

UNIVERSIDADE ESTADUAL PAULISTA

"Júlio de Mesquita Filho"

Pós-Graduação em Ciência da Computação

**Análise Multiescala de Formas Planas
Baseada em Estatísticas da
Transformada de Hough**

Lucas Alexandre Ramos

Bauru

2016

Lucas Alexandre Ramos

Análise Multiescala de Formas Planas Baseada em Estatísticas da Transformada de Hough

Dissertação de Mestrado elaborada junto ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual Paulista "Júlio de Mesquita Filho", Área de Concentração em Computação Aplicada, linha de Processamento de Imagens e Visão Computacional, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Aparecido Nilceu Marana

Bauru

2016

Ramos, Lucas Alexandre.

Análise multiescala de formas planas baseada em estatísticas da transformada de Hough / Lucas Alexandre Ramos. -- São José do Rio Preto, 2016

80 f. : il., tabs.

Orientador: Aparecido Nilceu Marana

Dissertação (mestrado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Computação - Matemática. 2. Processamento de imagens - Técnicas digitais. 3. Visão por computador. 4. Reconhecimento de padrões. 5. Formas (Matemática) I. Marana, Aparecido Nilceu. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 518.72:76

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Lucas Alexandre Ramos

**ANÁLISE MULTIESCALA DE FORMAS PLANAS BASEADA EM ESTATÍSTICAS DA
TRANSFORMADA DE HOUGH**

Área de Concentração: Computação Aplicada

Linha de Pesquisa: Processamento de Imagens e Visão Computacional

Banca Examinadora:

Prof. Dr. Aparecido Nilceu Marana
Faculdade de Ciências
UNESP - Bauru (Presidente)

Prof. Dr. Jurandy Gomes de Almeida Junior
UNIFESP
(Membro Externo)

Prof. Dr. João Paulo Papa
Faculdade de Ciências
UNESP - Bauru
(Membro Interno)

Aprovado em 12 de agosto de 2016.

Agradecimentos

Primeiramente agradeço aos meus pais pelo apoio e suporte ao longo da graduação e mestrado, sem eles nada disso seria possível.

Agradeço ao professor Aparecido Nilceu Marana, que sempre foi muito paciente e atencioso, me ajudando sempre a tomar as melhores decisões ao longo do mestrado e aos professores João P. Papa e Jurandy G. A. Jr, pelas críticas, sugestões e ideias.

Agradeço ao Luis pela parceria ao longo desses anos de graduação e mestrado.

Agradeço a minha noiva Marcia por sempre me apoiar e incentivar a perseguir os meus sonhos.

Agradeço também ao Programa de Pós-graduação em Ciência da Computação (PPGCC), à Unesp e à Fapesp pelo apoio: (Processo número: 2014/10611-0).

Resumo

Atualmente, dada a difusão dos computadores, a tarefa de se reconhecer padrões visuais está sendo cada vez mais automatizada, em especial para tratar a vasta e crescente quantidade de imagens digitais existentes. Aplicações de diversas áreas como biometria, recuperação de imagens baseada em conteúdo e diagnóstico médico, se valem do processamento de imagens, bem como de técnicas de extração e análise de características das mesmas, a fim de identificar pessoas, objetos, gestos, textos, etc. As características básicas que são utilizadas para a análise de imagens são: cor, textura e forma. Recentemente, foi proposto um novo descritor de formas denominado HTS (Hough Transform Statistics), o qual se baseia no espaço de Hough para representar e reconhecer objetos em imagens por suas formas. Os resultados obtidos pelo HTS sobre bases de imagens públicas têm mostrado que este novo descritor, além de apresentar altas taxas de acurácia, melhores do que muitos descritores tradicionais propostos na literatura, é rápido, pois tem um algoritmo de complexidade linear. O objetivo deste trabalho foi explorar as possibilidades de representação multiescala do HTS e, assim, propor novos descritores de formas. Escala é um parâmetro essencial em Visão Computacional e a teoria de espaço-escala refere-se ao espaço formado quando se observa, simultaneamente, os aspectos espaciais de uma imagem em várias escalas, sendo a escala a terceira dimensão. Os novos métodos multiescala propostos foram avaliados sobre várias bases de dados e seus desempenhos foram comparados com o desempenho do HTS e com os principais descritores de formas encontrados na literatura. Resultados experimentais mostraram que os novos descritores propostos neste trabalho são mais rápidos e em alguns casos também mais precisos.

Palavras chave: Análise de Imagens, Análise de Formas, Transformada de Hough, Espaço-Escala, Multiescala, HTS.

Abstract

Currently, given the widespread of computers through society, the task of recognizing visual patterns is being more and more automated, in particular to treat the large and growing amount of digital images available. Applications in many areas, such as biometrics, content-based image retrieval, and medical diagnostic, make use of image processing, as well as techniques for the extraction and analysis of their characteristics, in order to identify persons, objects, gestures, texts, etc. The basic features that are used for image analysis are: color, texture and shape. Recently, it was proposed a new shape descriptor called HTS (Hough Transform Statistics), which is based on the Hough space to represent and recognize objects in images by their shapes. The results obtained by HTS on public image databases have shown that this new shape descriptor, besides showing high accuracy levels, better than many traditional shape descriptors proposed in the literature, is fast, since it has an algorithm of linear complexity. In this dissertation we explored the possibilities of a multiscale and scale-space representation of this new shape descriptor. Scale is a key parameter in Computer Vision and the theory of scale-space refers to the space formed when observing, simultaneously, special aspects of an image at several scales, being the scale the third dimension. The multiscale HTS methods were evaluated on the same databases and their performances were compared with the main shape descriptors found in the literature and with the monoscale HTS. Experimental results showed that these new descriptors are faster and can also be more accurate in some cases.

Keywords: Image Analysis, Shape Analysis, Hough Transform, Scale-Space, Multiscale, HTS.

Conteúdo

1	Introdução	1
1.1	Objetivos	3
1.2	Organização da Dissertação	3
2	Análise Multiescala de Imagens	4
2.1	Técnicas de Representação Multiescala	6
2.1.1	Pirâmide de Imagens	6
2.1.2	Pirâmide Gaussiana	6
2.2	Pirâmide Laplaciana	10
3	Métodos de Reconhecimento Multiescala de Formas	12
3.1	Descritor Multiescala de Fourier	12
3.2	Descritor Fractal Multiescala	16
3.3	Descritor Curvature Scale-Space (CSS)	20
3.4	Descritor Baseado na Transformada Imagem-Floresta	23
4	Descritor Hough Transform Statistics (HTS)	27
4.1	O Descritor HTS	29
4.2	O Descritor HTSn	32
5	Métodos Propostos	34
5.1	Descritores MHTS e MHTSn	34
5.2	Descritores SHTS e SHTSn	36
5.3	Abordagem Hierárquica	38
6	Resultados Experimentais	39
6.1	Primeiro Experimento - Base de Dados Kimia-216	40
6.2	Segundo Experimento - Base de Dados MPEG-7 Part B	46
6.3	Terceiro Experimento - Base de Dados MNIST	53
6.4	Quarto Experimento - Base de Dados de Sinus Frontais	55

6.5	Quinto Experimento - Recuperação de Imagens	57
6.6	Sexto Experimento - SBAS e SMFD	60
7	Conclusões e Trabalhos Futuros	63
7.1	Trabalhos Futuros	64
7.2	Publicações Realizadas	64

Lista de Figuras

2.1	Representação multiescala de uma imagem.	5
2.2	Representação em formato de pirâmide após a divisão sucessiva da imagem por meio da combinação da sub-amostragem e da suavização (LINDEBERG, 1994).	7
2.3	Exemplo de imagens de uma pirâmide gaussiana. Adaptado de Adelson et al. (1984). . . .	8
2.4	Gaussianas utilizadas como função de peso nos diversos níveis da pirâmide. Adaptado de Adelson et al. (1984).	8
2.5	Imagem filtrada por um filtro Gaussiano passa-baixa com escala decrescente. Escalas - (a) $\sigma=20$, (b) $\sigma=15$, (c) $\sigma=11$, (d) $\sigma=8$, (e) $\sigma=5$, (f) $\sigma=3$. Extraído de (DIREKOGLU; NIXO, 2008).	9
2.6	Imagem filtrada por um filtro Gaussiano passa-alta com escala decrescente. Escalas - (a) $\sigma=15$, (b) $\sigma=11$, (c) $\sigma=8$, (d) $\sigma=5$, (e) $\sigma=3$, (f) $\sigma=1$. Extraído de (DIREKOGLU; NIXO, 2008).	9
2.7	Pirâmide Laplaciana, cada nível representa a diferença entre dois níveis sucessivos da pirâmide Gaussiana. Extraído de (ADELSON et al., 1984)	10
2.8	Níveis da pirâmide Laplaciana ampliados. Extraído de (ADELSON et al., 1984).	11
3.1	Imagem com aumento da intensidade (D) de ruído sal e pimenta. Extraído de (DIREKOGLU; NIXO, 2008).	16
3.2	Estrela de Koch, exemplo de fractal.	17
3.3	Dilatação de uma linha por um disco de raio r. Adaptado de (BRUNO; BACKES, 2008).	17
3.4	(a) Silhueta de um morcego da base de imagens MPEG-7 Part B (JEANNIN; BOBER, 1999). Em (b) mostra-se a curva logarítmica (pontos azuis) de $\log(Ar)$ em função de $\log(r)$ da silhueta do morcego, e a reta (em laranja) que melhor representa os pontos desta curva (interpolação linear da curva logarítmica). Adaptado de Souza e Marana (2013).	18
3.5	Para se obter a dimensão fractal multiescala do objeto utiliza-se a derivada da curva log-log em cada ponto desta. Adaptado de (SOUZA; MARANA, 2013).	19
3.6	(a) Baixa amostragem dos pontos iniciais da curva, (b) Fenômeno de Gibbs e (c) Solução para o Fenômeno de Gibbs.	20

3.7	(a) Processo de avaliação do contorno no CSS, e (b) Registro dos pontos nos quais houveram mudanças da curvatura (pontos de inflexão) (LORDÃO, 2009).	21
3.8	Curva com vetores tangentes (em vermelho) e vetores normais (azul).	21
3.9	Evolução do contorno durante o processo de suavização. $\sigma = 1, 4, 7, 10, 12, 14$ respectivamente. Extraído de (ABBASI et al., 1999).	22
3.10	Descritor CSS. (a) imagem com contorno normalizado e ponto inicial marcado, (b) Imagem do Descritor CSS, (c) máximos normalizados da imagem CSS. Extraído de (ABBASI et al., 1999).	23
3.11	Áreas de influência internas (em vermelho) e externas (em verde) de pontos de saliência. Os pontos A, B, D e E são pontos convexos, e suas áreas de influência externas são maiores que as internas. Os pontos C e F côncavos, e representam o inverso dos convexos. Extraído de (SOUZA; MARANA, 2013).	24
3.12	Processo de esqueletonização multiescala de imagens por propagação de rótulos (a) é o contorno rotulado, (b) é a mapa de rótulos, (c) é a imagem de diferença, e (d,e,f) são imagens do esqueleto em 3 escalas diferentes. Extraído de (TORRES et al., 2003).	25
3.13	(a) o contorno de saliências de uma folha e (b-c), saliências dos seus esqueletos internos e externos.	26
3.14	(a) contorno e esqueletos de um contorno com os pontos de saliência indicado por pontos, (b) os valores de saliência dos vértices do polígono pela sua posição relativa ao longo do contorno.	26
4.1	Equação da reta na imagem.	28
4.2	Representação da reta por coordenadas polares.	29
4.3	(a) Silhueta de uma maçã da base de imagens MPEG-7 PartB (JEANNIN; BOBER, 1999) (b) Borda da maçã segmentada; (c) espaço de Hough (matriz acumuladora) gerado a partir dos pontos da borda. (SOUZA; MARANA, 2013).	30
4.4	Histogramas associados a quatro pontos da borda de dois objetos de mesma classe (sinos) da base MPEG-7 Part B (JEANNIN; BOBER, 1999). Adaptado de (SOUZA; MARANA, 2013).	30
4.5	Funções obtidas para silhuetas da base Kimia-216 (SEBASTIAN et al., 2004). Adaptado de (SOUZA; MARANA, 2013).	31
4.6	Em (a) é exibida uma imagem da base Kimia-216 (SEBASTIAN et al., 2004) e em (b) são exibidas suas funções unidimensionais. As retas verticais tracejadas indicam as posições onde as funções foram amostradas para se gerar o vetor de características.	32

4.7	As matrizes com bordas verdes representam uma parte ampliada do espaço de Hough final de um dado objeto em dois momentos distintos, (a), no processo de construção (pelo descritor HTS) do histograma do ponto de borda associado à senóide cujas posições na matriz aparecem destacadas em vermelho. Os valores nas células com bordas azuis são armazenados, obedecendo a ordem entre as matrizes verdes, isto é, (a), nos <i>bins</i> do histograma do ponto. Já a matriz de (b) ilustra o processo de construção (agora pelo descritor HTSn) do histograma do ponto de borda associado à senóide cujas posições na matriz aparecem destacadas em vermelho. O valor armazenado em cada <i>bin</i> do histograma do ponto de borda corresponde à soma do valor na respectiva posição da senóide com os valores nas oito posições adjacentes a esta no espaço de Hough. Adaptado de (SOUZA; MARANA, 2013)	33
5.1	Espaço de Hough obtido para diferentes escalas (E): 1,2,3,4 e 5.	35
5.2	Diagrama dos passos do método MHTS: (a) - Extração das bordas da imagem de entrada; (b) - Cálculo dos espaços de Hough para cada uma das n escalas; (c) Funções calculadas a partir do espaço de Hough, (d) - Obtenção dos descritores HTS para cada uma das n escalas; (e) - Concatenação dos n descritores HTS em um único descritor MHTS.	35
5.3	Exemplos de imagens obtidas após a convolução com o núcleo gaussiano: (a) $\sigma = 2$, (b) $\sigma = 4$, (c) $\sigma = 8$, (d) $\sigma = 16$, (e) $\sigma = 32$, (f) $\sigma = 64$	36
5.4	Imagens binarizadas após a convolução com o núcleo gaussiano: (a) $\sigma = 2$, (b) $\sigma = 4$, (c) $\sigma = 8$, (d) $\sigma = 16$, (e) $\sigma = 32$, (f) $\sigma = 64$	37
5.5	Diagrama do método espaço-escala proposto, (a) convolução da imagem com filtro gaussiano, (b) extração da borda, (c) Funções calculadas a partir do espaço de Hough, (d) Obtenção dos descritores a partir da amostragem e (e) Descritores concatenados.	37
6.1	Exemplos de imagens da Base Kimia-216 (SEBASTIAN et al., 2004).	40
6.2	Curvas Precisão x Revocação das escalas individuais do método MHTS para a base Kimia-216 (JEANNIN; BOBER, 1999)	41
6.3	Curvas Precisão x Revocação do MHTS, MHTSv e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	42
6.4	Curvas Separabilidade Multiescala do MHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	42
6.5	Curvas Precisão x Revocação do MHTSn e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	43
6.6	Curvas Separabilidade Multiescala do MHTSn e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	44

6.7	Curvas Precisão x Revocação do SHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	44
6.8	Curvas Separabilidade Multiescala do SHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	45
6.9	Curvas Precisão x Revocação dos descritores propostos utilizando a base Kimia-216 (SEBASTIAN et al., 2004).	45
6.10	Exemplos de Imagens da Base MPEG-7 Part B (JEANNIN; BOBER, 1999).	46
6.11	Curvas Precisão x Revocação das escalas individuais do método MHTS para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)	47
6.12	Curvas Precisão x Revocação do MHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999)	48
6.13	Curvas Separabilidade Multiescala do MHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999)	48
6.14	Curvas Precisão x Revocação do método MHTSn, MHTSvn e outros métodos utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	49
6.15	Curvas Separabilidade Multiescala do MHTSn, MHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	49
6.16	Curvas Precisão x Revocação do SHTS, SHTSv e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	50
6.17	Curvas Separabilidade Multiescala do SHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	50
6.18	Curvas Precisão x Revocação do SHTSn, SHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	51
6.19	Curvas Separabilidade Multiescala do SHTSn, SHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	51
6.20	Curvas Precisão x Revocação dos descritores propostos utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	52
6.21	Exemplos de dígitos da base MNIST (LECUN et al., 1998): (a) Dígitos com boa grafia, mais simples de serem identificados (de 0 a 9, 3 exemplos cada) e (b) Dígitos com grafia precária, identificação complexa até mesmo para seres humanos. Extraído de (SOUZA; MARANA, 2014)	53
6.22	Curvas Precisão x Revocação do MHTS e demais descritores utilizando a base MNIST (LECUN et al., 1998).	54
6.23	Curvas Separabilidade Multiescala do MHTS, SHTS e demais descritores utilizando a base MNIST (LECUN et al., 1998).	54
6.24	(A) imagem de raios -X ; (B) imagem de Tomografia Computadorizada de um indivíduo. Os seios frontais estão delimitados por retângulos.	56

6.25	(a)-(c) Sequência de imagens de 3 exames de tomografia computadorizada da base FSCT.	56
6.26	Curvas ROC obtidas a partir dos descritores de formas para a base FSTC.	57
6.27	Cavalo em (a) pode ser mais parecido ao cachorro em (b) que ao cavalo em (c). Extraído de (BAI et al., 2012).	58
6.28	Curvas Precisão x Revocação para a recuperação de imagens da base MPEG-7 Part B.	59
6.29	Curvas Precisão x Revocação para a base Kimia-216.	60
6.30	Curvas Separabilidade Multiescala para a base Kimia-216.	61
6.31	Curvas Precisão x Revocação para a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	61
6.32	Curvas Separabilidade Multiescala para a base MPEG-7 Part B (JEANNIN; BOBER, 1999).	62

Lista de Tabelas

6.1	Resultados Bulls-Eye para cada escala do método MHTS para a base Kimia-216	41
6.2	Resultados Bulls-Eye para a base Kimia-216	46
6.3	Resultados Bulls-Eye para cada escala do método MHTS para a base MPEG-7 Part B . .	47
6.4	Resultados Bulls-Eye para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)	52
6.5	Resultados Bulls-Eye para a base MNIST (LECUN et al., 1998)	55
6.6	Valores de ERR para os descritores testados na base de sinus frontais.	57
6.7	Tempo médio em milissegundos e desvio padrão para a recuperação de imagens na base MPEG-7 Part B.	59
6.8	Tempo médio, desvio padrão e tempo total para a recuperação de imagens na base MNIST.	60
6.9	Resultados Bulls-Eye para a base Kimia-216.	61
6.10	Resultados Bulls-Eye para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)	62

Capítulo 1

Introdução

Durante o processo de evolução do homem, os sistemas neurais e cognitivos relacionados ao sistema visual humano (SVH) tornaram-se bastante sofisticados e muito efetivos na tarefa de processamento de imagens e reconhecimento de padrões visuais (DUDA et al., 2007).

Com o advento dos computadores, pesquisadores e empresas de todo o mundo direcionaram esforços para o desenvolvimento de sistemas de visão computacional (SVC) visando a automatização do processo de captura, processamento, extração de características e reconhecimento de padrões visuais presentes nas imagens digitais, dada a grande variedade de aplicações.

Inúmeras atividades que requerem reconhecimento de padrões visuais são realizadas atualmente pelos SVC de forma mais rápida e, em alguns casos, até mesmo com maior precisão do que se fossem realizadas por SVH, tais como a identificação de pessoas a partir de suas impressões digitais, o reconhecimento óptico de caracteres, a análise de sequências de DNA, a busca de imagens baseada em conteúdos em grandes bases de dados, dentre outras.

Dentre as características que podem ser analisadas a partir de uma imagem encontra-se a forma de objetos ou regiões da mesma. Apesar de o ser humano lidar com o reconhecimento de formas a todo momento, a definição do conceito “forma” não é fácil de ser enunciada. Uma definição é apresentada por Costa e Júnior (2000) para os quais uma forma pode ser entendida como um conjunto de pontos conectados (quer no espaço discreto ou no contínuo) os quais apresentam um padrão de distribuição (padrão visual).

Existem na literatura vários métodos para descrição de formas, dentre os quais destacam-se: Beam Angle Statistics (BAS) (ARICA; VURAL, 2003), Shape Context (SC) (BELONGIE; MALIK, 2000), Multiscale Fractal Dimension (MFD) (BACKES; BRUNO, 2006), Tensor Scale (TS) (MIRANDA et al., 2005), Fourier Descriptors (FD) (ZHANG; LU, 2002), Momentos de Zernike (HU, 1962) e Contour Saliency (CS) (TORRES; FALCÃO, 2007).

Inspirados pela abordagem do descritor BAS (ARICA; VURAL, 2003), Souza e Marana (2013) propuseram o descritor de formas HTS (Hough Transform Statistics). O HTS baseia-se em medidas estatísticas

extraídas do espaço de Hough (HOUGH, P. V. C., 1962) gerado a partir dos pontos de borda do objeto presente na imagem. Objetos semelhantes geram espaços de Hough semelhantes enquanto que objetos distintos geram espaços distintos, sendo possível realizar a discriminação entre eles (SOUZA; MARANA, 2013). Os resultados obtidos pelo HTS sobre bases de imagens públicas mostraram que este novo descritor, além de apresentar altas taxas de acurácia, melhores do que muitos descritores tradicionais propostos na literatura, é rápido, pois tem um algoritmo de complexidade linear.

No processamento e análise de imagens, ao se extrair características de interesse, é de extrema importância considerar que determinadas informações apenas fazem sentido sob certas condições de observação, como, por exemplo, a escala. Porém, a escolha da escala de observação mais adequada não é uma tarefa trivial, o que motivou o desenvolvimento de filtros espaço-escala para a representação multiescala de imagens (WITKIN, 1984), (LINDBERG, 1994). O espaço-escala refere-se ao espaço formado quando se observa os aspectos espaciais de uma imagem em várias escalas simultaneamente, sendo a escala considerada a terceira dimensão (WITKIN, 1984).

A ideia básica por trás de uma abordagem multiescala consiste em criar uma família de sinais (imagens) derivados cujas estruturas são sucessivamente simplificadas. Desta forma, as estruturas presentes nas imagens de escala menor devem representar as estruturas presentes em uma escala maior de maneira simplificada. Assim, é possível analisar os diferentes níveis de representação e utilizar aqueles que exibem apenas as características de interesse.

Diversas abordagens que utilizam a multiescala na detecção de objetos podem ser encontradas na literatura, como em (NOSATO et al., 2015), em que uma abordagem multiescala foi utilizada na detecção de objetos em imagens de colonoscopia. O exame de colonoscopia é fundamental para a detecção do câncer gastrointestinal, sendo um dos cânceres com maior índice de incidência. Porém, a precisão no diagnóstico é altamente dependente do médico que executa o exame, uma vez que a identificação deve ser feita observando-se imagens microscópicas enquanto é feito o exame. Além disso, os objetos presentes nas imagens podem variar de tamanho, o que torna interessante uma abordagem multiescala. Dessa forma, no método proposto por Nosato et al. (2015), são identificados objetos multiescala que indicam a presença de câncer gastrointestinal.

A grande dimensão das bases de dados atuais requer algoritmos de busca que sejam rápidos e eficientes (como o algoritmo de busca de imagens da Google, que requer resposta em tempo real) (LIU et al., 2012). Na análise de formas, estratégias que utilizam representação multiescala têm sido propostas por alguns pesquisadores (BACKES; BRUNO, 2006) e (TORRES et al., 2003). Nesta dissertação, propõe-se uma versão multiescala do descritor de formas HTS, proposto por Souza e Marana (2013). O desenvolvimento de uma versão multiescala do HTS é de grande relevância devido ao fato do descritor apresentar uma complexidade linear, o que o torna ideal para tarefas em que o tempo de processamento é um fator importante. Dessa forma, a partir das escalas mais grosseiras, o método multiescala pode eliminar imagens que são muito diferentes da imagem de busca, tornando o método ainda mais rápido. Além disso, o HTS pela sua concepção apresenta potencial multiescala.

1.1 Objetivos

O objetivo principal dessa dissertação de mestrado foi explorar as possibilidades de representação multiescala do descritor de formas HTS (Hough Transform Statistics).

O desempenho do HTS multiescala foi comparado com o desempenho do HTS original e de alguns dos principais descritores de formas propostos na literatura, dentre os quais: Beam Angle Statistics (BAS) (ARICA; VURAL, 2003), Shape Context (SC) (BELONGIE; MALIK, 2000), Multiscale Fractal Dimension (MFD) (BACKES; BRUNO, 2006), Tensor Scale (TS) (MIRANDA et al., 2005), Fourier Descriptors (FD) (ZHANG; LU, 2002) e Contour Saliency (CS) (TORRES; FALCÃO, 2007).

1.2 Organização da Dissertação

Esta dissertação está organizado em 9 capítulos, como descrito a seguir.

No Capítulo 1, apresenta-se a introdução ao tema da dissertação de mestrado e seus objetivos.

No Capítulo 2, apresenta-se o conceito de análise multiescala de imagens e uma revisão da literatura sobre técnicas de representação multiescala.

No Capítulo 3, apresentam-se métodos de reconhecimento que utilizam abordagem multiescala para o reconhecimento de formas.

No Capítulo 4, apresenta-se o método para análise de formas baseado na transformada de Hough.

No Capítulo 5, apresentam-se os métodos multiescala e espaço-escala baseados em estatísticas da transformada de Hough propostos nesta dissertação.

No Capítulo 6, apresentam-se os resultados obtidos.

No Capítulo 7, apresentam-se as conclusões desta dissertação, resumem-se as contribuições do trabalho desenvolvido, indicam-se trabalhos futuros e citam-se os artigos elaborados e publicados durante o desenvolvimento desta dissertação.

Capítulo 2

Análise Multiescala de Imagens

A visão computacional consiste na extração de características importantes a partir de imagens do mundo real. Ao se extrair características de interesse de uma imagem, é de extrema importância considerar que determinadas informações apenas fazem sentido sob certas condições de observação e o uso de métodos automáticos pode auxiliar a determinar quais informações são relevantes (LINDEBERG, 1994).

Como exemplo pode-se considerar o conceito de galho de uma árvore, que deve ser observado de uma distância variando de alguns centímetros a alguns metros, e da floresta a qual ela pertence, que deverá ser observada a alguns quilômetros de distância para que ambas observações sejam relevantes. Este fato está relacionado ao conceito de escala, ou seja, o conceito de um objeto pode variar dependendo da escala de observação (LINDEBERG, 1994).

A natureza Euclidiana do mundo impõe restrições naturais em um sistema de visão computacional, como as variações de iluminação, objetos em movimento, em diferentes escalas de acordo com a distância e muitos outros. Para um sistema de visão computacional simples, nos estágios iniciais pode-se efetuar a convolução da imagem com os núcleos Gaussianos, que podem ser definidos como filtros passa-baixa que fornecem uma estrutura uniforme e podem servir de entrada para outros módulos visuais (uma definição mais completa de núcleo Gaussiano é apresentada no Capítulo 4). KOENDERINK (1984) enfatiza que uma imagem digital pode se associar, em princípio, às escalas externa e interna, correspondendo ao seu tamanho e à sua densidade de amostragem, respectivamente.

Um problema comum na análise de imagens é definir que tipo de informação deverá ser analisada e quais recursos computacionais serão necessários para cada escala de representação. Uma abordagem axiomática sugere que os primeiros estágios de um sistema de visão computacional deverão funcionar sem exercer alterações irreversíveis nos dados de entrada. Como inicialmente é impossível se prever qual a escala de representação ideal, deve-se considerar todas as escalas possíveis, daí o conceito de representação multiescala, uma abordagem amplamente utilizada em diversas aplicações de análise e processamento de sinais (incluindo, portanto, imagens) (LINDEBERG, 1994).

A ideia básica por trás de uma abordagem de análise multiescala de imagens consiste em criar uma

família de sinais (imagens) derivados cujas estruturas são sucessivamente simplificadas. É importante salientar que estruturas em escalas inferiores devem representar simplificações das estruturas correspondentes em escalas superiores, ou seja, não podem simplesmente ser algo proveniente do método de suavização como o núcleo Gaussiano. Assim, é possível analisar os diferentes níveis de representação e utilizar aqueles que exibem apenas as características de interesse, conforme ilustra a Figura 2.1 (LINDBERG, 1994).

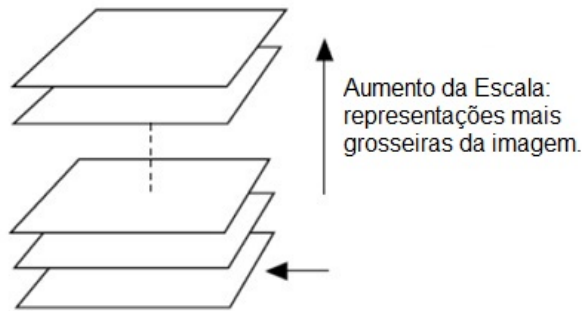


Figura 2.1: Representação multiescala de uma imagem.

Os principais problemas referentes ao uso de métodos multiescala se devem a dificuldade de relacionar informações significativas do sinal em diferentes níveis de observação. WITKIN (1984) propôs uma abordagem denominada espaço-escala, na qual uma característica de interesse descreve um caminho contínuo na representação, possibilitando que estruturas da imagem presentes em diferentes escalas possam ser manipuladas de maneira consistente. Assim, é possível relacionar informações obtidas em diferentes níveis de observação, bem como determinar sua localização precisa no sinal original.

Uma das principais razões para representar em múltiplos níveis uma informação que já está presente no sinal original é que a simplificação sucessiva remove detalhes indesejados, tais como ruído ou estruturas não-significativas, facilitando o processamento de tarefas futuras. Além disso, por lidar explicitamente com o parâmetro de escala, esta abordagem possibilita manipular apenas as estruturas significativas da imagem que se tornam explícitas em cada nível. O fato da diminuição da escala estar relacionada a redução da quantidade de informação implica em um aumento da eficiência computacional.

A teoria espaço-escala estabelece as condições necessárias para a definição de transformações que possibilitem a manipulação de características presentes em diferentes níveis de representação de maneira consistente.

Uma propriedade fundamental da teoria espaço-escala e multiescala é que a transformação para um nível mais grosseiro não introduza novas estruturas, ou seja, características pertencentes em uma determinada escala também estão presentes em todas as escalas mais finas (ou seja, nas escalas mais baixas). Tal propriedade conduz a uma simplificação do sinal (da imagem) original, considerando que a quantidade de características diminui e que as remanescentes em um dado nível correspondem a simplificações daquelas existentes inicialmente. Contudo, não é necessário que suas localizações sejam exatamente as mesmas ao longo das escalas. É preciso apenas que elas definam caminhos contínuos, ou seja, basta que seja possível

rastrear a localização de uma característica em cada escala. Desse modo, características em diferentes níveis estão relacionadas se elas pertencem ao mesmo caminho no espaço-escala resultante.

Outra propriedade exigida na teoria espaço-escala é a fidelidade (WITKIN, 1984). A fidelidade requer que a transformação espaço-escala convirja para o sinal original conforme a escala tenda a zero.

2.1 Técnicas de Representação Multiescala

Os primeiros trabalhos relacionados a abordagem multiescala foram publicados em (ROSENFELD; THURSTON, 1971), nos quais foi observada a vantagem da utilização de operadores de diferentes tamanhos na detecção de bordas. Propostas similares podem ser encontradas em (KLINGER, 2007), no qual é feita a representação de imagens utilizando diferentes escalas espaciais e sub amostras. Posteriormente aos trabalhos de (ROSENFELD; THURSTON, 1971), essas técnicas foram aprimoradas, principalmente por Burt (1981) e Crowley (1981), em uma das técnicas mais utilizadas para a representação multiescala, a pirâmide.

2.1.1 Pirâmide de Imagens

A pirâmide de imagens é uma estrutura de dados que oferece eficiência nas operações de convolução por meio da representação de imagens reduzidas. A pirâmide consiste em um conjunto de cópias da imagem original em que ambas resolução e densidade são reduzidas a cada passo. Os níveis da pirâmide são obtidos por meio de um algoritmo iterativo muito eficiente, no qual o nível inicial corresponde à imagem original e a cada nível a escala da imagem é reduzida por um fator k . Diversas técnicas para se obter uma representação da imagem em forma de pirâmide podem ser encontradas na literatura, como a pirâmide Gaussiana (BURT, 1981) e (CROWLEY 1981).

2.1.2 Pirâmide Gaussiana

Na técnica de representação de uma imagem em forma de pirâmide, o nível inicial da pirâmide (imagem original) é filtrado por um filtro passa-baixa e, depois, sub-amostrado por um fator (normalmente igual a dois), formando o segundo nível da pirâmide, e assim por diante, até o algoritmo convergir (ADELSON et al., 1984). Por exemplo, suponha uma imagem de dimensões $2K \times 2K$ com $f(K)=f$, sendo que no nível K inicial $f(K)$ é a imagem de entrada, a representação de $f(K)$ em um nível mais grosseiro $f(K-1)$ é definida por um operador de redução. Simplificando, assume-se que o filtro de suavização é separável e que o número de coeficientes para o filtro de uma dimensão é ímpar. Dessa forma, pode-se escrever a Equação 2.1.

$$f^{(K-1)}(x) = \sum_{n=-N}^N c(n) f^K(2x - n), \quad (2.1)$$

em que $c : Z \rightarrow R$ refere-se ao conjunto dos coeficientes do filtro, também chamado de núcleo gerador, e por razões computacionais ele deve ser uma função pequena e separável (ADELSON et al., 1984). O núcleo gaussiano 2D pode ser definido de acordo com a Equação 2.2

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (2.2)$$

Porém, esse processo é comumente chamado de Redução, logo a Equação 2.1 pode ser representada pela Equação 2.3. A Figura 2.2 ilustra o procedimento de redução, em que a base da pirâmide é a imagem de entrada e a cada nível superior a imagem é reduzida.

$$f^{(K-1)} = Reduce(f^K) \quad (2.3)$$

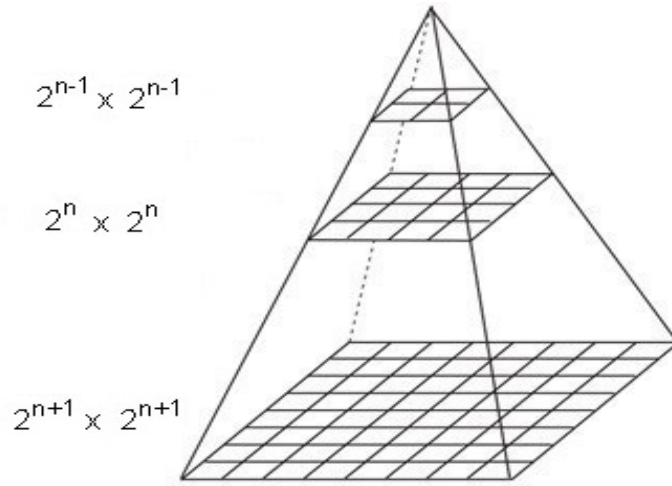


Figura 2.2: Representação em formato de pirâmide após a divisão sucessiva da imagem por meio da combinação da sub-amostragem e da suavização (LINDEBERG, 1994).

Esse tipo de pirâmide passa-baixa foi proposto por Burt (1981) e Crowley (1981) e sua principal vantagem é que tanto a escala quanto o tamanho da imagem diminuem exponencialmente, diminuindo também o esforço computacional para processar os dados. A construção de uma pirâmide de imagens é equivalente a convolução da imagem original com um conjunto de funções gaussianas. A convolução com as gaussianas age como um filtro passa baixa, suavizando a imagem a cada nível. Esse tipo de pirâmide é conhecida como pirâmide Gaussiana, e um exemplo pode ser observado na Figura 2.3, na qual $G(0)$ é a imagem de entrada e as demais representam cada nível da pirâmide.

Para se obter os diversos níveis da pirâmide é necessário variar a gaussiana. A Figura 2.4 mostra exemplos de gaussianas utilizadas em diferentes níveis. É importante notar que o tamanho da gaussiana dobra a cada nível.

Outra abordagem para se obter a pirâmide Gaussiana é a partir de um filtro passa-alta. A primeira abordagem utiliza um filtro passa-baixa, que suaviza a imagem e conforme a escala diminui, ocorre a perda das bordas e das regiões exteriores. Sendo assim, a utilização do filtro passa-baixa em diferentes escalas

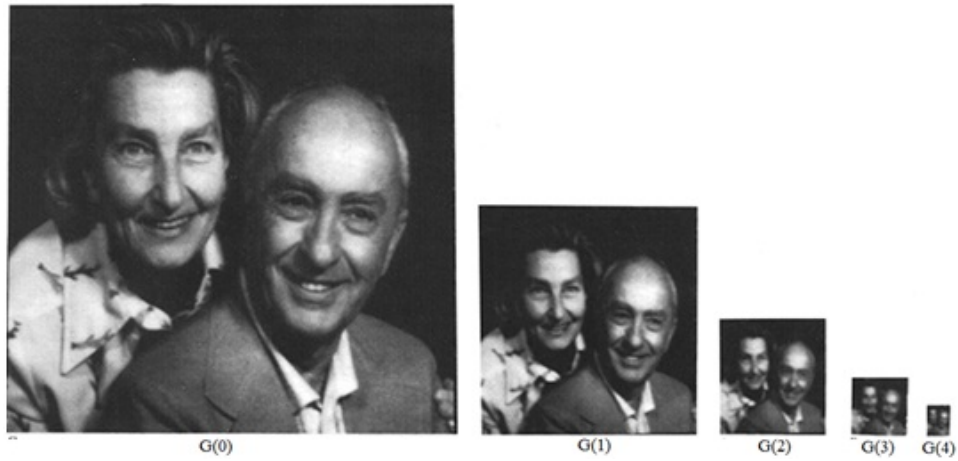


Figura 2.3: Exemplo de imagens de uma pirâmide gaussiana. Adaptado de Adelson et al. (1984).

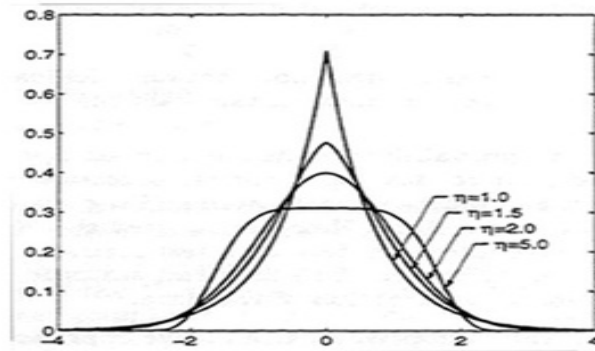


Figura 2.4: Gaussianas utilizadas como função de peso nos diversos níveis da pirâmide. Adaptado de Adelson et al. (1984).

tem como foco o destaque das regiões interiores e centrais do objeto. Já o filtro passa-alta aplicado em diferentes escalas tem como foco o destaque das bordas e das partes exteriores dos objetos (DIREKOGLU; NIXO, 2008). A Figura 2.5 mostra uma imagem após a aplicação de um filtro Gaussiano passa-baixa, nota-se que os componentes interiores são destacados e a borda perde foco conforme a escala diminui. A Figura 2.6 mostra uma imagem após a aplicação de um filtro Gaussiano passa-alta, pode-se observar que a borda torna-se mais definida conforme a escala diminui.

Ambas abordagens podem ser utilizadas para descrever uma imagem, podem ser aplicadas no reconhecimento de formas e servem como base para a aplicação de descritores como a Transformada de Fourier (DIREKOGLU; NIXO, 2008).

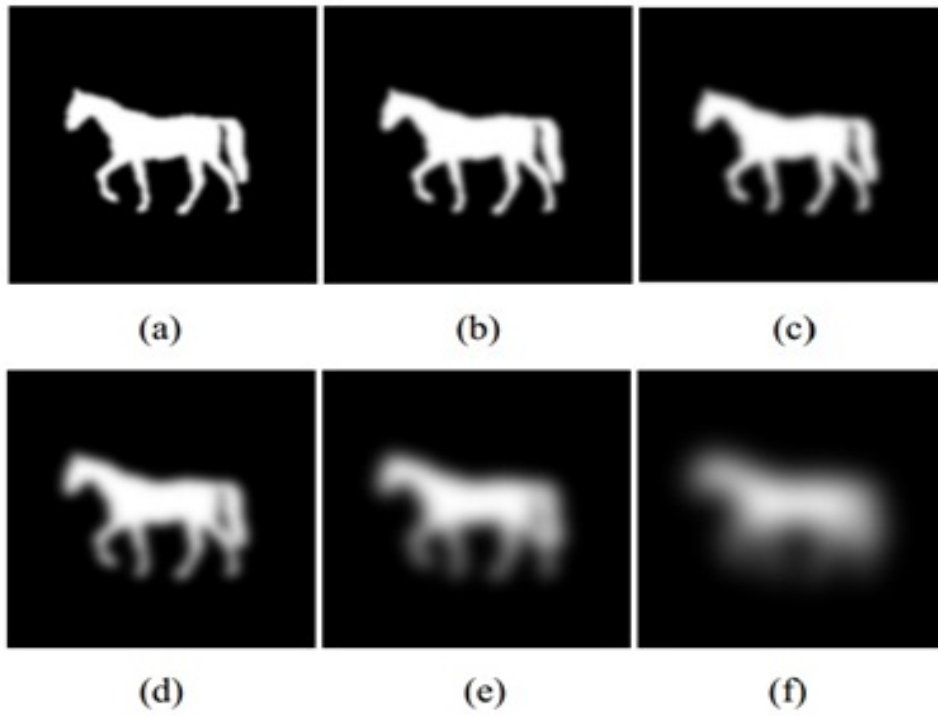


Figura 2.5: Imagem filtrada por um filtro Gaussiano passa-baixa com escala decrescente. Escalas - (a) $\sigma=20$, (b) $\sigma=15$, (c) $\sigma=11$, (d) $\sigma=8$, (e) $\sigma=5$, (f) $\sigma=3$. Extraído de (DIREKOGLU; NIXO, 2008).

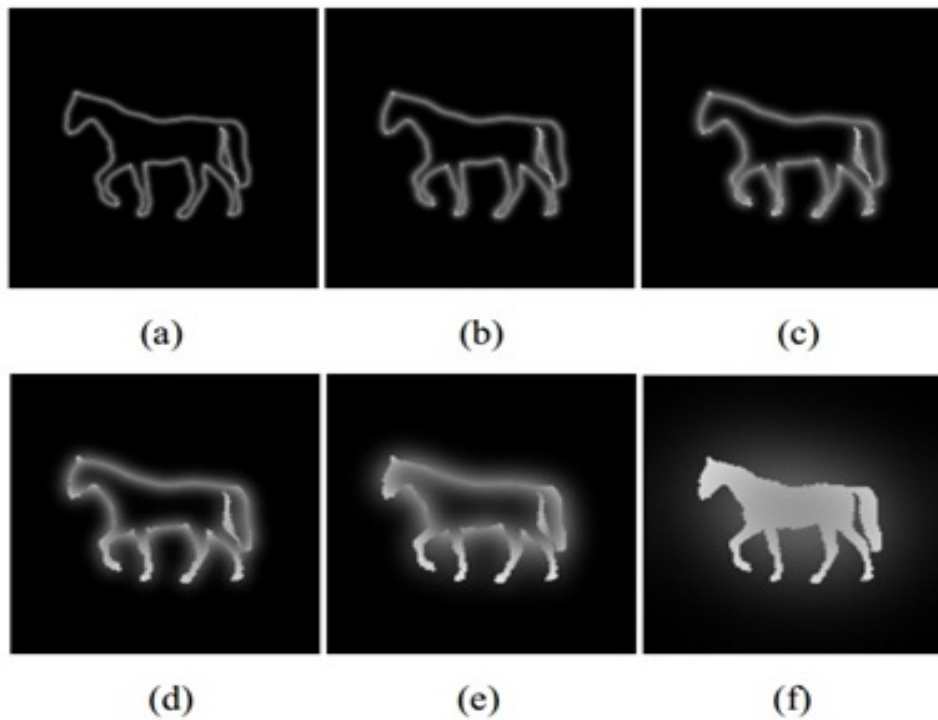


Figura 2.6: Imagem filtrada por um filtro Gaussiano passa-alta com escala decrescente. Escalas - (a) $\sigma=15$, (b) $\sigma=11$, (c) $\sigma=8$, (d) $\sigma=5$, (e) $\sigma=3$, (f) $\sigma=1$. Extraído de (DIREKOGLU; NIXO, 2008).

2.2 Pirâmide Laplaciana

As imagens pertencentes à pirâmide gaussiana são obtidas por meio da convolução da imagem com um filtro passa-baixa ou passa-alta. Porém, algumas aplicações, como extração de características e compressão de dados, requerem que as imagens pertencentes à pirâmide tenham sido obtidas a partir de um filtro passa-banda. O conceito de pirâmide Laplaciana surgiu para solucionar esse problema (LINDBERG, 1994) e (ADELSON et al., 1984).

Para se obter a pirâmide Laplaciana, subtrai-se cada imagem de um nível da pirâmide Gaussiana, do próximo nível inferior da pirâmide. Porém, as imagens em níveis diferentes da pirâmide Gaussiana apresentam densidade de amostras, sendo necessário a interpolação de novas amostras entre esses níveis para que a subtração possa ser efetuada. Essa interpolação é obtida a partir da operação inversa a Redução, chamada de Expansão, e é definida pela Equação 2.4.

$$f^{(k)}(x) = \sum_{n=-N}^N c(n) f^{k-1}\left(\frac{x-n}{2}\right) f^{(k)} = \text{Expand}(f^{k-1}) \quad (2.4)$$

A expansão dobra o tamanho da imagem a cada iteração. A Figura 2.7 mostra um exemplo de pirâmide Laplaciana, em que cada nível foi obtido a partir da operação de redução de duas imagens de níveis subseqüentes da pirâmide Gaussiana, e então subtraindo-se um nível do outro.



Figura 2.7: Pirâmide Laplaciana, cada nível representa a diferença entre dois níveis sucessivos da pirâmide Gaussiana. Extraído de (ADELSON et al., 1984)

As pirâmides Laplacianas vêm sendo usadas em diversas aplicações, como em (SUNKAVALLI et al., 2010), em que foi utilizada para a harmonização de imagens, e em (PARIS et al., 2011), em que foi utilizada para o processamento de imagens como mapeamento de tons e ampliação de detalhes. A Figura 2.8 mostra os níveis da pirâmide Laplaciana ampliados para o mesmo tamanho. Pode-se observar que a cada nível da pirâmide as bordas são melhoradas e segregadas por tamanho.

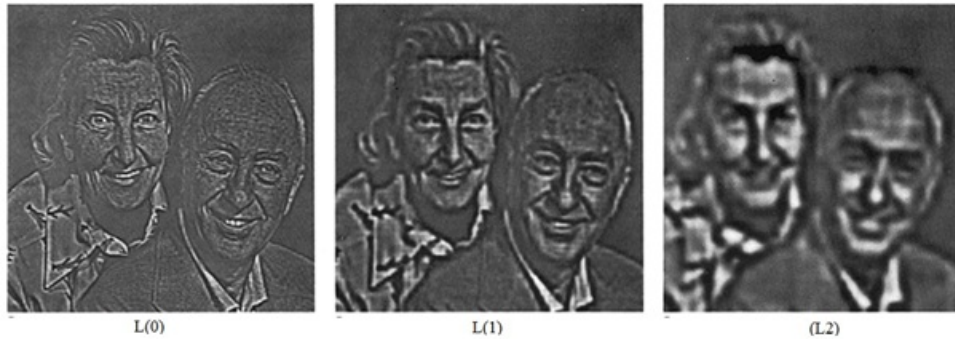


Figura 2.8: Níveis da pirâmide Laplaciana ampliados. Extraído de (ADELSON et al., 1984).

Capítulo 3

Métodos de Reconhecimento

Multiescala de Formas

A descrição da forma de um objeto é uma tarefa muito importante na análise e reconhecimento de objetos, na qual o maior problema consiste em determinar a similaridade entre dois formatos ou objetos.

Conforme citado anteriormente, na visão humana as estruturas são observadas em múltiplas escalas, portanto, uma representação multiescala é de extrema importância para a tarefa de classificação de imagens, além de melhorar o desempenho dos sistemas.

Diversos métodos para a descrição da forma podem ser encontrados na literatura e são divididos em duas categorias, baseados no contorno e baseados na região. Os métodos baseados em região levam em consideração a área da figura como um todo, enquanto os baseados no contorno trabalham apenas com as bordas da figura (GONZALEZ; WOODS, 2007). Tendo em vista que iremos trabalhar apenas com métodos baseados no contorno, os métodos baseados em região não serão abordados.

3.1 Descritor Multiescala de Fourier

O uso do descritor de Fourier para o reconhecimento de padrões é muito comum por ser invariante a rotação e ao ponto de início do contorno. Esse descritor utiliza uma abordagem baseada no contorno da imagem. Para a extração do contorno da imagem diversas técnicas podem ser utilizadas. Em (KUNTTU et al., 2003; KUNTTU et al., 2004) a função de coordenadas complexas representada pela Equação 3.1 foi utilizada, uma vez que por meio das coordenadas complexas o contorno pode ser representado independente da posição do objeto na imagem, dessa forma, temos uma invariância a translação.

$$z(k) = (x_k - x_c) + j(y_k - y_c) \quad (3.1)$$

em que $k=0,1,2,\dots,N-1$ em que N é o comprimento do contorno e (x_c, y_c) é o centróide do objeto.

A transformada de Fourier de um contorno é um conjunto de números complexos chamados de descritores de Fourier e descrevem a forma de um objeto no domínio da frequência (KUNTTU et al., 2003; KUNTTU et al., 2004). Para gerar os descritores de Fourier com um contorno complexo a transformada discreta de Fourier exibida na Equação 3.2 deve ser utilizada.

$$F(n) = \frac{1}{n} \sum_{k=0}^{N-1} z(k) e^{-j2\pi nk/N} \quad (3.2)$$

em que $n=0,1,2,\dots,N-1$.

O contorno geral do objeto é definido pelos descritores de baixa frequência, enquanto os descritores de alta frequência representam os detalhes do objeto. Normalmente, apenas um subconjunto dos descritores são utilizados para representar o contorno do objeto. Para isso, diversas técnicas podem ser utilizadas, como *Curvature Fourier*, *Radius Fourier* e *Contour Fourier*, porém, de acordo com os experimentos realizados em (KAUPPINEN et al., 1995), a técnica de *Contour Fourier* apresenta melhores resultados para a classificação de formas (KUNTTU et al., 2003; KUNTTU et al., 2004).

O método de *Contour Fourier* aplica a transformada de Fourier a partir das coordenadas complexas do contorno obtidas pela Equação 3.1. Nesse método, os descritores assumem eixos de frequência negativa e positiva. O escalonamento dos descritores é feito pela divisão absoluta dos valores dos descritores escolhidos pelo valor absoluto do primeiro componente diferente de zero. O vetor de características para este método é dado pela Equação 3.3.

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \dots \frac{|F_{-1}|}{|F_1|} \frac{|F_2|}{|F_1|} \dots \frac{|F_{L/2}|}{|F_1|} \right]^T, \quad (3.3)$$

em que L é uma constante que define o número de amostras do contorno utilizadas.

A representação multiescala da forma do objeto pode ser obtida a partir de uma transformada *wavelet*. A função da forma é transformada a partir de uma *wavelet* Ψ , nesse caso, a transformada *wavelet* complexa, que se baseia na transformada *wavelet* continua definida pela Equação 3.4 (KUNTTU et al., 2003; KUNTTU et al., 2004).

$$C_a(b) = \frac{1}{\sqrt{|a|}} \int_R z(k) \Psi \left(\frac{k-b}{a} \right) dk \quad (3.4)$$

Como na transformada de Fourier, o resultado da transformada *wavelet* complexa é um conjunto de coordenadas complexas $C_a(b)$ de escala a , sendo os coeficientes de b definidos para todas as posições de 1 a $N-1$.

O problema com os coeficientes obtidos a partir da transformada *wavelet* complexa é que estes são dependentes do ponto inicial da borda do objeto e o vetor de características pode variar de acordo com seu comprimento, sendo impossível a sua comparação com outros vetores. Dessa forma, a solução é aplicar a transformada de Fourier nos coeficientes complexos obtidos a partir da transformada *wavelet* complexa, dessa forma, a representação multiescala do objeto pode ser levada para o domínio da frequência. O

descriptor é então obtido aplicando-se a transformada de Fourier nos coeficientes obtidos a partir da Equação 3.4, conforme a Equação 3.5 (KUNTTU et al., 2003; KUNTTU et al., 2004).

$$F^a(n) = \frac{1}{N} \sum_{b=0}^{N-1} C_a(b) e^{-j2\pi nb/N}, \quad (3.5)$$

em que C_a é um conjunto de coordenadas complexas.

O descriptor multiescala $x^{(a)}$ pode então ser formado a partir da Equação 3.6 referente ao método *Contour Fourier*.

$$x^a = \left[\frac{|F_{(-L/2-1)}^a|}{|F_1^a|} \cdots \frac{|F_{-1}^a|}{|F_1^a|} \frac{|F_2^a|}{|F_1^a|} \cdots \frac{|F_{L/2}^a|}{|F_1^a|} \right]^T \quad (3.6)$$

A partir da Equação 3.6, uma representação multiescala do objeto pode ser obtida a partir de diversos valores de escala a e combinando os descritores em um único vetor de tamanho igual ao número de escalas utilizadas. Esses vetores podem então ser comparados a partir de uma medida de distância como a distância Euclidiana.

De acordo com os testes efetuados por KUNTTU et al. (2003), o método Fourier multiescala apresenta resultados melhores na classificação de imagens de três bases diferentes quando comparado ao método *Contour Fourier*, mostrando que a análise multiescala pode ser aplicada com melhores resultados em diversos tipos de bases de dados. Quanto ao custo computacional, o método de Fourier multiescala requer mais tempo de processamento que o método de *Contour Fourier* devido ao tamanho do vetor de características. Pode-se concluir que a abordagem multiescala combinada com uma abordagem típica como o descriptor de Fourier pode facilmente aprimorar o desempenho na classificação de imagens.

Outra abordagem para se obter descritores de imagens baseados na Transformada de Fourier é a proposta em (DIREKOGLU; NIXO, 2008), na qual a Transformada de Fourier é utilizada. A partir das imagens obtidas utiliza-se a técnica de pirâmide Gaussiana com filtro passa-alta e passa-baixa.

O primeiro passo dessa abordagem é a normalização da silhueta de um objeto usando interpolação bilinear. Dessa forma, caso um objeto apresente mais pontos na silhueta que outro, ambos passarão a ser representados pelo mesmo número de pontos. Então a Transformada de Fourier 2-D é aplicada de acordo com a Equação 3.7 e a partir dela, obtêm-se a imagem da magnitude de Fourier:

$$FT(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) e^{[-j2\pi(ux/M + vy/N)]}, \quad (3.7)$$

em que $FT(u, v)$ é a transformada de Fourier da silhueta da imagem $I(x, y)$ e $M \times N$ é o tamanho da silhueta da imagem.

A imagem da magnitude de Fourier $|FT(u, v)|$, é invariante à translação, porém não é invariante à rotação. A abordagem multiescala é obtida a partir dessa etapa, na qual para destacar a parte interior e central do objeto a técnica de pirâmide Gaussiana com filtro passa-baixa é aplicada segundo a Equação

3.8:

$$|FT(u, v)|^s = |FT(u, v)| \left(e^{-\sqrt{\frac{u^2+v^2}{2\sigma_s^2}}} \right), \quad (3.8)$$

em que $|FT(u, v)|^s$ e σ_s são a transformada de Fourier e o parâmetro de escala de índice s .

Esse método gera o espaço-escala da imagem conforme exemplificado na Figura 2.5. A pirâmide Gaussiana com filtro passa-baixa garante ênfase dos componentes de baixa frequência. Além disso, para representar as bordas da imagem, é aplicada também a técnica de pirâmide Gaussiana com filtro passa-alta, de acordo com a Equação 3.9:

$$|FT(u, v)|^s = |FT(u, v)| \left(1 - e^{-\sqrt{\frac{u^2+v^2}{2\sigma_s^2}}} \right) \quad (3.9)$$

Conforme a escala diminui, as bordas do objeto e as regiões exteriores se tornam mais evidentes, conforme o exemplo da Figura 2.6.

Porém, as imagens da magnitude obtidas ainda não são ideais para se efetuar a comparação uma vez que ainda sofrem com os problemas de rotação. Para remover a variação de rotação, as coordenadas de cada imagem da magnitude de Fourier são transformadas para coordenadas polares, para que as rotações sejam representadas por translações na nova imagem. Dessa forma cada ponto (x, y) passa a ser representado por um ponto (ρ, θ) . Por fim, a transformada de Fourier é aplicada novamente de acordo com a Equação 3.10 para se computar a magnitude e remover as componentes relacionadas à translação (DIREKOGLU; NIXO, 2008):

$$FPT^s(k, l) = \frac{1}{EF} \sum_{r=0}^{E-1} \sum_{\theta=0}^{F-1} P^s(r, \theta) e^{[-j2\pi(kr/E + l\theta/F)]} \quad (3.10)$$

em que $FPT^s(k, l)$ é a transformada de Fourier para a imagem em coordenadas polares $P^s(r, \theta)$ de tamanho $E \times F$ na escala de índice s .

A imagem da magnitude de Fourier resultante é invariante a translação, escala e rotação, e pode ser utilizada como descritor.

Para a classificação de imagens, os descritores podem ser comparados a partir do algoritmo do vizinho mais próximo. A distância Euclidiana (ED) é utilizada para calcular a similaridade entre dois objetos de acordo com a Equação 3.11:

$$ED^s(T, D) = \sqrt{\sum_{i=1}^E \sum_{j=1}^F (OD_T^s(i, j) - OD_D^s(i, j))^2}, \quad (3.11)$$

em que $ED^s(T, D)$ é a distância Euclidiana entre dois descritores de objetos e OD_T^s e OD_D^s são descritores das imagens de teste e da base de dados.

A distância média é calculada para cada objeto baseado no número de escalas utilizadas (DIREKO-

GLU; NIXO, 2008).

Experimentos com esse descritor foram realizados em (DIREKOGLU; NIXO, 2008), utilizando a base de dados MPEG-7 CE-Shape-1 Part B, que possui 1400 imagens, sendo divididas em 70 classes de 20 imagens. Para avaliação, ruído do tipo sal e pimenta foi adicionado a cada imagem da base de dados conforme ilustra a Figura 3.1.

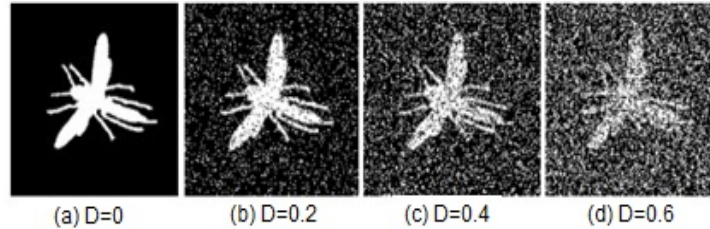


Figura 3.1: Imagem com aumento da intensidade (D) de ruído sal e pimenta. Extraído de (DIREKOGLU; NIXO, 2008).

De acordo com os experimentos realizados em (DIREKOGLU; NIXO, 2008), o método de reconhecimento baseado na transformada de Fourier a partir da aplicação da técnica de pirâmide Gaussiana com filtro passa-alta obteve os melhores resultados na classificação de imagens, seguidos pela pirâmide Gaussiana com filtro passa-baixa. De acordo com os resultados obtidos em (DIREKOGLU; NIXO, 2008), os descritores propostos apresentaram resultados superiores aos métodos Momentos de Zernike (HU, 1962) e o descritor Elíptico de Fourier na classificação de imagens da base de dados MPEG-7 CE-Shape-1 Part B (JEANNIN; BOBER, 1999).

3.2 Descritor Fractal Multiescala

A geometria Euclidiana define formas perfeitas conhecidas como retas, círculos, quadrados entre outras para representar os elementos presentes no mundo. Porém, essas formas não são suficientes para classificar alguns elementos presentes na natureza. Para isso, a geometria fractal estuda formas complexas conhecidas como fractais, que não podem ser explicadas pelos conceitos da geometria Euclidiana. Os fractais são objetos matemáticos gerados a partir de regras simples de construção, mas que ao serem repetidas diversas vezes geram figuras de grande complexidade (BRUNO et al., 2007).

Segundo Bruno et al. (2007), os fractais apresentam as seguintes propriedades: auto-semelhança em escala, complexidade infinita e dimensão fractal. Um exemplo de fractal pode ser observado na Figura 3.2.

Enquanto objetos Euclidianos possuem dimensão inteira, os objetos fractais possuem dimensão fracionada que indica o grau de complexidade e irregularidade de uma imagem, ou seja, o quanto do espaço físico ela ocupa. Essa medida é uma importante fonte de informação e pode ser utilizada na caracterização de formas para a identificação (BRUNO et al., 2007). Existem diversas abordagens para se efetuar o cálculo da Dimensão Fractal, como a de Hausdorff-Besicovitch e a de Minkowski-Bouligand, sendo que

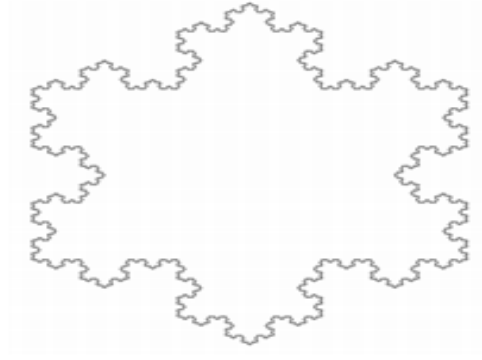


Figura 3.2: Estrela de Koch, exemplo de fractal.

esta última apresenta resultados mais precisos (BRUNO et al., 2007). O método de Minkowski-Bouligand é baseado no estudo da área de influência criada pela dilatação da forma utilizando um disco de raio r . A Figura 3.3 mostra o processo de dilatação de uma linha por um elemento estruturante em forma de disco.

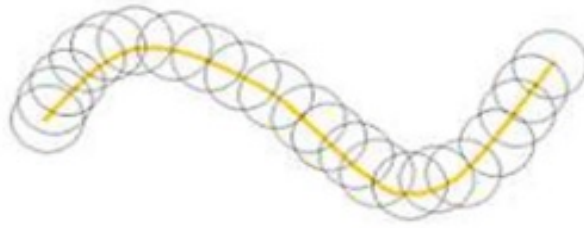


Figura 3.3: Dilatação de uma linha por um disco de raio r . Adaptado de (BRUNO; BACKES, 2008).

A área de influência gerada pela dilatação é chamada de $A(r)$, e sua relação com o raio é dada pela Equação 3.12.

$$A(r) = \mu r^{2-D} \quad (3.12)$$

Aplicando logaritmo de ambos os lados, a dimensão fractal pode ser obtida a partir da Equação 3.13:

$$D = 2 - \lim_{r \rightarrow 0} \frac{\log A(r)}{\log r}, \quad (3.13)$$

em que D é a dimensão fractal de Minkowski-Bouligand, sendo um número no intervalo de 0 a 2. Seu cálculo também é possível com base no coeficiente angular da reta α que melhor se aproxima da curva logarítmica de $A(r)$ em função do raio de dilatação r . Encontrada essa reta, é possível determinar o valor de D aproximado a partir do coeficiente angular da reta por meio da Equação 3.14 (BRUNO et al., 2007).

$$D = 2 - \alpha \quad (3.14)$$

Um problema da Dimensão Fractal é que na natureza os objetos possuem uma dimensão finita e a

sua complexidade tende a zero conforme a escala de observação aumenta. É importante observar que a dilatação do objeto fractal usando o método de Minkowski-Bouligand apresenta um comportamento diferente para diversos pontos da forma. Em algumas formas a dilatação ocorre livremente em alguns pontos, enquanto em outros, essa dilatação passa a ser saturada. A Dimensão Fractal indica como a forma se restringe para cada valor do raio (BRUNO; BACKES, 2008).

Esse comportamento garante à curva log-log, obtida a partir do método de Minkowski-Bouligand, uma riqueza de detalhes que não pode ser simplesmente expressada por um valor numérico. Um exemplo de curva log-log é mostrado na Figura 3.4 (BRUNO; BACKES, 2008).

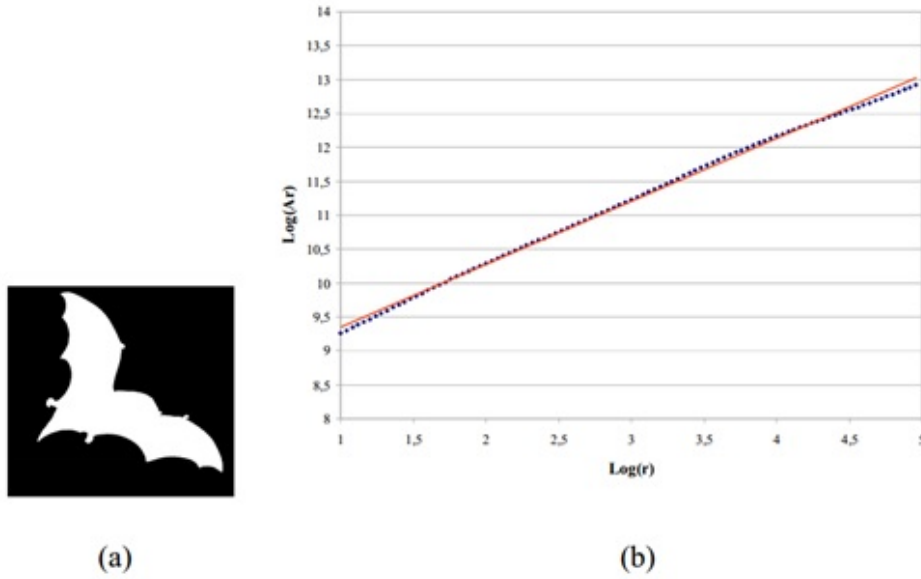


Figura 3.4: (a) Silhueta de um morcego da base de imagens MPEG-7 Part B (JEANNIN; BOBER, 1999). Em (b) mostra-se a curva logarítmica (pontos azuis) de $\log(Ar)$ em função de $\log(r)$ da silhueta do morcego, e a reta (em laranja) que melhor representa os pontos desta curva (interpolação linear da curva logarítmica). Adaptado de Souza e Marana (2013).

A interpolação linear da curva log-log é uma representação muito grosseira, o que causa a perda de muitas informações importantes. Para resolver este problema desenvolveu-se a Dimensão Fractal Multiescala. A partir da derivada, é possível obter-se uma função que une as mudanças da Dimensão Fractal às mudanças da dilatação do raio. Essa função chamada de Dimensão Fractal Multiescala (MFD) é definida pela Equação 3.15:

$$MFD = 2 - \frac{du(t)}{dt}, \quad (3.15)$$

em que $du(t)/dt$ é a derivada da curva log-log $u(t)$ calculada por meio da propriedade derivativa da Transformada de Fourier.

Essa propriedade permite calcular a curva derivada no espectro e tem um melhor desempenho quando comparado com métodos numéricos, uma vez que considera todos os pontos durante a computação da derivada. Um detalhe importante que requer atenção é que métodos derivados têm uma tendência para

ênfatizar ruídos de alta frequência. Assim, é necessário utilizar um filtro passa-baixa, como o Filtro Gaussiano, a fim de reduzir o ruído. A derivada baseada na transformada de Fourier é dada pelas Equações 3.16 e 3.17. A Figura 3.5 ilustra o conceito de Dimensão Fractal Multiescala.

$$\frac{du(t)}{dt} = F^{-1}Fu(t)Fg_{\sigma}(t)(2j\pi f) \quad (3.16)$$

$$g_{\sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right), \quad (3.17)$$

em que t e $u(t)$ são, respectivamente, o logaritmo do raio e a área de influência do método Minkowski-Bouligand, f é a frequência, j é o número imaginário e $g_{\sigma}(t)$ é a função gaussiana com desvio padrão σ .

Aspectos importantes devem ser considerados quando se calcula a curva MFD. Esses aspectos são relacionados ao comportamento da curva log-log $u(t)$ e é fundamental fornecer uma curva com uma boa amostragem e intervalo uniforme. Isso é realizado quando os pontos iniciais da curva não são considerados, uma vez que apresentam baixa amostragem, seguidos de uma interpolação linear, que é realizada a partir do preenchimento do espaço entre cada dois pontos da curva de amostragem por seu ponto médio. Outro problema comum é a descontinuidade do método nos limites da curvatura, sendo conhecido como fenômeno de Gibbs. Esse fenômeno é devido ao fato da transformada de Fourier não convergir uniformemente em descontinuidades. Uma solução eficaz para esse problema é a utilização de um esquema de duplicação e reflexão da curva, de modo que é possível fazê-lo contínua. A Figura 3.6 ilustra o problema de baixa amostragem dos pontos iniciais da curva (a), em (b) o fenômeno de Gibbs e em (c) a duplicação da curva e esquema de reflexão para fornecer uma curva contínua.

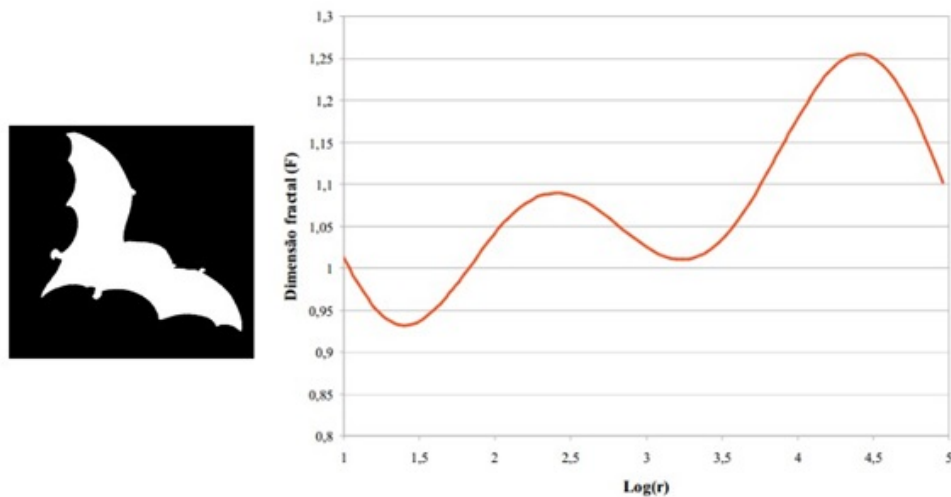


Figura 3.5: Para se obter a dimensão fractal multiescala do objeto utiliza-se a derivada da curva log-log em cada ponto desta. Adaptado de (SOUZA; MARANA, 2013).

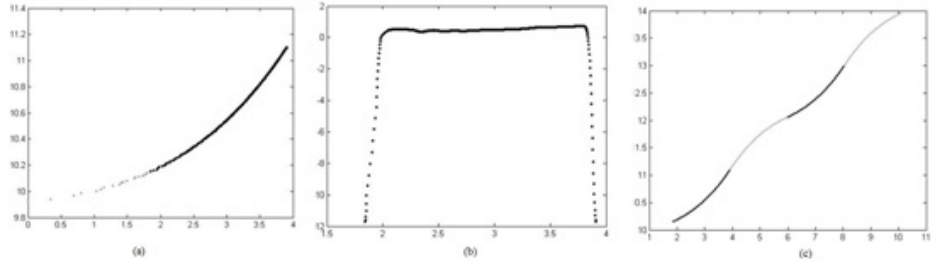


Figura 3.6: (a) Baixa amostragem dos pontos iniciais da curva, (b) Fenômeno de Gibbs e (c) Solução para o Fenômeno de Gibbs.

Apesar da curva MFD ser uma grande fonte de informação sobre os objetos de uma imagem, ela é composta por um grande conjunto de valores, o que torna a sua aplicação inviável para sistemas reais. Portanto, para tornar viável a sua utilização na discriminação de objetos é necessário uma nova representação da curva (FLORINDO et al., 2012). A partir da curva da Dimensão Fractal Multiescala, diversos métodos podem ser aplicados a fim de se extrair as melhores características da curva e gerar assim, o vetor de características da forma, como a abordagem utilizada em (TORRES et al., 2003), em que a curva log-log é aproximada por meio de uma curva polinomial de grau nove, obtendo então sua função derivada, cuja curva é amostrada em 100 posições para gerar o vetor de características da forma.

3.3 Descritor Curvature Scale-Space (CSS)

O descritor *Curvature Scale Space* (CSS) proposto por Mokhtarian e Mackworth (1986), tem como objetivo a representação da forma de um objeto por meio do registro dos locais onde o contorno muda de côncavo para convexo e vice-versa. O registro é feito a cada etapa de suavização do objeto, o qual vai perdendo a intensidade de variação de curvatura até tornar-se completamente côncavo (LORDÃO, 2009). A geração do descritor CSS consiste em avaliar e registrar os pontos de uma imagem na qual seu contorno fechado sofre mudanças no sentido da curvatura. A Figura 3.7 ilustra o processo de avaliação do contorno no CSS, sendo que em (a) as setas verdes indicam as regiões convexas e as setas amarelas as regiões côncavas e em (b) temos o registro dos pontos onde ocorreram mudanças da curvatura de convexa para côncava e vice-versa (ABBASI et al., 1999) e (LORDÃO, 2009).

A curvatura de uma curva é definida de acordo com a Equação 3.18.

$$k(s) = \lim_{h \rightarrow 0} \frac{\phi}{h}, \quad (3.18)$$

em que ϕ é o ângulo entre $t(s)$ e $t(s + h)$, sendo que t representa o vetor tangente e s é o comprimento do arco.

A Figura 3.8 ilustra uma curva com os vetores tangentes (vermelhos) a alguns pontos e em azul os vetores normais (ou seja, vetores perpendiculares a uma superfície).

A curva é então representada pela Equação paramétrica 3.19.

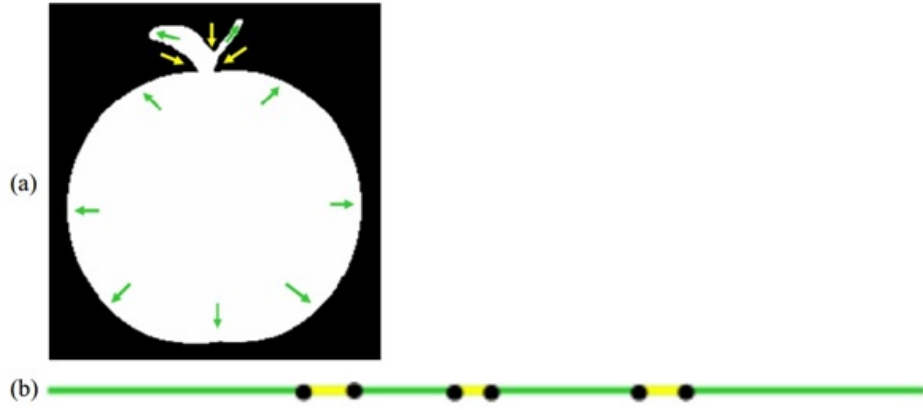


Figura 3.7: (a) Processo de avaliação do contorno no CSS, e (b) Registro dos pontos nos quais houveram mudanças da curvatura (pontos de inflexão) (LORDÃO, 2009).

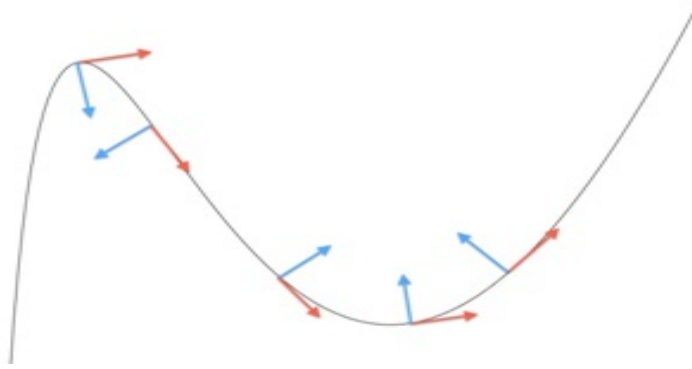


Figura 3.8: Curva com vetores tangentes (em vermelho) e vetores normais (azul).

$$\Gamma(u) = (x(u), (y(u)), \quad (3.19)$$

em que u apresenta posições arbitrárias do contorno e $x(u), y(u)$ representa um ponto do contorno indicado pelo parâmetro u .

A partir da convolução da imagem com um filtro Gaussiano suavizador, pode se reescrever a definição do contorno a partir das Equações 3.20 e 3.21:

$$\Gamma(u)_\sigma = (x(u, \sigma), (y(u, \sigma)) \quad (3.20)$$

$$x(u, \sigma) = x(u) \otimes g(u, \sigma), y(u, \sigma) = y(u) \otimes g(u, \sigma), \quad (3.21)$$

em que se assume $\otimes$ como operador de convolução e $g(u, \sigma)$ um filtro Gaussiano de desvio padrão σ para suavização do contorno definido pela Equação 3.22.

$$g(u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}} \quad (3.22)$$

A partir de $x(u, \sigma)$ e $y(u, \sigma)$, pode-se definir o cálculo da curvatura de $\Gamma(u)_\sigma$ de acordo com a Equação 3.23.

$$K(u, \sigma) = \frac{x_u(u, \sigma)y_{uu}(u, \sigma) - y_u(u, \sigma)x_{uu}(u, \sigma)}{(x_u(u, \sigma)^2 + y_u(u, \sigma)^2)^{2/3}}, \quad (3.23)$$

em que $x_u(u, \sigma)$ e $x_{uu}(u, \sigma)$ são as derivadas de primeira e segunda ordem de $x(u, \sigma)$ em relação a u assim como y_u e y_{uu} .

Em (LORDÃO, 2009), o processo de extração do descritor CSS é resumido da seguinte forma:

- Obter o contorno da imagem;
- Normalizar o contorno por meio de reamostragem, de forma que o contorno de todos os objetos seja representado por um mesmo número total de pontos;
- Definir σ inicial igual a zero;
- Calcular o contorno evoluído (convolução Equação 27);
- Calcular a função de curvatura $K(u, \sigma)$ sobre o contorno $\Gamma(u)_\sigma$;
- Determinar N amostras equidistantes em relação ao parâmetro normalizado u ;
- Identificar entre as N amostras os pontos onde houve mudança de sinal em relação ao ponto anterior;
- Registrar os pontos onde há mudança em um plano cartesiano de ordenadas σ e abscissas N;
- Se houve mudança de sinal incrementar σ e voltar ao passo 3. Caso contrário fim do algoritmo.

A Figura 3.9 mostra o processo de suavização do contorno e obtenção dos registros. Pode-se observar que o número de pontos nos quais o valor da função curvatura muda, diminui conforme σ aumenta.

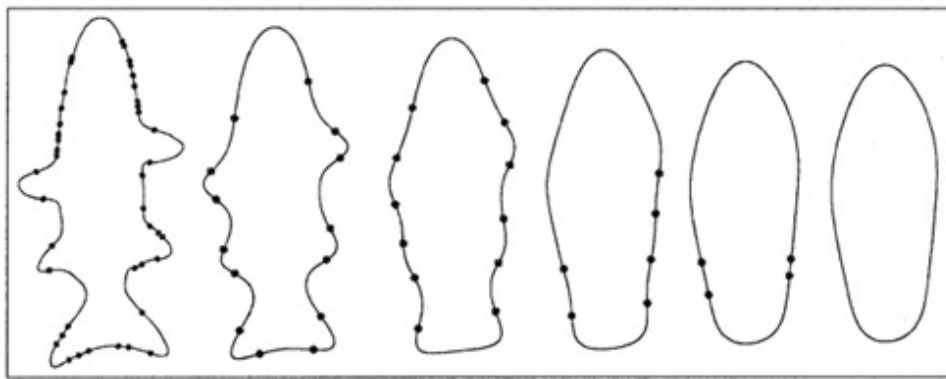


Figura 3.9: Evolução do contorno durante o processo de suavização. $\sigma = 1, 4, 7, 10, 12, 14$ respectivamente. Extraído de (ABBASI et al., 1999).

A Figura 3.10 mostra o descritor gerado pelo CSS, sendo que (a) apresenta a imagem com contorno normalizado, o ponto em que se inicia o cálculo da curvatura marcado por um X e os pontos nos quais a

função de curvatura muda de valor marcados pelos números de 1 a 8, (b) apresenta a imagem do descritor gerado na qual o eixo das ordenadas é composto pelos valores de σ e o das abscissas é composto pelos valores de N e conforme o valor de σ aumenta, o número de pontos de curvatura diminui, e (c) os pontos máximos de (b), que serão utilizados na comparação das imagens.

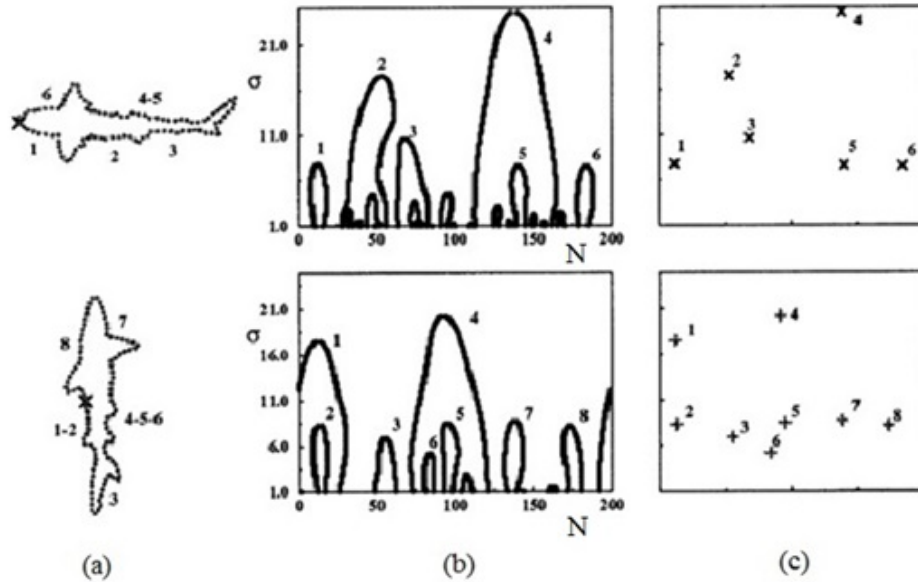


Figura 3.10: Descritor CSS. (a) imagem com contorno normalizado e ponto inicial marcado, (b) Imagem do Descritor CSS, (c) máximos normalizados da imagem CSS. Extraído de (ABBASI et al., 1999).

3.4 Descritor Baseado na Transformada Imagem-Floresta

A transformada imagem-floresta (IFT - *Image Foresting Transform*) proposta por Falcao et al. (2004) reduz problemas de processamento de imagem baseados em conectividade ao cálculo de uma floresta de caminhos ótimos em um grafo direcionado, no qual os nós são os pixels e as arestas são definidas de acordo com a adjacência dos pixels. O custo de um caminho é determinado a partir de um conjunto de sementes e uma função de custo que seja adequada.

Para se obter uma função de custo adequada, cada pixel da imagem recebe um valor de caminho de custo mínimo a partir do conjunto de sementes, de forma que a união desses caminhos forme uma floresta. A IFT atribui a cada pixel três atributos: o seu antecessor no caminho ótimo, o custo desse caminho, e a raiz correspondente ou algum rótulo associado a ele (TORRES; FALCÃO, 2007).

Na IFT, as raízes são os mínimos locais da função de conectividade. Essas raízes competem entre si e cada uma conquista os nós mais fortemente conectados a ela, o que gera o particionamento da imagem em uma floresta em que cada raiz forma uma árvore de caminhos ótimos. Essa floresta é uma partição ótima da imagem.

A IFT essencialmente generaliza o algoritmo de Dijkstra para funções de conectividade, nas quais

para qualquer nó existe um caminho ótimo.

Em (TORRES; FALCÃO, 2007), o algoritmo de IFT é aplicado aos esqueletos das imagens e, então, o método de Saliências do Contorno é aplicado a fim de se gerar o descritor multiescala.

As saliências de uma forma (contorno) são definidas como as áreas máximas de influência de seus pontos de borda de maiores curvaturas (TORRES et al., 2003). No contorno, a área de influência de pontos convexos tende a ser maior fora do contorno do que dentro dele e vice-versa para os pontos côncavos. A Figura 3.11 ilustra as áreas de influência associadas aos pontos de saliência (vértices) de um polígono. A área de influência de cada ponto de saliência relacionado ao ângulo de abertura θ é dada pela Equação 3.24:

$$Area = \frac{\theta + r^2}{2}, \quad (3.24)$$

em que r é o raio de dilatação.

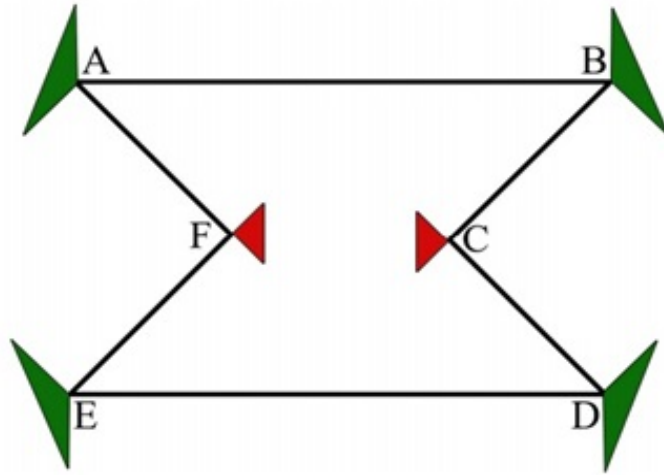


Figura 3.11: Áreas de influência internas (em vermelho) e externas (em verde) de pontos de saliência. Os pontos A, B, D e E são pontos convexos, e suas áreas de influência externas são maiores que as internas. Os pontos C e F côncavos, e representam o inverso dos convexos. Extraído de (SOUZA; MARANA, 2013).

Na abordagem de TORRES et al. (2003), cada imagem da base é esqueletonizada por meio da propagação de rótulos, na qual dado um contorno com N pixels, seu esqueleto interno é definido como a localização geométrica dos centros de máximos de discos contidos no contorno. Aplicando-se a IFT ao contorno cria-se um mapa de raiz R , esqueletos multiescala podem ser computados a partir de R se cada pixel do contorno p está atribuído a um valor de rótulo posterior $\lambda(p)$, variando de 1 a N enquanto se circunscreve o contorno. Um mapa de rótulos L pode ser criado pela computação de $L(R(p))$ para cada pixel p da imagem. No entanto, uma abordagem mais eficiente consiste em propagar os rótulos dos pixels do contorno durante a aplicação da IFT. Dessa forma, o mapa de rótulos L é criado simultaneamente ao mapa de Raízes R . Uma diferença de imagens D é computada a partir de cada pixels de dentro e de fora do contorno de acordo com a Equação 3.25.

$$D(p) = \max_{q \in A_4(p)} \{\min(\delta(p, q), N - \delta(p, q))\}, \quad (3.25)$$

em que $\delta(p, q) = L(q) - L(p)$ e $A_4(p)$ é o conjunto de pixels q que são 4 vizinhos de p .

A imagem de diferença D representa o esqueleto multiescala interno e externo da imagem de entrada. Esqueletos afinados (com largura de um pixel) podem ser obtidos pela limiarização da imagem diferença D . Quando maior o valor de limiar aplicado à limiarização, mais simplificado o esqueleto se torna, com a perda progressiva de pequenos detalhes. A Figura 3.12 ilustra o processo de esqueletonização multiescala da imagem, em que (a) é o contorno rotulado, (b) é a mapa de rótulos, (c) é a imagem de diferença e (d,e,f) são imagens do esqueleto em 3 escalas diferentes (TORRES et al., 2003).

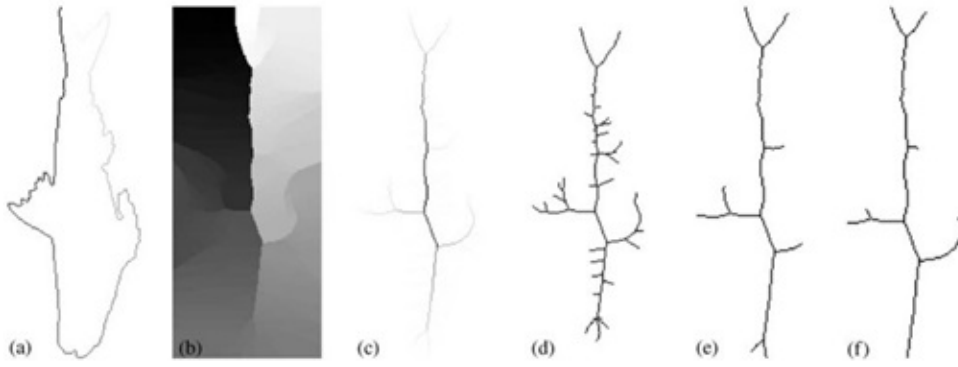


Figura 3.12: Processo de esqueletonização multiescala de imagens por propagação de rótulos (a) é o contorno rotulado, (b) é a mapa de rótulos, (c) é a imagem de diferença, e (d,e,f) são imagens do esqueleto em 3 escalas diferentes. Extraído de (TORRES et al., 2003).

A partir dos esqueletos multiescala das imagens, cada ponto de saliência do esqueleto interno corresponde a um ponto convexo do contorno e cada ponto de saliência do esqueleto externo corresponde a um ponto côncavo do contorno, conforme ilustrado na Figura 3.13. As áreas de influência de cada pixel são determinadas com base nos histogramas do contorno interno e externo do esqueleto obtido a partir da propagação de rótulos, restrito a pixels dentro de uma faixa estreita em torno dos esqueletos. Os pontos de saliência são aqueles com área de influência acima do limiar obtido com $\theta = 70$ na Equação 3.24.

Para localizar os pontos de saliência ao longo do contorno, o mapa de rótulos L é utilizado. A localização e a área de influência dos pontos de saliência ao longo do contorno representam informação local e global importante para a análise de formas. As áreas de influência (valores de saliência) são obtidas a partir do histograma restrito a uma estreita faixa ao redor do contorno, com $r = 10$ (Equação 3.24). São marcadas como negativas para pontos côncavos e positivas para pontos convexos. Um ponto arbitrário do contorno é tomado como ponto de referência e o algoritmo calcula a posição relativa de cada ponto de saliência em relação ao ponto de referência ao longo do contorno. A partir dessa operação, o descritor do objeto é definido por dois vetores de tamanho idêntico, um com os valores de saliência e outro com as posições relativas das saliência ao longo do contorno. Esses descritores podem ser então comparados para se obter uma medida de similaridade entre as imagens (TORRES et al., 2003). A Figura 3.14 ilustra as

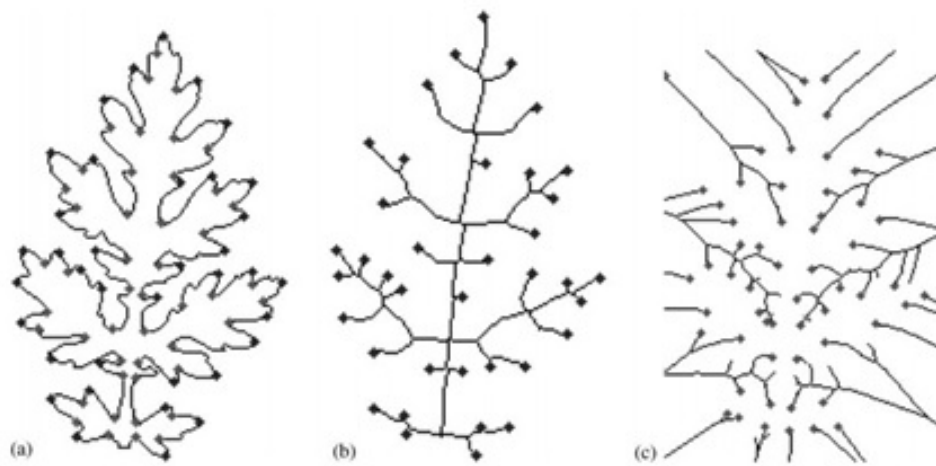


Figura 3.13: (a) o contorno de saliências de uma folha e (b-c), saliências dos seus esqueletos internos e externos.

saliências do contorno de um polígono (a) com seus esqueletos interno e externo e (b) indica os valores de saliência dos vértices do polígono, por sua posição relativa ao longo do contorno.

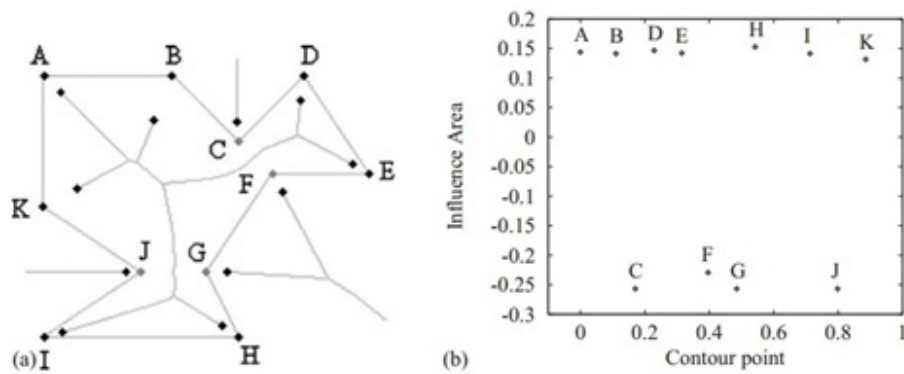


Figura 3.14: (a) contorno e esqueletos de um contorno com os pontos de saliência indicado por pontos, (b) os valores de saliência dos vértices do polígono pela sua posição relativa ao longo do contorno.

Capítulo 4

Descritor Hough Transform Statistics (HTS)

Um problema comum na análise de imagens é a identificação de retas caracterizadas por um conjunto de pontos colineares presentes na imagem. Uma possível solução para este problema consiste em testar todas as retas formadas a partir de todos os pares de pontos da imagem. Porém, esta é uma estratégia demorada e pouco inteligente de se solucionar o problema, especialmente se o número de pontos for grande (HOUGH, P. V. C., 1962).

A partir da Transformada de Hough, proposta por Paul Hough em 1962 (HOUGH, P. V. C., 1962), é possível detectar as retas de uma imagem de maneira mais eficiente. A Transformada de Hough define um mapeamento entre o espaço da imagem e o espaço de parâmetros *slope-intercept* para a detecção de retas, no qual cada reta da imagem é definida pela Equação 4.1:

$$y = a.x + b. \quad (4.1)$$

Cada reta é representada no espaço de parâmetros por um ponto (a, b) , em que a (*slope*) é a inclinação da reta e b (*intercept*) corresponde ao valor de y quando $x = 0$. A Figura 4.1 ilustra uma reta no espaço da imagem e sua respectiva equação.

A Transformada de Hough (HOUGH, P. V. C., 1962) apresenta uma importante propriedade: pontos colineares na imagem correspondem a retas concorrentes no espaço de parâmetros, as quais se interceptam na posição do espaço de parâmetros que representa a reta da imagem à que os pontos pertencem. A partir dessa propriedade, para detectar as retas em uma imagem basta representar todos os pontos presentes na imagem no espaço de parâmetros (a, b) e encontrar as posições em que há os maiores números de intersecções. Essas posições indicarão retas que satisfazem à Equação e cujos valores de a e b correspondem às coordenadas dos pontos de intersecção no espaço de parâmetros (HOUGH, P. V. C., 1962).

Como nos computadores deve-se utilizar um espaço de parâmetros discreto, uma matriz acumuladora

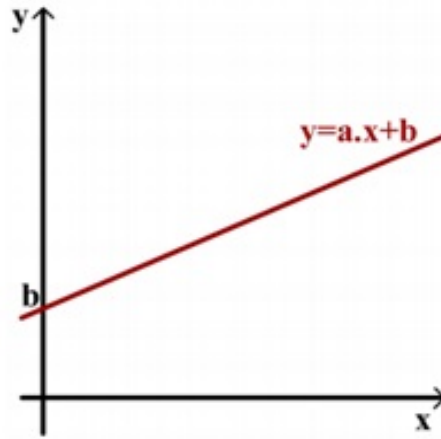


Figura 4.1: Equação da reta na imagem.

é utilizada para representar esse espaço. Dessa forma, para se detectar as retas de uma imagem basta procurar por posições na matriz com altos valores acumulados, pois estes representam as posições onde várias retas se interceptam.

O grande problema da Transformada de Hough (HOUGH, P. V. C., 1962) com o espaço de parâmetros *slope-intercept*, é que esses parâmetros são ilimitados (DUDA; HART, 1971), o que inviabiliza seu uso em computadores. Por exemplo, caso uma reta paralela ao eixo y seja representada no espaço *slope-intercept*, o parâmetro *intercept* não poderá ser calculado, inviabilizando o processo.

Para resolver o problema, Duda e Hart (1971) propuseram a utilização da equação da reta em coordenadas polares, na qual cada reta da imagem é representada por dois parâmetros (θ, ρ) (Equação 4.2) em que θ é o ângulo entre o vetor normal à reta e o eixo das abscissas e ρ é a distância da reta em relação à origem. A Figura 4.2 ilustra a representação de uma reta por coordenadas polares.

$$\rho = x.\cos(\theta) + y.\sen(\theta) \quad (4.2)$$

Restringindo θ ao intervalo $[0; \pi[$, os parâmetros θ e ρ de uma reta serão únicos. Dessa forma, toda reta no espaço $x - y$ corresponde a um único ponto no espaço de parâmetros (θ, ρ) . Cada ponto da imagem é então representado por uma curva senoidal.

Senóides que representam pontos colineares da imagem no espaço de parâmetros (θ, ρ) apresentam um ponto de intersecção (θ_0, ρ_0) em comum, o qual representa a reta da imagem à qual os pontos pertencem (DUDA; HART, 1971). Portanto, para identificar as retas da imagem, basta encontrar os pontos de intersecção das senóides no espaço de parâmetros.

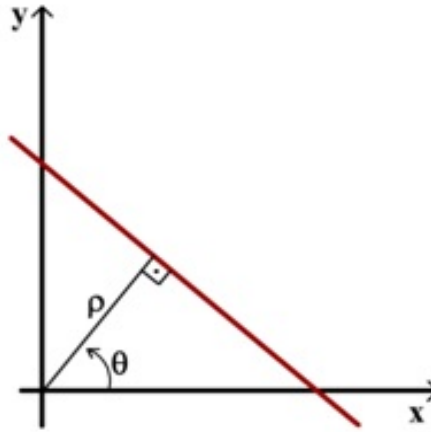


Figura 4.2: Representação da reta por coordenadas polares.

4.1 O Descritor HTS

O descritor de formas proposto em (SOUZA; MARANA, 2013) é baseado na transformada de Hough, cujo o espaço de Hough é gerado a partir dos pontos da borda do objeto de modo que pontos pertencentes a segmentos de reta da borda apresentarão picos em suas senóides do espaço de Hough e os pontos de regiões arredondadas da borda apresentarão quantidades de intersecções bem distribuída ao longo de suas senóides. Esse método pode ser dividido em três etapas: pré-processamento, extração de características e reconhecimento.

Na etapa de pré-processamento é segmentada a borda de cada imagem binária. Na etapa de extração de características é computada a matriz acumuladora, que possui 180 colunas referentes à variação do ângulo e um número de linhas que varia de acordo com o tamanho da imagem. Cada ponto da borda gera uma senóide no espaço de Hough, que é registrada no espaço de parâmetros incrementando-se uma unidade nas posições da matriz acumuladora por onde ela passa. Para cada coordenada (x, y) da borda, calcula-se um valor de ρ para cada ângulo θ variando de 0° a 179° . A Figura 4.3 mostra a imagem de entrada (a), a imagem após a extração das bordas (b) e o espaço de Hough (matriz acumuladora) gerada a partir da imagem binarizada, na qual quanto mais clara a região dessa imagem, mais senóides passam por essa região.

Após o cálculo da matriz acumuladora, cada ponto p pertencente à borda é associado a um histograma contendo a quantidade de intersecções encontradas em sua senóide e, então, calcula-se a média e o desvio padrão desse histograma. Cada ponto ρ da borda passa a ser representado por esses dois valores e os histogramas são descartados. A Figura 4.4 ilustra uma imagem e o respectivo histograma do ponto destacado. Nota-se que os histogramas de pontos de regiões semelhantes são parecidos, enquanto histogramas de regiões diferentes são diferentes. Além disso, pontos que pertencem a segmentos relativamente retos apresentam grandes picos em seus histogramas.

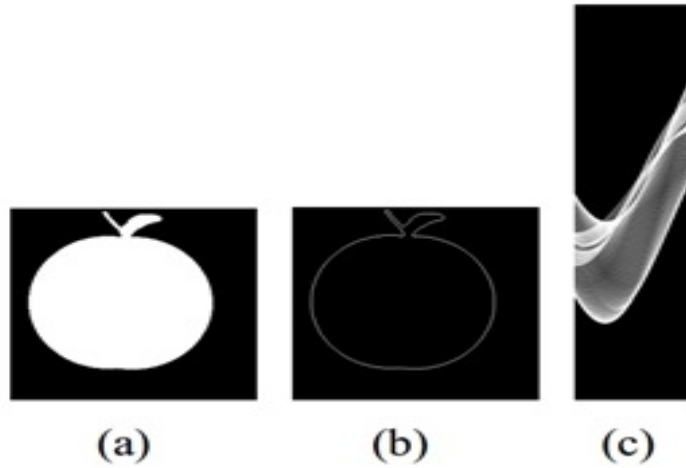


Figura 4.3: (a) Silhueta de uma maçã da base de imagens MPEG-7 PartB (JEANNIN; BOBER, 1999) (b) Borda da maçã segmentada; (c) espaço de Hough (matriz acumuladora) gerado a partir dos pontos da borda. (SOUZA; MARANA, 2013).

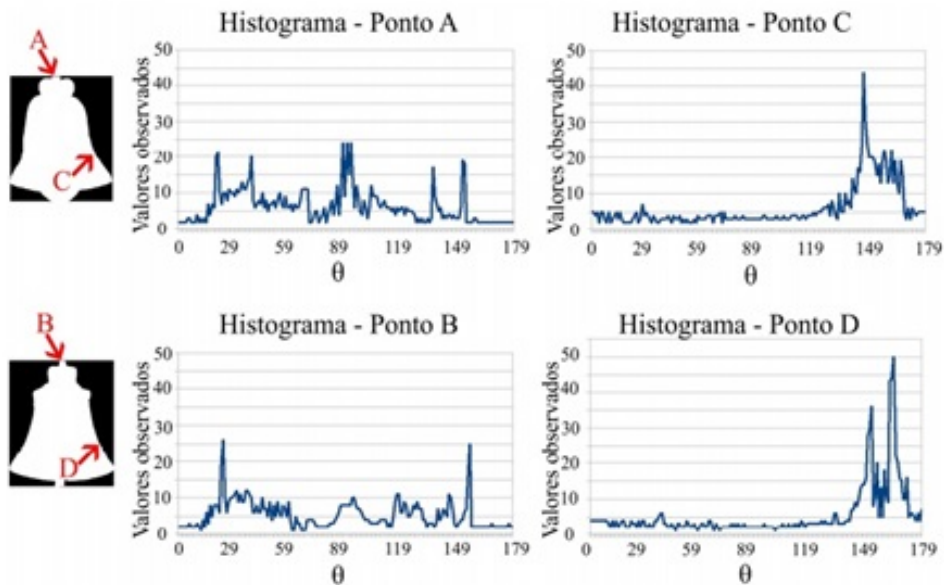


Figura 4.4: Histogramas associados a quatro pontos da borda de dois objetos de mesma classe (sinos) da base MPEG-7 Part B (JEANNIN; BOBER, 1999). Adaptado de (SOUZA; MARANA, 2013).

A partir de um ponto inicial da borda, percorrendo-se esta em sentido horário, pode-se construir duas funções unidimensionais para o objeto a partir da média e desvio padrão associados a cada ponto. A Figura 4.5 mostra um conjunto de objetos de diferentes classes (a,b,d) e de classes semelhantes (b,c) e suas respectivas funções associadas. Pode-se observar na Figura 4.5 que a imagem (a) apresenta quatro picos referentes aos quatro vértices do quadrado, pois cada um deles pertence a dois grandes segmentos de reta, enquanto os demais pontos do quadrado pertencem a apenas um e, apesar de apresentarem desvios elevados, estes são menores que os dos vértices.

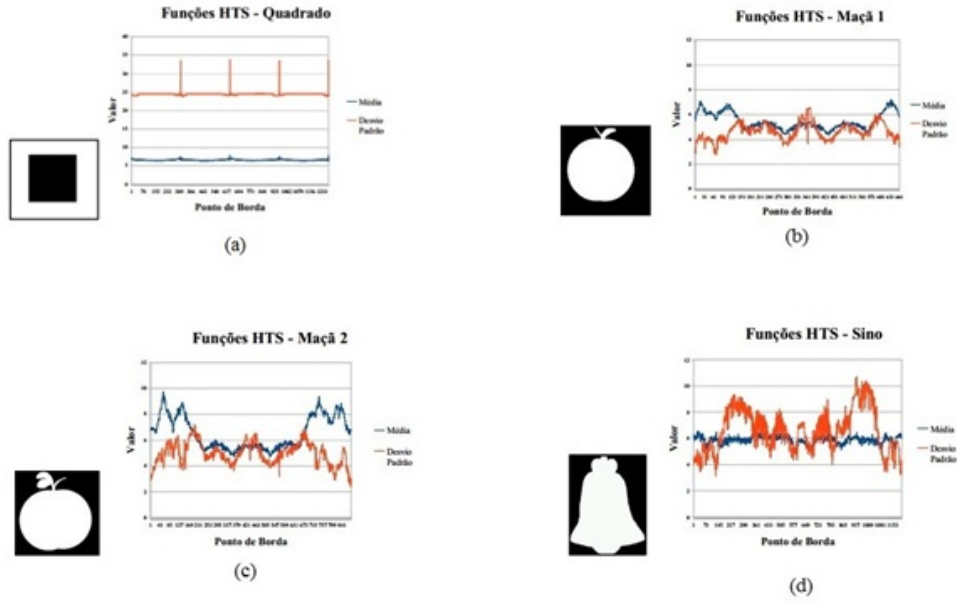


Figura 4.5: Funções obtidas para silhuetas da base Kimia-216 (SEBASTIAN et al., 2004). Adaptado de (SOUZA; MARANA, 2013).

Após a associação das funções unidimensionais a cada objeto, para gerar o vetor de características desse objeto, as funções unidimensionais são amostradas em posições igualmente espaçadas entre si. Dessa forma, cada posição do vetor de características é composta por dois valores, a média e o desvio padrão. A Figura 4.6 ilustra o processo de amostragem das funções unidimensionais de uma imagem da base Kimia-216 (SEBASTIAN et al., 2004).

A etapa final é o reconhecimento, em que os vetores de características da média e desvio padrão de cada objeto são comparados de acordo com a Equação 4.3:

$$L1(s, r) = \sum_{i=1}^N [|s_{im} - r_{im}| + |s_{id} - r_{id}|], \quad (4.3)$$

em que s e r são vetores de características, s_{im} e r_{im} são as posições da média presentes nas posições s_i e r_i dos vetores, s_{id} e r_{id} são os valores de desvio padrão presentes na posição s_i e r_i dos vetores de características e N corresponde ao tamanho do vetor.

Cada objeto é comparado com objetos da base de dados para que se possa determinar a sua classe a partir da menor distância encontrada entre os vetores de características utilizando a Equação 4.3.

Uma característica importante do método é que cada comparação entre dois vetores de características é realizada k vezes, em que k corresponde ao número de elementos do vetor. A cada comparação um dos vetores é rotacionado, ou seja, troca-se a posição dos dados do vetor de forma circular. A menor distância encontrada nos k casamentos é o resultado final. Dessa forma, obtém-se invariância à rotação.

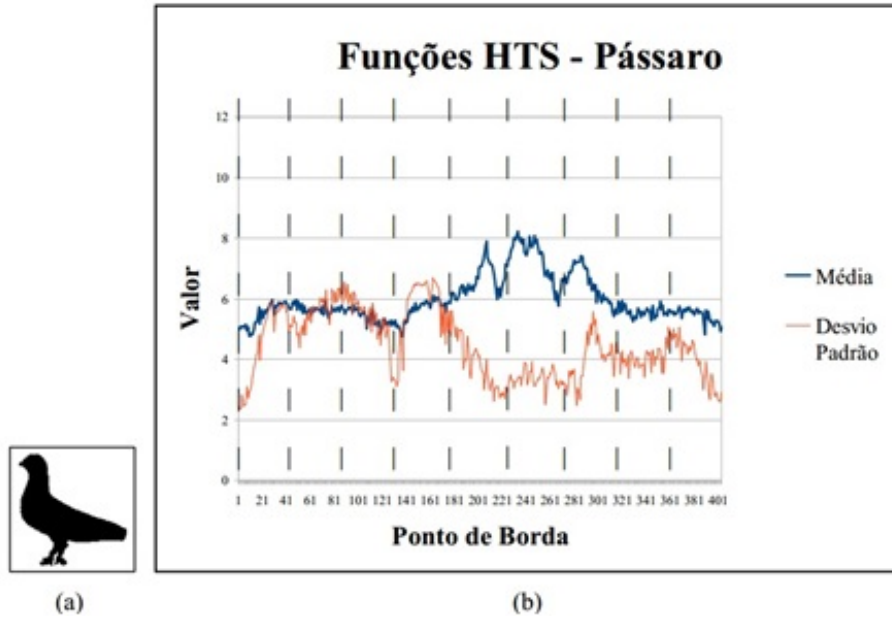


Figura 4.6: Em (a) é exibida uma imagem da base Kimia-216 (SEBASTIAN et al., 2004) e em (b) são exibidas suas funções unidimensionais. As retas verticais tracejadas indicam as posições onde as funções foram amostradas para se gerar o vetor de características.

4.2 O Descritor HTSn

O descritor HTSn foi criado a partir de uma modificação na etapa de extração de características do HTS. Essa modificação consiste em, ao se percorrer uma senóide no espaço de Hough referente a um ponto da borda, calcula-se o histograma somando-se todos os valores do espaço de Hough encontrados em uma vizinhança $n \times n$. Portanto, o método foi chamado de HTSn (neighborhood) (SOUZA; MARANA, 2013).

As Figuras 4.7(a) e 4.7(b) ilustram o processo de construção do histograma utilizando o HTS. O valor do histograma calculado em uma determinada posição (θ, ρ) corresponde a soma de todos os valores encontrados na vizinhança 3×3 desta posição, representada na figura.

Experimentos apresentados em (SOUZA; MARANA, 2013) mostraram que o HTSn apresenta melhores resultados do que o HTS. Esta melhora era esperada, pois ao se empregar um espaço discreto (matriz acumuladora) para a representação de um espaço contínuo, trabalha-se com aproximações e muitas vezes pontos pertencentes a segmentos de retas da imagem possuem senóides que não passam exatamente nas posições do espaço correspondentes aos segmentos na imagem, mas sim nas posições vizinhas.

Experimentos realizados em (SOUZA; MARANA, 2013), utilizando a base de dados Kimia-216 (SEBASTIAN et al., 2004) e a base MPEG-7 Part B, compararam o HTS e sua variação HTSn com diversos métodos da literatura, como o Beam Angle Statistics (BAS) (ARICA; VURAL, 2003), Shape Context (SC) (BELONGIE; MALIK, 2000), Multiscale Fractal Dimension (MFD) (BACKES; BRUNO, 2006), Tensor Scale (TS) (MIRANDA et al., 2005), Fourier Descriptors (FD) (ZHANG; LU, 2002) e Contour Saliency (CS) (TORRES; FALCÃO, 2007) e os resultados obtidos mostraram que esses descritores além

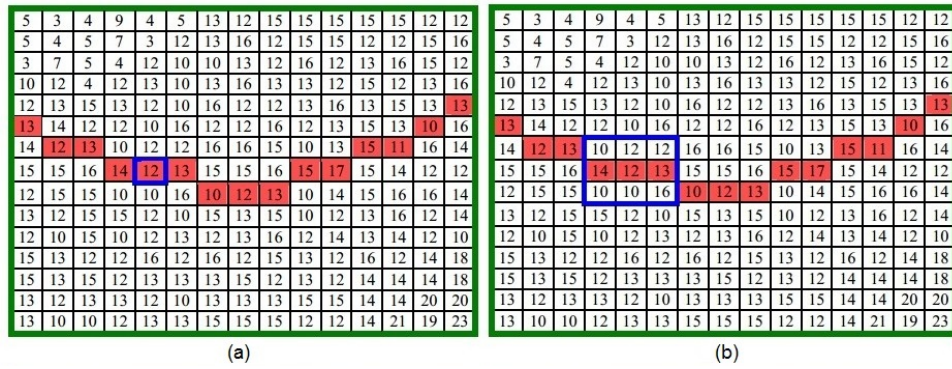


Figura 4.7: As matrizes com bordas verdes representam uma parte ampliada do espaço de Hough final de um dado objeto em dois momentos distintos, (a), no processo de construção (pelo descritor HTS) do histograma do ponto de borda associado à senóide cujas posições na matriz aparecem destacadas em vermelho. Os valores nas células com bordas azuis são armazenados, obedecendo a ordem entre as matrizes verdes, isto é, (a), nos *bins* do histograma do ponto. Já a matriz de (b) ilustra o processo de construção (agora pelo descritor HTSn) do histograma do ponto de borda associado à senóide cujas posições na matriz aparecem destacadas em vermelho. O valor armazenado em cada *bin* do histograma do ponto de borda corresponde à soma do valor na respectiva posição da senóide com os valores nas oito posições adjacentes a esta no espaço de Hough. Adaptado de (SOUZA; MARANA, 2013)

de serem precisos, são rápidos, pois possuem complexidade linear (SOUZA; MARANA, 2013).

Capítulo 5

Métodos Propostos

5.1 Descritores MHTS e MHTSn

MHTS e MHTSn são descritores criados a partir de uma abordagem multiescala dos métodos HTS e HTSn originais, propostos por Souza e Marana (2013).

Para a criação do MHTS (Multiscale HTS), na etapa de extração de características, na qual é computada a matriz acumuladora que possui 180 colunas respectivas a variação de 0° a 179° , alterou-se o valor de incremento de θ para cada cálculo de ρ . No HTS original, varia-se θ de 0° a 179° e, para cada valor de θ , calcula-se um valor de ρ . No MHTS, o método HTS é aplicado E vezes, sendo que em cada nova iteração os parâmetros de discretização do espaço de Hough são modificados conforme as Equações 5.1 e 5.2, em que Δ é a variação de θ e ρ para a escala i que varia de 1 a E . Dessa forma, o espaço de Hough irá diminuir a cada iteração.

$$\Delta_\theta(i) = i, \quad \text{com } i = 1, \dots, E \quad (5.1)$$

$$\Delta_\rho(i) = i, \quad \text{com } i = 1, \dots, E \quad (5.2)$$

A Figura 5.1 ilustra os espaços de Hough obtidos para $n = 5$.

Pode-se observar que tanto ρ quanto θ diminuem a cada iteração, mantendo a estrutura do espaço de Hough. Dessa forma, conforme a escala de incremento aumenta, as informações de região se condensam, tornando o método mais robusto.

Como foi dito, para cada um dos espaços de Hough calculados, executam-se as mesmas etapas do HTS original. Assim, obtém-se um descritor para cada uma das escalas e, então, concatena-se esses descritores formando um único descritor de tamanho $E \times N$, em que E é o número de iterações (escalas) e N é a dimensão do descritor HTS em cada escala. O diagrama apresentado na Figura 5.2 resume os passos do método MHTS.

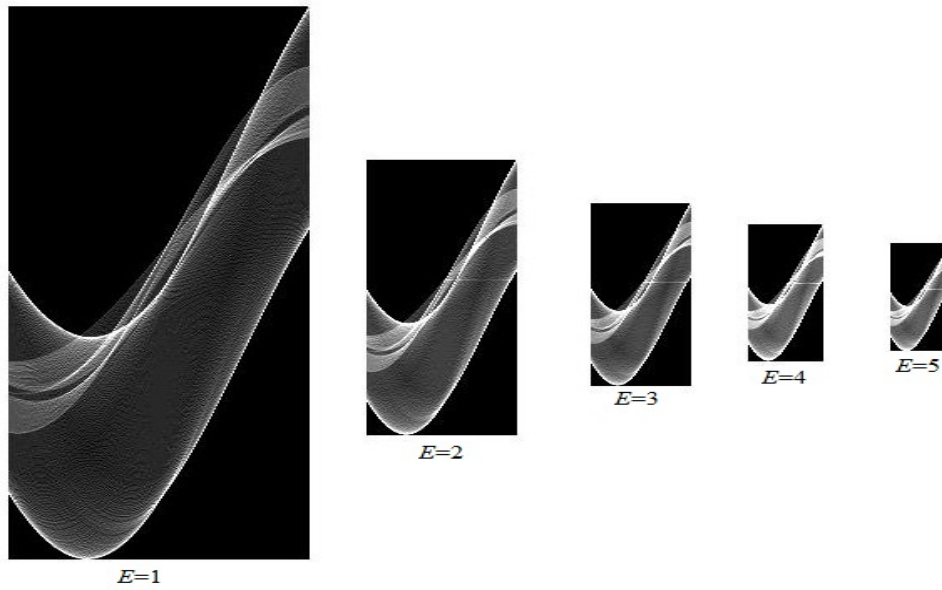


Figura 5.1: Espaço de Hough obtido para diferentes escalas (E): 1,2,3,4 e 5.

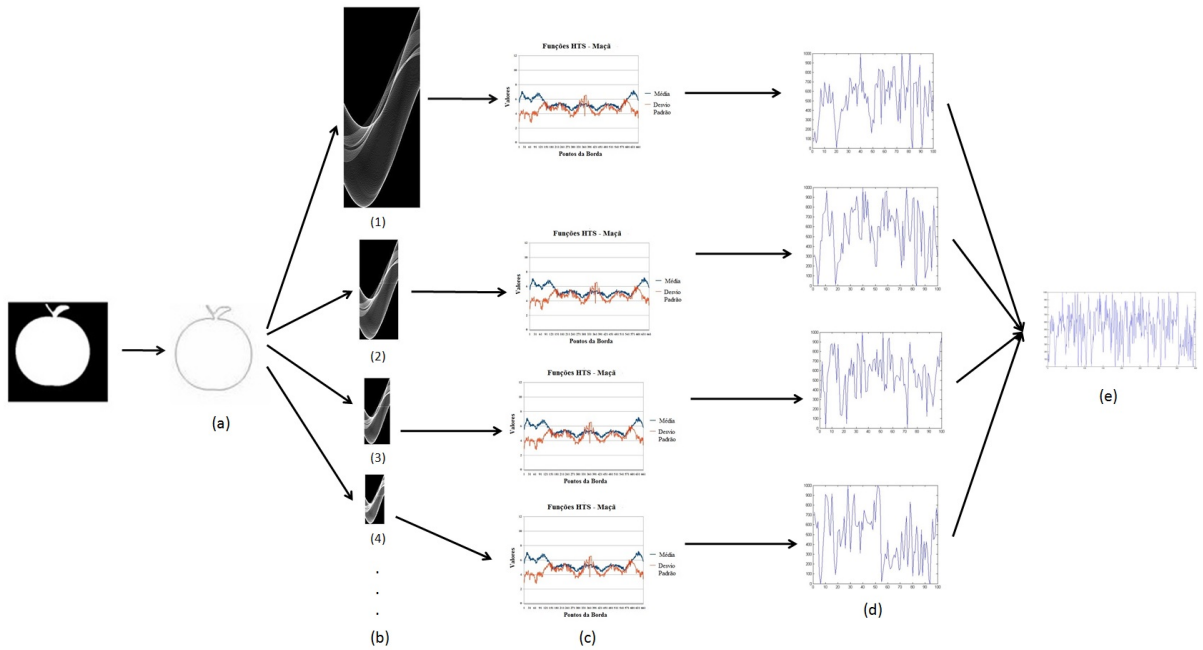


Figura 5.2: Diagrama dos passos do método MHTS: (a) - Extração das bordas da imagem de entrada; (b) - Cálculo dos espaços de Hough para cada uma das n escalas; (c) Funções calculadas a partir do espaço de Hough, (d) - Obtenção dos descritores HTS para cada uma das n escalas; (e) - Concatenação dos n descritores HTS em um único descritor MHTS.

O MHTSn (Multiscale HTSn) foi criado a partir do MHTS. Porém, como o HTSn já utiliza informações da vizinhança de cada ponto pertencente a uma senoide no espaço de Hough, condensar essas informações em escalas inferiores não apresentou melhores resultados que o HTSn. Dessa forma, optou-se por utilizar o HTSn para a escala 1 e para as demais escalas, o MHTS. Assim, obtêm-se um descritor baseado no

HTSn para a escala de variação 1, e outros 4 descritores para as outras 4 escalas de variação do MHTS. Esses n descritores são então concatenados, gerando-se um único descritor de dimensão $n \times N$, de forma análoga ao MHTS.

5.2 Descritores SHTS e SHTSn

SHTS e SHTSn são descritores criados a partir de uma abordagem espaço-escala dos métodos HTS e HTSn.

Para a criação do SHTS e do SHTSn, foi adicionado um novo passo na etapa de pré-processamento, na qual antes da imagem ter sua borda segmentada e binarizada, é feita a sua convolução com o núcleo gaussiano da Equação 2.2.

De maneira análoga a técnica da pirâmide gaussiana, é feita a convolução da imagem original com núcleos gaussianos definidos por $\sigma = 2^i$, para $i = 1, 2, \dots$. Porém, diferentemente da técnica da pirâmide gaussiana, a imagem não é reduzida a cada iteração, ou seja, mantém-se a sua resolução original.

A Figura 5.3 ilustra o procedimento. Pode-se observar que a cada convolução com o núcleo gaussiano a imagem se torna mais borrada. Após a convolução, cada imagem resultante é binarizada conforme ilustra a Figura 5.4.

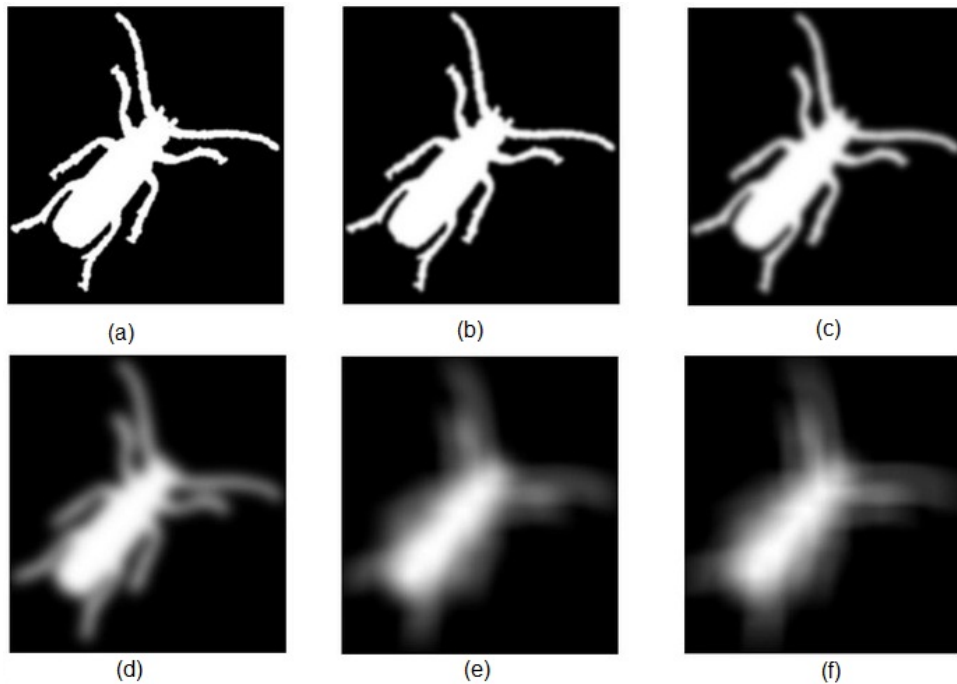


Figura 5.3: Exemplos de imagens obtidas após a convolução com o núcleo gaussiano: (a) $\sigma = 2$, (b) $\sigma = 4$, (c) $\sigma = 8$, (d) $\sigma = 16$, (e) $\sigma = 32$, (f) $\sigma = 64$.

Com esse procedimento, a borda é suavizada e, com isso, eliminam-se detalhes e ruídos que muitas vezes podem prejudicar a eficiência do método. O HTS e HTSn são aplicados em cada uma das imagens

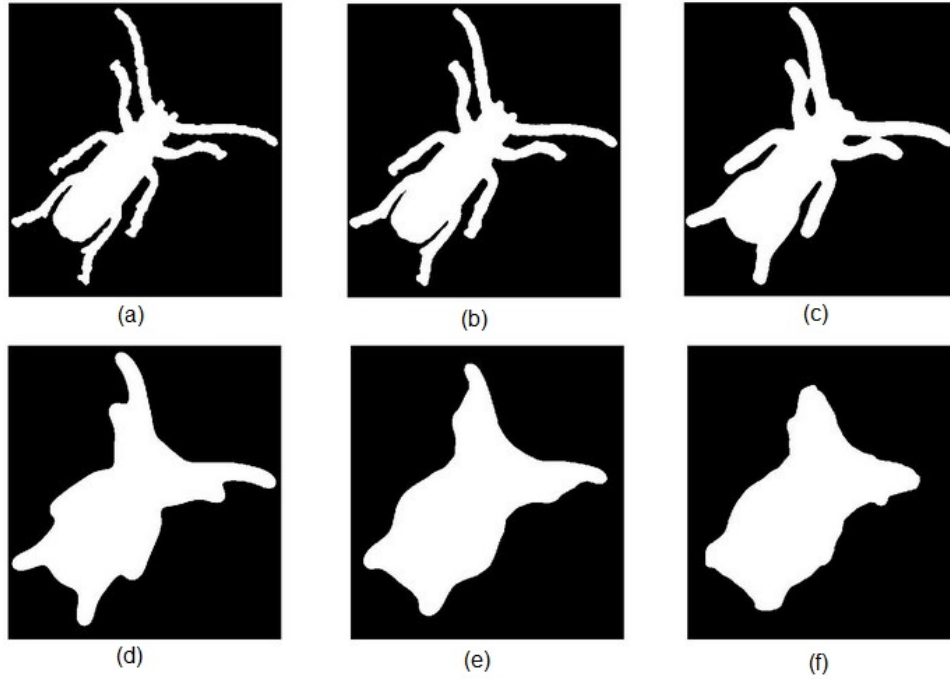


Figura 5.4: Imagens binarizadas após a convolução com o núcleo gaussiano: (a) $\sigma = 2$, (b) $\sigma = 4$, (c) $\sigma = 8$, (d) $\sigma = 16$, (e) $\sigma = 32$, (f) $\sigma = 64$.

resultantes, seus descritores são extraídos, concatenados e então comparados. O diagrama apresentado na Figura 5.5 resume os passos do método SHTS e SHTSn.

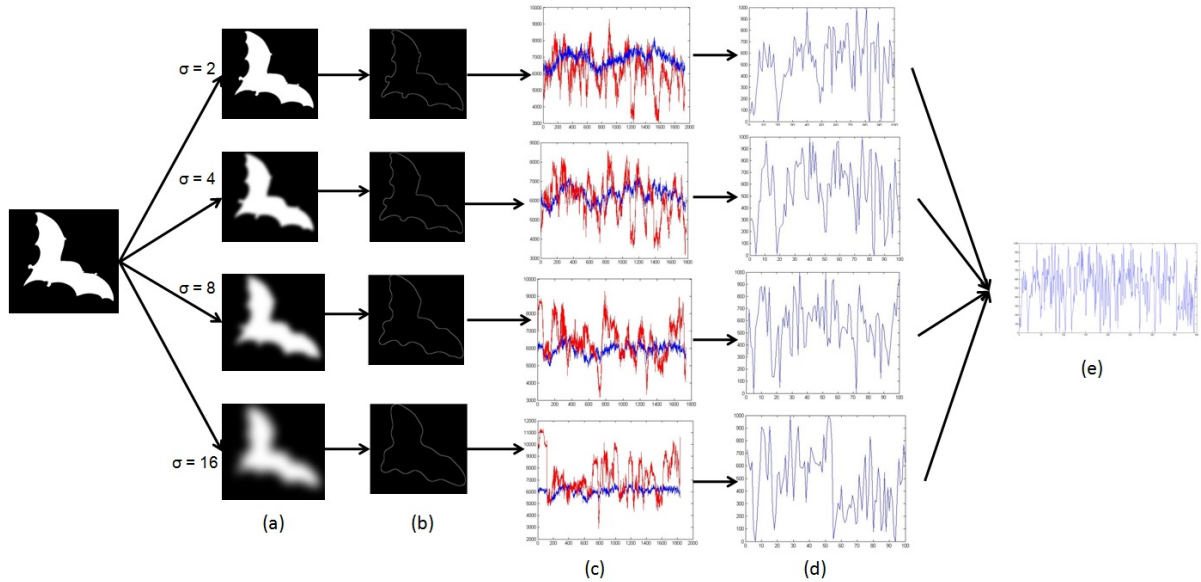


Figura 5.5: Diagrama do método espaço-escala proposto, (a) convolução da imagem com filtro gaussiano, (b) extração da borda, (c) Funções calculadas a partir do espaço de Hough, (d) Obtenção dos descritores a partir da amostragem e (e) Descritores concatenados.

Os métodos SHTS e SHTSn podem ser muito úteis em tarefas como a recuperação de imagens, já que as imagens que não satisfazem o critério de recuperação podem ser descartadas a partir de suas escalas

mais grosseiras tornando o processo mais rápido.

5.3 Abordagem Hierárquica

Outras versões dos métodos foram criadas aplicando-se a multiescala em seus vetores de características. Dessa forma, ao invés de todos os vetores possuírem o mesmo tamanho e cada escala contribuir de maneira igualitária, o tamanho do vetor de características passou a diminuir conforme a escala aumenta e o espaço de Hough se torna menor. Com essa abordagem a etapa de comparação se torna muito mais eficiente, uma vez que há um número menor de valores a serem comparados, tornando as versões multiescala altamente indicadas para a tarefas de recuperação de imagens. Essas novas versões são chamadas MHTS_v, MHTS_{nv}, SHTS_v e SHTS_{nv}.

Capítulo 6

Resultados Experimentais

Os métodos propostos no Capítulo 6 foram avaliados sobre as seguintes bases de dados: Kimia-216 (SEBASTIAN et al., 2004), MPEG-7 Part B (JEANNIN; BOBER, 1999) e na MNIST (LECUN et al., 1998), e na FSCT (*Frontal Sinus Computed Tomography*) database, criada pelo grupo de pesquisa Recogna-UNESP.

É importante salientar que nos experimentos realizados os melhores resultados foram obtidos quando os vetores de características apresentavam tamanho 50. Esse tamanho foi obtido a partir de testes efetuados em (SOUZA; MARANA, 2013), em que o tamanho desses vetores foi variado de 10 a 70.

Para analisar o desempenho dos descritores testados sobre as bases de dados, foram utilizadas duas curvas, a Precisão x Revocação (MÜLLER et al., 2001) e a Separabilidade Multiescala (TORRES; FALCÃO, 2007), utilizadas em outros trabalhos encontrados na literatura para avaliar o desempenho de descritores de formas.

A curva Precisão x Revocação (MÜLLER et al., 2001) mede a taxa de Precisão, ou seja, a razão do número de imagens da classe correta retornadas em uma consulta sobre o total de imagens retornadas pela consulta, e a taxa de Revocação, ou seja, a razão entre o total de imagens da classe correta retornadas por uma consulta sobre o total de imagens da classe correta presentes na base de dados.

A curva Separabilidade Multiescala (TORRES; FALCÃO, 2007) indica quão separados estão os agrupamentos de vetores de características de classes distintas no espaço de características, dessa forma, quanto mais separados, melhor o descritor. Por exemplo, dada uma base de dados B com k objetos organizados em classes, a Separabilidade S de um descritor para uma classe C será definida da seguinte forma: seleciona-se arbitrariamente um objeto O de uma classe C , em seguida normaliza-se as distâncias calculadas entre o objeto O e todos os demais objetos da base para o intervalo de 0 a 1. Após a normalização, o raio de busca R é incrementado por um valor fixo a cada iteração, nesse caso o valor inicial é de 0,02, e o incremento a cada iteração é de 0,02. A cada iteração verifica-se o número de objetos N_o cujas distâncias são menores que R e não pertencem a classe C . Para valor do raio de busca R , a Separabilidade S da classe C será definida de acordo com a Equação 6.1.

$$S = 1 - \frac{No}{k} \quad (6.1)$$

em que k é o número total de objetos na base menos o total de objetos pertencentes a classe C .

A partir da definição da Separabilidade de uma classe C , a Separabilidade de um descritor D é dada pela média das Separabilidades para todas as classes da base B .

Para a avaliação da precisão na recuperação de imagens o critério de avaliação chamado Bulls-Eye foi utilizado (LIU et al., 2012). Nesse critério, dado um conjunto de dados S , com c classes de n imagens, cada imagem de S é utilizada como imagem de consulta e o número de imagens corretas (que pertencem a mesma classe da imagem de consulta) é computado para as $2n$ melhores comparações. A melhor pontuação é obtida quando $n^2 \times c$ casos positivos são encontrados na base. Dessa forma, o total de objetos recuperados corretamente é dividido pelo total de recuperações corretas possíveis, sendo esse valor, a taxa de acerto de acordo com o critério Bulls-Eye (LIU et al., 2012).

6.1 Primeiro Experimento - Base de Dados Kimia-216

O primeiro experimento foi realizado sobre a base de dados Kimia-216, composta por 216 imagens divididas em 18 classes de 12 imagens. A Figura 6.1 mostra algumas imagens dessa base.

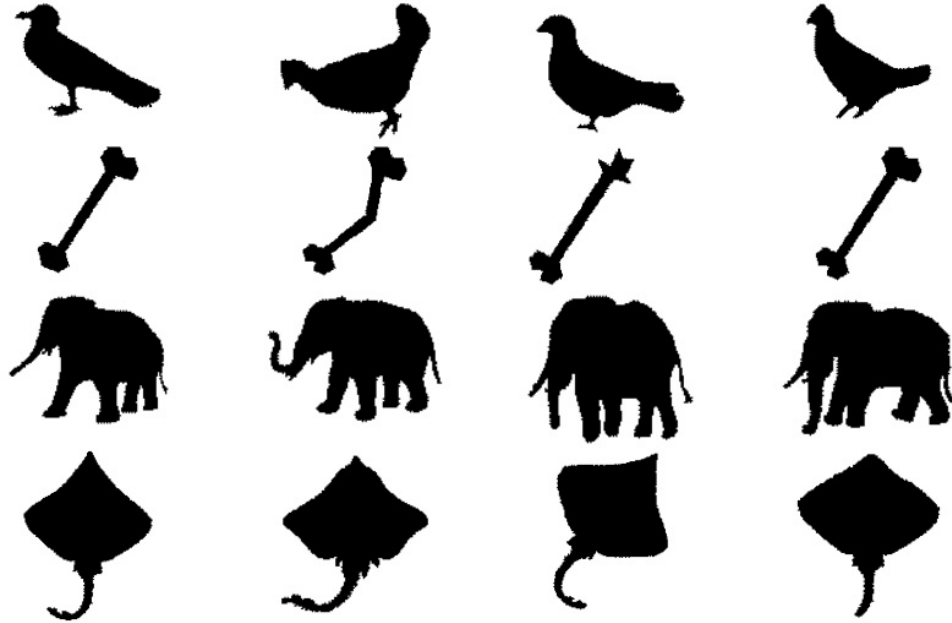


Figura 6.1: Exemplos de imagens da Base Kimia-216 (SEBASTIAN et al., 2004).

Inicialmente, com o objetivo de verificar a precisão de cada escala, optou-se por aplicar cada uma das cinco escalas dos MHTS individualmente na base de dado Kimia-216 e então, foram calculadas suas curvas Precisão x Revocação e seus valores de Bulls-Eye.

A Figura 6.2 mostra os resultados obtidos para cada uma das escalas do MHTS na base Kimia-216. Pode-se observar que quando aplicadas na base Kimia-216 a perda de precisão é bem reduzida ao longo das escalas, sendo mais evidente apenas entre as duas primeiras escalas. O mesmo pode ser observado para os resultados utilizando medida Bulls-Eye, conforme pode ser observado na Tabela 6.1.

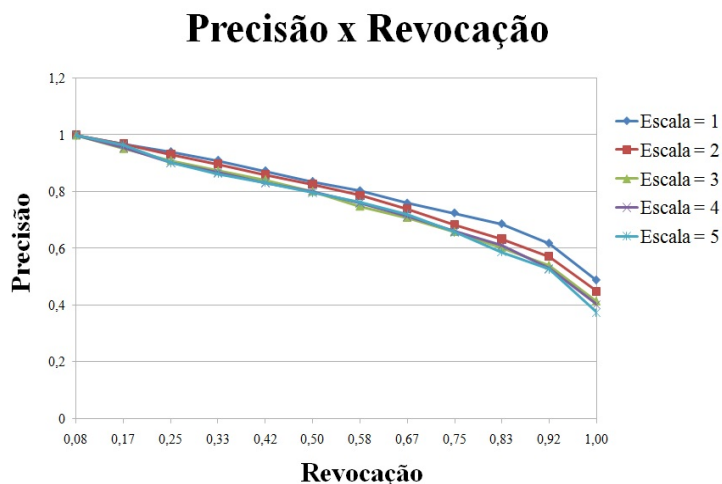


Figura 6.2: Curvas Precisão x Revocação das escalas individuais do método MHTS para a base Kimia-216 (JEANNIN; BOBER, 1999)

Tabela 6.1: Resultados Bulls-Eye para cada escala do método MHTS para a base Kimia-216

Escala	Taxa de Acerto
1	85,22%
2	83,06%
3	80,44%
4	80,59%
5	80,83%

A Figura 6.3 ilustra as curvas Precisão x Revocação da base Kimia-216 utilizando o método MHTS e MHTSv (MHTS com variação no tamanho dos descritores). Quanto mais acima está a curva, melhor o descritor. Pode-se observar que os melhores resultados foram obtidos pelos métodos BAS (ARICA; VURAL, 2003), Momentos de Zernike (HU, 1962), MHTS, MHTSv, HTS e HTSn, sendo que o MHTS apresentou resultados bem próximos do Momentos de Zernike e HTSn.

A Figura 6.4 mostra as curvas Separabilidade Multiescala de alguns descritores para a base Kimia-216. Quanto mais acima está a curva, melhor o método. Pode-se observar que o MHTS e o MHTSv apresentaram resultados bastante superiores aos do BAS (ARICA; VURAL, 2003) e Momentos de Zernike (HU, 1962).

Em relação às curvas Precisão x Revocação, os descritores MHTS e MHTSv, junto com o BAS (ARICA; VURAL, 2003), apresentaram curvas superiores aos demais descritores testados, inclusive suas versões monoescala.

A curva Precisão x Revocação do descritor MHTS ficou um pouco abaixo da curva do BAS (ARICA; VURAL, 2003). Quando comparado às curvas do HTSn e do Momentos de Zernike (HU, 1962), o MHTS apresentou uma curva superior à de ambos.

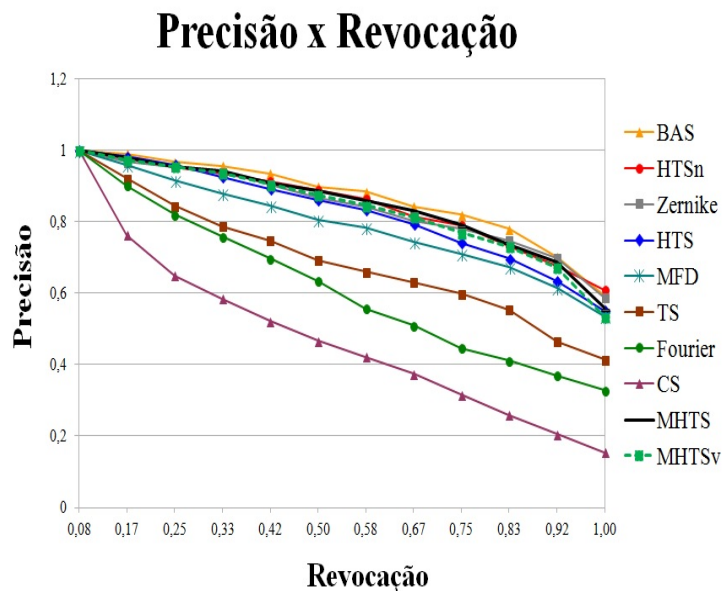


Figura 6.3: Curvas Precisão x Revocação do MHTS, MHTSv e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

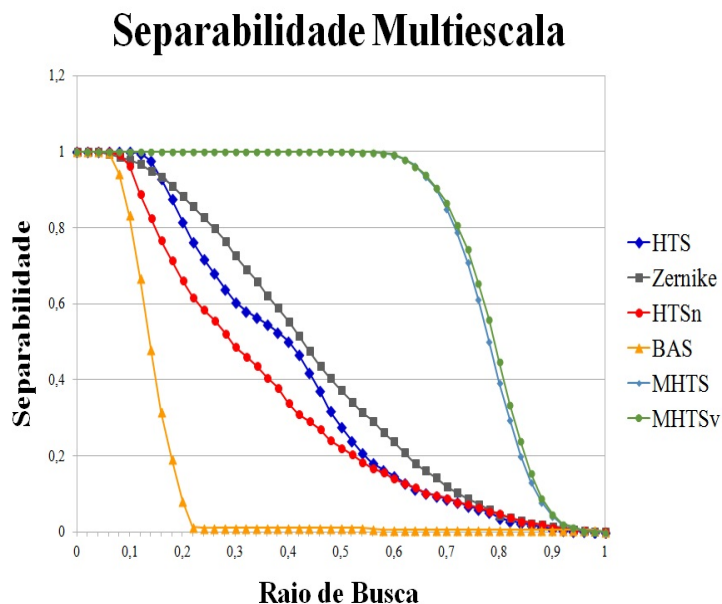


Figura 6.4: Curvas Separabilidade Multiescala do MHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

Além disso, para os valores de revocação até 25% (início das curvas), o desempenho do BAS (ARICA; VURAL, 2003), do HTS, do HTSn e do MHTS foram praticamente iguais. Isto é um fato importante

porque várias aplicações que utilizam descritores de formas como sistemas de busca de imagens por conteúdo (*Content-Based Image Retrieval* - CBIR) operam principalmente nesta região da curva (SOUZA; MARANA, 2013). Nesse tipo de sistema, a grande prioridade é que as imagens obtidas pela busca sejam da mesma classe, independente do fato de todas as imagens da classe de consulta presentes na base terem sido retornadas.

Tratando-se dos resultados Separabilidade Multiescala (avaliada apenas para os métodos que apresentaram os melhores resultados na Precisão X Revocação), o HTS, HTSn e MHTS superaram os resultados do BAS (ARICA; VURAL, 2003), sendo que os melhores resultados foram obtidos pelo MHTS e MHTSv.

As Figuras 6.5 e 6.6 mostram os resultados obtidos para os métodos MHTSn e MHTSv. Os métodos apresentaram resultados um pouco melhores que o MHTS, mas sem grandes melhorias para ambas as curvas Precisão x Revocação e Separabilidade Multiescala.

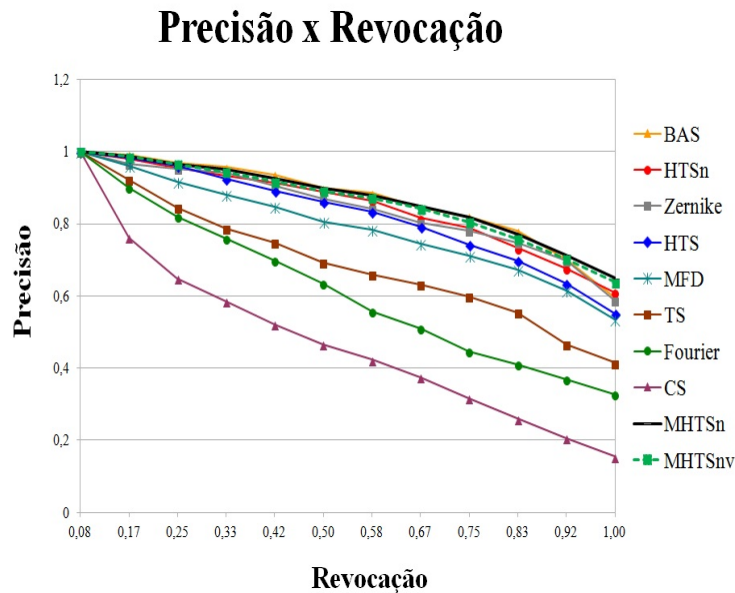


Figura 6.5: Curvas Precisão x Revocação do MHTSn e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

As Figuras 6.7 e 6.8 mostram os resultados obtidos para os métodos SHTS e SHTSv. Pode-se observar nas curvas Precisão x Revocação que ambos os métodos apresentaram resultados bem próximos do BAS (ARICA; VURAL, 2003), sendo que suas curvas ficam praticamente sobrepostas em alguns valores de revocação para o SHTS, principalmente para os valores até 25%, enquanto o SHTSv apresentou uma curva superior à do BAS (ARICA; VURAL, 2003) na maioria dos pontos. Porém, a grande melhoria pode ser observada nas curvas Separabilidade Multiescala, na qual o SHTS e o SHTSv apresentaram os melhores resultados, superando as curvas do método Momentos de Zernike (HU, 1962), que apresenta um bom desempenho segundo este critério. Os resultados dos métodos SHTSn e SHTSv foram muito similares aos do SHTS, não mostrando nenhuma melhoria significativa.

Por fim, a Figura 6.9 mostra uma comparação entre as curvas de Precisão x Revocação das abordagens

Separabilidade Multiescala

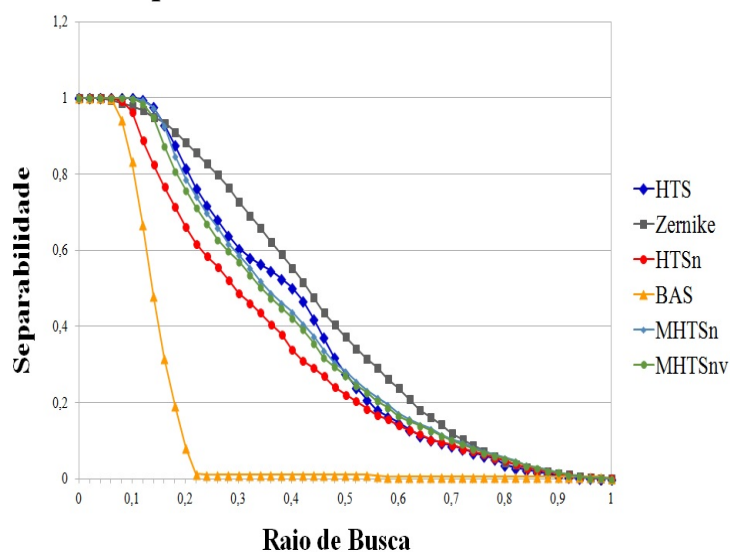


Figura 6.6: Curvas Separabilidade Multiescala do MHTSn e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

Precisão x Revocação

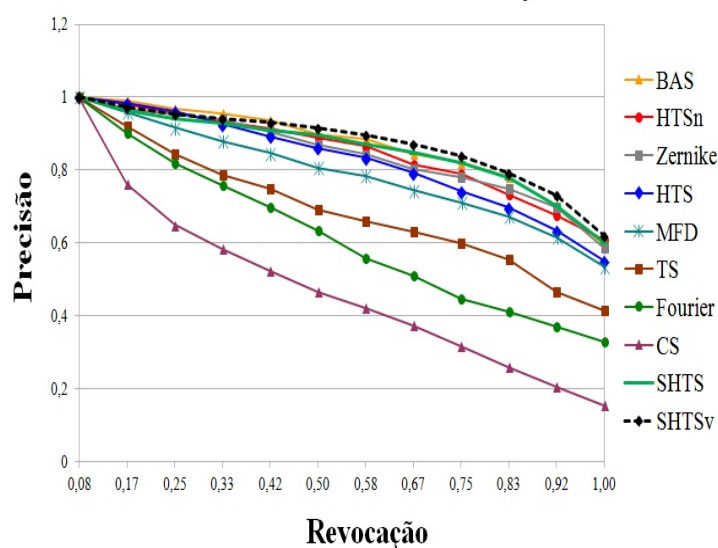


Figura 6.7: Curvas Precisão x Revocação do SHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

propostas. Pode-se observar que as curvas mais altas são dos métodos SHTSv e MHTSn. A Tabela 6.2 apresenta uma comparação da medida de similaridade Bulls-Eye para os métodos propostos nesta dissertação com os demais métodos multiescala apresentados. Pode-se observar na Tabela 6.2 que os métodos MHTSn, MHTSnv e SHTSv obtiveram os melhores resultados entre os métodos propostos, sendo superiores até mesmo ao BAS. A versão multiescala do HTSn apresentou uma melhoria de precisão de aproximadamente 1,72%, enquanto a versão espaço-escala obteve uma melhoria de aproximadamente

Separabilidade Multiescala

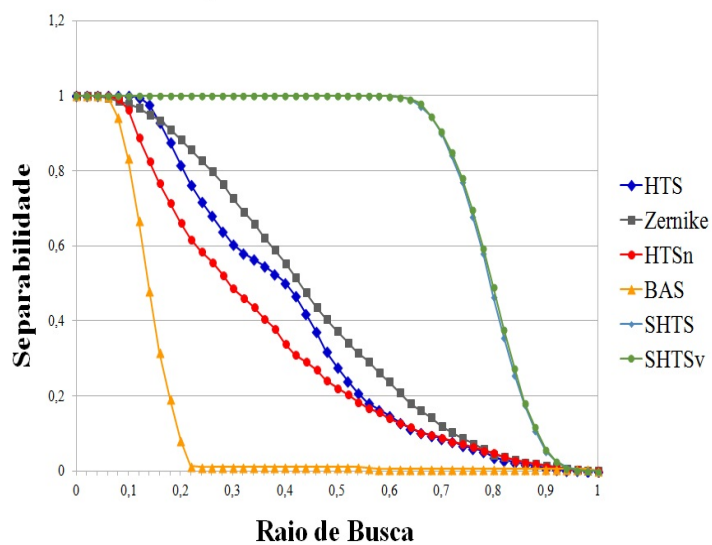


Figura 6.8: Curvas Separabilidade Multiescala do SHTS e outros descritores utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

1,42% quando comparados com suas versões monoescala.

Precisão x Revocação

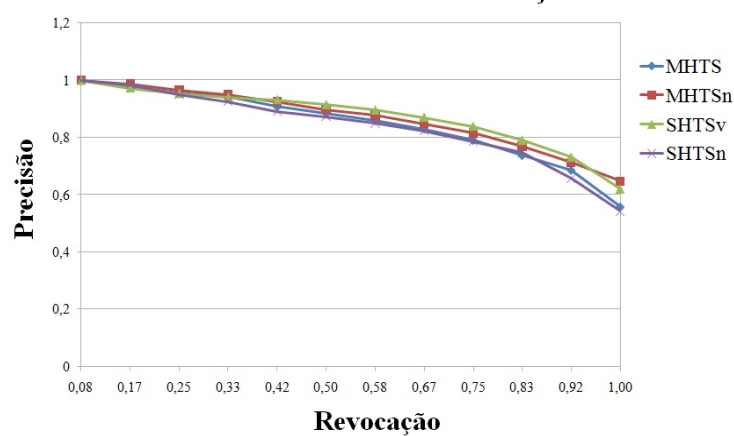


Figura 6.9: Curvas Precisão x Revocação dos descritores propostos utilizando a base Kimia-216 (SEBASTIAN et al., 2004).

Tabela 6.2: Resultados Bulls-Eye para a base Kimia-216

Método	Bulls-Eye
MHTSn	90,09 %
SHTSv	89,82 %
MHTSnv	89,70 %
BAS	88,85 %
HTS	88,54 %
SHTSnv	88,08 %
MHTS	87,73 %
MHTSv	86,84 %
SHTSn	85,96 %
MFD	85,49 %
TS	77,05 %

6.2 Segundo Experimento - Base de Dados MPEG-7 Part B

O segundo experimento foi realizado sobre a base de dados MPEG-7 Part B (JEANNIN; BOBER, 1999) composta por 1400 imagens de resoluções diferentes divididas em 70 classes com 20 imagens cada. A Figura 6.10 mostra algumas imagens dessa base de dados. Pode-se observar que a base MPEG-7 Part B é uma base (JEANNIN; BOBER, 1999) bastante desafiadora, uma vez que apresenta diversos casos de variabilidade intraclasse (Figura 6.10 (a)), similaridade interclasses (Figura 6.10 (b)), transformações de escala e rotação (Figura 6.10 (c)) e deformações e ruídos (Figura 6.10 (d)).

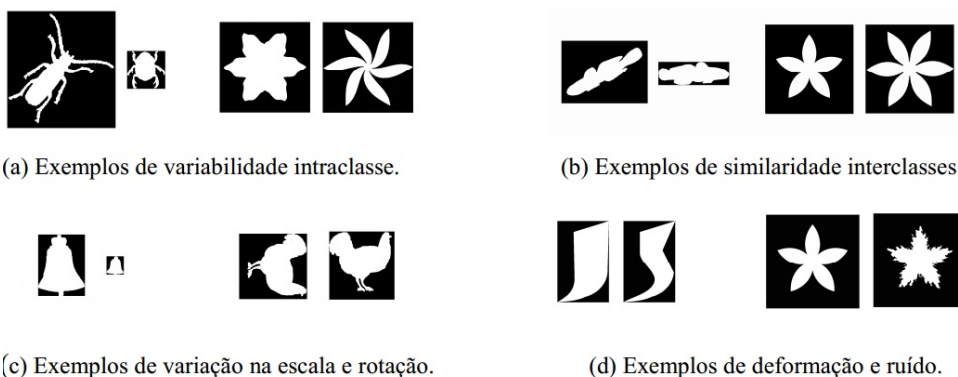


Figura 6.10: Exemplos de Imagens da Base MPEG-7 Part B (JEANNIN; BOBER, 1999).

O descritores MHTS, MHTSn, SHTS e SHTSn, bem como suas variações e os demais descritores utilizados no primeiro experimento, também foram avaliados sobre a base MPEG-7 Part B (JEANNIN; BOBER, 1999), que é uma base mais complexa que a Kimia-216 (SEBASTIAN et al., 2004).

Com o objetivo de verificar a precisão de cada escala do MHTS, optou-se por aplicar cada uma das cinco escalas individualmente na base de dados utilizada nesse experimento e, então, foram calculadas suas curvas Precisão x Revocação e seus valores de Bulls-Eye.

A Figura 6.11 mostra os resultados obtidos para cada uma das escalas do MHTS na base MPEG-7 Part B. Pode-se observar que quando aplicadas na base de dados MPEG-7 Part B, a precisão das escalas diminui gradativamente conforme a escala se torna maior. O mesmo pode ser observado para os resultados utilizando a medida Bulls-Eye, conforme pode ser observado na Tabela 6.3.

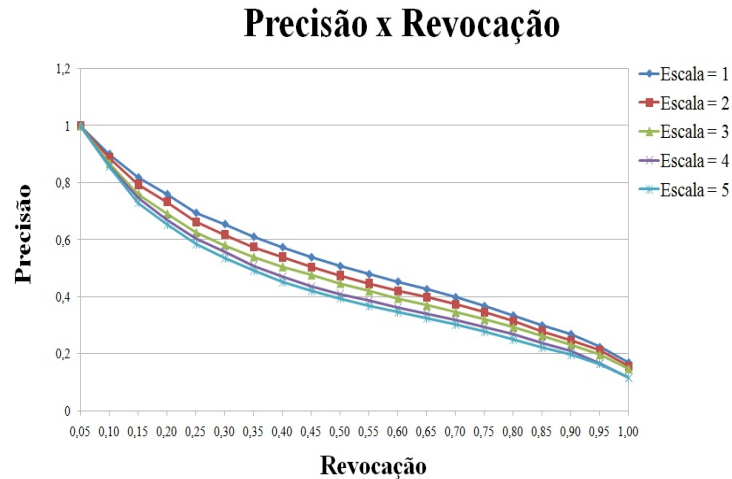


Figura 6.11: Curvas Precisão x Revocação das escalas individuais do método MHTS para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)

Tabela 6.3: Resultados Bulls-Eye para cada escala do método MHTS para a base MPEG-7 Part B

Escala	Bulls-Eye
1	58,69 %
2	56,12 %
3	53,78 %
4	51,51 %
5	50,05 %

As Figuras 6.12 e 6.13 mostram os resultados obtidos para o método MHTS e sua variação MHTSv. Pode-se observar que no critério Precisão x Revocação, o MHTS apresentou resultados próximos aos do HTSn monoescala, superando-o na maioria dos pontos, além de apresentar uma curva superior a diversos métodos, sendo inferior apenas ao BAS (ARICA; VURAL, 2003) e Momentos de Zernike (HU, 1962). Além disso, o método MHTSv apresentou resultados bem próximos ao MHTS e HTSn, sendo mais rápido que esses dois últimos. Já para o critério Separabilidade Multiescala o MHTS e o MHTSv apresentaram uma curva muito acima dos demais métodos.

As Figuras 6.14 e 6.15 mostram os resultados obtidos para o descritor MHTSn e MHTSvn. Pode-se observar que o mesmo apresentou resultados pouco superiores ao HTSn na curva de Precisão x Revocação. analisando as curvas de Separabilidade Multiescala, pode-se observar que as curvas do MHTSn e MHTSvn ficaram bem próximas de sua versão monoescala (HTSn).

As Figuras 6.16 e 6.17 mostram os resultados obtidos com os descritores SHTS e SHTSv. Tratando-se

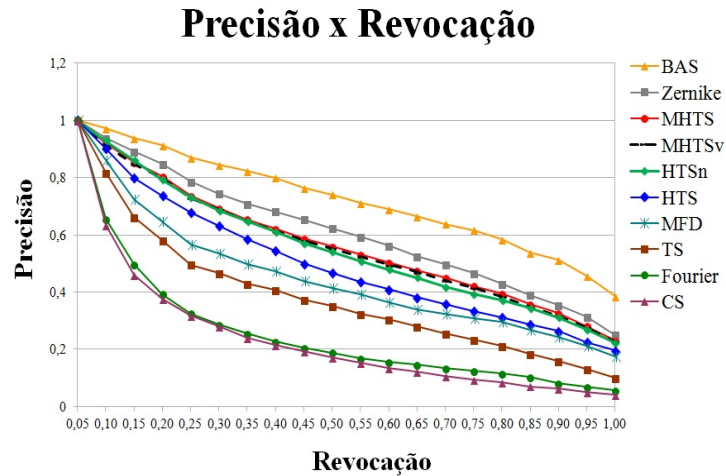


Figura 6.12: Curvas Precisão x Revocação do MHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999)

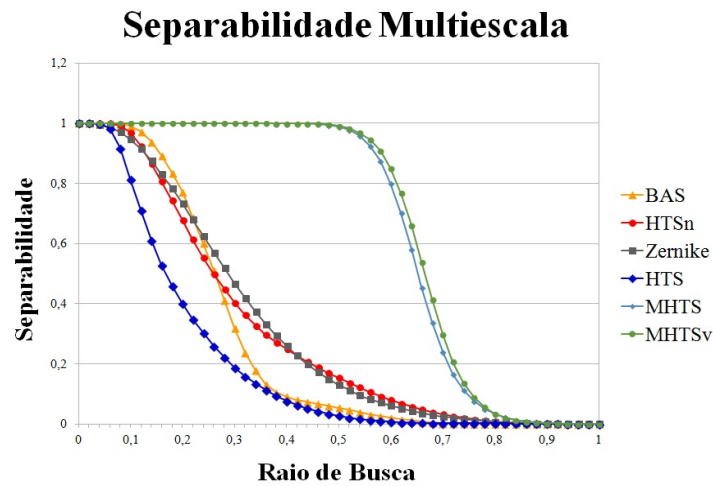


Figura 6.13: Curvas Separabilidade Multiescala do MHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999)

dos resultados de Precisão x Revocação, o descritor SHTS apresentou uma curva superior às curvas da maioria dos métodos, inclusive as dos métodos HTS e HTSn e ficou muito próximo do Momentos de Zernike (HU, 1962). Tratando-se dos resultados da Separabilidade Multiescala, o SHTS apresentou uma curva bem acima das obtidas pelos demais métodos, inclusive a do BAS (ARICA; VURAL, 2003).

As Figuras 6.18 e 6.19 mostram os resultados obtidos pelos descritores SHTSn e SHTSv. A partir dos resultados de Precisão x Revocação pode-se observar que o SHTSn e o SHTSv obtiveram resultados superiores a maioria dos métodos, inclusive a sua versão monoescala HTSn, ficando muito próximo ao descritor Momentos de Zernike (HU, 1962) e sendo inferior apenas ao BAS. De acordo com o critério Separabilidade Multiescala, os descritores SHTSn e SHTSv apresentaram os melhores resultados, superando os métodos que apresentaram as melhores curvas de Precisão x Revocação, como o BAS (ARICA; VURAL, 2003), Momentos de Zernike, sua versão monoescala HTSn e a versão espaço-escala SHTS.

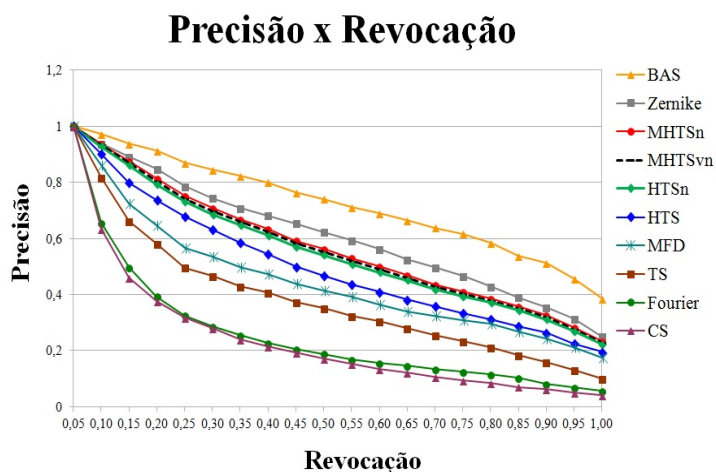


Figura 6.14: Curvas Precisão x Revocação do método MHTSn, MHTSvn e outros métodos utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

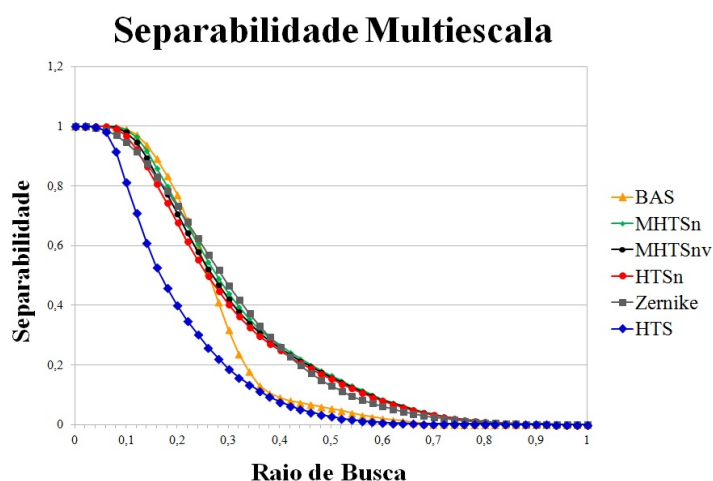


Figura 6.15: Curvas Separabilidade Multiescala do MHTSn, MHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Por fim, a Figura 6.20 mostra uma comparação entre as curvas de Precisão x Revocação das abordagens propostas. Pode-se observar que as curvas mais altas são dos métodos SHTS e SHTSn. A Tabela 6.4 apresenta uma comparação dos resultados Bulls-Eye para os métodos propostos com os demais métodos apresentados. Pode-se observar a partir da Tabela 6.4 que os métodos SHTS e SHTSv obtiveram os melhores resultados entre os métodos propostos, ficando atrás apenas do BAS e Momentos de Zernike (HU, 1962), validando os resultados das curvas Precisão x Revocação. A versão multiescala do HTS apresentou uma melhoria de precisão de aproximadamente 11,52%, enquanto a versão espaço-escala apresentou uma melhoria de aproximadamente 19% quando comparados com suas versões monoescala. Para o HTSn a versão multiescala apresentou uma melhoria de precisão de 2,7% e 8,1% para a versão espaço-escala quando comparados ao HTSn monoescala.

Precisão x Revocação

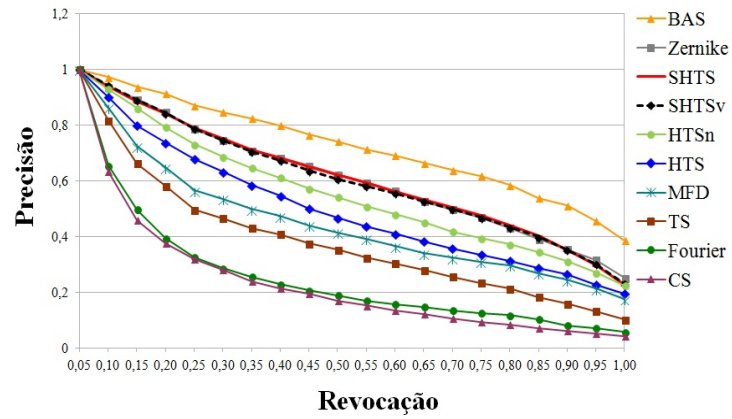


Figura 6.16: Curvas Precisão x Revocação do SHTS, SHTSv e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Separabilidade Multiescala

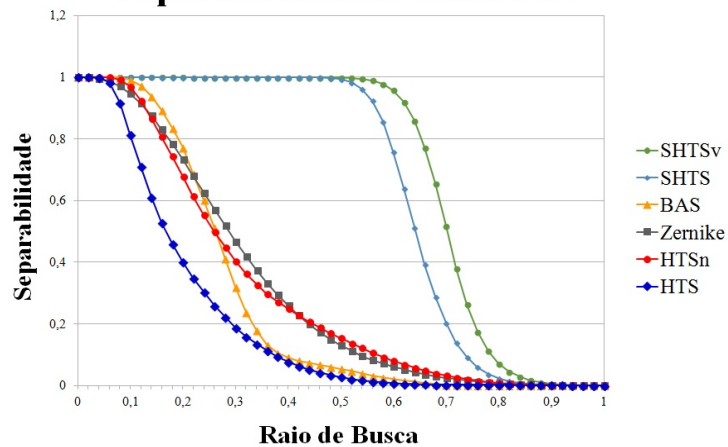


Figura 6.17: Curvas Separabilidade Multiescala do SHTS e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Precisão x Revocação

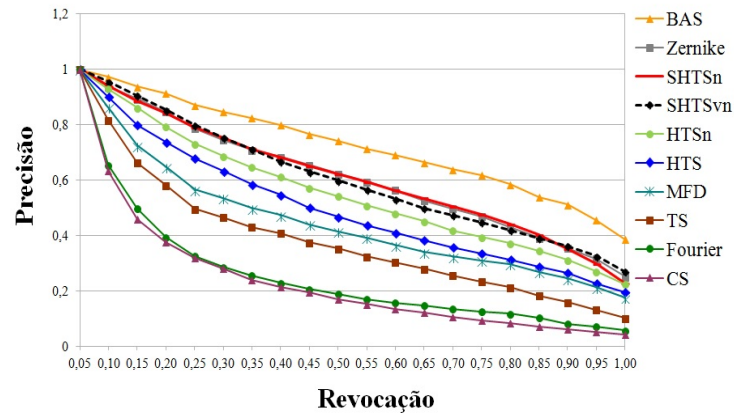


Figura 6.18: Curvas Precisão x Revocação do SHTSn, SHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Separabilidade Multiescala

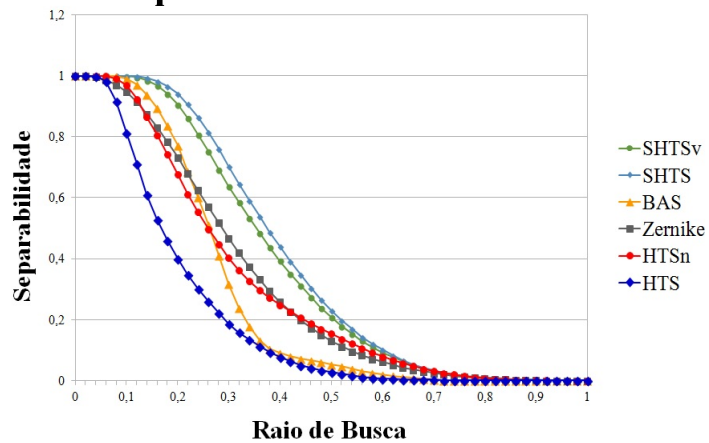


Figura 6.19: Curvas Separabilidade Multiescala do SHTSn, SHTSvn e outros descritores utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

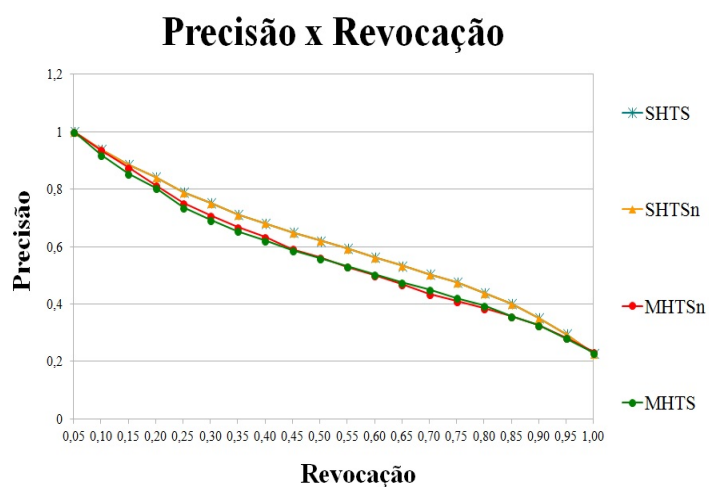


Figura 6.20: Curvas Precisão x Revocação dos descritores propostos utilizando a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Tabela 6.4: Resultados Bulls-Eye para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)

Método	Bulls-Eye
BAS	76,65 %
Mom. Zernike	70,22 %
SHTS	66,66 %
SHTSv	66,48 %
SHTSn	64,79 %
SHTSnv	64,66 %
MHTS	62,46 %
MHTSv	62,00 %
MHTSn	61,59 %
MHTSnv	60,83 %
HTSn	59,97 %
HTS	56,01 %
MFD	52,57 %
TS	45,90 %

6.3 Terceiro Experimento - Base de Dados MNIST

O terceiro experimento foi realizado sobre a base de dados MNIST (*Mixed National Institute of Standards and Technology*) (LECUN et al., 1998), que é uma base de imagens públicas composta por 10.000 imagens dos 10 dígitos escritos à mão que variam de 0 a 9. Todas as imagens possuem a mesma resolução de 28x28 pixels.

Embora a maioria das imagens desta base seja composta por dígitos bem escritos que são relativamente fáceis de serem identificados (Figura 6.21 (a)), ela apresenta algumas imagens que são difíceis de serem classificadas até mesmo pelos seres humanos (Figura 6.21 (b)) devido à caligrafia precária de alguns escritores que contribuíram com a criação da base.

Devido à irregularidade da base, algumas classes de dígitos possuem mais imagens que as demais, dessa forma optou-se por utilizar apenas 8000 imagens do conjunto total de 10000 imagens. Além disso, na etapa inicial de extração de bordas algumas imagens apresentaram uma borda externa e outra interna. Portanto, efetuou-se um pré-processamento na base para a remoção de todas as bordas internas.

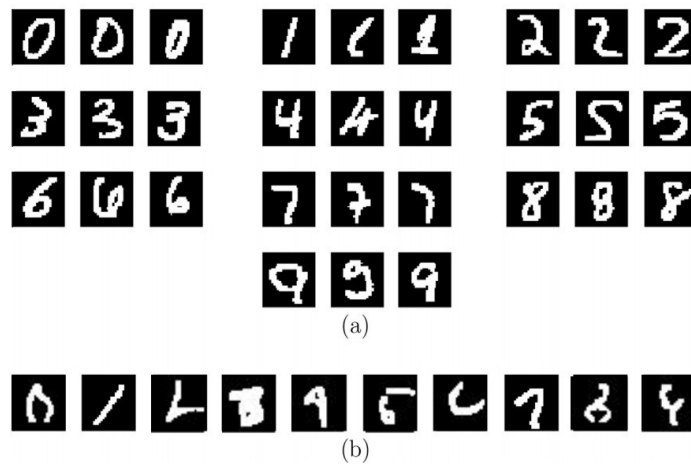


Figura 6.21: Exemplos de dígitos da base MNIST (LECUN et al., 1998): (a) Dígitos com boa grafia, mais simples de serem identificados (de 0 a 9, 3 exemplos cada) e (b) Dígitos com grafia precária, identificação complexa até mesmo para seres humanos. Extraído de (SOUZA; MARANA, 2014)

Para avaliar o desempenho do descritor MHTS efetuou-se a comparação com alguns descritores como BAS (Beam Angle Statistics), Zernike Moments (HU, 1962), Multiscale Fractal Dimension (MFD), Tensor Scale (TS), Fourier e Contour Saliency (CS). A Figura 6.22 apresenta as curvas Precisão x Revocação de cada método. Pode-se observar que o melhor resultado foi obtido pelo BAS, seguido pelo método proposto MHTS. O terceiro melhor foi o HTSn seguido pelo Fourier, HTS e outro descritor proposto SHTS. A Figura 6.23 apresenta as curvas Separabilidade Multiescala dos métodos. Apesar da melhoria de acordo com a Separabilidade Multiescala, o MHTS e SHTS apresentaram resultados inferiores ao BAS segundo o critério Bulls-Eye como pode ser observado na Tabela 6.5

A melhoria obtida pelo MHTS em relação ao HTS foi de 5,1% de acordo com a medida de similaridade

Bulls-Eye. Se comparado ao BAS, o MHTS apresentou desempenho 11,43% inferior. Isso prova que a versão multiescala do HTS apresenta melhores resultados em bases de dados grandes e complexas que sua versão monoescala e pode ser mais eficiente que muitos descritores. Já o SHTS não apresentou resultados superiores ao HTS.

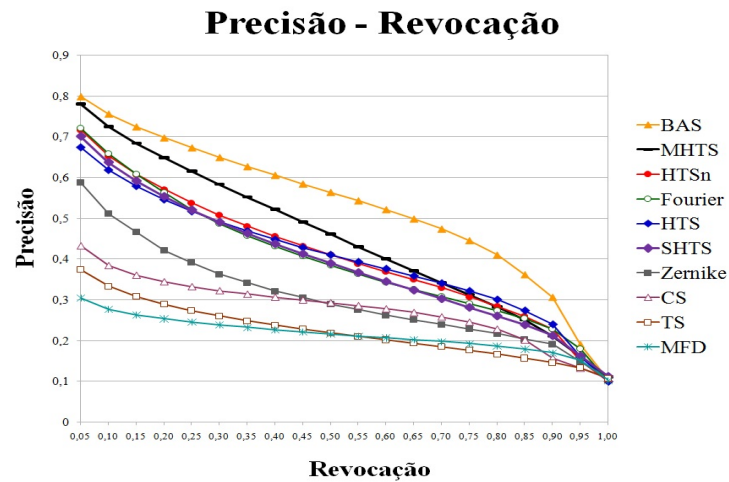


Figura 6.22: Curvas Precisão x Revocação do MHTS e demais descritores utilizando a base MNIST (LECUN et al., 1998).

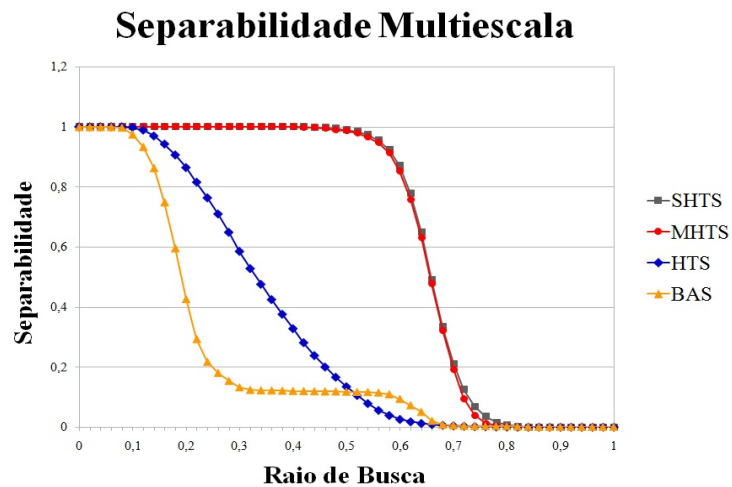


Figura 6.23: Curvas Separabilidade Multiescala do MHTS, SHTS e demais descritores utilizando a base MNIST (LECUN et al., 1998).

Tabela 6.5: Resultados Bulls-Eye para a base MNIST (LECUN et al., 1998)

Método	Bulls-Eye
BAS	67,12 %
MHTS	60,23 %
HTS	57,32 %
SHTS	55,35 %

6.4 Quarto Experimento - Base de Dados de Sinus Frontais

Outro experimento foi realizado com base de tomografias computadorizadas FSCT, criada pelo grupo de pesquisa Recogna-UNESP, a fim de se testar a eficácia dos métodos em uma aplicação na área de Biometria.

Para a identificação de pessoas pós-morte, a maior parte das características biométricas (impressões digitais, íris, veias da mão, face, retina) não podem ser utilizadas devido à rápida decomposição dos tecidos moles do corpo humano. Para este fim específico, partes do esqueleto humano, tais como dentes, tórax, vértebras, ombro e cavidades ósseas são mais adequados.

Alguns estudos têm sido feitos a fim de se identificar pessoas a partir de medidas dos sinus frontais de imagens de raios-X ou tomografia computadorizada. Os sinus frontais são compostos por pares de cavidades loculadas de forma irregular, localizadas no osso frontal, que se comunicam com a fossa nasal.

Os sinus frontais são únicos para cada indivíduo e são controlados por fatores genéticos e ambientais. Variações nos sinus frontais têm sido relatados na literatura mesmo para gêmeos monozigóticos (SAUNDERS, 1974).

A Figura 6.24 mostra uma imagem de raio-x (à esquerda) e uma imagem de tomografia computadorizada (TC) (à direita) de dois indivíduos distintos, com seus sinus frontais demarcados por retângulos. Pode-se observar que a imagem de TC permite melhor análise das estruturas do sinus frontal quando comparada com a imagem de raios-x.

A base FSCT é composta por 310 imagens obtidas a partir de dois exames de tomografia computadorizada de 31 indivíduos diferentes. Cada exame é composto por uma sequência de 5 imagens. A Figura 6.25 apresenta uma sequência de 5 imagens para 3 exames diferentes.

Nesse experimento, após a segmentação da região do sinus frontais de cada imagem de tomografia computadorizada, que foi efetuada manualmente (*Ground Truth*) o MHTS, bem como outros métodos, foram aplicados na base de dados.

Cada método foi testado individualmente e avaliado em termos de sua curva ROC (*Receiver Operating Characteristic*), de sua taxa EER (*Equal Error Rate*).

A curva ROC é muito utilizada na avaliação de sistemas biométricos uma vez que ela permite avaliar o sistema de acordo com a Taxa de Falsa Aceitação (*FAR - False Acceptance Rate*), ou seja, um indivíduo

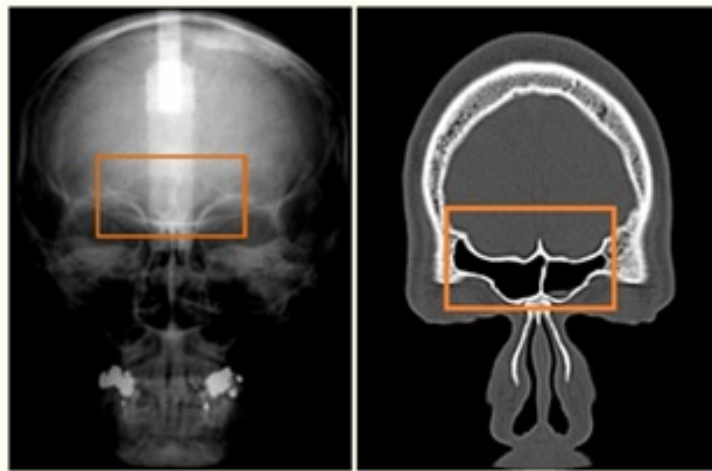


Figura 6.24: (A) imagem de raios -X ; (B) imagem de Tomografia Computadorizada de um indivíduo. Os seios frontais são delimitados por retângulos.

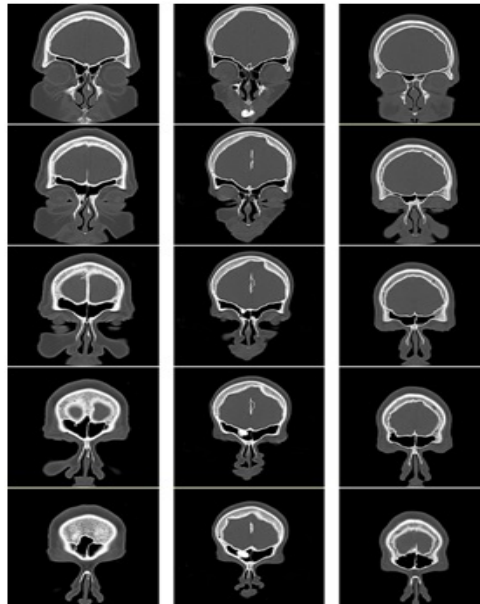


Figura 6.25: (a)-(c) Sequência de imagens de 3 exames de tomografia computadorizada da base FSCT.

impostor sendo reconhecido como genuíno, e a Taxa de Aceitação Verdadeira (*TAR - True Acceptance Rate*), em que um indivíduo genuíno é classificado corretamente como genuíno.

A Figura 6.26 mostra as curvas ROC obtidas para os descritores testados. Quanto mais alta a curva, melhor é o desempenho do método. Pode-se observar que apesar do MHTS não ser o melhor método, pois não superou o BAS, ele superou os demais métodos, como o HTS monoescala e o MFD.

Além da curva ROC, utilizou-se o EER para avaliar o desempenho de cada descritor. O valor do EER corresponde ao ponto em que a Taxa de Falsa Aceitação e a Taxa de Aceitação Verdadeira são iguais. Quanto menor for o valor do EER melhor é o desempenho do descritor. A Tabela 6.6 apresenta os valores de EER obtidos para cada método. Percebe-se que o descritor BAS obteve os melhores resultados para

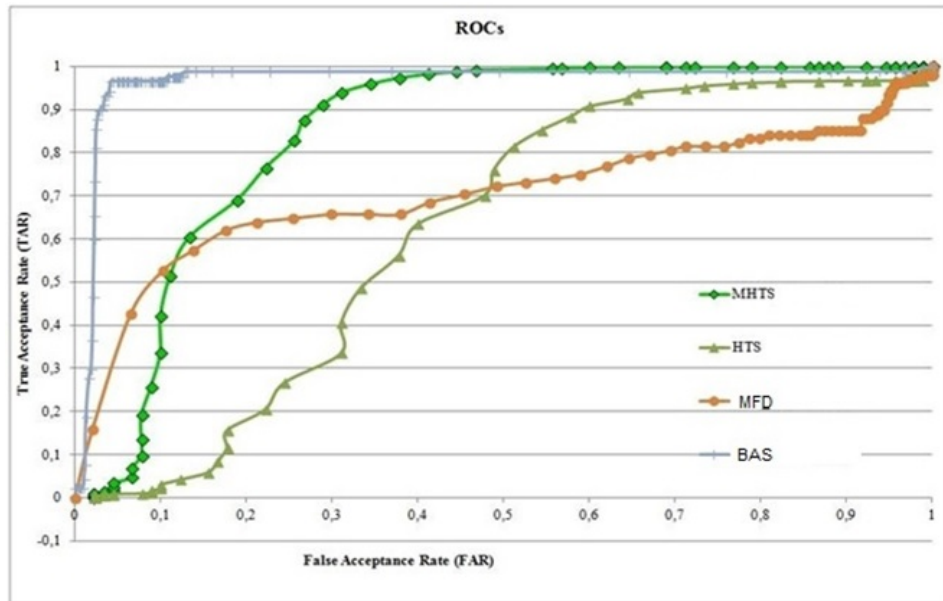


Figura 6.26: Curvas ROC obtidas a partir dos descritores de formas para a base FSTC.

ambas as métricas, seguido pelo MHTS, que superou os descritores MFD e HTS.

Tabela 6.6: Valores de ERR para os descritores testados na base de sinus frontais.

Método	EER %
BAS	4,1
MHTS	20,7
MFD	33,5
HTS	40,1

Os resultados desse experimento comprovam as melhorias do MHTS em relação ao HTS, uma vez que o primeiro teve desempenho inferior apenas ao BAS, obteve um EER 19,4% melhor que sua versão monoescala HTS para a base de sinus frontais.

6.5 Quinto Experimento - Recuperação de Imagens

O quinto experimento foi realizado sobre a base de dados MPEG-7 Part B (JEANNIN; BOBER, 1999) e MNIST (LECUN et al., 1998) e objetivou avaliar o desempenho de descritores de formas, focando as aplicações de recuperação de imagens baseada em conteúdo, utilizando-se a forma.

A recuperação de imagens a partir da forma dos objetos é uma tarefa muito importante na visão computacional. Dado uma imagem de busca, a recuperação de imagens consiste em retornar todas as imagens da base de dados que são similares à imagem de busca (BAI et al., 2012);

Porém, a escolha da função de distância e do descritor a serem utilizados para determinar se uma imagem é ou não similar à imagem de busca é uma tarefa desafiadora, como pode-se observar na Figura

6.27, em que a forma do cavalo (a) é mais parecida com a forma do cachorro (b) do que com a forma do cavalo (c) (BAI et al., 2012).

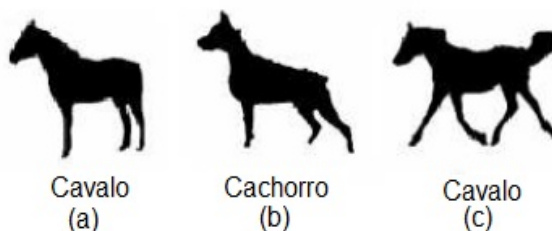


Figura 6.27: Cavalo em (a) pode ser mais parecido ao cachorro em (b) que ao cavalo em (c). Extraído de (BAI et al., 2012).

Atualmente, pode-se encontrar dois tipos de técnicas de recuperação de imagens, as baseadas em palavras chave e as baseadas em conteúdo. Embora a recuperação de imagens baseada em palavras chave seja importante e simples de se utilizar, ela pode muitas vezes ser subjetiva, uma vez que palavras diferentes podem ser utilizadas para indexar a mesma imagem. Portanto, sua precisão na recuperação de imagens pode ser limitada. Dessa forma, a recuperação de imagens baseada em características da imagem, como cor, textura e forma, é muito relevante. De todas as características que podem ser analisadas, o forma apresenta ótimos resultados na recuperação de imagens (LIU et al., 2012).

Devido ao crescimento das bases de dados, sejam elas bases públicas na Internet ou bases de dados biométricos, a tarefa de recuperação de imagens tem se tornado crucial para um grande número de aplicações (LIU et al., 2012), como em (BADAWY; KAMEL, 2002), em que o descritor *Zernike moment edge gradient technique* (ZMEG), foi utilizado na recuperação de imagens de marcas comerciais. Em (MOUINE et al., 2012) a recuperação de imagens utilizando o método *Advanced Shape Context* foi aplicada em uma base de folhas para auxiliar na identificação de espécies de plantas.

Neste experimento, utilizou-se as versões dos métodos propostos com a multiescala aplicada nos vetores de características (MHTSv, MHTS_{nv}, SHTSv). Os métodos foram avaliados sobre duas bases, a MPEG-7 Part B e a MNIST.

Para cada imagem pertencente a uma das base iniciou-se a tarefa de recuperação a partir das escalas inferiores, em que o tamanho do espaço de Hough e do vetor de características é menor. A partir da comparação das imagens na escala mais inferior, apenas as imagens que obtiveram na comparação uma distância inferior a um limiar eram levadas para a próxima escala de avaliação, sendo eliminadas as imagens que apresentaram grande diferença da imagem de busca. Essa eliminação continua ocorrendo até que a última escala seja atingida e o resultado final seja retornado.

Como pode ser visto na Figura 6.11, a perda de precisão de cada uma das escalas individuais do MHTS é pequena, portanto, para a base MPEG-7 Part B, optou-se por utilizar cada uma das 5 escalas do MHTS, sendo a quinta escala a menor delas, com um vetor de características de tamanho igual a 10. O tamanho do vetor foi incrementado a cada escala em 10 unidades, dessa forma, o vetor de características da quarta escala era composto por 20 valores, o da terceira por 30 valores e assim por diante até a

primeira escala composta por 50 valores.

Cada uma das 1400 imagens da base MPEG-7 Part B e 8000 imagens da base MNIST foram utilizadas na tarefa de recuperação. O tempo de recuperação foi calculado para cada uma das imagens presentes nas bases de dados e a média e desvio padrão desses tempos foram calculados para cada uma das bases, além da avaliação dos resultados finais por meio das curvas de Precisão x Revocação. O tempo de processamento foi calculado utilizando-se um notebook com as seguintes especificações: Ubuntu versão 14.04 LTS 32 bits, 2 GB de memória, 260 GB de HD e processador Intel i5 3337U de 1,80 Ghz.

A Tabela 6.7 apresenta o tempo médio em milissegundos e seu desvio padrão para os testes efetuados com a base MPEG-7 Part B. Nota-se que o MHTS apresentou os melhores resultados de tempo médio, seguido pelo SHTS e por fim sua versão monoescala HTS.

A Figura 6.28 mostra as curvas de Precisão x Revocação para a base MPEG-7 Part B. Nota-se que apesar da pequena queda de precisão dos métodos MHTS e SHTS, eles ainda mantêm-se superiores a sua versão monoescala. A queda de precisão já era esperada devido ao fato de as escalas inferiores determinarem as imagens que são utilizadas pelas demais escalas superiores e como a precisão das escalas inferiores é pior, algumas imagens verdadeiras acabam sendo eliminadas por essas escalas.

Tabela 6.7: Tempo médio em milissegundos e desvio padrão para a recuperação de imagens na base MPEG-7 Part B.

Método	Média	D. Padrão
MHTS	48,88 ms	9,19 ms
SHTS	52,77 ms	8,48 ms
HTS	92,25 ms	4,81 ms
BAS	892,04 ms	37,01 ms

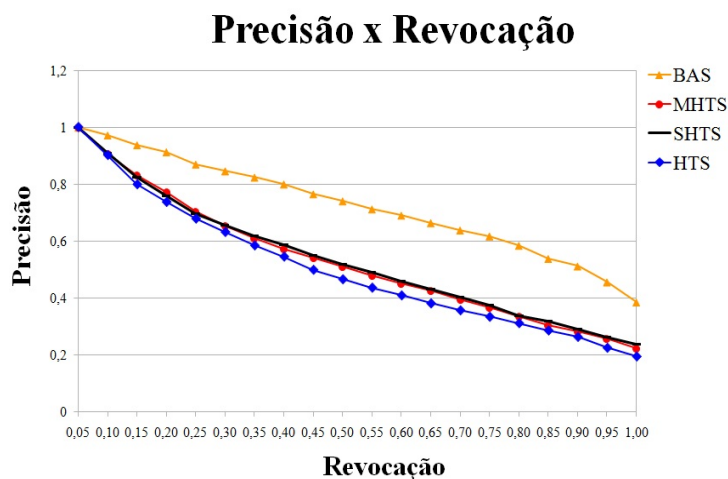


Figura 6.28: Curvas Precisão x Revocação para a recuperação de imagens da base MPEG-7 Part B.

A Tabela 6.8 apresenta o tempo médio em milissegundos, seu desvio padrão e o tempo total para os testes efetuados com a base MNIST. Nota-se que o MHTS apresentou os melhores resultados. Como a

base MNIST apresenta mais imagens, torna-se evidente a importância da velocidade na recuperação de imagens, uma vez que o MHTS levou 984 mil milissegundos (aproximadamente 16 minutos) a menos para recuperar todas as imagens da base, tornando-o muito relevante para aplicações voltadas para grandes bases de dados.

Tabela 6.8: Tempo médio, desvio padrão e tempo total para a recuperação de imagens na base MNIST.

Método	Média	D. Padrão	Tempo Total
MHTS	326,18 ms	36,92 ms	2.610.000 ms
SHTS	380,17 ms	20,88 ms	3.041.000 ms
HTS	449,05 ms	39,97 ms	3.594.000 ms

6.6 Sexto Experimento - SBAS e SMFD

Uma vez que a abordagem espaço-escala ocorre na etapa de pré-processamento, ou seja, na imagem de entrada, ela pode ser aplicada também em métodos como o BAS (ARICA; VURAL, 2003) e o MFD (BACKES; BRUNO, 2006), resultando nos métodos SBAS e SMFD, respectivamente.

A Figura 6.29 apresenta os resultados de Precisão x Revocação para a base Kimia-216, a Figura 6.30 apresenta as curvas Separabilidade multiescala, e a Tabela 6.9 apresenta os valores de Bulls-Eye. Pode-se observar que o método SBAS apresentou os melhores resultados de Precisão x Revocação, Bulls-Eye e Separabilidade multiescala. O método SMFD apresentou resultados superiores à sua versão monoescala.

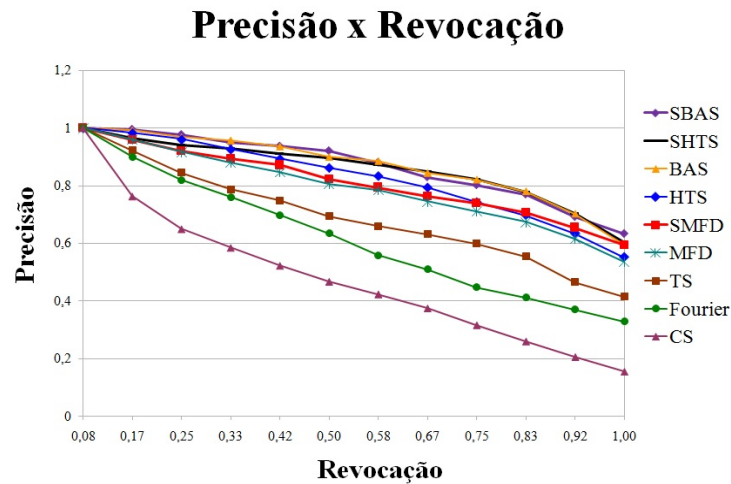


Figura 6.29: Curvas Precisão x Revocação para a base Kimia-216.

A Figura 6.31 apresenta as curvas Precisão x Revocação para a base MPEG-7 Part B obtidas pelos métodos SBAS e SMFD e a Figura 6.32 apresenta as curvas Separabilidade Multiescala para os métodos que apresentaram os melhores resultados. A Tabela 6.10 apresenta os valores de Bulls-Eye para a base MPEG-7 Part B. Pode-se observar que ambos os métodos apresentaram um ganho de precisão em suas

Multiscale Separability

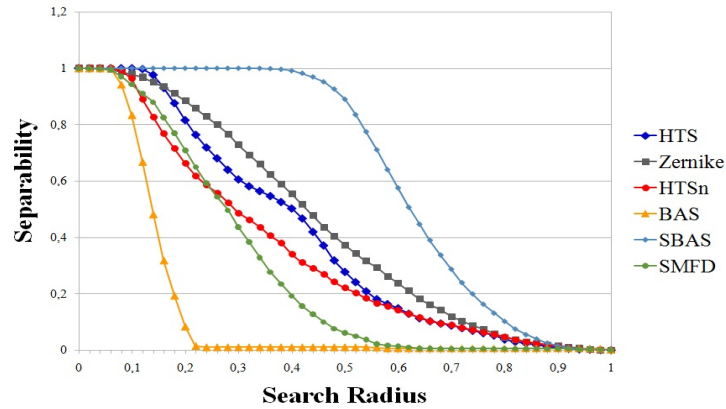


Figura 6.30: Curvas Separabilidade Multiescala para a base Kimia-216.

Tabela 6.9: Resultados Bulls-Eye para a base Kimia-216.

Método	Bulls-Eye
SHTS	89,82 %
SBAS	89,43 %
BAS	88,85 %
SMFD	85,73 %
MFD	85,49 %

curvas Precisão x Revocação, sendo que o método SMFD superou sua versão monoescala. De acordo com a medida Bulls-Eye o método SBAS apresentou uma melhoria de 4,81% quando comparado à sua versão monoescala, já o SMFD apresentou melhoria de aproximadamente 9,82% quando comparada à sua versão monoescala. De acordo com as curvas Separabilidade multiescala, o método SBAS apresentou grande melhoria, sendo inferior apenas ao SHTS.

Precisão x Revocação

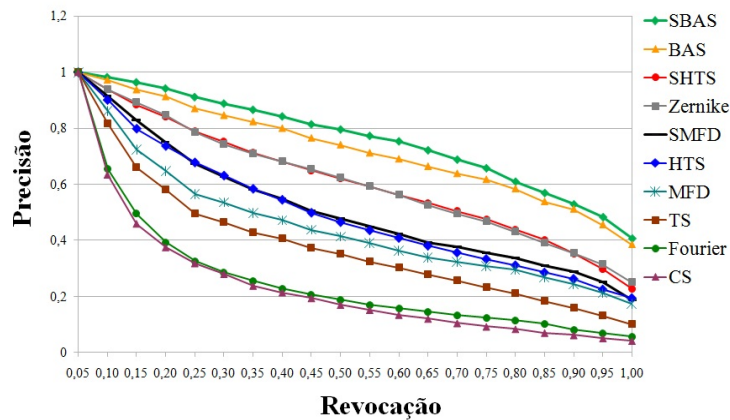


Figura 6.31: Curvas Precisão x Revocação para a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Separabilidade Multiescala

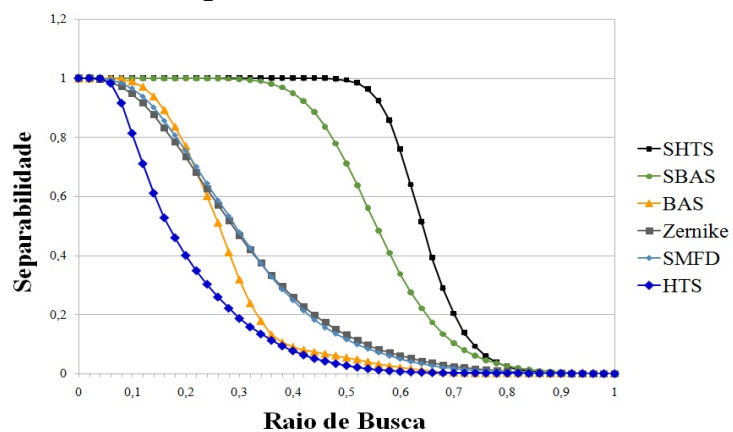


Figura 6.32: Curvas Separabilidade Multiescala para a base MPEG-7 Part B (JEANNIN; BOBER, 1999).

Tabela 6.10: Resultados Bulls-Eye para a base MPEG-7 Part B (JEANNIN; BOBER, 1999)

Método	Bulls-Eye
SBAS	80,34 %
BAS	76,65 %
SHTS	66,66 %
SMFD	57,75 %
MFD	52,57 %

Capítulo 7

Conclusões e Trabalhos Futuros

Neste trabalho foram propostos os seguintes descritores multiescala: MHTS, MHTSn e suas abordagens hierárquicas MHTSv e MHTSnv, e os seguintes descritores espaço-escala: SHTs, SHTSn e suas abordagens hierárquicas SHTSv e SHTSnv. Esses descritores apresentaram ótimo desempenho sobre as bases KIMIA-216 (SEBASTIAN et al., 2004), MPEG-7 Part B (JEANNIN; BOBER, 1999) e MNIST (LECUN et al., 1998), o que prova que esses novos descritores são muito eficazes e eficientes, sendo mais rápidos e precisos que o HTS na tarefa de recuperação de imagens, com resultados de precisão muito próximos ao BAS (ARICA; VURAL, 2003), porém muito mais rápido que ele e mantendo a complexidade linear do HTS monoescala.

Com relação ao método MHTS, ele apresentou resultados bem próximos ao BAS (ARICA; VURAL, 2003) e em alguns experimentos chegou a apresentar os melhores resultados em termos de acurácia, sendo o melhor dentre os demais descritores avaliados. Apesar do método HTS (SOUZA; MARANA, 2013) já ser mais rápido que o BAS (ARICA; VURAL, 2003) a abordagem hierárquica MHTSv foi mais rápida que ambos na recuperação de imagens, reduzindo o tempo de busca na base MNIST em 27% quando comparado ao HTS, provando sua eficiência em relação ao tempo e precisão na busca de imagens em grandes bases de dados.

O SHTS, outro descritor proposto nesse trabalho, utiliza uma abordagem espaço-escala nas imagens de entrada, que resulta em uma representação multiescala da imagem de entrada. A partir dos resultados obtidos pode-se concluir que esse novo descritor é mais preciso e veloz que o HTS, sendo que apresentou ótimos resultados quando comparado ao BAS nos experimentos realizados com a base Kimia-216 (SEBASTIAN et al., 2004) de acordo com o critério Bulls-Eye, sendo inferior apenas ao outro método proposto nesse trabalho, o MHTS. Além disso, uma vez que a abordagem espaço-escala ocorre na etapa de pré-processamento de imagens, outros métodos puderam ser empregados a partir dessa estratégia, resultando em dois novos métodos chamados SBAS e SMFD, sendo que ambos apresentaram resultados superiores às suas versões monoescala, melhorando sua precisão segundo os critérios Precisão x Revocação, Bulls-Eye e Separabilidade multiescala.

No experimento de recuperação de imagens as vantagens do uso de uma abordagem multiescala ficaram evidentes, uma vez que os métodos propostos nesse trabalho foram mais rápidos que suas versões monoescala e muito mais rápidos que outros métodos que apresentam uma melhor precisão, como o BAS (ARICA; VURAL, 2003). Em algumas aplicações que envolvem grandes bases de dados, como a busca de imagens na Internet e a busca de imagens de logos de marcas comerciais, o tempo de processamento é um fator muito importante e deve ser levado em conta no momento da escolha do método ideal.

Os resultados obtidos a partir do SBAS e SMFD provaram que a abordagem espaço-escala pode ser estendida para outros métodos, sendo capaz de melhorar sua eficiência de acordo com diversos critérios, como a Precisão x Revocação, Separabilidade multiescala e Bulls-Eye.

Os métodos MHTS e SHTS foram analisados sobre diversas bases de dados e de diferentes formas, apresentando bons resultados quando comparados às suas versões monoescala e, em alguns casos, apresentaram os melhores resultados quando comparados à todos os métodos utilizados nessa dissertação. A abordagem multiescala dos métodos de análise de imagens apresentou melhorias em todos os estudos efetuados com outros métodos. É importante ressaltar que as versões multiescala e espaço-escala do HTS preservam sua complexidade linear.

7.1 Trabalhos Futuros

Como trabalhos futuros pretende-se explorar outras possibilidades de representação multiescala do HTS e suas variações, como por exemplo, a representação utilizando redes complexas a partir dos picos do espaço de Hough.

Neste trabalho realiza-se a fusão dos descritores obtidos nas várias escalas gerando um único descritor. Esta estratégia é conhecida como fusão no nível de características. Entretanto, existem outras estratégias de fusão que poderiam ser exploradas em trabalhos futuros.

Como os métodos apresentados no Capítulo 3 não foram utilizados nos experimentos, um estudo que compare o MHTS e SHTS aos demais métodos multiescala apresentados no Capítulo 3 seria muito relevante.

Além disso, uma vez que o propósito da transformada de Hough (HOUGH, P. V. C., 1962) é identificar retas em uma dada imagem, uma nova abordagem poderia explorar a diferença de ângulos entre as retas identificadas em uma imagem, e, a partir disso, criar-se um descritor. A quantidade de retas detectadas pode variar de acordo com um limiar r que determina se uma reta faz ou não parte do conjunto. Variando-se esse limiar r , pode-se criar uma nova abordagem multiescala.

7.2 Publicações Realizadas

RAMOS, L. A.; SOUZA, G. B.; MARANA A. N.; Shape Analysis Using Multiscale Hough Transform Statistics. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Vol.

9423, p. 452-459, 2015.

RAMOS, L. A.; MARANA, A. N. Análise de Formas com HTS Multiescala. WORKSHOP DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DA UNESP. Nº 5, Bauru, 2015.

Bibliografia

- ABBASI, S.; MOKHTARIAN, F.; KITTLER, J. Curvature scale space image in shape similarity retrieval. *Multimedia Syst.*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 7, n. 6, p. 467–476, Novembro 1999. ISSN 0942-4962. Citado 4 vezes nas páginas ix, 20, 22 e 23.
- ADELSON, E. H.; ANDERSON, C. H.; BERGEN, J. R.; BURT, P. J.; OGDEN, J. M. *Pyramid Methods in Image Processing*. 1984. Citado 6 vezes nas páginas viii, 6, 7, 8, 10 e 11.
- ARICA, N.; VURAL, F. T. Y. Bas: A perceptual shape descriptor based on the beam angle statistics. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 24, n. 9-10, p. 1627–1639, Junho 2003. ISSN 0167-8655. Citado 11 vezes nas páginas 1, 3, 32, 41, 42, 43, 47, 48, 60, 63 e 64.
- BACKES, A. R.; BRUNO, O. M. *Implementação e Comparação de Métodos de Estimativa da Dimensão Fractal e sua Aplicação à Análise e Processamento de Imagens*. Dissertação (Mestrado) — Universidade de São Paulo (USP), São Carlos - SP, 2006. Citado 5 vezes nas páginas 1, 2, 3, 32 e 60.
- BADAWY, O. E.; KAMEL, M. Shape-based image retrieval applied to trademark images. *International Journal of Image and Graphics*, v. 2, n. 3, p. 375–393, 2002. Citado na página 58.
- BAI, X.; WANG, B.; YAO, C.; LIU, W.; TU, Z. Co-transduction for shape retrieval. *IEEE Transactions on Image Processing*, v. 21, n. 5, p. 2747–2757, Maio 2012. Citado 3 vezes nas páginas xii, 57 e 58.
- BELONGIE, S.; MALIK, J. Matching with shape contexts. In: *Content-based Access of Image and Video Libraries (CBAIVL)*. Hilton Head, SC: [s.n.], 2000. p. 20–26. Citado 3 vezes nas páginas 1, 3 e 32.
- BRUNO, O. M.; BACKES, A. R. Fractal and multi-scale fractal dimension analysis: A comparative study of bouligand-minkowski method. *INFOCOMP Journal of Computer Science*, v. 7, n. 2, p. 74–83, 2008. Citado 3 vezes nas páginas viii, 17 e 18.
- BRUNO, O. M.; BRAZIL, L. P.; ANDRADE, L. A. Método da determinação da dimensão fractal por fourier e análise multiescala para reconhecimento de padrões. In: *Workshop de Visão Computacional*. [S.l.: s.n.], 2007. p. 233–238. Citado 2 vezes nas páginas 16 e 17.
- BURT, P. J. Fast filter transforms for image processing. *Computer Graphics and Image Processin*, v. 16, n. 1, p. 20–51, 1981. Citado 2 vezes nas páginas 6 e 7.
- COSTA, L. F.; JÚNIOR, R. M. C. *Shape Analysis and Classification: Theory and Practice*. 1st. ed. Boca Raton, FL, USA: CRC Press, Inc., 2000. ISBN 0849334934. Citado na página 1.
- CROWLEY, J. L. *A Representation for Visual Information*. Tese (Doutorado) — CarnegieMellon University, Pittsburgh, Pennsylvania, 1981. Citado 2 vezes nas páginas 6 e 7.
- DIREKOGLU, C.; NIXO, M. S. Image-based multiscale shape description using gaussian filter. In: *Sixth Indian Conference on Computer Vision, Graphics & Image Processing, ICVGIP 2008, Bhubaneswar, India, 16-19 December 2008*. [S.l.: s.n.], 2008. p. 673–678. Citado 6 vezes nas páginas viii, 8, 9, 14, 15 e 16.
- DUDA, R. O.; HART, P. E. *Use of the Hough Transformation to Detect Lines and Curves in Pictures*. [S.l.], 1971. Citado na página 28.

- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. *J. Classif.*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 24, n. 2, p. 305–307, Setembro 2007. ISSN 0176-4268. Citado na página 1.
- FALCAO, A. X.; STOLFI, J.; LOTUFO, R. A. The image foresting transform: theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 1, p. 19–29, 2004. Citado na página 23.
- FLORINDO, J. B.; BACKES, A. R.; CASTRO, M.; M, B. O. A comparative study on multiscale fractal dimension descriptors. *Pattern Recognition Letters*, v. 33, n. 6, p. 798–806, 2012. Citado na página 20.
- GONZALEZ, R. C.; WOODS, R. *Digital Image Processing*. 3rd. ed. [S.l.]: Prentice Hall, 2007. ISBN 013168728X. Citado na página 12.
- HOUGH, P. V. C. *A Method and Means for Recognizing Complex Patterns*. 1962. US Patent: 3,069,654. Citado 4 vezes nas páginas 2, 27, 28 e 64.
- HU, M. K. Visual pattern recognition by moment invariants. In: . [S.l.: s.n.], 1962. v. 8, n. 2, p. 179–187. Citado 9 vezes nas páginas 1, 16, 41, 42, 43, 47, 48, 49 e 53.
- JEANNIN, S.; BOBER, M. *"Description of Core Experiments for MPEG-7 Motion/Shape"*. [S.l.], 1999. Citado 22 vezes nas páginas viii, ix, x, xi, xii, xiii, 16, 18, 30, 39, 41, 46, 47, 48, 49, 50, 51, 52, 57, 61, 62 e 63.
- KAUPPINEN, H.; SEPPÄNEN, T.; PIETIKÄINEN, M. An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 17, n. 2, p. 201–207, 1995. Citado na página 13.
- KLINGER, A. *Pattern and search statistics*. 3rd. ed. [S.l.]: Prentice Hall, 2007. ISBN 013168728X. Citado na página 6.
- KUNTTU, L.; LEPISTÖ, L.; RAUHAMAA, J.; VISA, A. Multiscale fourier descriptor for shape classification. In: *Image Analysis and Processing, 2003.Proceedings. 12th International Conference on*. [S.l.: s.n.], 2003. p. 536 – 541. Citado 3 vezes nas páginas 12, 13 e 14.
- KUNTTU, L.; LEPISTÖ, L.; RAUHAMAA, J.; VISA, A. Multiscale fourier descriptor for shape-based image retrieval. In: CAMBRIDGE, UK. *Proceedings of 17th International Conference on Pattern Recognition*. [S.l.], 2004. v. 2, p. 765 – 768. Citado 3 vezes nas páginas 12, 13 e 14.
- LECUN, Y.; C., C.; C.J.C., B. *"The MNIST database of handwritten digits"*. [S.l.], 1998. Citado 8 vezes nas páginas xi, xiii, 39, 53, 54, 55, 57 e 63.
- LINDBERG, L. *Scale-Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994. ISBN 0792394186. Citado 6 vezes nas páginas viii, 2, 4, 5, 7 e 10.
- LIU, M.; VEMURI, B. C.; AMARI, S.; NIELSEN, F. Shape retrieval using hierarchical total bregman soft clustering. *IEEE Transactions on Pattern Analysis and Machine Learning*, v. 34, n. 12, p. 2407–2419, Dezembro 2012. Citado 3 vezes nas páginas 2, 40 e 58.
- LORDÃO, F. A. F. *Reconhecimento de Formas Utilizando Modelos de Compressão de Dados e Espaço de Escalas de Curvatura*. Dissertação (Mestrado) — Universidade Federal de Paraíba, João Pessoa, 2009. Citado 4 vezes nas páginas ix, 20, 21 e 22.
- MIRANDA, P. A. V.; TORRES, R. S.; FALCÃO, A. X. Tsd: A shape descriptor based on a distribution of tensor scale local orientation. In: . [S.l.: s.n.], 2005. v. 0, p. 139–146. ISSN 1530-1834. Citado 3 vezes nas páginas 1, 3 e 32.
- MOKHTARIAN, F.; MACKWORTH, A. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 8, n. 1, p. 34–43, 1986. ISSN 0162-8828. Citado na página 20.

- MOUINE, S.; YAHIAOUI, I.; A., V.-B. Advanced shape context for plant species identification using leaf image retrieval. *International Conference on Multimedia Retrieval*, v. 12, p. 1–8, 2012. Citado na página 58.
- MÜLLER, H.; MÜLLER, W.; SQUIRE, D. M.; MAILLET, S. M.; PUN, T. Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, v. 22, n. 5, p. 593–601, 2001. Citado na página 39.
- NOSATO, H.; SAKANASHI, H.; TAKAHASHI, E.; MURAKAWA, M. Method of retrieving multi-scale objects from optical colonoscopy images based on image-recognition techniques. *Biomedical Circuits and Systems Conference (BioCAS), IEEE*, IEEE, p. 1–4, Outubro 2015. Citado na página 2.
- PARIS, S.; HASINOFF, S. W.; KAUTZ, J. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. In: *ACM SIGGRAPH 2011 Papers*. New York, NY, USA: ACM, 2011. (SIGGRAPH '11), p. 68:1–68:12. ISBN 978-1-4503-0943-1. Citado na página 10.
- ROSENFELD, A.; THURSTON, M. Edge and curve detection for visual scene anaiysis. *IEEE Transactions on Computer*, IEEE Computer Society, v. 20, n. 5, p. 562 – 569, 1971. ISSN 0018-9340. Citado na página 6.
- SAUNDERS, W. B. *Dorland's Illustrated Medical Dictionary*. 25th. ed. Philadelphia: CRC Press, Inc., 1974. ISBN 0849334934. Citado na página 55.
- SEBASTIAN, T.; KLEIN, P.; KIMIA, B. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 26, n. 5, p. 550–571, Maio 2004. ISSN 0162-8828. Citado 13 vezes nas páginas ix, x, xi, 31, 32, 39, 40, 42, 43, 44, 45, 46 e 63.
- SOUZA, G. B.; MARANA, A. N. *Análise de Formas Planas em Imagens Digitais: Uma Nova Abordagem Baseada na Transformada de Hough*. Dissertação (Mestrado) — Unesp/Faculdade de Ciências, Bauru/SP, 2013. Citado 17 vezes nas páginas viii, ix, x, 1, 2, 18, 19, 24, 29, 30, 31, 32, 33, 34, 39, 43 e 63.
- SOUZA, G. B.; MARANA, A. N. HTS and HTSn: New shape descriptors based on hough transform statistics. *Computer Vision and Image Understanding*, v. 127, p. 43–56, 2014. Citado 2 vezes nas páginas xi e 53.
- SUNKAVALI, K.; JOHNSON, M. K.; MATUSIK, W.; PFISTER, H. Multi-scale image harmonization. *ACM Trans. Graph.*, ACM, New York, NY, USA, v. 29, n. 4, p. 125:1–125:10, Julho 2010. ISSN 0730-0301. Citado na página 10.
- TORRES, R. S.; FALCÃO, A. X. Contour salience descriptors for effective image retrieval and analysis. *Image and Vision Computing*, v. 25, p. 3–13, 2007. Citado 6 vezes nas páginas 1, 3, 23, 24, 32 e 39.
- TORRES, R. S.; FALCÃO, A. X.; COSTA, L. F. *A Graph-based Approach for Multiscale Shape Analysis*. 2003. Citado 5 vezes nas páginas ix, 2, 20, 24 e 25.
- WITKIN, A. P. Scale-space filtering: a new approach to multi-scale description. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.]: IEEE, 1984. (ICASSP'84), p. 150 – 153. Citado 3 vezes nas páginas 2, 5 e 6.
- ZHANG, D.; LU, G. A comparative study of fourier descriptors for shape representation and retrieval. In: *Proc. of 5th Asian Conference on Computer Vision (ACCV)*. [S.l.]: Springer, 2002. p. 646–651. Citado 3 vezes nas páginas 1, 3 e 32.