\lambda_{M,i}}{\lambda_{m,i}} \quad (2.40)$$

- ♦ **Ângulo normal** – Menor ângulo de valor absoluto entre o vector próprio do eixo principal maior do sector i , ($e_{M,i}$), e o vector próprio do eixo principal maior do sector raiz, equivalente ao objecto original, ($e_{M,r}$);
- ♦ **Centro** – Distância normalizada segundo o valor próprio maior, $\lambda_{M,r}$, entre c_i , o centro do sector i e c_r , o centro do sector raiz:

$$centro(i) = \frac{\|c_i - c_r\|}{\lambda_{M,r}} \quad (2.41)$$

- ♦ **Solidez** – Relação entre A_i , a área do sector i , e A_{bb} , a área da *bounding box* correspondente ao sector i :

$$solidez(i) = \frac{A_i}{A_{bb}} \quad (2.42)$$

O parâmetro de forma baseado em vectores próprios multi-nível é então constituído pelos quatro parâmetros acima descritos, sendo que para o sector raiz apenas a relação entre os valores próprios e a solidez são calculados. Este parâmetro é insensível a transformações geométricas como a mudança de escala, a rotação e a translação.

2.3.2.2 Medidas de semelhança para o parâmetro baseado em vectores próprios multi-nível

A medida de semelhança sugerida para o parâmetro *MLEV* é simplesmente a soma das diferenças absolutas dos quatro elementos (distância entre valores próprios, ângulo normal, centro e solidez) acima definidos para todos os sectores definidos ou seja a distância de *Minkowsky* [KiKi1999c][KiKi1999d].

2.3.3 Parâmetros de forma baseados em transformadas

Os parâmetros baseados em transformadas que se apresentam de seguida têm sido utilizados em vários domínios de aplicação, nomeadamente para a descrição da forma de uma região ou regiões disjuntas de um objecto complexo.

2.3.3.1 Definição de parâmetros baseados em transformadas

A transformada aqui apresentada como parâmetro de descrição da forma é denominada transformada Angular-Radial [KiKi2000].

▪ Transformada Angular-Radial (ART)

A transformada *ART* baseia-se numa transformada ortogonal unitária definida sobre um disco circular unitário. A transformada *ART* aplicada à descrição da forma implica a normalização da forma a um disco circular de diâmetro unitário, já que os seus coeficientes se encontram definidos sobre um disco circular do mesmo tipo. Assim sendo, o disco de raio unitário R é definido de modo a que a forma do objecto complexo seja completamente envolvida pelo disco, ou seja, desde o seu centróide ao seu *shapel* mais extremo (a forma normalizada terá como área πR^2 *shapels*). De seguida, extraem-se os coeficientes *ART*, sendo as suas amplitudes usadas como o vector de valores para o parâmetro de forma.

O cálculo dos coeficientes *ART*, T_{nm} , é efectuado através da expressão (2.43):

$$T_{nm} = \int_0^{2\pi} \int_0^1 B_{nm}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta \quad (2.43)$$

onde n é um número inteiro positivo ou zero, m é um número positivo ou negativo, (ρ, θ) são coordenadas polares, e $B_{nm}^*(\rho, \theta)$ é a função base sinusoidal e ortonormal completa da transformada *ART* de ordem (n, m) definida sobre o disco circular de dimensão unitária e $f(\rho, \theta)$ é a função do suporte binário da forma do objecto em função de ρ e θ . A Figura 2.21 mostra a parte real da função de base *ART*, semelhante à da parte imaginária, calculadas usando a expressão (2.44):

$$B_{nm}^*(\rho, \theta) = A_m(\theta) R_n(\rho), \quad \rho \leq 1 \quad (2.44)$$

onde $A_m(\theta)$ é o polinómio angular *ART* definido de acordo com a expressão (2.45):

$$A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta) \quad (2.45)$$

e $R_n(\rho)$ é o polinómio radial *ART* definido de acordo com a expressão (2.46).

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2 \cos(\pi n \rho) & n \neq 0 \end{cases} \quad (2.46)$$

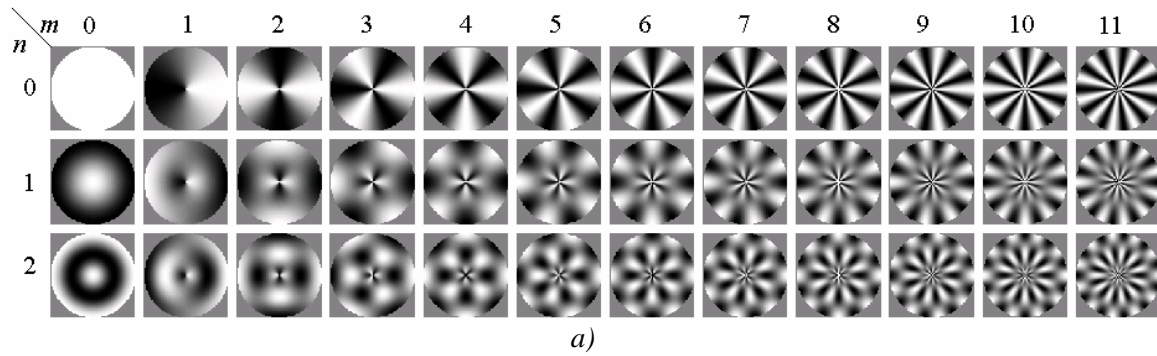


Figura 2.21 – Parte real da função de base da transformada ART [KiKi2000]

Este parâmetro permite a descrição multi-escalar da forma do objecto complexo através de um certo número de coeficientes, de modo que quanto maior for a ordem da descrição, ou seja, mais os coeficientes usados maior será o detalhe usado na descrição. A transformada ART é insensível a transformações geométricas como, por exemplo, a rotação, a mudança de escala e o ponto de começo.

2.3.3.2 Medidas de semelhança para o parâmetro baseado em transformada

A medida de semelhança usada para a transformada ART, D_{ART} (ver expressão (2.47)) [KiKi2000], é a soma ponderada das diferenças absolutas de cada coeficiente no parâmetro de forma:

$$D_{ART} = \sum_{i=0}^{N-1} W_i \times \|T_I[i] - T_Q[i]\| \quad (2.47)$$

onde N é o número máximo de coeficientes no parâmetro e $T_I[i]$ e $T_Q[i]$ são o i -ésimo coeficiente do parâmetro da forma correspondente às formas descritas disponíveis e à forma fornecida ao mecanismo de procura. O peso W_i poderá ser a variância de cada coeficiente calculado para as formas previamente descritas ou simplesmente 1.

2.3.4 Parâmetros de forma baseados em momentos

Os momentos têm sido utilizados em vários domínios de aplicação, nomeadamente para a descrição da forma de uma região ou regiões disjuntas de um objecto complexo.

2.3.4.1 Definição de parâmetros baseados em momentos

Os momentos aqui descritos como parâmetro de descrição da forma são denominados momentos de *Zernike* [TeCh1988][KhHo1990][KiKi1999a] [KiKi1999b].

▪ Momentos de *Zernike*

Os momentos de *Zernike* baseiam-se em polinómios que permitem que estes momentos sejam independentes e facilmente calculados até uma ordem elevada. Os momentos de *Zernike* aplicados à representação da forma implicam a normalização da forma a um disco circular de diâmetro unitário, já que os momentos de *Zernike* se encontram definidos sobre um disco circular do mesmo tipo. Assim sendo, o disco de raio unitário R é definido de modo a que a forma do objecto complexo seja completamente envolvida pelo disco, ou seja, desde o seu centróide ao seu *shapel* mais extremo (a forma normalizada terá como área πR^2 *shapels*). De seguida, extraem-se os momentos de *Zernike* relativos à forma do objecto sendo os seus valores usados como o vector de valores para o parâmetro de forma.

O cálculo dos momentos de *Zernike*, Z_{nm} , é efectuado através da expressão (2.48):

$$Z_{nm} = \frac{n+1}{\pi} \sum_{\text{disco unitário}} B_{nm}^*(\rho, \theta) f(\rho, \theta) \quad (2.48)$$

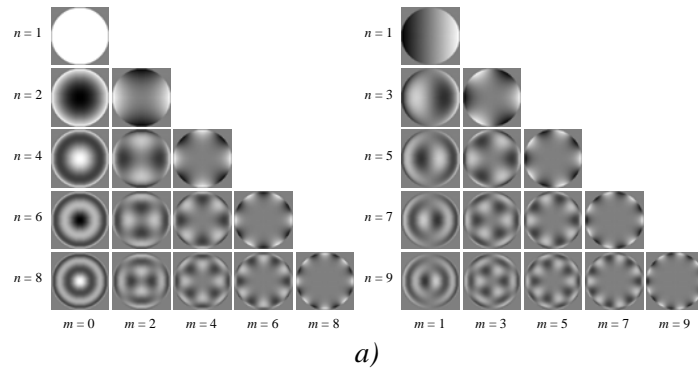
onde n é um número inteiro positivo ou zero, m é um número positivo ou negativo, (ρ, θ) são coordenadas polares, $B_{nm}^*(\rho, \theta)$ é a função base de *Zernike* de ordem (n, m) definida sobre o disco circular de dimensão unitária e $f(\rho, \theta)$ é a função do suporte binário da forma do objecto em função de ρ e θ . A Figura 2.22 mostra as funções de base de *Zernike* calculados, até à ordem 8, usando a expressão (2.49):

$$B_{nm}^*(\rho, \theta) = R_{nm}(\rho) e^{jm\theta}, \quad \rho \leq 1 \quad (2.49)$$

onde $R_{nm}(\rho)$ é o polinómio radial de *Zernike* definido de acordo com a expressão (2.55):

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s \right)! \left(\frac{n-|m|}{2} - s \right)!} \rho^{n-2s} \quad (2.55)$$

e onde $n-|m|$ é um número par maior que zero.



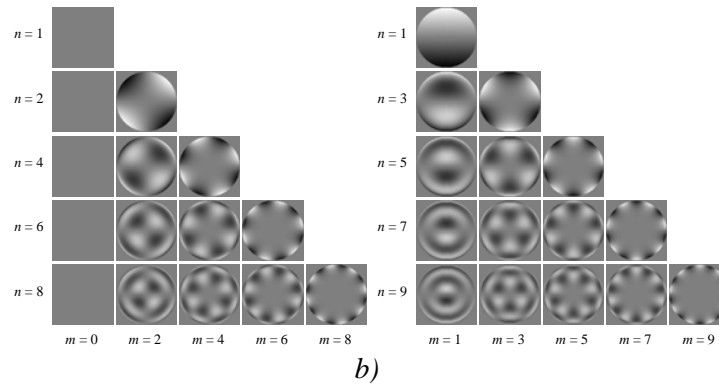


Figura 2.22 – Funções de base de *Zernike* até à ordem 8: a) parte real b) parte imaginária [KiKi1999b]

Este parâmetro permite ainda a descrição multi-escalar da forma do objecto complexo através de um conjunto de momentos, de tal modo que quanto maior for a ordem ou seja maior o número de momentos usados maior será o detalhe usado. Os momentos de *Zernike* são insensíveis a transformações geométricas como, por exemplo, a rotação, a mudança de escala e o ponto de começo.

2.3.4.2 Medida de semelhança para o parâmetro baseado em momentos

A medida de semelhança usada para os momentos de *Zernike*, D_Z (ver expressão (2.50)) [KiKi1999a][KiKi1999b], é a soma ponderada das diferenças absolutas de cada momento no parâmetro de forma:

$$D_Z = \sum_{i=0}^{N-1} W_i \times \|M_I[i] - M_Q[i]\| \quad (2.50)$$

onde N é o número máximo de momentos de *Zernike* no parâmetro e $M_I[i]$ e $M_Q[i]$ são o i -ésimo momento do parâmetro da forma correspondente à forma fornecida ao mecanismo de procura e correspondente às formas descritas disponíveis. O peso W_i poderá ser a variância de cada momento calculada para as formas previamente descritas ou simplesmente 1.

2.4 Conclusões

O estudo dos parâmetros de descrição de forma feito neste capítulo veio evidenciar a sua importância no domínio da análise de imagem, especialmente na descrição de conteúdo visual composto por objectos de forma arbitrária, e permitir a familiarização com as suas definições matemáticas e correspondentes medidas de semelhança. A revisão bibliográfica dos parâmetros de forma permitiu:

1. Analisar o significado das expressões matemáticas que definem os parâmetros e, nalguns casos, visualizar graficamente o seu resultado;

2. Avaliar as propriedades dos parâmetros de descrição em relação às transformações geométricas mais comuns, nomeadamente translação, rotação, mudança de escala e ponto de começo; e
3. Estabelecer medidas de semelhança adequadas para os vários parâmetros de modo a permitir a avaliação da sua eficácia em termos da procura de formas semelhantes.

Como conclusão do estudo feito, apresentam-se nas Tabela 2.2 e Tabela 2.3, de forma resumida, as principais propriedades dos parâmetros de forma baseados no contorno e na região relativamente à sua insensibilidade a transformações geométricas (T - translações, R - rotações, ME - mudança de escala e PC - ponto de começo). Para além disso, indica-se se as medidas de semelhança mais comuns, D_E – distância Euclidiana e D_M – distância de *Minkowsky*, podem ser ou não usadas para cada descritor.

O símbolo ✓ representa a existência de insensibilidade a uma das transformações geométricas ou a possibilidade de usar uma das medidas de semelhança, caso contrário, na ausência de uma destas propriedades usar-se-á o símbolo ✕.

Analisando-se estas tabelas pode-se concluir que a maioria dos parâmetros de forma são insensíveis às transformações geométricas consideradas; noutros casos (CSS e ângulos de curvatura), é possível obter essa insensibilidade através do uso de medidas de semelhança adequadas.

Parâmetros de forma baseados no contorno	Transformações geométricas				Medidas de semelhança		
	T	R	ME	PC	D_E	D_M	Outras
Geométricos							
Perímetro	✓	✓	✕	✓	✓	✓	✕
Corda Máxima	✓	✓	✕	✓	✓	✓	✕
Circularidade	✓	✓	✓	✓	✓	✓	✕
Convexidade	✓	✓	✓	✓	✓	✓	✕
Excentricidade	✓	✓	✓	✓	✓	✓	✕
Transformadas							
<i>Fourier</i>	✓	✓	✓	✓	✓	✓	✓
<i>Wavelets</i>	✓	✕	✓	✓	✓	✓	✕
Momentos invariantes	✓	✓	✓	✓	✓	✓	✕
CSS	✓	✓	✓	✕	✕	✕	✓
Contornos normalizados	✓	✓	✓	✓	✓	✓	✕
Ângulos de curvatura	✓	✓	✓	✕	✓	✓	✕

Tabela 2.2 – Resumo das propriedades dos parâmetros de forma baseados no contorno

Parâmetros de forma baseados na região	Transformações geométricas				Medidas de semelhança		
	T	R	ME	PC	D _E	D _M	Outras
Geométricos							
<i>Bounding box</i>	✓	✓	✗	✓	✓	✓	✗
Área	✓	✓	✗	✓	✓	✓	✗
Centróide	✗	✓	✓	✓	✓	✓	✗
Projecções	✓	✓	✗	✓	✓	✓	✗
Diâmetro circular equivalente	✓	✓	✓	✓	✓	✓	✗
Solidez	✓	✓	✓	✓	✓	✓	✗
Compactação	✗	✓	✓	✓	✓	✓	✗
Transformadas							
<i>ART</i>	✓	✓	✓	✓	✗	✓	✗
<i>Fourier</i>	✓	✓	✓	✓	✓	✓	✓
Momentos de Zernike	✗	✓	✓	✓	✗	✓	✗
<i>MLEV</i>	✓	✓	✓	✓	✗	✓	✗

Tabela 2.3 – Resumo das propriedades dos parâmetros de forma baseados na região

Uma outra propriedade interessante para a descrição da característica de forma é a possibilidade de usar vários níveis de detalhe, nomeadamente através de uma representação escalável; este tipo de requisito depende da aplicação. Dos parâmetros analisados, os que permitiam este tipo de descrição foram o parâmetro *CSS*, a transformada *ART* e os momentos de *Zernike*.

De entre os parâmetros de descrição de forma apresentados foram propostos ao MPEG-7 os seguintes parâmetros: transformadas *wavelet*, *CSS*, transformada *ART*, momentos de *Zernike* e *MLEV*. Os parâmetros que de entre estes foram seleccionados pelo MPEG-7 como mais prometedores serão estudados em detalhe no capítulo 5, nomeadamente em termos da sintaxe e semântica da sua representação. Com a especificação completa da sintaxe e da semântica, passar-se-á a falar de descritores e já não de parâmetros de descrição (nos quais não se inclui a especificação sintáctica).

Capítulo 3

Mecanismos de descrição e procura

Estando-se, actualmente, perante uma explosão da produção, profissional e doméstica, de conteúdos audiovisuais é natural que surja a necessidade de descrever, armazenar e procurar todos estes conteúdos, eficiente e eficazmente, a fim de poderem ser disponibilizados e usados nos mais diversos domínios de aplicação do processamento de imagem e vídeo. Mas a enorme dificuldade em aceder aos conteúdos audiovisuais pretendidos pode facilmente desmotivar qualquer consumidor menos determinado. Para minimizar esta dificuldade, foram surgindo ao longo dos anos, nomeadamente na *Internet*, vários mecanismos de descrição e de procura de imagens e vídeo. Entre os mecanismos disponíveis na *Internet* mais interessantes contam-se o *NETRA* [NETRA], o *Query by image content (QBIC)* [QBIC_Art][QBIC_Trade][QBIC_Stamps], o *Multimedia analysis and retrieval system (MARS)* [MARS], o *VideoQ* [VIDEOQ], o *Photobook* [Photobook], o *Virage* [Virage], o *VisualSeek* [VisualSEEK], o *Excalibur* [Excalibur], o *WebSeek* [Webseek], o *Candid* [CANDID], o *Chabot* e o *Viscomp* [Bimb1999][Ma1997][Viscomp]. Estes mecanismos têm como objectivos:

- Descrever conteúdos visuais (ou audiovisuais) de acordo com as suas características de forma, cor, textura, relação espacial e movimento;
- Armazenar os descritores gerados numa base de dados (para acesso directo ou posterior transmissão), segundo estruturas de descrição conhecidas (esquemas de descrição); e
- Permitir procurar conteúdos audiovisuais nas bases de dados disponíveis de acordo com as necessidades do utente expressas através de exemplos ou através de parâmetros de descrição ou descritores adequados para a procura em questão.

As primeiras abordagens a estes mecanismos basearam-se na descrição textual da informação audiovisual, dando origem a anotações textuais, como no caso do mecanismo *WebSeek* [Webseek]. Esta abordagem não só não era eficaz e eficiente porque consumia muito tempo, como podia ser extremamente subjectiva e não se adequar a alguns tipos de procura. O facto de a anotação implicar sempre alguma subjectividade e a suposição de eventuais aplicações, limitava largamente à partida as possibilidades de encontrar conteúdo audiovisual para uso em aplicações não inicialmente previstas. Este facto fez crescer a necessidade de desenvolver mecanismos que descrevessem os dados de um modo menos subjectivo e mais relacionado com a percepção humana [Bimb1999]. Como consequência, surgiram nos anos mais recentes vários mecanismos que permitem descrever e procurar informação audiovisual armazenada remotamente, p.e. numa base de dados, de acordo com as suas características audiovisuais de baixo-nível. Ainda que os resultados deste tipo de descrição sejam prometedores, há ainda muito por fazer a fim de completar, melhorar e otimizar os mecanismos já existentes.

Este capítulo debruçar-se-á sobre mecanismos de descrição e procura que usam apenas características audiovisuais de baixo-nível. Depois da apresentação de uma arquitectura genérica para estes mecanismos, seguir-se-á a descrição de alguns dos mecanismos mais interessantes disponíveis na *Internet*, concluindo-se com uma análise comparativa entre a arquitectura genérica proposta e a arquitectura dos mecanismos apresentados. Uma vez que este trabalho se debruça especialmente sobre a característica de forma, é natural que os mecanismos sejam mais detalhadamente analisados e comparados em tudo o que diz respeito à característica de forma.

3.1 Arquitectura de um mecanismo de descrição e procura

As novas abordagens dos mecanismos de descrição e procura suportando descrições e procuras através de características audiovisuais, permitem hoje em dia o acesso aos conteúdos não só ao nível conceptual, usando palavras-chave (texto), mas também ao nível perceptual, usando parâmetros de descrição que funcionam como medidas objectivas das características. Exemplos relevantes disso mesmo são, por exemplo, os momentos de *Zernike* aplicados à forma, os histogramas aplicados à cor ou os vectores de movimento aplicados ao movimento. O uso combinado de vários tipos de parâmetros e de medidas de semelhança adequadas permite procuras mais ou menos elaboradas, muitas vezes sem que o utente tenha sequer consciência que está a usar este tipo de abordagem.

Actualmente, a caracterização das aplicações é muitas vezes feita, nomeadamente pelo MPEG-7, de acordo com o modo e local onde se faz a comparação das descrições, junto ao utente ou remotamente ao utente. Após a revisão dos mecanismos existentes e da literatura respeitante a esta área, concluiu-se que a categorização pode ser feita, de forma simplificada, essencialmente em dois tipos de aplicações: aplicações do tipo *pull* e do tipo *push*. Dos dois tipos de aplicações referidos, as aplicações do tipo *Pull* são as de maior interesse no âmbito deste trabalho por serem aquelas onde mais facilmente se usa a característica de forma para descrever a informação visual. Assim sendo, propõe-se de seguida uma arquitectura genérica para os mecanismos de descrição e procura do tipo *Pull*, com o objectivo de ajudar a definir os diferentes módulos dos mecanismos deste tipo. A Figura 3.1 apresenta a arquitectura proposta e usa as linhas tracejadas para definir as fronteiras entre os três módulos principais que a compõem:

- **Módulo de descrição** – Módulo que faz a descrição da informação audiovisual através das suas características audiovisuais e parâmetros correspondentes; as descrições são então armazenadas no módulo da base de dados para posterior comparação;
- **Módulo da base de dados** – Módulo onde as descrições da informação audiovisual geradas pelo módulo de descrição são armazenadas e onde se faz a comparação dessas descrições com a descrição (ou apenas parâmetros) disponibilizada pelo módulo de procura através de medidas de semelhança adequadas;
- **Módulo de procura** – Módulo de interface onde: i) o utente fornece informação que caracteriza de algum modo o conteúdo que se pretende procurar; essa informação pode ser fornecida através de um exemplo, através de alguns parâmetros de descrição ou através de outras propriedades mais intuitivas para o utente e que o mecanismo se encarrega de traduzir de forma conveniente em termos de parâmetros de descrição; nalguns mecanismos, esta formulação pode ser posteriormente refinada através de *feedback* baseado nos resultados obtidos; ii) são calculados os parâmetros de descrição para a informação de procura dada pelo utilizador; os parâmetros assim obtidos ou aqueles directamente fornecidos são então enviados ao módulo da base de dados para comparação com as descrições disponíveis (aplicação tipo *pull*); iii) são visualizados os resultados da procura feita através de comparação com as descrições disponíveis na(s) base(s) de dados, podendo ainda este módulo aceitar *feedback* do utente para refinamento da procura. Exemplos: no caso de fotografias em venda, pode ser visualizada uma versão de menor qualidade/resolução das fotografias; no caso de vídeo, pode ser visualizado um conjunto de tramas chave representativas do vídeo.

Estes três módulos serão analisados pormenorizadamente nas sub-secções que se seguem.

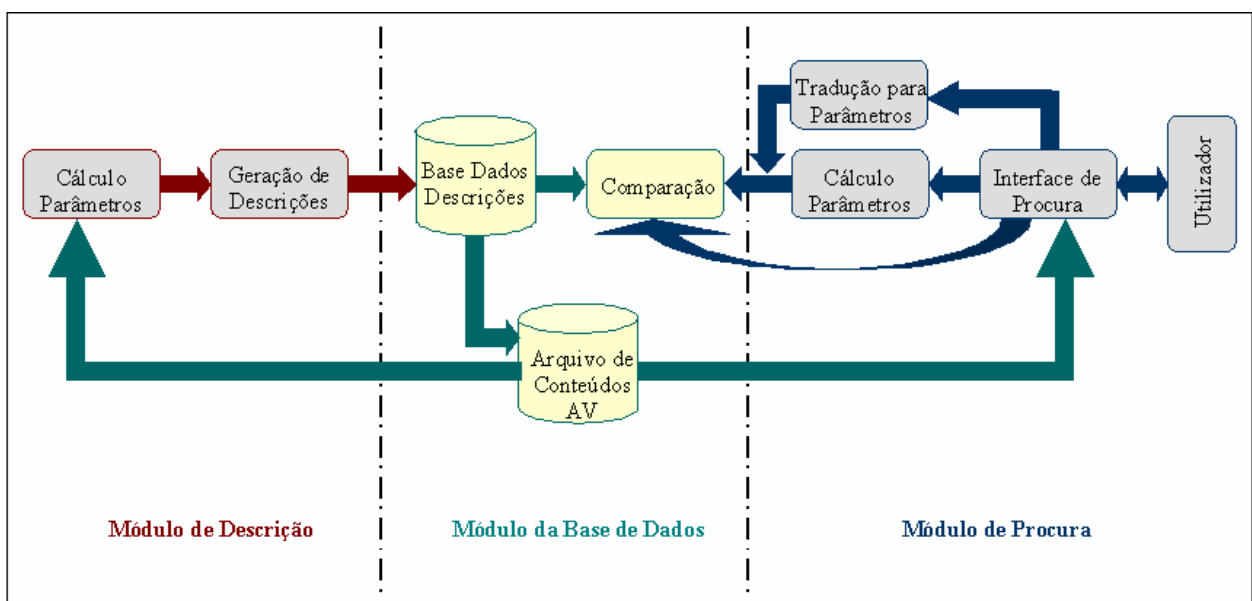


Figura 3.1 – Arquitectura genérica de um mecanismo de descrição e procura para aplicações do tipo *Pull*

3.1.1 Módulo de descrição

O módulo de descrição da arquitectura genérica proposta é o módulo onde a informação visual é processada para que sejam calculados os parâmetros que representam as suas características de forma, cor, textura, relação espacial (para imagens estáticas e vídeo); movimento de câmara e de objectos (para vídeo) ou timbre, tom e pico (para áudio). De seguida, as descrições são enviadas para o módulo da base de dados para que possam, no futuro, ser comparadas de acordo com as necessidades dos utentes expressas através do módulo de procura. O módulo de descrição proposto é composto por dois grandes blocos:

- **Extracção dos parâmetros** – Bloco responsável pela extracção dos parâmetros para as características a utilizar. Os valores extraídos para estes parâmetros representarão numericamente as características em análise;
- **Geração da descrição** – Bloco responsável pela geração das descrições para o conteúdo em questão, a partir dos parâmetros já calculados e usando uma dada estrutura de descrição [Bimb1999] [MPEG7-XM1999]; a geração de uma descrição (com representação textual ou binária) requer:
 - ♦ **Sintaxe** – Especifica as regras (gramática) que determinam a ordem (posição) de todos os componentes numa descrição, nomeadamente para os descritores e para a estrutura de descrição;
 - ♦ **Semântica** – Define o significado dos vários componentes da descrição bem como outras regras a que devem obedecer para além da posição na descrição especificada pela sintaxe, p.e. conjunto de valores possíveis para um descritor.

As normas para descrição de informação audiovisual especificam necessariamente a sintaxe e a semântica a usar para as descrições geradas já que isso é essencial para garantir a interoperabilidade entre os vários mecanismos usando a mesma norma. É o caso da futura norma MPEG-7 que especificará estes elementos mas não o bloco de geração da descrição propriamente dito.

As descrições geradas são de seguida armazenadas para que futuramente possam ser usadas pelo mesmo ou outros mecanismos de descrição e procura. Estes mecanismos de procura terão, em princípio, que conhecer a sintaxe e a semântica das descrições armazenadas, para que possam expressar as propriedades da informação que procuram através da mesma ‘linguagem’ usada para as descrições disponíveis. O uso da mesma ‘linguagem de descrição’ permite mais facilmente a comparação entre as descrições disponíveis e os parâmetros que caracterizam a informação procurada.

3.1.2 Módulo da base de dados

O módulo da base de dados é o módulo do mecanismo de descrição e procura onde se encontram armazenadas as descrições dos conteúdos audiovisuais indexados/descritos. O próprio conteúdo poderá estar ou não armazenado na mesma base de dados das descrições, consoante os requisitos da aplicação em causa. É também neste módulo que as descrições

provenientes do módulo de procura e as armazenadas na base de dados (geradas pelo módulo de descrição) são comparadas, através de medidas de semelhança adequadas aos parâmetros usados.

De acordo com as suas funções, este módulo poderá incluir três blocos principais, descritos de seguida:

- **Arquivo de conteúdos audiovisuais** – Armazena a informação audiovisual descrita e que pode mais tarde ser disponibilizada aos utentes, através ou não, do módulo de procura. Este arquivo pode ser de vários tipos, consoante o formato dos conteúdos que armazena, influenciando de forma determinante o processo de extracção de parâmetros e de disponibilização dos conteúdos ao utilizador. Exemplos: arquivo electrónico com conteúdo codificado em MPEG-1, 2 ou 4, arquivo de filmes em celulóide, arquivo de fotografias em papel, arquivo de áudio em fita (digital ou analógico). Como é evidente, o conteúdo audiovisual encontrado só poderá ser imediatamente disponibilizado ao utente através do módulo de procura se encontrar em formato electrónico (e a banda o permitir).
- **Base de dados de descrições** – Armazena de forma altamente organizada, segundo estruturas de indexação adequadas, as descrições da informação audiovisual geradas pelo módulo de descrição. A organização eficiente desta base de dados é feita através do recurso a estruturas de indexação, como por exemplo as *hash tables*, *k-d trees*, *R-trees*, *R⁺-trees*, *R^{*}-trees* e *SS-trees* [Bimb1999]. A utilização destas estruturas de indexação permite acelerar o processo de comparação das descrições que é determinante para a satisfação do utente.
- **Comparação** – Bloco fundamental do mecanismo de descrição e procura por ser o responsável pela definição do significado de ‘semelhança’ usando os parâmetros de descrição disponíveis. Este significado determina fortemente os resultados da procura e logo o nível de satisfação das necessidades do utente. O significado de ‘ser semelhante’ é expresso através de uma medida de semelhança entre os valores dos parâmetros para as descrições disponíveis na base de dados e os valores dos parâmetros calculados ou fornecidos pelo utilizador no processo de procura. Exemplos destas medidas de semelhança são a distância Euclidiana e a distância de *Minkowsky*.

Após o cálculo da medida de semelhança, a decisão sobre a semelhança entre as descrições disponíveis na base de dados e a descrição fornecida pelo módulo de procura, faz-se segundo dois critérios de comparação [GaGr1992]:

- ♦ **Critério exacto** – A comparação só é bem sucedida se os parâmetros de descrição para as duas descrições acima referidas são totalmente iguais ou seja a medida de semelhança (que também é uma medida da diferença) terá que ser nula o que indica uma semelhança perfeita em termos dos parâmetros usados na comparação.
- ♦ **Critério aproximado** – A comparação é bem sucedida para o conjunto das descrições mais semelhantes ou seja com menores valores para a medida de semelhança, definido através de um limiar para o valor da medida de semelhança; neste caso, encontra-se na base de dados conteúdo que é bastante semelhante ao procurado, segundo os parâmetros definidos

como relevantes, ainda que não completamente igual; como seria de esperar, este é o critério usado mais frequentemente.

Consequentemente, este é o módulo da arquitectura genérica responsável por toda a gestão das descrições. Tanto o módulo de descrição como o módulo de procura terão forçosamente que ser compatíveis entre si, nomeadamente em termos dos parâmetros usados, de modo a conseguirem ‘comunicar’ através do bloco de comparação. Assim sendo, os valores para os parâmetros gerados por estes módulos poderão ser comparados a fim de propor ao utilizador do mecanismo de descrição e procura um conjunto de conteúdos audiovisuais que poderão ser a resposta às suas necessidades.

3.1.3 Módulo de procura

O módulo de procura de um mecanismo de descrição e procura é o módulo responsável pela interface com o utente – formulação da procura e visualização dos resultados - e normalmente também pelo cálculo dos parâmetros de descrição que expressam a procura no caso desses parâmetros não serem directamente fornecidos.

No seguimento deste trabalho, designar-se-á por interface de procura o conjunto dos dois blocos onde se dá a interacção entre o utente e o mecanismo ou seja os blocos de formulação da procura e de visualização dos resultados. A sub-secção que se segue pretende definir mais pormenorizadamente os requisitos a considerar no desenvolvimento de uma interface de procura, os níveis de abstracção envolvidos na procura e os tipos de procura existentes.

3.1.3.1 Interface de procura

Devido ao papel fundamental da interface de procura no contexto de qualquer mecanismo deste tipo, nomeadamente do modo como o utente pode expressar as suas necessidades em termos de conteúdo vídeo, far-se-á de seguida uma análise mais pormenorizada de vários aspectos associados à formulação da procura e à visualização dos resultados bem como à interface correspondente.

Como imediatamente se perceberá, é a interface de procura que estabelece o tipo de interacção/relação entre o utilizador e o mecanismo, através da definição dos modos como se pode expressar a procura e dos modos como se podem visualizar os resultados, eventualmente fornecendo ainda *feedback* a partir deles. Normalmente, a interface permitirá a navegação simples do utilizador pela base de dados (simples visualização sem qualquer procura associada) bem como a navegação orientada/selectiva através da formulação da procura e, possivelmente, vários tipos de *feedback*.

Como é natural, as interfaces de procura de cada mecanismo são fortemente determinadas pelo tipo de aplicação e o nível de conhecimentos e experiência dos utilizadores à qual se destinam. Num âmbito mais geral, os mecanismos de descrição e procura têm que se adequar a um largo número de utilizadores, desde utilizadores com pouca ou nenhuma experiência (amadores), até utilizadores com muita experiência (profissionais). Consequentemente, e para que o acesso à informação visual seja eficiente e eficaz no sentido de satisfazer as necessidades específicas de cada utente, a interface de procura deve considerar os seguintes requisitos [Bimb1999]:

- Ser o mais visual possível (menos textual) ou seja mais adequada ao tipo de informação que se procura;
- Possuir vários níveis de complexidade para se poder adaptar ao tipo de utente;
- Possibilitar a compreensão e a navegação pela informação visual na base de dados;
- Possibilitar procuras de nível sintáctico e semântico ou seja mais de baixo-nível e mais conceptuais; e
- Possibilitar a integração de outras ferramentas de descrição e procura.

Na sequência da evolução dos mecanismos de descrição e procura e da inclusão de parâmetros de descrição de baixo nível, distinguem-se hoje dois grandes tipos de procura consoante o nível de abstracção envolvido na procura [ChEl1998]:

- **Procura de nível sintáctico** – Procura essencialmente baseada no uso de parâmetros de baixo-nível, requerendo algumas vezes, por parte dos utilizadores, uma boa familiaridade com as ferramentas de descrição e com o significado dos valores dos parâmetros calculados no módulo de descrição. Exemplo: procuras baseadas na forma dos objectos, por exemplo objectos bem circulares, na quantidade de movimento ou na cor; e
- **Procura de nível semântico** – Procura baseada em conceitos mais abstractos e parâmetros de mais alto nível (podendo eventualmente o mecanismo fazer internamente o mapeamento em parâmetros de mais baixo-nível), explorando os conhecimentos adquiridos pelo utilizador no ‘mundo real’. Exemplo: a procura de uma imagem alegre ou com chuva.

Uma outra forma de classificar as procuras é através do tipo de informação que o utilizador usa para especificar a procura. Entre as formas mais relevantes destacam-se:

▪ **Procuras baseadas em estatísticas do sinal**

As procuras baseadas em estatísticas do sinal são aquelas em que o utilizador insere valores estatísticos, percentagens ou outros valores numéricos associados aos parâmetros usados para a comparação, a fim de definir o que procura. Este tipo de procura baseia-se essencialmente na definição de valores para parâmetros de baixo-nível, significando que se encontra mais orientada para utilizadores com um bom conhecimento das ferramentas de descrição e da informação armazenada na base de dados, ou seja, é uma procura de nível sintáctico. Neste tipo de procura, o utilizador poderá usar apenas uma característica ou várias características para compor a sua procura, consoante as características usadas pelo mecanismo para descrever a informação audiovisual.

Essencialmente, as procuras baseadas em estatísticas do sinal podem ser especificadas de dois modos:

- **Valores, pesos e outros** – Define-se a procura através de um conjunto de valores associados a cada parâmetro, nomeadamente a gama de valores relevantes, o peso desses parâmetros em relação aos outros, etc.; exemplo: o peso de uma

característica em relação a outra no acto da procura para as quais se deu também a gama relevante de valores.

- **Gráficos** – Define-se a procura através das propriedades da representação gráfica de certos parâmetros; exemplo: definindo-se algumas propriedades do histograma de cor para a característica cor.

A Figura 3.2 mostra um exemplo de uma procura baseada na estatística do sinal. A Figura 3.2 mostra um exemplo de uma procura usando apenas a cor como característica visual. Nesta interface de procura usando apenas a cor como característica de descrição, constrói-se a procura definindo os pesos para cada cor escolhida (através dos seus valores RGB ou de um mostruário de cores), o que significa que se procuram imagens com uma composição de cores semelhante.

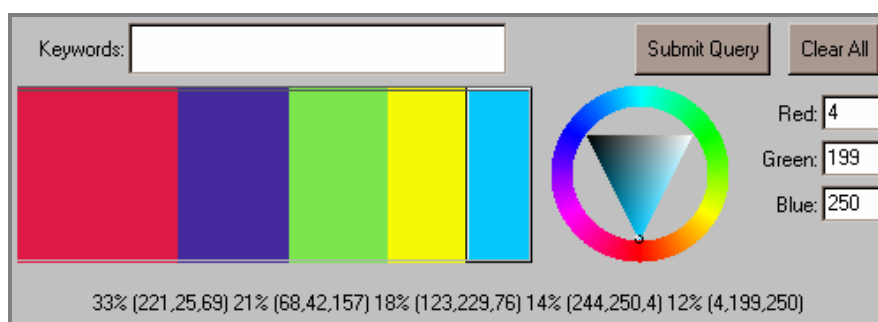


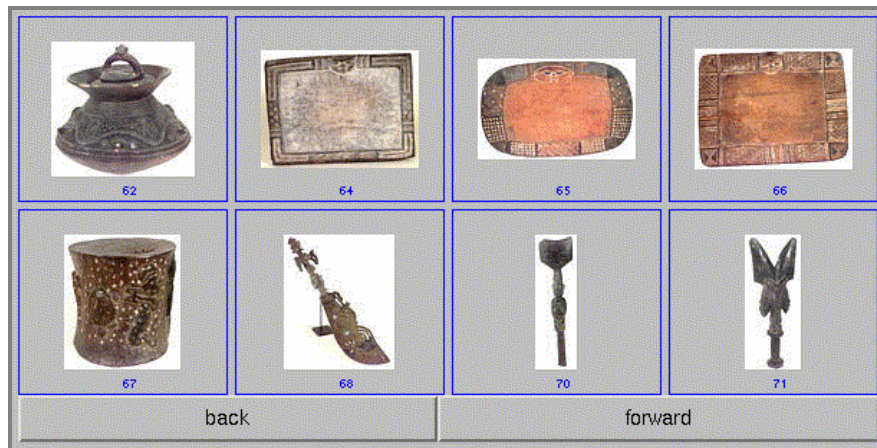
Figura 3.2 – Exemplo de procura por estatística do sinal, nomeadamente percentagens ou outros valores numéricos [QBIC_Stamps]

▪ **Procuras por exemplo**

As procuras por exemplo são aquelas em que o utilizador fornece um exemplo do que procura ao mecanismo, pedindo-lhe conteúdo semelhante segundo um ou mais parâmetros a escolher consoante as suas necessidades. Fornecer um exemplo é uma forma eficiente e simples de exprimir aspectos perceptuais associados a parâmetros de baixo-nível e que podem ser extraídos automaticamente sem precisar de qualquer conhecimento da sua representação no ‘mundo real’. Exemplos de parâmetros tipicamente usados são a forma, a cor e o movimento para vídeo, ou o timbre, o tom e o pico de uma peça de áudio. Tal como a procura baseada em estatísticas do sinal, este tipo de procura pode ser classificado como uma procura de nível sintáctico.

Assim, as procuras por exemplo para imagens podem ser definidas essencialmente de três maneiras diferentes:

- **Conteúdo como exemplo** – O exemplo de referência para a procura é conteúdo visual (imagem ou vídeo), proveniente de uma procura prévia ou de um sub-conjunto de imagens ou vídeos disponíveis na base de dados para esse fim específico. Este tipo de especificação da procura é normalmente utilizado para procuras que se baseiam na cor ou textura dos objectos, regiões ou imagens, nomeadamente quando a procura não pode ser facilmente especificada através de um esboço. A Figura 3.3 mostra alguns exemplos de procuras baseadas em exemplos de conteúdo.



a)



b)

Figura 3.3 – Procuras baseadas em exemplos por conteúdo: interfaces para os mecanismos a) MARS [MARS] e b) QBIC [QBIC_Stamps]

- **Esboço como exemplo** – O exemplo que define a procura é um esboço (desenho) mais ou menos completo da imagem que se procura. Este esboço define, de modo aproximado, a forma, cor e posição de uma ou mais regiões da imagem procurada. Este esboço pode ser desenhado manualmente ou ser obtido a partir de uma imagem nomeadamente desenhando o esboço sobre um fundo com uma imagem. Este tipo de definição da procura é normalmente utilizado na procura de imagens que contenham objectos ou regiões com forma, cor e posição semelhantes às do esboço feito. Para proceder à comparação com as descrições disponíveis na base de dados, o mecanismo de procura deve extrair do esboço os parâmetros

relevantes. A Figura 3.4, a Figura 3.5 e a Figura 3.6 mostram alguns exemplos de procuras baseadas num exemplo por esboço.

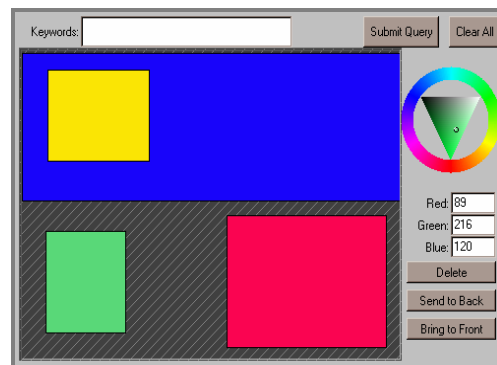


Figura 3.4 – Exemplo de procura baseada num exemplo por esboço usando como parâmetros a relação espacial, a cor e o tamanho dos objectos ou regiões a procurar [QBIC_Stamps].

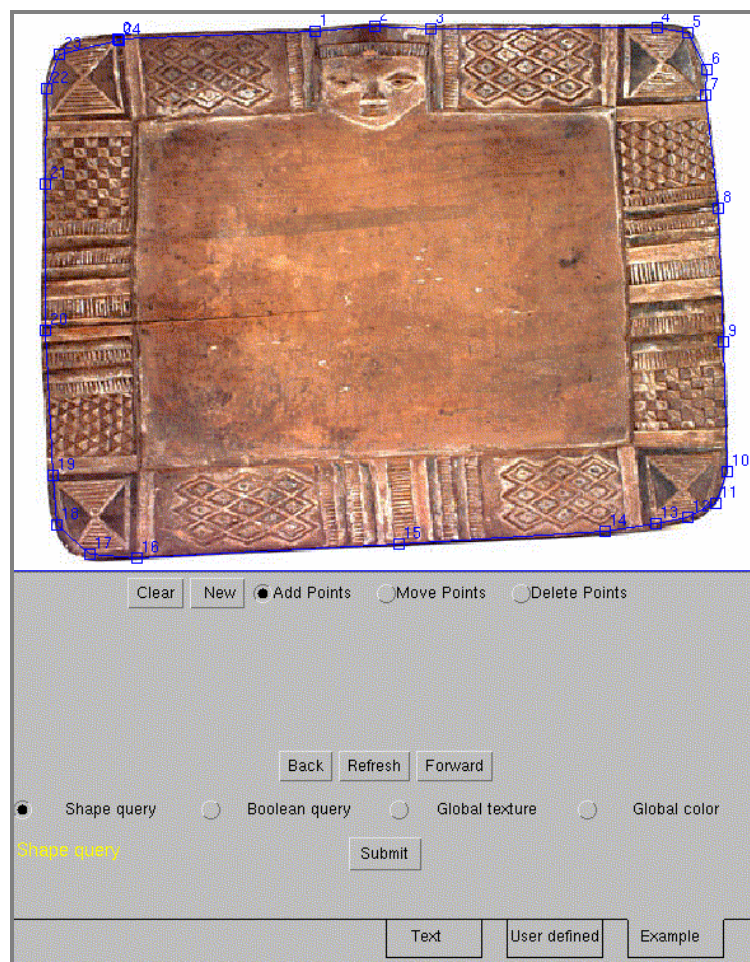


Figura 3.5 – Exemplo de procura baseada num exemplo por esboço delineando o contorno sobre uma imagem de referência [MARS]

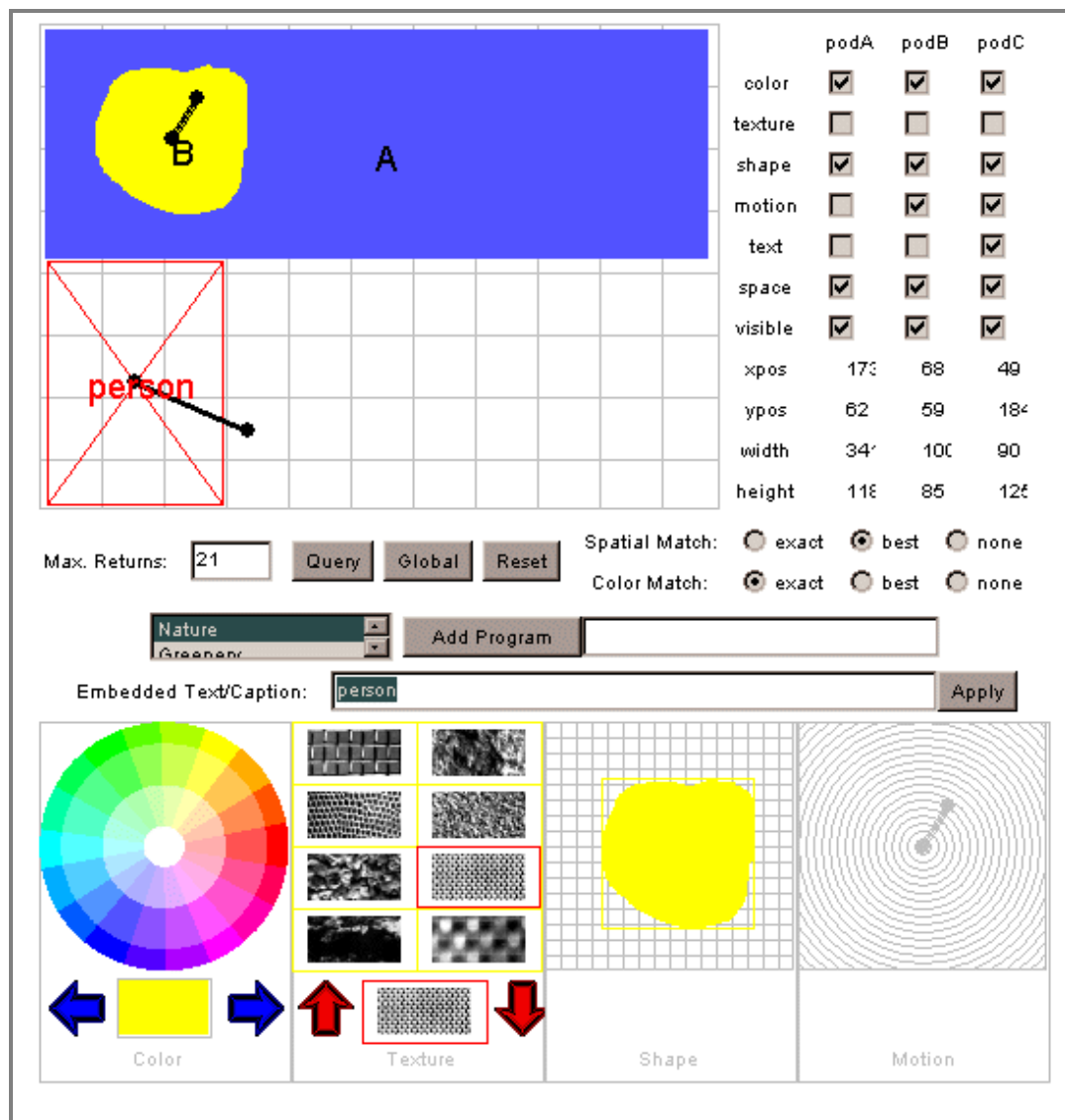


Figura 3.6 – Exemplo de procura baseada num exemplo por esboço: o esboço é desenhado com figuras geométricas, cor, textura e movimento para cada um dos objectos na imagem [VisualSEEK]

- Procuras semânticas** - As procuras semânticas são aquelas que exploram o significado do conteúdo visual através de conhecimentos adquiridos no ‘mundo real’ ou seja usando conceitos de mais alto-nível. Estas procuras baseiam-se normalmente em texto ou listas de palavras chave/conceitos que descrevem o conteúdo visual. Ao contrário dos primeiros mecanismos de descrição e procura em que texto era comparado com texto, aqui as palavras chave são traduzidas/mapeadas pelo módulo de procura para um conjunto de parâmetros de baixo-nível e respectivos valores, associados às características de descrição relevantes para o conceito em questão; nessa altura podem proceder-se às comparações, agora já em termos de parâmetros de baixo-nível; por exemplo um golo num jogo de futebol pode ser traduzido por um conjunto de propriedades em termos de parâmetros de baixo nível, nomeadamente movimento e cor para o vídeo e amplitude para o áudio.

Como nem sempre estes tipos de procura usados isoladamente podem ser capazes de exprimir todas as necessidades dos diferentes utilizadores, amadores ou profissionais, é frequente que se combinem mais do que um tipo de procura, assim como várias características, com vista a obter mecanismos de descrição e procura que possam satisfazer os requisitos do maior número possível de aplicações e utentes.

3.2 Alguns mecanismos de descrição e procura de vídeo

O aparecimento ao longo dos anos, de vários mecanismos de descrição e de procura baseados em características visuais, na *Internet*, como, por exemplo, o *NETRA*, *Photobook*, *QBIC*, *MARS*, *VideoQ*, *VisualSeek*, *Excalibur*, *Candid*, *Chabot* e *Viscomp* [Bimb1999] [Ma1997], causaram um impacto profundo no mundo da descrição e procura de conteúdos visuais.

Deste modo, e tentando seguir a arquitectura genérica proposta anteriormente para estes mecanismos de descrição e procura, analisar-se-ão de seguida alguns dos mecanismos de descrição e procura disponíveis e mais relevantes, nomeadamente o *NETRA*, o *QBIC*, o *MARS* e o *VideoQ*. A escolha dos mecanismos a analisar incidiu sobretudo naqueles que usam a característica de forma ou seja extraem parâmetros de forma, por ser a característica de forma o tema central desta tese. A análise consistirá essencialmente na avaliação da arquitectura escolhida para cada mecanismo, nos algoritmos utilizados para a extracção dos parâmetros de forma no módulo de descrição, nos diversos tipos de interfaces de procura do módulo de procura e, finalmente, nas medidas de semelhança utilizadas para a comparação entre as descrições na base de dados e a formulação da procura feita pelo utilizador.

3.2.1 NETRA

O *NETRA* é um mecanismo de descrição e procura desenvolvido no contexto de um projecto denominado por *Alexandria Digital Library (ADL)* da Universidade da Califórnia em Santa Barbara (UCSB). Este mecanismo tem como domínio de aplicação a descrição e procura de fotografias da natureza, nomeadamente flores, vegetais, plantas tropicais, oceanos, neve, paisagens naturais, desertos, entre muitas outras coisas.

A arquitectura do *NETRA* assemelha-se à arquitectura genérica definida na Figura 3.1, tanto no módulo de descrição como no módulo da base de dados. No entanto, o seu módulo de procura difere do proposto pela arquitectura genérica por não conter o bloco de tradução para parâmetros uma vez que este não é necessário para o tipo de procuras implementadas.

Conhecido o domínio de aplicação deste mecanismo, é importante conhecer as características visuais usadas por este com o intuito de melhor descrever a informação visual armazenada na sua base de dados: a forma, a cor, a textura e a relação espacial. Para representar cada uma destas características foram escolhidos os seguintes parâmetros de descrição [Ma1997]:

- **Forma** – Parâmetros baseados em contornos através da transformada de *Fourier*: função de curvatura, função da distância ao centróide, função de coordenadas complexas;

- **Cor** – Algoritmo generalizado de *Lloyd* baseado na vectorização de um conjunto de cores *RGB* quantificadas;
- **Textura** – Parâmetros baseados na *Gabor wavelet*;
- **Relação Espacial** – Parâmetros baseados em cadeias de símbolos 2D.

Consequentemente, a interface de procura deste mecanismo explora todas estas características e parâmetros, tendo os seus implementadores optado por uma interface que permite procuras por exemplo.

As sub-seções que se seguem descrevem mais pormenorizadamente os detalhes de cada um dos módulos da arquitectura do *NETRA*, com ênfase nos aspectos ligados aos parâmetros de forma.

3.2.1.1 Módulo de descrição

Este mecanismo apresenta um aspecto importante, especialmente no que diz respeito à característica de forma, que é a incorporação de um algoritmo de segmentação automática em regiões homogéneas em termos de cor. A segmentação da imagem permite assim a descrição e a procura baseada em objectos ou regiões. Este algoritmo foi precisamente incorporado no mecanismo para melhorar o seu desempenho em termos da característica de forma, tanto ao nível da descrição como da procura, para imagens com múltiplos e complexos objectos.

Uma vez segmentadas as imagens em objectos ou regiões são então calculados os seus parâmetros de forma. Os parâmetros de forma usados pelo *NETRA* baseiam-se na transformada de *Fourier*, coeficientes F_i , dos contornos e são os apresentados de seguida [Ma1997]:

- **Função de curvatura** – Definida como a taxa de variação da direcção segundo as tangentes ao contorno. Este parâmetro precisa apenas de considerar o eixo real positivo das frequências para representação da função de curvatura do contorno, f_k , definido por:

$$f_k = \left\{ |F_1|, |F_2|, |F_3|, \dots, |F_{M/2}| \right\} \quad (3.1)$$

onde F_i representa o i -ésimo coeficiente da transformada de *Fourier*.

- **Função da distância ao centróide** – Definida como a distância dos *pixels* (x_s, y_s) do contorno ao centróide (x_c, y_c) , segundo a seguinte função, $R(s)$:

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (3.2)$$

Este parâmetro é ainda definido segundo os coeficientes da transformada de *Fourier*, através da função f_R que se segue:

$$f_R = \left\{ \left| \frac{F_1}{F_0} \right|, \left| \frac{F_2}{F_0} \right|, \left| \frac{F_3}{F_0} \right|, \dots, \left| \frac{F_{M/2}}{F_0} \right| \right\} \quad (3.3)$$

- **Função de coordenadas complexas, $Z(s)$** – Obtida através da representação das coordenadas dos *pixels* do contorno como números complexos:

$$Z(s) = (x_s - x_c) - j(y_s - y_c) \quad (3.4)$$

onde $j = \sqrt{-1}$;

Este parâmetro é ainda definido segundo os coeficientes da transformada de *Fourier*, através da função f_Z que se segue:

$$f_Z = \left\{ \left| \frac{F_{-(M/2-1)}}{F_1} \right|, \dots, \left| \frac{F_{-1}}{F_1} \right|, \left| \frac{F_2}{F_1} \right|, \dots, \left| \frac{F_{M/2}}{F_1} \right| \right\} \quad (3.5)$$

3.2.1.2 Módulo da base de dados

O módulo da base de dados deste mecanismo tem funções idênticas ao módulo correspondente da arquitectura genérica proposta anteriormente. As suas funções são o armazenamento da informação visual descrita, o armazenamento das respectivas descrições (geradas pelo módulo de descrição) e a comparação destas descrições com a descrição fornecida pelo módulo de procura.

As imagens são armazenadas na base de dados em formato *JPEG (Joint Photographic Experts Group)*. Quanto às descrições geradas pelo módulo de descrição, encontram-se organizadas segundo estruturas de indexação semelhantes às das *SS-trees* [WhJa1996], existindo uma estrutura para cada uma das características usadas, de modo a facilitar o processo de comparação. Para organizar as descrições segundo a estrutura de indexação adoptada, um algoritmo modificado de agrupamento de *k*-médias é usado, a fim de equilibrar a estrutura e tornar mais eficiente a procura. Este equilíbrio da estrutura é conseguido impondo-se um limiar mínimo ao número de nós em cada agrupamento. Se o agrupamento de uma dada iteração tiver um número de nós inferior ao limiar mínimo imposto, os seus nós serão eliminados desse agrupamento e atribuídos a agrupamentos cujo número de nós é superior ao limiar mínimo imposto mas inferior comparativamente com os outros. Os centros dos agrupamentos são de seguida recalculados e usados como condição inicial para a próxima iteração. O processo repete-se até que nenhum agrupamento tenha um número de elementos inferior ao limiar imposto ou o número de iterações exceda um dado valor [Ma1997]. Como medida de semelhança entre as diversas descrições é usada a distância Euclidiana.

3.2.1.3 Módulo de procura

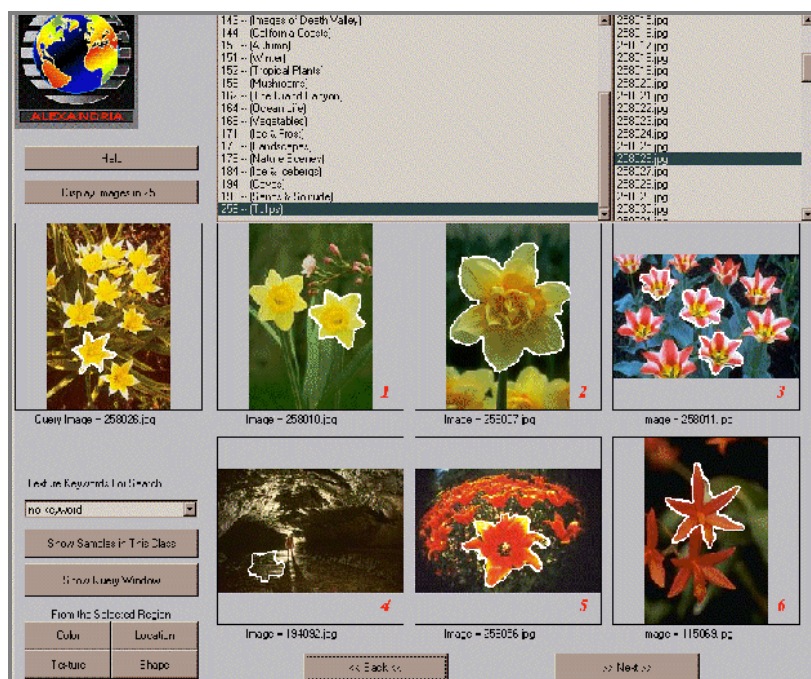
O módulo de procura deste mecanismo tem como principal função a conversão da procura definida pelo utilizador, através da interface disponível, num conjunto de valores dos parâmetros de forma para que sejam comparados no módulo da base de dados. Como já foi

dito, o módulo de procura deste mecanismo difere do sugerido pela arquitectura genérica por não ter o bloco de tradução para parâmetros, ou seja como não se podem fazer procuras semânticas não é necessário fazer a sua tradução em parâmetros de baixo nível. A interface de procura deste mecanismo permite apenas procuras baseadas em exemplos por conteúdo, não permitindo qualquer ajuste em termos de valores estatísticos (ver Figura 3.7a)). A formulação de uma procura no *NETRA*, baseada na característica de forma, segue as seguintes etapas:

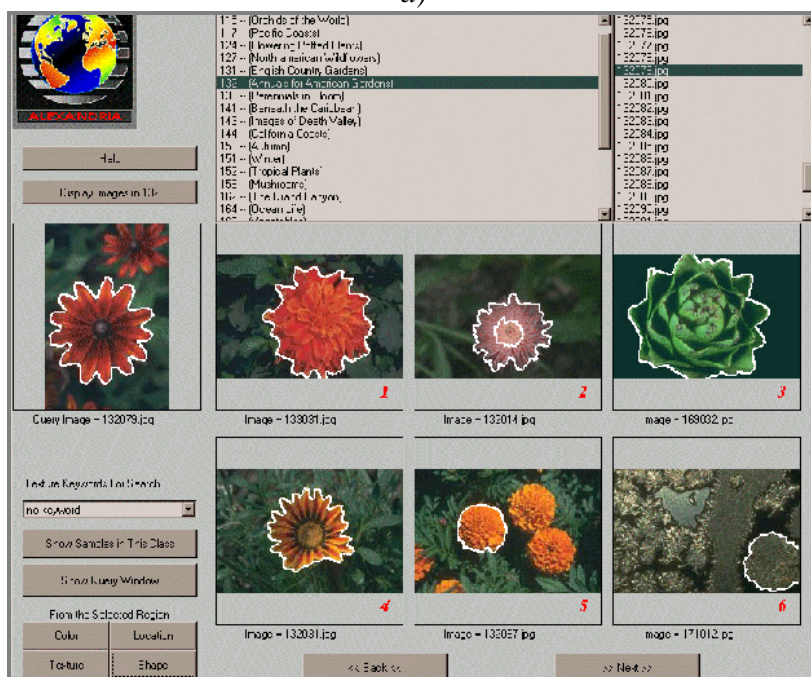
1. Escolha do género de informação visual que se procura como, por exemplo, flores, vegetais, plantas tropicais, oceanos, neve, paisagens naturais, desertos, etc.;
2. Escolha da imagem de referência (exemplo) para definição da procura (ver Figura 3.7a));
3. Escolha de uma das regiões segmentadas na imagem exemplo (ver Figura 3.7b));
4. Escolha dos parâmetros de forma a usar na fase de comparação (ou seja em relação aos quais a semelhança deve ser medida), por exemplo a função de curvatura, a função da distância ao centróide ou a função das coordenadas complexas (ver Figura 3.7b)).

Este mecanismo permite refinar a procura através de *feedback* a partir dos resultados, repetindo-se o processo de procura a partir da etapa 2. Uma vez definida a procura (que não permite o controlo de qualquer tipo de valores estatísticos), é calculada a semelhança entre a descrição de procura e as descrições na base de dados, apresentando-se os resultados por ordem decrescente da sua semelhança com a descrição de procura.

O exemplo da Figura 3.7 pretende demonstrar a definição de uma procura baseada em exemplos por conteúdo usando a característica de forma, na interface do *NETRA*. Nesta procura começou-se por definir o género a procurar (no caso, tulipas), tendo-se escolhido a imagem visualizada no canto à esquerda, da Figura 3.7a) e b), como exemplo de referência para cada uma das procuras efectuadas. De seguida, são escolhidas a região exemplo com a forma que se procura e o parâmetro de forma a usar para a descrição da procura (função de curvatura, função da distância ao centróide (raio) ou função das coordenadas complexas). É de salientar que o mecanismo apenas permite escolher um e só um dos seus três parâmetros de forma para descrever o conteúdo. As Figura 3.7a) e b) mostram os resultados obtidos para cada uma das procuras usando-se as funções das coordenadas complexas e da distância ao centróide, respectivamente.



a)



b)

Figura 3.7 – Interface de procura baseada num exemplo por conteúdo para o *NETRA*: resultados obtidos usando os parâmetros a) função de coordenadas complexas; b) função da distância ao centróide [NETRA]

3.2.2 QBIC

O *QBIC* é um projecto da *IBM* e foi o primeiro mecanismo de descrição e procura a usar características visuais para a descrição da informação visual. Pela sua inovação, este mecanismo teve um impacto profundo nos mecanismos que surgiram posteriormente

[AsBa1995]. Hoje em dia, o *QBIC* é um mecanismo comercial [QBIC], tendo como principais domínios de aplicação o auxílio na venda de têxteis, na procura de selos, imagens com paisagens naturais ou peças de arte em museus, na identificação de logótipos semelhantes ou mesmo iguais. Segundo informações obtidas junto da equipa dentro da IBM que é responsável por este mecanismo, a descrição e a procura de vídeo (e não só imagens estáticas) já foi incluída no QBIC; contudo, as ferramentas de descrição de vídeo, nomeadamente usando o movimento, ainda não estão suficientemente optimizadas, pelo que esta versão do mecanismo ainda não se encontra disponível na *Internet*.

A arquitectura do *QBIC* é semelhante à do *NETRA* e logo semelhante à arquitectura genérica proposta na secção 3.1. No entanto, as características visuais usadas pelo QBIC com o objectivo de melhor descrever e procurar os conteúdos visuais armazenados na sua base de dados são a forma, a cor, a textura e a relação espacial. Os parâmetros escolhidos para representar cada uma destas características são:

- **Forma** – Parâmetros geométricos como a área, a circularidade, a excentricidade e a direcção do eixo maior, momentos invariantes e os ângulos de curvatura do contorno;
- **Cor** – Parâmetros de *Munsell* e histogramas de cor [AsBa1995][QBIC_Stamps];
- **Textura** – Granularidade, contraste e direccionalidade [QBIC_Stamps][TaMo1978];
- **Relação espacial** – Coordenadas (x,y) normalizadas do centro de massa do objecto ou região [QBIC_Stamps].

Tendo em conta os domínios de aplicação deste projecto e as características visuais escolhidas, os tipos de procura possíveis neste mecanismo são: i) procura por exemplo usando esboço, ii) procura por exemplo usando conteúdo e iii) procura baseada em estatísticas do sinal, nomeadamente percentagens e valores numéricos representativos das componentes *RGB* da cor.

As sub-secções que se seguem descrevem mais pormenorizadamente cada um dos módulos da arquitectura deste mecanismo, com maior ênfase na sua aplicação à identificação de logótipos por ser esta que usa a característica de forma para descrever a informação visual.

3.2.2.1 Módulo de descrição

O módulo de descrição do *QBIC* é o módulo onde são calculados os parâmetros visuais associados aos conteúdos armazenados na sua base de dados. O *QBIC* descreve o seu conteúdo extraindo os parâmetros de descrição de duas maneiras distintas:

1. Globalmente onde os parâmetros são calculados ao nível da trama (no caso do vídeo usa-se a trama representativa de uma cena) e
2. Localmente onde os parâmetros são calculados para objectos ou regiões segmentadas.

Na aplicação dos logótipos, a descrição das formas é feita através de um conjunto de parâmetros geométricos como a área, circularidade, excentricidade e direcção do eixo maior, de um conjunto de parâmetros de momentos invariantes e da representação do contorno de cada objecto ou região em cada imagem através dos ângulos de curvatura do contorno.

3.2.2.2 Módulo da base de dados

O módulo da base de dados deste mecanismo tem funções idênticas ao mesmo módulo da arquitectura genérica sugerida anteriormente. A informação visual armazenada na base de dados são fotografias ou no caso específico da aplicação dos logótipos imagens estáticas binárias (preto e branco), assim como as suas diferentes regiões ou objectos. As descrições geradas pelo módulo de descrição encontram-se organizadas segundo uma estrutura de indexação multi-dimensional *R-tree* [RuHu1997], na base de dados das descrições. Ainda nesta base de dados, encontram-se armazenados todos os contornos de cada objecto ou região de um logótipo para que futuramente possam ser usados durante uma procura.

A medida de semelhança usada por este mecanismo é a distância Euclidiana, definida como a raiz quadrada da soma das diferenças quadráticas entre as descrições armazenadas na base de dados e a proveniente do módulo de procura. Os resultados obtidos pelo cálculo desta medida são ordenados por ordem crescente, a fim de enviar para a interface de procura do mecanismo as *N* imagens mais semelhantes.

3.2.2.3 Módulo de procura










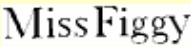





O módulo de procura deste mecanismo tem como principal função a conversão da formulação da procura definida pelo utilizador em parâmetros de forma para que sejam comparados com as descrições disponíveis na base de dados. Tal como o *NETRA*, também este mecanismo não inclui o bloco de tradução para parâmetros por não ser possível fazer procuras de nível semântico. A Figura 3.8 mostra as diferentes interfaces de procura para a aplicação de identificação de logótipos. A Figura 3.8a) apresenta a interface inicial onde é escolhido o tipo de logótipo que se procura, enquanto que a Figura 3.8b) mostra a possibilidade de fazer a procura usando-se apenas um dos elementos da forma inicialmente escolhida, obtidos por segmentação. Uma vez escolhido um dos elementos de forma segmentados, passa-se à fase de comparação, obtendo-se os resultados numa interface final como a apresentada na Figura 3.8c). Esta última interface permite ao utilizador do mecanismo rejeitar falsos resultados e refinar os resultados obtidos.

To query for parameters such as Owner or International Class codes, click [here](#) To get random trademarks, click [here](#)

For each trademark below, you can click on:

- I** to display detailed information about the trademark.
- F** to display the full page scanned trademark form.
- D** to find trademarks with the same Design Code(s).
- S** to show shape components for this trademark. Not all patterns have shape components. If they do, you will be able to ask for similar shapes.

Randomly selected patterns

				
I F D S	I F D S	I F D S	I F D S	I F D
				
I F D	I F D	I F D	I F D	I F D
				
I F D	I F D	I F D	I F D	I F D S

a)

Trademark pattern 1392952



- One or more shapes identified in the above pattern are shown below. Click on [Q](#) below a shape to find similar shapes.
- Use your browser's "Back" button (usually in the upper left) to return to the previous page.

Shape components in this pattern:



[Q](#) mask.0



[Q](#) mask.1

b)



c)

Figura 3.8 – *QBIC* – Interfaces de procura da aplicação dos logótipos: a) interface por exemplo inicial; b) interface por exemplo intermédia (escolheu-se como exemplo a *mask 0*); e c) interface que apresenta os resultados da procura [*QBIC_Trade*]

Apesar das procuras permitidas por esta aplicação do *QBIC* gerarem bons resultados e de se pretender que estes mecanismos sejam completamente automáticos (ou seja, não necessitando de qualquer interacção humana), poderia ainda ser usado outro tipo de procura ou seja por exemplo com esboço. Em aplicações mais antigas do *QBIC*, este tipo de procura foi também usado afim de se encontrar conteúdo visual semelhante em termos da característica de forma. Exemplos deste tipo de procura no *QBIC* são apresentados na Figura 3.9a), onde o utilizador esboça (desenha) a forma que procura e o mecanismo devolve os conteúdos mais parecidos (ver Figura 3.9b)).

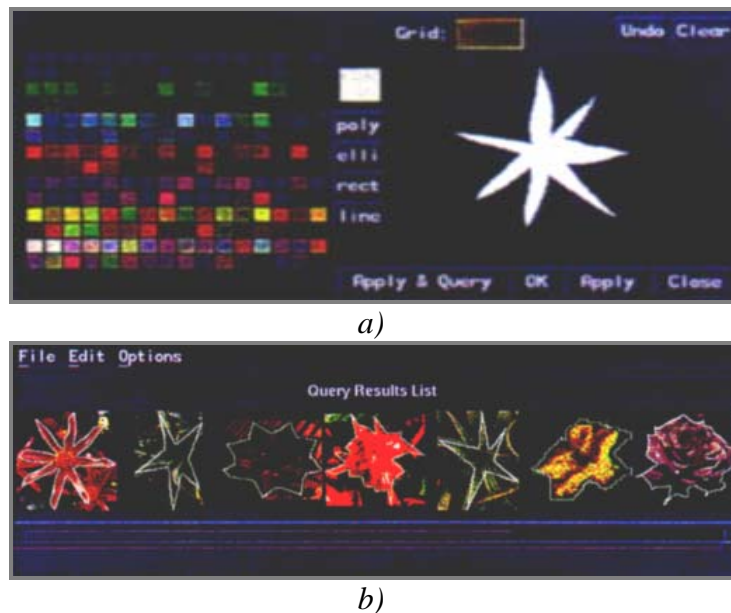


Figura 3.9 – *QBIC*: a) interface com procura do tipo esboço como exemplo; b) interface com o resultado da procura [Finn1996]

3.2.3 *MARS*

O *MARS* é um projecto desenvolvido pela Universidade de *Illinois* (EUA) que visa implementar um mecanismo de descrição e procura com uma arquitectura semelhante à proposta na secção 3.1. Este mecanismo não só procura a melhor característica, como a forma, cor, textura e relação espacial, para descrever o conteúdo da imagem mas organiza as várias características visuais de modo a obter-se um método de procura da informação dinamicamente adaptável às diferentes aplicações e utilizadores, pois cada um destes exige graus de importância diferentes relativamente às características usadas. Neste contexto, o *MARS* propõe um método, para o módulo de procura, de *feedback* relevante [OrRu1997] [RuCh1997][RuHu1997b]. Este método, integrado em vários níveis da procura de informação, permite a interacção entre o utilizador e o mecanismo de modo a refinar as procuras através das descrições geradas. Esta interacção entre o utilizador e o mecanismo é feita em duas fases distintas, que se repetem até que o utilizador esteja satisfeito com o conteúdo devolvido pelo mecanismo:

1. Utilizador informa o mecanismo do que é relevante ou não para a sua procura, ou seja assinala o que mais se assemelha com o que procura; e
2. Mecanismo aprende, interpretando a informação e atribuindo pesos às várias características.

Neste tipo de mecanismo, as preocupações normalmente passadas ao utilizador de expressar conceptualmente o que procura e da atribuição de pesos às características são aqui eliminadas ao nível do utilizador e passadas para o mecanismo. Os pesos normalmente envolvidos na procura efectuada pelo utilizador são actualizados dinamicamente de modo a moldarem-se em conceitos de alto nível [RuHu1997a].

Os tipos de procura usando a característica de forma que este mecanismo permite baseiam-se apenas em exemplos por conteúdo e por esboço. A semelhança é medida através da distância Euclidiana, entre os valores dos parâmetros da procura e os das descrições existentes em base de dados.

3.2.3.1 Módulo de descrição

Os parâmetros escolhidos para a descrição da informação visual neste mecanismo foram os coeficientes de *Fourier*, estudados e descritos na secção 2.2.2.1, caracterizados pelos implementadores do *MARS* como rápidos e exactos [RuCh1997].

3.2.3.2 Módulo da base de dados

O módulo da base de dados do *MARS* armazena as imagens originais e as suas descrições que em termos de forma se resumem aos valores dos coeficientes da transformada de *Fourier*. Como medida de semelhança entre os valores dos parâmetros da descrição de procura e os residentes na base de dados, o *MARS* propõe a distância Euclidiana.

3.2.3.3 Módulo de procura

O módulo de procura do *MARS* oferece ao utilizador o método de *feedback* relevante em relação aos resultados da procura, permitindo assim ao utilizador do mecanismo efectuar procuras baseadas em exemplos de conteúdo visual e de esboço, pois estes dois tipos de procura pouco ou nada necessitam de interactividade por parte do utilizador. O exemplo da Figura 3.10 demonstra o funcionamento dos dois tipos de procura adoptados neste mecanismo: procura por exemplo (Figura 3.10a)) e procura por esboço (Figura 3.10b)). Neste exemplo, começou-se por seleccionar uma das imagens da base de dados de um museu, procura por exemplo (Figura 3.10a)), podendo de seguida desenhar-se a forma que se pretende procurar sobre esta imagem, procura por esboço (Figura 3.10b)), salientando-se o facto de se poderem refinar os resultados obtidos.

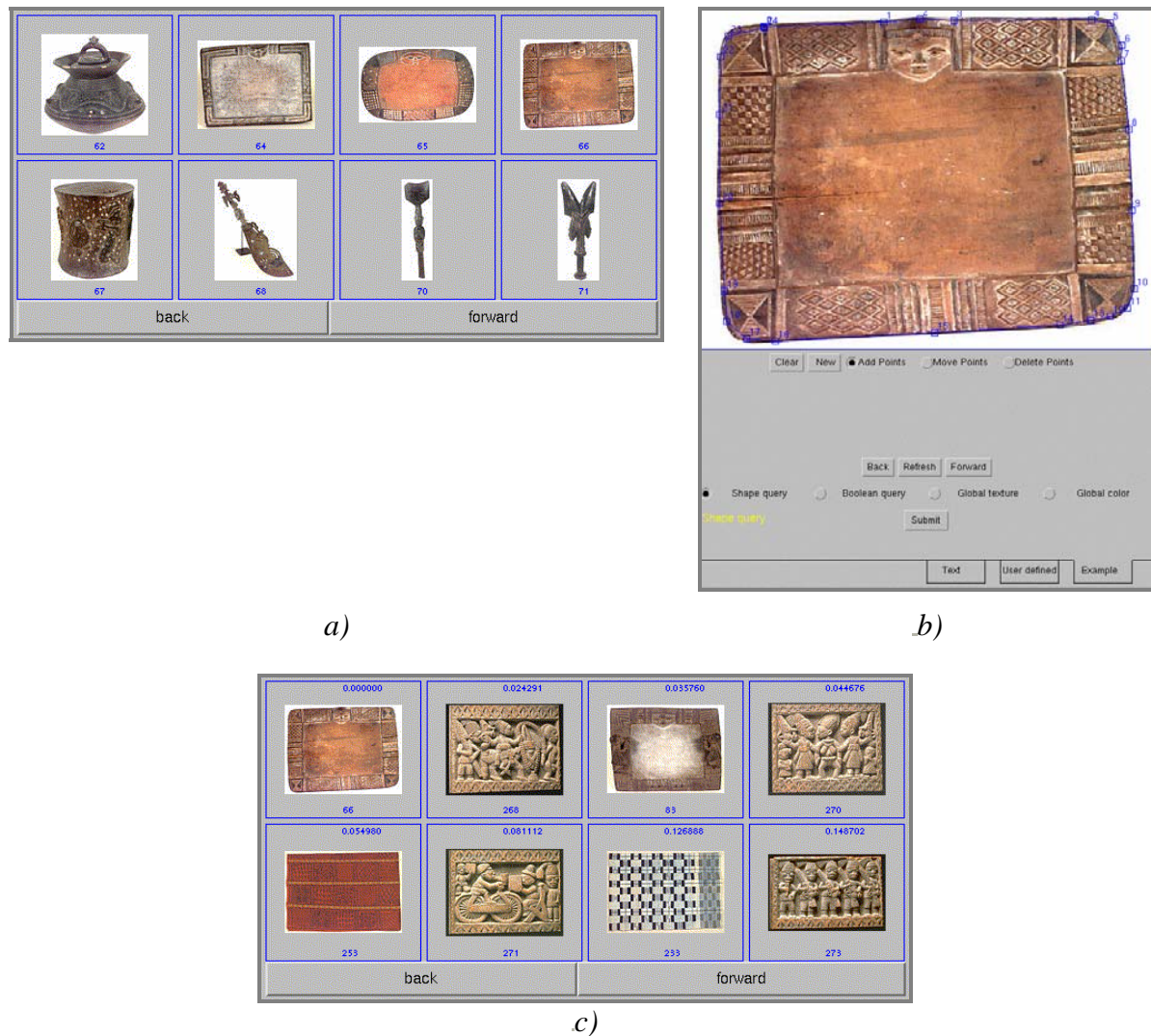


Figura 3.10 – MARS – Interface de procura: a) procura baseada em exemplo por conteúdo; b) procura baseada em exemplo por esboço; c) resultados da procura efectuada usando a característica de forma [MARS]

3.2.4 VideoQ

O *VideoQ* é um mecanismo de descrição e procura que pretende inovar sobre outros mecanismos mais convencionais através da utilização de um conjunto de características visuais e de relações espaço-temporais para vídeo. A inovação deste mecanismo face aos anteriores tem essencialmente a ver com o facto de a sua interface de procura permitir procuras baseadas em exemplos por esboço animadas ou seja permitir a formulação de procuras usando a característica de movimento [ChCh1997].

À semelhança dos mecanismos anteriormente apresentados, o *VideoQ* descreve a informação visual através das usuais características de forma, cor e textura, acrescentado ainda o movimento. Os parâmetros escolhidos para representar cada uma destas características foram os seguintes [ChCh1997]:

- **Forma** – Componentes (eixos) principais, momentos de primeira e segunda ordem, área normalizada em relação à imagem e percentagem de área da imagem ocupada pelo objecto ou região;
- **Cor** – Cor dominante ou seja cor mais representativa/importante na trama ou objecto em análise;
- **Textura** – Medidas de *Tamura* de granularidade, contraste e direcção usadas como parâmetro para a representação do conteúdo textural do objecto ou região;
- **Movimento** – Vectores de movimento calculados a partir do fluxo óptico; o parâmetro armazena um vector de movimento por trama, calculado como a translação média do centróide entre duas tramas sucessivas.

Consequentemente, a interface de procura deste mecanismo explora todas estas características e parâmetros, tendo os seus implementadores optado por uma interface que permite procuras por esboço como exemplo e por estatísticas de sinal usando pesos.

As sub-seções que se seguem descrevem mais pormenorizadamente os detalhes de cada um dos módulos da arquitectura do *VIDEOQ*, com ênfase nos aspectos ligados aos parâmetros de forma.

3.2.4.1 Módulo de descrição

No *VideoQ*, o processo de descrição da informação visual inicia-se com a segmentação temporal – detecção de mudanças de cena - e espacial – identificação de objectos dentro da cena - da informação de vídeo. A segmentação temporal do vídeo é feita através da detecção das mudanças de cena, armazenando-se o número da primeira trama de cada cena. A segmentação espacial dos objectos ou regiões usa várias características da informação de vídeo, como a cor e as transições (*edges*), assim como o seguimento das trajetórias de regiões com cores dominantes para obter coerência temporal das regiões segmentadas. Tendo-se os objectos ou regiões segmentados, o mecanismo está em condições de fazer a descrição dos objectos ou regiões através da característica de forma.

Assim sendo, os diferentes parâmetros de forma a serem calculados para cada objecto ou região serão [ChCh1997]:

- Eixos principais do objecto ou região tal como definido na sub-seção 2.2.5.1;
- Momentos de primeira e segunda ordem da forma do objecto ou região (não se sabendo precisamente qual o tipo de momentos usados, por esta informação não se encontrar disponível);
- Área normalizada (a razão entre a área do objecto e a área da circunferência envolvente) e percentagem da área (a percentagem ocupada pelo objecto ou região na trama); e
- Proximidade da forma do objecto ou região a uma elipse.

Após o cálculo dos parâmetros de forma, as descrições geradas são armazenadas na base de dados de descrições no módulo da base de dados.

3.2.4.2 Módulo da base de dados

No módulo da base de dados do *VideoQ* são armazenadas as descrições geradas pelo módulo de descrição, com as suas respectivas segmentações e são calculadas as medidas de semelhança entre as descrições na base de dados e a fornecida pelo mecanismo de procura.

O mecanismo permite a procura de informação visual na sua base de dados calculando a semelhança existente entre as descrições na base de dados e a descrição gerada pelo módulo de procura usando como medidas de semelhança para a forma uma medida baseada nas componentes principais (3.6) e outra baseada na percentagem da área ocupada pelo objecto (3.7), definidas como:

$$D(p, q) = \left| \frac{\lambda_{2_p}}{\lambda_{1_q}} - \frac{\lambda_{2_q}}{\lambda_{1_p}} \right| \quad (3.6)$$

$$D_A(p, q) = 1 - \frac{\min(A_p, A_q)}{\max(A_p, A_q)} \quad (3.7)$$

onde λ_1 e λ_2 são os valores próprios dos eixos principais do objecto ou região e A_p e A_q são as percentagens das áreas ocupadas pelos objectos ou regiões na base de dados e os fornecidos pelo utilizador do mecanismo, respectivamente.

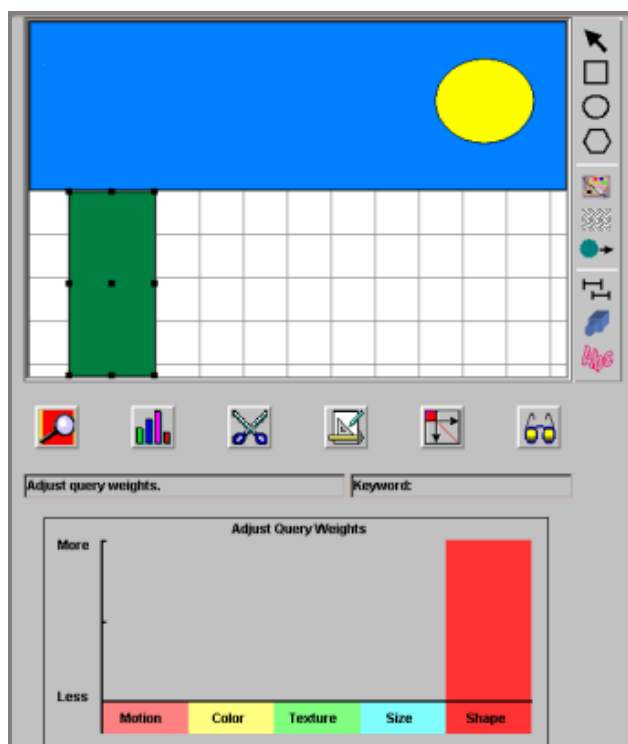
3.2.4.3 Módulo de procura

O módulo de procura do *VideoQ* é aquele que apresenta maior inovação já que a sua interface de procura permite ao utilizador formular procuras usando parâmetros que descrevem variações ao longo do tempo e logo mais adequados a procuras de vídeo. Embora esta inovação não se encontre directamente relacionada com a característica que esta tese pretende estudar, não deixa no entanto de ser um facto importante porque permite a descrição de vídeo através de características visuais mais adequadas e não só através de texto ou interpretando o vídeo como uma sucessão de imagens-chave (*key-frames*).

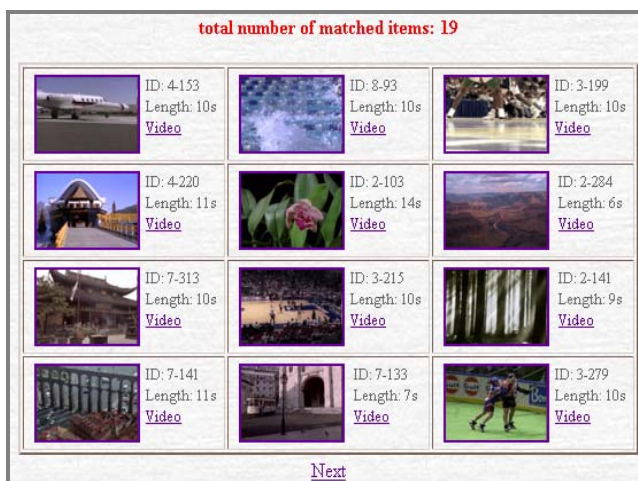
As procuras permitidas por este mecanismo baseiam-se essencialmente em exemplos por esboço e baseados em estatísticas do sinal usando pesos. Uma procura baseada na característica de forma do *VideoQ* poderá ser definida através de formas geométricas (circunferências, elipses ou rectângulos) de tamanho arbitrário; a influência das várias características na procura pode ser determinada através de pesos associados a cada uma delas [VideoQ].

A Figura 3.11a) mostra um exemplo de uma procura no *VideoQ* definida usando várias características, entre as quais a característica de forma. Embora as formas se encontrem coloridas é possível verificar que não foi atribuído qualquer peso à característica de cor (zero) nem a qualquer outra característica para além da forma, o que significa que esta terá o peso

total e logo determinará sozinha o resultado da procura. Na Figura 3.11b) é apresentado o resultado da procura formulada através da interface da Figura 3.11a). Para cada resultado apresentado, é possível visualizar o vídeo correspondente seleccionando-se o botão 'Video'.



a)



b)

Figura 3.11 – VideoQ: a) Interface de entrada usando esboço como exemplo; b) resultados da pesquisa definida em a) [VideoQ]

3.3 Conclusões

A crescente invasão de informação audiovisual nos mais variados domínios de aplicação, aumentou enormemente as necessidades em termos de descrição e procura levando a que estes temas se tenham tornado em importantes áreas de pesquisa e estudo, sobretudo nos anos mais recentes. Assim, muitos parâmetros de descrição associados a características audiovisuais bem como as correspondentes ferramentas de análise foram desenvolvidos e aperfeiçoados. Como consequência, surgiram mecanismos de descrição e procura cada vez mais sofisticados, usando parâmetros de baixo-nível para além de descritores textuais, o que abre novas possibilidades em termos de descrição e procura. Lembre-se que a adopção de parâmetros de descrição de baixo-nível tem sobretudo as grandes vantagens de poderem ser calculados automaticamente (o que é essencial face à crescente quantidade de informação) e de evitarem o mapeamento da informação em termos subjectivos ou específicos de um domínio.

De entre os vários mecanismos disponíveis na Internet foram analisados o NETRA, o QBIC, o MARS e o VideoQ por usarem a característica de forma que esta tese trabalho tem como objectivo estudar e desenvolver.

Ao longo deste capítulo sobre mecanismos de descrição e procura, pretenderam-se analisar e avaliar dois aspectos importantes:

- Propriedades mais relevantes de cada um dos módulos da arquitectura genérica proposta na Figura 3.1; e
- Soluções adoptados pelos mecanismos analisados para os vários módulos desta arquitectura, com especial incidência em todas as questões associadas à característica de forma, nomeadamente parâmetros adoptados, medidas de semelhança, tipos de interface, etc.

Tendo-se analisado detalhadamente quatro mecanismos de descrição e procura entre os mais relevantes actualmente existentes nos meios comercial e académico, é agora possível fazer a sua análise comparativa. Os mecanismos serão analisados e avaliados qualitativamente de acordo com os requisitos descritos anteriormente no capítulo 1 e na secção 3.1.3.1. Esta avaliação dos vários mecanismos pretende contribuir para a especificação do mecanismo de descrição e procura de forma a implementar no contexto desta tese.

Relativamente ao módulo de descrição, todos os mecanismos têm em comum a descrição da informação de forma, pois foi a condição básica da escolha de cada um dos mecanismos aqui analisados, distinguindo-se os mecanismos apenas no facto de permitirem descrever imagens estáticas ou também vídeo (*VideoQ*). Pode também concluir-se que a maioria dos parâmetros de forma usados não variam significativamente dos apresentados no capítulo anterior. A Tabela 3.1 resume os parâmetros de forma escolhidos por cada um dos mecanismos.

Mecanismo de descrição e procura	Parâmetros de forma
<i>NETRA</i>	<ul style="list-style-type: none"> • Função de curvatura; • Função da distância ao centróide; e • Função complexa, • baseadas nos coeficientes de <i>Fourier</i>
<i>QBIC</i>	<ul style="list-style-type: none"> • Área • Circularidade • Excentricidade • Direcção do eixo maior • Momentos invariantes • Ângulos de curvatura do contorno

MARS	<ul style="list-style-type: none"> • Coeficientes de <i>Fourier</i>
VideoQ	<ul style="list-style-type: none"> • Componentes principais • Momentos • Área

Tabela 3.1 – Parâmetros de forma escolhidos para cada um dos mecanismos analisados

O módulo da base de dados de cada um dos mecanismos, à semelhança do da arquitectura genérica, para além de armazenar a informação a descrever e as respectivas descrições, é ainda responsável pelas comparações entre descrições e logo pelas medidas de semelhança. Dos mecanismos estudados apenas o *QBIC* e o *MARS* usam a distância Euclidiana como medida de semelhança, usando os restantes mecanismos medidas mais específicas aos parâmetros escolhidos.

Finalmente, os módulos de procura destes mecanismos têm propriedades e funções semelhantes distinguindo-se da arquitectura genérica por nenhum incluir o bloco de tradução de texto ou palavras-chave para parâmetros. No entanto, os parâmetros usados por este módulo são os mesmos que os usados pelo módulo de descrição de modo a garantir a interoperabilidade entre eles. Quanto às interfaces de procura de cada mecanismo, todas são maioritariamente gráficas e visuais (incluindo métodos de procura que usufruem da capacidade visual do ser humano para definir o que se pretende encontrar), facilitando e possibilitando a compreensão do conteúdo e descrições armazenadas na base de dados assim como a sua navegação e ainda a definição de procuras de nível sintáctico. Dos mecanismos estudados apenas o *VideoQ* permite o controlo da influência de vários parâmetros através da atribuição de pesos, o que já não acontece com os restantes mecanismos onde o processo de procura, após a selecção da imagem exemplo, é completamente automático. Do que se observou em relação à procura baseada na característica de forma, os tipos de formulação da procura mais usados são as procuras por conteúdo como exemplo e por esboço como exemplo. A Tabela 3.2 apresenta as propriedades principais dos módulos da base de dados e de procura para os mecanismos analisados. O símbolo ✓ representa a existência do tipo de informação visual que cada mecanismo descreve, o uso de uma das medidas de semelhança e o tipo de interface de procura usada por cada mecanismo, caso contrário, na ausência de uma destas propriedades usar-se-á o símbolo ✕.

Mecanismo de descrição e procura	Informação visual		Medidas de semelhança		Tipos de procura		
					Por exemplo		Por estatísticas
	Imagens estáticas	Vídeo	Dist. Euclidiana	Outras	Conteúdo	Esboço	Pesos
NETRA	✓	✕		✓	✓	✕	✕
QBIC	✓	✕	✓		✓	✓	✕
MARS	✓	✕	✓		✓	✓	✕
VideoQ	✕	✓		✓	✕	✓	✓

Tabela 3.2 – Propriedades dos módulos de base de dados e de procura dos mecanismos analisados

Uma vez analisados estes mecanismos e tendo-se em vista as propriedades dos módulos de descrição, da base de dados e de procura, pode-se concluir que um mecanismo de descrição e procura que pretende descrever e procurar conteúdos através da forma deverá obedecer aos seguintes requisitos:

- Incluir ferramentas automáticas para a extracção de parâmetros de forma para que seja possível realizar procuras baseadas na forma através de interfaces simples mas eficazes na aproximação ao conceito humano de forma;
- Adoptar medidas de semelhança que traduzam de maneira adequada a semelhança entre a informação visual descrita e a informação procurada pelo utilizador do mecanismo usando os parâmetros escolhidos; e
- Possuir interfaces de procura que exijam o mínimo de esforço por parte do utilizador, passando para o módulo de procura o processo de traduzir em parâmetros de baixo-nível as necessidades do utente expressas o mais possível de forma ‘intuitiva’.

Existe no entanto um problema fundamental relativamente às descrições geradas por cada mecanismo de descrição e procura: as descrições são em princípio específicas de cada mecanismo e logo não pode existir interoperabilidade entre os vários mecanismos. Apesar das ferramentas de cálculo dos parâmetros de forma nos módulos de descrição e de procura em cada um destes mecanismos poderem até ser idênticas, a interoperabilidade exige mais do que isso ou seja a adopção das mesmas sintaxe e semântica para as descrições para que estas possam ser compreensíveis para todos os mecanismos em questão.

É neste contexto que tem crescido ultimamente a necessidade de especificar normas de descrição de informação audiovisual, permitindo o acesso o mais generalizado possível à informação audiovisual disponível. A norma MPEG-7 é o exemplo maior da resposta a esta necessidade, sendo também a única que até hoje decidiu adoptar parâmetros de baixo-nível.

Capítulo 4

Mecanismo de descrição e procura de formas

Este capítulo tem como objectivo apresentar e descrever a aplicação desenvolvida no âmbito desta tese, ou seja, um mecanismo de descrição e procura baseado principalmente na característica de forma dos objectos de vídeo. Esta aplicação permite a avaliação e comparação do desempenho dos descritores de forma do MPEG-7 assim como de outros descritores alternativos de descrição de forma como os ângulos de curvatura do contorno [IBMRe1999], os momentos de *Zernike* [KiKi1999a] [KiKi1999b] e os vectores próprios multi-nível (*MLEV*) [KiKi1999c] [KiKi1999d]. A aplicação desenvolvida funciona essencialmente como uma interface gráfica entre o utente do mecanismo de descrição e procura e o *software Experimentation Model (XM)* do MPEG-7, onde foram integrados os descritores de forma alternativos (os descritores de forma do MPEG-7 já estavam integrados). A novidade desta aplicação relativamente a outros mecanismos existentes na *Internet*, e os descritos no capítulo 3, recai precisamente sobre o facto de se terem usado os descritores normalizados pelo MPEG-7, garantindo-se deste modo a interoperabilidade com outros mecanismos em termos das descrições de forma se apenas descritores MPEG-7 forem usados (e nenhuns dos alternativos).

O mecanismo de descrição e procura, denominado por *DREVIST - Description and Retrieval of Visual content by IST*, permite descrever e procurar imagens estáticas, sobretudo usando a característica da forma, estando a aplicação no entanto preparada para futuramente descrever e procurar vídeo, nomeadamente usando as características de forma e movimento. Apesar de a característica relevante no âmbito desta tese ser a forma, a aplicação permite que o utente use os descritores de cor disponíveis no MPEG-7 para a realização de procuras mais

elaboradas, p.e. usando a cor e a forma para formular uma procura. Para além disso, a aplicação tem já também integrados alguns dos descritores de movimento do MPEG-7 (ainda que o seu uso não seja ainda possível por a aplicação só lidar com imagens estáticas e não com vídeo). Todas as restantes ferramentas normativas do MPEG-7, como os esquemas de descrição, poderão também, no futuro, ser facilmente integradas nesta aplicação devido à sua estrutura modular.

Uma vez que a maioria dos utentes tem conhecimentos e facilidade em utilizar o sistema operativo *Windows*, a aplicação foi desenvolvida para funcionar em ambientes *Windows95/98/NT*, tendo-se por isso optado pelo uso da linguagem C++ e do compilador *Microsoft Visual C++v6.00*. A interface gráfica desenvolvida para esta aplicação usa a Língua Inglesa por esta ser a língua utilizada pela maior parte dos utentes do mundo informático e tecnológico, facilitando assim a sua utilização por um maior número de pessoas, não só em Portugal mas também no estrangeiro. Esta opção permitirá a sua apresentação ao grupo MPEG tal como já foi apresentada na *Electronic Conference on Digital Libraries (ECDL)* [ZiPe2000].

As secções que se seguem descrevem e analisam pormenorizadamente a aplicação desenvolvida, de acordo com a seguinte ordem:

- **Arquitectura geral da aplicação** – Descreve-se a estrutura da aplicação com base no modelo de arquitectura genérico para aplicações do tipo *Pull* apresentado e definido no capítulo 3; nesta secção serão também descritas, de modo genérico, as interfaces gráficas correspondentes aos dois módulos principais da aplicação;
- **Módulo de descrição** – Descreve-se pormenorizadamente cada opção dos menus, barra de ferramentas e caixas de diálogo disponíveis na interface relativa ao módulo de descrição;
- **Módulo de procura** – Descreve-se pormenorizadamente cada opção dos menus, barra de ferramentas e caixas de diálogo disponíveis na interface relativa ao módulo de procura.

4.1 Arquitectura geral da aplicação

A arquitectura usada para o mecanismo de descrição e procura desenvolvido baseia-se na estrutura sugerida pela arquitectura genérica para uma aplicação do tipo *Pull* apresentada no capítulo 3. Esta arquitectura tem uma estrutura modular e permite o armazenamento, descrição e procura de conteúdo visual. Deste modo, e à semelhança da arquitectura genérica previamente apresentada, a aplicação é composta por três módulos ou seja: o módulo da base de dados, o módulo de descrição e o módulo de procura. De acordo com a arquitectura modular sugerida, esta secção irá descrever e definir a sua estrutura de dados assim como a estrutura das respectivas interfaces gráficas.

4.1.1 Estrutura de dados

Para que os objectivos da aplicação levados a cabo através das suas interfaces sejam alcançados, foi definida uma estrutura de dados que inclui alguns ficheiros de informação comuns às duas interfaces, para além do conteúdo de cada base de dados de conteúdo visual e de descrições.

A estrutura de dados sugerida a partir da directoria em que é instalada a aplicação é a que se mostra na Figura 4.1:

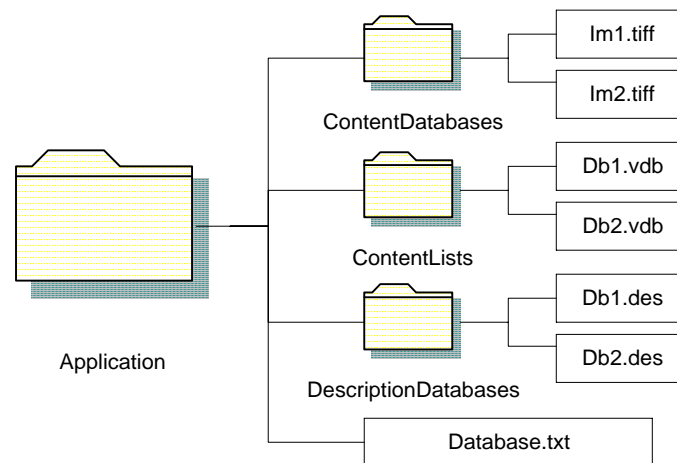


Figura 4.1 – Estrutura de directórios para a aplicação desenvolvida

Na directoria onde é instalada a aplicação *Application* são criadas três sub-directorias, nomeadamente:

- ***ContentDatabases*** – Armazena as bases de dados de conteúdo visual a ser descrito; o conteúdo usado por esta aplicação está todo em formato *TIFF* (*Tag Image File Format*) [MuRy1996]) e cada imagem corresponde a um ficheiro;
- ***ContentLists*** - Contém ficheiros com a extensão “vdb” (*visual databases*) , cada um deles com a lista de imagens pertencentes a cada base de dados; são estas listas que definem as bases de dados de conteúdo a partir do conjunto de imagens disponíveis em *ContentDatabases*;
- ***DescriptionDatabases*** – Armazena as descrições geradas para o conjunto de imagens das bases de dados definidas através dos ficheiros com extensão “vdb”; a cada base de dados estará associado um ficheiro com o nome identificativo da base de dados e com extensão “des”.

Ainda na directoria da aplicação existe um ficheiro de informação para ambas as interfaces de descrição e procura: *Database.txt*. Este ficheiro fornece a estas interfaces a informação necessária relativamente ao módulo da base de dados ou seja informa sobre as bases de dados de conteúdo existentes e a existência ou não de descrições para cada uma dessas bases de dados.

O ficheiro “*Database.txt*” criado para auxiliar o funcionamento das interfaces de descrição e de procura tem as seguintes funções relativamente a cada uma das interfaces, respectivamente:

- Informar a interface de descrição sobre as bases de dados de conteúdos visuais existentes para descrição e as descrições que terão já sido geradas para cada uma das bases de dados de conteúdo visual. Este ficheiro é inicializado pela aplicação, durante a sua primeira execução, com a lista das bases de dados de conteúdo existentes para descrição e procura. Daí por diante, cada vez que surge uma nova base de dados de conteúdo esta é acrescentada ao ficheiro assim como o nome dos descritores usados para descrever as várias bases de dados (p.e. MPEG-7 CSS, MPEG-7 ART, MLEV, Zernike Moments, Turning Angles). Com base nesta informação, a aplicação consegue sempre saber as bases de dados de conteúdo existentes e as suas respectivas descrições mantendo o utente da aplicação informado; e
- Informar a interface de procura sobre as bases de dados de descrições existentes para procura e os descritores usados para as descrever. Sempre que uma base de dados for descrita, esta será incluída neste ficheiro com o(s) respectivo(s) descritor(es) usados; este ficheiro será também actualizado sempre que novas descrições sejam geradas para bases de dados já com descrições. Com base nesta informação, a aplicação consegue sempre saber as bases de dados de descrições existentes e os descritores usados para as descrever, mantendo o utente da aplicação informado.

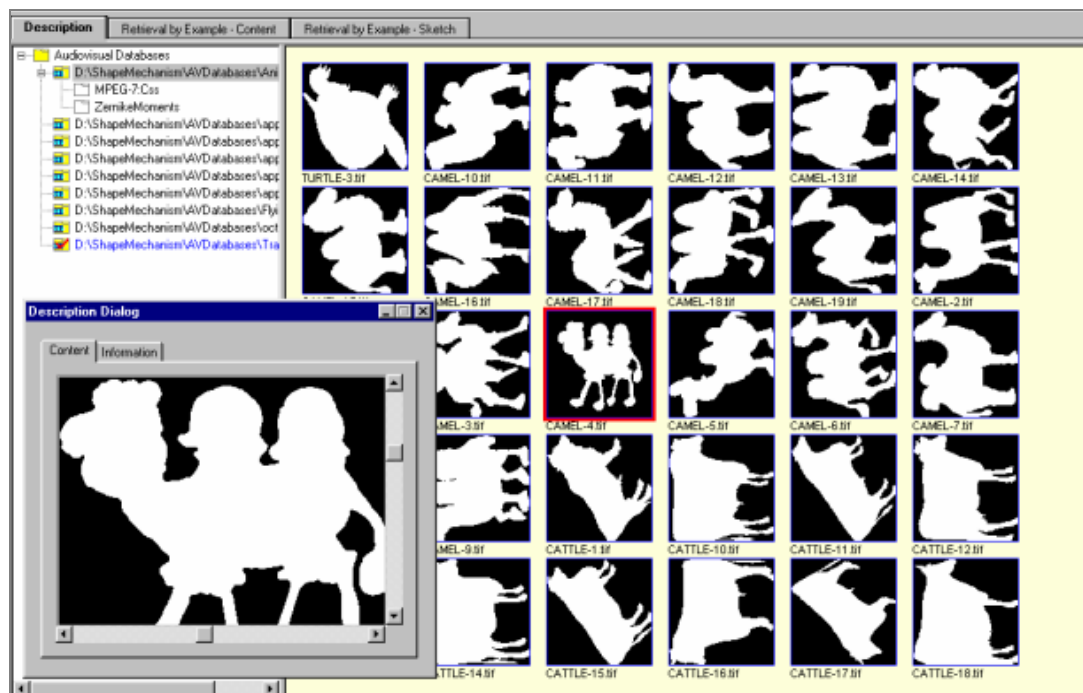
Com base nesta estrutura de dados, as interfaces de descrição e procura desta aplicação poderão estar sempre actualizadas sobre o conteúdo e descrições disponíveis tendo como ponto de contacto as bases de dados.

4.1.2 Estrutura da interface gráfica

A interface gráfica desenvolvida para esta aplicação apresenta três interfaces elementares:

1. **Interface de descrição - *Description*** - Permite gerar descrições do conteúdo disponível usando os descritores implementados (ver Figura 4.2a));
2. **Interface de procura baseada em conteúdo como exemplo - *Retrieval By Example-Content*** - Permite efectuar procuras sobre as descrições geradas usando uma imagem ou imagem binária como exemplo (ver Figura 4.2b));
3. **Interface de procura baseada em esboço como exemplo - *Retrieval By Example-Sketch*** - Permite efectuar procuras sobre as descrições geradas usando um esboço como exemplo (ver Figura 4.2c)).

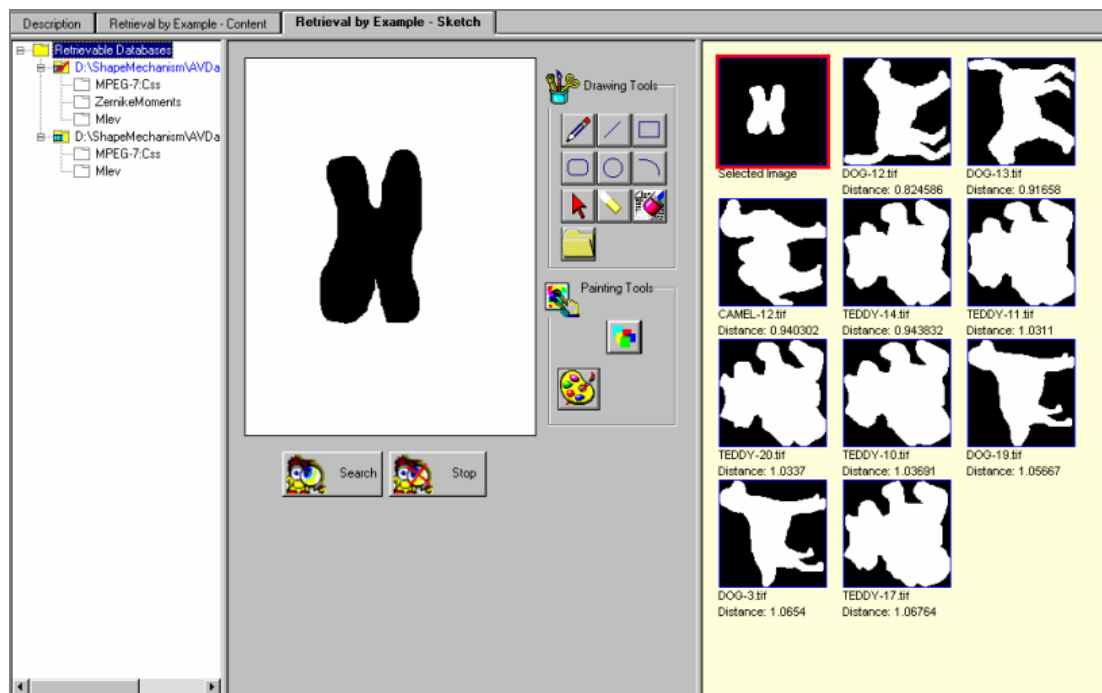
As duas interfaces de procura são usadas alternadamente na mesma área do ecrã ou seja a procura é feita usando conteúdo como exemplo OU um esboço como exemplo.



a)



b)



c)

Figura 4.2 – Interfaces da aplicação desenvolvida - a) de descrição; b) de procura baseada em conteúdo como exemplo e c) de procura baseada em esboço como exemplo

Como suporte à descrição da interface gráfica da aplicação, vai-se usar a organização espacial das interfaces no ecrã de modo a descrever mais facilmente os vários elementos que a compõem. Assim sendo, o ecrã da aplicação encontra-se dividido em quatro zonas, conforme mostra a Figura 4.3:

- Menu principal;
- Barra de ferramentas;
- Janela da base de dados;
- Janela de descrição e procura.

Note-se que a interface de descrição e as duas interfaces de procura acima referidas se alternam na zona do ecrã designada por janela de descrição e procura.

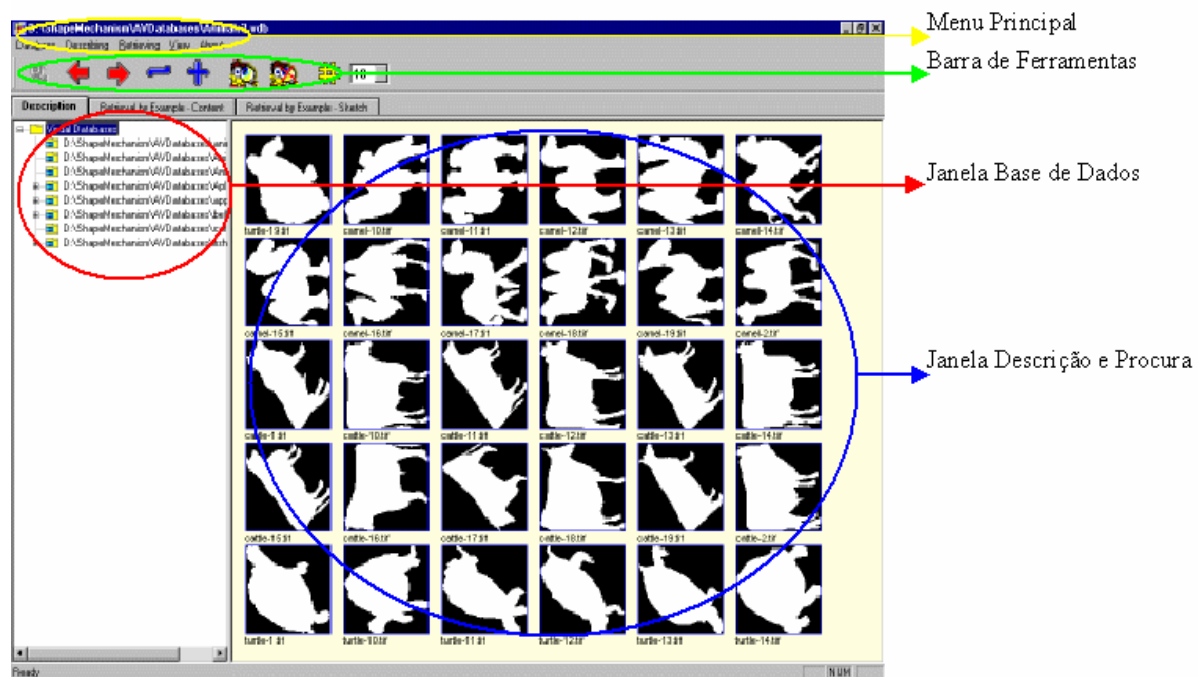


Figura 4.3 – Organização espacial do ecrã da aplicação

As secções seguintes descrevem cada uma destas zonas do ecrã a fim de conhecer as suas funcionalidades na aplicação.

4.2 Menu principal

O menu principal, apresentado na Figura 4.4, é uma componente importante da aplicação por incluir todas as operações da aplicação disponíveis para o utente, quer para o módulo da base de dados quer para os módulos de descrição e procura.

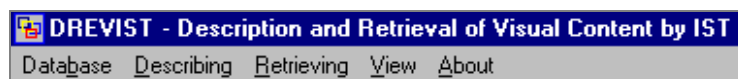


Figura 4.4 – Menu principal da aplicação desenvolvida

Assim, as operações permitidas por este menu subdividem-se essencialmente em três tipos:

- ♦ Operações sobre as bases de dados de conteúdo (sub-menu *Database*);
- ♦ Operações específicas à interface de descrição (sub-menu *Describing*);
- ♦ Operações específicas à interface de procura (sub-menu *Retrieving*).

O sub-menu ***Database***, apresentado na Figura 4.5, permite ao utente da aplicação realizar as seguintes funções sobre as bases de dados:

- ***Create*** - Permite a criação de novas bases de dados de conteúdo a partir das bases de dados de conteúdo já existentes ou a partir de conteúdo novo;

- **Ddelete** – Permite ao utente apagar bases de dados de conteúdo existentes;
- **Open** – Permite a visualização na estrutura em árvore das bases de dados de conteúdo escolhidas pelo utente da aplicação;
- **Open All** – Permite ao utente visualizar simultaneamente todas as bases de dados de conteúdo existentes;
- **Exit** – Permite a saída da aplicação.



Figura 4.5 – Sub-menu *Database* da aplicação desenvolvida

O sub-menu *Describing*, apresentado na Figura 4.6, permite ao utente da aplicação realizar as seguintes funções de descrição:

- **Selecting Databases** – Permite a selecção das bases de dados de conteúdo que se pretendem descrever; o utente poderá escolher simultaneamente todas as bases de dados através do elemento *All* do sub-menu ou escolher apenas algumas das bases de dados existentes através do elemento *By Choice*;
- **Removing Databases** – Permite a remoção das bases de dados de conteúdo que já não se pretendem usar no módulo de descrição; à semelhança do elemento anterior deste sub-menu, o utente poderá escolher simultaneamente todas as bases de dados disponíveis através do elemento *All* do sub-menu ou escolher apenas algumas delas através do elemento *By Choice*;
- **Description** – Permite descrever o conteúdo (bases de dados) previamente escolhido, usando os descritores de forma e/ou cor seleccionados para o efeito.

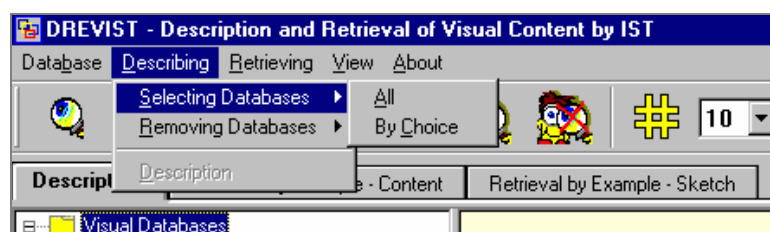


Figura 4.6 – Sub-menu *Describing* da aplicação desenvolvida

O sub-menu *Retrieving*, apresentado na Figura 4.7, permite ao utente da aplicação realizar as seguintes funções de procura:

- **By Content** – Permite seleccionar/activar a interface de procura baseada em conteúdo como exemplo (em alternativa à procura por esboço);
- **By Sketch** - Permite seleccionar/activar a interface de procura baseada em esboço como exemplo (em alternativa à procura por conteúdo);
- **Selecting Databases** – Permite a selecção das bases de dados de descrições que se pretendem usar para efectuar eventuais procuras; o utente poderá escolher simultaneamente todas as bases de dados de descrições disponíveis através do elemento *All* do sub-menu ou escolher apenas algumas delas através do elemento *By Choice*;
- **Removing Databases** – Permite a remoção das bases de dados de descrições que já não se pretendem usar no módulo de procura; à semelhança do elemento anterior deste sub-menu, o utente poderá remover simultaneamente todas elas através do elemento *All* do sub-menu ou remover apenas algumas delas através do elemento *By Choice*;
- **Retrieval** – Permite procurar conteúdo semelhante ao fornecido pelo utente ao mecanismo de procura, usando os descritores de forma e/ou cor seleccionados para o efeito.

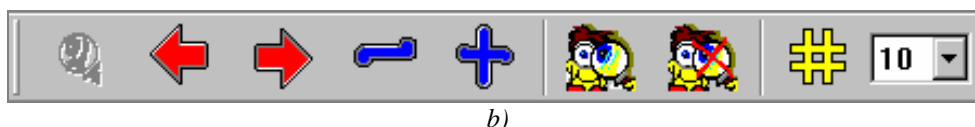


Figura 4.7 – Sub-menu *Retrieving* da aplicação desenvolvida

Os sub-menus *View* e *About* têm como funções a visualização ou não da barra de ferramentas e a visualização numa caixa de diálogo de informação relativa à versão da aplicação e à sua autora, respectivamente.

4.3 Barra de ferramentas

A barra de ferramentas da aplicação desenvolvida, apresentada na Figura 4.8, permite ao utente da aplicação navegar pelas bases de dados de conteúdo e de descrições, quer a partir da interface de descrição, quer a partir das interfaces de procura.



b)

Figura 4.8 – Barra de ferramentas da aplicação desenvolvida

Os botões desta barra de ferramentas permitem ao utente as seguintes funções:



- Visualizar a primeira base de dados de conteúdo na estrutura em árvore da janela da base de dados; nesta estrutura em árvore, as bases de dados encontram-se ordenadas por ordem alfabética;



- Visualizar o conteúdo da base de dados de conteúdo anterior relativamente à lista ordenada da estrutura em árvore da janela da base de dados;



- Visualizar o conteúdo da base de dados de conteúdo seguinte relativamente à lista ordenada da estrutura em árvore da janela da base de dados;



- Visualizar as imagens anteriores da base de dados de conteúdo aberta na janela de descrições e procura;



- Visualizar as imagens seguintes da base de dados de conteúdo aberta na janela de descrições e procura; e



- Define o número dos primeiros resultados correctos a visualizar resultantes de uma procura.

4.4 Janela da base de dados

A janela da base de dados representa nesta aplicação o módulo da base de dados na arquitectura proposta para o mecanismo de descrição e procura. As funções desta janela são listar numa estrutura em árvore as bases de dados de conteúdo existentes e que poderão vir a ser descritas e, posteriormente, as bases de dados de descrições existentes que poderão ser usadas para a realização de procuras através de medidas de semelhança adequadas aos descritores usados.

Para que esta janela possa desempenhar as funções inerentes ao módulo da base de dados foi usada uma estrutura em árvore, como mostra a Figura 4.9. Esta estrutura em árvore é responsável por apresentar toda a informação relativa às bases de dados (p.e. localização das bases de dados de conteúdo e de descrições, tipo de descrições geradas pelo módulo de descrição ou seja simultaneamente tipo de descritores que podem ser usados nas procuras a realizar através do módulo de procura). A informação armazenada nesta estrutura é usada para fornecer à aplicação todos os dados necessários para que se gere uma descrição ou se efectue uma procura.

A Figura 4.9 mostra as diferenças existentes relativamente à informação apresentada entre as janelas da base de dados usadas para a interface de descrição, Figura 4.9a) e de procura, Figura 4.9b). É de salientar que a diferença principal existente entre estas duas janelas é que na janela correspondente à interface de descrição aparecem todas as bases de dados seleccionadas pelo utente, apesar de algumas não terem ainda sido descritas, enquanto que na janela da base de dados relativa à interface de procura aparecem apenas as bases de dados para as quais já foram geradas descrições. Esta diferença permite ao utente no acto da procura escolher apenas bases de dados que já tenham sido descritas. Na Figura 4.9a) pode-se ver que a informação armazenada na estrutura em árvore são as bases de dados de conteúdo

escolhidas pelo utente da aplicação para serem descritas e o tipo de descritores até ao momento usados para descrever duas das bases de dados de conteúdo. Na Figura 4.9b) o que se pode ver são apenas as duas bases de dados para os quais foram geradas descrições e os descritores usados para as descrever.

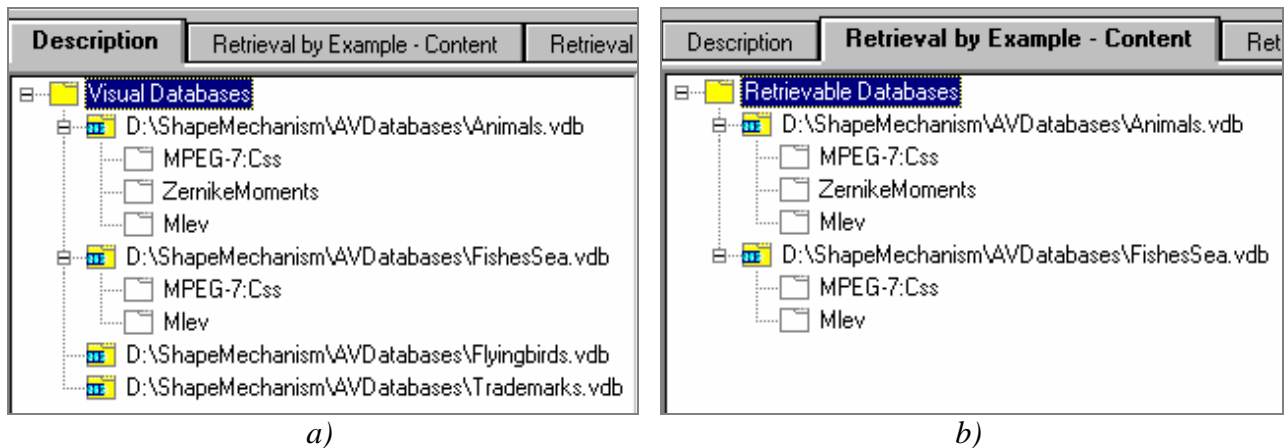


Figura 4.9 – Janela da base de dados: a) para a interface de descrição e b) para a interface de procura

Os vários níveis desta estrutura em árvore, relacionados com o tipo de informação armazenada, distinguem-se através do uso de ícones diferentes como os que se apresentam de seguida:



- Corresponde ao primeiro nível da estrutura em árvore de cada interface, identificando-a por *Visual Databases* ou *Retrievable Databases*;



- Corresponde ao segundo nível da estrutura em árvore de cada interface, definindo a localização das bases de dados de conteúdo ou de descrições escolhidas;



- Corresponde também ao segundo nível da estrutura em árvore de cada interface; neste caso, a base de dados está marcada o que significa que vai ser usada para a próxima descrição ou procura a ser efectuada;



- corresponde ao terceiro nível da estrutura em árvore da janela da base de dados, indicando os descritores usados na geração das descrições já disponíveis.

4.5 Janela de descrição e procura

A janela de descrição e procura representa nesta aplicação a interface do utente para o módulo de descrição e de procura. Esta janela está alternadamente associada a funções de descrição ou procura. As funções desta janela são de permitir a navegação e a visualização das bases de dados de conteúdo no caso da descrição e da procura baseada em conteúdo como exemplo, e o esboço da informação a procurar no caso da procura baseada em esboço como exemplo.

Para que esta janela possa desempenhar as funções inerentes ao módulo de descrição e de procura tem associadas três interfaces, a usar alternadamente, com funções e aspecto visual

distintos: a descrição e as procuras por conteúdo e por esboço. Na verdade, a interface de descrição é idêntica à interface de procura baseada em conteúdo como exemplo. Como se pode ver nas Figura 4.10 e Figura 4.11, as diferenças existentes entre as duas interfaces de descrição e procura por esboço são significativas e justificam-se pelas funções que lhes estão associadas.

A interface de descrição e de procura baseada em conteúdo como exemplo permite a navegação e a visualização do conteúdo das bases de dados tanto na descrição como na procura; no entanto, na procura é também nesta janela que a definição da procura é feita (neste caso, simples escolha de uma imagem) e de seguida apresentados os resultados respectivos. Durante a navegação e a visualização das bases de dados, na interface de descrição e de procura, as imagens na base de dados são apresentadas sequencialmente, uma a seguir à outra, aparecendo imediatamente em baixo o seu nome ou identificação. A única diferença entre as duas interfaces surge na procura, uma vez que a seguir à definição de uma procura, aparecem na interface os resultados ordenados por ordem decrescente da semelhança com a imagem ou esboço exemplo (a primeira imagem a aparecer é a imagem exemplo), aparecendo por debaixo de cada imagem o seu nome e o valor do grau de semelhança existente entre esta e a imagem/esboço exemplo.

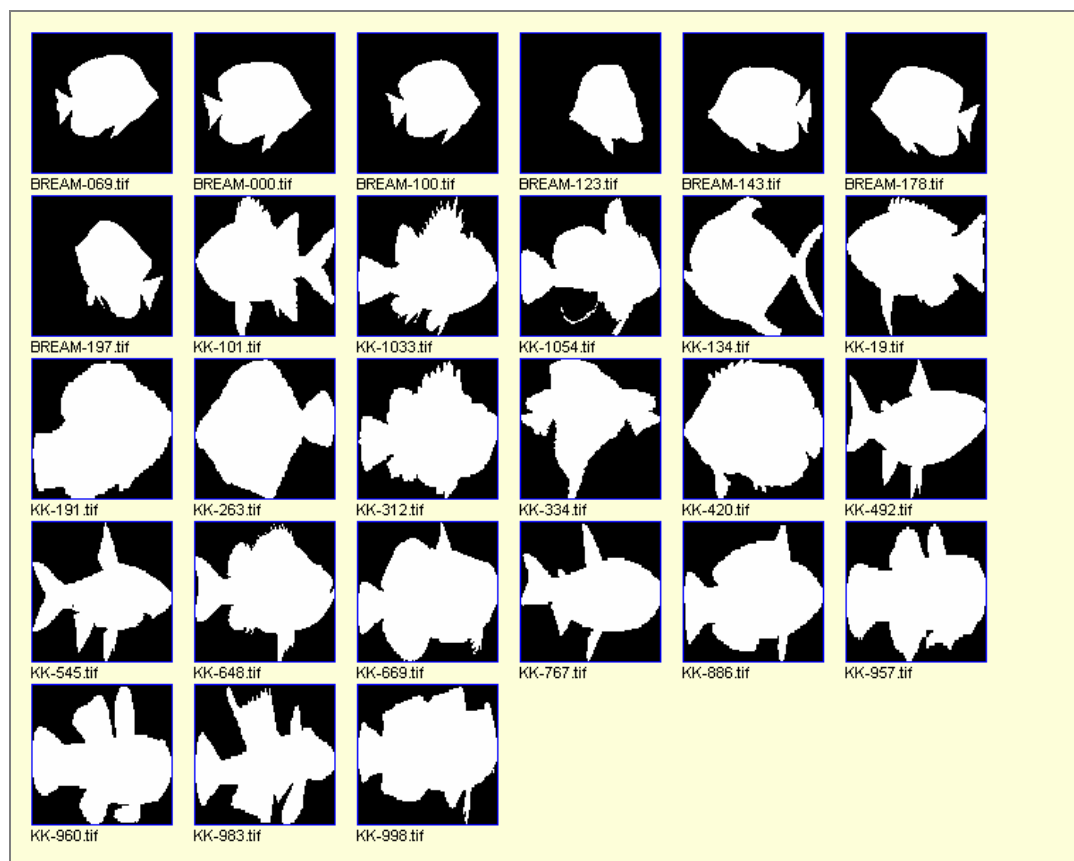


Figura 4.10 – Interface de descrição e de procura baseada em conteúdo como exemplo

A interface na Figura 4.11 é a interface usada na procura baseada em esboço como exemplo. Esta interface está sub-dividida em duas regiões: a primeira região permite definir a procura através de um esboço usando as ferramentas de desenho disponíveis enquanto a segunda região permite ainda a navegação e visualização das bases de dados de conteúdo e dos resultados de uma procura.

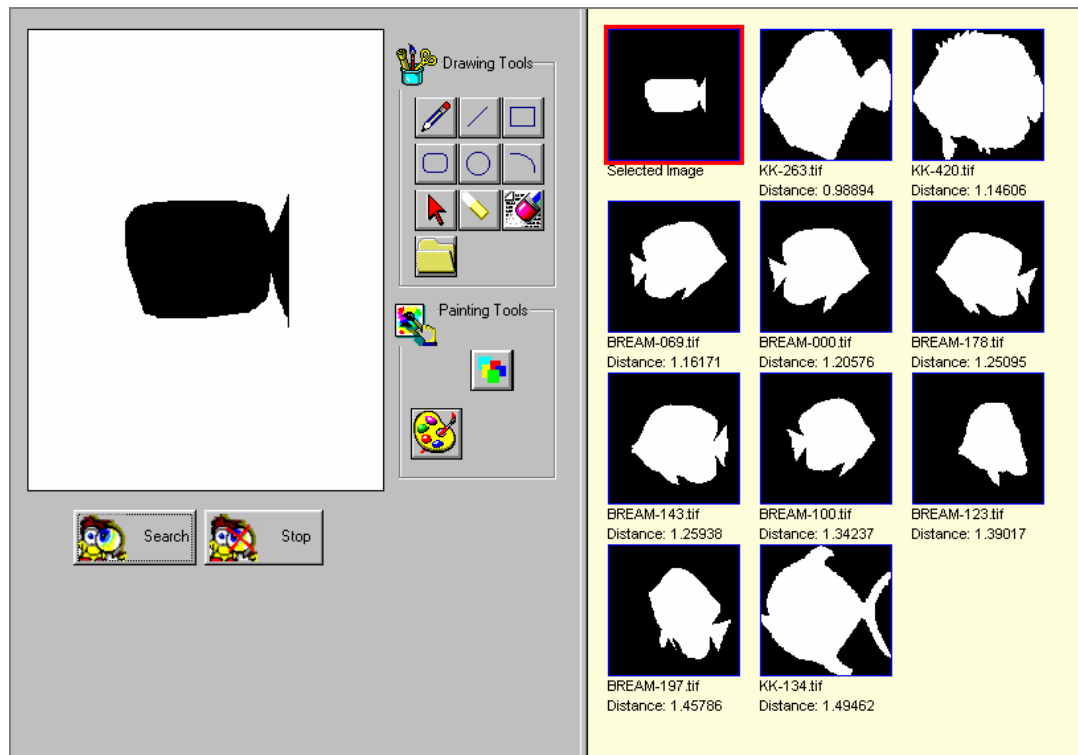


Figura 4.11 – Interface de procura baseada em esboço como exemplo

Uma vez definidas e apresentadas a arquitectura e a organização gráfica da aplicação desenvolvida, resta apresentar mais pormenorizadamente as propriedades e funções dos módulos de descrição e procura desta aplicação através das suas interfaces. Para além das funções específicas de cada interface, existem algumas funções comuns às três interfaces de descrição e procura que são as seguintes:

- ♦ Visualização das bases de dados disponíveis para descrição ou procura;
- ♦ Escolha ou remoção da(s) base(s) de dados da lista de bases de dados a descrever ou onde procurar;
- ♦ Escolha dos descritores a usar na descrição ou na procura.

As secções seguintes descrevem as propriedades e funções dos módulos de descrição e procura, tendo em conta as funções específicas e comuns às três interfaces, a fim de se conhecerem todas as funcionalidades desta aplicação.

4.6 Módulo de descrição

A interface de descrição, correspondente ao módulo de descrição da arquitectura adoptada, é a interface onde o utente do mecanismo de descrição e procura gera as descrições para a informação visual armazenada na base de dados de conteúdos usando um dado conjunto de características visuais e descritores correspondentes. Estas descrições são então armazenadas na base de dados de descrições para posterior comparação durante o processo de procura.

A interface de descrição definida para esta aplicação disponibiliza dois tipos de funcionalidades ao utente da aplicação:

- ♦ Navegação pelas bases de dados de conteúdo;
- ♦ Geração de descrições para as bases de dados de conteúdo seleccionadas usando os descritores escolhidos entre os disponíveis.

As secções seguintes descrevem detalhadamente estas duas funcionalidades, de modo a perceber-se melhor como operar para as realizar no contexto da aplicação desenvolvida.

4.6.1 Navegação pelas bases de dados de conteúdo

A navegação pelas bases de dados de conteúdo pode, nesta aplicação, ser feita globalmente (entre bases de dados), ou localmente (dentro de cada base de dados). A navegação global pode ser feita usando-se a barra de ferramentas apresentada na secção 4.1.2 ou carregando-se duas vezes com o botão esquerdo do rato sobre o elemento da estrutura em árvore com o nome da base de dados a visualizar. A denominada navegação local é feita na janela de descrição e procura, permitindo ao utente escolher, com o botão esquerdo do rato, a imagem que deseja ver, em tamanho original, numa caixa de diálogo; isto porque, na verdade, a imagem que aparece na janela de visualização da base de dados foi reajustada (aumentada ou diminuída) de acordo com as dimensões especificadas pela aplicação (100x100 *pixels*) de modo a caberem várias imagens de cada vez no ecrã da aplicação. O utente poderá eliminar esta caixa de diálogo carregando, com o botão direito do rato, sobre a imagem seleccionada na janela de visualização do conteúdo. Alternativamente à visualização da imagem estática original, ter-se-á acesso, numa outra janela da caixa de diálogo, a informação relativa à imagem, nomeadamente o seu nome, localização e dimensões, como se pode verificar na Figura 4.12.

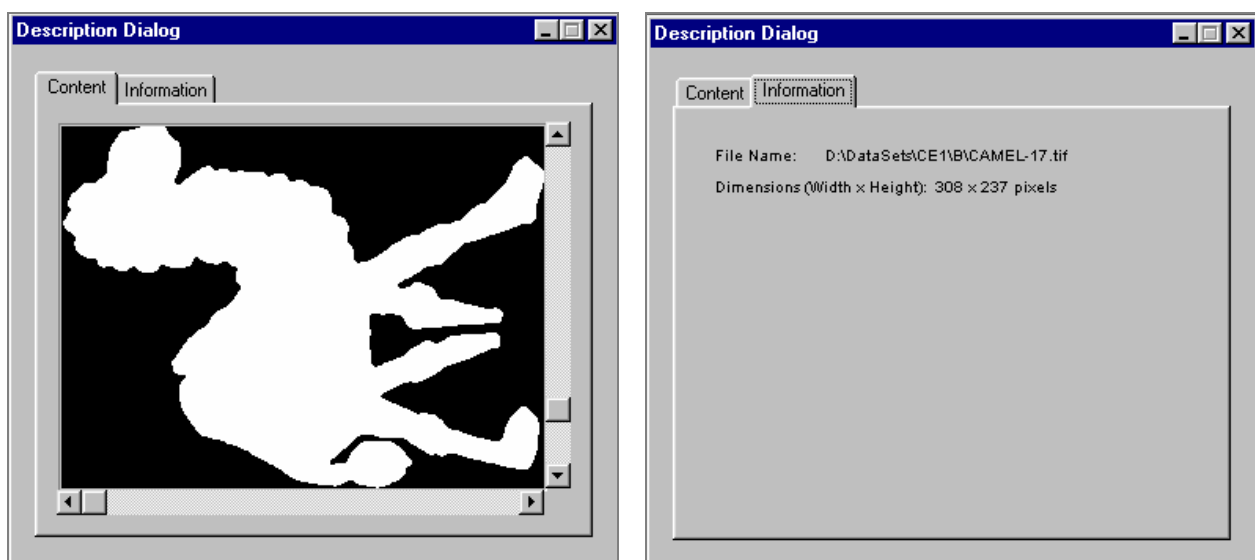


Figura 4.12 – Caixa de diálogo com a) imagem binária original e b) informação relativa à imagem






4.6.2 Geração de descrições


A geração de descrições na aplicação desenvolvida é feita em duas etapas:



1. Selecção das base de dados de conteúdo que se pretendem descrever, e
2. Escolha dos descritores a usar para descrever as bases de dados de conteúdo previamente seleccionadas.

▪ Selecção das bases de dados de conteúdo a descrever

A selecção das bases de dados de conteúdo a descrever pode ser feita de dois modos diferentes:

- **Usando o menu principal da aplicação** – A selecção da(s) base(s) de dados de conteúdo a descrever através do menu principal é feita através do sub-menu *Describing* (ver Figura 4.6). Neste sub-menu, o utente tem as seguintes possibilidades:
 - ♦ *Selecting Databases -> ByChoice* – Escolhem-se apenas as base de dados de conteúdo que o utente pretende descrever, através de uma caixa de diálogo e assinalando-as de imediato com o símbolo .
 - ♦ *Selecting Databases -> All* – Escolhem-se todas as bases de dados de conteúdo disponíveis, assinalando-as de imediato com o símbolo .
 - ♦ *Removing Databases-> ByChoice* – Remove(m)-se a(s) base(s) de dados que não deseja se descrever, marcando-a novamente com o símbolo de defeito ou seja o símbolo ; e
 - ♦ *Removing Databases-> All* – Removem-se todas as bases de dados de conteúdo seleccionadas para serem descritas, marcando-as novamente com o símbolo de defeito ou seja o símbolo .
- **Usando o menu *popup* da janela das bases de dados** - A selecção da(s) base(s) de dados de conteúdo a descrever pode também ser feita através do menu *popup* da janela das bases de dados, apresentado na Figura 4.13. Neste menu, o utente tem à disposição as seguintes funções associadas ao módulo de descrição:
 - ♦ *Selecting All* – Permite escolher simultaneamente todas as bases de dados de conteúdo na estrutura em árvore, assinalando-as de imediato com o símbolo .
 - ♦ *Selecting ByChoice* – Permite escolher as base de dados de conteúdo a descrever, seleccionando com o botão esquerdo do rato o elemento representativo dessa base de dados na árvore das bases de dados, e

pressionando de seguida com o botão direito do rato para activar o menu *popup*, ficando imediatamente marcada com o símbolo .

- ♦ *Removing All* – Permite desfazer a selecção de todas as bases de dados de conteúdo seleccionadas, marcando-as novamente com o símbolo de defeito ou seja o símbolo .
- ♦ *Removing ByChoice* – Permite escolher uma a uma as bases de dados de conteúdo cuja selecção deve ser removida, marcando-as novamente com o símbolo .

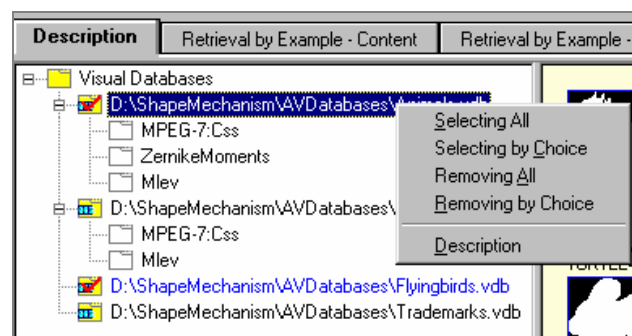


Figura 4.13 – Menu *popup* da janela das bases de dados

▪ Selecção dos descritores

A selecção dos descritores a usar na descrição a criar é feita no menu principal, seleccionando o elemento *Describing->Description* ou o elemento *Description* do menu *popup* (accionado através do botão direito do rato). A selecção de um destes elementos faz aparecer uma caixa de diálogo com a lista dos descritores que a aplicação permite usar para descrever a(s) base(s) de dados de conteúdo seleccionadas, ver Figura 4.14.

De entre as características visuais mais comumente conhecidas, i.e. cor, textura, forma, movimento e relação espacial, a cor e a forma são, neste momento, as características disponíveis nesta aplicação. A caixa de diálogo com a lista dos descritores que podem ser usados, aparece por defeito com dois descritores do MPEG-7 seleccionados: o *MPEG-7 Dominant Color* para a cor e o *MPEG-7 ART* para a forma. Caso o utente queira alterar estas sugestões, poderá fazê-lo retirando os descritores que não quer usar e seleccionando outros disponíveis na aplicação ou seja os descritores *CSS*, *Turning Angles*, momentos de *Zernike* e *MLEV* para a característica de forma e *Compact Color*, *Color Layout* e *Color Structure Histogram* para a característica de cor.

Após a selecção do descritor ou descritores pretendidos para a descrição do conteúdo, o utente terá apenas que pressionar o botão *OK* para dar início à geração das descrições que no fim serão armazenadas na directoria *DescriptionsDatabases*, para futura utilização pelo módulo de procura.

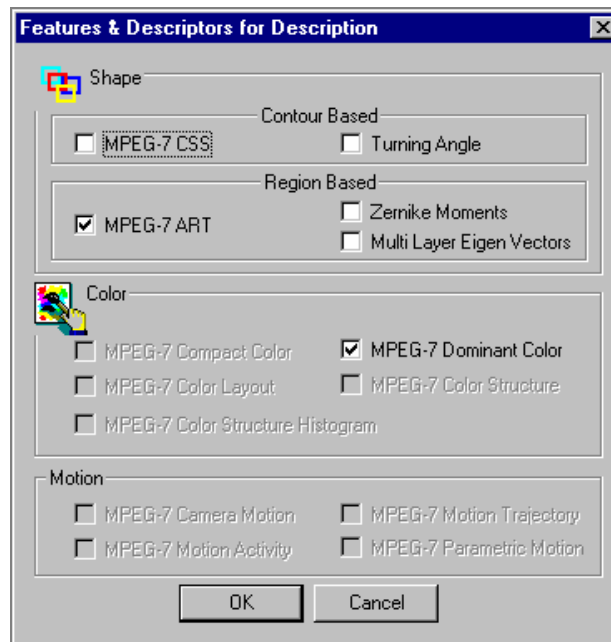


Figura 4.14 – Caixa de diálogo para a selecção dos descritores a usar no processo de descrição

4.7 Módulo de procura

A interface de procura, correspondente ao módulo de procura na arquitectura adoptada, é a interface onde:

- O utente do mecanismo de procura fornece informação que caracteriza de algum modo o conteúdo que se pretende encontrar; essa informação pode ser fornecida através de um exemplo, através de alguns parâmetros de descrição ou através de outras propriedades mais intuitivas para o utente e que o mecanismo se encarrega de traduzir de forma conveniente em termos de parâmetros de descrição (ver secção 3.1.3);
- Numa procura por exemplo (como aquelas usadas no mecanismo implementado nesta tese), são calculados os parâmetros de descrição para a informação de procura dada pelo utente; os parâmetros assim obtidos ou aqueles directamente fornecidos são então enviados ao módulo da base de dados (eventualmente codificados como uma descrição MPEG-7 se apenas descritores MPEG-7 forem usados) para comparação com as descrições disponíveis;
- Os resultados da procura feita através de comparação com as descrições disponíveis na(s) base(s) de dados são visualizados.

Uma vez que as interfaces de procura de cada mecanismo são fortemente determinadas pelo tipo de aplicação e o nível de conhecimentos e experiência dos utilizadores à qual se destinam, optou-se por criar dois tipos de interface de procura nesta aplicação: **procura baseada em conteúdo como exemplo** e **procura baseada em esboço como exemplo**.

Apesar de ambas as interfaces definirem a procura a partir de um exemplo, o modo como o fazem é bastante diferente apresentando consequentemente capacidades bem distintas.

As interfaces de procura definidas nesta aplicação, disponibilizam três tipos de funcionalidades ao utente:

- ♦ Navegação pelas bases de dados descritas;
- ♦ Definição da procura baseada num exemplo (por conteúdo ou por esboço);
- ♦ Visualização dos resultados da procura obtidos nas bases de dados de conteúdo previamente seleccionadas.

As secções seguintes descrevem, detalhadamente, estes três tipos de funcionalidades, para cada um dos dois modos de definir a procura, de modo a definir precisamente os passos necessários para a sua execução.

4.7.1 Navegação pelas bases de dados com descrições

A navegação pelas bases de dados com descrições pode, à semelhança da navegação pelas bases de dados de conteúdo, ser feita usando a barra de ferramentas definida na secção 4.1.2 ou carregando-se duas vezes com o botão esquerdo do rato sobre o elemento da estrutura em árvore com o nome da base de dados correspondente a visualizar. Neste tipo de interface de procura, a navegação pelas bases de dados pode durar até que se encontre uma imagem que exprima de forma suficientemente adequada a procura que se pretende realizar, de modo a poder encontrar-se conteúdo realmente semelhante ao pretendido.









4.7.2 Definição da procura

O processo de definição da procura nesta aplicação é feito de dois modos diferentes o que dá origem a duas interfaces com capacidades distintas: uma baseada em conteúdo como exemplo e outra baseada em esboço como exemplo.

O modo como cada uma das interfaces de procura disponíveis na aplicação desenvolvida permite definir a procura será descrito de seguida. No entanto, existem duas funções comuns a ambos os tipos de procura que é a selecção das bases de dados de descrições onde se pretende realizar a procura e a selecção dos descritores a usar para efectuar a procura, pelo que a descrição destas duas tarefas será feita em primeiro lugar, definindo-se de seguida os métodos disponíveis para a definição de uma procura.

▪ Selecção das bases de dados de descrições

A selecção das bases de dados de descrições onde a procura se vai efectuar pode ser feita de dois modos diferentes:

- **Usando o menu principal da aplicação** – A selecção da(s) base(s) de dados de descrições através do menu principal é feita usando o sub-menu *Retrieving* (ver Figura 4.7). Neste sub-menu, o utente tem as seguintes possibilidades:
 - ♦ *Selecting Databases -> ByChoice* – Escolhem-se apenas as base de dados de conteúdo que o utente pretende usar para a procura, através de uma caixa de diálogo, ficando estas imediatamente assinaladas com o símbolo ;
 - ♦ *Selecting Databases -> All* – Escolhem-se todas as bases de dados de conteúdo disponíveis para uso durante a fase de procura, ficando estas imediatamente assinaladas com o símbolo ;
 - ♦ *Removing Databases-> ByChoice* – Remove(m)-se a(s) base(s) de dados que não se deseja usar na procura, através de uma caixa de diálogo, ficando estas novamente marcadas com o símbolo de defeito ou seja o símbolo ; e
 - ♦ *Removing Databases-> All* – Removem-se todas as bases de dados de conteúdo seleccionadas para o uso durante a fase de procura, ficando estas novamente marcadas com o símbolo de defeito ou seja o símbolo .
- **Usando o menu *popup* da janela das bases de dados** - A selecção da(s) base(s) de dados de descrições pode também ser feita através do menu *popup* da janela das bases de dados, apresentado na Figura 4.13, e activado ao carregar-se no botão direito do rato. Neste menu, o utente tem à disposição as seguintes funções sobre o módulo de procura e consequentemente nas interfaces de procura:
 - ♦ *Selecting All* – Permite seleccionar simultaneamente todas as bases de dados de descrições na estrutura em árvore, marcando-as com o símbolo ;
 - ♦ *Selecting ByChoice* – Permite escolher as bases de dados de descrições a usar, seleccionando com o botão esquerdo do rato o elemento representativo dessa base de dados na árvore das bases de dados, e pressionando de seguida com o botão direito da rato para activar o menu *popup*, ficando estas marcadas com o símbolo ;
 - ♦ *Removing All* – Permite desfazer simultaneamente a selecção de todas as bases de dados de descrições seleccionadas, marcando-as novamente com o símbolo ;
 - ♦ *Removing ByChoice* – Permite seleccionar as bases de dados de descrições cuja selecção deve ser removida, marcando-as novamente com o símbolo ;

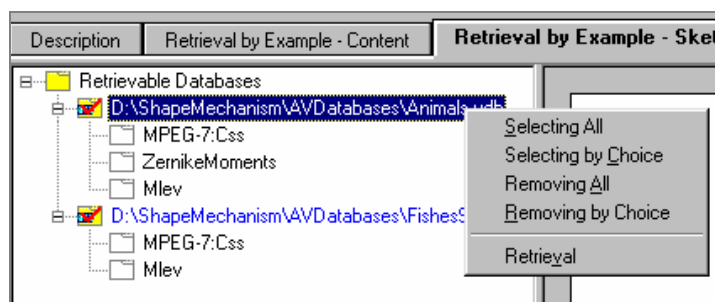


Figura 4.15 – Menu *popup* da janela das bases de dados

▪ Selecção dos descritores a usar no acto da procura

A selecção dos descritores a usar para a realização da procura é feita através de uma caixa de diálogo (ver Figura 4.16), onde aparecem já seleccionados todos os descritores comuns às descrições para as bases de dados seleccionadas para o processo de procura. Caso o utente da aplicação não queira utilizar algum(ns) dos descritores, terá apenas que remover a sua selecção. Aparece ainda nesta caixa de diálogo três barras deslizantes que permitirão definir o peso a atribuir à medida de semelhança de cada uma das características usadas na procura.

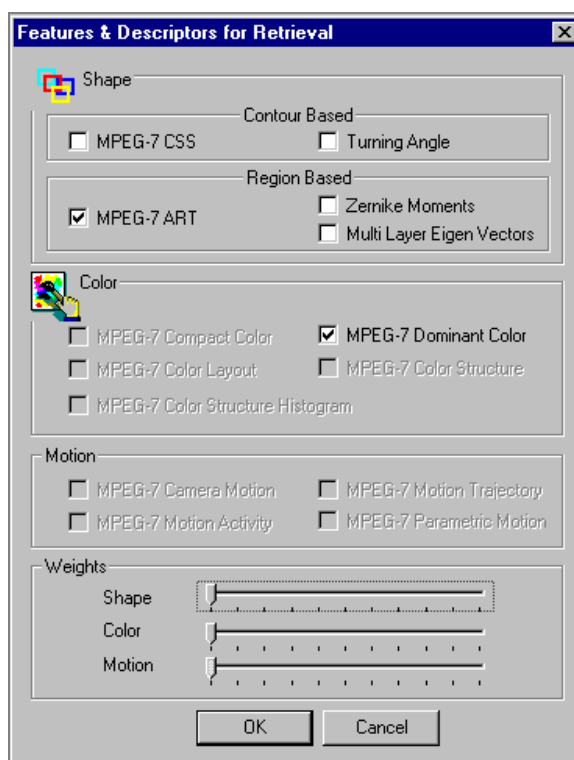


Figura 4.16 – Caixa de diálogo para selecção dos descritores a usar no processo de procura

▪ Definição da procura baseada em conteúdo como exemplo

A definição da procura através da interface de procura baseada em conteúdo como exemplo é feita fornecendo ao mecanismo de procura uma das imagens binárias com a forma de um objecto disponíveis; esta imagem binária deve ser representativa do tipo de forma que se

pretende encontrar. Podem também utilizar-se imagens de objectos com textura, e nesse caso poderão também utilizar-se na procura descritores de cor.

A escolha da imagem exemplo é feita após a navegação nas bases de dados disponíveis, escolhendo-se a imagem exemplo que melhor representa o conteúdo visual que se procura. Esta imagem é seleccionada de entre as imagens visualizadas na janela de descrição e procura, pressionando-se o botão esquerdo do rato sobre a imagem escolhida. Uma vez seleccionada a imagem a fornecer ao mecanismo de procura, esta passará a apresentar uma moldura de cor vermelha à sua volta, expressando o facto de ser esta a imagem exemplo. De seguida, o sistema compara a sua descrição (obtida a partir do *software* de descrição disponível) com as descrições disponíveis nas bases de dados onde se pretende fazer a procura e previamente seleccionadas e envia à interface de procura o resultado da procura, ordenando as imagens segundo o seu grau de semelhança com a imagem exemplo, medido usando uma métrica pré-estabelecida. De seguida, o utente poderá refinar a sua procura escolhendo de entre os resultados da primeira procura, uma outra imagem exemplo que se assemelhe mais ao que procura e assim sucessivamente.

▪ Definição da procura baseada em esboço como exemplo

A interface de procura baseada em esboço como exemplo fornece ao mecanismo de procura um esboço desenhado pelo utente; este esboço deve ser representativo das formas que este pretende encontrar. O esboço poderá ser obtido de dois modos distintos:

- **Esboço livre** – O utente faz o esboço daquilo que pretende encontrar, de modo livre, sobre a tela de desenho, como mostra a Figura 4.17b), usando ferramentas de desenho definidas através dos botões que se encontram sobre a janela de esboço:



- Desenha traços contínuos, guiados pelo utente;



- Desenha uma circunferência de dimensão definida pelo utente;



- Desenha um rectângulo com os cantos redondo de dimensão definida pelo utente;



- Desenha um rectângulo de dimensão definida pelo utente;



- Desenha rectas de direcção e comprimento definidos pelo utente;







- Desenha arcos de direcção e comprimento definidos pelo utente;

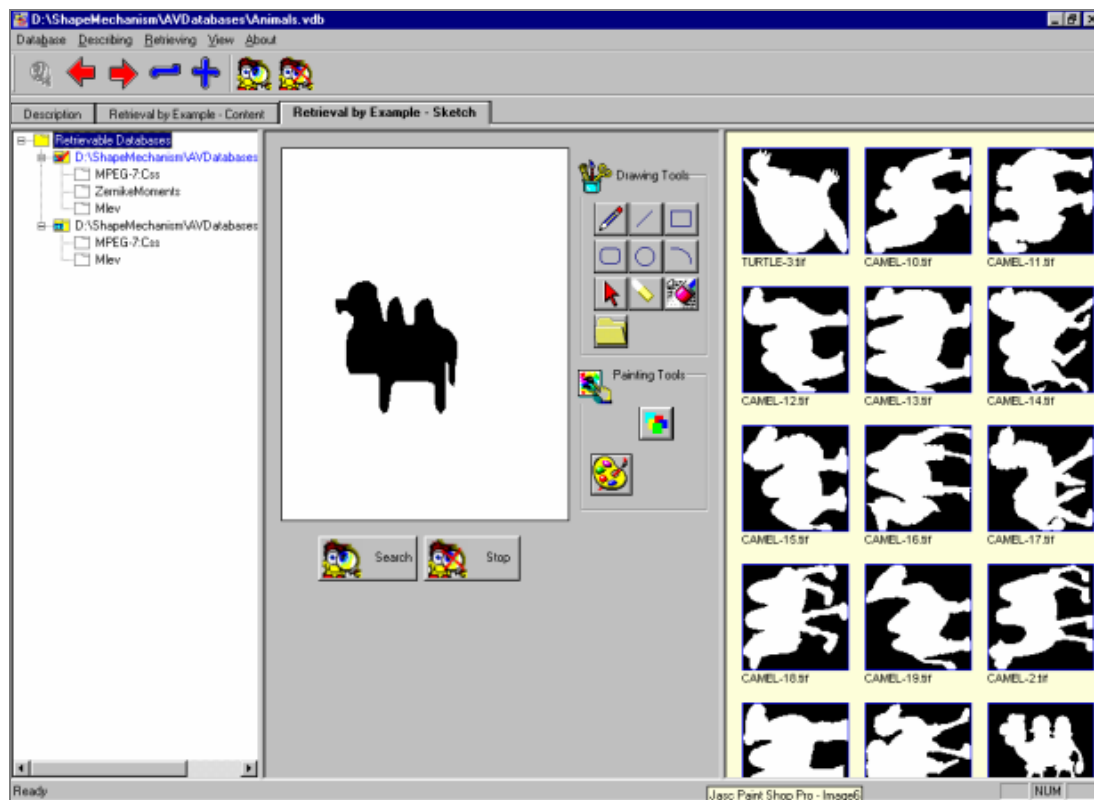


- Selecciona o objecto sobre o qual se pretende executar uma dada acção específica;



- Apaga o objecto seleccionado;

-  - Apaga todos os objectos sobre a tela de desenho;
-  - Chama a paleta de cores permitindo seleccionar uma cor para pintar o esboço feito;
-  - Colora o interior do objecto seleccionado com a cor escolhida através da paleta; e
-  - Abre a caixa de diálogo que permite seleccionar e abrir uma imagem.

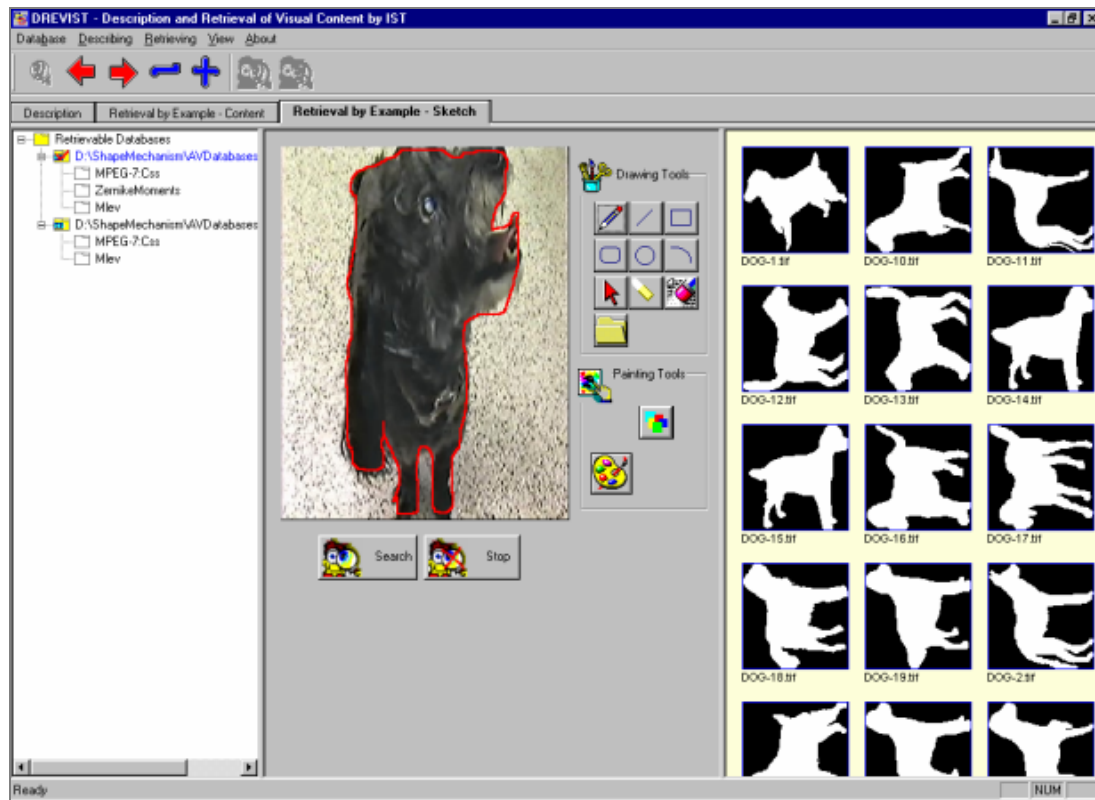


b)

Figura 4.17 – Interface de procura baseada em esboço livre como exemplo

- **Esboço sobre imagem** – O utente faz o esboço do que pretende encontrar sobre uma imagem que deverá conter algo semelhante ao que se procura, o que melhora normalmente a qualidade do esboço; este método é particularmente adequado para os utentes com menos jeito para o desenho, ver a Figura 4.18;

Uma vez desenhado o esboço (que pode ser colorido se se pretenderem usar na procura também descritores de cor) com a qual se pretende executar o processo de procura, o sistema compara a sua descrição com as descrições disponíveis nas bases de dados seleccionadas para a procura e mostra, à semelhança do que também acontece para as procuras baseadas em conteúdo como exemplo. Finalmente, aparece na janela de descrição e procura o resultado das comparações, com as imagens resultado ordenadas segundo o seu grau de semelhança com o esboço feito.







a)

Figura 4.18 – Interface de procura baseada em esboço sobre uma imagem referência como exemplo

A definição da procura através de qualquer um destes métodos será feita usando as ferramentas de desenho disponibilizadas pela interface. Com estas ferramentas, o utente poderá desenhar a forma do objecto que procura e eventualmente dar-lhe cor para efectuar uma procura conjunta de forma e cor.

Os processos de procura baseados em esboço sobre uma imagem (incluída ou não nas bases de dados disponíveis) e em esboço livre são semelhantes, com excepção de que no caso do esboço livre não é necessário seleccionar a imagem de referência sobre a qual se desenha o esboço. Neste caso, ignora-se para a procura definida através de um esboço livre a primeira das etapas definidas de seguida:

1. Selecciona-se a imagem sobre a qual se vai fazer o esboço através do botão  que fará aparecer uma caixa de diálogo onde se poderá indicar a localização dessa imagem;
2. De seguida, usando a ferramenta de desenho , desenha-se sobre a imagem aberta a forma do objecto a procurar;
3. No caso de uma procura conjunta de forma e cor, o utente poderá ainda usar a *pallette* de cores, seleccionada através da ferramenta de desenho  com o objectivo de

seleccionar a cor pretendida, e dar ordem para colorir o objecto usando a ferramenta de desenho ;

4. Por último, são escolhidos os descritores, de entre os comuns às bases de dados de descrições seleccionadas, a usar na procura através de uma caixa de diálogo igual à apresentada na Figura 4.16.

A diferença entre a definição da procura através de esboço sobre imagem e esboço livre tem apenas a ver com o fundo usado para a tela de desenho. No caso do esboço livre, a tela é branca, não fornecendo qualquer ajuda ao utente e dando-lhe total liberdade em termos de desenho, enquanto que no caso do esboço sobre imagem existe uma imagem a fazer de tela que auxilia o desenho do esboço.

4.7.3 Apresentação dos resultados da procura

A apresentação dos resultados após a definição da procura e da escolha dos descritores a usar no processo de procura é feita na janela mais à direita do ecrã. Nesta janela, aparecerá em primeiro lugar e com uma moldura vermelha, a imagem ou esboço exemplo fornecidos ao mecanismo de procura, seguida dos resultados obtidos, ordenados segundo o seu grau de semelhança com a imagem ou esboço exemplo para a medida de semelhança usada. A Figura 4.19 apresenta a janela de resultados: esta janela inclui o nome das imagens e o valor da distância de semelhança entre cada imagem resultado e a imagem exemplo.



Figura 4.19 – Janela para apresentação dos resultados da procura

4.8 Comentários Finais

Este capítulo teve como objectivo apresentar e descrever a aplicação desenvolvida no âmbito desta tese, ou seja, um mecanismo de descrição e procura essencialmente baseado na característica de forma dos objectos de vídeo. A aplicação desenvolvida vai permitir a avaliação e comparação do desempenho dos descritores de forma do MPEG-7 assim como de outros descritores de forma alternativos, como os ângulos de curvatura do contorno, os momentos de *Zernike* e os vectores próprios multi-nível (*MLEV*). A novidade desta aplicação relativamente a outros mecanismos existentes na *Internet* recaiu precisamente sobre o facto de se terem usado os descritores normalizados pelo MPEG-7, garantindo-se deste modo a interoperabilidade com outros mecanismos em termos das descrições de forma se apenas descritores MPEG-7 forem usados.

Capítulo 5

Avaliação dos descritores de forma

considerados e adoptados pelo MPEG-7

O panorama tecnológico na área das aplicações audiovisuais mudou consideravelmente com a transição das representações analógicas dos sinais para as representações digitais, sendo por isso essencial que as normas correspondentes o reconheçam e se adaptem, ajustando o seu objectivo de especificação. As normas devem oferecer interoperabilidade entre as várias áreas, serviços e aplicações, integrando horizontalmente, ou seja, não se vocacionando apenas para sistemas específicos de uma única área, serviço ou aplicação. De acordo com o desenvolvimento tecnológico e com a consciencialização das partes envolvidas neste processo, o grupo MPEG (*Moving Picture Experts Group*) da ISO (*International Standards Organization*) adoptou uma metodologia de trabalho para a definição de normas na área da representação audiovisual que considera estes aspectos e se tem revelado extremamente bem sucedida pelos resultados alcançados. O grande princípio estruturante das normas MPEG baseia-se na especificação de um número mínimo de ferramentas - *uma ferramenta por cada funcionalidade oferecida, sem redundância* - que ao serem combinadas de acordo com os requisitos do domínio de aplicação em questão, garantem a máxima interoperabilidade, minimizando a complexidade e o custo dos sistemas envolvidos [PeKo1998].

Esta metodologia de trabalho do MPEG tem proporcionado enormes sucessos no que diz respeito à definição de normas, nomeadamente as normas MPEG-1 [MPEG1-1993] onde se inclui o famoso MP3 para a codificação de áudio, as normas MPEG-2 [MPEG2-1996] mundialmente usadas para a televisão digital e DVD e as normas MPEG-4 [MPEG4-1999] cada vez mais usadas para aplicações interactivas, móveis e na *Internet*.

A metodologia de desenvolvimento das normas adoptada pelo MPEG baseia-se numa sequência de seis processos principais:

1. Identificação de aplicações relevantes para a norma em questão;
2. Identificação das funcionalidades necessárias para as aplicações acima identificadas;
3. Identificação dos requisitos técnicos relevantes, de acordo com as funcionalidades anteriormente identificadas, de forma que requisitos comuns possam ser identificados para aplicações diferentes;
4. Identificação dos requisitos que são comuns às diferentes áreas de interesse, e ainda dos que não sendo comuns são ainda assim relevantes;
5. Especificação das ferramentas necessárias para satisfazer os requisitos técnicos identificados através de:
 - i. Chamada e avaliação de propostas – Pedido a todas as entidades interessadas, empresas, universidades, centros de investigação, etc. que submetam tecnologia relevante para o problema em questão, com o intuito de reunir a melhor tecnologia disponível nesse domínio científico, para avaliação pelo MPEG. Esta fase é normalmente designada por fase competitiva.
 - ii. *Core Experiments (CE)* – Processo de desenvolvimento da melhor tecnologia a partir das melhores propostas técnicas submetidas à chamada de propostas, baseada na avaliação e comparação com critérios pré-estabelecidos e precisos, adequados ao objectivo em questão. Esta fase é normalmente designada por fase colaborativa.
6. Verificação das ferramentas desenvolvidas, integradas num modelo de trabalho, *Working Model*, para garantir que estas possam ser combinadas de modo a oferecer as funcionalidades necessárias às aplicações relevantes. Para a norma MPEG-7, este modelo de trabalho é designado por *eXperimentation Model (XM)*.

Este processo não é rígido: alguns passos podem ser repetidos mais do que uma vez e várias iterações dos mesmos podem inclusive ser necessárias. No entanto, o calendário de desenvolvimento deste processo é estabelecido no início do processo e seguido à risca para que as soluções tecnológicas possam aparecer no mercado de forma rápida e eficiente e as empresas possam efectivamente fazer planos baseando-se na disponibilidade atempada das normas relevantes.

E é seguindo este processo de normalização que o MPEG se encontra de momento a desenvolver a norma MPEG-7, de modo a alcançar os objectivos estabelecidos pelo grupo no seu documento intitulado “*MPEG-7: Context, Objectives and Technical Roadmap*” [MPEG7-Cont1999]. Neste contexto, os objectivos deste capítulo são:

- Apresentar pormenorizadamente os passos dados no desenvolvimento desta norma, com especial ênfase nos parâmetros de forma propostos ao MPEG-7 como candidatos a descritores;
- Analisar e comparar o desempenho dos descritores de forma considerados na fase de competição;

- Reproduzir todos os core experiments associados à descrição de forma no contexto do eXperimentation Model (XM) do MPEG-7 (descrito adiante) ou seja obter resultados para o desempenho de todos os descritores em questão e implementar em software os descritores não disponíveis no XM (atendendo a que alguns dos descritores propostos não chegaram a ser integrados no XM);
- Descrever e analisar a sintaxe e a semântica referentes a cada descritor adoptado pelo MPEG-7.

5.1 MPEG-7: Processo de Normalização

Ao longo do processo de identificação dos domínios de aplicação relevantes para a norma MPEG-7 foram identificados como oferecendo grande potencial, a educação, o jornalismo, o turismo, o entretenimento, a biomédica, a vigilância, etc. [MPEG7-App1998]. Esta lista tem tendência a aumentar e consequentemente a ser enriquecida no futuro, aparecendo apenas como uma lista de sugestões e possibilidades para despertar a apetência das empresas e utentes relevantes. Uma vez que a análise detalhada das funcionalidades e requisitos definidos para a norma MPEG-7 está fora do âmbito desta tese, sugere-se a leitura do documento MPEG-7 denominado “*MPEG-7 Applications*” [MPEG7-App1998] para um conhecimento mais aprofundado desta matéria.

Após a identificação das aplicações mais relevantes, bem como das funcionalidades e requisitos, o MPEG-7 fez de seguida uma chamada de propostas técnicas a todas as entidades interessadas para que estas submetessem as suas tecnologias relevantes para avaliação.

A chamada de propostas feita pelo MPEG, em Outubro de 1998, incluiu tanto ferramentas normativas como ferramentas não-normativas (ferramentas cuja especificação normativa não é essencial para garantir interoperabilidade), bem como metodologias de trabalho, nomeadamente:

- Ferramentas normativas:
 - ♦ Descritores (*Descriptors, D*)
 - ♦ Esquemas de descrição (*Description Schemes, DS*)
 - ♦ Linguagem de definição de descrições (*Description Definition Language, DDL*)
 - ♦ Esquemas de codificação para representação compacta das descrições (*Coding Schemes, CS*)
 - ♦ Ferramentas ao nível de sistema, nomeadamente para multiplexagem e sincronização das descrições e do conteúdo descrito
- Ferramentas não-normativas:
 - ♦ Métodos de extracção de características

- ♦ Métodos de procura (medidas de semelhança)
- Metodologias de trabalho
- ♦ Técnicas de avaliação e validação dos descritores propostos ao MPEG-7

As propostas recebidas foram avaliadas, em Fevereiro de 1999, numa reunião de avaliação realizada em Lancaster – Reino Unido, englobando várias dezenas de especialistas. Esta avaliação pretendia fundamentalmente avaliar a extensão e o mérito tecnológico de cada proposta através de sete critérios: aplicabilidade, eficiência, domínios de aplicação, eficiência de expressão e processamento, escalabilidade e representação multi-escalar de modo a filtrar as propostas mais fracas.

Para os descritores, a categorização das propostas foi efectuada, segundo a metodologia de avaliação previamente definida ou seja baseando-se nos seguintes passos [MPEG7-Eva1998]:

- **Avaliação da importância da característica descrita** – compreensão da característica descrita, verificando o modo como se obtém a informação de descrição relativamente ao conteúdo audiovisual;
- **Avaliação da informação escrita submetida** – primeiro contacto com a tecnologia proposta através da informação escrita associada à proposta;
- **Apresentações de demonstrações** – demonstração das funcionalidades da ferramenta proposta e indicados os seus domínios de aplicação;
- **Procuras baseadas na semelhança** – esta fase efectuou-se sempre que o parâmetro de descrição proposto permitisse a procura baseada na semelhança, de modo a avaliar a sua eficácia (precisão do descritor na procura, comparativamente com os outros) e eficiência de expressão (capacidade do descritor em expressar precisa e completamente a característica em questão); e
- **Avaliação global conclusiva** – elaboração de um relatório final com base nos resultados das etapas anteriores, a fim de se seleccionarem as tecnologias para inclusão na primeira versão do XM, XM 1.0, ou para definição de *core experiments* comparando ferramentas oferecendo a mesma funcionalidade.

A actualização e optimização do XM foi feita, iterativamente, através da realização de *core experiments*, que contribuíram com novas ferramentas para a norma. No entanto, e de modo a facilitar o processo de actualização e optimização, foram utilizadas duas ferramentas de trabalho já mencionadas, XM e CEs, cujas funções foram de extrema importância durante a fase de colaboração e que são:

- **eXperimentation Model (XM)** – Modelo de trabalho para a descrição e procura de conteúdo audiovisual, construído com a tecnologia aceite após avaliação minuciosa, normalmente através de *core experiments*; o XM permite verificar, comparar e melhorar o desempenho das várias ferramentas integradas no modelo ou das ferramentas que tentam nele entrar ‘competindo’ com as que já lá estão [Pere1998]. O XM inclui ferramentas normativas (descritores, esquemas de descrição, métodos de codificação e a linguagem de definição de descrições) e não-normativas (métodos de extracção de características e medidas de

semelhança), o que permite avaliar e comparar continuamente as ferramentas integradas neste modelo de trabalho. Caso uma ferramenta seja seleccionada para integrar o XM, o código fonte correspondente deverá ser fornecido para integração no software deste modelo de trabalho, de acordo com o exemplo fornecido pelo MPEG [MPEG7-SWINT2000];

- **Core Experiment (CE)** – Processo através da qual se seleccionam as ferramentas a integrar no XM; o primeiro conjunto de *CEs* foi definido com base nos resultados da avaliação das propostas. Muitas vezes, após a selecção das ferramentas a integrar no XM, os *CE* passam a funcionar como um processo de optimização das mesmas. Os *CEs* permitem a realização de múltiplas experiências, independentes e directamente comparáveis, de modo a concluir qual das várias ferramentas em competição num dado *core experiment* apresenta o melhor desempenho. As ferramentas em análise poderão substituir uma ferramenta já incluída no XM ou ser integradas directamente no XM se oferecerem novas funcionalidades, ainda não oferecidas por nenhuma ferramenta já incluída no XM. Consequentemente, os *CEs* têm que ser definidos de forma precisa, descrevendo as ferramentas e as condições em que estas serão comparadas, para que os resultados sejam inequívocos [MPEG7-Dev1999] [Pere1998].

No contexto da fase de avaliação das propostas, os descritores de forma, mais relevantes no contexto desta tese, foram divididos em dois grupos de acordo com as suas funcionalidades: descritores de forma baseados no contorno e descritores de forma baseados na região². Ao primeiro tipo de descritor compete a descrição de formas simples, conexas, através do uso dos seus *shapels* mais externos (contorno), enquanto que ao segundo tipo de descritor compete a descrição de formas complexas, eventualmente incluindo várias regiões não conexas, através de todos os seus *shapels*. A Figura 5.1 ilustra o que cada um destes dois tipos de descritores pretende descrever e os resultados que cada um deles permite obter.

² Ao longo deste capítulo, e sempre que isso não criar problemas, usa-se muitas vezes o termo ‘descritor’ para designar seja um parâmetro de descrição seja um descritor, segundo as definições apresentadas na secção 1.2. Na verdade, um descritor e um parâmetro de descrição diferem no facto de o primeiro ao contrário do segundo incluir uma sintaxe e semântica precisas. Quando for importante distinguir os dois termos, usar-se-ão os termos ‘parâmetro de descrição’ e ‘descritor’.

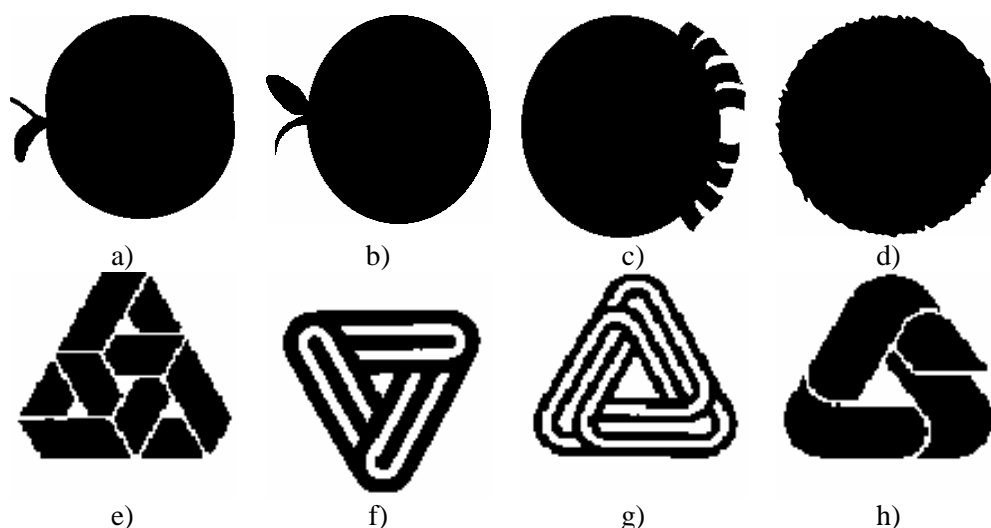


Figura 5.1 – Descritor baseado no contorno: a) imagem binária que define a forma que se pretende encontrar, b), c) e d) resultados da procura; Descritor baseado na região – e) imagem binária que define a forma que se pretende encontrar, f), g) e h) resultados da procura (pouco semelhantes em termos de contorno)

A Tabela 5.1 apresenta, resumidamente, todas as propostas, relacionadas com a característica de forma, apresentadas como resposta à chamada de proposta para a norma MPEG-7, para cada um dos elementos chamados, indicando para os mais relevantes o número da proposta respectiva:

<i>Elementos Normativos (Descritores)</i>		<i>Elementos Não-Normativos (Medidas de Semelhança)</i>	<i>Metodologias de Trabalho</i>
Baseados no Contorno	<ul style="list-style-type: none"> ▪ <i>Curvature Scale Space (CSS)</i> – P320 ▪ <i>Wavelet</i>- P567 ▪ Ângulos de contorno – P162 ▪ Contorno normalizado – P579 ▪ Momento geométrico– P010 ▪ Corda (ângulo) – P007 ▪ Corda (ângulo) – P008 ▪ Corda (raio) – P009 	<ul style="list-style-type: none"> ▪ Distância Euclidiana ▪ Distância <i>Minkowsky</i> 	<ul style="list-style-type: none"> ▪ Avaliação da insensibilidade a: <ul style="list-style-type: none"> – Mudanças de escala e rotações – Deformações geométricas – Robustez a movimentos não-rígidos
Baseados na Região	<ul style="list-style-type: none"> ▪ <i>Angular Radial Transform (ART)</i> ▪ Momentos <i>Zernike</i> – P687 ▪ Vectores próprios multi-nível (<i>MLEV</i>) – P517 ▪ <i>Hashing</i> – P144 ▪ <i>Bitmap</i> normalizado – P578 	<ul style="list-style-type: none"> ▪ Distância Euclidiana ▪ Distância <i>Minkowsky</i> 	<ul style="list-style-type: none"> ▪ Avaliação da insensibilidade a: <ul style="list-style-type: none"> – Mudanças de escala e rotações no domínio digital e analógico, e alterações de perspectiva – Deformações geométricas

Tabela 5.1 – Elementos normativos, não-normativos e metodologias de trabalho para a característica de forma apresentados no seguimento da chamada de propostas do MPEG-7 [VeHa1999]

Após a avaliação, os relatórios elaborados pelos especialistas para cada uma das propostas apresentaram resultados com base numa pontuação de 0-5 [VeHa1999], com o seguinte significado:

- 0** – não foi possível fazer a avaliação, p.e. por não haver informação suficiente
- 1** – a proposta não é de interesse para o MPEG-7
- 2** – a proposta é de interesse para o MPEG-7 mas necessita de algumas mudanças a fim de melhorar a proposta
- 3** – a proposta é de interesse para o MPEG-7 tendo que sofrer apenas pequenas alterações
- 4** – a proposta apresenta propriedades bastante vantajosas e é uma boa candidata dentro da sua categoria para o XM
- 5** – a proposta é de nível superior dentro da sua categoria e é fortemente recomendada para integração no XM

A classificação atribuída a cada uma das propostas com descritores de forma, baseados no contorno ou região, é apresentada nas Tabela 5.2 e Tabela 5.3, respectivamente:

Número proposta	Nome do descritor proposto	Proponente	Avaliação
P144	Estrutura afim do objecto	IBM, E.U.A.	4
P145	<i>Location Hashing</i>	IBM, E.U.A.	4
P010	Momento	National Research Council, Canadá	3
P162	Ângulos de curvatura do contorno	IBM, E.U.A.	3
P567	<i>Wavelet</i>	HHL, Alemanha	3
P007	Corda (ângulo)	National Research Council, Canadá	2.5
P008	Corda (ângulo)	National Research Council, Canadá	2.5
P009	Corda (raio)	National Research Council, Canadá	2.5
P320	<i>Curvature Scale Space (CSS)</i>	Mitsubishi Electric, Reino Unido	2.5
P568	Momentos Invariantes	HHL, Alemanha	2.5
P105	Circularidade, momento central e curvatura	RICOH, Japão	2
P298	Espaço tangencial	Siemens, Alemanha	2
P579	Contorno normalizado	Tektronix, Inc., E.U.A.	2
P177	Momentos esféricos	IBM, E.U.A.	1.5
P178	Pontos característicos	IBM, E.U.A.	1.5

Tabela 5.2 – Avaliação das propostas a descritores de forma baseados no contorno [VeHa1999]

Número proposta	Nome do descritor proposto	Proponente	Avaliação
P517	Vectores próprios multi-nível (MLEV)	Hyundai Electronics Ind., Coreia	4
P687	Momentos de <i>Zernike</i>	Hyundai Electronics Ind., Coreia	4
P578	Bitmap normalizado	Tektronix, Inc., E.U.A.	3

Tabela 5.3 – Avaliação das propostas a descritores de forma baseados na região [VeHa1999]

Com base nestas classificações, o grupo de especialistas recomendou a todos os proponentes dos descritores baseados no contorno e aos proponentes das propostas P687 e P517 com descritores baseados na região que realizassem *core experiments* com um conjunto de

imagens maior e mais genérico a fim de se seleccionar e otimizar a tecnologia de descrição de forma a integrar no XM do MPEG-7.

Uma vez que foi com base nos resultados dos *core experiments* que se tomaram as grandes decisões referentes à tecnologia de descrição de forma a incluir na norma MPEG-7, as secções seguintes analisarão de forma detalhada esta fase do processo de normalização no que diz respeito aos descritores de forma.

5.2 Core Experiments: Metodologias de Avaliação para a Característica de Forma

Os parâmetros de forma e as correspondentes medidas de semelhança disponíveis na literatura jamais foram comparados com testes definidos de forma precisa e em condições comuns, a fim de se analisar e avaliar o seu desempenho de forma credível.

Face às propostas recebidas em termos da característica de forma, o MPEG-7 decidiu que os descritores de forma fossem divididos em dois tipos, de acordo com as suas funcionalidades: descritores baseados no contorno e descritores baseados na região, seguindo as definições já apresentadas no capítulo 1. Esta divisão dos descritores em dois grupos levou à definição de cinco *CEs* que permitiram analisar, avaliar, comparar e otimizar cada uma das propostas, segundo as suas funcionalidades. Os *core experiments* definidos foram:

- **Core Experiment 1 (CE-1)** – avalia o desempenho do descritor, relativamente à descrição e procura de formas simples (conexas), na presença de transformações geométricas tais como mudanças de escala, rotações, deformações, simetria e movimentos não-rígidos [MPEG7-DCE1999a];
- **Core Experiment 2 (CE-2)** – avalia o desempenho do descritor, relativamente à descrição e procura de formas complexas (incluindo não conexas), na presença de transformações geométricas, no domínio digital e analógico, tais como mudanças de escala, rotações e alterações de perspectiva [MPEG7-DCE1999a];
- **Core Experiment CE-1/2 Otimizado** – otimiza as tecnologias, em termos de método de extracção e/ou método de procura, até à data integradas no XM;
- **Core Experiment MotionShape-3 (MS-3 CE)** – otimiza a dimensão dos descritores de forma baseados no contorno e na região, adoptados pelo MPEG-7 e integrados no XM, analisando o efeito da sua quantificação (número de bits) em função do seu desempenho de procura [MPEG7-DCE1999b] [MPEG7-DCE2000a] (há que salientar que o movimento aparece no nome desta *core experiment* porque as *core experiments* para a característica de forma e movimento foram definidas no mesmo documento);
- **Core Experiment MotionShape-5 (MS-5 CE)** – otimiza os parâmetros globais e as transformações efectuadas aos picos da imagem CSS do descritor de forma baseado no contorno, adoptado pelo MPEG-7 e integrado no XM, nas condições do CE-1 [MPEG7-DCE2000b].

Estes *CE* avaliaram, essencialmente, o desempenho de cada um dos descritores propostos ao MPEG-7 de acordo com três categorias de medida:

- **Emparelhamento exacto** – a imagem dada como exemplo para a procura diferirá das resultantes da procura apenas por transformações de mudança de escala e/ou rotação. A Figura 5.2 mostra dois casos de emparelhamento exacto;

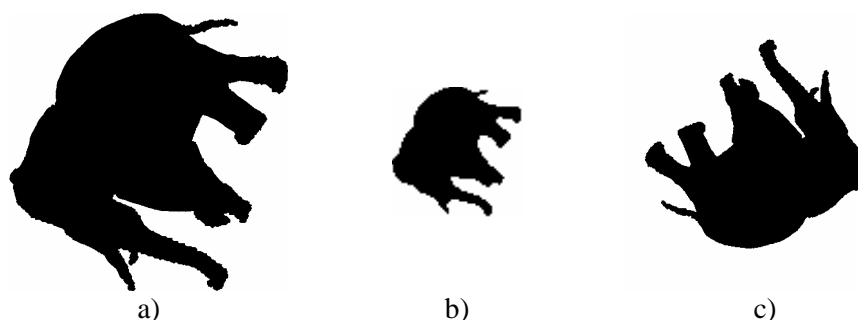


Figura 5.2 – Exemplo de um emparelhamento exacto: a) *Elephant.tif*: imagem dada como exemplo para a procura, b) e c) imagens resultantes da procura: imagem reduzida (0.33) e imagem rodada (150°), respectivamente

- **Emparelhamento baseado na semelhança perceptiva** – a imagem dada como exemplo para a procura assemelhar-se-á às resultantes da procura, por exemplo através de deformações do objecto exemplo ou qualquer outra semelhança perceptiva. A Figura 5.3 mostra dois casos de emparelhamento baseado na semelhança perceptiva;

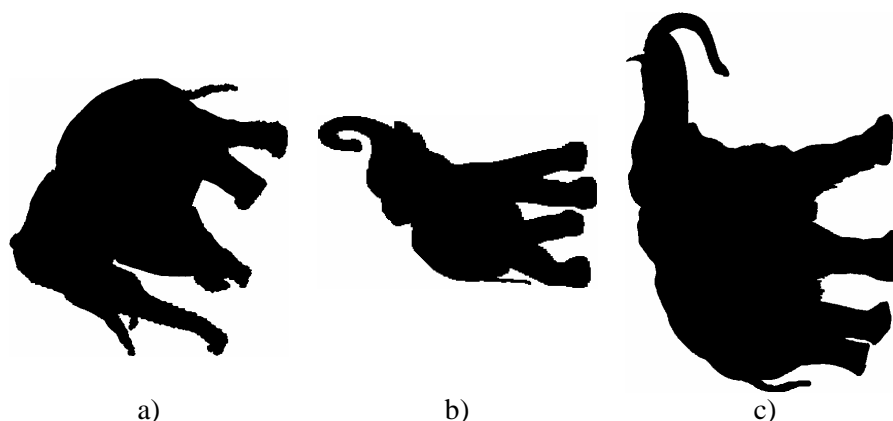


Figura 5.3 – Exemplo de um emparelhamento baseado na semelhança perceptiva: a) *Elephant.tif*: imagem dada como exemplo para a procura, b) e c) imagens resultantes da procura

- **Robustez a movimentos não rígidos** – a imagem dada como exemplo assemelhar-se-á às resultantes da procura apesar destas terem sofrido alterações de perspectiva ou outras deformações. A Figura 5.4 mostra algumas tramas da sequência de teste do MPEG-7, *Bream*, a fim de demonstrar o tipo de robustez pretendida com este teste.

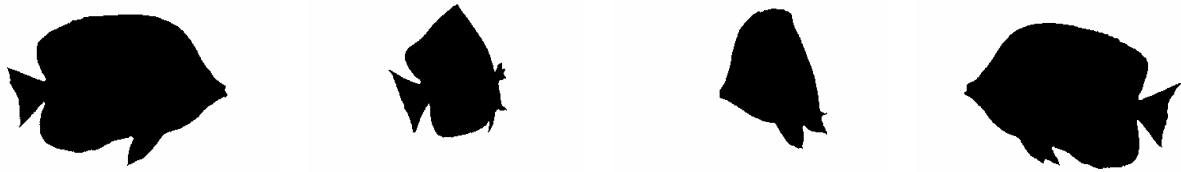


Figura 5.4 – Exemplo de uma sequência de movimentos não-rígidos da sequência de teste do MPEG-7, *Bream*

Estes *CEs* e categorias de medida serão usados pelos descritores de forma, para objectos 2D, propostos ao MPEG-7. Definem-se de seguida, de modo mais pormenorizado, os vários *core experiments*.

5.2.1 Core Experiment-1 (*CE-1*)

O *CE-1* avalia o desempenho das ferramentas em questão - descritores baseados no contorno - relativamente à descrição e procura de formas simples (conexas), usando as três categorias de medida acima definidas:

1. Emparelhamento exacto na presença de transformações geométricas como mudanças de escala e rotação;
2. Emparelhamento baseado na semelhança perceptiva com base em imagens semelhantes por deformações e/ou transformações geométricas; e
3. Robustez a movimentos não-rígidos.

Cada uma destas categorias do *CE-1* é testada individualmente a fim de se avaliar o comportamento individual de cada proposta perante estas condições. Deste modo, foi decidido pelo MPEG que o *CE-1* fosse dividido em três partes, A, B e C, sendo cada uma delas responsável pela avaliação dos descritores em relação às três categorias acima referidas, respectivamente.

Descreve-se, de seguida, cada uma das partes do *CE-1* para que se conheçam as condições de teste aplicadas aos descritores de forma em avaliação, bem como as propriedades que distinguem as imagens binárias (base de dados), usadas para cada uma das partes do *CE-1*. As bases de dados de conteúdo usadas para o *core experiment CE-1* são duas:

- 1) Base de dados com 1400 imagens binárias (correspondentes aos suporte binário da forma) agrupadas segundo 70 classes de 20 imagens binárias semelhantes cada;
- 2) Base de dados com 200 tramas binárias do objecto *Bream* e ainda 1000 imagens binárias de criaturas marinhas.

5.2.1.1 CE-1: Parte A

O objectivo da parte A do CE-1, subdivido em duas partes, A1 e A2, é o de testar o desempenho de cada um dos parâmetros de descrição de forma em termos da sua robustez a mudanças de escala e a rotações no domínio digital, através de procuras por semelhança exacta.

As bases de dados de conteúdo (imagens binárias de teste) para cada uma das duas sub-partes da parte A, foram extraídas da base de dados de 1400 imagens binárias existente para o CE-1. A base de dados de conteúdo para a parte A1 foi obtida extraíndo-se 1 imagem binária de referência das 20 imagens pertencentes a cada uma das 70 classes, à qual se aplicaram os factores de mudança de escala $\times 2$, $\times 0.3$, $\times 0.25$, $\times 0.2$ e $\times 0.1$ ou seja, por exemplo, $\times 0.25$ significa que a imagem fica com um quarto do seu tamanho original. Este processo resulta numa base de dados com 420 (70×6) imagens binárias composta por 70 classes com 6 imagens binárias cada. A Figura 5.5 mostra o tipo de transformação geométrica sofrida pela imagem binária de referência após a aplicação dos factores de mudança de escala.

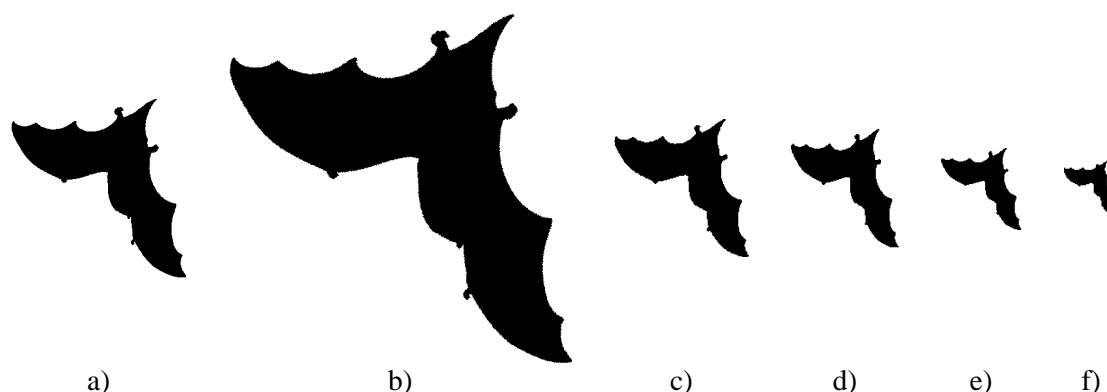


Figura 5.5 – Classe *Bat* - a) imagem binária de referência; imagens binárias resultantes após aplicação dos factores de mudança de escala b) $\times 2$, c) $\times 0.3$, d) $\times 0.25$, e) $\times 0.2$ e f) $\times 0.1$

À semelhança da base de dados para a parte A1, a base de dados de conteúdo para a parte A2 foi conseguida do mesmo modo com a variante de se terem aplicado rotações com os ângulos de rotação 9° , 36° , 45° , 90° e 150° de modo a criar uma base de dados com 420 (70×6) imagens binárias composta por 70 classes com 6 imagens binárias cada. A Figura 5.6 mostra o tipo de transformação geométrica sofrida pela imagem binária de referência após a aplicação dos ângulos de rotação.

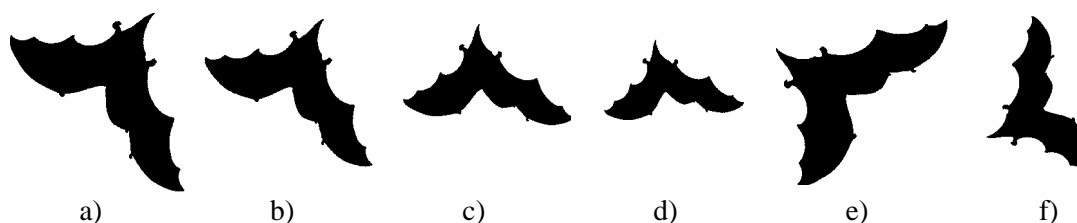


Figura 5.6 – Classe *Bat* - a) imagem binária de referência; imagens binárias resultantes após aplicação dos ângulos de rotação b) 9° , c) 36° , d) 45° , e) 90° e f) 150°

Para cada uma das bases de dados, para as partes A1 e A2, com 420 imagens binárias, foram efectuadas 70 procuras com a imagem binária de referência de cada uma das 70 classes criadas a fim de se verificar o desempenho do descritor na presença de transformações de mudança de escala e de rotação. O desempenho do descritor para cada uma das partes será calculado de acordo com o número de resultados correctos entre as 6 primeiras formas resultantes da procura. Definem-se como resultados correctos aqueles onde a imagem binária encontrada é uma versão escalada ou rodada da imagem binária de referência usada no processo de procura. Após efectuadas as procuras, o desempenho do parâmetro de descrição para as partes A1 e A2 é medido em termos percentuais através da expressão (5.1):

$$Desempenho (A1/ A2) = \frac{100}{420} \sum_{m=1}^{70} \sum_{n=1}^6 f(x_n) \quad (5.1)$$

$$\text{onde } f(x_n) = \begin{cases} 1 & \text{resultado correcto} \\ 0 & \text{c.c.} \end{cases}$$

Finalmente, o desempenho médio do descritor para a parte A do CE-1 é calculado através da expressão (5.2):

$$Desempenho (A) = \frac{Desempenho (A1) + Desempenho (A2)}{2} \quad (5.2)$$

5.2.1.2 CE-1: Parte B

O objectivo da parte B da metodologia de avaliação para o CE-1 é o de testar o desempenho de cada um dos parâmetros de forma em termos de procuras baseadas em semelhanças perceptivas, ou seja avaliar a capacidade do descritor em encontrar formas semelhantes à dada como referência ao processo de procura. Neste caso consideram-se formas semelhantes, as formas de algum modo parecidas com a forma fornecida pelo mecanismo de procura e não formas exactamente iguais após terem sofrido transformações geométricas como na parte A.

A base de dados definida para esta parte do CE-1 é exactamente a mesma base de dados de 1400 imagens usada na parte A do CE-1. Para a parte B do CE-1 serão efectuadas 1400 procuras usando cada uma das 20 imagens semelhantes pertencentes a cada uma das 70 classes. A Figura 5.7 mostra o tipo de imagens binárias existentes (e semelhantes) em cada uma das 70 classes.



Figura 5.7 – Classe *Bat* – 6 das 20 imagens binárias pertencentes a uma das 70 classes de imagens binárias disponíveis

Tendo em conta que nesta parte do *CE-1* se pretende medir o desempenho baseado na semelhança perceptiva, usou-se o *Bull's Eye Performance (BEP)* como medida de desempenho [MPEG7-DCE1999a]. O *BEP* define-se como a percentagem de resultados correctos entre os $2N$ primeiros resultados encontrados e correspondentes à mesma classe dentro da base de dados, sendo N a dimensão da classe correspondente; neste caso, N vale 20 uma vez que cada classe tem 20 imagens binárias (ver expressão (5.3)).

$$BEP = \frac{100}{N} \sum_{n=1}^{2N} f(x_n) \quad (5.3)$$

$$\text{onde } f(x_n) = \begin{cases} 1 & \text{resultado correcto} \\ 0 & \text{c.c.} \end{cases}$$

Após a execução das 1400 procuras, é calculada uma média das percentagens de *BEP* para as 1400 procuras executadas como mostra a expressão (5.4):

$$Desempenho(B) = \frac{1}{1400} \sum_{m=1}^{1400} BEP_m \quad (5.4)$$

5.2.1.3 CE-1: Parte C

A parte C do *CE-1* tem como objectivo testar o desempenho de cada um dos parâmetros de forma em termos da sua robustez a movimentos não-rígidos, como mudanças de perspectiva ou deformações e transformações mais difíceis de definir geometricamente, por exemplo resultantes do movimento natural do objecto não-rígido e consequentemente da sua forma.

A base de dados de conteúdo usada para a parte C do *CE-1* é o conjunto das 200 tramas binárias consecutivas do objecto de vídeo *Bream* e as 1100 imagens binárias de criaturas marinhas. A Figura 5.8 mostra algumas das imagens binárias pertencentes a esta base de dados. As imagens binárias a), b) c) e d) mostram ainda uma sequência de movimentos não-rígidos.

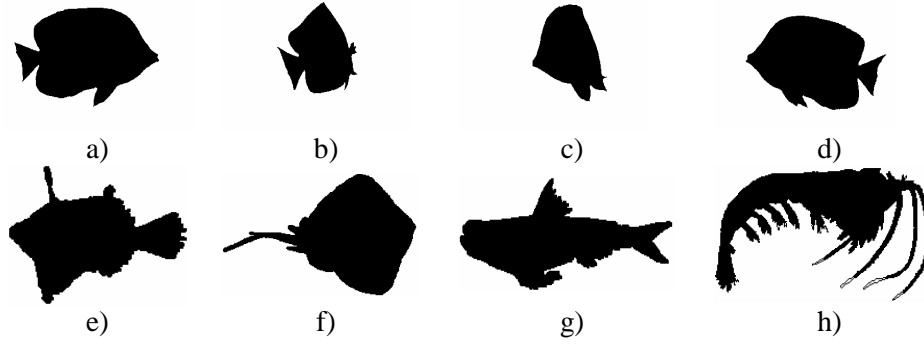


Figura 5.8 – Exemplos de imagens binárias pertencentes a), b), c) e d) à sequência *Bream*, e e), f), g) e h) à base de dados de criaturas marinhas

A parte C do *CE-1* efectua apenas uma procura usando a trama 0 da sequência *Bream* como imagem binária de referência no processo de procura. O desempenho do descritor (ver expressão (5.5)) para as condições da parte C é medido contando-se os resultados correctos entre os 200 primeiros resultados obtidos (entendendo-se como resultados correctos as 200 tramas da sequência *Bream*), os quais serão divididos por 200 e multiplicados por 100 a fim de gerar a medida de desempenho em termos percentuais:

$$Desempenho(C) = \frac{100}{200} \sum_{n=1}^{200} f(x_n) \quad (5.5)$$

$$\text{onde } f(x_n) = \begin{cases} 1 & \text{resultado correcto} \\ 0 & \text{c.c.} \end{cases}.$$

Às restantes 1100 imagens binárias pertencentes à base de dados definida e que não foram agrupadas na única classe que contém as 200 tramas da sequência *Bream*, chamam-se imagens binárias desagrupadas.

Finalmente, o desempenho total médio para o *core experiment CE-1* é calculado como a média dos desempenhos obtidos para cada uma das partes A, B e C [MPEG7-DCE1999a], ou seja:

$$Desempenho \text{ Total Médio} = \frac{Desempenho(A) + Desempenho(B) + Desempenho(C)}{3}$$

A Tabela 5.4 mostra, de um modo resumido, o conjunto de imagens binárias usadas em cada uma das partes do *core experiment CE-1* e já explicados nas sub-secções anteriores.

<i>Core Experiment</i>			# de classes	# de imagens em cada classe	# de resultados correctos	# total de imagens em classes	# de imagens de referência	# de imagens desagrupadas	# total de imagens na base de dados
CE-1	A	A1	70	6	6	420	70	0	420
		A2	70	6	6	420	70	0	420
	B		70	20	40	1400	1400	0	1400
	C		1	200	200	200	1	1100	1300

Tabela 5.4 – Imagens binárias para as várias partes do *core experiment CE-1*

5.2.2 Core Experiment-2 (CE-2)

O *CE-2* avalia o desempenho das ferramentas em questão - descritores baseados em região - relativamente à descrição e procura de formas complexas (incluindo formas não conexas), usando duas das categorias de medida acima definidas:

1. Emparelhamento exacto na presença de transformações geométricas como mudanças de escala e rotação; e
2. Emparelhamento baseado na semelhança perceptiva com base em imagens semelhantes por deformações e/ou transformações geométricas.

Cada uma destas categorias do *CE-2* é testada individualmente a fim de se avaliar o comportamento individual de cada ferramenta de descrição perante estas condições. Deste modo, foi decidido pelo MPEG que o *CE-2* fosse dividido em duas partes, A e B, sendo cada uma delas responsável pela avaliação dos descritores em termos das duas categorias acima referidas, respectivamente.

Descrevem-se, de seguida, cada uma das duas partes do *CE-2* para que se conheçam as condições de teste aplicadas aos descritores de forma avaliados, bem como as propriedades que distinguem as imagens binárias (base de dados), usadas para cada uma das partes do *CE-2*. A base de dados de conteúdo usada para o *core experiment CE-2* é a base de dados de logótipos, *S8*, que faz parte do material de teste do MPEG-7 [MPEG7-DCE1999a]; esta base de dados foi disponibilizada pelo MPEG aos participantes neste *core experiment*.

O desempenho global de cada descritor para o *CE-2* é calculado como a média dos desempenhos para as duas partes: A e B.

5.2.2.1 CE-2: Parte A

O objectivo da parte A do *CE-2*, subdivida em quatro partes, A1, A2, A3 e A4, é o de testar o desempenho de cada um dos descritores de forma em termos da sua robustez a mudanças de escala e a rotações no domínio digital e analógico, através de procuras por semelhança

exacta. O teste à robustez a mudanças de escala e a rotações no domínio analógico é efectuada sobre imagens binárias adquiridas através de uma máquina de filmar, de um *CCD* e de uma câmara digital.

As bases de dados usadas para cada uma das quatro partes da parte A do *CE-2* foram extraídas da base de dados de logótipos (*S8*) pertencente à base de dados de teste do MPEG-7; estes logótipos correspondem a imagens binárias de 2800 objectos complexos.

A base de dados de conteúdo definida para a parte A1 foi obtida escolhendo-se aleatoriamente 20 imagens de referência de *S8*, às quais se aplicaram os seguintes factores de mudança de escala: $\times 3$, $\times 1.5$, $\times 0.5$ e $\times 0.33$. Este processo resultou numa base de dados de 2881 imagens binárias, ou seja 20 classes definidas pelas imagens de referência escolhidas aleatoriamente com 5 imagens binárias cada e ainda as restantes 2781 imagens binárias de *S8* desagrupadas. A Figura 5.9 mostra o tipo de transformação geométrica sofrida pela imagem binária de referência após a aplicação dos factores de mudança de escala.

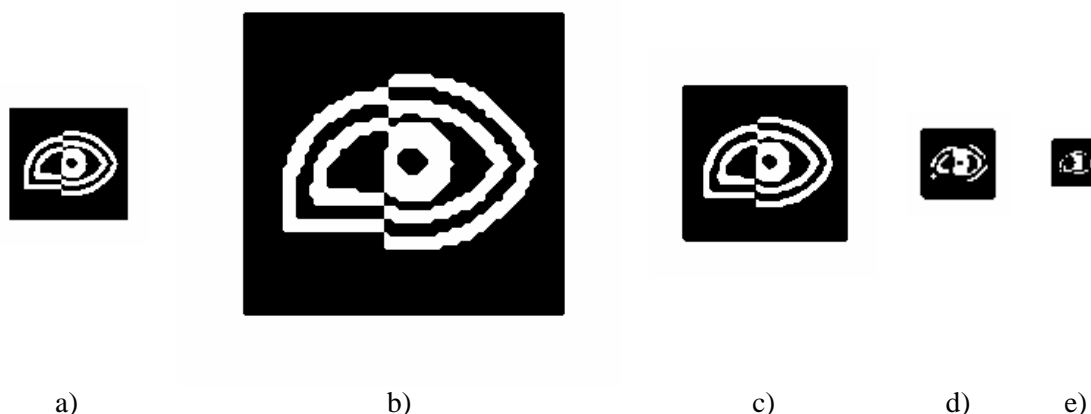


Figura 5.9 – Classe 368 - a) imagem binária de referência; imagens binárias resultantes após aplicação dos factores de mudança de escala b) $\times 3$, c) $\times 1.5$, d) $\times 0.5$, e) $\times 0.33$

À semelhança da base de dados para a parte A1, a base de dados de conteúdo para a parte A2 foi conseguida do mesmo modo com a variante de se terem aplicado os ângulos de rotação 9° , 36° , 45° , 90° , 120° e 150° de modo a criar uma base de dados de 2921 imagens binárias composta por 20 classes com 7 imagens binárias cada e ainda das restantes 2781 imagens binárias de *S8* desagrupadas. A Figura 5.10 mostra o tipo de transformação geométrica sofrida pela imagem binária de referência após a aplicação dos ângulos de rotação.

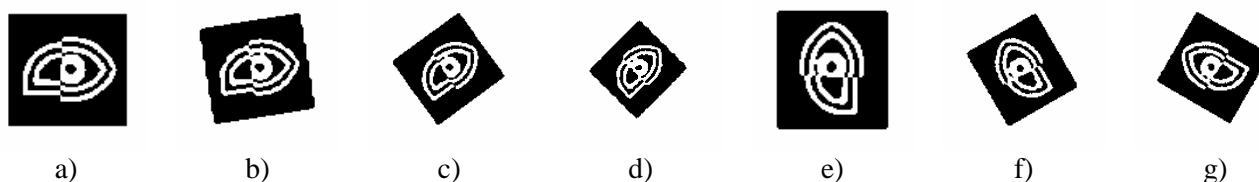


Figura 5.10 – Classe 368 - a) imagem binária de referência; imagens binárias resultantes após aplicação dos ângulos de rotação b) 9° , c) 36° , d) 45° , e) 90° , f) 120° e g) 150°

Quanto às partes A3 e A4, as suas bases de dados foram também geradas a partir de *S8*, mas no domínio analógico, começando por se escolher 10 logótipos de referência, aleatoriamente.

A partir destes 10 logótipos foram geradas três versões em vídeo, através de uma máquina de filmar, de um CCD e de uma câmara digital, que foram de seguida digitalizadas. Entre as imagens binárias digitalizadas foram escolhidas 10 de cada objecto (100 no total mais as 10 imagens de referência escolhidas aleatoriamente), onde devem constar imagens binárias com mudanças de escala e rotação (para a parte A3) e imagens com transformações de perspectiva (para a parte A4). Este processo resulta em duas bases de dados, para A3 e A4, com 3111 imagens binárias, composta por 30 classes com 11 imagens binárias cada e ainda 2781 imagens binárias desagrupadas. A Figura 5.11 e a Figura 5.12 mostram o tipo de transformações geométricas (mudança de escala, rotação e perspectiva) sofridas pela imagem binária de referência nas partes A3 e A4.

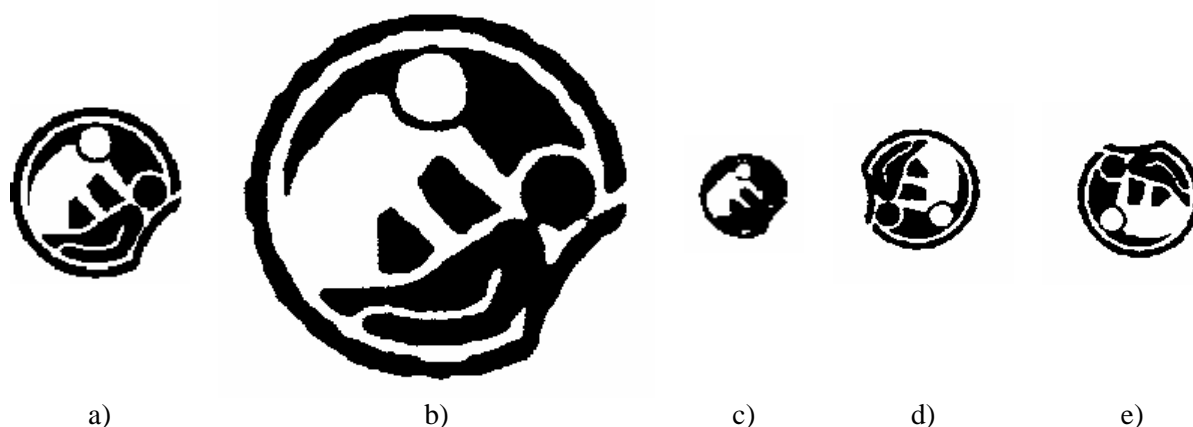


Figura 5.11 – Classe 616 – a) imagem binária de referência; b), c) imagens binárias resultantes após mudança de escala; d) e e) imagens binárias resultantes após rotação no domínio analógico

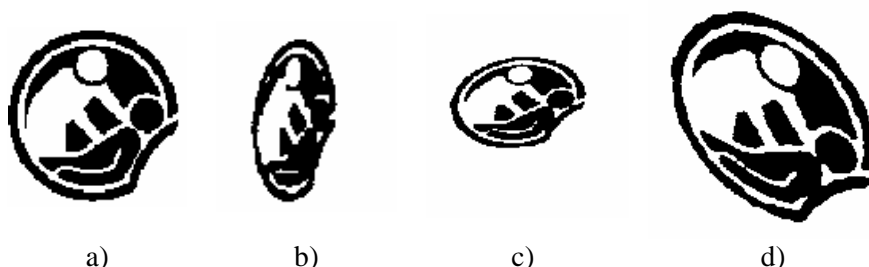


Figura 5.12 – Classe 616 - a) imagem binária de referência; b), c) e d) imagens binárias resultantes após transformações de perspectiva no domínio analógico

A avaliação do desempenho dos descritores de forma para as partes A1 e A2 do CE-2 é idêntica à das partes A1 e A2 do CE-1 (ver expressão (5.6)), variando o número de procuras K efectuadas para 100 e 140, respectivamente, a contagem do número de resultados correctos N entre os 5 ou 7 primeiros, respectivamente, e sendo o número de classes M de 20 para ambas as partes. Quanto às partes A3 e A4, a avaliação do desempenho é idêntica às definidas para as partes A1 e A2 deste *core experiment*, variando o número de procuras K efectuadas para 330, a contagem do número de resultados correctos $2N$ para os 22 primeiros e o número de classes M de 30 para ambas as partes.

$$Desempenho(A1 / A2 / A3 / A4) = \frac{100}{K} \sum_{m=1}^M \sum_{n=1}^N f(x_n) \quad (5.6)$$

$$\text{onde } f(x_n) = \begin{cases} 1 & \text{resultado correcto} \\ 0 & \text{c.c.} \end{cases}$$

Consequentemente, o desempenho médio do descritor para a parte A é calculado através da expressão (5.7):

$$\text{Desempenho (A)} = \frac{\text{Desempenho (A1)} + \text{Desempenho (A2)} + \text{Desempenho (A3)} + \text{Desempenho (A4)}}{4} \quad (5.7)$$

5.2.2.2 CE-2: Parte B

O objectivo da parte B do CE-2 é o de testar o desempenho de cada um dos descritores de forma em termos de procuras baseadas em semelhanças perceptivas, ou seja avaliar a capacidade do descritor em encontrar formas semelhantes à dada como referência ao processo de procura. É de salientar novamente que o que se entende aqui por formas semelhantes é que estas sejam de algum modo parecidas com a forma fornecida ao mecanismo de procura e não formas apenas sujeitas a transformações geométricas simples no domínio analógico ou digital, como na parte A deste *core experiment*.

A base de dados definida para esta parte da metodologia de avaliação consiste em 2800 imagens agrupadas em 10 classes ou grupos, representadas pela letra **G** na Tabela 5.5, com formas semelhantes e de uma 11ª classe com as 2122 imagens binárias desagrupadas da base de dados S8. Posteriormente, todas as imagens pertencentes às 10 primeiras classes, ou seja 678 são usadas para efectuar procuras, sendo o cálculo do desempenho igual ao usado para a parte B do CE-1 (ver expressões (5.3) e (5.4)).

O desempenho médio total será calculado como a média dos desempenhos para cada uma das partes A e B do *Core Experiment CE-2* [MPEG7-DCE1999a].

Uma vez definidas as duas categorias de medida do desempenho para o CE-2, a Tabela 5.5 mostra de um modo resumido o conjunto das imagens binárias usadas em cada uma destas categorias e explicadas nas sub-secções anteriores.

<i>Core Experiments da Forma</i>			# de classes	# de imagens em cada classe	# de resultados correctos	# total de imagens em classes	# de imagens de referência	# de imagens desagrupadas	# total de imagens na base de dados
<i>CE-2</i>	A	A1	20	5	5	100	100	2781	2881
		A2	20	7	7	140	140	2781	2921
		A3	30	11	11	330	330	2781	3111
		A4	30	11	11	330	330	2781	3111
	B	G1	1	68	136	68	68	2122	2800
		G2	1	244	488	244	244		
		G3	1	22	44	22	22		
		G4	1	28	56	28	28		
		G5	1	17	34	17	17		
		G6	1	22	44	22	22		
		G7	1	45	90	45	45		
		G8	1	145	290	145	145		
		G9	1	45	90	45	45		
		G10	1	42	84	42	42		

Tabela 5.5 – Imagens binárias para as várias partes do *core experiment CE-2*

5.2.3 *Core Experiment CE-1/2 Optimizado (CE-1/2)*

O objectivo do *CE-1/2 Optimizado* é o de avaliar as optimizações efectuadas nos métodos de extracção e de procura dos descritores de forma baseados no contorno e na região, adoptados pelo MPEG-7 e integrados no *XM*, em função das condições definidas para os *core-experiments CE-1* e *CE-2*. É de salientar que a metodologia de cálculo do desempenho do descritores em nada se altera, ou seja é a mesma aplicada no *CE-1* e no *CE-2*.

5.2.4 *Core Experiment MS-3 (CE MS-3)*

O objectivo do *CE MS-3* é o de optimizar a dimensão dos descritores de forma baseados no contorno e na região avaliando o desempenho do descritor de forma em função da quantificação dos seus vários componentes. Quando este *CE* estava em desenvolvimento surgiu uma nova proposta para descritor de forma baseado na região, o *Angular-Radial Transform (ART)*, considerada muito interessante. Por este motivo, este *CE* foi aplicado tanto aos descritores já incluídos no *XM* como o *Curvature Scale Space (CSS)* e os momentos de *Zernike*, como à nova proposta *ART* que visava substituir no *XM* os momentos de *Zernike* (o que veio a acontecer) [MPEG7-DCE2000a][MPEG7-DCE2000b].

5.2.5 Core Experiment MS-5 (CE MS-5)

O *CE MS-5* visava melhorar o desempenho do descritor baseado no contorno, *CSS*, alterando o método de extracção e de procura. Neste *CE* foram avaliados e optimizados os valores dos parâmetros globais e as transformações efectuadas aos picos do descritor para que estes descrevam a forma de um objecto de modo optimizado. Novamente, o *CE MS-5* foi efectuado usando as condições do *CE-1* [MPEG7-DCE2000b].

5.3 Selecção dos Descritores de Forma Propostos ao MPEG-7

Na secção 5.1 apresentaram-se as propostas a descritores de forma enviadas ao MPEG no seguimento da chamada de propostas feita para a norma MPEG-7, em Outubro de 1998. O relatório com as avaliações destas propostas, elaborado pelos próprios membros do MPEG em Fevereiro de 1999 [VeHa1999], foi conhecido na reunião do MPEG, em Março de 1999, em Seúl. Com base nestas avaliações, o grupo de especialistas avaliadores decidiu que todos os proponentes de descritores baseados no contorno e que os proponentes das propostas P687 e P517 com descritores baseados na região realizassem *core experiments* com um conjunto de formas maior e mais genérico a fim de se seleccionar e optimizar a tecnologia a integrar no *XM*.

Apesar da recomendação, nem todas as propostas participaram nos *core experiments*, das seis propostas apresentadas inicialmente aos *core experiments*, para os dois tipos de descritores de forma, apenas quatro foram avaliadas no contexto dos *CE* definidos (ver Tabela 5.6), por desistência dos ângulos de curvatura do contorno e do contorno normalizado. Das quatro propostas restantes, duas foram classificadas como descritores de forma baseados no contorno (*CSS* e transformada de *wavelet*) e as restantes como descritores de forma baseados na região (momentos de *Zernike* e *vectores próprios multi-nível*) [MPEG7-DCE1999a].

Descritores de Forma Propostos ao MPEG-7	Proponentes
<i>Curvature Scale Space (CSS)</i>	<i>Mitsubishi ITE-VIL (Reino Unido)</i>
<i>Wavelet</i>	<i>HHI (Alemanha)</i>
Momentos de <i>Zernike</i>	<i>ETRI (Coreia)</i>
Vectores Próprios Multi-Nível (<i>Multi Layer Eigen Vectors – MLEV</i>)	<i>Hyundai Electronics Ind. (Coreia)</i>

Tabela 5.6 – Descritores de forma que participaram nos *core experiments* MPEG-7 [MPEG7-DCE1999a]

Note-se que enquanto os descritores baseados no contorno não podem ser avaliados nas condições do *CE-2* por este usar como material de teste formas complexas (incluindo formas não conexas) que estes descritores não estão preparados para descrever, os descritores baseados na região podem ser avaliados quer nas condições do *CE-1* (formas simples), quer nas condições do *CE-2* (formas complexas).

As sub-secções que se seguem apresentam as quatro fases que até agora ocorreram no processo de desenvolvimento dos descritores de forma para a norma MPEG-7. O desempenho dos descritores será apresentado com base nos resultados obtidos a partir de documentos MPEG disponíveis e com base nos resultados gerados pela autora desta tese a partir do *software* instalado e desenvolvido.

As quatro fases do processo de desenvolvimento/especificação dos descritores de forma incluíram vários *core experiments* e distinguem-se do seguinte modo:

- **Primeira fase** – Foram avaliados os quatro descritores submetidos aos *core experiments*, usando os parâmetros de definição e as medidas de semelhança tal como foram definidos no capítulo 2, a fim de se escolherem dois descritores, um baseado no contorno e outro baseado na região, para integrar no XM. A avaliação destes descritores foi efectuada usando as condições de avaliação definidas para os *core experiments CE-1* e *CE-2*;
- **Segunda fase** – Os descritores de forma com melhor desempenho na primeira fase, e já integrados no XM, melhoraram os seus métodos de extracção e de procura a fim de optimizarem o seu desempenho global, de acordo com o *core experiment CE-1/2*;
- **Terceira fase** – Nesta fase, a dimensão das descrições correspondentes a cada um dos descritores (em *bits*) foi optimizada, ajustando-se os níveis de quantificação de componentes dos descritores em função do desempenho dos descritores de forma, de acordo com o *core experiment MS-3 CE*;
- **Quarta fase** – O desempenho do descritor de forma baseado no contorno foi novamente optimizado através da optimização dos seus parâmetros globais e de extracção conforme definido no *core experiment MS-5 CE*.

5.3.1 Avaliação independente do desempenho dos descritores

A obtenção independente de resultados para esta tese implicou:

1. Instalar o *software* do XM MPEG-7, bem como outro *software* necessário, e corrigir alguns dos erros existentes, atendendo a que o processo de desenvolvimento se encontrava numa fase muito inicial;
2. Obter as bases de dados de conteúdo (ver Tabela 5.7) definidas para os vários *core experiments*;

<i>Core Experiment</i>	Bases de dados de conteúdo
CE-1	<ul style="list-style-type: none"> ▪ 1400 imagens binárias diversas agrupadas segundo 70 classes com 20 imagens binárias semelhantes cada ▪ 200 tramas binárias consecutivas (0-199) do objecto de vídeo <i>Bream</i> ▪ 1000 imagens binárias de criaturas marinhas
CE-2	<ul style="list-style-type: none"> ▪ 2800 imagens binárias de logótipos, S8, cedida pelo grupo MPEG aos participantes

Tabela 5.7 – Bases de dados a usar para a execução dos *core experiments*

3. Comparar os descritores em questão replicando as metodologias precisamente definidas através dos *core experiments*, a fim de se avaliar o seu desempenho de forma credível, com vista à sua posterior integração no XM; e
4. Implementar o *software* necessário para os descritores para os quais não havia *software* disponível.

O pacote de programas e bibliotecas a instalar para que se pudesse analisar, avaliar e verificar, nesta tese, os resultados obtidos pelos proponentes dos vários descritores, inclui:

- ***XM MPEG-7*** – O XM MPEG-7 é um programa que simula o ambiente de um mecanismo de descrição e procura, permitindo verificar, comparar e melhorar o desempenho das ferramentas integradas ou que tentam nele entrar ‘competindo’ com as que já lá estão. O XM inclui ferramentas normativas e não-normativas (p.e. métodos de extracção e medidas de semelhança), o que permite a avaliação e comparação completa das ferramentas nele integradas. Este programa foi desenvolvido de forma colaborativa por todos os participantes nos trabalhos do MPEG-7 uma vez que os proponentes de cada técnica aceite para a norma deviam automaticamente oferecer o respectivo software para o XM. O XM pode ser instalado em três ambientes de trabalho: *Linux*, *Solaris* e *Windows NT/95/98* de modo a facilitar o seu uso e desenvolvimento por todos os participantes na especificação da norma MPEG-7 [MPEG7-SWINT2000]. A obtenção deste programa foi feita através do site na *World Wide Web* (WWW): <http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/mpeg7.html> e as actualizações do mesmo através de um servidor CVS (*Concurrent Versions System*);
- **Descodificador MPEG-2** – O descodificador MPEG-2 é um programa compilado e que é obtido conjuntamente com o XM. O descodificador MPEG-2 é necessário durante a execução do XM para a descodificação de sequências de vídeo codificadas em MPEG-2. Apesar do mecanismo de descrição e procura implementado no âmbito desta tese não permitir ainda a descrição e procura de vídeo, já permite a visualização de conteúdo de vídeo codificado em MPEG-2 (sendo para isso necessário o descodificador MPEG-2), e encontra-se preparado para futuramente incluir o módulo de descrição e procura para informação de vídeo;

- **ImageMagick** – O ImageMagick é uma colecção robusta de ferramentas e bibliotecas para ler, escrever e manipular imagens, como redimensionar e rodar, em vários formatos, como o *TIFF*, *JPEG*, *PNG*, *PDF* e *GIF*. Esta biblioteca encontra-se disponível para vários ambientes de trabalho nomeadamente *Linux*, *Unix*, *Windows 95/98/2000*, *Macintosh*, *VMS* e *OS2*. Para poder ser compilada com o XM, obteve-se a versão 4.2.9 para o *Windows* a partir do site: <ftp://ftp.cdrom.com/pub/ImageMagick/win2k/>. As funcionalidades oferecidas por este software são importantes para obter as variações necessárias das imagens binárias para os vários *core experiments*.

Uma vez obtidos os programas e bibliotecas necessárias para a instalação do XM procedeu-se à sua compilação, de modo a criar o executável *XMWinExe.exe*. Finalmente, testou-se o sucesso da instalação e compilação correndo o programa através da linha de comandos da aplicação. Apesar de as primeiras tentativas terem fracassado pela existência ainda de alguma instabilidade no XM, devido às alterações constantes feitas na sequência das melhorias das ferramentas incluídas no XM, foi finalmente possível pôr tudo a funcionar conforme descrito e suposto.

Tal como descrito no capítulo 4, o XM e *software* associado foram integrados num mecanismo de descrição e procura, desenvolvido no âmbito desta tese, e que permite explorar o XM através de uma interface gráfica flexível e amigável. Os resultados apresentados no resto deste capítulo com vista ao estudo exaustivo dos descritores de forma MPEG-7 foram obtidos usando o mecanismo de descrição e procura desenvolvido.

As sub-seções seguintes apresentam os resultados obtidos para cada um dos descritores propostos, com base nos métodos de extracção e nas medidas de semelhança apresentadas pelos proponentes e já descritas no capítulo 2.

5.3.2 Primeira fase do processo de desenvolvimento

A primeira fase do processo de desenvolvimento dos descritores correspondeu aos *core experiments CE-1* e *CE-2*, ou seja à avaliação dos quatro descritores submetidos aos *core experiments*, usando a especificação e as medidas de semelhança tal como foram apresentadas no capítulo 2, a fim de se escolherem dois descritores, um baseado no contorno e outro baseado na região, para integrar no XM. Como é óbvio, a escolha do descritor de forma baseado na região é também determinada pelo desempenho para o *core experiment CE-2*.

5.3.2.1 Descritores de forma baseados no contorno

No capítulo 2 foram apresentados e definidos alguns dos parâmetros de descrição de forma baseados no contorno mais comuns na área da análise e processamento de imagem, como os parâmetros de *Fourier*, de *wavelet*, os momentos invariantes, a imagem *CSS*, o contorno normalizado e os ângulos de curvatura do contorno. No entanto, e como já foi dito, apenas duas das propostas de descritores de forma baseados no contorno participaram nos *core experiments* e logo apresentaram resultados durante esta primeira fase do processo de desenvolvimento do MPEG-7: a transformada de *wavelet* e a imagem *CSS*.

Para além destas duas propostas, será ainda avaliado no âmbito deste capítulo o desempenho de uma terceira proposta, os ângulos de curvatura do contorno, que apesar de ter sido submetida ao processo de avaliação após a chamada de propostas acabou por desistir de participar no processo de desenvolvimento por *core experiments*. Esta técnica, implementada completamente no contexto desta tese, foi escolhida face à boa pontuação e avaliação global feita pelos especialistas no final do processo de avaliação. A pontuação atribuída a esta proposta (3 pontos em 5) encontra-se acima da média, e ao mesmo nível da transformada de *wavelet* e acima do descritor *CSS* (2.5); note-se que o descritor *CSS* acabou por ser adoptado pela norma MPEG-7. Este facto, juntamente com o facto de esta técnica já ter sido usada por bons mecanismos de descrição e procura na *Internet*, como por exemplo o *QBIC* [AsBa1995], justifica o interesse em conhecer melhor as suas potencialidades e fraquezas, nomeadamente face aos descritores entretanto adoptados pelo MPEG-7.

Assim, estes três descritores – transformada *wavelet*, *CSS* e ângulos de curvatura - serão analisados mais pormenorizadamente nas sub-secções que se seguem, nomeadamente em termos do desempenho de procura e do número de *bits* usado para a sua codificação (caso o método de codificação tenha sido especificado na proposta), a fim de se avaliarem as vantagens e desvantagens de uns em relação aos outros.

▪ Curvature Scale Space (CSS)

O parâmetro de descrição de forma baseado na imagem *CSS* pretende representar a forma de um objecto simples ou região através de propriedades do seu contorno fechado, e foi proposto ao MPEG-7 tal como descrito nas secções 2.2.6.1 e 2.2.6.2. O descritor de forma proposto com base neste parâmetro inclui os seguintes componentes, segundo a estrutura sintáctica apresentada de seguida, codificadas com o número de *bits* indicado [MPEG7-XM1999]:

CSS(){	Número de bits
PeakCount	6
GlobalCurvatureVector [2]	32
if (PeakCount!=0){	
PrototypeCurvatureVector [2]	32
}	
HighestPeak	16
For (k=1; k< PeakCount, k++) {	
Xpeak[k]	16
Ypeak[k]	16
}	
}	

A semântica de cada um dos componentes deste descritor é a que se apresenta de seguida:

- **PeakCount** – especifica o número de picos na imagem *CSS*; os picos da imagem *CSS* correspondem aos zeros da função de curvatura do contorno filtrado;
- **GlobalCurvatureVector** – especifica os parâmetros geométricos e globais do contorno, nomeadamente a excentricidade e circularidade;
- **PrototypeCurvatureVector** – especifica os parâmetros globais da curva protótipo (i.e. a curva filtrada com os parâmetros correspondentes ao pico máximo); a curva

protótipo é definida como a curva alisada através de uma filtragem até que se obtenha uma forma convexa;

- **HighestPeak** – especifica os parâmetros do filtro Gaussiano correspondentes ao pico máximo;
- **Xpeak[k] e Ypeak[k]** – especificam os parâmetros dos restantes picos tendo como referência o pico máximo. O **Xpeak[k]** é a distância normalizada entre dois pontos do contorno, o ponto *k* e o **HighestPeak**, e o **Ypeak[k]** é a altura normalizada do pico *k* [MPEG7-XM1999].

Nesta fase, a avaliação do descritor de forma **CSS** foi feita através do **CE-1** atrás definido, tendo competido com a transformada *wavelet* a fim de integrar o **XM**. Usando as condições anteriormente descritas para este *core experiment*, os proponentes do **CSS** apresentaram os resultados parciais e totais ilustrados na Tabela 5.8 que se segue:

CE-1			CSS	
Partes	A	A1	89.760%	94.565%
		A2	99.370%	
	B		75.439%	
	C		96.000%	
	Média Total			88.668%

Tabela 5.8 – Resultados obtidos pelos proponentes do descritor **CSS** de acordo com o método de avaliação do **CE-1** [Bobe1999b][BoKi1999]

▪ Transformada *Wavelet*

O parâmetro de descrição de forma baseado na transformada *wavelet* pode ser usado para objectos ou regiões simples e caracteriza-se pela sua capacidade de descrição de propriedades locais e globais da forma bem como pela sua capacidade de representar a forma com vários níveis de detalhe. Este parâmetro foi proposto ao MPEG-7 com as propriedades descritas nas secções 2.2.2.1 e 2.2.2.2. O descritor proposto com base neste parâmetro tem a seguinte sintaxe, não sendo apresentado simultaneamente o número de *bits* por componente em virtude de esta informação não ter sido fornecida pelo proponente deste parâmetro [MuOh1999b] [MuOh1999d]:

Wavelet() {
Image_URL
Contour_ID
Modification_ratio
Buffer_size
for (k=0; k<Buffer_size; k++){
Magnitude_wavelet_buffer[k]
}
FilterSet(wavelet_type)
}

A semântica de cada um dos componentes deste descritor é a que se apresenta de seguida:

- **Image_URL** – especifica o endereço URL onde se pode encontrar o contorno descrito;
- **Contour_ID** – especifica o identificador do contorno;
- **Modification_ratio** – especifica a razão entre a diagonal inscrita na bounding box e a diagonal máxima da bounding box;
- **Buffer_size** – especifica o número de coeficientes usados pelo filtro, passa-baixo ou passa-alto;
- **Magnitude_wavelet_buffer[k]** – especifica o coeficiente *wavelet k* do contorno;
- **FilterSet(wavelet_type)** – especifica o tipo de filtros usados, por exemplo passa-baixo ou/e passa-alto.

Nesta fase, a avaliação do descritor de forma baseado na transformada *wavelet* foi feita através do *CE-1*, atrás definido. Usando as condições anteriormente descritas para este *core experiment*, os proponentes do parâmetro de forma baseado na transformada *wavelet* apresentaram os resultados parciais e totais ilustrados na Tabela 5.9 que se segue:

CE-1			Wavelet	
Partes	A	A1	88.040%	97.460%
		A2	92.750%	
	B		67.760%	
	C		93.000%	
	Média Total			84.503%

Tabela 5.9 – Resultados obtidos pelos proponentes e outros especialistas para o descritor baseado na transformada *wavelet* de acordo com o método de avaliação do *CE-1* [BoKi1999] [MuOh1999d]

▪ Ângulos de Curvatura do Contorno (ACC)

O parâmetro de descrição de forma baseado nos ângulos de curvatura do contorno adequado a formas simples, conexas, 2D, foi proposto ao MPEG-7, tal como descrito nas secções 2.2.5.1 e 2.2.5.2. Este descritor tem como objectivo descrever a forma de um objecto através dos valores dos ângulos de curvatura do seu contorno ou seja através de uma aproximação poligonal do contorno. Apesar de os proponentes deste descritor terem desistido de participar no processo de desenvolvimento do MPEG-7, ficou disponível a informação que os seus proponentes apresentaram na fase de avaliação na sequência da chamada de propostas. Como esta informação estava muito incompleta, foi necessário completar, no contexto desta tese, a especificação deste descritor.

Com base na informação fornecida pelos proponentes na fase de avaliação, o descritor ângulos de curvatura do contorno foi implementado no âmbito desta tese. A Figura 5.13 apresenta um esquema simplificado do algoritmo implementado para o cálculo do parâmetro de forma baseado nos ângulos de curvatura do contorno.

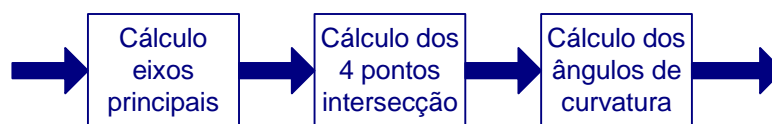


Figura 5.13 – Diagrama de blocos para as 3 etapas do cálculo dos ângulos de curvatura do contorno

Descreve-se de seguida o algoritmo implementado de acordo com o esquema da Figura 5.13:

- **Cálculo dos eixos principais** – Etapa em que são calculados os eixos principais da forma do objecto, ou seja eixos perpendiculares entre si e centrados no centróide do objecto tendo em conta que os eixos têm que ser paralelos aos lados da menor bounding box que contém o objecto;
- **Cálculo dos pontos de intersecção do contorno com os eixos principais** – Etapa em que são determinados os quatro pontos de intersecção entre o contorno do objecto e os eixos principais. Estes pontos são, aqui, também denominados por pontos de começo, ou seja os pontos a partir dos quais se começam a medir os ângulos da curvatura do contorno resultando em quatro vectores de ângulos, um para cada ponto de começo; e
- **Cálculo dos ângulos de curvatura do contorno** – Etapa em que são determinados um conjunto de pontos representativos do contorno e a partir dos quais são calculados os ângulos de curvatura do contorno. Estes ângulos são os ângulos definidos por dois vectores: um vector une dois pontos representativos consecutivos do contorno e o outro corresponde ao eixo principal onde se situa o ponto de começo; são estes ângulos que constituem a informação principal deste descritor.

Descreve-se, de seguida, a implementação feita para cada uma destas etapas salientando as escolhas que tiveram de ser feitas uma vez que neste caso não havia *software* disponível e a proposta inicial do descritor estava muito incompleta. Assim, tudo teve de ser desenvolvido de raiz.

Como o próprio tipo do descritor indica, trata-se de um descritor baseado no contorno, e logo a informação a descrever é essencialmente o contorno do suporte binário do objecto. Deste modo, e visto que a base de dados de conteúdo inclui apenas os suportes binários dos objectos, foi necessário começar por extrair o contorno do objecto a fim de se poder descrever a sua forma.

O contorno do objecto foi extraído a partir do seu suporte binário considerando apenas os *pixels* pertencentes ao objecto, ou seja, os *shapels*. De entre estes *shapels*, extraíram-se os *shapels* extremos do objecto, ou seja os *shapels* que definem o contorno do objecto de acordo com a definição do contorno nas arestas dos *pixels*, considerando uma vizinhança de ordem 4. Define-se um contorno nas arestas como o conjunto de arestas conexas pertencentes ao objecto (ver Figura 5.14a)) e com vizinhança de ordem 4 do *pixel* (i,j) tal que (i,j) seja um *pixel* extremo da forma pertencente ao objecto. A vizinhança de ordem 4 do *pixel* (i,j) , $V_4(i,j)$, é o conjunto dos quatro *pixels* que estão associados ao *pixel* (i,j) (ver Figura 5.14b)) do seguinte modo [Ferr1995]:

$$V_4(i, j) \equiv \{(i, j + 1), (i + 1, j), (i, j - 1), (i - 1, j)\}$$

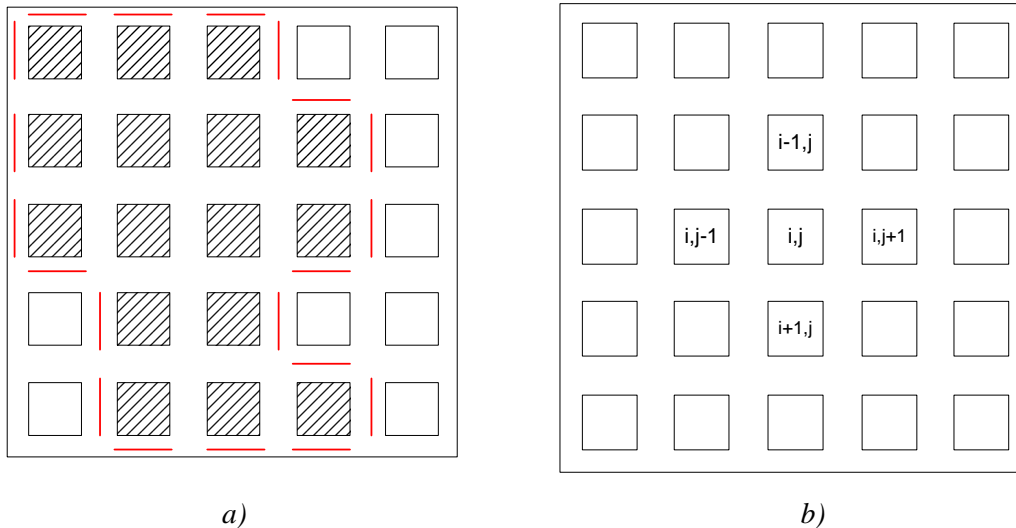


Figura 5.14 – a) Contornos definidos nas arestas dos *pixels* e b) Vizinhança de ordem 4 do *pixel* (i,j) [Ferr1995]

A Figura 5.15 mostra o contorno obtido para a imagem binária *Apple*, focando em particular uma das zonas do contorno para que se possa verificar em detalhe a aparência de um contorno definido nas arestas.

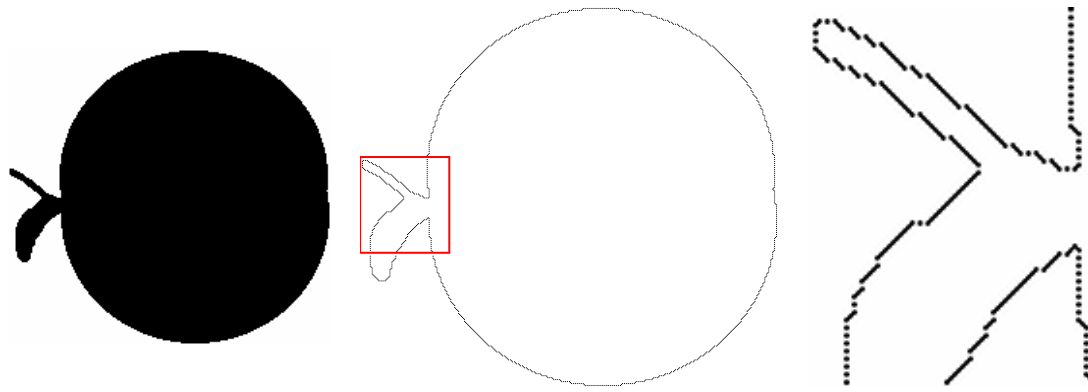


Figura 5.15 – *Apple* - a) Imagem binária; b) contorno correspondente; c) zona ampliada do contorno (marcada em b))

Cálculo dos eixos principais

Após a extracção do contorno do objecto, segue-se o cálculo dos eixos principais do objecto definidos matematicamente pelos seus vectores próprios. Os eixos principais do objecto são determinados pelos vectores próprios da matriz de covariância do contorno, definida através da matriz V com a expressão (5.7). Os eixos principais são identificados através dos valores próprios, $\lambda_{1,2}$, da matriz de covariância; o valor próprio de maior amplitude definirá o *eixo-x*, enquanto que o segundo valor próprio definirá o *eixo-y*.

$$V = \begin{bmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{bmatrix}$$

$$V_{xx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(x_i - m_x)$$

$$V_{xy} = V_{yx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(y_i - m_y)$$

$$V_{yy} = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)(y_i - m_y)$$
(5.7)

$$\lambda_{1,2} = (V_{xx} + V_{yy})^2 \pm \sqrt{(V_{xx} + V_{yy})^2 - 4(V_{xx}V_{yy} - V_{xy}^2)}$$

Nas expressões acima, N é o número de *shapels* pertencentes ao contorno, (x_i, y_i) a coordenada do *shapel* i do contorno e (m_x, m_y) o centróide do objecto definido através da expressão (5.8):

$$m_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad e \quad m_y = \frac{1}{N} \sum_{i=0}^{N-1} y_i$$
(5.8)

Cálculo dos pontos de intersecção do contorno com os eixos principais

Uma vez definidos os eixos principais do objecto, o cálculo dos ângulos de curvatura não é imediato porque um ponto de começo é preciso ser escolhido. Pode-se, eventualmente, começar por escolher qualquer ponto do contorno mas esta escolha seria problemática para o processo de procura. Consequentemente, e de modo a garantir que a descrição seja insensível

à rotação, escolheu-se um método que escolhe o ponto de começo baseado nos momentos (valores próprios) da forma [NiYi1995]. Segundo este critério, o ponto de começo escolhido é o ponto mais distante do centro de massa do objecto. No entanto, esta escolha pode facilmente afectar a descrição da forma e o valor de futuras medidas de semelhança, já que pequenas variações na forma do objecto podem ser responsáveis por grandes alterações nas descrições e pela diminuição de robustez no processo de procura. De modo a tornar o processo de procura mais robusto, um conjunto de até 4 pontos de começo pode ser usado para descrever a forma [NiYi1995].

A proposta feita ao MPEG-7 pelos proponentes deste descritor definiu estes quatro pontos de começo do seguinte modo: Sejam s_1 e s_2 os eixos principais, com momentos correspondentes m_1 e m_2 e seja p_{12} o ponto mais próximo do centro de massa e sobre s_1 . Então, define-se p_{11} como o ponto mais distante do centróide sobre s_1 e no sentido oposto de p_{12} . Analogamente, p_{22} e p_{21} são os pontos mais próximo e mais distante do centróide sobre s_2 e posicionados na direcção oposta, um relativamente ao outro. Consequentemente, múltiplos conjuntos de ângulos de curvatura são calculados apenas em situações de ambiguidade, como quando $m_1 \approx m_2$ e as distâncias entre o centróide e p_{12} e o centróide e p_{11} , o centróide e p_{22} e o centróide e p_{21} são aproximadamente iguais. No entanto, e apesar dos proponentes proporem um critério de selecção que varia o número de vectores de ângulos a usar conforme as situações de ambiguidade, ir-se-á começar por analisar o desempenho do descritor usando-se sempre quatro vectores de ângulos por descrição uma vez que o factor de proximidade destas variáveis não foi definido pelos proponentes deste descritor.

A Figura 5.16 apresenta um exemplo de um objecto onde a presença desta ambiguidade existe, ou seja, os dois pontos de intersecção, p_{12} e p_{11} , sobre o mesmo eixo principal encontram-se à mesma distância D do centróide do objecto. Neste caso, o envio de apenas um vector de ângulos seria problemático porque não se teria a certeza se se estaria a enviar o vector correspondente ao ponto de intersecção p_{12} ou p_{11} .

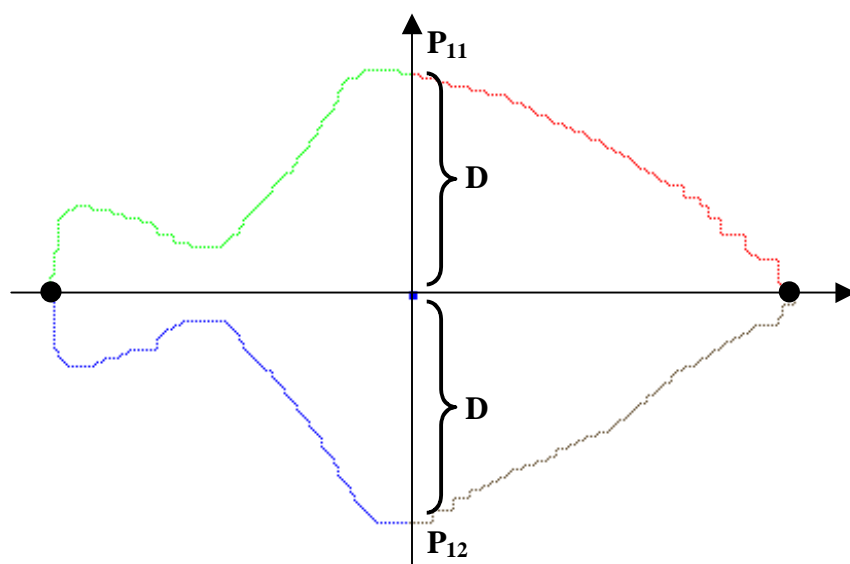


Figura 5.16 – Exemplo de um objecto ambíguo relativamente ao ponto de começo a usar para o cálculo dos ângulos de curvatura

Assim, propõe-se nesta tese, e como ponto de partida, o uso de quatro pontos de começo, resultando consequentemente no uso de quatro vectores de ângulos (um para cada ponto de começo), situados nas intersecções do contorno com os eixos principais de modo a ultrapassar o problema apresentado anteriormente e tornando o parâmetro de descrição mais robusto, essencialmente à rotação.

A Figura 5.17 apresenta os pontos de começo para a forma da imagem binária *Apple*. Na Figura 5.17b) podem ver-se os quatro quadrantes, assinalados através de cores diferentes, bem como os quatro pontos de intersecção/comoço detectados.

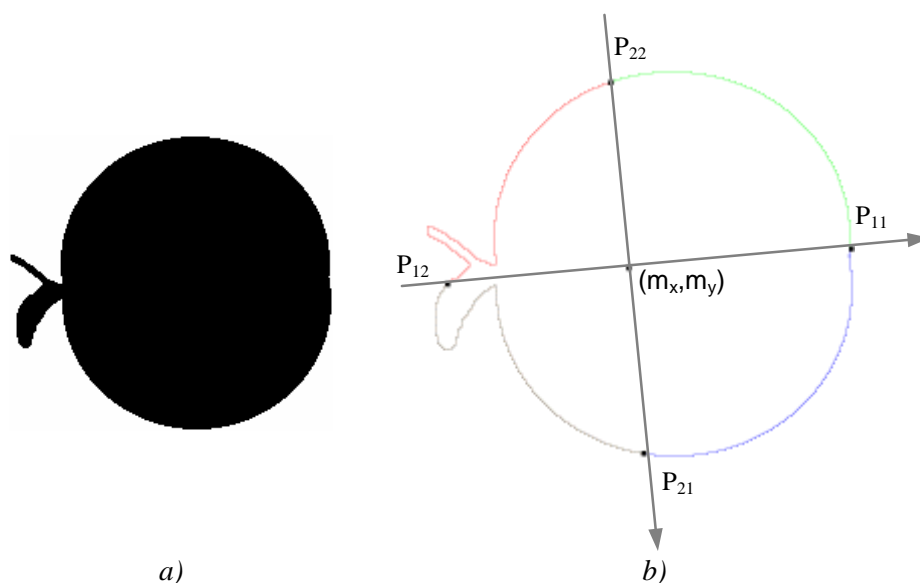


Figura 5.17 – *Apple* – a) imagem binária; e b) identificação dos quatro pontos de começo

Cálculo dos ângulos de curvatura do contorno

Após a determinação de todos os elementos necessários para a fase seguinte do processo de cálculo deste parâmetro, segue-se finalmente o cálculo dos ângulos de curvatura do contorno. Os ângulos de curvatura do contorno são ângulos locais (como mostra a Figura 5.18), em número fixo de 64, segundo a proposta apresentada pelos proponentes deste parâmetro, apesar de se pode usar outro valor [IBMRe1999]. Estes ângulos são calculados a partir de pontos representativos do contorno, calculados a partir de todos os pontos de contorno que estão situados entre pontos do contorno igualmente espaçados de k pontos de contorno: k é a relação existente entre o perímetro e o número de ângulos a usar para descrever a forma do objecto. O contorno deve ser percorrido segundo a direcção contrária à dos ponteiros do relógio.

Sejam $t(i)$ os ângulos de curvatura do contorno, onde $i=0, \dots, N-1$, $N=64$ e (x_j, y_j) com $j=0, \dots, w$, os pontos do contorno fechado do objecto simples ou região. Caso $w < N$, ou seja o perímetro for menor que o número de ângulos a calcular, o y é interpolado e x é fixo de modo a obter-se um conjunto com N pontos. Caso contrário, ou seja se $w > N$, são calculadas as coordenadas médias do ponto representativo, somando-se as coordenadas dos k pontos e dividindo-se finalmente por k , de cada conjunto de $w/N=k$ pontos do contorno fechado, resultando num conjunto de N pontos, $(x'(i), y'(i))$, $i=0, \dots, N-1$, designados por pontos representativos do

contorno (e que não têm de estar sobre o próprio contorno). Os N valores $t(i)$ dos ângulos de curvatura serão então calculados a partir da expressão (5.9):

$$t(i) = \arctg \frac{(y'(i+1) - y'(i))}{(x'(i+1) - x'(i))} \quad (5.9)$$

resultando em quatro vectores de 64 ângulos cada, onde o ponto inicial de cálculo para cada vector corresponde a um dos quatro pontos de começo determinados anteriormente.

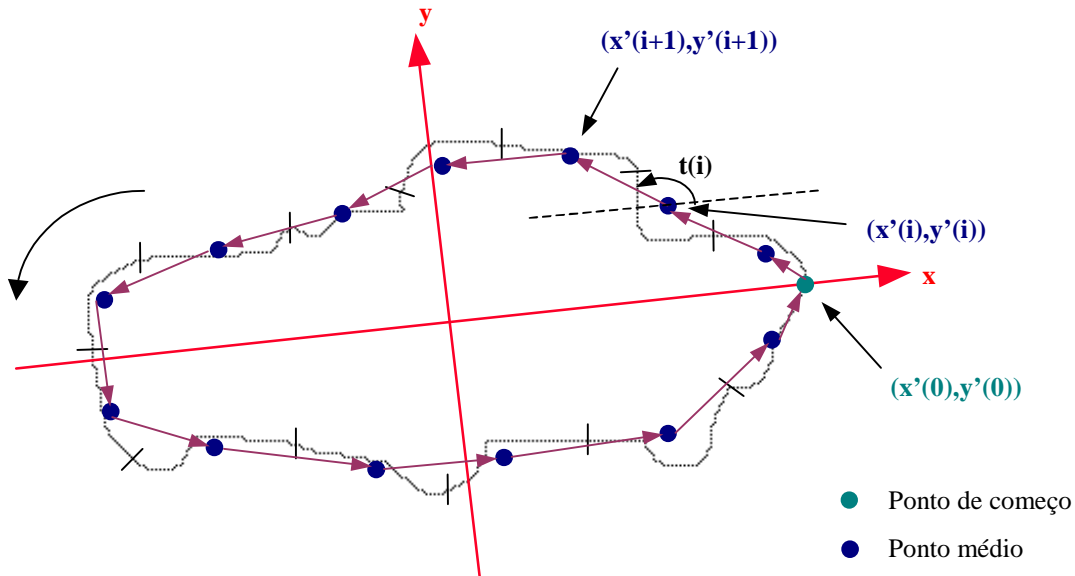


Figura 5.18 – Exemplo da determinação dos ângulos de curvatura, $t(i)$, segundo o algoritmo descrito

Medida de semelhança

A medida de semelhança sugerida pelos proponentes iniciais deste parâmetro foi a distância de *Minkowsky* ou seja o somatório das diferenças absolutas entre os dois vectores de ângulos em questão, um para as imagens na base de dados e o outro para a imagem de procura fornecida pelo utilizador, como mostra a expressão (5.10):

$$D_M = \sum_{k=0}^{N-1} |A_I[k] - A_B[k]| \quad (5.10)$$

onde N é o número de ângulos de curvatura usados para definir a forma do objecto (sempre 64 neste caso), A_I é o vector dos ângulos para a imagem fornecida ao mecanismo de procura e A_B é o vector dos ângulos que descreve as imagens nas bases de dados onde se faz a procura.

Dado que cada descrição conterá quatro vectores de ângulos, existirão dezasseis combinações possíveis de vectores a ser comparadas, para cada par de imagens em comparação. Uma vez calculadas as diferenças entre as dezasseis combinações, a medida final de semelhança entre duas formas corresponde ao menor valor da distância de *Minkowsky* associado a estas combinações, ou seja:

$$D = \min_i \{D_{Mi}\} \quad (5.11)$$

onde i corresponde ao número da combinação.

Uma vez implementados o método de extracção e a medida de semelhança aqui descritos, procedeu-se à sua integração no *XM*, de acordo com as especificações de integração definidas pelo grupo MPEG. Deste modo, o parâmetro de descrição aqui apresentado pôde ser avaliado e comparado, em pé de igualdade, com os restantes parâmetros de descrição avaliados nesta primeira fase do processo de desenvolvimento dos descritores de forma MPEG-7.

A sintaxe inicialmente sugerida pelos proponentes deste parâmetro foi [IBMR99]:

TurningAngles() {
Number_angles
Moments_vector[2]
Distance_vector[4]
Index_vector[4]
for (x=0; x<4; x++){
for (k=0; k<Number_angles; k++){
Turning_angle[x][k]
}
}
}

A semântica de cada um dos seus componentes é a que se apresenta de seguida:

- **Number_angles** – especifica o número de ângulos, 64, usados para descrever a forma do objecto (se o número de ângulos usados for sempre 64, não há necessidade de especificar este componente sintáctico);
- **Moments_vector** – especifica os dois momentos da forma, ou seja, os valores próprios correspondentes aos vectores próprios dos eixos principais do objecto;
- **Distance_vector** – especifica a distância de cada um dos quatro pontos de começo do objecto (pontos de intersecção entre os eixos principais e o contorno do objecto e em relação aos quais se medem os ângulos de curvatura do contorno ao longo do contorno) ao centróide do objecto;
- **Index_vector** – especifica o índice do ângulo calculado a partir do ponto de começo escolhido e o ponto seguinte representativo do contorno;
- **Turning_angle** – especifica o conjunto dos ângulos de curvatura do contorno do objecto; cada um destes ângulos é o ângulo entre um vector que une dois pontos representativos médios consecutivos e o eixo principal onde se situa o ponto de começo usado, sendo os pontos representativos do contorno calculados a partir do contorno percorrido no sentido contrário ao dos ponteiros do relógio.

Finalmente, a avaliação do desempenho do descritor de forma baseado nos ângulos de curvatura do contorno tal como foi implementado de acordo com a informação fornecida pelos proponentes iniciais desta proposta, e completada nesta tese, foi feita nas condições

definidas para o *core experiment CE-1* atrás definido, apresentando os resultados parciais e totais ilustrados na Tabela 5.10:

CE-1			CSS		Wavelet		Ângulos de Curvatura	
Partes	A	A1	89.760%	94.565%	88.040%	97.460%	66.667%	79.288%
		A2	99.370%		92.750%		91.910%	
	B		75.439%	67.760%	54.14%			
	C		96.000%	93.000%	63.500%			
	Média Total (Bits)			88.668%	84.503%	65.643%		

Tabela 5.10 – Resultados obtidos no âmbito desta tese para os descritores *CSS*, baseado na transformada *wavelet* e ângulos de curvatura do contorno segundo as condições do *CE-1*

Com base nos resultados de desempenho apresentados, nesta fase, pelos proponentes dos descritores *CSS* e transformada *wavelet*, os especialistas do grupo MPEG responsáveis pelos *core experiments* da forma escolheram o descritor *CSS* como o descritor de forma baseado no contorno a integrar no *XM*. A preferência por parte dos especialistas pelo *CSS* deve-se unicamente ao facto de este ter um melhor desempenho, 88.668%, face ao descritor associado à transformada *wavelet*, 84.503% como se pode ver na Tabela 5.10. Quanto ao descritor baseado nos ângulos de curvatura do contorno, o seu desempenho comparativamente com os outros é bastante fraco, contrariamente ao que se poderia esperar analisando o relatório do processo de avaliação. Pensa-se que esta elevada diferença de desempenho se fica a dever à especificação pouco optimizada do descritor submetida pelos proponentes iniciais. Por este motivo, este descritor será desenvolvido ao longo desta tese tanto em termos do método de extracção como do método de procura, de modo a melhorar o seu desempenho relativamente aos restantes descritores avaliados.

5.3.2.2 Descritores de forma baseados na região

No capítulo 2 foram apresentados e definidos alguns dos parâmetros de descrição de forma baseados na região mais comumente usados na área da análise e processamento de imagem como os parâmetros baseados nas transformadas, nomeadamente os momentos de *Zernike* e os *MLEV*. Como já foi dito neste capítulo, apenas duas das propostas de descritores de forma baseados na região participaram na primeira fase do processo de desenvolvimento do MPEG-7, os momentos de *Zernike* e os *MLEV*. Neste contexto, estes dois descritores serão avaliados mais pormenorizadamente nas sub-secções que se seguem, nomeadamente em termos do seu desempenho de procura e do número de *bits* usado para a sua codificação (caso o método de codificação tenha sido especificado na proposta), a fim de se avaliarem as vantagens e desvantagens de um descritor em relação ao outro.

▪ Momentos de Zernike

O parâmetro de descrição da forma baseado nos momentos de *Zernike* pode ser usado para descrever formas simples ou complexas, tal como descrito nas secções 2.3.4.1 e 2.3.4.2. O descritor de forma proposto com base neste parâmetro inclui os seguintes componentes, segundo a estrutura sintáctica apresentada de seguida, codificados com o número de *bits* indicado [KiKi1999b][MPEG7-XM1999]:

MomentosZernike(){	Número de bits
ZernikeMomentsNumber	16
for (k=0; k< ZernikeMomentsNumber; k++){	
Magnitude[k]	32
}	
}	

A semântica dos seus componentes é a que se apresenta de seguida:

- ***ZernikeMomentsNumber*** – especifica o número de momentos de *Zernike* usados para a descrição da forma do objecto em questão;
- ***Magnitude[k]*** - especifica a amplitude do momento de *Zernike* de ordem *k*.

Nesta fase, a avaliação do descritor de forma baseado nos momentos de *Zernike* foi feita através dos *core experiments CE-1* e *CE-2*, atrás definidos. Note-se que enquanto os descritores baseados no contorno não podem ser avaliados nas condições do *CE-2* por este incluir nas suas bases de dados de conteúdo formas complexas (nomeadamente não conexas), os descritores baseados na região podem ser avaliados quer nas condições do *CE-1* (formas simples), quer nas condições do *CE-2* (formas complexas).

Usando as condições anteriormente descritas para os *core experiments CE-1* e *CE-2*, os proponentes do parâmetro de forma baseado nos momentos de *Zernike* apresentaram os resultados parciais e totais ilustrados na Tabela 5.11:

Core Experiments			Momentos de Zernike	
CE-1	A	A1	92.540%	96.072%
		A2	99.603%	
	B		70.221%	
	C		94.500%	
	Média Total		86.931%	
CE-2	A	A1	69.800%	83.350%
		A2	94.184%	
		A3	70.455%	
		A4	70.455%	
	B		67.931%	
	Média Total		75.640%	

Tabela 5.11 – Resultados obtidos para os CE-1 e CE-2 com o descritor momentos de Zernike [BoKi1999]

▪ Multi-Layer Eigen Vectors (MLEV)

O parâmetro de descrição da forma baseado em vectores próprios multi-nível (*MLEV* de *Multi-Layer Eigen Vectors*) é uma representação da forma baseada numa estrutura multi-escalar, onde cada nível introduz mais detalhe/informação relativamente à forma do objecto complexo. Este parâmetro foi proposto ao MPEG-7 tal como descrito nas secções 2.3.2.1 e 2.3.2.2. O descritor proposto com base neste parâmetro tem a seguinte sintaxe, não sendo apresentado simultaneamente o número de *bits* por componente devido à não disponibilização desta informação pelo seu proponente [KiKi1999c]:

```

MLEV(){
    Number_regions
    for (k=0; k<Number_regions; k++){
        Eigen_ratio[k]
        Normal_angle[k]
        Center[k]
        Compactness[k]
    }
}

```

A semântica dos componentes deste descritor é a que se apresenta de seguida:

- ***Number_regions*** – especifica o número de regiões em que a forma do objecto foi subdividida para a descrição em questão;
- ***Eigen_ratio[k]*** – especifica a razão entre os valores próprios calculados para os eixos principais de cada região em que a forma do objecto foi subdividida;
- ***Normal_angle[k]*** – especifica o menor ângulo absoluto feito entre o maior vector próprio da região k e o vector próprio do eixo principal;
- ***Center[k]*** – especifica a distância entre o centro de massa da forma do objecto e o centro da região k ;
- ***Compacteness[k]*** – especifica o factor de compactação relativo à região k ; a compactação define a relação existente entre a área ocupada por uma região k e a sua *bounding box* mais justa, ou seja, a *bounding box* cujos lados são paralelos aos eixos principais do objecto;

Nesta fase, o desempenho do descritor de forma baseado nos *MLEV* foi medido através dos *CE-1* e *CE-2* atrás definidos. Usando as condições anteriormente descritas para estes *core experiments*, os proponentes do parâmetro de forma baseado nos *MLEV* apresentaram os resultados parciais e totais ilustrados na Tabela 5.12:

Core Experiments			MLEV	
CE-1	A	A1	92.421%	96.211%
		A2	100%	
	B		70.332%	
	C		88.000%	
	Média Total		84.848%	
CE-2	A	A1	81.600%	82.694%
		A2	98.256%	
		A3	98.223%	
		A4	52.686%	
	B		41.818%	
	Média Total		62.256%	

Tabela 5.12 – Resultados obtidos para os *CE-1* e *CE-2* com o descritor *MLEV*
[BoKi1999][KiKi1999e]

Com base nos resultados de desempenho apresentados nesta fase pelos proponentes dos parâmetros de descrição momentos de *Zernike* e *MLEV*, os especialistas do grupo MPEG responsáveis pelos *core experiments* da forma adoptaram os momentos de *Zernike* como o

descriptor de forma baseado na região a integrar no *XM* MPEG-7 e logo na norma. A preferência por parte dos especialistas pelos momentos de *Zernike* deve-se unicamente ao facto de este descriptor ter um desempenho melhor em ambos os *core experiments* *CE-1* e *CE-2*, 88.931% e 75.640%, relativamente ao *MLEV* com 84.848% e 62.256%, respectivamente.

5.3.3 Segunda fase do processo de desenvolvimento

A segunda fase do processo de desenvolvimento dos descritores correspondeu ao *core experiment CE-1/2* Optimizado, ou seja à optimização dos descritores de forma baseados no contorno e na região, melhorando os métodos de extracção e procura/semelhança com vista a obter um melhor desempenho. É de salientar que a metodologia de cálculo do desempenho dos descritores em nada se altera neste *core experiment* em relação àquela já aplicada nos *core experiments CE-1* e *CE-2*.

5.3.3.1 Descritores de forma baseados no contorno

Na segunda fase do processo de desenvolvimento, e já depois de escolhido o descriptor de forma baseado no contorno com base no seu desempenho em termos de procura (*CE-1*), foi realizado um *core experiment* com vista a otimizar o descriptor em termos da sua medida de semelhança. Ainda que não seja normativa, a medida de semelhança é essencial para o bom desempenho do descriptor e, como tal, o MPEG tenta sugerir uma boa medida de semelhança para que os utilizadores da norma possam facilmente alcançar bons desempenhos.

Assim sendo, e observando os resultados que o descriptor *CSS* apresentou durante a primeira fase de desenvolvimento, pode constatar-se que estes tinham uma taxa de falsos resultados (ou seja resultados maus bem colocados em termos de medida de semelhança) significativa, especialmente no que diz respeito à parte B do *CE-1*, ou seja a procura baseada na semelhança perceptiva. Na sequência destas observações, foi especificado um *core experiment*, *CE-1/2* Optimizado, com vista a melhorar a medida de semelhança sugerida inicialmente para este descriptor e tentar eliminar esta ‘debilidade’ do descriptor.

As optimizações feitas à medida de semelhança do *CSS* centraram-se essencialmente nos limiares dos componentes de excentricidade e circularidade, E_L e C_L , respectivamente. Estes limiares estão associados ao primeiro nível da medida de semelhança, correspondente a uma primeira selecção da base de dados, ou seja depende da escolha destes limiares que as formas dos objectos mais diferentes sejam de imediato eliminadas do conjunto de formas candidatas à análise do segundo nível com a medida de semelhança de *Minkowsky* tal como explicado na secção 2.2.1.2.

Com base nos resultados deste *core experiment*, os proponentes deste descriptor decidiram alterar os valores iniciais de E_L e C_L de **0.36** e **0.77** para **0.8** e **0.9** a fim de melhorarem o desempenho do descriptor. Face a esta alteração, o descriptor *CSS* conseguiu ver o seu desempenho, para as partes A1 e B do *core experiment CE-1*, melhorado em cerca de 1.5% [Bobe1999c], ou seja, passou de 89.76% para 92.42% e de 75.44% para 77.82%, respectivamente, como mostra a Tabela 5.13.

Core Experiments			CSS – antes da optimização (1ª fase do processo de desenvolvimento)		CSS – depois da optimização (2ª fase do processo de desenvolvimento)	
CE-1	A	A1	89.760%	94.565%	92.42%	96.210%
		A2	99.370%		100%	
	B		75.439%	77.820%		
	C		96.000%	96.000%		
	Média Total		88.668%	90.010%		

Tabela 5.13 – Desempenho do descritor de forma baseado no contorno, CSS, antes e após a optimização dos limiares E_L e C_L da sua medida de semelhança [Bobe1999c]

5.3.3.2 Descritores de forma baseados na região

Na segunda fase do processo de desenvolvimento, e já com o descritor de forma baseado na região escolhido a partir do desempenho em termos de procura, foi realizado um *core experiment* com vista a optimizar o descritor em termos do seu método de extracção (não-normativo); as alterações em termos do método de extracção implicam alterações na semântica do descritor, essa sim normativa. A optimização esteve associada à alteração da forma sobre a qual se calculam os momentos de *Zernike* que deixou de ser a forma do objecto para passar a ser uma forma ou região exterior e uma forma ou região interior definidas através de um raio de circunferência r , menor que o valor do raio da menor circunferência que engloba toda a forma do objecto centrada no seu centro de massa (ver Figura 5.19). Assim sendo, a região exterior é definida através dos *shapels* cuja distância ao centro de massa é maior ou igual a r e a forma interior é definida através dos *shapels* cuja distância ao centro de massa é inferior a r .

A Figura 5.19 apresenta um exemplo para a determinação das formas interior e exterior para uma dada forma. A Figura 5.19a) representa a forma de um dado objecto e a Figura 5.19b) mostra a menor circunferência que engloba toda a forma e o centro de massa do objecto. Por sua vez, as Figura 5.19c) e d) apresentam as duas novas formas exterior e interior geradas para um dado valor de r . Os momentos de *Zernike* são então calculados para as duas novas regiões, determinando-se o valor de r de acordo com o raio da função base dos momentos de *Zernike*; enquanto a parte interior da região exterior é preenchida com *shapels*, de valor igual aos pertencentes a um objecto, calculando-se assim os momentos da região exterior, os *shapels* pertencentes à região exterior são preenchidos como fundo no cálculo dos momentos de *Zernike* para a região interior [KiKi1999f].

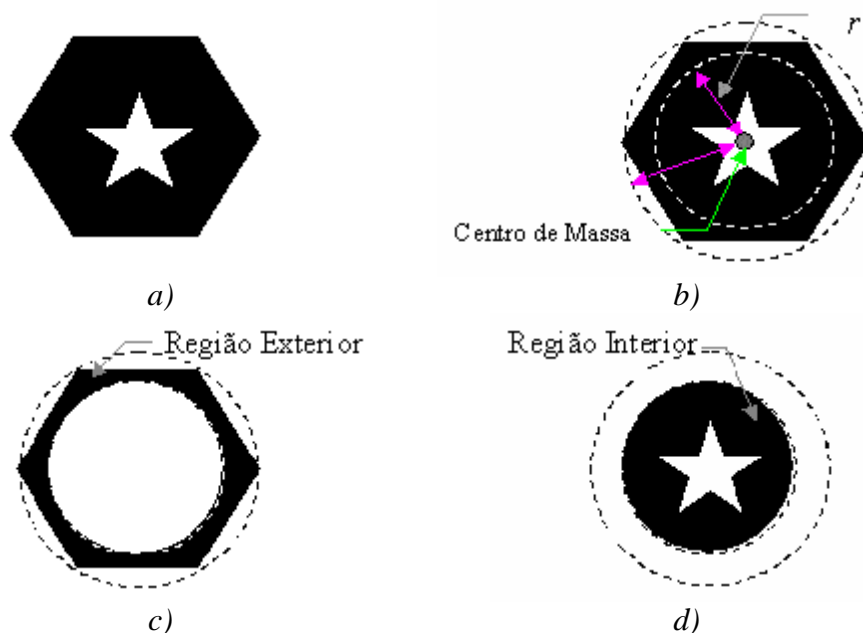


Figura 5.19 – a) forma do objecto; b) menor circunferência que engloba o objecto; c) região interior para um dado raio r ; e d) região exterior para um dado raio r [KiKi1999f]

De acordo com o novo método de extracção proposto, o descritor de forma baseado na região viu as suas sintaxe e semântica serem alteradas, passando a exibir a sintaxe que se apresenta de seguida [KiKi1999f]:

Region_based_Shape_Descriptor(){	Número de bits
Outer_Number_of_Zernike_moments	16
For (k=0; k< Outer_Number_of_Zernike_moments; k++){	
Outer_Magnitude_of_Zernike_moments[k]	32
}	
Inner_Number_of_Zernike_moments	16
For (k=0; k< Inner_Number_of_Zernike_moments; k++){	
Inner_Magnitude_of_Zernike_moments[k]	32
}	
}	

A semântica dos seus componentes é a que se apresenta de seguida:

- **Outer_Number_of_Zernike_moments** – especifica o número de momentos de *Zernike* usados para descrever a região exterior da forma do objecto;
- **Inner_Number_of_Zernike_moments** – especifica o número de momentos de *Zernike* usados para descrever a região interior da forma do objecto;
- **Outer_Magnitude_of_Zernike_moments[]** – contém os valores dos momentos de *Zernike* para a região exterior da forma do objecto;
- **Inner_Magnitude_of_Zernike_moments[]** – contém os valores dos momentos de *Zernike* para a região interior da forma do objecto.

Uma vez que a especificação do descritor foi modificada, a medida de semelhança aconselhada foi consequentemente alterada de modo a contabilizar os novos componentes do descritor. Assim sendo, a medida de semelhança entre duas descrições e logo entre duas formas é agora calculada através das expressões (5.12), (5.13) e (5.14):

$$D_{ZO} = \sum_{i=0}^{O-1} W_i \times (M_I[i] - M_Q[i]) \quad (5.12)$$

$$D_{ZI} = \sum_{i=0}^{I-1} W_i \times (M_I[i] - M_Q[i]) \quad (5.13)$$

$$D_Z = W_O \times D_{ZO} + (1 - W_O) \times D_{ZI} \quad (5.14)$$

onde D_{ZO} , D_{ZI} e D_Z são as distâncias para as regiões exteriores, interiores e conjunta das formas em comparação, respectivamente, W_i é o peso do momento i , W_O é o peso da região exterior no cálculo da medida de semelhança total (conjunta), O e I são o número de momentos usados para as regiões exteriores e interiores, respectivamente, e M_I e M_Q são os vectores com os momentos das imagens binárias na base de dados e da imagem dada como referência no processo de procura, respectivamente [KiKi1999f].

Os resultados obtidos pelos proponentes para o desempenho após esta modificação do parâmetro de forma baseado nos momentos de *Zernike*, foram calculados tendo W_i [KiKi1999f] que obedecer a dois tipos de distribuições diferentes, para selecção da melhor distribuição: 1) desvio padrão e 2) *Fischer*. Deste modo os proponentes dos momentos de *Zernike* optaram por avaliar o desempenho deste parâmetro de descrição atribuindo pesos diferentes aos momentos de cada uma das regiões, resultando nos valores apresentados na Tabela 5.14.

Core Experiments			Momentos de Zernike (antes da modificação)		Momentos de Zernike (W _i desvio padrão)		Momentos de Zernike (W _i Fischer)	
CE-1	A	A1	95.159%	97.580%	94.881%	97.441%	95.000%	97.500%
		A2	100%		100%		100%	
	B		70.843%	70.268%	68.564%			
	C		94.000%	95.000%	95.000%			
	Média Total		87.474%	87.570%	87.021%			
CE-2	A	A1	74.400%	74.813%	72.800%	82.973%	73.400%	82.411%
		A2	97.857%		97.551%		97.347%	
		A3	97.300%		96.804%		97.906%	
		A4	55.179%		64.738%		60.992%	
	B		68.441%	64.668%	68.085%			
	Média Total		74.813%	73.821%	75.248%			

Tabela 5.14 – Desempenho dos momentos de Zernike usando como pesos: a) desvio padrão e b) distribuição de Fischer [KiKi1999h][KiKi1999i]

Como se pode verificar através dos resultados apresentados na Tabela 5.14, os ganhos obtidos no desempenho do descritor para os CE-1 e CE-2 usando o desvio padrão e os pesos de Fischer foram, na generalidade, melhores em 0.639% e -1.819%, e 0.09% e -0.392%, respectivamente, em relação à situação anterior. No entanto, o aumento da quantidade de informação de descrição necessário para obter ganhos de desempenho desta ordem de grandeza (pequenos), não justifica a alteração proposta ao descritor. Consequentemente, a alteração estudada neste core experiment não foi adoptada e logo o descritor baseado nos momentos de Zernike ficou como estava.

5.3.4 Terceira fase do processo de desenvolvimento

A terceira fase do processo de desenvolvimento consistiu essencialmente na realização do core experiment MS-3, visando a optimização da dimensão codificada dos descritores (número de bits) em função dos níveis de quantificação usados para alguns dos componentes dos descritores.

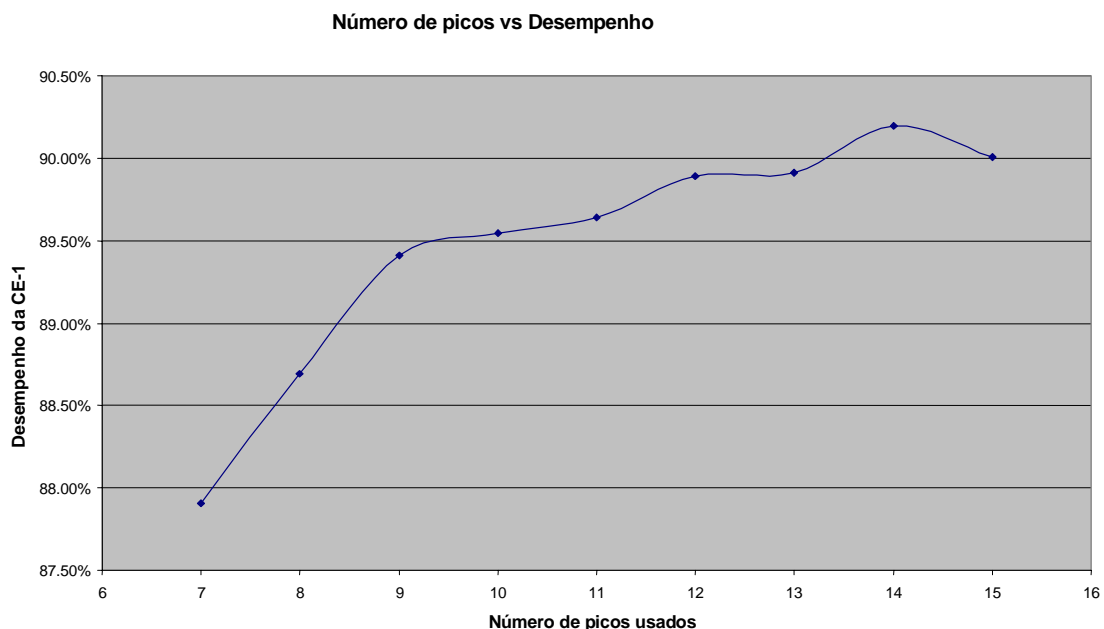
5.3.4.1 Descritores de forma baseados no contorno

A terceira fase do processo de desenvolvimento do descritor *CSS* está associada à implementação do *core experiment MS-3*. Durante esta fase de desenvolvimento, o descritor de forma baseado no contorno e integrado no *XM, CSS*, foi optimizado em termos da sua dimensão codificada, segundo as condições definidas para o *core experiment MS-3*. Neste contexto, a atribuição de *bits* a cada componente do descritor foi rigorosamente testada de modo a conseguir-se o melhor compromisso em termos do desempenho de procura e da dimensão em *bits* do descritor, medida como a média do número de *bits* usado para cada descrição das bases de dados de teste.

Para o descritor *CSS*, a dimensão codificada depende fortemente de dois factores:

1. Número de picos usados para descrever a forma do objecto, consoante se quer maior ou menor detalhe na descrição;
2. Níveis de quantificação aplicados a cada componente do descritor (i.e. parâmetros globais e valores de x_{peak} e y_{peak}).

O *core experiment* desenrolou-se fixando os níveis de quantificação dos vários componentes relevantes para que se pudesse determinar o número de picos *CSS* óptimo para descrever a forma de um objecto, tendo-se posteriormente variado o número de picos a usar na descrição; neste contexto, os proponentes do descritor obtiveram os resultados apresentados na Figura 5.20a) e b). A partir destes resultados, pode concluir-se que o melhor compromisso entre o desempenho de procura e a dimensão do descritor é conseguido usando-se 14 picos *CSS* para descrever a forma de qualquer objecto da base de dados de conteúdo definida para o *CE-1*.



a)

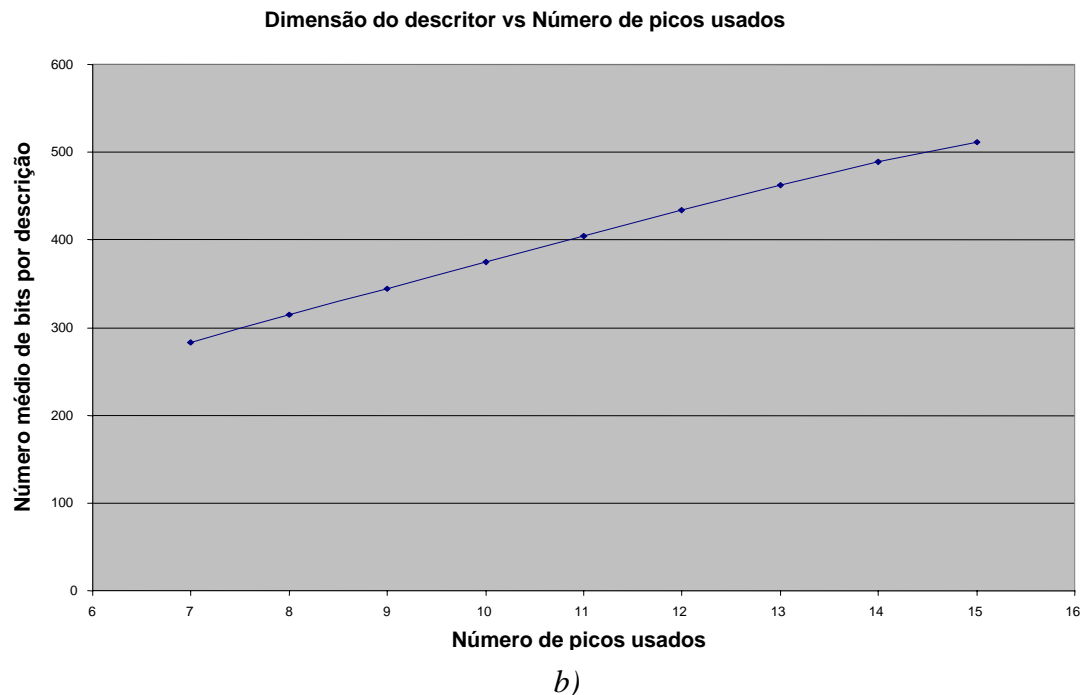
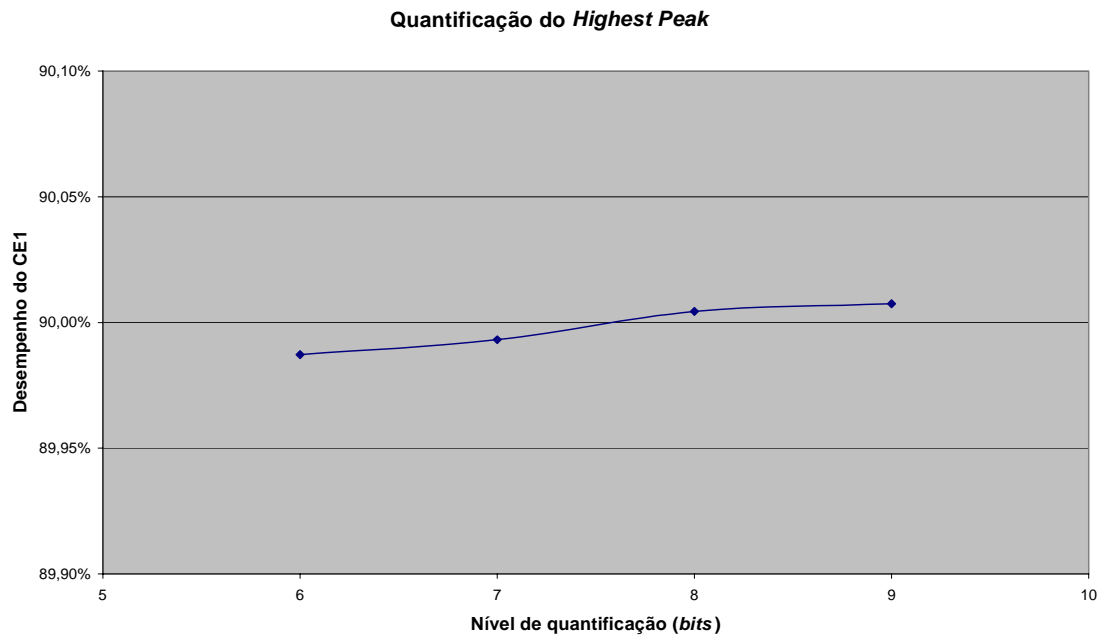
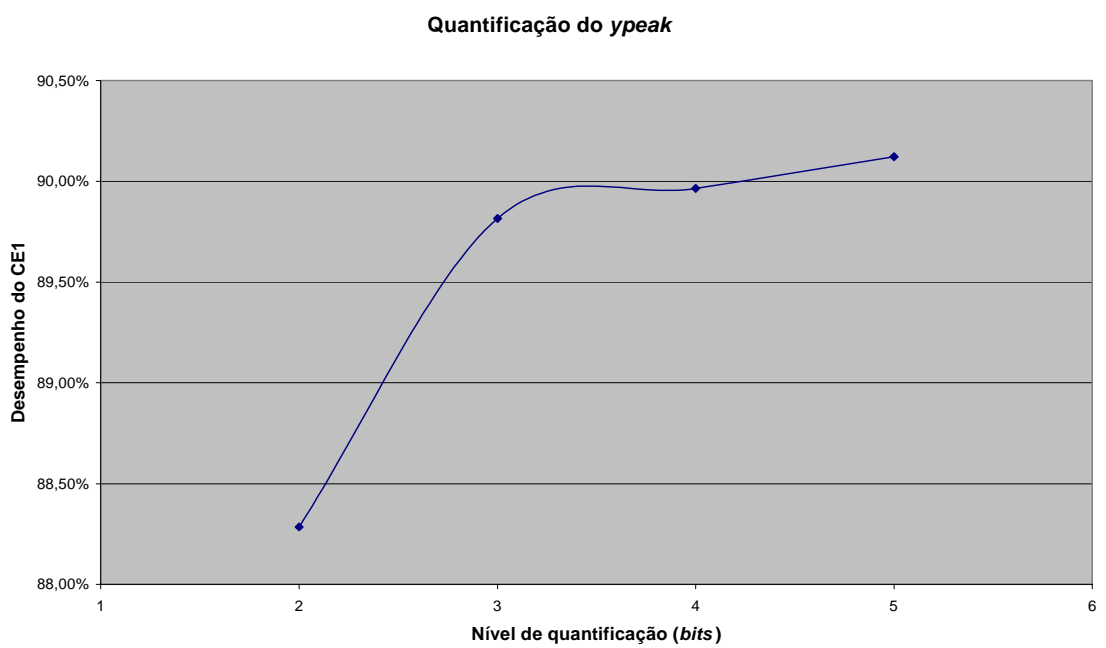


Figura 5.20 – Resultados obtidos para: a) desempenho do descritor CSS em função do número de picos e b) dimensão do descritor em função do número de picos [BoAt2000a]

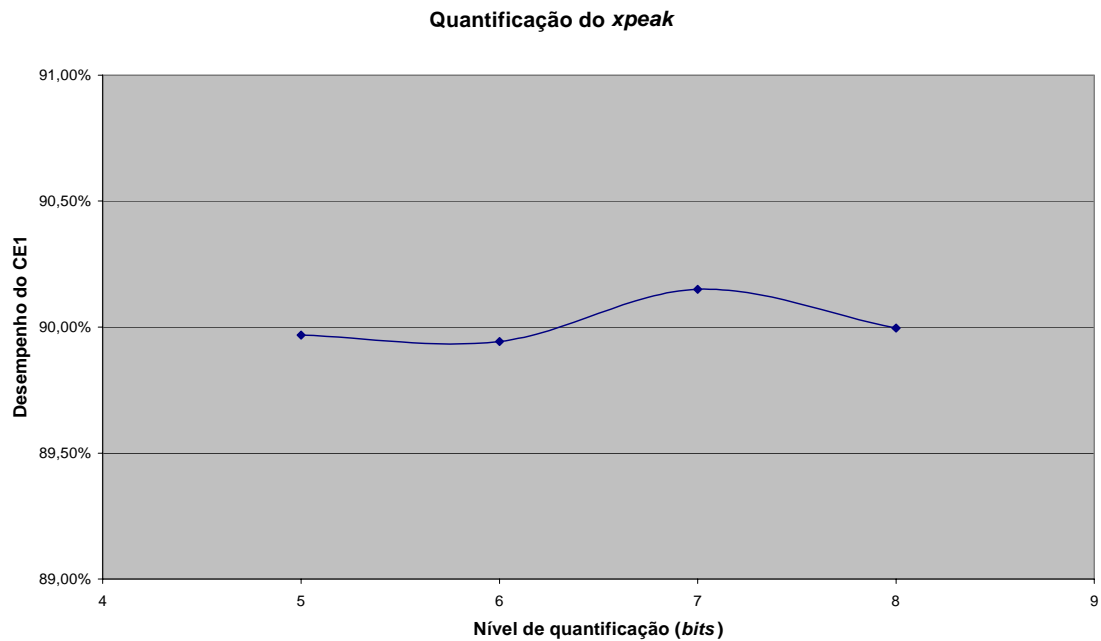
Com base neste valor óptimo para o número de picos a usar, os proponentes avaliaram a seguir o desempenho em função dos níveis de quantificação (medidos em *bits* de codificação para esses níveis) para os vários componentes do descritor. A avaliação dos níveis de quantificação foi feita de modo a conseguir o melhor compromisso em termos do desempenho e da dimensão em *bits* para um dado componente do descritor e consequentemente para o descritor. Deste modo, os componentes do descritor para os quais se variaram os níveis de quantificação foram o *HighestPeak*, *y_peak*, *x_peak* e o *Curvature_Vector*, que tinham anteriormente como número de *bits* 16, 16, 16 e 32, respectivamente (agora também os valores de defeito quando se varia o nível de quantificação para um dos componentes). A variação do desempenho em função do nível de quantificação para cada um dos componentes, é apresentada nas Figura 5.21a), b), c) e d), respectivamente.



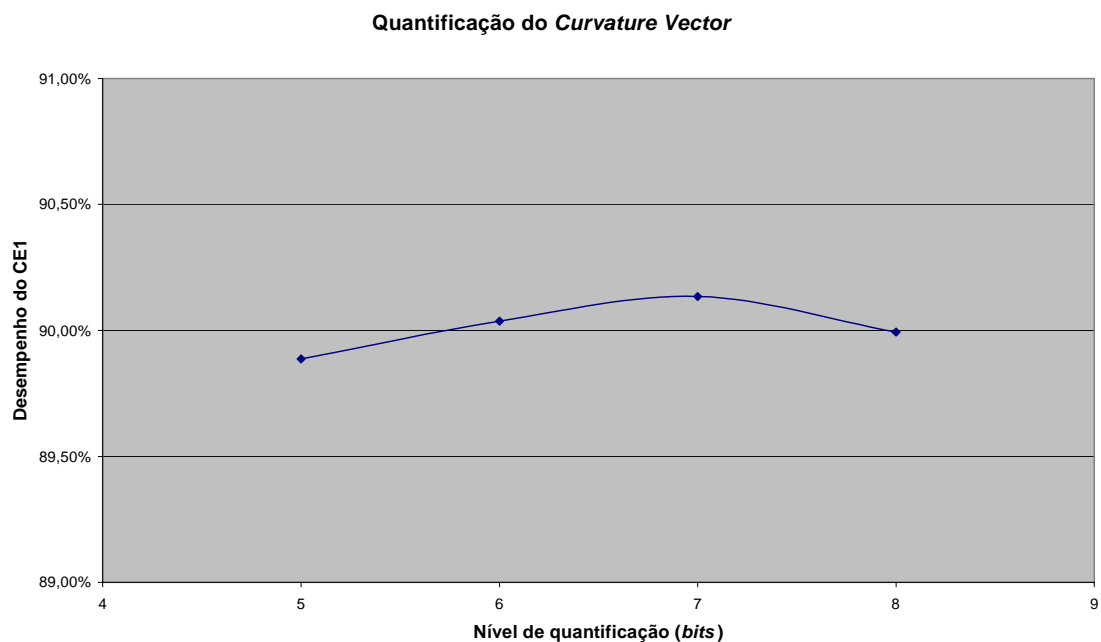
a)



b)



c)



d)

Figura 5.21 – Desempenho do descriptor *CSS* em função do nível de quantificação dos componentes: a) *HighestPeak*, b) *ypeak*, c) *xpeak* e d) *Curvature_Vector* [BoAt2000a]

Face aos resultados observados nos gráficos acima e à comparação entre o número médio de *bits* usados por descrição antes e depois da optimização dos níveis de quantificação (ver Tabela 5.15), constatou-se que apesar de o número médio de *bits* por descrição diminuir cerca de 77.34%, ou seja 4.4 vezes após a optimização, o desempenho médio total do descriptor decresce apenas 1.25% (ver Tabela 5.16) o que não parece à partida positivo. No entanto, face ao elevado ganho conseguido relativamente à dimensão do descriptor aceita-se esta perda de desempenho que é considerada como pouco significativa, com vista a obter um bom compromisso entre os dois elementos em análise: desempenho versus dimensão do descriptor.

Deste modo, os proponentes decidiram atribuir aos vários componentes do descritor o seguinte número de *bits*: *HighestPeak* - 7 *bits*, *y_peak* - 3 *bits*, *x_peak* - 6 *bits* e ao *Curvature_Vector* - 12 *bits*, visto que são estes valores os que oferecem o melhor compromisso entre a compactação do descritor e o seu desempenho em termos de procura (ver Tabela 5.17) para as bases de dados definidas para o *core experiment CE-1*.

CE MS-3			Bits por descrição (antes optimiza.)		Bits por descrição (depois optimiza.)	
Partes	A	A1	476	485.5	111	111.5
		A2	495		112	
	B		500		114	
	C		524		117	
	Média Total (Bits)			503		114

Tabela 5.15 – Número médio de *bits* usados antes e depois da optimização do descritor CSS [BoAt2000a][HeKa2000]

Core Experiment			Variação do desempenho após CE MS-3	
CE-1	A	A1	-4.49%	-2.245%
		A2	0%	
	B		-1.51%	
	C		0%	
	Média Total		-1.25%	

Tabela 5.16 – Perdas de desempenho do descritor CSS após a optimização da sua dimensão

A Tabela 5.17 apresenta os resultados obtidos para o desempenho nesta fase de desenvolvimento assim como os da fase anterior. Há aqui que salientar que foi pela primeira vez possível confirmar os resultados obtidos pelos proponentes através do mecanismo de descrição e procura desenvolvido no âmbito desta tese. Este facto deve-se unicamente ao facto de não se ter tido acesso às versões iniciais do *software* para os vários descritores porque as suas primeiras versões surgiram antes de se ter iniciado esta tese.

CE MS-3			Desempenho CSS – antes da optimização com o CE MS-3		Desempenho CSS – depois da optimização com o CE MS-3		Desempenho CSS – resultados depois do CE-MS3 confirmados pelo mecanismo de descrição e procura desta tese	
Partes	A	A1	92.42%	96.210%	87.930%	93.970%	89.760%	94.880%
		A2	100%		100%		100%	
	B		77.820%		76.310%		76.260%	
	C		96.000%		96.000%		96.000%	
Média Total)			90.010%		88.760%		88.960%	

Tabela 5.17 – Desempenho do descritor CSS antes e depois da optimização em termos da dimensão do descritor [BoAt2000a][HeKa2000]

Como se pode verificar na Tabela 5.17, os resultados obtidos através do mecanismo de descrição e procura implementado nesta tese são muito idênticos aos resultados obtidos pelos proponentes deste descritor, com excepção do desempenho para a parte A1 que é 1.83% superior para o mecanismo da tese. Esta diferença deve-se ao facto de existirem inicialmente algumas imagens binárias na base de dados com irregularidades e que provocavam alguns falsos resultados. Estas irregularidades foram posteriormente corrigidas e foi com base nesta base de dados corrigida que os resultados gerados pelo mecanismo da tese foram obtidos o que justifica o ligeiramente melhor desempenho.

Na sequência desta optimização, o descritor de forma baseado no contorno, CSS, adoptado pelo MPEG-7 é o seguinte [BoAt2000a]:

CSS ()	Número de bits
PeakCount	6
GlobalCurvatureVector [2]	12
If (PeakCount!=0){	
PrototypeCurvatureVector [2]	12
}	
HighestPeak	7
For (k=1; k< PeakCount, k++) {	
xpeak[k]	6
ypeak [k]	3
}	

A semântica é a mesma já anteriormente descrita.

5.3.4.2 Descritores de forma baseados na região

Nesta terceira fase do processo de desenvolvimento dos descritores de forma baseados na região em que o objectivo era optimizar o desempenho do descritor em função da sua dimensão, surgiu uma nova proposta para descritor de forma baseado na região. O descritor proposto, baseado na transformada *Angular-Radial Transform (ART)* e também definido no

capítulo 2, foi apresentado com o objectivo de substituir o descritor até então escolhido ou seja o descritor baseado nos momentos de *Zernike*. Contudo, para que essa substituição pudesse acontecer, o *ART* teve que demonstrar o seu bom desempenho em todos os testes já passados pelo descritor baseado nos momentos de *Zernike* ou seja apresentar melhor desempenho nas condições definidas para os *CE* relevantes, ou seja *CE-1*, *CE-2* e *CE MS-3*. Assim sendo, durante esta terceira fase do processo de desenvolvimento, os proponentes do *ART* tiveram que apresentar os resultados relativos à primeira e terceira fase do processo de desenvolvimento, como se verá nesta secção (a 2ª fase do processo de desenvolvimento não é relevante por ter correspondido apenas a uma optimização específica dos momentos de *Zernike*).

▪ ***Angular-Radial Transform (ART) versus momentos de Zernike***

O descritor de forma baseado na transformada *ART* usa um parâmetro que pretende descrever formas simples e complexas 2D através de uma transformada de função complexa, como foi descrito nas secções 2.3.3.1 e 2.3.3.2. O descritor de forma proposto com base neste parâmetro apresenta a seguinte estrutura sintáctica [KiKi1999k]:

ART(){	Número de bits
ART_coefficient[k]	k*4
}	

A semântica deste descritor é a que se apresenta de seguida:

- ***ART_coefficient*** – especifica os k coeficientes da transformada *ART* que descrevem a forma do objecto.

Para demonstrar o seu desempenho, nomeadamente em relação ao descritor ‘desafiado’, e provisoriamente aceite na norma, o descritor *ART* teve que se submeter a um processo de avaliação que englobou os *core experiments CE-1* e *CE-2*. No contexto das condições anteriormente descritas para estes *core experiments*, os proponentes do descritor de forma baseado na transformada *ART* obtiveram os resultados parciais e totais, ilustrados na Tabela 5.18 para uma ordem angular e radial de 11 e 4, respectivamente e usando 60 coeficientes [KiKi1999g]. Os momentos de *Zernike* apresentados nesta tabela foram calculados considerando $n=14$ e $m=14$ e 64 coeficientes. Mais uma vez, os resultados obtidos pelos proponentes foram confirmados com o mecanismo de descrição e procura desenvolvido no âmbito desta tese.

Core Experiments			Momentos de Zernike		ART		Momentos de Zernike – resultados confirmados pelo mecanismo de descrição e procura	
CE-1	A	A1	98.130%	99.065%	98.650%	99.325%	98.100%	99.050%
		A2	100%		100%		100%	
	B		70.230%	70.190%	69.160%			
	C		94.500%	94.500%	95.500%			
	Média Total		87.932%	88.005%	87.903%			
	CE-2	A	A1	80.600%	84.005%	87.400%	86.830%	78.200%
A2			97.960%	98.370%		95.510%		
A3			98.620%	98.820%		97.840%		
A4			58.840%	62.730%		65.450%		
B		59.220%	64.990%	64.000%				
Média Total		71.613%	75.910%	74.125%				

Tabela 5.18 – Desempenho para os *core experiments* CE-1 e CE-2: momentos de *Zernike* versus transformada ART [KiKi2000]

Como se pode verificar através da Tabela 5.18, o desempenho do descritor *ART* comparativamente com o dos momentos de *Zernike* é efectivamente globalmente superior, tanto para o CE-1 como para o CE-2, justificando-se pelo menos a sua participação nos *core experiments*. Quanto aos resultados obtidos pelo mecanismo de descrição e procura para os momentos de *Zernike*, estes são ligeiramente inferiores no CE-1 e superiores para o CE-2; isto deve-se ao problema previamente identificado nas bases de dados de teste definidas e que provocaram alguns falsos resultados.

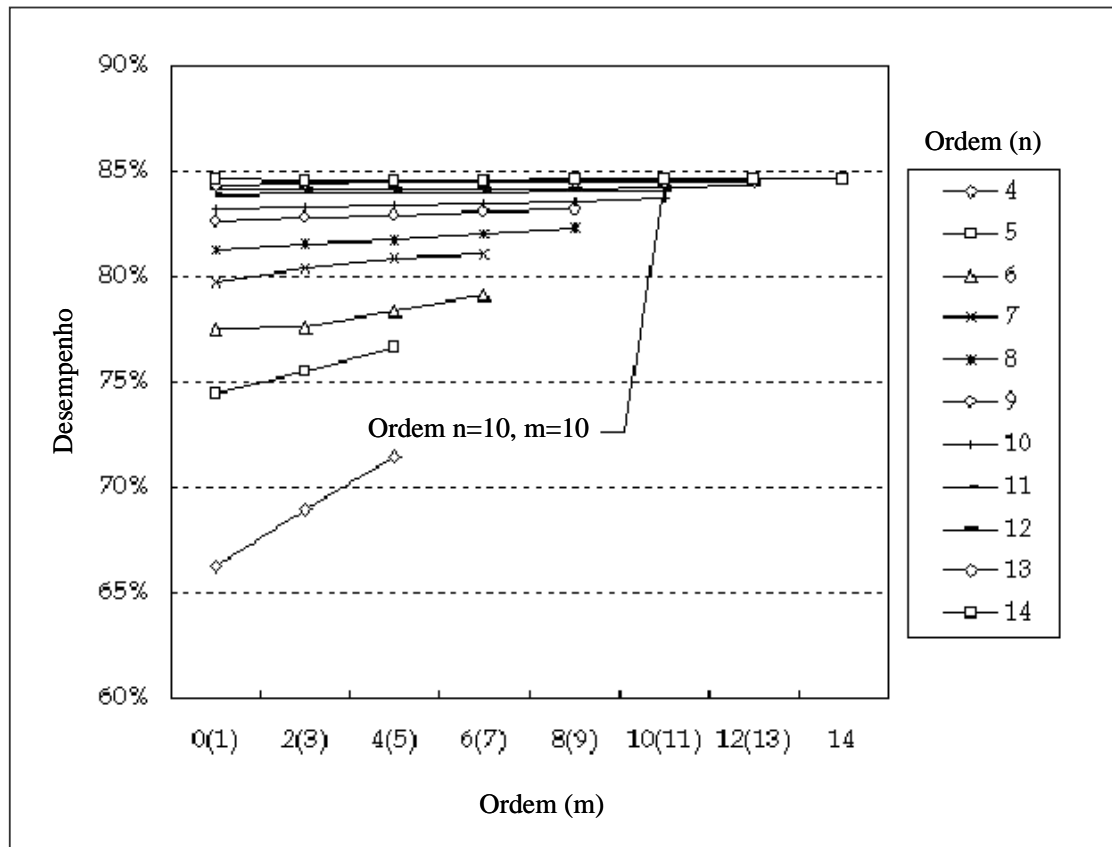
Uma vez apresentados os resultados do *ART* relativos à primeira fase, os resultados do desempenho obtidos para a terceira fase, *CE MS-3*, foram apresentados em simultâneo com os resultados para os momentos de *Zernike*, a fim de poderem ser mais facilmente comparados.

Nesta terceira fase que tem como objectivo otimizar a dimensão do descritor de forma baseado na região, os proponentes verificaram que dois factores influenciavam a dimensão dos descritores *ART* e momentos de *Zernike*: o número de momentos ou coeficientes usados e o nível de quantificação (*bits*) de cada momento ou coeficiente.

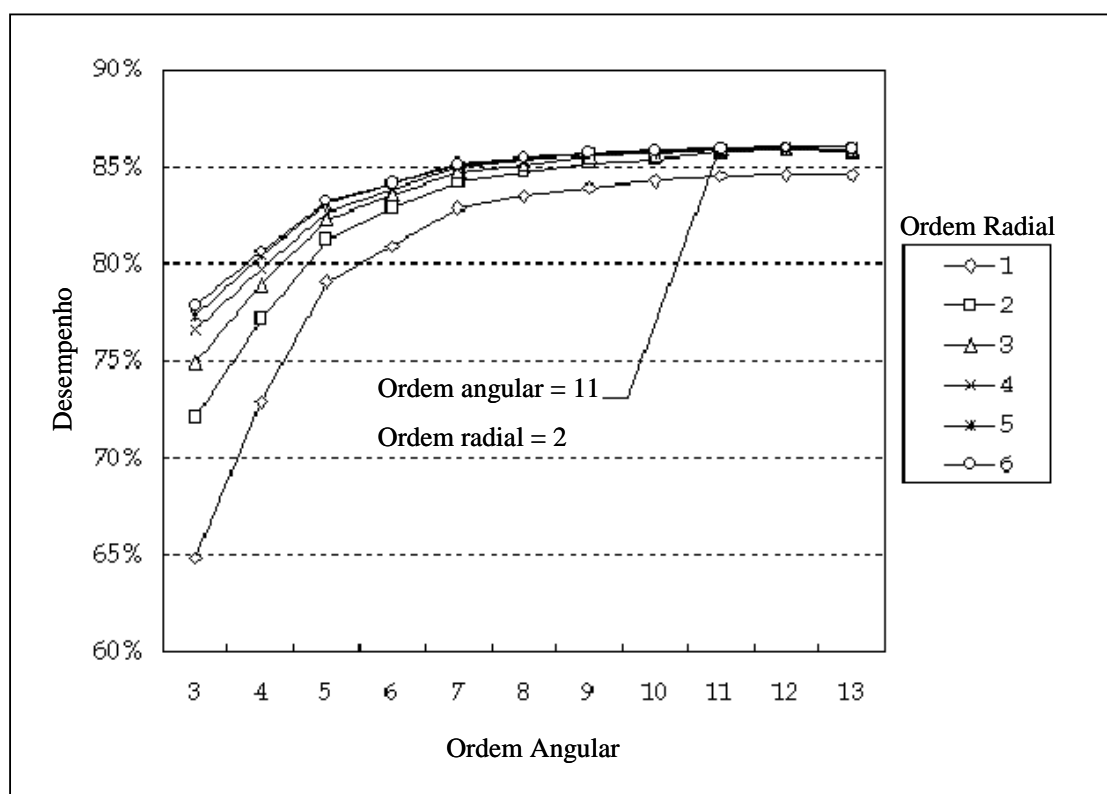
Assim, o desempenho de ambos os descritores em função da ordem dos momentos de *Zernike* e da ordem dos coeficientes *ART* pode ser visto nas Figura 5.22a) e b), respectivamente. Para os momentos de *Zernike*, o melhor desempenho em termos de procura foi conseguido para $n=14$ e $m=14$, como mostra a Figura 5.22a), resultando no uso de 63 momentos, excluindo o momento DC que é normalmente usado para normalizar os restantes momentos. O uso de um

maior número de momentos permitiria descrever a forma do objecto com maior detalhe e vice-versa. No entanto, a ordem recomendada é a de $n=10$ e $m=10$, o que implica o uso de apenas 35 momentos, uma queda de quase para metade com uma perda de desempenho de apenas cerca de 1% [KiKi2000l].

Para a transformada *ART*, o melhor desempenho em termos de procura foi conseguido para a ordem angular= 11 e radial= 4 , como mostra a Figura 5.22b); neste caso foram usados 59 coeficientes, excluindo o DC normalmente usado para a normalização dos restantes coeficientes. Estes valores garantem que não há perda de informação já que o desempenho do descritor é máximo ainda que há custo de um número de coeficientes elevado; no entanto, podem não ser estes valores os que garantem o melhor compromisso entre a quantidade de informação necessária para descrever a forma do objecto e o seu desempenho. Na verdade, os proponentes deste descritor recomendam ordens de valor angular= 11 e radial= 2 , para um número de coeficientes reduzido para 35 e uma perda de desempenho de apenas 0.24% [KiKi2000l].



a)



b)

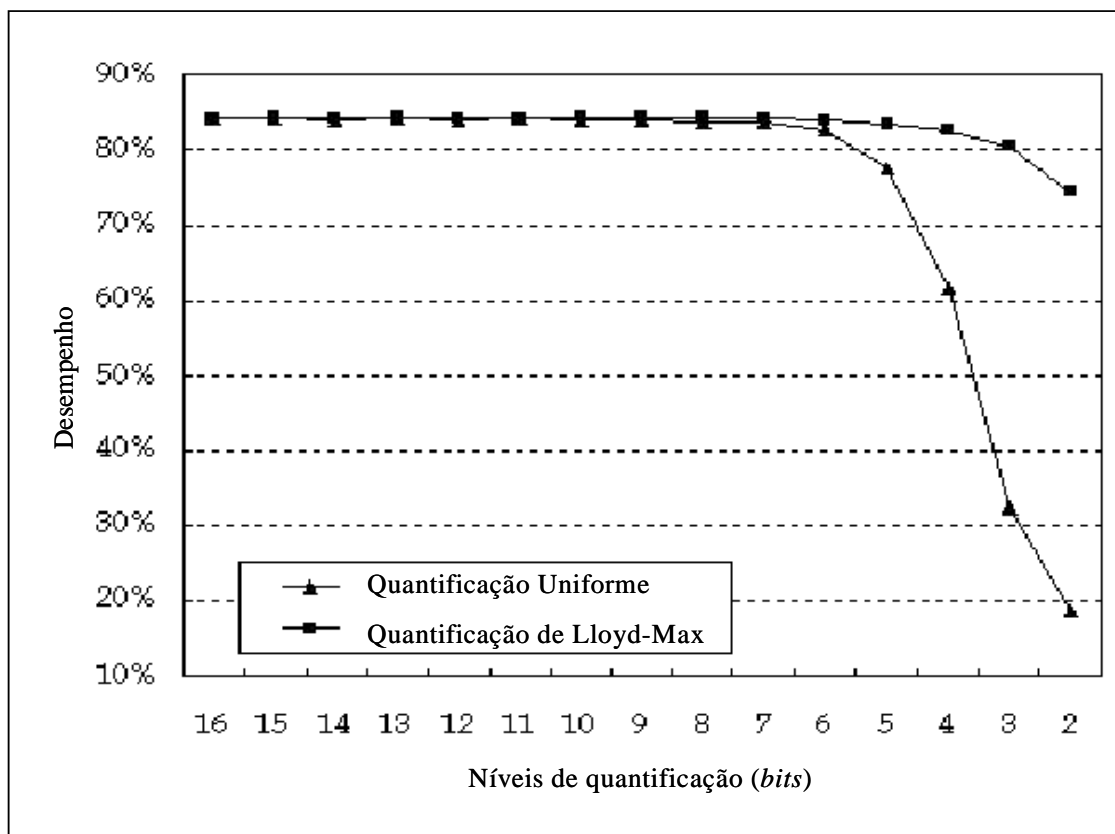
Figura 5.22 – Desempenho em função da ordem a) momentos de *Zernike*; e b) transformada *ART* [KiKi2000l]

Como os níveis de quantificação para cada momento ou coeficiente influenciam a dimensão do descritor, a metodologia usada para medir o seu efeito começou por estudar a densidade de probabilidade de ocorrência dos vários momentos ou coeficientes. Por conseguinte, foram extraídos 35 momentos ou coeficientes das imagens binárias nos *CE-1* e *CE-2* e obtido um conjunto de dados representativos da ocorrência de cada nível de quantificação dos momentos e dos coeficientes. Verificou-se então que a distribuição deste conjunto de dados se assemelhava a uma distribuição exponencial (ver expressão (5.15)),

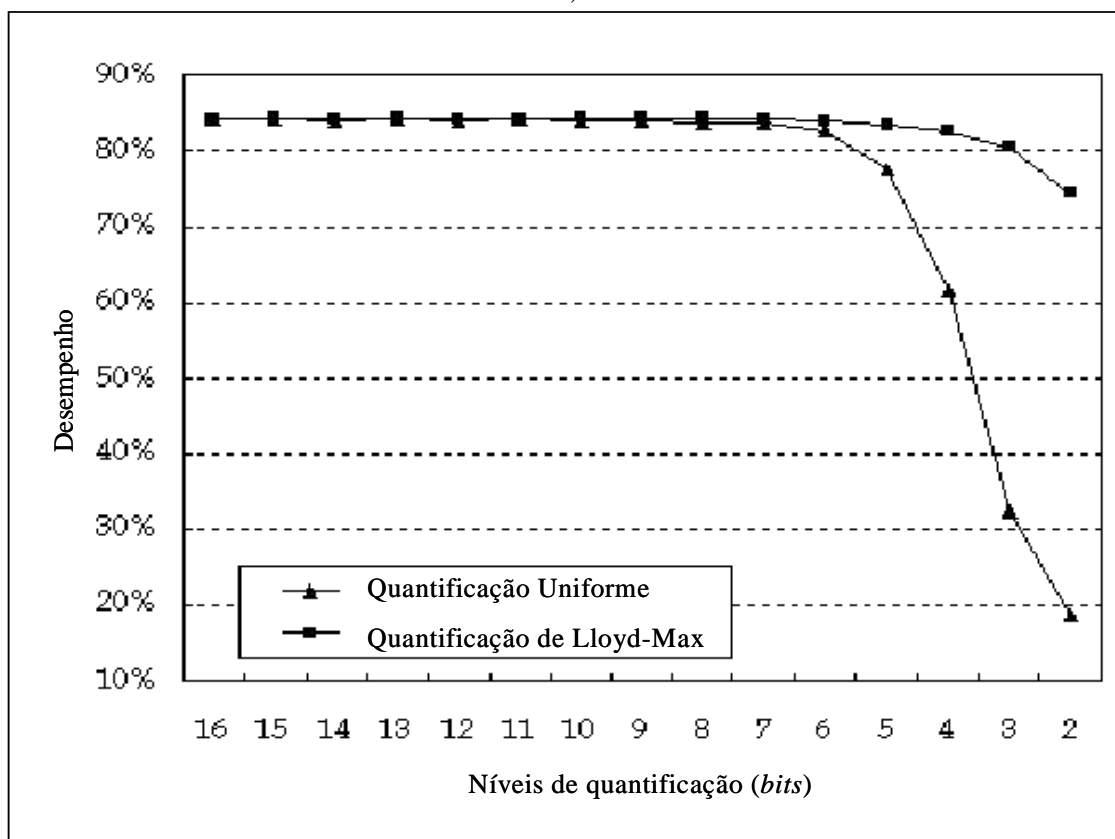
$$P(x) = \lambda e^{-\lambda x} \quad (5.15)$$

tendo a sua validade sido testada pelo teste de *Kolmogorov-Smirnov* [KiKi2000l].

Com base neste comportamento/distribuição, foram usados dois quantificadores para a implementação do *core experiment*: o quantificador uniforme e o quantificador de *Lloyd-Max* [KiKi2000l] de modo a definir-se o quantificador que oferece o melhor compromisso entre o desempenho do descritor e a sua dimensão. Para cada um destes quantificadores, variou-se o número de níveis de quantificação (*bits*) dos momentos e coeficientes, tendo-se obtido os desempenhos de procura apresentados nas Figura 5.23a) e b).



a)



b)

Figura 5.23 – Desempenho em função dos níveis de quantificação a) momentos de Zernike; e b) transformada ART [KiKi2000]

Com base nos resultados apresentados nos gráficos acima e na Tabela 5.19, os participantes neste *core experiment* propuseram como valores óptimos para os momentos de *Zernike*, 35 momentos codificados com 5 *bits* e para a transformada *ART*, 35 coeficientes codificados com 4 *bits*. O desempenho médio total obtido para cada um destes descritores (ver Tabela 5.20) com estes valores otimizados e para as bases de dados definidas pelos *CE-1* e *CE-2* decrescem -0.876% e -0.775% para os momentos de *Zernike*, respectivamente e -0.637% e -0.098% para o *ART*, respectivamente (ver Tabela 5.21). No entanto, face ao ganho conseguido relativamente à dimensão do descritor aceita-se a perda de desempenho considerada pouco significativa como um bom compromisso entre os dois elementos em análise.

	Momentos de <i>Zernike</i>		<i>ART</i>	
	Q. Uniforme	Q. Lloyd-Max	Q. Uniforme	Q. Lloyd-Max
Número de momentos/coeficientes	35	35	35	35
Número de <i>bits</i> para a quantificação dos momentos/coeficientes	7	5	5	4
Desempenho total para o conteúdo dos <i>CE-1</i> e <i>CE-2</i>	83.59%	83.44%	85.84%	85.44%
Número total de <i>bits</i> sem codificação entrópica	245	175	175	140
Número médio de <i>bits</i> com codificação de <i>Huffman</i>	140	–	134	–

Tabela 5.19 – Número de momentos/coeficientes e número de *bits* necessários para a codificação de cada momento de *Zernike* ou coeficiente *ART* [KiKi2000][Yama2000]

Relativamente aos resultados obtidos pelos participantes nas condições dos *core experiments CE-1* e *CE-2* e apresentados na Tabela 5.14, concluiu-se que o *ART* apresentava globalmente uma taxa de desempenho superior à dos momentos de *Zernike* tendo-se consequentemente adoptado o descritor *ART* e substituído os momentos de *Zernike* na versão *Committee Draft* da norma [MPEG7-CDVisual2000] e no *XM*. Mais uma vez, os resultados obtidos pelos proponentes do descritor *ART* foram confirmados através do mecanismo de descrição e procura desenvolvido no âmbito desta tese.

Core Experiments			Momentos de Zernike		ART		ART – resultados confirmados pelo mecanismo de descrição e procura	
CE-1	A	A1	97.860%	98.930%	98.570%	99.285%	98.100%	99.050%
		A2	100%		100%		100%	
	B		68.170%		68.340%		68.330%	
	C		94.500%		94.500%		92.000%	
	Média Total		87.200%		87.375%		86.460%	
CE-2	A	A1	79.000%	83.035%	85.080%	85.765%	83.400%	86.450%
		A2	97.960%		98.060%		96.330%	
		A3	97.930%		98.540%		98.510%	
		A4	57.250%		61.380%		67.560%	
	B		59.800%		66.070%		65.460%	
	Média Total		71.418%		75.916%		75.955%	

Tabela 5.20 – Desempenho dos descritores momentos de *Zernike* e *ART* após a quantificação com os valores otimizados [KiKi2000][Yama2000]

Core Experiments			Momentos de Zernike		ART	
CE-1	A	A1	-0.270%	-0.270%	-0.080%	-0.080%
		A2	0%		0%	
	B		-2.060%	-1.85%		
	C		0%	0%		
	Média Total		-0.876%	-0.637%		
CE-2	A	A1	-1.600%	-0.970%	-1.600%	-0.885%
		A2	0%		-0.310%	
		A3	-0.690%		-0.280%	
		A4	-1.590%		-1.350%	
	B		-0.580%	1.080%		
	Média Total		-0.775%	-0.098%		

Tabela 5.21 – Perdas de desempenho dos descritores momentos de Zernike e *ART* após optimização da sua dimensão

5.3.5 Quarta fase do processo de desenvolvimento

Na quarta e última fase do processo de desenvolvimento foi realizado o *core experiment MS-5* com o objectivo de melhorar o desempenho do descritor de forma baseado no contorno, CSS, podendo dividir-se esta melhoria em duas partes: a melhoria dos parâmetros usados no cálculo dos picos CSS antes da medida de semelhança e a melhoria dos limites de quantificação. Esta melhoria, sugerida pelos proponentes deste descritor, não alterou em nada a sintaxe e a dimensão do descritor, melhorando no entanto o seu desempenho.

A optimização do valor dos parâmetros usados no cálculo dos picos foi proposta uma vez que os valores usados alteram a altura dos picos a serem comparados. A expressão usada para o cálculo dos picos é a expressão (5.16):

$$y_{peak}(i) = k_1 * \left(\frac{y_{css}(i)}{N_{samples}^2} \right)^{k_2} \quad (5.16)$$

onde $y_{peak}(i)$ é a altura do pico i , $y_{css}(i)$ é o número de passagens do filtro binomial pelo núcleo (0.25,0.5,0.25), $N_{samples}$ é o número de pontos equidistantes usados no processo de filtragem e k_1 e k_2 são as variáveis a otimizar nesta fase de desenvolvimento do MPEG-7 [BoAt2000b].

Quanto à optimização dos limites de quantificação, estudaram-se unicamente os componentes *CurvatureVector* do descritor, linearmente quantificados entre 0 e o valor máximo respectivo. A proposta de modificação é mais precisamente ao *GlobalCurvatureVector* revendo-se os valores mínimos dos seus componentes, excentricidade e circularidade. Isto porque, a excentricidade, definida como a relação entre o raio máximo e o raio mínimo do objecto, tem como valor mínimo 1.0 e a circularidade um valor mínimo de 4π para um objecto com raio 1. Esta revisão dos valores mínimos de quantificação demonstra que apesar do contra-domínio abrangido ser menor, a cobertura de todo o domínio do *CurvatureVector* é mesmo assim verificada e que as componentes do elemento, a excentricidade e a circularidade, são definidas com maior resolução e sem qualquer custo adicional em termos de *bits*.

Com base nestas conclusões e nas condições definidas para este *core experiment*, os resultados de desempenho obtidos pelos proponentes após a optimização foram os que se apresentam na Tabela 5.22:

CE MS-5			CSS – antes da optimização (3ª fase do processo de desenvolvimento)		CSS – depois da optimização (4ª fase do processo de desenvolvimento)	
Partes	A	A1	87.930%	93.970%	90.830%	95.420%
		A2	100%		100%	
	B		76.310%	79.390%		
	C		96.000%	94.500%		
	Média Total (Bits)			88.760%	89.770%	

Tabela 5.22 – Desempenho do descritor de forma baseado no contorno, CSS, antes e depois da optimização em relação aos parâmetros globais do descritor e às transformadas efectuadas aos picos [Bobe2000e][BoAt2000b]

Tendo em conta os resultados obtidos, pode-se concluir que as melhorias de desempenho observadas foram relativas às partes A1 de 2.9%, realçando a sua robustez a mudanças de escala, e à parte B de 2.77%, realçando a sua robustez a procuras baseadas na semelhança perceptiva. Já a parte C perdeu em desempenho 1.5%, ainda que em termos gerais o desempenho do descritor tenha melhorado cerca de 0.9%.

5.4 Descritores de Forma Adoptados pelo MPEG-7

Após a avaliação e optimização em colaboração dos descritores de forma baseados no contorno e na região propostos ao MPEG-7, os resultados mostraram quais aqueles que oferecem melhor desempenho segundo os vários critérios definidos como relevantes. Neste contexto, foi escolhido para descritor de forma baseado no contorno da norma MPEG-7 o descritor de forma baseado na imagem CSS e como descritor de forma baseado na região o descritor de forma baseado na transformada ART.

As sub-secções que se seguem descrevem a sintaxe e a semântica destes descritores para as representações textual (MPEG-7 DDL) e binária (MPEG-7 BIM) tal como foram normalizadas pelo MPEG-7 no contexto dos *Committee Draft* (CD) aprovados em Outubro de 2000 [MPEG7-CD2000].

5.4.1 Descritor de forma baseado no contorno

O descritor de forma baseado no contorno descreve um contorno fechado e conexo 2D. O parâmetro de forma usado para a definição deste descritor foi a imagem CSS. As secções que seguem apresentam a especificação normativa deste descritor em termos da sintaxe e semântica das suas representações textual e binária com base no documento “*Visual Committee Draft*” editado pelo grupo de vídeo do MPEG [MPEG7-CD2000].

5.4.1.1 Representação textual

A representação textual conforme definida pelo grupo de vídeo do MPEG baseia-se na *Description Definition Language*, DDL [MPEG7-DDL2000]. A DDL tem como objectivo especificar esquemas de descrição, permitindo ainda criar novos esquemas de descrição a partir dos esquemas de descrição já existentes. A DDL proporciona assim maior flexibilidade e garante a longevidade da norma. Atendendo a que a linguagem de estruturação de texto *eXtensible Markup Language*, XML, é hoje largamente usada em múltiplos domínios de aplicação e ao facto de esta linguagem responder à maioria dos requisitos definidos para a DDL, decidiu o grupo MPEG usar o XML como linguagem base para definir a DDL. Significa isto que o MPEG poderá definir algumas extensões ao XML, específicas do MPEG-7, se isso for necessário para satisfazer requisitos não considerados pelo XML.

Deste modo, a sintaxe da representação DDL do descritor CSS é a que se apresenta de seguida [MPEG7-CDVisual2000]:

```
<element name="ContourShape">
  <complexType>
    <attribute name="PeakCount" type="unsigned6"/>
    <element name="GlobalCurvatureVector">
      <simpleType base="unsigned6" derivedBy="list">
        <length value="2"/>
      </simpleType>
    </element>
    <element name="PrototypeCurvatureVector" minOccurs="0" maxOccurs="1">
      <simpleType base="unsigned6" derivedBy="list">
        <length value="2"/>
      </simpleType>
    </element>
    <element name="HighestPeak" type="HighestPeakType"/>
    <element name="ContourPeak" minOccurs="1" maxOccurs="1">
      <complexType>
        <element name="xpeak" type="unsigned6"/>
        <element name="ypeak" type="unsigned3"/>
      </complexType>
    </element>
  </complexType>
</element>
```



```

    </complexType>
</element>
<simpleType name="HighestPeakType" base="nonNegativeInteger">
  <!--strictly speaking this should be enumeration of 0, 8, ..., 1016 -->
  <minInclusive value="0"/>
  <maxInclusive value="1023"/>
</simpleType>

```

onde [MPEG7-DDL2000]:

- **<element>** – define que a informação armazenada/guardada por este elemento é uma parte constituinte do elemento pai;
- **<attribute>** - define propriedades dos elementos;
- **<simpleType>** - define um elemento ou atributo;
- **<complexType>** - define um conjunto de elementos ou atributos;
- **<name>** - define o nome do elemento em questão;
- **<value>** - define o valor atribuído ao elemento;
- **<type>** - define os possíveis tipos de um elemento, por exemplo *unsigned6*, *unsigned3*, *nonNegativeInteger*;
- **<units>** - define a unidade a usar para o atributo *value*; e
- **<size>** - define o número de elementos de um descritor com número variável de elementos.

5.4.1.2 Representação binária

Uma vez que a representação textual não foi projectada para ser muito eficiente em termos do uso de recursos limitados de banda ou de armazenamento, o MPEG-7 definiu também uma representação binária para cada descritor com o objectivo de obter descrições binárias mais eficientes. Assim, o descritor de forma baseado no contorno, *CSS*, tem a representação binária que se segue [MPEG7-CD2000]:

ContourShape(){	Número de bits	Mnemónica
PeakCount	6	Uimsbf
GlobalCurvatureVector	2*6	Uimsbf
if (PeakCount == 1){		
PrototypeCurvatureVector	2*6	Uimsbf
}		
HighestPeak	7	Uimsbf
for (k=1; k<PeakCount; k++){		
Xpeak[k]	6	Uimsbf
Ypeak[k]	3	Uimsbf
}		
}		

A semântica de cada um destes componentes é [MPEG7-CD2000]:

- **PeakCount** – número de picos usados pela imagem CSS para definir a forma do objecto; caso a forma do objecto seja convexa, a imagem CSS não terá qualquer pico e esta componente do descritor toma o valor 0;
- **GlobalCurvatureVector** – parâmetros globais do contorno, nomeadamente a excentricidade e a circularidade; a circularidade é calculada através da expressão (5.17)

$$circularidade = \frac{perimetro^2}{\acute{area}} \quad (5.17)$$

e definida num intervalo entre [0,110] usando-se 6 *bits* com quantificação uniforme; a excentricidade é calculada através da expressão (5.18):

$$excentricidade = \sqrt{\frac{i_{20} + i_{02} + \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}{i_{20} + i_{02} - \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}} \quad (5.18)$$

onde $i_{02} = \sum (y - y_c)^2$, $i_{11} = \sum (x - x_c)(y - y_c)$ e $i_{20} = \sum (x - x_c)^2$ são calculados para cada ponto (x, y) dentro do contorno e para o centro de massa (x_c, y_c) . A excentricidade é definida num intervalo entre [0,10] e representada usando 6 *bits* com quantificação uniforme.

- **PrototypeCurvatureVector** – excentricidade e circularidade da curvatura protótipo; a curvatura protótipo é definida como a curva alisada da forma do objecto através de filtragens, sendo essa a curva descrita pelo CSS;
- **HighestPeak** – parâmetros dos filtros correspondentes ao pico máximo, calculados através da expressão (5.19):

$$ypeak[0] = 3.65 * \left(\frac{y_{css}(0)}{Nsamples^2} \right)^{0.46} \quad (5.19)$$

onde $y_{css}[0]$ é o número de passagens do filtro pelo núcleo (0.25,0.5,0.25) correspondente ao pico máximo e $Nsamples$ é o número de pontos equidistantes usados no processo de filtragem. Este parâmetro está definido no intervalo de [0,2.4691] e é representado por 7 *bits* com quantificação uniforme;

- **Xpeak, Ypeak** – valores tomados pelos restantes picos da imagem CSS e normalizados em função do **HighestPeak**; os picos encontram-se armazenados nos vectores por ordem decrescente da altura dos picos; a semântica exacta destes dois componentes é: $Xpeak[k]$ representa a distância normalizada ao longo do contorno entre o ponto do **HighestPeak** e o ponto k , encontrando-se definido num intervalo de [0,1] e representado por 6 *bits*; $Ypeak[k]$ representa a altura do pico k e é calculado através da expressão (5.20):

$$y_{peak}(k) = 3.65 * \left(\frac{y_{css}(k)}{N_{samples}^2} \right)^{0.46} \quad (5.20)$$

onde $y_{css}[k]$ é o número de passagens do filtro binomial pelo núcleo (0.25,0.5,0.25) correspondente ao pico k e $N_{samples}$ é o número de pontos equidistantes usados no processo de filtragem. Este componente do descritor encontra-se definido num intervalo entre $[0, y_{peak}[k-1]]$ e é representado por 3 *bits*.

5.4.2 Descritor de forma baseado na região

O descritor de forma baseado na região descreve qualquer forma 2D numa imagem ou numa sequência segmentada de vídeo mesmo que esta contenha mais do que uma região (forma não conexa). O parâmetro de forma usado para a definição deste descritor foi a transformada *ART*. As secções que se seguem apresentam a especificação normativa deste descritor em termos da sintaxe e semântica das representações *DDL* e binária com base no documento “*Visual Committee Draft*” editado pelo grupo de vídeo do MPEG [MPEG7-CDVisual2000].

5.4.2.1 Representação textual

À semelhança do que foi dito na secção 5.4.1.1 sobre a representação textual, o documento *Committee Draft* para a parte visual apresenta a seguinte representação *DDL* para o descritor de forma baseado na região adoptado pelo MPEG-7 [MPEG7-CDVisual2000]:

```
<element name="RegionShape">
  <complexType>
    <element name="ArtDE">
      <simpleType base="unsigned4" derivedBy="list">
        <length value="35"/>
      </simpleType>
    </element>
  </complexType>
</element>
```

5.4.2.2 Representação binária

A representação binária do descritor *ART* é especificada no documento *Committee Draft* para a parte visual do seguinte modo [MPEG7-CDVisual2000]:

RegionShape(){	Número de <i>bits</i>	Mnemónica
For(k=0; k<35; k++) {		
ArtDE[k]	4	uimsbf
}		
}		

A semântica do seu único componente é a seguinte [MPEG7-CD2000]:

- *ArtDE[]* – é um vector de 35 coeficientes normalizados e quantificados; a quantificação destes coeficientes é feita segundo a Tabela 5.23;

<i>ArtDE</i>	Contra-Domínio	Semântica (valor representativo)
0000	0.000000000 ~ 0.003585473	0.001763817
0001	0.003585473 ~ 0.007418411	0.005468893
0010	0.007418411 ~ 0.011535520	0.009438835
0011	0.011535520 ~ 0.015982337	0.013714449
0100	0.015982337 ~ 0.020816302	0.018346760
0101	0.020816302 ~ 0.026111312	0.023400748
0110	0.026111312 ~ 0.031964674	0.028960940
0111	0.031964674 ~ 0.038508176	0.035140141
1000	0.038508176 ~ 0.045926586	0.042093649
1001	0.045926586 ~ 0.054490513	0.050043696
1010	0.054490513 ~ 0.064619488	0.059324478
1011	0.064619488 ~ 0.077016351	0.070472849
1100	0.077016351 ~ 0.092998687	0.084434761
1101	0.092998687 ~ 0.115524524	0.103127662
1110	0.115524524 ~ 0.154032694	0.131506859
1111	0.154032694 ~ 1.000000000	0.192540857

Tabela 5.23 – Tabela de quantificação usada para os coeficientes *ART* [MPEG7-VWD2000b][MPEG7-CDVisual2000]

5.5 Conclusões

A necessidade de incluir nas normas ferramentas com elevado desempenho e que possam responder às necessidades de vários domínios de aplicação requer a definição de uma metodologia de avaliação e desenvolvimento sólida e credível. O grupo MPEG adoptou uma metodologia de trabalho baseada em seis etapas. As seis etapas desta metodologia foram definidas e analisadas no início deste capítulo, tendo-se implementado a etapa 5 onde se avaliam as propostas submetidas à consideração do grupo de normalização.

Para além do exaustivo trabalho de pesquisa realizado neste capítulo, foi instalado e desenvolvido *software* de forma a poder replicar o processo de desenvolvimento do MPEG-7 em termos de descritores de forma, a fim de poder confirmar o seu desempenho nomeadamente comparativamente a outros descritores. Neste contexto, implementou-se também um descritor de forma baseado no contorno, os ângulos de curvatura do contorno,

uma vez que este descritor não participou no processo de desenvolvimento do MPEG-7 depois de ter sido submetido para avaliação.

Para os parâmetros de forma baseados no contorno, observou-se que tanto o parâmetro baseado no CSS como o parâmetro baseado na transformada *wavelet* apresentam, logo na primeira fase do processo de desenvolvimento, desempenhos médios elevados relativamente ao parâmetro baseado nos ângulos de curvatura do contorno. Este facto poderá dever-se ao facto de os primeiros serem parâmetros multi-nível permitindo a descrição da forma do objecto com vários níveis de detalhe enquanto que o último mantém constante o tipo de informação a enviar para qualquer forma (por exemplo, envia sempre 64 ângulos de curvatura). Ao enviar-se um número fixo de ângulos, o método de descrição encontra-se de imediato limitado, restringindo o envio de mais informação de descrição quando a forma assim o exige e exagerando no caso de uma forma com necessidade de menos ângulos, enviando informação de descrição exagerada face ao seu perímetro e podendo inclusive estar-se a enviar ruído como descrição. Uma vez que os primeiros dois descritores apresentam desempenhos elevados fará sentido investir-se, numa fase posterior, no desenvolvimento e optimização do descritor baseado nos ângulos de curvatura do contorno, pois este aparenta ter sido menos ‘trabalhado’ em relação aos outros dois.

Para os parâmetros de forma baseados na região, observou-se que os três parâmetros estudados apresentam taxas de desempenho elevadas. No entanto, o *MLEV* mostrou ser o mais fraco relativamente às condições avaliadas pelo *core experiment CE-2*, nomeadamente em termos da sua insensibilidade a rotações, mudanças de escala e de perspectiva.

Com base nos resultados obtidos para os vários descritores propostos, os especialistas MPEG concluíram que o descritor de forma baseado no contorno com melhor desempenho é o descritor baseado na imagem CSS e que o descritor de forma baseado na região com melhor desempenho é o descritor baseado na transformada ART. Ambos os descritores apresentaram um bom desempenho relativamente ao *core experiment CE-1*, cerca de 90% e 87.4%, respectivamente, e o desempenho global do ART relativamente ao *core experiment CE-2* foi de 76%. Os descritores mostraram-se insensíveis à rotação, à mudança de escala e a pequenas deformações não-rígidas, como era pretendido.

Quanto à optimização da dimensão dos descritores, esta provocou uma perda de desempenho para ambos os descritores, na maioria das partes de cada *core experiment*; no entanto, estas perdas não foram significativas em termos globais, resultando num bom compromisso entre o desempenho do descritor e a sua dimensão.

Com base nos resultados obtidos, pelos proponentes e pelo mecanismo de descrição e procura implementado no âmbito desta tese, e ainda as conclusões da avaliação, os melhores descritores de forma adoptados pelo MPEG-7, integrados no modelo de trabalho experimental, o *XM*, e no documento da norma denominado como *Visual Committee Draft (CD)* foram o CSS e o ART.

Apesar de nesta altura serem estes os descritores de forma adoptados pelo MPEG-7, nada impede que novas propostas venham a ser feitas no futuro próximo para melhorar ainda estes descritores antes da versão final da norma ou num futuro menos próximo que novos descritores de forma venham a ser acrescentados à norma MPEG-7 se se comprovar que o seu desempenho e funcionalidades são definitivamente melhores do que os dos descritores existentes. Após esta análise das vantagens/forças e desvantagens/fraquezas dos descritores testados e escolhidos pelo MPEG-7, passar-se-á no capítulo seguinte ao desenvolvimento de

um dos descritores mais prometedores mas com pior desempenho até ao momento, com vista a avaliar até que ponto o seu desempenho se pode aproximar do desempenho dos descritores escolhidos pelo MPEG-7: os ângulos de curvatura do contorno.

Capítulo 6

Desenvolvimentos ao descritor ângulos de curvatura do contorno

A aplicação do grande princípio estruturante das normas MPEG que preconiza a especificação de um número mínimo de ferramentas - *uma ferramenta por cada funcionalidade oferecida, sem redundância* – que, ao serem combinadas de acordo com os requisitos do domínio de aplicação em questão, garantem a máxima interoperabilidade enquanto minimizam a complexidade e o custo dos sistemas envolvidos [PeKo1998], resultou para o caso da característica de forma, no contexto da norma MPEG-7, na adoção de dois descritores com diferentes funcionalidades, adotados após avaliação exaustiva. Estes dois descritores distinguem-se pelas propriedades que descrevem relativamente à forma, contorno fechado ou região, e são, respectivamente, o *CSS* e o *ART*, apresentados e avaliados nos capítulos 2 e 5 desta tese. Cada um destes descritores passou por uma fase de avaliação e de desenvolvimento, a fim de garantir os requisitos mínimos de interoperabilidade, complexidade e desempenho, antes de serem finalmente aprovados pelo grupo MPEG. No capítulo 5, apresentaram-se, analisaram-se e compararam-se os resultados obtidos para o desempenho dos vários candidatos a descritores, e que resultaram na integração dos descritores *CSS* e *ART*, no modelo de trabalho *XM* e na própria norma MPEG-7, logo na sua fase inicial denominada por *Committee Draft (CD)* (ver capítulo 1).

Tendo-se concluído, no capítulo 5, que o descritor de forma baseado nos ângulos de curvatura do contorno (*ACC*), submetido inicialmente ao MPEG [MPEG7-Dev1999] e implementado no âmbito desta tese, apresentava propriedades interessantes mas não estava ainda suficientemente desenvolvido, pretende-se neste capítulo apresentar o trabalho realizado no sentido de melhorar este descritor, se possível levando-o a um desempenho semelhante aos

descritores incluídos na norma MPEG-7. Após a sua implementação de acordo com o método de extracção sugerido pelos seus proponentes [IBMRe1999], verificou-se que o seu desempenho médio se encontrava longe dos desempenhos obtidos pelos seus descritores ‘adversários’, ou seja o CSS e a transformada *wavelet*, tal como se mostrou na secção 5.3.2.1. Consequentemente, e por este descritor apresentar claras ‘debilidades’ relativamente aos outros dois, optou-se por investir no seu desenvolvimento, tendo sempre em mente as propriedades dos outros descritores que os levaram a superar as debilidades que o descritor baseado nos ângulos de curvatura do contorno ainda apresenta.

Tendo em conta as boas propriedades dos vários descritores estudados ao longo de todo o processo, ir-se-á tentar melhorar o desempenho do descritor baseado nos ângulos de curvatura do contorno através de várias acções, nomeadamente:

- Melhoria das suas propriedades relativamente à insensibilidade a transformações geométricas
- Inclusão de parâmetros complementares de descrição da forma que possam auxiliar o descritor principal
- Optimização dos valores de algumas das suas componentes,
- Melhoria da medida de semelhança
- Optimização da sua dimensão.

Nas secções que se seguem, serão apresentados sequencialmente as várias melhorias propostas e os correspondentes resultados alcançados. Algumas das melhorias implicarão alterações na sintaxe e semântica do descritor. Todos os resultados apresentados serão obtidos em condições relevantes e adequadas, nomeadamente usando o material de teste do MPEG-7 e seguindo o mais de perto possível as condições dos *core experiments* MPEG-7 apresentadas na secção 5.2.

6.1 Optimização inicial da dimensão do descritor ACC

Após a implementação do descritor baseado nos ângulos de curvatura do contorno para a primeira fase do processo de desenvolvimento do MPEG-7 e da avaliação do seu desempenho no capítulo 5, verificou-se que este se encontrava bastante longe do desempenho dos outros descritores que com ele foram comparados (ver Tabela 6.1), ou seja o CSS e a transformada *wavelet*, nomeadamente em termos de desempenho de procura (65.643% versus 89.770% e 84.503% para o *core experiment CE-1*).

CE-1			CSS		Wavelet		ACC	
Partes	A	A1	90.830%	95.420%	88.040%	97.460%	66.667%	79.288%
		A2	100%		92.750%		91.910%	
	B		79.390%	67.760%	54.140%			
	C		94.500%	93.000%	63.500%			
	Média Total			89.770%	84.503%	65.643%		

Tabela 6.1 – Desempenho de procura obtido nesta tese para os descritores *CSS*, transformada *wavelet* e ângulos de curvatura do contorno (ACC) nas condições do *core experiment CE-1*

Ainda no capítulo 5, foram apresentadas as sintaxe e semântica dos descritores que foram comparados com o descritor ACC. No entanto, os proponentes do ACC não propuseram a sintaxe e semântica completas para o descritor pois faltou-lhes especificar o número de *bits* usado para cada componente do descritor. Consequentemente, durante a implementação do descritor ACC para a obtenção dos resultados apresentados no capítulo 5 foi especificado, no contexto do trabalho desta tese, o número de *bits* para cada componente do descritor. Assim, a sintaxe então proposta para o descritor ACC foi:

TurningAngles(){	Número de bits
Number_angles	7
Moments_vector[2]	9*2
Distance_vector[4]	9*4
Index_vector[4]	7*4
for (x=0; x<4; x++){	
for (k=0; k<Number_angles; k++){	
Turning_angle[x][k]	9*4*Number_angles
}	
}	
}	

Com base nesta sintaxe, pode-se então comparar a dimensão, em termos de número médio de *bits* por cada descrição de forma, obtida para os descritores *CSS* (final, otimizado) e ACC, como mostra a Tabela 6.2:

CE-1			CSS (bits por descrição)		ACC (bits por descrição)	
Partes	A	A1	111	111.5	2392	2392
		A2	112		2392	
	B		114	2389		
	C		117	2390.7		
	Média Total (bits)			114	2390.6	

Tabela 6.2 – Número médio de *bits* usado por cada descrição usando os descritores CSS e ACC

Tendo em conta as dimensões apresentadas na Tabela 6.2 foi possível constatar que, nesta fase, as dimensões dos dois descritores são muito diferentes, uma vez que, em média, uma descrição gerada pelo descritor ACC tem uma dimensão 21 vezes maior do que a gerada pelo descritor CSS. No entanto, é preciso ter em conta que os resultados apresentados para o descritor CSS são valores otimizados, o que não é verdade para o caso do descritor ACC uma vez que o número de *bits* especificado para cada componente do descritor garante antes de mais que não há perdas de precisão e de desempenho devido à codificação binária.

Com base nas conclusões acima extraídas a partir dos valores apresentados, a secção seguinte tem como principal objectivo diminuir a dimensão do descritor ACC através de uma primeira redefinição da sua sintaxe, tomando como referência a sintaxe inicialmente implementada no âmbito desta tese, obtida da proposta feita pelos seus proponentes ao MPEG-7.

6.1.1 Eliminação das componentes não essenciais do descritor ACC

A redefinição do descritor ACC em termos de sintaxe surge na sequência da análise e avaliação feita ao descritor no capítulo 5 que levou à conclusão que algumas das suas componentes são claramente supérfluas não contribuindo com qualquer funcionalidade ou para qualquer melhoria do desempenho do descritor. Consequentemente, optou-se por ‘limpar’ a sintaxe do descritor ACC, eliminando três das suas componentes não essenciais, nomeadamente *Moments_vector*, *Distance_vector* e *Index_vector* uma vez que nenhuma das três tem qualquer impacto no bom desempenho do descritor já que não são de nenhum modo contabilizadas na medida de semelhança proposta pelos proponentes e adoptada por esta tese.

A componente *Número_Ângulos*, apesar de o seu valor ser fixo e pré-determinado e logo não necessitar em princípio de ser transmitido, não será para já eliminada de modo a permitir futuras melhorias ao descritor, nomeadamente a descrição de formas com um número variável de ângulos de contorno conforme a sua complexidade. A complexidade do contorno está associada ao número e intensidade das variações ao longo do contorno.

Eliminando-se as três componentes acima referidas, a nova sintaxe proposta nesta secção para o descritor ACC e que optimiza um pouco a sua dimensão é a seguinte:

Ângulos_Curvatura()	Número de bits
Número_Ângulos	7
for (x=0; x<4; x++){	
for (k=0; k< Número_Ângulos; k++){	
Ângulos_Curvatura[x][k]	9*4*Número_Ângulos
}	
}	
}	

A semântica para cada uma das componentes do descritor ACC é a seguinte:

- **Número_Ângulos** – Especifica o número de ângulos (64 por enquanto) usados para descrever a forma do objecto em questão usando 7 bits; e
- **Ângulos_Curvatura** – Define o conjunto dos ângulos de curvatura do contorno do objecto usando 9 bits por ângulo; note-se que são enviados 4 conjuntos de ângulos, um por cada ponto de começo, tal como definido na secção 5.3.2.1

6.1.2 Análise da dimensão do descritor ACC

Uma vez feita esta alteração à sintaxe do descritor, verificou-se (como se esperava) que o seu desempenho de procura em nada se alterou, reduzindo-se apenas a dimensão do descritor, tal como pretendido. A Tabela 6.3 apresenta a nova dimensão, em bits médios por descrição, para o descritor ACC.

CE-1			ACC (bits médios por descrição antes da optimização)		ACC (bits médios por descrição após a optimização)	
Partes	A	A1	2392	2392	2310	2310
		A2	2392		2310	
	B		2389		2307	
	C		2390.7		2308	
	Média Total (bits)			2390.6		2308.3

Tabela 6.3 – Número médio de bits por descrição antes e após a optimização da sintaxe do descritor ACC proposta nesta secção

Como se pode constatar a partir dos resultados apresentados na Tabela 6.3, a alteração proposta à sintaxe do descritor permitiu diminuir a dimensão do descritor em cerca de 3.5%. Apesar da dimensão média alcançada não ser ainda um valor ao nível do descritor CSS,

considera-se este ganho em termos de dimensão um primeiro passo para obter uma sintaxe mais optimizada, adoptando-se esta alteração ao descritor para o resto deste capítulo.

6.2 Melhoria do desempenho de procura e computacional através da inclusão de parâmetros geométricos

A alteração proposta nesta secção tem como objectivo melhorar o desempenho e a rapidez da procura ‘filtrando’ à partida grande parte das formas descritas disponíveis significativamente diferentes da forma fornecida como exemplo ao mecanismo de procura. Deste modo, passarão à etapa seguinte da procura apenas as formas descritas disponíveis cujo grau de semelhança com a forma fornecida ao mecanismo de procura seja elevado, acelerando-se assim o processo de procura e eliminando alguns falsos positivos (ou seja imagens que são dadas como semelhantes mas não o são efectivamente). Na prática, é como determinar o conjunto de imagens semelhantes em termos de forma usando três conceitos (parâmetros) de semelhança diferentes em vez de apenas um e logo compensando algumas ‘fraquezas’ de um com os outros. Contudo, para que as ‘fraquezas’ dos descritores que fazem a filtragem inicial não determinem demasiado o desempenho final é necessário que os critérios de filtragem não sejam exageradamente apertados.

À semelhança do que foi feito para o descritor de forma baseado no contorno do MPEG-7, o descritor CSS, propõe-se aqui que a filtragem das formas significativamente diferentes seja feita usando dois parâmetros geométricos, a excentricidade e a circularidade, definidos no capítulo 2 através das expressões 2.31 e 2.39. A inclusão destes dois parâmetros geométricos no descritor ACC vai alterar a sua sintaxe e semântica.

6.2.1 Redefinição da sintaxe e semântica do descritor ACC

A alteração aqui proposta ao descritor ACC acrescenta-lhe duas novas componentes, ambas representando de algum modo a forma a ser descrita: a excentricidade e a circularidade. O número de *bits* proposto para cada uma destas novas componentes do descritor ACC é o mesmo número de *bits* que já é usado pelo descritor CSS uma vez que a dimensão destas componentes já foi então optimizada. Deste modo, o novo descritor ACC proposto nesta secção passa a ter a seguinte sintaxe:

AngulosCurvatura(){	Número de <i>bits</i>
Número_Ângulos	7
Excentricidade	6
Circularidade	6
For (x=0; x<4; x++){	
For (k=0; k<Número_Ângulos; k++){	
Ângulo_Curvatura[x][k]	9*4*Número_Ângulos
}	
}	
}	

A semântica das novas componentes é a seguinte:

- **Excentricidade** – Exprime as propriedades de alongamento do contorno do objecto. Uma vez que esta componente foi definida à semelhança da componente correspondente no descritor CSS, a expressão usada para o cálculo da excentricidade será novamente a expressão 6.1 ou seja,

$$excentricidade = \sqrt{\frac{i_{20} + i_{02} + \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}{i_{20} + i_{02} - \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}} \quad (6.1)$$

onde $i_{02} = \sum (y - y_c)^2$, $i_{11} = \sum (x - x_c)(y - y_c)$ e $i_{20} = \sum (x - x_c)^2$ são calculados para cada ponto (x, y) do objecto e para o centróide (x_c, y_c) (para o centróide ver expressões 2.41 e 2.42). A excentricidade é definida num intervalo entre $[0, 10]$ e representada usando 6 *bits*, tal como para o descritor CSS;

- **Circularidade** – Exprime as propriedades de circularidade do contorno do objecto, ou seja quão próximo este está de uma circunferência. A expressão usada para definir esta nova componente do descritor ACC é a expressão 2.31 apresentada no capítulo 2, idêntica à usada pelo descritor CSS mas normalizada para uma circunferência de raio unitário,

$$circularidade = \frac{4\pi \text{Área}}{\text{Perímetro}^2} \quad (6.2)$$

A circularidade é definida num intervalo $[0, 1]$ e representada usando 6 *bits*, tal como para o descritor CSS.

Com base nesta nova sintaxe e semântica do descritor, espera-se que a dimensão do descritor aumente (inevitavelmente) um pouco. A Tabela 6.4 apresenta as novas dimensões parciais e totais do descritor ACC:

CE-1			ACC (bits por descrição antes da inclusão)		ACC (bits por descrição após a inclusão)	
Partes	A	A1	2310	2310	2322	2322
		A2	2310		2322	
	B		2307		2329	
	C		2308		2320	
	Média Total (bits)			2308.3		2323.7

Tabela 6.4 – Número médio de *bits* por descrição antes e após a melhoria do descritor ACC através da inclusão dos parâmetros geométricos, excentricidade e circularidade

Os valores apresentados na Tabela 6.4 confirmam a afirmação relativamente ao aumento da dimensão do descritor, em cerca de 0.5%, considerando-se no entanto que o aumento não é significativo face às vantagens em termos de desempenho de procura e computacional (rapidez da procura).

6.2.2 Definição da nova medida de semelhança

Para que esta nova versão do descritor ACC pudesse ser analisada e avaliada foi necessário modificar o código dos módulos de descrição e procura do *XM*. No módulo de descrição foi necessário implementar duas funções que calculassem os parâmetros geométricos incluídos no descritor, a excentricidade e a circularidade. Por outro lado, para o módulo de procura teve de ser definida e implementada uma nova medida de semelhança, considerando também os parâmetros geométricos agora usados que devem, inicialmente, filtrar as imagens antes de as passar ao descritor ACC propriamente dito.

A definição e implementação da nova medida de semelhança foi feita de dois modos alternativos:

i. Filtragem inicial com base na excentricidade e circularidade

Para a eliminação das formas significativamente diferentes da forma fornecida como exemplo ao mecanismo de procura, é primeiro calculada uma medida de semelhança entre as formas disponíveis e a forma fornecida ao mecanismo de procura, baseada na excentricidade e circularidade, através das seguintes expressões:

$$\frac{|E_I - E_B|}{\max(E_I, E_B)} \leq E_L \quad e \quad \frac{|C_I - C_B|}{\max(C_I, C_B)} \leq C_L \quad (6.3)$$

onde E_B e E_I , e C_B e C_I são os valores de excentricidade e circularidade para as descrições disponíveis nas bases de dados e aquela fornecida como exemplo ao mecanismo de procura, respectivamente. E_L e C_L são os respectivos limiares de filtragem das descrições em base de dados.

A filtragem das descrições é determinada através dos dois limiares, E_L e C_L , com valores iguais aos propostos para o descritor *CSS*, 0.9 e 0.8, respectivamente [Bobe1999][MPEG-7XM1999]. Caso ambas as medidas de semelhança de excentricidade e circularidade de uma dada forma na base de dados sejam inferiores aos respectivos limiares, então essa forma passará de imediato à fase seguinte. Caso contrário, essa forma será automaticamente excluída do processo de procura. O uso de valores para os limiares iguais aos usados pelo descritor *CSS* deve-se exclusiva e unicamente ao facto de as condições de teste serem as mesmas para a avaliação dos dois descritores; para além disso, estes limiares dependem apenas dos parâmetros de filtragem usados o que significa que mantendo-se estes não há qualquer razão para alterar os limiares.

Neste primeiro modo de integração dos parâmetros geométricos para a filtragem das descrições permitiu-se apenas que estes determinassem a filtragem, não tendo qualquer impacto na medida de semelhança final, D_{final} , ou seja a própria distância de *Minkowsky* que tem vindo a ser usada até aqui:

$$D_{final} = D_{\text{ânguloscurvatura}} = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^{\text{Número-Ângulos}} |A_{B_c}(t) - A_{I_c}(t)| \right\} \quad (6.4)$$

onde A_B e A_I são os vectores com os ângulos de curvatura do contorno para as descrições de forma disponíveis e para a forma exemplo fornecida ao mecanismo de procura, respectivamente, $Número_Ângulos$ é o número de ângulos usado nas descrições, t é o índice de cada ângulo no vector e c é o índice de cada combinação de vectores de ângulos.

ii. **Medida de semelhança tendo em conta os parâmetros de excentricidade e circularidade**

Numa segunda fase da integração dos parâmetros geométricos, definiu-se uma nova medida de semelhança levando também em conta as novas componentes geométricas do descritor ACC. A medida de semelhança final será então uma medida conjunta, composta por $D_{ânguloscurvatura}$ e medidas de semelhança associadas aos parâmetros geométricos:

$$D_{final} = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^{Número_Ângulos} |A_{B_c}(t) - A_{I_c}(t)| + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.5)$$

onde w_1 e w_2 são parâmetros associados ao peso, i.e. importância, atribuído a cada um dos parâmetros geométricos em questão. Também aqui, os valores propostos para w_1 (0.7) e w_2 (0.8) são os mesmos que foram adoptados para o descritor CSS [Bobe1999][MPEG-7XM1999].

6.2.3 Análise do desempenho do descritor

Após a implementação dos parâmetros geométricos e depois de efectuadas as alterações necessárias ao descritor ACC, apresenta-se e analisa-se aqui o desempenho de procura do descritor, para os dois casos definidos acima em termos do uso dos parâmetros geométricos no processo de procura.

Os resultados apresentados na Tabela 6.5 correspondem aos primeiro e segundo caso acima definidos ou seja ao caso onde os parâmetros geométricos determinam a filtragem mas não têm qualquer influência na medida de semelhança final e ao caso onde os parâmetros geométricos determinam a filtragem mas também influenciam a medida de semelhança final, respectivamente.

	A			B	C	Desempenho Médio Total
	A1	A2	A			
Sem filtragem	66.67%	91.91%	79.29%	54.14%	63.50%	65.64%
Com filtragem mas sem influência na medida de semelhança	63.81%	91.91%	77.86%	57.06%	63.50%	66.14%
Com filtragem e com influência na medida de semelhança	64.76%	91.67%	78.22%	57.41%	63.50%	66.38%

Tabela 6.5 – Desempenho de procura do descritor ACC antes e depois da filtragem baseada nos parâmetros geométricos sem e com influência destes na medida de semelhança final

Os resultados apresentados na Tabela 6.6 correspondem à taxa percentual de formas eliminadas pelos parâmetros geométricos introduzidos no descritor o que deverá dar uma ideia, ainda que apenas aproximada, das vantagens desta solução em termos da diminuição do peso computacional.

	A			B	C	Taxa Média Total
	A1	A2	A			
Formas ‘eliminadas’ pelos parâmetros geométricos	12.12%	12.26%	12.19%	12.58%	1.23%	8.67%

Tabela 6.6 – Taxa percentual de formas eliminadas pelos parâmetros geométricos incluídos no descritor ACC

Os resultados apresentados para o primeiro caso mostram a influência da filtragem associada aos parâmetros geométricos introduzidos no desempenho final do descritor ACC. Verificou-se uma melhoria, ainda que pequena e na ordem dos 0.5%, no desempenho médio total uma vez que, em média, apenas 8.67% das formas são filtradas; contudo, a grande vantagem da alteração proposta nesta secção está na redução do número de formas cuja comparação em termos dos ângulos de curvatura tem de ser processada (computacionalmente muito mais pesada que a comparação dos parâmetros geométricos), o que diminui significativamente o tempo de processamento correspondente ao processo de procura.

Já os resultados para o segundo caso mostram o impacto do uso dos parâmetros de excentricidade e circularidade no desempenho de procura obtido não só através da filtragem mas também através da medida de semelhança final. Como se pode ver na Tabela 6.5 o desempenho médio total do descritor volta a melhorar, desta vez cerca de 0.25% em relação ao desempenho do descritor com filtragem mas sem influência dos parâmetros geométricos na medida de semelhança, o que significa que estes parâmetros geométricos são boas medidas globais da forma do objecto mas incapazes de melhorar significativamente o desempenho do descritor ACC uma vez que o descritor apresenta ainda algumas debilidades em termos da insensibilidade a transformações geométricas, como é o caso da simetria e das mudanças de escala.

6.3 Diminuição da sensibilidade à simetria do descritor ACC

A sensibilidade à simetria foi outra das ‘fraquezas’ detectadas no descritor ACC através do seu fraco desempenho para as partes B e C do *core experiment CE-1*, isto porque ambas as bases de dados de conteúdo usadas para estas duas partes do *core experiment* contêm várias imagens simétricas de outras também existentes na base de dados. A ideia de melhorar o descritor através da compensação de transformações de simetria surgiu após a análise do desempenho e do conteúdo da base de dados usada sobretudo pela parte C do *core experiment CE-1*, principalmente nas 200 tramas binárias da sequência *Bream*. Esta análise evidenciou que existem muitos falsos resultados, aproximadamente 50%, o que se deve ao facto de durante a segunda metade da sequência o objecto apresentar uma forma simétrica (em relação ao eixo vertical) relativamente às formas da primeira metade da sequência, e o descritor ACC não conseguir detectar estas imagens como semelhantes.

A incapacidade do descritor de encontrar este tipo de imagens ou seja de as considerar semelhantes levou a que se tentasse resolver este problema, adoptando dois métodos que compensam a transformação de simetria no acto da procura:

- **Compensação da simetria através do uso de uma descrição de procura analiticamente simétrica** – Para além das 16 comparações de ângulos de curvatura efectuadas (4 conjuntos de ângulos de descrição ‘contra’ 4 conjuntos de ângulos extraídos para a forma exemplo na procura), determina-se analiticamente para cada vector de ângulos correspondente à forma exemplo um novo vector de ângulos que descreve uma forma simétrica, tal como se explicará de seguida. Estes novos vectores de ângulos (em número de 4) serão também comparados com os conjuntos de ângulos correspondentes às formas descritas, permitindo assim detectar formas simétricas daquela dada como exemplo ao processo de procura.
- **Compensação da simetria através do uso de uma descrição de procura gerada no sentido dos ponteiros do relógio** – Para além das 16 comparações de ângulos de curvatura efectuadas (4 conjuntos de ângulos de descrição ‘contra’ 4 conjuntos de ângulos extraídos para a forma exemplo na procura), determina-se um novo conjunto de ângulos de curvatura, para cada ponto de começo da forma exemplo mas agora gerado segundo o sentido dos ponteiros do relógio. Estes novos vectores de ângulos (em número de 4) serão também comparados com os conjuntos de ângulos correspondentes às formas descritas, permitindo assim detectar formas simétricas daquela dada como exemplo ao processo de procura.

Ambos os métodos acima apresentados não implicam qualquer alteração à definição do próprio descritor ACC mas apenas ao processo de procura e à medida de semelhança o que significa que a dimensão do descritor se mantém.

As secções seguintes descrevem mais pormenorizadamente e avaliam cada um dos métodos de compensação da simetria propostos, focando as suas implementações, propriedades e vantagens face à versão inicial deste parâmetro de descrição.

6.3.1 Definição da nova medida de semelhança

A ideia sobre a forma de melhorar o desempenho de procura do descritor ACC surge da própria noção de simetria. Diz-se que duas formas são simétricas se uma forma é igual à outra forma depois de a inverter/rodar em relação a um eixo horizontal ou vertical. A Figura 6.1 mostra exemplos de algumas formas e das suas formas simétricas.

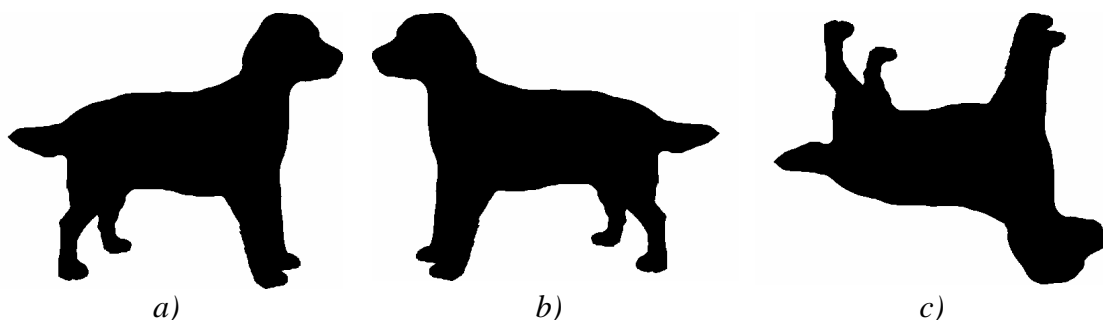


Figura 6.1 – Exemplo de simetria: Cão a) forma original, b) e c) formas simétricas relativamente ao eixo vertical e horizontal, respectivamente

6.3.1.1 Compensação da simetria através do uso de uma descrição de procura analiticamente simétrica

Com base na noção de simetria, uma forma cujo cálculo dos ângulos de curvatura é feito no sentido contrário ao dos ponteiros do relógio, como definido para o descritor ACC, teria que ver o cálculo dos ângulos da sua imagem simétrica segundo o eixo sobre o qual se localiza o ponto de começo em questão ser feito na direcção contrária, sentido dos ponteiros do relógio, para que a medida de semelhança entre os dois conjuntos de ângulos fosse nula.

Neste caso a compensação da simetria é feita através da geração analítica de um novo conjunto de ângulos para cada ponto de começo, no sentido contrário dos ponteiros do relógio, do seguinte modo:

- i. Cada um dos quatro vectores de ângulos, A_i , extraídos (segundo o sentido contrário aos ponteiros do relógio) para a forma exemplo fornecida ao mecanismo de procura é invertido segundo a expressão (6.6):

$$W[i] = A_i[(N-1)-i] \quad (6.6)$$

onde W representa o novo vector de ângulos invertido, N é o número de ângulos nos vectores e i varia no intervalo $[0, N-1]$. Uma vez invertido, o vector de ângulos da forma fornecida ao mecanismo de procura, transformam-se os ângulos em ângulos simétricos subtraindo-lhes 180° a fim de compensar o sentido de medição dos ângulos. A expressão (6.7) define esta compensação como:

$$W'[i] = 180^\circ - W[i] \quad (6.7)$$

onde W' é o vector dos ângulos compensados de simetria e correspondentes aos da forma simétrica fornecida ao mecanismo de procura e i varia novamente entre $[0, N-1]$. A Figura 6.2 apresenta graficamente este método de compensação da simetria, através das suas várias fases e transformações;

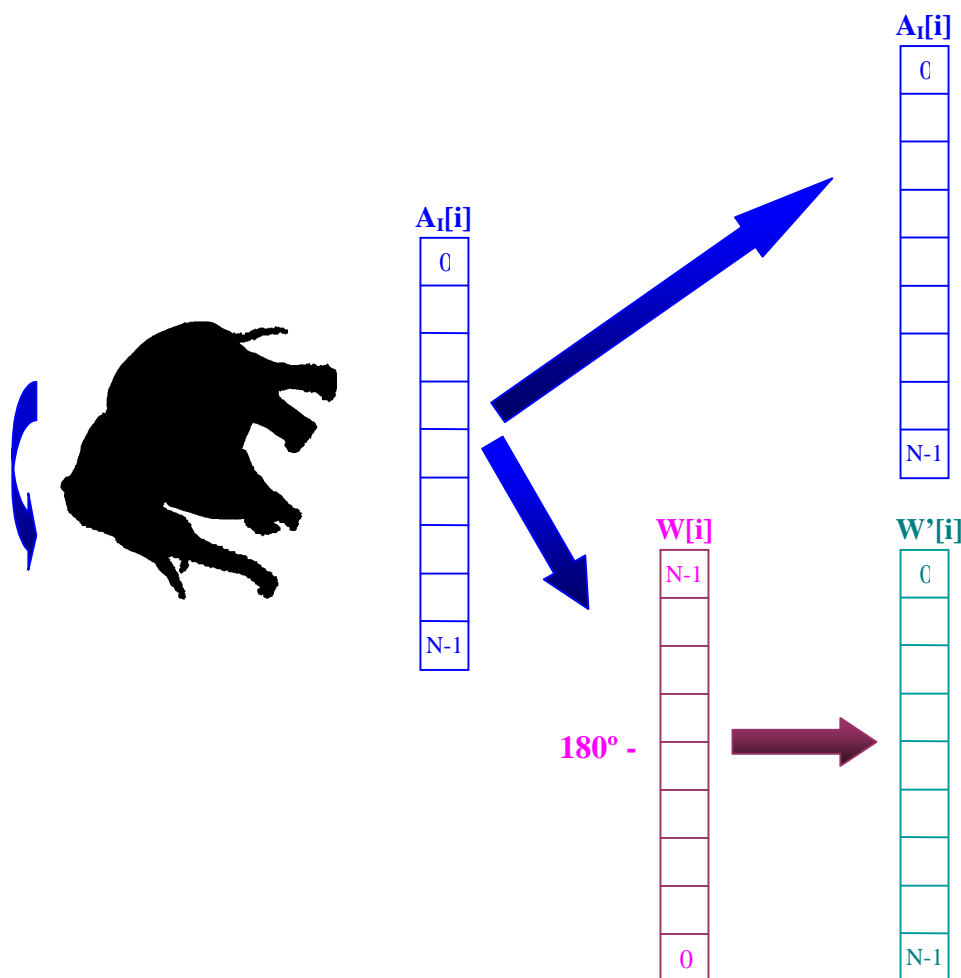


Figura 6.2 – Esquema representativo do método de compensação da simetria através do uso de uma descrição de procura analiticamente simétrica

- iii. Após o cálculo analítico do segundo vector de ângulos para cada ponto de começo da forma exemplo, cada um dos vectores de ângulos da forma fornecida ao mecanismo de procura (8 no total) será comparado com cada um dos 4 vectores de ângulos, A_B , correspondentes a cada uma das forma em base de dados usando a medida de semelhança definida na secção 6.2.2. Desta comparação resultam dois valores, D_1 e D_2 , determinados de acordo com as expressões (6.8) e (6.9), respectivamente, que traduzem o grau de semelhança entre cada uma das descrições das formas em base de dados e cada uma das descrições para a forma exemplo fornecida ao mecanismo de procura e para as correspondentes descrições simétricas analiticamente determinadas. De entre estes dois valores é escolhido o de menor valor, D_{final} , uma vez que este é o que melhor traduz o grau de semelhança existente entre as duas formas, como mostra a expressão (6.10).

$$D_1 = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^k |A_{B_c}(t) - A_{I_c}(t)| + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.8)$$

$$D_2 = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^k |A_{B_c}(t) - W_c'(t)| + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.9)$$

$$D_{final} = \min\{D_1, D_2\} \quad (6.10)$$

6.3.1.2 Compensação da simetria através do uso de uma descrição de procura gerada no sentido dos ponteiros do relógio

O segundo método proposto para a compensação da simetria segue novamente a noção de simetria. Assim, uma imagem cuja extracção dos ângulos de curvatura é feita no sentido contrário ao dos ponteiros do relógio teria que ver o cálculo dos ângulos da sua imagem simétrica segundo o eixo sobre o qual se localiza o ponto de começo em questão ser feito na direcção contrária, ou seja no sentido dos ponteiros do relógio para que a medida de semelhança entre eles seja nula.

Também com este segundo método não se introduziu qualquer alteração na especificação do descritor ACC, alterando-se apenas a medida de semelhança usada no processo de procura de modo a compensar a transformação de simetria. Esta compensação da simetria foi feita do seguinte modo:

- i. Durante o processo de procura, a forma exemplo fornecida ao mecanismo de procura é descrita de modo semelhante ao modo usado para descrever as formas na base de dados gerando-se um vector de ângulos de curvatura, A_{I1} , para cada um dos quatro pontos de começo percorrendo o contorno no sentido contrário ao dos ponteiros do relógio mas também um outro vector de ângulos, A_{I2} , para cada um dos mesmos quatro pontos de começo, percorrendo o contorno no sentido dos ponteiros do relógio. A Figura 6.3 apresenta este método de compensação da simetria, através das suas várias fases e transformações;

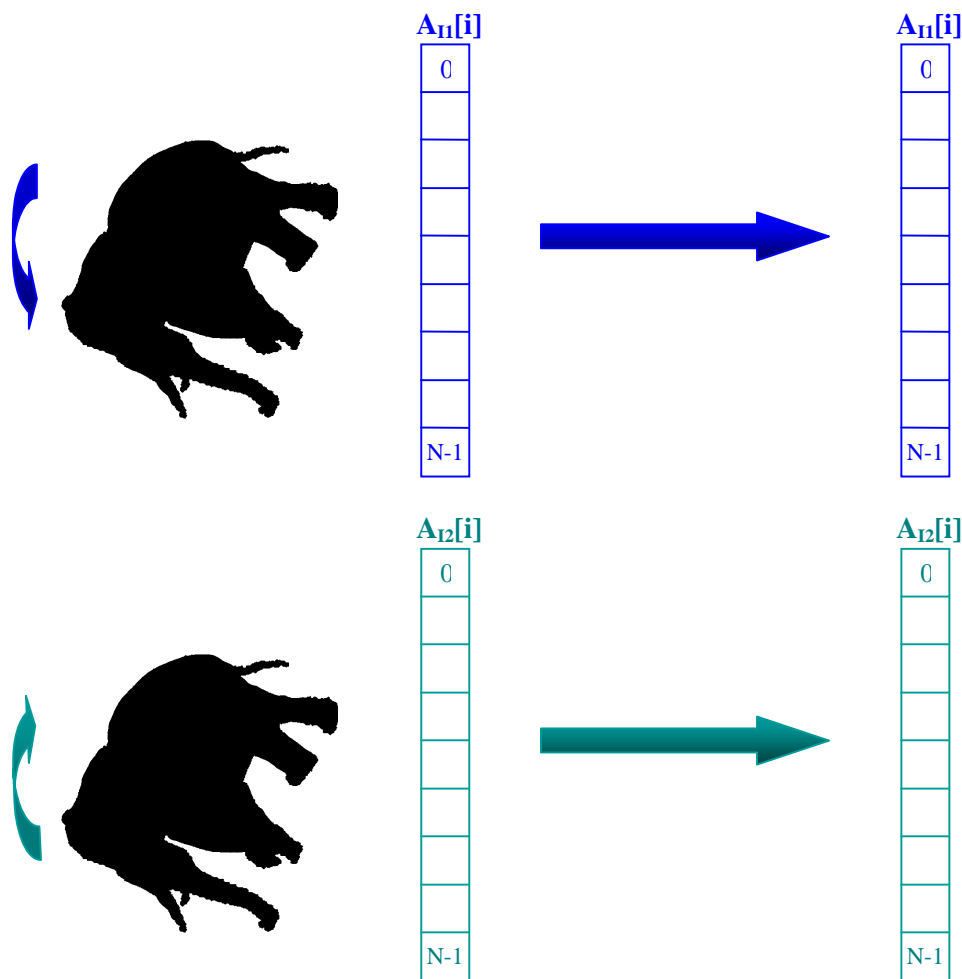


Figura 6.3 – Esquema representativo do método de compensação da simetria através do uso de uma descrição de procura gerada no sentido dos ponteiros do relógio

- ii. Uma vez gerados os oito vectores de ângulos (quatro para cada um dos dois sentidos percorridos) para a forma fornecida ao mecanismo de procura, comparam-se então os vectores de ângulos em ambas as direcções com os quatro vectores de ângulos correspondentes às descrições das formas nas bases de dados, ou seja, cada um dos oito vectores de ângulos da procura com cada um dos quatro vectores da descrição na base de dados. Desta comparação resultam novamente dois valores, D_1 e D_2 , calculados através das expressões (6.11) e (6.12), respectivamente, que traduzem o grau de semelhança entre cada uma das descrições das formas em base de dados e cada uma das descrições para a forma exemplo fornecida ao mecanismo de procura e para as correspondentes descrições simétricas determinadas segundo o sentido dos ponteiros do relógio. Finalmente, a forma que minimiza a medida final será escolhida como aquela que melhor aproxima a forma fornecida ao mecanismo de procura, de entre aquelas disponíveis na(s) bases de dados (ver expressão (6.13)).

$$D_1 = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^k |A_{B_c}(t) - A_{I_{1c}}(t)| + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.11)$$

$$D_2 = \min_{1 \leq c \leq 16} \left\{ \sum_{t=1}^k |A_{B_c}(t) - A_{I_{2c}}(t)| + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.12)$$

$$D_{final} = \min\{D_1, D_2\} \quad (6.13)$$

6.3.2 Nova Análise do desempenho do descritor ACC

Os resultados obtidos para o desempenho de procura do descritor ACC usando os métodos de compensação da simetria descritos nas secções anteriores, são os apresentados na Tabela 6.7:

	A			B	C	Desempenho Médio Total
	A1	A2	A			
Antes da alteração	64.76%	91.67%	78.22%	57.41%	63.50%	66.38%
Depois da alteração usando o 1º método	64.76%	93.81%	79.29%	64.85%	83.00%	75.71%
Depois da alteração usando o 2º método	65.24%	93.81%	79.53%	64.84%	83.50%	75.96%

Tabela 6.7 – Desempenho do descritor ACC antes e depois das alterações à medida de semelhança para a compensação de transformações de simetria conforme os dois métodos anteriormente descritos

Como se pode verificar, o desempenho médio total do descritor para o primeiro método melhora significativamente, apresentando um ganho de 9.33% em relação à situação anterior. Como era de esperar, o desempenho melhorou sobretudo para as partes B e C pois cada uma delas viu o seu desempenho melhorar em 7.4% e 19.5%, respectivamente.

No entanto e para os resultados obtidos para o segundo método proposto, verifica-se que o desempenho médio total do descritor ACC aumentou apenas 0.25% em relação ao primeiro método sugerido. Apesar do ligeiro aumento de desempenho do descritor para o segundo método, pode-se concluir que ambos os métodos têm um desempenho semelhante. Nestas condições, propõe-se o uso do primeiro método como solução para o problema da compensação de simetria uma vez que o segundo método apresenta as seguintes desvantagens: maior complexidade computacional associada à extracção de novos conjuntos de ângulos e aumento do tempo de processamento na fase de procura associado especialmente ao facto de ter de se gerar mais uma descrição e comparar mais 16 pares de vectores de ângulos de curvatura.

6.4 Optimização do número de ângulos de curvatura usados por descrição

A alteração proposta nesta secção ao descritor ACC tem como principal objectivo melhorar o seu desempenho para a parte A1 do *core experiment CE-I*, optimizando o número de ângulos de curvatura a usar por descrição.

Com base neste objectivo, existem duas alternativas possíveis para melhorar o desempenho do descritor em termos do número de ângulos usados por descrição de forma:

1. Variar o *Número_Ângulos* de curvatura usados para descrever a forma consoante a complexidade da forma do objecto a descrever; ou
2. Fixar à partida o valor do *Número_Ângulos* de curvatura no valor que maximiza o desempenho do descritor ACC ou seja que permita descrever mais fielmente a noção de forma através do descritor ACC; esse valor poderá ser eventualmente diferente do valor actual de 64.

Ainda que tanto um como o outro método permitam optimizar de algum modo o desempenho do descritor, é claro que o primeiro método é bastante mais flexível e como tal deveria dar origem a maiores ganhos de desempenho em virtude da capacidade de adaptar o descritor às características específicas de cada forma. Contudo, a primeira alternativa apresenta grandes dificuldades no que diz respeito à definição de uma medida de semelhança capaz de lidar com a variação do número de ângulos de curvatura usados. Como consequência destas dificuldades, teve de se optar pela segunda alternativa para tentar optimizar o número de ângulos de curvatura usados nas descrições em relação à complexidade das formas a descrever, nomeadamente o conteúdo correspondente ao *core experiment CE-I*.

Estando a ser usados actualmente 64 ângulos de curvatura para descrever a forma através do descritor ACC, o que se pode concluir observando o desempenho relativo à parte A1 do *CE-I* obtido é que este número de ângulos é extremamente elevado para o grau de complexidade de algumas das formas da base de dados correspondentes à parte A1 do *core experiment CE-I*, sobretudo as formas associadas a reduções de outras formas que apresentam um número total de pontos de contorno inferior a 64 pontos; este facto implica que os ângulos de contorno para estas formas são gerados através do processo de interpolação apresentado na secção 5.3.2.1 o que dá origem a algum ruído.

A optimização proposta nesta secção ao descritor ACC não alterará a sua sintaxe e implicará apenas uma ligeira alteração semântica, correspondente à alteração do valor previamente definido para a componente *Número_Ângulos* que determina o número de ângulos de curvatura a usar na criação de qualquer descrição de forma usando o descritor ACC.

6.4.1 Optimização da componente *Número_Ângulos* do descritor ACC

A optimização da componente *Número_Ângulos* usando o mesmo valor para todas as descrições de forma, foi feita analisando-se;

- O desempenho do descritor ACC para um conjunto de valores de *Número_Ângulos* para todo o conteúdo correspondente ao *core experiment CE-1* apresentado na secção 5.2.1;
- O desempenho do descritor ACC para um conjunto de valores de *Número_Ângulos* relativamente às reduções existentes na base de dados correspondente à parte A1 do *core experiment CE-1*;
- O desempenho do descritor ACC para um conjunto de valores de *Número_Ângulos* relativamente às ampliações existentes na base de dados correspondente à parte A1 do *core experiment CE-1*.

Para a escolha do conjunto de valores de *Número_Ângulos* a testar foi necessário definir, em primeiro lugar, os valores extremos do intervalo a testar, ou seja o seu valor máximo e mínimo. Uma vez que um *Número_Ângulos* igual a 64 e um conteúdo composto maioritariamente por reduções, o descritor apresentou um baixo desempenho relativamente à parte A1 do *CE-1*, optou-se por definir como valor máximo deste conjunto o valor 64, evitando a introdução de ruído nas descrições de objectos pequenos. Como valor mínimo do conjunto de valores a testar, escolheu-se o valor 8 uma vez que a possibilidade de existirem objectos com perímetros iguais ou inferiores a este valor é pequena. Por conseguinte, o conjunto de valores irá variar entre 8 e 64, de 8 em 8 unidades, ou seja 8, 16, 24, 32, 40, 48, 56 e 64, uma vez que se considera que esta resolução de teste é já suficientemente baixa.

Após a análise do desempenho de procura do descritor ACC para cada um destes valores de *Número_Ângulos*, optar-se-á pelo valor que consiga o melhor compromisso entre o desempenho de procura do descritor e o número de ângulos a usar na descrição da forma que determina a dimensão binária do descritor.

6.4.2 Nova análise do desempenho do descritor ACC

Considerando o conjunto de valores para *Número_Ângulos* definido na secção anterior, começou-se por avaliar o desempenho médio total do descritor ACC nas condições do *core experiment CE-1*; os resultados obtidos são apresentados na Tabela 6.8:

<i>Número_Ângulos</i>	A			B	C	Desempenho Médio Total
	A1	A2	A			
8	77.62%	90.95%	84.29%	63.47%	89.00%	78.92%
16	80.48%	93.81%	87.15%	65.67%	84.50%	79.10%
24	77.86%	94.05%	85.96%	65.60%	82.00%	77.85%
32	76.67%	94.05%	85.36%	65.25%	83.50%	78.04%
40	73.33%	93.81%	83.57%	65.39%	84.00%	77.65%
48	72.86%	94.29%	83.57%	65.30%	83.50%	77.46%
56	69.52%	94.05%	81.79%	65.03%	82.50%	76.44%
64	64.76%	93.81%	79.29%	64.85%	83.00%	75.71%

Tabela 6.8 – Desempenho médio total do descritor ACC para o conjunto de valores de *Número_Ângulos* definido

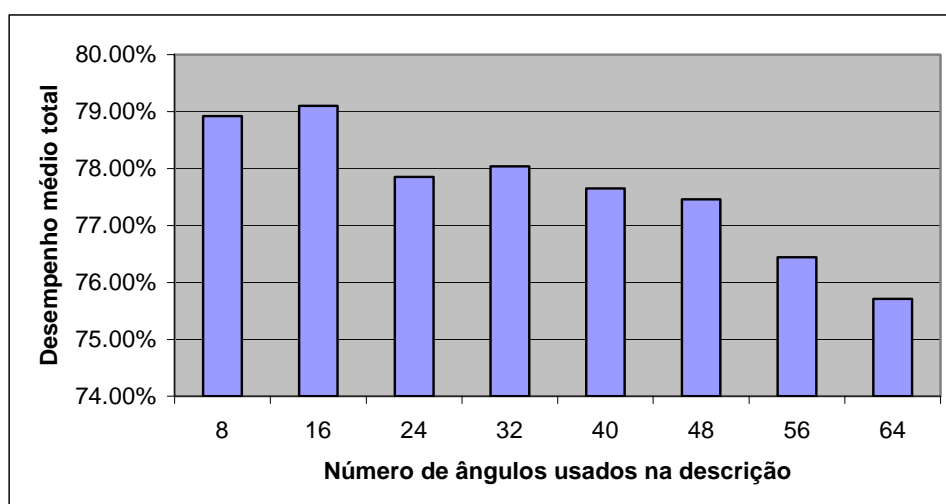


Figura 6.4 – Desempenho médio total do descritor ACC para o conjunto de valores de *Número_Ângulos* definido

Como se pode constatar a partir da Tabela 6.8 e da Figura 6.4, o número de ângulos de curvatura para o qual o desempenho médio total é máximo é 16; para 16 ângulos existe um ganho de desempenho de 3.4% em relação ao uso de 64 ângulos, tal como foi proposto ao MPEG pelos proponentes iniciais do descritor ACC.

Numa segunda etapa, avaliou-se o desempenho do descritor ACC relativamente às reduções e ampliações da imagem de referência de cada classe existente na base de dados correspondente à parte A1 do *core experiment CE-1* usando o mesmo conjunto de valores para *Número_Ângulos*. O desempenho apresentado para as reduções foi calculado com base na capacidade do algoritmo detectar todas as reduções pertencentes a uma mesma classe (4 reduções e o original) e existentes na base de dados. Analogamente, o desempenho apresentado para as ampliações foi calculado com base na capacidade do algoritmo detectar todas as ampliações pertencentes a uma mesma classe (1 ampliação e o original) e existentes

na base de dados. O desempenho médio abaixo calculado corresponde ao valor médio do desempenho conseguido para as reduções e ampliações como mostra a expressão (6.14), sendo meramente ilustrativo do desempenho conseguido para cada valor de *Número_Ângulos*.

$$Desempenho\ Médio = \frac{A1_{reduções} + A1_{ampliações}}{2} \quad (6.14)$$

A Tabela 6.9 apresenta os resultados obtidos para esta segunda etapa:

<i>Número_Ângulos</i>	A1 (reduções)	A1 (ampliações)	Desempenho Médio
8	74.57%	91.43%	83.00%
16	77.71%	95.71%	86.71%
24	75.71%	95.71%	85.71%
32	74.00%	95.71%	84.86%
40	70.57%	95.71%	83.14%
48	70.00%	95.71%	82.86%
56	66.00%	95.71%	80.86%
64	60.86%	95.71%	78.29%

Tabela 6.9 – Desempenho do descritor ACC relativamente às reduções e ampliações existentes na base de dados correspondente à parte A1 do CE-1

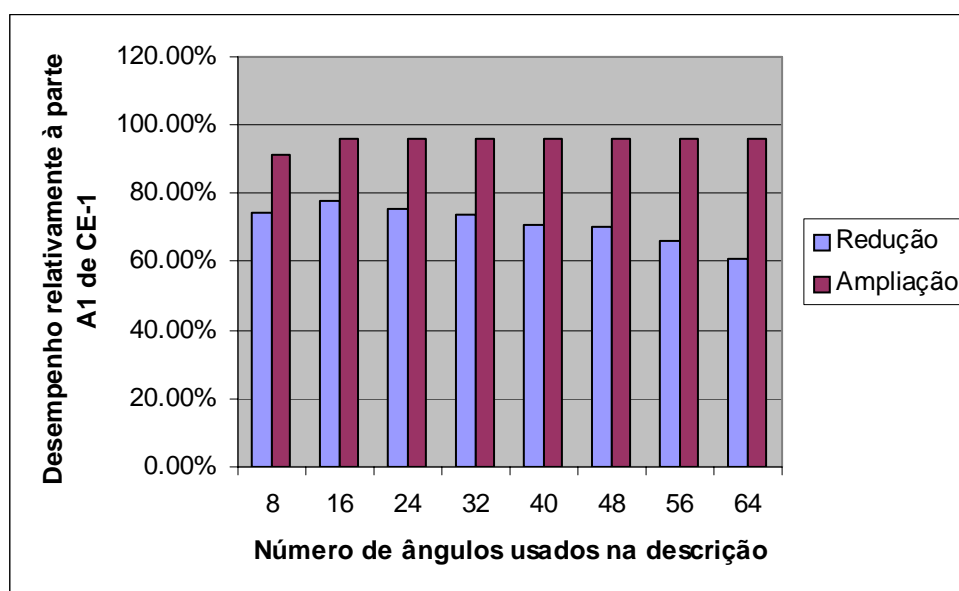


Figura 6.5 – Desempenho do descritor relativamente às reduções e ampliações existentes na base de dados correspondente à parte A1 do CE-1

Como se pode constatar a partir da Tabela 6.9 e da Figura 6.5, o melhor desempenho do descritor relativamente às reduções existentes na base de dados A1 é para um número de ângulos de 16. No que diz respeito ao desempenho relativamente às ampliações existentes na base de dados A1, verifica-se que para um número de ângulos superior ou igual a 16 o desempenho de procura estabiliza o que significa que o uso de um número de ângulos superior a 16 em nada melhora a capacidade de descrição da forma e logo o desempenho. Com base nesta conclusão, pode-se ainda afirmar que ao reduzir o número de ângulos a usar por descrição, em relação aos 64 ângulos inicialmente usados, se diminui a sensibilidade do descritor a mudanças de escala. Conclui-se assim que a alteração do número de ângulos de curvatura a usar nas descrições baseadas no descritor ACC, *Número_Ângulos*, de 64 para 16 resultou numa melhoria global do desempenho de procura de cerca de 3.39%.

A diminuição do número de ângulos de curvatura a usar nas descrições baseadas no descritor ACC de 64 para 16 permite também uma diminuição significativa da dimensão do descritor, como se pode constatar através dos dados apresentados na Tabela 6.10:

<i>CE-1</i>			ACC (<i>bits</i> por descrição antes de melhoria)		ACC (<i>bits</i> por descrição após melhoria)	
Partes	A	A1	2322	2322	594	594
		A2	2322		594	
	B		2329		601	
	C		2320		592	
	Média Total (<i>bits</i>)			2323.7		595.7

Tabela 6.10 – Número médio de *bits* por descrição antes e após a optimização do descritor ACC relativamente ao valor da componente *Número_Ângulos*

O novo valor médio do número de *bits* usado para cada descrição de forma baseada no descritor ACC, apresenta uma diminuição significativa relativamente ao valor anterior, cerca de 75%, uma vez que a dimensão do descritor ACC passa para aproximadamente 4 vezes menos. Este decréscimo na dimensão do descritor ACC é extremamente importante uma vez que permite aproximar a dimensão média duma descrição baseada no descritor ACC dos valores típicos duma descrição baseado no descritor CSS normalizado pelo MPEG-7 (114 *bits*).

6.5 Estudo do descritor ACC usando a distância Euclidiana

As alterações efectuadas ao descritor ACC na secção anterior permitiram melhorar o seu desempenho médio significativamente ou seja em cerca de 3.4%. No entanto, melhorias adicionais podem ainda ser obtidas adoptando uma distância para a medida de semelhança entre a forma exemplo e as formas descritas nas bases de dados mais adequada ao descritor em questão. Assim, analisar-se-á nesta e na próxima secção o papel fundamental de duas distâncias de semelhança no desempenho de procura para o descritor ACC. A medida de

semelhança exprime o conceito de semelhança de forma que se pretende usar e como tal determina fortemente o resultado das procuras.

Dada a importância da medida de semelhança para o desempenho final do descritor, decidiu-se avaliar, no âmbito desta e da próxima secção, o desempenho de procura do descritor ACC usando duas distâncias de semelhança diferentes daquela que tem vindo a ser usada ou seja a distância de *Minkowsky* já incluindo os parâmetros geométricos: a distância Euclidiana e a distância *alpha-trimmed average* [IaPu1999].

6.5.1 Definição e implementação da medida de semelhança baseada na distância Euclidiana

A distância Euclidiana é de entre todas as distâncias usadas nas medidas de semelhança a mais imediata devido ao seu significado óbvio nos espaços 2D e 3D e por isso mesmo a distância mais frequentemente usada para a comparação de dois vectores. A distância Euclidiana, D_E , entre dois vectores A_B e A_I , neste caso vectores com os ângulos de curvatura do contorno referentes à forma exemplo fornecida ao mecanismo de procura e às formas descritas na base de dados, respectivamente, é definida através da expressão (6.15) que se segue:

$$D_E = \sqrt{\sum_{i=0}^{N-1} (A_B[i] - A_I[i])^2} \quad (6.15)$$

onde N é o número de ângulos no vector e i é o índice de cada ângulo no vector.

No entanto, esta medida de semelhança não é das mais fiáveis em termos da medição do grau de semelhança entre duas formas, sobretudo devido à sua sensibilidade a pequenas variações na forma [IaPu1999]. A sensibilidade desta medida de semelhança a pequenas variações é explicada através da sua própria expressão que soma o quadrado das diferenças entre cada dois elementos do vector o que significa que, quando a distribuição do ruído pela forma do objecto é não-uniforme a contribuição deste para o aumento arbitrário da diferença entre duas formas semelhantes poderá ser elevado, uma vez que se soma o quadrado das diferenças [IaPu1999]. O desempenho correspondente à distância Euclidiana irá ser comparado com o desempenho correspondente à actual distância de *Minkowsky*, ambas incluindo na distância final duas parcelas relativas aos parâmetros geométricos (como mostra a expressão (6.16)); note-se que na distância de *Minkowsky* as diferenças entre os ângulos não são elevadas ao quadrado o que é uma vantagem em termos da sensibilidade a pequenas variações da forma. A comparação destas duas distâncias deverá permitir confirmar ou não as afirmações feitas em relação ao desempenho do descritor ACC em combinação com cada uma das duas medidas de semelhança.

$$D_{final} = \min_{1 \leq c \leq 16} \left\{ \sqrt{\sum_{i=0}^{N-1} (A_{B_c}[i] - A_{I_c}[i])^2} + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.16)$$

6.5.2 Nova análise do desempenho do descritor ACC

Após a implementação da distância Euclidiana no contexto do módulo de procura do XM, fez-se uma análise comparativa do desempenho desta distância em relação à distância de *Minkowsky* a fim de verificar as vantagens e desvantagens de cada uma delas. A análise do desempenho será feita com base nos resultados apresentados na Tabela 6.11.

	A			B	C	Desempenho Médio Total
	A1	A2	A			
Distância <i>Minkowsky</i>	80,48%	93,81%	87,15%	65,67%	84,50%	79,10%
Distância Euclidiana	78,57%	93,57%	86,07%	64,24%	83,00%	77,77%

Tabela 6.11: Desempenho de procura para as distâncias de *Minkowsky* e Euclidiana

Os resultados da Tabela 6.11 mostram que o desempenho médio total relativo à distância Euclidiana é inferior, em cerca de 1.3%, ao desempenho da distância de *Minkowsky*, o que confirma os problemas da distância Euclidiana acima referidos (grande sensibilidade a pequenas variações na forma do objecto) e as suas consequências no desempenho de procura. Com base nestes resultados, conclui-se que a distância de *Minkowsky* traduz melhor o conceito de semelhança para descrições de forma baseadas no descritor ACC, nomeadamente por ser mais insensível a pequenas variações no contorno do objecto.

6.6 Estudo do descritor ACC usando a distância *alpha-trimmed average*

À semelhança do que foi dito na secção anterior, a medida de semelhança exprime o conceito de semelhança de forma que se pretende usar e como tal determina fortemente o resultado das procuras. Neste sentido, esta secção irá definir e analisar o desempenho do descritor ACC quando usado em combinação com a distância *alpha-trimmed average*.

6.6.1 Definição e implementação da medida de semelhança baseada na distância *alpha-trimmed average*

Nesta secção ir-se-á introduzir o conceito de mediana com o intuito de melhorar o desempenho do descritor ACC. A mediana é matematicamente definida como o valor central de um conjunto de valores ordenados, por ordem crescente ou decrescente. A expressão (6.17) define a mediana de um vector de valores ordenados X de N elementos:

$$Mediana = \begin{cases} X\left(\frac{N}{2}\right) & \text{se } N \text{ ímpar} \\ \frac{X\left(\frac{N-1}{2}\right) + X\left(\frac{N}{2}\right)}{2} & \text{se } N \text{ par} \end{cases} \quad (6.17)$$

onde $X(n)$ é o valor do vector correspondente à posição n .

Com base na definição de mediana, o que acontece ao ordenar os valores do vector é que se forçam as amostras associadas aos valores mais elevados do ruído (normalmente nos extremos da gama de variação) a ficarem nas extremidades da variação e por conseguinte longe do valor mediano, minimizando o seu impacto no valor da mediana o que não acontece, por exemplo, no valor da média. Tentando explorar este efeito, apresenta-se de seguida a distância *alpha-trimmed average* que tem como um dos seus casos particulares o valor mediano: a distância *alpha-trimmed average*, D_{ATA} , é uma distância baseada na soma de uma percentagem dos componentes de um vector em torno do valor mediano (vizinhança) do vector em análise.

A Figura 6.6 representa o conceito aqui definido para a distância *alpha-trimmed average*. O vector da figura é composto por 11 componentes tendo como valor mediano e consequentemente D_{ATA} para uma vizinhança de 0% a componente de índice 6. Para os restantes casos apresentados na figura, o símbolo de chaveta abraça todas as componentes consideradas no cálculo de D_{ATA} consoante a vizinhança definida em torno do valor mediano.

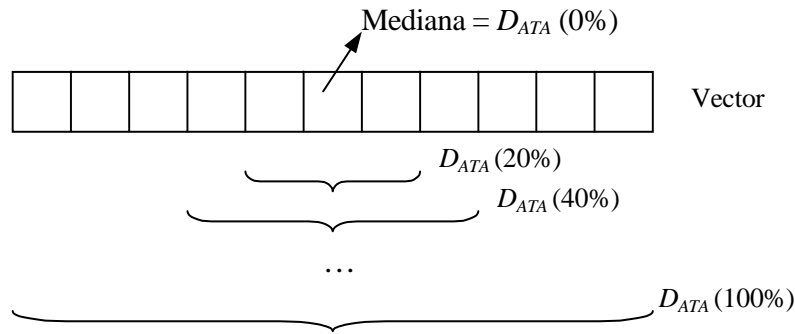


Figura 6.6 – Exemplo representativo do cálculo da distância D_{ATA}

A implementação da medida de semelhança usando a distância *alpha-trimmed average* foi feita em cinco etapas:

- i. Para cada uma das combinações de cada um dos 2 conjuntos de combinações, 16 combinações considerando os 4 vectores de ângulos de descrição com 4 de procura sem compensação de simetria e 16 combinações considerando os 4 vectores de ângulos de descrição com 4 de procura com compensação de simetria), é calculada a diferença absoluta entre cada componente dos respectivos vectores de ângulos de curvatura sendo estes valores armazenados num vector do tipo D_{DA} :

$$D_{DA}[j][i] = |A_B(i) - A_I(i)| \quad (6.18)$$

onde A_B e A_I são os vectores de ângulos de curvatura correspondentes à descrição da forma fornecida ao mecanismo de procura e às descrições na base de dados, respectivamente, i é o índice do ângulo respectivo e j é o índice da respectiva combinação. Consequentemente, resultam desta etapa dois vectores do tipo D_{DA} , um relativo às 16 combinações sem compensação de simetria, D_{DA1} , e o outro vector relativo às restantes 16 combinações com compensação de simetria, D_{DA2} .

- ii. Ordena-se por ordem crescente cada uma das componentes dos vectores D_{DA} que contém as diferenças absolutas entre os componentes correspondentes dos vectores de ângulos.
- iii. Calcula-se o valor mediano de cada componente dos vectores D_{DA} de acordo com a expressão (6.17).
- iv. Calcula-se a distância *alpha-trimmed average*, D_{ATA} , como o valor mediano médio usando todas as componentes do vector em análise dentro de uma certa vizinhança da posição correspondente à mediana. O valor mediano médio é assim definido como a média dos valores de D_{DA} para cada combinação numa vizinhança $2xM_L$ (limiar de vizinhança) do valor mediano:

$$D_{ATA_{M_L}}[c] = \frac{1}{N - 2 \cdot \lfloor (0.5 - M_L) \cdot (N-1) \rfloor} \sum_{n=\lfloor (0.5 - M_L) \cdot (N-1) \rfloor + 1}^{N - \lfloor (0.5 - M_L) \cdot (N-1) \rfloor} D_{DA}[c][n] \quad (6.19)$$

onde N é o número de ângulos de curvatura usados por descrição e M_L , definido entre 0 e 0.5 (assumindo $2xM_L$ um valor máximo de 100%), é o limite da vizinhança considerada para o cálculo da medida de semelhança e c é o índice da combinação. Novamente, resultam dois vectores de distâncias *alpha-trimmed average*, D_{ATA1} e D_{ATA2} , com 16 componentes cada.

- v. Calculam-se os valores finais parciais, D_{final1} e D_{final2} , contabilizando ainda a distância dos parâmetros geométricos introduzidos no descritor ACC à semelhança da expressão (6.19), ou seja:

$$D_{final1,2} = \min_{1 \leq c \leq 16} \left\{ D_{ATA1,2}[c] + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\} \quad (6.20)$$

- vi. O valor final da medida de semelhança, D_{final} , é o valor mínimo dentre estes dois valores finais parciais e que melhor traduz o grau de semelhança entre as duas formas:

$$D_{final} = \min\{D_{final1}, D_{final2}\} \quad (6.21)$$

Após a implementação desta medida de semelhança, resta analisar o desempenho do descritor ACC nestas condições, nomeadamente comparativamente à distância de *Minkowsky* até agora usada.

6.6.2 Nova análise de desempenho do descritor ACC

A análise do desempenho de procura obtido para a nova medida de semelhança teve também como objectivo determinar o valor óptimo para o parâmetro M_L , que pode variar entre 0 e 0.5, ou seja do valor de M_L que maximiza o desempenho do descritor ACC. A escolha do valor a atribuir ao limiar de vizinhança, M_L , está ligada ao facto de que as amostras associadas a valores mais elevados do ruído se devem encontrar nas extremidades do vector ordenado e

por conseguinte longe do valor mediano e logo valores mais elevados de M_L tendem a aumentar o impacto no valor da distância de semelhança das amostras mais ruidosas.

Assim, foram escolhidos alguns valores entre 0% e 50% para o limiar de vizinhança M_L , a fim de determinar qual o melhor compromisso entre o desempenho do descritor ACC e o limiar de vizinhança. Os valores atribuídos ao limiar M_L para análise do desempenho do descritor ACC foram: 0%, 10%, 20%, 30%, 40%, e 50%.

Após a implementação da medida de semelhança baseada na distância *alpha-trimmed average*, obtiveram-se os resultados apresentados na Tabela 6.12 e na Figura 6.7 para os vários limiares avaliados.

	A			B	C	Desempenho Médio Total
	A1	A2	A			
$M_L = 50\%$	80.48%	93.81%	87.15%	65.67%	84.50%	79.10%
$M_L = 40\%$	80.95%	94.05%	87.50%	65.36%	85.00%	79.29%
$M_L = 30\%$	81.19%	93.81%	87.50%	64.67%	84.50%	78.89%
$M_L = 20\%$	81.19%	93.81%	87.50%	64.10%	84.00%	78.53%
$M_L = 10\%$	80.71%	93.33%	87.03%	62.07%	83.00%	77.36%
$M_L = 0\%$	79.52%	93.10%	86.31%	61.03%	82.00%	76.45%

Tabela 6.12 - Desempenho do descritor ACC para a distância *alpha-trimmed average* usando vários valores para o limiar M_L

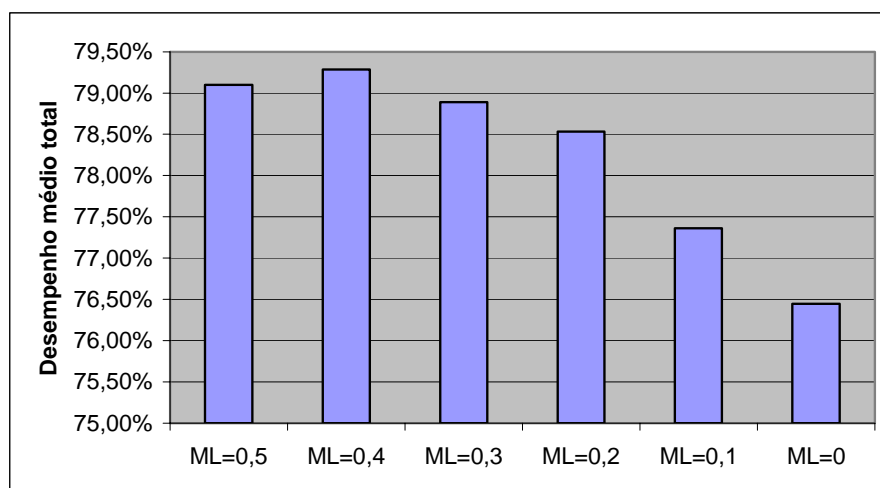


Figura 6.7 –Desempenho do descritor ACC para a distância *alpha-trimmed average* usando vários valores para o limiar M_L

Dos resultados obtidos para o desempenho de procura em função do valor de M_L , pode constatar-se que aquele que conseguiu o melhor desempenho médio total foi o que considera uma vizinhança correspondente a 80% dos valores das amostras (diferenças) em torno da mediana, ou seja M_L igual a 40%. Estes resultados confirmam que as diferenças dos ângulos

associadas às variações do contorno por influência do ruído, se encontram de facto nas extremidades do vector ordenado das diferenças e que a sua eliminação melhora o desempenho do descritor em aproximadamente 0.2% face ao desempenho médio inicial medido usando-se a distância de *Minkowsky*, equivalente a M_L igual a 50%.

6.7 Optimização do número de vectores de ângulos de curvatura no descritor ACC

O descritor ACC até agora usado neste capítulo descreve uma forma usando quatro vectores de ângulos de curvatura, um para cada ponto de começo, tal como descrito na secção 5.3.2.1. A solução de gerar quatro vectores de ângulos e não apenas um aumenta inevitavelmente a dimensão do descritor ACC tornando-o pouco competitivo em termos de dimensão relativamente ao descritor CSS mas foi alegadamente feita pelos proponentes iniciais do descritor ACC de modo a melhorar o seu desempenho de procura uma vez que uma solução alternativa visando a geração de um menor número de vectores de ângulos nunca foi precisamente definida e testada.

A optimização do número de vectores de ângulos a usar por descrição ACC baseia-se num facto de extrema importância relativamente aos pontos de começo escolhidos e usados para descrever a forma de um objecto. Até aqui, a escolha dos quatro pontos de começo foi feita com base nos momentos (valores próprios) da forma [NiYi1995] de modo a garantir a insensibilidade à rotação. No entanto, nem sempre se justifica o uso de quatro vectores de ângulos de curvatura por descrição, uma vez que parece existir uma repetição de informação sobre a forma descrita uma vez que cada um destes 4 conjuntos de ângulos corresponde a uma aproximação poligonal do contorno; assim, a existência de 4 vectores de ângulos corresponde à existência de 4 aproximações poligonais do mesmo contorno o que constitui uma solução claramente redundante em termos de descrição do contorno.

Esta secção assim como as duas que se seguem pretendem estudar o melhor compromisso entre a dimensão do descritor ACC e o seu desempenho médio total. Partindo do pressuposto de que o uso sistemático de 4 vectores de ângulos não parece ser uma boa solução, esta secção procura definir um critério de selecção para o número de vectores de ângulos a usar por descrição de forma, apresentando também a análise do desempenho e de dimensão do descritor ACC para esta nova solução.

6.7.1 Definição do critério de selecção do número de vectores de ângulos no descritor ACC

O critério de selecção para o número de vectores de ângulos a usar na descrição ACC bem como para os respectivos pontos de começo proposto nesta secção depende de dois tipos de parâmetros característicos da forma:

- i. Os momentos, valores próprios calculados pela matriz de covariância definida pela expressão (5.7), relativos aos eixos principais da forma, M_1 e M_2 ; e

$$V = \begin{bmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{bmatrix}$$

$$V_{xx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(x_i - m_x)$$

$$V_{xy} = V_{yx} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)(y_i - m_y)$$

$$V_{yy} = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - m_y)(y_i - m_y)$$
(6.22)

$$\lambda_{1,2} = M_{1,2} = (V_{xx} + V_{yy})^2 \pm \sqrt{(V_{xx} + V_{yy})^2 - 4(V_{xx}V_{yy} - V_{xy}^2)} \quad (6.23)$$

- ii. As distâncias, D_{11} , D_{12} , D_{21} e D_{22} , entre cada ponto de intersecção do contorno, P_{11} , P_{12} , P_{21} e P_{22} , com os eixos principais do objecto e o seu centróide. D_{11} é a distância do ponto de começo, situado no eixo principal, mais distante do centro de massa e D_{12} a distância do ponto situado no extremo oposto do eixo e cuja distância é inferior. Analogamente, D_{21} e D_{22} são as maiores e menores distâncias, respectivamente, entre os restantes dois pontos de começo situados sobre o eixo perpendicular ao eixo principal e o centróide. A Figura 6.8 define graficamente estes pontos e as respectivas distâncias relativamente ao centróide da forma.

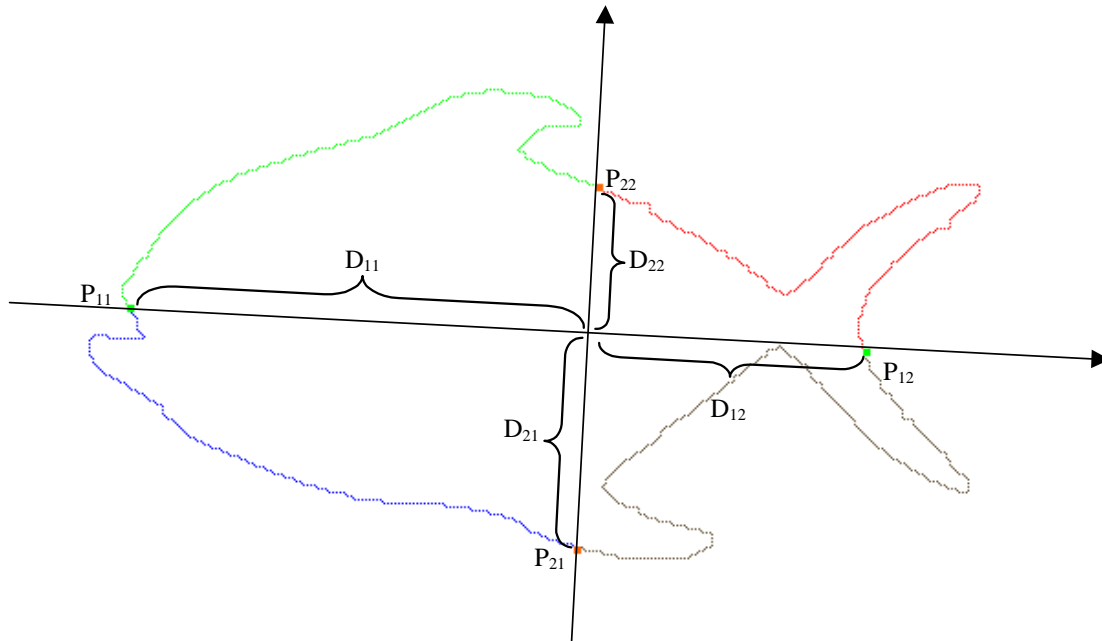


Figura 6.8 – Exemplo ilustrativo da localização dos pontos de começo P_{11} , P_{12} , P_{21} e P_{22} , e das respectivas distâncias, D_{11} , D_{12} , D_{21} e D_{22}

Embora tenham sido usados até aqui sempre quatro vectores de ângulos por descrição, parece ser possível enviar um menor número de vectores de ângulos, tendo os parâmetros acima referidos um papel fundamental na selecção do número de vectores e correspondentes pontos de começo adequados. O envio de um menor número de vectores de ângulos por descrição depara-nos com alguns casos de ambiguidade até agora solucionado pelo algoritmo enviando

sempre quatro vectores de ângulos por descrição que cobrem assim todos os possíveis casos de ambiguidade. Estas ambiguidades resultam do elevado grau de proximidade existente entre as distâncias dos pontos de começo, no mesmo eixo e verticalmente opostos, ao centróide e os momentos do objecto, uma vez que pequenas variações e/ou deformações do objecto podem causar alterações a estes valores calculados computacionalmente. A influência destas alterações assim como de possíveis erros de arredondamento contribuem para estes casos de ambiguidade quando os valores das distâncias e dos momentos são realmente próximos. As expressões abaixo definem o grau de proximidade entre os momentos (M_1 e M_2) e as distâncias (D_{k1} e D_{k2}) em causa com base nos limiares de vizinhança M_L e D_L em torno das suas médias, respectivamente.

$$M_1 \cong M_2 \Rightarrow |M_L| \leq \frac{2M_1}{M_1 + M_2} - 1 \quad \wedge \quad |M_L| \leq \frac{2M_2}{M_1 + M_2} - 1$$

$$D_{k1} \cong D_{k2} \Rightarrow |D_L| \leq \frac{2D_{k1}}{D_{k1} + D_{k2}} - 1 \quad \wedge \quad |D_L| \leq \frac{2D_{k2}}{D_{k1} + D_{k2}} - 1$$

em que os limiares de vizinhança M_L e D_L se encontram definidos no intervalo $[-1,1]$.

De acordo com o grau de proximidade existente entre M_1 e M_2 , e D_{k1} e D_{k2} pode-se então determinar o número de vectores a enviar por descrição através das seguintes condições:

Se $M_1 \cong M_2 \Rightarrow P_{11}$ e P_{21} serão usados como pontos de começo

Se $D_{k1} \cong D_{k2} \Rightarrow P_{k1}$ e P_{k2} serão usados como pontos de começo

Caso Contrário P_{11} será usado como o único ponto de começo

em que o número máximo de vectores a usar na descrição é 4 e o número mínimo é 1, podendo usar-se 1, 2, ou 4 vectores de ângulos. No caso de não existir qualquer tipo de semelhança (proximidade) entre M_1 e M_2 ou entre D_{k1} e D_{k2} , apenas será necessário usar um ponto de começo que neste caso será o ponto de começo mais distante do centróide, ou seja P_{11} . No pior caso, o número de vectores de ângulos a usar será 4 (como já acontecia no método até aqui usado) uma vez que no pior caso existe ambiguidade relativamente às duas propriedades (momentos e distâncias) ter-se-ão que usar tantos vectores quantos os pontos de começo iniciais.

6.7.2 Redefinição da sintaxe e semântica do descritor ACC

Variando o número de vectores de ângulos por descrição segundo o critério apresentado nesta secção, há que adicionar uma nova componente sintáctica ao descritor ACC que indica o número de vectores de ângulos de curvatura usados por descrição. A inclusão desta componente no descritor ACC implica uma alteração ao nível da sintaxe e da semântica do descritor, como era de se esperar. Há que salientar que a escolha do número de *bits* para esta nova componente, 2, considerou os valores mínimo e máximo do número de vectores de ângulos a usar, 1 e 4, numa descrição ACC da forma de um objecto.

Assim sendo, a nova sintaxe e semântica do descritor ACC será:

Ângulos_Curvatura(){	Número de bits
Número_Ângulos	7
Número_Vectores	2
Excentricidade	6
Circularidadde	6
for (x=0; x<Número_Vectores; x++){	
for (k=0; k< Número_Ângulos; k++){	
Ângulos_Curvatura[x][k]	9*Número_Vectores*Número_Ângulos
}	
}	
}	

A semântica para a nova componente do descritor ACC, uma vez que as restantes permanecem inalteradas, é a seguinte:

- **Número_Vectores** – Define o número de vectores de ângulos de curvatura (até um máximo de quatro vectores e um mínimo de um vector) usados por descrição ACC de forma.

6.7.3 Nova análise do desempenho do descritor ACC

Com base na proposta feita para otimizar a dimensão do descritor ACC variando o número de vectores de ângulos a usar por descrição, segue-se uma análise do desempenho do descritor ACC a fim de determinar os valores dos limiares M_L e D_L que melhor alcançam o compromisso entre o número de vectores a usar e o desempenho de procura do descritor.

Tendo em conta que a proximidade entre os momentos M_1 e M_2 e as distâncias D_{k1} e D_{k2} traduzem situações de ambiguidade relativamente ao ponto de começo usado para gerar a descrição da forma, é necessário escolher os valores de teste para os limiares de proximidade dos momentos e das distâncias, M_L e D_L , respectivamente, de modo a poder escolher os seus valores óptimos. Neste caso, ambos os limiares encontram-se definidos num intervalo $[-1,1]$. Sabendo-se *a priori* que para valores de M_L e D_L iguais a 0 o descritor ACC usará em média um vector de ângulos por descrição (diminuindo em cerca de 75% a dimensão do descritor), deduz-se também que será para este caso que apresentará o pior desempenho médio total (uma vez que não resolve os casos de ambiguidade nas formas). No entanto, para valores de M_L e D_L iguais a 1 ou -1 o desempenho do descritor é máximo apesar de dar origem a descrições com maior dimensão (em média quatro vectores de ângulos). Partindo do conhecimento destas regras, ir-se-á avaliar o desempenho do descritor ACC para valores de $|M_L|$ e $|D_L|$ entre 0 e 1 de modo a obter-se o melhor compromisso entre o desempenho de procura e a dimensão do descritor ACC.

Assim sendo, geraram-se resultados para 36 combinações diferentes de valores de M_L e D_L nesse intervalo, nomeadamente para os valores 0, 0.2, 0.4, 0.6, 0.8 e 1. A Tabela 6.13 apresenta os resultados obtidos para estas combinações de valores, segundo as condições de teste definidas para o *core experiment CE-1* do MPEG-7.

A medida de semelhança aqui usada para o descritor ACC foi exactamente a mesma da secção 6.6.2 ou seja a distância *alpha-trimmed average* tendo-se no entanto acrescentado um ciclo de controlo devido à variação do número de vectores por descrição. Este ciclo é controlado através de duas componentes, **Número_Vectores** da forma em base de dados e

Número_Vectores da forma fornecida pelo mecanismo de procura. Assim sendo, combinam-se apenas o número de vectores enviados para cada uma das descrições mesmo que estes não sejam de igual número.

				A1	A2	A	B	C	Desempenho Médio Total	Número Médio de Vectores
Antes da otimização (M _L e D _L = 1)				80.95%	94.05%	87.50%	65.36%	85.00%	79.29%	4.0
M _L	0	D _L	0	76.19%	92.86%	84.53%	61.07%	83.50%	76.36%	1.0
			0.2	79.76%	93.81%	86.79%	62.67%	85.00%	78.15%	3.0
			0.4	80.00%	93.57%	87.00%	63.73%	85.00%	79.00%	3.5
			0.6	80.00%	94.05%	87.00%	64.21%	85.00%	79.00%	3.7
			0.8	80.00%	94.05%	87.00%	64.59%	85.00%	79.00%	3.8
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0
M _L	0.2	D _L	0	77.14%	92.86%	85.00%	61.61%	83.50%	76.70%	1.2
			0.2	80.00%	93.81%	87.00%	62.83%	85.00%	78.00%	3.1
			0.4	80.24%	93.57%	86.91%	63.76%	85.00%	78.55%	3.5
			0.6	80.24%	94.05%	87.15%	64.23%	85.00%	78.79%	3.7
			0.8	80.24%	94.05%	87.15%	64.60%	85.00%	78.92%	3.8
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0
M _L	0.4	D _L	0	77.14%	92.86%	85.00%	61.96%	85.50%	77.49%	1.4
			0.2	80.00%	93.81%	87.00%	63.10%	85.00%	78.00%	3.1
			0.4	80.24%	93.57%	86.91%	63.90%	85.00%	78.60%	3.3
			0.6	80.24%	94.05%	87.15%	64.33%	85.00%	78.82%	3.7
			0.8	80.24%	94.05%	87.15%	64.59%	85.00%	78.91%	3.8
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0
M _L	0.6	D _L	0	77.38%	92.86%	85.12%	62.25%	85.50%	77.62%	1.6
			0.2	80.24%	93.81%	87.03%	63.18%	85.00%	78.40%	3.2
			0.4	80.24%	93.57%	86.91%	63.98%	85.00%	78.63%	3.6
			0.6	80.24%	94.05%	87.15%	64.43%	85.00%	78.86%	3.7
			0.8	80.24%	94.05%	87.15%	64.59%	85.00%	78.91%	3.8
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0
M _L	0.8	D _L	0	78.10%	92.86%	85.48%	62.69%	84.50%	77.56%	1.8
			0.2	80.24%	93.81%	87.03%	63.46%	85.00%	78.50%	3.3
			0.4	80.24%	93.57%	86.91%	64.15%	85.00%	78.69%	3.6
			0.6	80.24%	94.05%	87.15%	64.60%	85.00%	78.92%	3.8
			0.8	80.24%	94.05%	87.15%	64.73%	85.00%	78.96%	3.8
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0
M _L	1.0	D _L	0	78.10%	92.86%	85.48%	63.03%	84.50%	77.67%	2.0
			0.2	80.24%	93.81%	87.03%	63.70%	85.00%	78.58%	3.3
			0.4	80.24%	93.57%	86.91%	64.33%	85.00%	78.74%	3.6
			0.6	80.24%	94.05%	87.15%	64.67%	85.00%	78.94%	3.8
			0.8	80.24%	94.05%	87.15%	64.74%	85.00%	78.96%	3.9
			1.0	80.95%	94.05%	87.50%	65.27%	85.00%	79.26%	4.0

Tabela 6.13: Resultados antes e após a optimização do número de vectores de ângulos enviados em cada descrição em função dos limiares de proximidade $|M_L|$ e $|D_L|$

Dos resultados apresentados na Tabela 6.13 pode constatar-se que o desempenho médio total é inferior ao obtido antes desta alteração, como seria de esperar uma vez que se inclui menos informação na descrição. No entanto, verifica-se que o parâmetro determinante para o número médio de vectores a usar por descrição é o $|D_L|$ uma vez que à medida que este parâmetro aumenta, aumenta também rapidamente o número médio de vectores a serem

usados por descrição e consequentemente o desempenho médio total. Com base nos valores obtidos, conclui-se que os valores de $|M_L|$ e $|D_L|$ que permitem alcançar o melhor compromisso entre o desempenho de procura e a dimensão da descrição são os valores 0.4 e 0, respectivamente para $|M_L|$ e $|D_L|$ porque são para estes valores que o número médio de vectores de ângulos diminui para aproximadamente 1, não ultrapassando o valor médio de vectores de 1.5 representativo de um envio de vectores superior a 1 por descrição. Neste caso o número médio de vectores usados é 1.4 e o desempenho médio total é de, aproximadamente, 77.5%. Neste caso acha-se conveniente sacrificar um pouco o desempenho de procura do descritor face à enorme redução em termos da sua dimensão binária uma vez que o desempenho máximo seria 79.26% para uma média de 4 vectores de ângulos.

A Tabela 6.14 apresenta o número de *bits* usado por descrição ACC antes e após a optimização da dimensão do descritor.

<i>CE-1</i>			ACC (<i>bits</i> por descrição antes de melhoria)		ACC (<i>bits</i> por descrição após melhoria)	
Partes	A	A1	594	594	222.6	222.6
		A2	594		222.6	
	B		601		231.6	
	C		592		220.6	
	Média Total (<i>bits</i>)			595.7		224.9

Tabela 6.14– Número de *bits* usado por descrição ACC, antes e depois da optimização da dimensão do descritor

Como se pode verificar a dimensão do descritor ACC diminui significativamente, cerca de 62.24%. De acordo com estes resultados pode-se dizer que cada vez mais a dimensão do descritor ACC se aproxima da do descritor CSS adoptado pelo MPEG-7.

6.8 Optimização adicional do número de vectores de ângulos para o descritor ACC

Uma vez que na secção anterior se verificou que o melhor compromisso entre o desempenho de procura e a dimensão do descritor ACC correspondia ao uso médio de 1.4 vectores de ângulos por descrição, surge então a ideia de avaliar o desempenho de procura do descritor ACC se se usar sempre um único vector de ângulos por descrição o que evitaria a complexidade associada à determinação dinâmica do número de vectores a usar.

Contudo, se se optar pela geração de um único vector de ângulos por descrição ACC, existe a necessidade de definir qual o critério de selecção para o ponto de começo a usar em todas as descrições. Assim sendo, e uma vez que os quatro pontos de começo até agora usados foram escolhidos com base nos momentos da forma, optou-se por definir um critério que aproveita estes quatro pontos de começo e faz a escolha com base na relação existente entre os pontos de começo e o centróide da forma em questão.

As secções que se seguem definem o critério usado para a selecção do único ponto de começo a usar nas descrições da forma dos objectos quando se usa um único conjunto de vectores de ângulos e apresentam os resultados obtidos para o desempenho de procura segundo as condições de teste definidas para o *core experiment CE-1* do MPEG-7.

6.8.1 Definição do critério de selecção para o único ponto de começo

No capítulo 5 mencionou-se a sugestão de Niblack e Yin [NiYi1995] relativamente à escolha de um só ponto de começo para descrever a forma de um objecto com base nos ângulos de curvatura. A escolha deste ponto (ponto de intersecção entre o contorno e os eixos principais mais próximo do centróide da forma) era mais uma vez baseada nos momentos da forma mas a medida de semelhança então usada não conseguia compensar eventuais ambiguidades relativas aos pontos de começo, como a existência de dois pontos de intersecção sobre o mesmo eixo e cujas distâncias ao centróide são iguais, porque a descrição gerada para a forma exemplo fornecida ao mecanismo de procura usava o mesmo método de descrição (um único vector de ângulos correspondente a um único ponto de começo). O método proposto nesta secção pretende explorar esta ideia, ultrapassando as suas limitações através da melhoria da medida de semelhança usada que tomará a seu cargo a solução dos problemas inerentes à presença de ambiguidades.

Conforme se disse acima, será proposto por esta tese que a escolha do único ponto de começo a usar para a descrição da forma de um objecto se baseie nos momentos da forma em análise. A escolha do único ponto de começo será feita com base nas distâncias dos 4 pontos de começo candidatos (aqueles usados até aqui) relativamente ao centróide da forma, sendo de seguida avaliado o desempenho para quatro critérios diferentes de escolha do único ponto de começo:

- **Critério 1** – O único ponto de começo (ponto 1) é o ponto de intersecção entre o contorno e os eixos principais que fica mais distante do centróide;
- **Critério 2** – O único ponto de começo (ponto 2) é o ponto de intersecção entre o contorno e os eixos principais que fica mais próximo do centróide;
- **Critério 3** – O único ponto de começo é o ponto de intersecção entre o contorno e os eixos principais que fica mais próximo do centróide e situado no eixo perpendicular ao ponto 1;
- **Critério 4** – O único ponto de começo é o ponto de intersecção entre o contorno e os eixos principais que fica mais próximo do centróide e situado no eixo perpendicular ao ponto 2.

Para garantir alguma insensibilidade à rotação que não pode agora ser garantida através dos ângulos associados a 4 pontos de começo, há que alterar a medida de semelhança que passa agora a exigir a geração de quatro vectores de ângulos de curvatura correspondentes a cada um dos quatro pontos de começo anteriormente definidos para a forma exemplo fornecida ao mecanismo de procura. Deste modo garante-se que, mesmo na presença de ambiguidades ou seja de formas que poderão ser muito semelhantes mas que poderão ser declaradas como não semelhantes se forem descritas usando pontos de começo diferentes, o método de procura garante que todos os pontos de começo serão testados (neste caso para a forma exemplo) e consequentemente sempre detectado o vector de ângulos da forma de procura que mais se

assemelha ao vector de ângulos da forma em base de dados mesmo que esta tenha sido descrita usando apenas um vector de ângulos.

Apesar deste novo critério, a expressão usada para o cálculo da distância de semelhança continua ser a mesma do que a usada na secção 6.6.2, no entanto esta apenas considera 4 combinações de vectores de ângulos de curvatura, ou seja,

$$D_{final} = \min_{1 \leq c \leq 4} \left\{ \frac{1}{N - 2 \cdot \lfloor (0.5 - M_L) \cdot (N - 1) \rfloor} \sum_{n=\lfloor (0.5 - M_L) \cdot (N - 1) \rfloor + 1}^{N - \lfloor (0.5 - M_L) \cdot (N - 1) \rfloor} D_{DA}[n] + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\}$$

6.8.2 Nova análise do desempenho do descritor ACC

Tendo-se definido quatro critérios para a escolha do único ponto de começo a ser usado nas descrições e avaliado o desempenho de procura do descritor ACC associado a cada um desses critérios, vão-se agora analisar os resultados obtidos. A Tabela 6.15 apresenta os desempenhos médios de procura conseguidos para cada um dos quatro critérios de selecção do ponto de começo previamente definidos.

	A			B	C	Desempenho Médio Total
	A1	A2	A			
Critério 1	76.19%	92.86%	84.53%	61.09%	83.00%	76.21%
Critério 2	76.67%	92.86%	84.76%	57.39%	76.00%	72.72%
Critério 3	74.76%	90.71%	82.74%	57.71%	85.00%	75.15%
Critério 4	77.62%	93.57%	85.60%	56.19%	83.00%	74.93%

Tabela 6.15 – Desempenho médio de procura do descritor AAC usando os quatro critérios de selecção do ponto de começo propostos

Os resultados apresentados indicam que o melhor desempenho de procura é conseguido para o critério de selecção 1, com um desempenho de 76.21%; no entanto, este desempenho médio é inferior ao desempenho conseguido na secção anterior, 77.49%, como seria de esperar, uma vez que se usava então um número de vectores variável por descrição, em média um pouco superior ao aqui usado (1.4). Se se considerar que a perda de desempenho aqui verificada de cerca de 1.28% é pouco significativa comparativamente ao que se ganha em termos de dimensão do descritor (redução média de 0.4 vectores por descrição ou seja aproximadamente 25 ângulos o que corresponde a 150 *bits*), então deverá optar-se pelo uso simples de um único ponto de começo escolhido com base no critério 1 anteriormente definido. A Tabela 6.16 apresenta as consequências desta opção relativamente à dimensão do descritor ACC (114 *bits* para o descritor CSS do MPEG-7). Há no entanto que salientar que para o cálculo dos valores na tabela seguinte já se tirou da sintaxe do descritor ACC a componente *Número_Vectores* que agora já não é precisa.

<i>CE-1</i>			ACC (<i>bits</i> por descrição antes da melhoria)		ACC (<i>bits</i> por descrição após melhoria)	
Partes	A	A1	594	594	162	162
		A2	594		162	
	B		601	169		
	C		592	160		
	Média Total (<i>bits</i>)			595.7		163.7

Tabela 6.16 – Número médio de *bits* usado por descrição ACC, antes e depois da optimização da dimensão do descritor

Os resultados aqui apresentados mostram que esta alteração do descritor ACC permitiu diminuir a dimensão do descritor em cerca de 72.5% à custa de uma perda de desempenho de 3.08% relativamente ao desempenho médio usando-se sempre 4 vectores de ângulos e 76.21% usando-se em média 1 vector de ângulos.

6.9 Optimização do número de *bits* para cada uma das componentes do descritor ACC

As alterações introduzidas na secção anterior permitiram que a dimensão do descritor ACC diminuísse cerca de 72.5% com uma perda de desempenho de apenas 3%. Esta diminuição foi alcançada à custa do decréscimo do número de vectores de ângulos de curvatura gerados por descrição não se tendo ainda tentado optimizar a dimensão do descritor ACC em função do número de *bits* usados para cada uma das suas componentes, de modo a conseguir-se um ainda melhor compromisso entre a dimensão do descritor e o seu desempenho médio total.

Neste contexto, as componentes associadas aos parâmetros geométricos, *excentricidade* e *circularidade*, manterão o número de *bits* adoptado e optimizado para o descritor CSS do MPEG-7, ou seja 6 *bits*. A dimensão da componente *Ângulo_Curvatura* será avaliada nesta secção por ser esta aquela que mais influencia a dimensão do descritor assim como o seu desempenho. Neste sentido, esta secção terá como principal objectivo analisar o desempenho de procura do descritor ACC em função do número de *bits* usados para esta componente ou seja em função da quantificação usada para os valores dos ângulos de curvatura, tudo isto nas condições do *core experiment CE-1* do MPEG-7.

6.9.1 Nova análise do desempenho do descritor ACC

A análise do desempenho de procura do descritor ACC em função do número de *bits* usados para a componente *Ângulo_Curvatura* será feita considerando-se uma variação entre 1 e 10 *bits* por ângulo. A Tabela 6.17 e a Figura 6.9 apresentam os resultados obtidos nas condições de avaliação atrás definidas.

Número de bits	A			B	C	Desempenho Médio Total
	A1	A2	A			
10	76.43%	92.86%	84.65%	61.12%	83.00%	76.25%
9	76.19%	92.86%	84.53%	61.09%	83.00%	76.21%
8	76.43%	92.86%	84.65%	61.05%	83.00%	76.23%
7	76.19%	92.62%	84.41%	61.04%	83.00%	76.15%
6	76.19%	92.62%	84.41%	61.00%	84.00%	76.47%
5	75.24%	92.86%	84.05%	60.60%	84.00%	76.22%
4	75.48%	90.95%	83.22%	57.45%	80.00%	73.56%
3	71.43%	88.33%	79.88%	49.73%	53.50%	61.04%
2	64.76%	71.67%	68.22%	41.94%	45.00%	51.72%
1	55.24%	29.05%	42.15%	37.25%	56.50%	45.30%

Tabela 6.17 – Desempenho médio de procura do descritor ACC em função do número de *bits* para a componente *Ângulo_Curvatura*

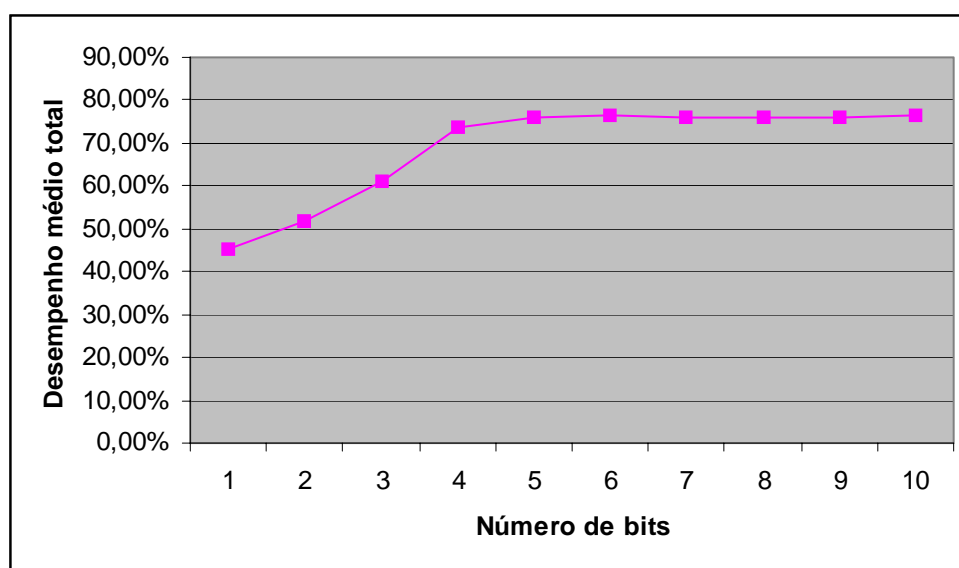


Figura 6.9 – Desempenho médio do descritor ACC em função do número de *bits* para a componente *Ângulo_Curvatura*

Com base nos resultados apresentados, pode-se concluir que para um número de *bits* superior a 4 o desempenho médio de procura do descritor ACC tende a estabilizar um pouco acima dos 76%. Portanto, com um número de *bits* por ângulo de curvatura igual a 5 consegue-se o melhor compromisso entre o desempenho de procura e a dimensão do descritor, diminuindo-se o desempenho médio total em apenas 0.25% relativamente ao mais elevado valor obtido (76.47% para um número de *bits* igual a 6) e diminuindo-se a dimensão do descritor em cerca de 40%, como se pode constatar através dos dados apresentados na Tabela 6.18. Note-se que a evolução do desempenho com o número de *bits* não é perfeitamente crescente, ainda que as

variações em relação a um crescimento monotónico sejam pequenas, em virtude da variação da localização dos extremos dos intervalos de quantificação.

De acordo com os dados apresentados na Tabela 6.18, esta diminuição da dimensão do descritor ACC torna-o competitivo em termos de dimensão com o descritor CSS ainda que o mesmo não se passe infelizmente em termos do desempenho de procura (89.7% versus 76.2%).

<i>CE-1</i>			ACC (<i>bits</i> por descrição antes da melhoria)		ACC (<i>bits</i> por descrição após a melhoria)		<i>CSS</i>	
Partes	A	A1	162	162	98	98	111	111.5
		A2	162		98		112	
	B		169		105		114	
	C		160		96		117	
	Média Total (Bits)			163.7		96.7		114

Tabela 6.18 – Número de *bits* usado por descrição ACC, antes e depois da optimização da dimensão da componente *Ângulos_Curvatura*

6.10 Proposta final para o descritor ACC

Após o estudo exaustivo do descritor baseado nos ângulos de curvatura do contorno feito neste capítulo, pode concluir-se que o melhor compromisso conseguido através de todas as alterações propostas conduziu a um desempenho médio total de 76.22% e a uma dimensão média de 96.7 *bits* por descrição para o conteúdo correspondente ao *core experiment CE-1* do MPEG-7. Ainda que fosse possível alcançar outros compromissos com um desempenho de procura um pouco maior à custa de uma dimensão também um pouco maior, o compromisso acima apresentado pareceu o mais razoável.

Assim sendo as secções seguintes definem a proposta final para o descritor ACC, nomeadamente em termos de sintaxe e semântica, assim como a medida de semelhança proposta para o processo de procura.

6.10.1 Definição final da sintaxe e semântica do descritor ACC

A sintaxe e a semântica correspondentes ao descritor ACC finalmente proposto são as seguintes:

AngulosCurvatura({	Número de bits
Excentricidade	6
Circularidade	6
for (k=0; k<16; k++){	
Ângulo_Curvatura[k]	5*16
}	
}	

- **Excentricidade** – Especifica as propriedades de alongamento do contorno da forma do objecto; a excentricidade é definida no intervalo [0,10] e representada usando 6 *bits*;
- **Circularidade** – Especifica as propriedades de circularidade do contorno do objecto, ou seja quão próximo este está de uma circunferência; a circularidade é definida no intervalo [0,1] e representada usando 6 *bits*;
- **Ângulo_Curvatura** – Especifica cada um dos ângulos de curvatura do contorno do objecto em questão; cada um destes ângulos é o ângulo entre um vector que une dois pontos representativos consecutivos do contorno e o eixo principal onde se situa o ponto de começo usado (ver secção 5.3.2.1), sendo os pontos representativos do contorno calculados a partir do contorno percorrido no sentido contrário ao dos ponteiros do relógio. Os ângulos de curvatura são definidos no intervalo [0°, 360°] e codificados com 5 *bits*.

6.10.2 Medida de semelhança final para o descritor ACC

A medida de semelhança finalmente proposta para o descritor ACC baseia-se em duas distâncias: a distância *alpha-trimmed average*, D_{ATA} , e a distância para os parâmetros geométricos, D_{geom} . Para o cálculo desta medida é importante relembrar que os métodos de extracção dos ângulos propostos para as formas a descrever e para a forma exemplo são diferentes. Assim, enquanto as formas nas bases de dados são descritas usando um único vector de ângulos de curvatura por descrição de forma (extraídos no sentido contrário aos ponteiros do relógio), escolhendo como ponto de começo o ponto de intersecção entre o contorno e os eixos principais que fica mais distante do centróide, a forma exemplo fornecida ao mecanismo de procura é descrita usando quatro vectores de ângulos, um para cada um dos quatro pontos de começo definidos na secção 5.3.2.1. Para cada um destes quatro vectores de ângulos é ainda calculado o vector simétrico (ver secção 6.3.1.1) de modo a compensar possíveis transformações de simetria.

Tendo em conta estas regras para o método de procura, a medida de semelhança final, D_{final} , é então calculada através da seguinte expressão:

$$D_{final} = \min_{1 \leq c \leq 8} \{ D_{ATA} + D_{geom} \}$$

$$D_{final} = \min_{1 \leq c \leq 8} \left\{ D_{ATA} + \left(w_1 \times \frac{|E_{I_c} - E_{B_c}|}{\max(E_{I_c}, E_{B_c})} + w_2 \times \frac{|C_{I_c} - C_{B_c}|}{\max(C_{I_c}, C_{B_c})} \right) \right\}$$

onde N é o número de ângulos de curvatura, ou seja 16, usados por descrição e M_L é o limiar de vizinhança, 40%, considerado para o cálculo da medida de semelhança, A_B é o vector de ângulos das formas descritas na base de dados, A_I é o vector de ângulos da forma fornecida ao mecanismo de procura, E_B e E_I , e C_B e C_I são os valores de excentricidade e circularidade para as descrições disponíveis nas bases de dados e aquela fornecida como exemplo ao mecanismo de procura, respectivamente, e finalmente w_1 e w_2 são parâmetros associadas ao peso atribuído a cada um dos parâmetros geométricos em questão. Os valores propostos para w_1 e w_2 , são os mesmos que para o descritor CSS, ou seja 0.7 e 0.8, respectivamente.

6.11 Conclusões

Este capítulo apresentou um conjunto de alterações ao descritor ACC e à medida de semelhança associada com vista a melhorar o seu desempenho de procura e a reduzir a sua dimensão binária, aproximando-o do descritor ‘adversário’ escolhido para a norma MPEG-7 ou seja o descritor CSS. As alterações propostas foram sempre avaliadas nas condições definidas pelo *core experiment CE-1* do MPEG-7.

Analisando a evolução dos resultados obtidos com as sucessivas alterações ao descritor ACC apresentados na Tabela 6.19, verifica-se que cada uma das alterações propostas apresentou uma mais-valia significativa em termos do desempenho de procura ou em termos da dimensão binária do descritor ACC ainda que não necessária e simultaneamente nas duas vertentes.

	A			B	C	Desempenho Médio Total	Número bits por Descrição
	A1	A2	A				
Primeira fase de desenvolvimento do MPEG7	66.67%	91.91%	79.29%	54.14%	63.50%	65.64%	2390.6
Inclusão da excentricidade e circularidade	64.76%	91.67%	78.22%	57.41%	63.50%	66.38%	2323.7
Obtenção da insensibilidade à simetria	64.76%	93.81%	79.29%	64.85%	83.00%	75.71%	2323.7
Optimização do número de ângulos usados	80.48%	93.81%	87.15%	65.67%	84.50%	79.10%	595.7
Adopção da distância <i>alpha-trimmed average</i> como medida de semelhança	80.95%	94.05%	87.50%	65.36%	85.00%	79.29%	595.7
Adopção de um único vector de ângulos	76.19%	92.86%	84.53%	61.09%	83.00%	76.21%	163.7
Optimização do número de bits para cada componente do descritor	75.24%	92.86%	84.05%	60.60%	84.00%	76.22%	96.7

Tabela 6.19: Resumo dos resultados obtidos para as alterações efectuadas ao descritor ACC ao longo deste capítulo

Com base nas alterações introduzidas no descritor ACC ao longo deste capítulo, conseguiu-se melhorar o seu desempenho médio total em cerca de 10.6% (de 65.6% para 76.2%), podendo-se considerar uma boa melhoria face ao desempenho inicial obtido. Quanto à dimensão média do descritor por descrição, em *bits*, esta passou de 2390.6 para 96.7 *bits*, ou seja uma

melhoria de 96%. No entanto, o desempenho de procura médio final poderia ainda ser melhorado, nomeadamente no que diz respeito às partes A1 e B do *core experiment CE-1* do MPEG-7 uma vez que os desempenhos parciais para estas partes estão longe dos valores obtidos para outras partes. Uma vez que o conteúdo incluído na parte B do *core experiment CE-1* apresenta algumas deformações face à forma fornecida ao mecanismo de procura, poder-se-ia pensar na melhoria do desempenho através da extracção dos ângulos de curvatura do contorno para um polígono aproximado da forma do objecto. Este polígono conteria os pontos de contorno mais importantes representando de forma mais independente as deformações do contorno do objecto em análise.

A comparação entre os descritores de forma ACC e CSS em termos do desempenho de procura médio total (76.2% versus 89.8%) e da dimensão (96.7 versus 114 *bits*) do descritor permite concluir que a melhoria alcançada neste capítulo em termos do desempenho de procura para o descritor ACC deixou-o ainda algo distante do descritor MPEG-7 CSS ainda que o mesmo não se possa dizer em termos da dimensão do descritor onde o descritor ACC acabou por superar claramente o descritor CSS. Se se levar em conta que o descritor CSS foi optimizado ao longo de muitos meses por várias equipas de investigação ao nível mundial, poderá concluir-se que os resultados aqui obtidos não deixam de ser muito satisfatórios.

Capítulo 7

Conclusões

Como foi dito no capítulo 1, algumas das tendências mais importantes de hoje em termos de tecnologia audiovisual como a interactividade, a personalização de conteúdos e o acesso universal, estão intimamente ligadas a um modelo de representação da informação audiovisual baseado no conceito de ‘objecto’. Em termos visuais, um objecto é representado através da sua informação de textura (luminância e crominância) e através da sua informação de forma. A informação de forma aparece pois como a grande novidade associada à representação, codificação ou descrição, de objectos uma vez que a informação de textura já era usada nos modelos de representação anteriores baseados em tramas rectangulares.

Devido à grande importância dos conceitos de ‘objecto’ e de ‘forma’ no mundo visual e também à sua relativa novidade, escolheu-se como tema central desta tese a descrição de formas visuais, em íntima relação com o que está a ser desenvolvido nesta área no contexto da norma MPEG-7.

Neste sentido e de acordo com os grandes objectivos desta tese (definidos no capítulo 1) pôde-se verificar e reafirmar, durante o desenvolvimento desta tese, a enorme importância dos conceitos de ‘objecto’ e ‘forma’ em virtude das imensas possibilidades que surgem em termos de novas funcionalidades e consequentemente novas e melhoradas aplicações. Este facto tem-se também consubstanciado no crescimento do investimento em termos de investigação e desenvolvimento que têm vindo a ser feitos em áreas que possam beneficiar da força dos conceitos acima referidos. Resumidamente, cada capítulo desta tese contribuiu para rever, estudar, aplicar e desenvolver conceitos e métodos relacionados com a característica de forma.

No capítulo 2 foi feito um estudo bibliográfico com o objectivo de definir os principais requisitos relevantes para a exploração de parâmetros de descrição de informação de forma.

Na sequência deste estudo foram apresentados de modo pormenorizado os principais parâmetros de descrição de forma e medidas de semelhança. Estes parâmetros foram divididos em dois grandes tipos: parâmetros de forma baseados no contorno e parâmetros de forma baseados em regiões. Desta revisão bibliográfica relativamente aos parâmetros de forma actualmente disponíveis, concluiu-se que entre as suas propriedades ideais está a insensibilidade a transformações geométricas como a translação, rotação, mudança de escala e ponto de começo. Para além disso, há que definir medidas de semelhança adequadas para os vários parâmetros de tal modo que permitam a avaliação eficaz da noção de semelhança em termos de forma usando os parâmetros de descrição em questão.

O capítulo 3 começou por definir uma arquitectura genérica para um mecanismo de descrição e procura de informação visual tendo-se de seguida debruçado sobre os principais mecanismos de descrição e procura disponíveis na *Internet* e que usam apenas características visuais de baixo-nível ou seja características automaticamente extraíveis da informação visual. Da análise comparativa entre a arquitectura proposta e os mecanismos existentes, concluiu-se que um mecanismo de descrição e procura que pretenda descrever e procurar conteúdo visual através da característica de forma deverá incluir ferramentas automáticas para a extracção de parâmetros de forma, adoptar medidas de semelhança que traduzam de maneira adequada a semelhança entre a informação visual descrita e a procurada pelo utilizador do mecanismo e finalmente, possuir interfaces que exijam o mínimo de esforço por parte do utilizador durante o processo de procura.

Com base nestas conclusões, o capítulo 4 apresentou o mecanismo de descrição e procura de informação visual desenvolvido no âmbito desta tese, explorando principalmente a característica de forma dos objectos de vídeo. A aplicação desenvolvida permitiu a análise e a comparação do desempenho dos descritores de forma estudados no contexto do processo de desenvolvimento da norma MPEG-7 assim como de outros descritores alternativos de descrição da forma.

Usando a aplicação desenvolvida, o capítulo 5 apresentou pormenorizadamente, e de acordo com o processo de trabalho que o MPEG adoptou para o desenvolvimento da norma MPEG-7, os principais passos dados com vista à especificação desta norma, com especial ênfase nos parâmetros de forma propostos ao MPEG-7 como candidatos a descritores. Daqui resultou um conhecimento aprofundado das vantagens/forças e desvantagens/fraquezas dos descritores de forma propostos ao MPEG-7 permitindo que estes conhecimentos fossem posteriormente usados para o melhoramento de outros parâmetros de descrição da forma.

O capítulo 6 apresentou e descreveu o trabalho realizado nesta tese no sentido de melhorar o descritor de forma baseado nos ângulos de curvatura, de modo a que este apresentasse um desempenho semelhante aos descritores incluídos na norma MPEG-7. O descritor de forma baseado nos ângulos de curvatura do contorno foi inicialmente implementado de acordo com o método de extracção sugerido pelos seus proponentes originais, apresentando um desempenho médio que se encontrava muito longe dos desempenhos médios obtidos pelos outros descritores seus ‘adversários’. Consequentemente, e por este descritor aparentar claras ‘debilidades’ relativamente aos descritores em competição no MPEG-7, investiu-se na sua melhoria, quer em termos de sintaxe e semântica, quer em termos de medidas de semelhança, tendo sempre presentes as propriedades dos descritores MPEG-7 adoptados que os levaram a superar as debilidades que o descritor baseado nos ângulos de curvatura do contorno ainda apresentava. Deste modo, conseguiu-se melhorar significativamente o descritor ACC tornando-o inclusive competitivo com o descritor CSS adoptado pela norma MPEG-7 em

termos de dimensão porque relativamente ao seu desempenho este precisa de ainda algumas melhorias especialmente no que diz respeito à parte B da CE-1.

Apesar de todo o trabalho desenvolvido em torno dos descritores de forma resta ainda muito por ser feito nesta área. Algumas das sugestões de trabalho futuro que aqui ficam, estão essencialmente relacionadas com a aplicação desenvolvida baseada numa arquitectura genérica de descrição e procura. Apesar da inovação que esta aplicação apresenta relativamente aos mecanismos existentes na *Internet* sobretudo por se basear nos descritores de forma e cor especificados pela norma MPEG-7 e logo de poder beneficiar das vantagens inerentes à adopção de uma norma mundial, esta aplicação poderia ainda ser melhorada através da inclusão de:

- ♦ Ferramentas que permitam gerar descrições para vídeo e não só para imagens estáticas, de modo a poderem ser efectuadas descrições e procuras também baseadas na característica de movimento;
- ♦ Restantes elementos normativos da norma MPEG-7, nomeadamente os esquemas de descrição, conseguindo-se um mecanismo de descrição e procura mais potente através da combinação no espaço e no tempo de vários descritores;
- ♦ Todos os descritores textuais especificados pela norma MPEG-7;
- ♦ Ferramentas de descrição para voz e áudio e procuras combinadas de áudio e vídeo;
- ♦ Mapeamento semântico ou seja mapeamento entre parâmetros de baixo nível e descritores de alto nível como forma de poder gerar automaticamente descrições com maior poder semântico, potenciando assim também procuras semânticas de conteúdo audiovisual sem o envolvimento de anotadores humanos;
- ♦ Novos tipos de interfaces de procura tais como procuras semânticas abrangendo um bom número de domínios de aplicação e de necessidades dos utentes; e
- ♦ Medidas de semelhança mais eficientes, permitindo procuras mais elaboradas e precisas, nomeadamente de modo a permitir procuras usando vários descritores.

Espera-se que estas sugestões possam futuramente vir a ser implementadas, resultando numa aplicação mais potente, flexível e robusta, garantindo contudo sempre total interoperabilidade relativamente às descrições geradas por outros mecanismos de descrição conformes com a norma MPEG-7.

Referências

- [AnGa1994] M. Antonini, T. Gaidon, P. Mathieu e M. Barlaud, “Wavelet Transform and Image Coding”, *Advances in Image Communication*, Elsevier, Vol. 5, 1994.
- [AsBa1995] J. Ashley, R. Barber, M. Flickner, J. Hafner, D. lee, W. Niblack e D. Petkovic, “Automatic and Semi-Automatic Methods for Image Annotation and Retrieval in QBIC”, *IBM Research Report*, Abril 1995.
- [Bimb1999] A. del Bimbo, “Visual Information Retrieval”, *Morgan Kaufmann Publishers, Inc.*, 1999.
- [BoAt2000a] M. Bober, J. Atkinson, W. Price, “Results of the CE on the Contour Based Shape Descriptor focusing on the Retrieval Accuracy versus Descriptor Size”, *ISO/IEC JTC1/SC29/WG11/M5871*, Noordwijkerhout, Março 2000.
- [BoAt2000b] M. Bober e J. Atkinson, “Report on further Optimization of the Contour Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M6039*, Geneva, Maio 2000.
- [Bobe1999a] M. Bober, *ISO/IEC JTC1/SC29/WG11/P320*, Lancaster, Fevereiro 1999.
- [Bobe1999b] M. Bober, “Performance Evaluation of the CSS Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M4731*, Vancouver, Julho 1999.
- [Bobe1999c] M. Bober, “Improvements to the Similarity Retrieval Rate of the XM Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5166*, Melbourne, Outubro 1999.
- [Bobe1999d] M. Bober, “Retrieval Performance of the XM Shape Descriptors under Partial Occlusion”, *ISO/IEC JTC1/SC29/WG11/M5167*, Melbourne, Outubro 1999.

- [Bobe2000e] M. Bober, "Report on results of the CE-5 on Contour-based Shape", *ISO/IEC JTC1/SC29/WG11/M6293*, Beijing, Julho 2000.
- [BoKi1999] M. Bober, J. Kim, H. Kim, Y. Kim, W. Kim e K. Mueller, "Summary of the Results in Shape Descriptor Core Experiment", *ISO/IEC JTC1/SC29/WG11/M4869*, Vancouver, Julho 1999.
- [CaHa1987] G. L. Cash e M. Hatamian, "Optical Character Recognition by the Method of Moments", *Computer Vision, Graphics and Image Processing*, N° 39, 1987, pp. 291-310.
- [ChCh1997] S.F. Chang, W. Chen, H.J. Meng, H. Sundaram e D. Zhong, "VideoQ: An Automated Content based Video Search System Using Visual Cues", 1997.
- [ChEl1998] S. Chang, A. Eleftheriadis e R. McClintock, "Next-Generation Content Representation, Creation, and Searching for New-Media Applications in Education", *Proceedings of the IEEE*, Vol. 86, N°5, Maio 1998.
- [ChRo1988] T. Cho-Huak e T. Roland, "On Image Analysis by the Method of Moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, N° 4, 1988, pp. 496-513.
- [Excalibur] "Excalibur Visual RetrievalWare", <http://vrw.excalib.com:8015/cst>.
- [Ferr1995] D. Ferreira, "Classificação e Codificação de Contornos", Tese de Mestrado, Universidade Técnica de Lisboa, Instituto Superior Técnico, Lisboa, Maio 1995.
- [Finn1996] R. Finn, "Querying by Image Content", *IBM Research Report*, N°3, 1996.
- [FoDa1994] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes e R.L. Phillips, "Introduction to Computer Graphics", *Addison-Wesley Publishing Company, Inc.*, 1994.
- [GaGr1992] S.R. Garber e M.B. Grunes, "The Art of Search: A Study of Art Directors", *Proceedings CHI'92 Human Factors in Computer Systems*, Monterey, CA: ACM, 1992, pp.157-163.
- [GoWo1993] R. C. Gonzalez e R. E. Woods, "Digital Image Processing", *Addison-Wesley Publishing Company*, 1993.
- [IBMRe1999] IBM, "Technical Summary of Turning Angle Shape Descriptors Proposed by IBM", *ISO/IEC JTC1/SC29/WG11/P162*, Lancaster, Fevereiro 1999.
- [Hall1987] E. Hall, *Computer Image Processing and Recognition*, Academic Press, n° 39, pp. 291-310, 1987.
- [HeKa2000] J. Heuer, A. Kaup, "Results on Core Experiment on the Contour based Shape Descriptor", *ISO/IEC JTC1/SC29/WG11/M5905*, Noordvijkerhout, Março 2000.
- [Jain1989] A. K. Jain, "Fundamentals of Digital Image Processing", *Prentice-Hall International Editions*, 1989.
- [KhHo1990] A. Khotanzad e Y. H. Hong, "Invariant Image Recognition by Zernike Moments", *IEEE Transactions on Pattern and Machine Intelligence*, Vol. 12, N° 5, Julho 1990.

[KiKi1999a] W. Kim e Y. Kim, “A Region-Based Shape Descriptor using Zernike Moments”, 1999.

[KiKi1999b] W. Kim, “A Rotation Invariant Geometric Shape Descriptor using Zernike Moments”, *ISO/IEC JTC1/SC29/WG11/P687*, Lancaster, Fevereiro 1999.

[KiKi1999c] W. Y. Kim e Y. S. Kim, “Shape Descriptor Based on Multi-Layer Eigen Vector”, *ISO/IEC JTC1/SC29/WG11/P517*, Lancaster, Fevereiro 1999.

[KiKi1999d] H. K. Kim e J. D. Kim, “Region-Based Shape Descriptor Invariant to Rotation, Scale and Translation”, Número Especial sobre MPEG-7, *Signal Processing: Image Communication*, Processing: Image Communication, Vol. 16, Nº 1-2, pp. 87-93, 2000.

[KiKi1999e] J. Kim and H. Kim, “Result of MPEG-7 Shape Core-Experiments: Multi-Layer Eigenvector (P517)”, *ISO/IEC JTC1/SC29/WG11/M4793*, Vancouver, Julho 1999.

[KiKi1999f] W. Kim, Y. Kim e D. Sim, “Modified Zernike Shape Descriptor in XM for Enhancing Retrieval Accuracy and Adding New Functionality”, *ISO/IEC JTC1/SC29/WG11/M5010*, Melbourne, Outubro 1999.

[KiKi1999g] W. Kim e Y. Kim, “A new region-based shape descriptor: The ART (Angular Radial Transform) Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5472*, Maui, Dezembro 1999.

[KiKi1999h] W. Kim e Y. Kim, “Cross-verification results of Modified Zernike Moments”, *ISO/IEC JTC1/SC29/WG11/M5473*, Maui, Dezembro 1999.

[KiKi1999i] H. Kim, J. Kim, N. Kim e D. Sim, “Mini-Core Experiment Results - Modified Zernike Moments Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5487*, Maui, Dezembro 1999.

[KiKi1999j] H. Kim e J. Kim, “Cross-verification results of ART Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5609*, Maui, Dezembro 1999.

[KiKi1999k] W. Kim, Y. Kim, Y. Kim e D. Sim, “The Detailed Description of ART (Angular Radial Transform) Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5977*, Geneva, Maio 1999.

[KiKi2000l] W. Kim, Y. Kim, D. Sim e Y. Kim, “Results of the CE on the Region-based Shape Descriptors in Terms of the Retrieval Accuracy and Descriptor Size”, *ISO/IEC JTC1/SC29/WG11/M5979*, Geneva, Maio 2000.

[Kim2000a] M. Kim, “Report for Cross-verification Results on a Region-based Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M5862*, Noordwijkerhout, Março 2000.

[Kim2000b] M. Kim, “Cross-verification Results of Region-based Shape Descriptors”, *ISO/IEC JTC1/SC29/WG11/M6068*, Geneva, Maio 2000.

[Ma1997] W.Y. Ma, “NETRA: A Toolbox for Navigating Large Image Databases”, *Tese de Doutoramento*, Cap.7, Junho 1997.

<http://vivaldi.ece.ucsb.edu/users/wei/mypapers/thesis.html>.

[MARS] Multimedia Analysis and Retrieval System (MARS), <http://www-db.ics.uci.edu/pages/demos/index.shtml>.

[MoAb1999] F. Mokhtarian e S. Abbasi, "Shape-Based Indexing using Curvature Scale Space with Affine Curvature", *Proc. of the First European Workshop on Content-Based Multimedia Indexing*, IRIT, Toulouse, pp. 255-262, Outubro 1999.

[MPEG1-1993] ISO/IEC 11172, "Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s", 1993

[MPEG2-1996] ISO/IEC 13818, "Information technology - Generic coding of moving pictures and associated audio information", 1996

[MPEG4-1999] ISO/IEC 14496, "Information technology - Coding of audio-visual objects", 1999

[MPEG7-App1998] Requirements Group, "MPEG-7 Applications Document V.7", *ISO/IEC JTC1/SC29/WG11/N2462*, Atlantic City, Outubro 1998.

[MPEG7-CDVisual2000] L. Cieplinski, M. Kim, J. Ohm, M. Pickering e A. Yamada, Video Group, "CD 15938-3 MPEG-7 Multimedia Content Description Interface – Part 3 Visual", *ISO/IEC JTC1/SC29/WG11/W3703*, La Baule, Outubro 2000.

[MPEG7-Cont1998] Requirements Group, "MPEG-7: Context and Objectives", *ISO/IEC JTC1/SC29/WG11/N2460*, Atlantic City, Outubro 1998.

[MPEG7-DCE1999a] S. Jeannin e M. Bober, "Description of CE for MPEG-7 Motion/Shape", *ISO/IEC JTC1/SC29/WG11/N2690*, Seoul, Março 1999.

[MPEG7-DCE1999b] S. Jeannin e M. Bober, "Description of CE for MPEG-7 Motion/Shape", *ISO/IEC JTC1/SC29/WG11/N3099*, Maui, Dezembro 1999.

[MPEG7-DCE2000a] S. Jeannin e M. Bober, "Description of CE for MPEG-7 Motion/Shape", *ISO/IEC JTC1/SC29/WG11/N3319*, Noordwijkerhout, Março 2000.

[MPEG7-DCE2000b] S. Jeannin e M. Bober, "Description of CE for MPEG-7 Motion/Shape", *ISO/IEC JTC1/SC29/WG11/N3397*, Geneva, Junho 2000.

[MPEG7-Dev1999] Requirements Group, "MPEG-7 Development Process", *ISO/IEC JTC1/SC29/WG11/N2999*, Melbourne, Outubro 1999.

[MPEG7-Eva1998] Requirements Group, "MPEG-7 Evaluation Process Document", *ISO/IEC JTC1/SC29/WG11/N2463*, Atlantic City, Outubro 1998.

[MPEG7-DDL2000] MDS, Video e Audio Group, "Guidelines for DDL representation of MPEG-7 DSs and Ds ver.2.0", *ISO/IEC JTC1/SC29/WG11/N3625*, La Baule, Outubro 2000.

[MPEG7-Req1999] Requirements Group, "MPEG-7 Requirements Document V.10", *ISO/IEC JTC1/SC29/WG11/N2996*, Melbourne, Outubro 1999.

[MPEG7-SWINT2000] Multimedia Description Schemes Group, "Workplan for DS Software Integration", *ISO/IEC JTC1/SC29/WG11/N3249*, Noordwijkerhout, Março 2000.

[MPEG7-VWD2000a] Video group, “Visual Working Draft Version 3.0”, *ISO/IEC JTC1/SC29/WG11/N3399*, Geneva, Junho 2000.

[MPEG7-VWD2000b] Video group, “Visual Working Draft Version 4.0”, *ISO/IEC JTC1/SC29/WG11/N3399*, Beijing, Julho 2000.

[MPEG7-XM1999] S. Jeannin, L. Cieplinski, J. Ohm e M. Kim, “MPEG-7 Visual Part of eXperimentation Model Version 3.0”, *ISO/IEC JTC1/SC29/WG11/N2931*, Melbourne, Outubro 1999.

[MPEG7-XM2000c] S. Jeannin, L. Cieplinski, J. Ohm e M. Kim, “MPEG-7 Visual Part of eXperimentation Model Version 6.0”, *ISO/IEC JTC1/SC29/WG11/N3398*, Geneva, Junho 2000.

[MuOh1999a] K. Muller e J. Ohm, “Contour Description using Wavelets”, *WIAMIS '99*, Berlim, Maio 1999.

[MuOh1999b] K. Muller e J. Ohm, “Descriptor for Arbitrarily Shaped Objects”, *ISO/IEC JTC1/SC29/WG11/P567*, Lancaster, Fevereiro 1999.

[MuOh1999c] K. Muller e J. Ohm, “Descriptor for Arbitrarily Shaped Objects”, *ISO/IEC JTC1/SC29/WG11/P568*, Lancaster, Fevereiro 1999.

[MuOh1999d] K. Muller e J. Ohm, “Results on MPEG-7 Core Experiment Shape-1 (for P567 simple shapes)”, *ISO/IEC JTC1/SC29/WG11/M4740*, Vancouver, Julho 1999.

[MuPa1999] A. Murching, E. Pacquet, T. Naveen e A. Tabatabai, “Binding Box Based Shape Descriptor for 2D and 3D Visual Objects”, *ISO/IEC JTC1/SC29/WG11/N4438*, Seoul, Março 1999.

[NETRA] “NETRA: A Content Based Image Retrieval System”, <http://maya.ece.ucsb.edu/Netra/>.

[NiYi1995] W. Niblack e J. Yin, “A Pseudo-Distance measure for 2D Shapes Based on Turning Angle”, *Proc. of ICIP-95*, Outubro 1995, pp. 352-355.

[OrRu1997] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra e T. Huang, “Supporting Similarity Queries in Mars”, *ACM Multimedia '97*, 1997.

[PeKo1998] F. Pereira, R. Koenen, “MPEG-7: Status and Directions”, *Advances in Multimedia: Systems, Standards and Networks*, editado por Atul Puri e Tsuhan Chen, 1998.

[Pere1996] F. Pereira, “MPEG-4: Why, What, How and When?”, artigo convidado para o Tutorial Issue sobre a norma MPEG-4, *Signal Processing: Image Communication*, Vol. 15, Nº 4-5, Janeiro 2000

[PrTe1994] W. Press, S. Teukolsky, W. Vetterling e B. Flannery, “Numerical Recipes in C”, *Cambridge University Press*, 2ª edição, 1994, pp. 591-598.

[QBIC_Art] “Fine Arts Museum in San Francisco”, <http://www.thinker.org/imagebase/index-2.html>.

- [QBIC_Trade] “IBM’s Trademark Server”, <http://www.qbic.almaden.ibm.com/tmdemo>.
- [QBIC_Stamps] “IBM’s Stamps Server”, <http://www.qbic.almaden.ibm.com/cgi-bin/stamps-demo>.
- [RuCh1997] Y. Rui, K. Chakrabarti e T. Huang, “Relevance Feedback Techniques in Interactive Content-Based Image Retrieval”, *SPIE ’97*, 1997.
- [RuHu1997a] Y. Rui, T. Huang, S. Chang, “Image Retrieval: Past, Present and Future”, *JVCIR Special Issue*, 1997.
- [RuHu1997b] Y. Rui, T. Huang e S. Mehrotra, “Content-Based Image retrieval with Relevance Feedback in MARS”, *ICIP ’97*, 1997;
- [Russ1995] J. Russ, “The Image Processing Handbook”, *IEEE Press, Inc.*, 2ª edição, 1995.
- [SoHl1993] M. Sonka, V. Hlavac e R. Boyle, “Image Processing, Analysis and Machine Vision”, *Chapman and Hall Computing*, 1ª edição, Londres, 1993.
- [StDi1996] M. Stricker e A. Dimai, “Color Indexing with Weak Spatial Constraints”, *SPIE Proceedings 2670*, Fevereiro 1996, pp. 29-40.
- [Taba1999] A. Tabatabai, “Normalized Contour as a Shape Descriptor for Visual Objects”, *ISO/IEC JTC1/SC29/WG11/P579*, Lancaster, Fevereiro 1999.
- [TaMo1978] H. Tamura, S. Mori e T. Yamawaki, “Texture Features Corresponding to Visual Perception”, *IEEE T. Sys. Man and Cyber*, SMC, Vol. 8, Nº 6, 1978, pp. 460-473.
- [TeCh1988] C. Teh e R. T. Chin, “On Image Analysis by the Methods of Moments”, *IEEE Transactions on Pattern and Machine Intelligence*, Vol. 10, Nº 4, Julho 1988.
- [Tek1999] Tektronix Inc., “Binding-box Based Shape Descriptor for Visual Objects”, *ISO/IEC JTC1/SC29/WG11/P577*, Lancaster, Fevereiro 1999.
- [VeHa1999] M. Vetter, E. Hartley e P. Salembier, “Report of the Ad-hoc Group on MPEG-7 Evaluation Logistics”, *ISO/IEC JTC1/SC29/WG11/M4582*, Seul, Março 1999.
- [VIDEOQ] “VideoQ, An Object Oriented Video Search Engine”, <http://www.ctr.columbia.edu/videoq/>.
- [VisualSEEK] “VisualSEEK: Content-Based Image Retrieval System”, <http://www.ctr.columbia.edu/~jrsmith/VisualSEEK/VisualSEEK.html>.
- [Webseek] “Webseek: Content-Based Image and Video Search and Catalog Tool for the Web”, <http://www.ctr.columbia.edu/webseek>.
- [WhJa1996] D.A. White e R. Jain, “Similarity Indexing with the SS-tree”, *Proceedings 12th IEEE Int. Conf. on Data Engineering*, New Orleans, Louisiana, Fevereiro 1996.
- [Yama2000] A. Yamada, “Results of Core Experiment MS-3: Region-based Shape Descriptor”, *ISO/IEC JTC1/SC29/WG11/M6181*, Beijing, Julho 2000.

[ZiPe2000] C.Zibreira e F. Pereira, “Image Description and Retrieval using MPEG-7 Shape Descriptors”, *European Conference on Digital Libraries 2000*, Lisboa, Setembro 2000.

