**TSS e TSB: Novos descritores de
forma baseados em *Tensor Scale***

Anderson Meirelles Freitas

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Mestrado

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Paulo André Vechiatto de Miranda

São Paulo, 22 de dezembro de 2017

TSS e TSB: Novos descritores de forma baseados em *Tensor Scale*

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 24/10/2017. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Paulo André Vechiatto de Miranda (orientador) – IME-USP
- Prof. Dr. Aparecido Nilceu Marana – UNESP-Bauru
- Prof. Dr. Jurandy Gomes de Almeida Jr. – UNIFESP

Agradecimentos

Eu gostaria de agradecer ao professor Paulo Miranda que orientou este trabalho com paciência e dedicação, pelos ensinamentos e pela confiança a mim depositada, dando a oportunidade de realizar este trabalho. Agradeço também a minha família, meus pais Gilmar e Tânia, e meu irmão Aluizio, pelo suporte fornecido e pela compreensão nos momentos de ausência deste período. Agradeço também o professor Ricardo da Silva Torres, do IC-Unicamp, pela colaboração na publicação do artigo. E por fim, sou grato a todos os docentes do IME-USP, em especial aos professores Paulo Miranda, Roberto Hirata Jr., Marcel Jackowski e Roberto Marcondes, por me servirem de inspiração, sendo modelos de bons professores e pesquisadores, demonstrando paixão pelo que fazem, que é irradiada a todos os alunos.

Resumo

FREITAS, Anderson M. **TSS e TSB: Novos descritores de forma baseados em *Tensor Scale***. 2017. 72 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Neste trabalho são apresentados dois novos descritores de forma para tarefas de recuperação de imagens por conteúdo (CBIR) e análise de formas, que são construídos sobre uma extensão do conceito de *tensor scale* baseada na Transformada de Distância Euclidiana (EDT). Primeiro, o algoritmo de *tensor scale* é utilizado para extrair informações da forma sobre suas estruturas locais (espessura, orientação e anisotropia) representadas pela maior elipse contida em uma região homogênea centrada em cada pixel da imagem. Nos novos descritores, o limite do intervalo das orientações das elipses do modelo de *tensor scale* é estendido de π para 2π , de forma a melhor discriminar a descrição das estruturas locais. Então, com base em diferentes abordagens de amostragem, visando resumir informações mais relevantes, os novos descritores são construídos. No primeiro descritor proposto, *Tensor Scale Sector* (TSS), a distribuição das orientações relativas das estruturas locais em setores circulares é utilizada para compor um vetor de características de tamanho fixo, para uma caracterização de formas baseada em região. No segundo descritor, o *Tensor Scale Band* (TSB), foram considerados histogramas das orientações relativas extraídos de bandas concêntricas, formando também um vetor de características de tamanho fixo, com uma função de distância de tempo linear. Resultados experimentais com diferentes bases de formas (MPEG-7 e MNIST) são apresentados para ilustrar e validar os métodos. TSS demonstra resultados comparáveis aos métodos estado da arte, que geralmente dependem de algoritmos custosos de otimização de correspondências. Já o TSB, com sua função de distância em tempo linear, se demonstra como uma solução adequada para grandes coleções de formas.

Palavras-chave: recuperação de imagens baseada em conteúdo, CBIR, descritores de forma, análise de formas, escala tensorial, *tensor scale*, visão computacional.

Abstract

FREITAS, Anderson M. **TSS & TSB: New Shape Descriptors based on Tensor Scale**. 2017. 72 p. Dissertation (Masters) - Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2017.

In this work, two new shape descriptors are proposed for tasks in Content-Based Image Retrieval (CBIR) and Shape Analysis tasks, which are built upon an extended tensor scale based on the Euclidean Distance Transform (EDT). First, the tensor scale algorithm is applied to extract shape attributes from its local structures (thickness, orientation, and anisotropy) as represented by the largest ellipse within a homogeneous region centered at each image pixel. In the new descriptors, the upper limit of the interval of local orientation of tensor scale ellipses is extended from π to 2π , to better discriminate the description of local structures. Then, the new descriptors are built based on different sampling approaches, aiming to summarize the most relevant features. In the first proposed descriptor, *Tensor Scale Sector* descriptor (TSS), the local distributions of relative orientations within circular sectors are used to compose a fixed-length feature vector, for a region-based shape characterization. For the second method, the *Tensor Scale Band* (TSB) descriptor, histograms of relative orientations are considered for each circular concentric band, to also compose a fixed-length feature vector, with linear time distance function for matching. Experimental results for different shape datasets (MPEG-7 and MNIST) are presented to illustrate and validate the methods. TSS can achieve high retrieval values comparable to state-of-the-art methods, which usually rely on time-consuming correspondence optimization algorithms, but uses a simpler and faster distance function, while the even faster linear complexity of TSB leads to a suitable solution for very large shape collections.

Keywords: content based image retrieval, CBIR, shape descriptors, shape analysis, tensor scale, computer vision.

Sumário

Lista de Abreviaturas	xi
Lista de Símbolos	xiii
Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Contribuições	4
1.2 Organização do Trabalho	4
2 Descritores de forma	5
2.1 Fundamentos	5
2.2 Descritores de forma por contorno	7
2.2.1 <i>Beam Angle Statistics</i> (BAS)	7
2.2.2 <i>Hough Transform Statistics</i> (HTS) e <i>HTS-neighborhood</i> (HTSn)	8
2.2.3 <i>Inner-Distance Shape Context</i> (IDSC-DP)	8
2.2.4 Descritor de saliências de forma com aspectos globais (SSD+GF)	8
2.2.5 <i>Multi-scale Convexity Concavity</i> (MCC)	8
2.2.6 <i>Height Functions</i> (HF)	9
2.2.7 <i>Complex Network Spectrum Signature</i> (CNSS)	9
2.3 Descritores de forma por região	9
2.3.1 <i>Multi-layer Eigenvector Shape Descriptor</i> (MLEV)	9
3 Fundamentos de <i>Tensor Scale</i>	11
3.1 Conceitos de Escala	11
3.2 Conceitos de <i>Tensor Scale</i>	13
3.2.1 Visualização da <i>Tensor Scale</i>	13
3.3 Algoritmos para cálculo da <i>Tensor Scale</i>	14
3.3.1 Algoritmo de Saha	14
3.3.2 Algoritmo de Miranda <i>et al.</i>	16
3.3.3 Algoritmo de Andaló <i>et al.</i>	17
3.4 Trabalhos prévios com <i>tensor scale</i> para descrição de formas	19
3.4.1 Descritor <i>Tensor Scale</i> (TSD)	19
3.4.2 Descritor baseado em saliências de contorno por <i>Tensor Scale</i> (TSCS)	19

3.4.3	Descritor <i>Tensor Scale</i> por zonas de influência (TSDIZ)	21
4	Novos descritores de forma baseados em <i>Tensor Scale</i>	23
4.1	Extensão da <i>Tensor Scale</i> para 360°	23
4.2	TSS: Descritor por setores circulares	25
4.2.1	Função de dissimilaridade TSS	27
4.3	TSB: Descritor baseado em histogramas por banda	28
4.4	Descritor Composto com Características Geométricas Globais	32
4.5	Análise de Complexidade	33
5	Resultados Experimentais	35
5.1	Medidas de avaliação	35
5.1.1	Precisão vs. Revocação	35
5.1.2	Medida <i>Bullseye Precision</i> (BEP)	36
5.1.3	Taxa de acerto via classificador 1-NN	36
5.2	Experimentos	37
5.2.1	Análise de Formas da base <i>MPEG-7 Core Experiment Shape-1</i>	37
5.2.2	Análise de Formas da base de dígitos MNIST	38
6	Conclusão	45
6.1	Trabalhos Futuros	46
	Referências Bibliográficas	47
	Índice Remissivo	54

Lista de Abreviaturas

BEP	Medida <i>bullseye</i> ou <i>Recall@40</i> (<i>Bullseye Precision</i>)
CBIR	Recuperação de Imagens por Conteúdo (<i>Content-Based Image Retrieval</i>)
EDT	Transformada de Distância Euclidiana (<i>Euclidean Distance Transform</i>)
HSI	Espaço de cores <i>Hue-Saturation-Intensity</i>
IFT	Transformada Imagem-Floresta (<i>Image Foresting-Transform</i>)
PCA	Análise de Componentes Principais (<i>Principal Component Analysis</i>)
SURF	Características Robustas Aceleradas (<i>Speeded Up Robust Features</i>)
TS	Escala tensorial (<i>Tensor Scale</i>)
TSB	Descritor de formas <i>Tensor Scale Band</i>
TSCS	Descritor de formas <i>Tensor Scale Contour Saliences</i>
TSD	Descritor de formas <i>Tensor Scale Descriptor</i>
TSDIZ	Descritor de formas <i>Tensor Scale Descriptor by Influence Zones</i>
TSS	Descritor de formas <i>Tensor Scale Sector</i>

Lista de Símbolos

$\vec{x}$	Vetor geométrico
$\vec{v}_x$	Componente x do vetor $\vec{v}$
$\ \vec{v}\ $	Magnitude do vetor $\vec{v}$
$ \cdot $	Valor absoluto
ℓ_1	Soma dos valores absolutos ou norma L_1
δ_D	Função de dissimilaridade de um descritor D
γ	Ângulo relativo ao setor
$L(p)$	Valor do rótulo no pixel p
$\bar{\mathbf{x}}$	Média amostral

Lista de Figuras

2.1	Uso de um descritor para computar a similaridade entre duas imagens	7
2.2	Exemplo de subdivisão de uma forma pelo descritor MLEV	9
2.3	Representação <i>quadtree</i> do descritor MLEV	10
3.1	Representação da escala <i>ball scale</i>	12
3.2	Representação da <i>tensor scale</i>	14
3.3	Mapeamento de valores do modelo HSI para a orientação das elipses da TS	14
3.4	Passos do algoritmo de <i>tensor scale</i> proposto por Saha	15
3.5	Exemplo de elipse não contida no objeto	16
3.6	Esquema de localização dos pixels de borda proposto por Miranda <i>et al.</i>	17
3.7	Perseguição de bordas com auxílio da EDT	18
3.8	Ilustração sobre técnica de <i>sphere tracing</i>	18
3.9	Exemplos de TS calculado pelo método apresentado por Miranda <i>et al.</i>	20
3.10	Descritor TSCS	21
3.11	Descritor TSCS	21
3.12	Descritor TSDIZ	22
4.1	Ilustração sobre o critério utilizado para determinar o TS com 360 graus	24
4.2	Comparação da TS com 180 e 360 graus no espaço de cores HSI	24
4.3	Setores circulares do TSS e o mapeamento no vetor de características	25
4.4	Ilustração sobre os ângulos relativos dos setores	26
4.5	Características extraídas de setores com diferentes padrões de orientação	27
4.6	Ilustração de bandas concêntricas e setores circulares	29
4.7	Vetores de características TSB, exemplificando robustez a deformações na forma.	29
4.8	Capacidade de distinção entre formas com buracos com histogramas TSB.	30
4.9	Vetores de características TSB, exemplificando invariância a rotação.	31
4.10	Atributos globais considerados no descritor composto	32
5.1	Amostra de imagens da base <i>MPEG-7</i>	38
5.2	Variações nas imagens da coleção <i>MPEG-7</i>	39
5.3	Exemplos de consulta por similaridade com TSS (<i>MPEG-7</i>)	40
5.4	Curvas de Precisão vs. Revocação comparando diferentes descritores (<i>MPEG-7</i>)	41
5.5	Matriz representando a distância mútua entre os pares de imagem (<i>MPEG-7</i>)	41
5.6	Amostra de imagens da base MNIST	42
5.7	Pré-processamento aplicado na base MNIST	42

5.8	Curvas Precisão vs. Revocação (base MNIST).	43
6.1	Imagens da mesma classe com variações da forma em partes que se tocam.	45

Lista de Tabelas

4.1	Complexidade computacional das funções de distância de diferentes descritores . . .	34
5.1	Taxa de acertos do classificador 1-NN na base <i>MPEG-7</i>	38
5.2	Valores BEP obtidos com diferentes descritores (<i>MPEG-7</i>)	40

Capítulo 1

Introdução

Uma das principais características utilizadas pela percepção visual humana para detectar e identificar objetos é a *forma*, que carrega consigo conteúdo semântico suficiente para o seu reconhecimento, na maioria das vezes, mesmo na ausência de outras informações visuais, como textura, cor ou profundidade. Formas são de extrema importância na natureza, sendo as suas variações em organismos e suas partes, ocasionadas por processos biológicos, como mutação, crescimento, adaptação a fatores climáticos e geográficos, ou ainda, diversificação pela seleção natural (Zelditch *et al.*, 2012). Já na percepção visual humana, os resultados dos experimentos realizados por Landau *et al.* (1988) sugerem que a extrapolação da associação de novas palavras a objetos, tanto em crianças como em adultos, está mais relacionada à forma dos objetos do que a cor ou textura (Diesendruck e Bloom, 2015; Landau *et al.*, 1988). De um modo geral, formas são arquétipos de objetos pertencentes a uma mesma categoria, e muitas vezes, formas bidimensionais são projeções de objetos tridimensionais mais complexos (Costa e Cesar Jr., 2009).

Diferentes definições para *forma* são apresentadas na literatura e não há uma definição universalmente aceita ou que possa ser aplicada a todas as situações. Os primeiros trabalhos para delimitar a área de conhecimento que ficou depois conhecida como “análise de formas” foram desenvolvidos por três pesquisadores independentes. Kendall (1977) buscava uma definição de forma para uso na astronomia e arqueologia. Bookstein (1978) iniciou sua teoria sobre formas baseado em problemas da zoologia, e Ziezold (1977) formalizou como estimar distâncias entre formas bidimensionais (Dryden e Mardia, 1998; Kendall *et al.*, 2009).

No trabalho de Kendall (1977), define-se *forma* como toda informação geométrica que permanece quando são removidas de um objeto informações sobre sua posição no espaço, transformações de escala e rotação. Neste trabalho será adotada uma definição mais específica para forma, que será a região delimitada pela silhueta bidimensional de um objeto.

A tomada de decisões de maneira quantitativa e precisa sobre objetos, a partir da coleta de informações geométricas e análise de suas formas, é assunto de grande importância em aplicações para diversas áreas, como biologia, medicina, arqueologia, bioinformática, geologia, geografia e direito (Dryden e Mardia, 2016). Por exemplo, a análise de formas é empregada na medicina no estudo de áreas do cérebro afetadas por doenças neurológicas (Ng *et al.*, 2014), para diagnóstico de patologias e na avaliação da resposta de pacientes à novos tratamentos.

Sistemas de recuperação de imagens por conteúdo, do inglês *Content-Based Image Retrieval* (CBIR), tratam do problema de localizar imagens similares a uma imagem de consulta, armazenadas em grandes bases de dados. Um sistema CBIR pode empregar diversos descritores

para caracterizar as propriedades visuais de imagens. Assim, um descritor é composto por dois componentes: (1) um algoritmo extrator de características, que codifica numericamente uma ou mais propriedades visuais (cor, textura ou forma) em um vetor; e (2) uma medida de dissimilaridade (Torres e Falcão, 2006).

Um tipo particular de sistema CBIR explora informações de formas presentes em imagens para descrevê-las e caracterizá-las, tendo aplicações importantes em diversas áreas, por exemplo: identificação de espécies similares em pesquisas biológicas (Torres *et al.*, 2006), identificação de logomarcas (Jain e Vailaya, 1998; Wei *et al.*, 2009), reconhecimento biométrico de impressões digitais (Karu e Jain, 1996) e busca de imagens de tumores por forma (Korn *et al.*, 1998).

Para que esses objetivos sejam alcançados, um descritor de forma deve ser compacto, robusto a ruídos e pequenas deformações, invariante a transformações geométricas, como rotação, escala, translação e cisalhamento, e ao mesmo tempo conter informações relevantes capazes de distinguir e medir mudanças perceptuais de formas diferentes (Kim e Kim, 2000). Preferencialmente, a função de dissimilaridade usada pelo descritor de formas deve também ser rápida para que a sua utilização seja apropriada em grandes coleções, já que a extração de características geralmente é executada previamente.

Apesar da grande quantidade de trabalhos recentemente propostos neste tópico, ainda existem muitas questões abertas referentes ao desenvolvimento de técnicas eficientes para descrição e discriminação de formas. Isso se dá pelo fato de que cada aplicação apresenta características muito peculiares, como os objetos são representados, quantidade de classes, deformações e possíveis variações intraclasses permitidas, então diferentes descritores e classificadores podem desempenhar melhor que outros em diferentes contextos.

Ao longo dos últimos anos, diversos descritores de forma foram propostos, porém, poucos estudaram aplicações com bases de dados em grande escala (Bai *et al.*, 2016). Neste caso específico, são necessários descritores com custo computacional baixo, ou seja, que apresentem uma função de distância com complexidade computacional próxima à $O(n)$. Muitos descritores de forma propostos na literatura atingem um alto poder de discriminação, mas sacrificam desempenho por demandar elevados custos computacionais, seja de tempo ou espaço, muitas vezes envolvendo programação dinâmica na etapa de casamento (*matching*) para determinar a correspondência entre duas imagens, chegando a apresentar complexidade $O(n^3)$. Como exemplo, os descritores *Inner-Distance Shape Context* (IDSC) (Ling e Jacobs, 2007) e *Height Functions* (Wang *et al.*, 2012) utilizam abordagens baseadas em programação dinâmica para o casamento de seus vetores de características.

Nesta direção, este trabalho apresenta novos descritores de forma projetados para alcançar um alto poder de discriminação com funções de distância de baixo custo computacional. Além de não utilizar programação dinâmica, as funções de distância são mais simples que os algoritmos: OCS (*Optimal Correspondent Subsequence*) usado pelo descritor BAS (Arica e Vural, 2003), OSB (*Optimal Subsequence Bijection*) e algoritmo do Húngaro usado por Bai e Latecki (2008). Comparado com outros métodos especificamente propostos para lidar com grandes bases de dados, tais como *Hough Transform Statistics* (HTS) e *HTS-neighborhood* (HTSn) (Souza e Marana, 2014), os métodos aqui apresentados possuem menor complexidade computacional no caso de formas não-alinhadas e ainda apresentam taxas de recuperação superiores.

Os novos descritores de forma apresentados neste trabalho são baseados no conceito de escala tensorial (do inglês *tensor scale* – TS) (Saha, 2003, 2005), um parâmetro morfométrico que representa de

forma unificada uma estrutura local (com orientação, espessura e anisotropia) em cada ponto de uma imagem. Diferentes aplicações ao TS já foram estudadas, como em filtragem de imagens (Saha, 2005; Saha e Xu, 2010; Xu *et al.*, 2012), segmentação (Saha e Udupa, 2003), registro (Saha *et al.*, 2003), descritores de forma (Andaló *et al.*, 2010; Miranda *et al.*, 2005), detecção de pontos de saliência de contorno (Andaló *et al.*, 2007; Andaló *et al.*, 2007), interpolação em imagens médicas (Xu *et al.*, 2009, 2011), remoção de efeitos de volume parcial em renderização (Souza *et al.*, 2005) e também na quantificação de morfometria local em redes quasi-aleatórias de ossos trabeculados (Saha e Wehrli, 2004). O algoritmo para o cálculo da TS como originalmente proposto por Saha (2003, 2005) possui um alto custo computacional. Como solução para o problema, Andaló *et al.* (2010) propõem uma implementação simples e eficiente do método original. Mais recentemente, o trabalho de Xu *et al.* (2012) apresentou uma abordagem analítica para imagens volumétricas.

Trabalhos recentes apresentaram o potencial da TS para descrição de formas (Andaló *et al.*, 2007; Andaló *et al.*, 2007; Miranda *et al.*, 2005), porém sem explorar a construção de métodos que apresentem bons resultados aliados à um baixo custo computacional, a fim de uma busca eficiente em grandes bases de dados. O descritor baseado em TS mais promissor até então, o TSDIZ, além de apresentar as limitações inerentes de um descritor de contorno, possui cálculo de dissimilaridade de complexidade superior à $O(n^2)$.

Os trabalhos de recuperação de formas podem ser divididos em quatro categorias: casamento par-a-par (*pairwise matching*), similaridade de contexto (*context based similarity*), baseados em aprendizado computacional, e abordagens utilizando combinação de classificadores. Para os métodos de casamento par-a-par, um vetor de características é extraído de cada forma e uma função de distância é usada para calcular a similaridade (ou dissimilaridade/distância) entre um par de formas. Métodos de reclassificação (*re-ranking*) e agregação de listas, exploram a similaridade contextual de toda coleção de imagens, provenientes de uma consulta geralmente par-a-par para reclassificar os resultados, e são importantes para lidar com bases com muita variação intraclasse (Bai *et al.*, 2010; Pedronette e Torres, 2010, 2012; Yang *et al.*, 2009). Abordagens de aprendizado não-supervisionado são geralmente inspiradas no paradigma de *Bag-of-Words* (BoW) (Csurka *et al.*, 2004; Sivic e Zisserman, 2003, 2009), em que um vocabulário é construído a partir do particionamento de um espaço de características. A qualidade deste particionamento é um fator essencial para definir o poder de discriminação do descritor resultante, e envolve o processamento de características de várias imagens da coleção (Bai *et al.*, 2014). Já abordagens de combinação de classificadores, fazem a fusão de características obtidas por diferentes descritores, combinando seus pontos fortes com o propósito de melhorar a acurácia do método (Shekar *et al.*, 2015).

Existem três tipos de descritores para comparações entre pares de formas (*pairwise matching*) (Latecki *et al.*, 2003): baseados no contorno da forma, de região e baseados em esqueletos (eixo medial). Para objetos articulados, as abordagens baseadas em esqueleto geralmente desempenham melhor que as outras, porém exigem representações complexas, na forma de estruturas hierárquicas de árvores (grafos de *shock*), exigindo operações de edição em grafos (Di Ruberto, 2004; Geiger *et al.*, 2003) na etapa de casamento, tornando-as de elevada complexidade computacional (Sebastian *et al.*, 2004; Sebastian e Kimia, 2005). Diferentes dos métodos baseados em contorno, como o BAS (Arica e Vural, 2003), TSDIZ (Andaló *et al.*, 2010) e *Height Functions* (Wang *et al.*, 2012), os descritores apresentados neste trabalho são descritores de região, apresentando vantagens

como por exemplo, não possuir limitação a espaços topológicos específicos com apenas um contorno fechado (podendo lidar com formas disjuntas), e também menor sensibilidade a erros de segmentação, ruído e oclusão parcial (Oliveira *et al.*, 2015). Assim, os métodos apresentados são versáteis, com perspectivas em outras aplicações, podendo tratar imagens não-binárias (com múltiplos rótulos), formas com buracos e, teoricamente, podem ser ainda estendidos para imagens tridimensionais.

Neste trabalho, os esforços foram concentrados nas atividades de descrição e casamento de formas, por meio da proposta de novos descritores com funções de distância para casamento par-a-par, podendo ser utilizados em conjunto com técnicas de reclassificação para melhorar os resultados de consultas CBIR, e ainda, em abordagens com classificadores combinados.

1.1 Contribuições

O desenvolvimento deste trabalho resultou nas seguintes contribuições:

- i. Desenvolvimento de um descritor de forma por região, *Tensor Scale Sector* (TSS), utilizando características de *tensor scale* extraídas de setores de uma grade circular, com função de distância mais rápida que os métodos anteriores por *tensor scale*;
- ii. Desenvolvimento de um segundo descritor de forma, *Tensor Scale Band* (TSB), baseado em histogramas, com função de distância de complexidade linear;
- iii. Estudo dos descritores TSS e TSB combinados com características geométricas globais;
- iv. Extensão da *tensor scale* até 360°, em relação ao originalmente proposto por Saha (2003, 2005), em que as elipses resultantes possuíam orientação até 180° apenas, resultando em maior poder de discriminação dos descritores derivados.

1.2 Organização do Trabalho

Esta dissertação de mestrado é organizada como descrito a seguir. O [Capítulo 2](#) apresenta uma breve introdução sobre descritores de forma, juntamente às metodologias de descritores dos diferentes tipos. O [Capítulo 3](#) expõe os conceitos teóricos da *tensor scale* e algoritmos para o seu cálculo, assim como métodos de caracterização de formas anteriores para fundamentar as abordagens propostas neste trabalho. O [Capítulo 4](#) detalha a metodologia proposta, principal contribuição deste trabalho, na qual os novos processos de caracterização de formas usando *tensor scale* são desenvolvidos. O [Capítulo 5](#) descreve a metodologia utilizada para avaliação dos experimentos e demonstra os resultados obtidos. Finalmente, o [Capítulo 6](#) discute as conclusões obtidas, bem como possíveis desdobramentos a serem realizados em trabalhos futuros.

Capítulo 2

Descritores de forma

Neste capítulo serão apresentados os conceitos básicos sobre descritores de formas e o seu uso em sistemas de recuperação de imagens por conteúdo (CBIR) será formalizado, com especial destaque para as funções de dissimilaridade. Ao final, uma pequena amostra de descritores de forma da literatura será detalhada e classificada conforme as diferentes categorias.

2.1 Fundamentos

Sistemas de recuperação de imagens por conteúdo empregam um ou mais descritores, geralmente complementares entre si. Cada descritor sumariza um conjunto de características visuais extraídas da imagem, como cor, textura ou forma. A etapa de extração de características está vinculada ao contexto de uma aplicação e ao método de representação da imagem.

No caso de descritores de formas, estes podem ser construídos considerando os seguintes métodos de representação: pelo contorno das formas (i.e., silhueta ou borda do objeto), pelo interior do contorno, ou seja, a intensidade da região de dentro do objeto, e também, pelo eixo medial (esqueleto da forma) delimitado pelo contorno do objeto (Latecki *et al.*, 2003). Ainda, descritores podem ser híbridos e utilizar informações obtidas simultaneamente de diferentes representações.

O processo de aquisição de formas das imagens se dá por diferentes métodos. Alguns objetos já são inerentemente formas que existem apenas no plano bidimensional, por exemplo um logotipo, um desenho ou um caractere de um alfabeto. Outras formas bidimensionais, são nada mais que projeções de um objeto 3D nas quais diferentes ângulos de observação resultam em formas diferentes. Ainda, para imagens naturais, a extração das formas nela presentes está atrelada a difícil tarefa de segmentação, sendo muito sensível a variações de iluminação, sombras nos objetos, irregularidades do fundo, mudanças do ponto de observação e transformações de perspectiva. Com isso, um descritor de imagem deve ser robusto o suficiente para garantir o seu desempenho mesmo nessas condições.

Descritores de formas devem preferencialmente possuir algumas qualidades quanto ao seu funcionamento. Diferentes autores apontam diversas características (Costa e Cesar Jr., 2009; Kimia, 2002; Loncaric, 1998; Marques e Furht, 2012; Zhang e Lu, 2004). De um modo geral, para a etapa de extração de características de um descritor de imagens pode-se enumerar:

1. **similaridade perceptiva:** a distância entre dois vetores de características obtidos pelo descritor deve representar uma medida da dissimilaridade entre duas imagens de forma coerente e consistente com a nossa percepção;

2. **economia:** como um vetor de características não necessariamente precisa permitir a reconstrução de uma imagem, as informações capturadas podem ser bastante resumidas e assim, é desejável que seja compacto no tamanho para seu armazenamento, representando apenas a essência do que torna aquele objeto único ou diferente dos demais;
3. **eficiência:** o cálculo da comparação entre vetores de características deve ser rápido;
4. **estabilidade:** capacidade de tolerar alterações significativas ocasionadas por transformações afins e ainda ser capaz de capturar a similaridade entre as imagens de uma mesma classe;
5. **escalabilidade:** ser independente ao tamanho da base de imagens, capturando informações que possam distinguir classes de objetos similares e distantes.

Para formalizar descritores de imagem no contexto de um sistema CBIR, serão adotadas as definições tais como apresentadas por [Torres e Falcão \(2006, 2008\)](#), como segue:

Definição 1. Uma *imagem* $\hat{I}$ é um par $(D_I, \vec{I})$, na qual:

- D_I é um conjunto finito de pixels (pontos em $\mathbb{Z}^2$, tal que, $D_I \subset \mathbb{Z}^2$), e
- $\hat{I} : D_I \mapsto D'$ é uma função que atribui a cada pixel p em D_I um vetor $\vec{I}(p)$ de valores em algum espaço arbitrário D' (p. ex., $D' = \mathbb{R}^3$ quando uma cor é atribuída a um pixel no sistema de cores RGB).

Definição 2. Um *vetor de características* $\vec{v}_{\hat{I}}$ de uma imagem $\hat{I}$ é um ponto no espaço $\mathbb{R}^n$:

- $\vec{v}_{\hat{I}} = (v_1, v_2, \dots, v_n)$, em que n é a dimensão do vetor.

Definição 3. Um *descriptor de imagem simples* D , ou simplesmente *descriptor* é definido como um par (ϵ_D, δ_D) , no qual:

- $\epsilon_D : \hat{I} \mapsto \mathbb{R}^n$ é uma função que extrai um vetor de características $\vec{v}_{\hat{I}}$ de uma imagem $\hat{I}$;
- $\delta_D : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ é uma função de dissimilaridade (podendo ser uma métrica de distância) que calcula a dissimilaridade entre duas imagens a partir da distância entre seus respectivos vetores de características.

Definição 4. Um *descriptor composto* $\hat{D}$ é um par $(\mathcal{D}, \delta_{\mathcal{D}})$, no qual:

- $\mathcal{D} = \{D_1, D_2, \dots, D_k\}$ é um conjunto de k descritores simples pré-definidos;
- $\delta_{\mathcal{D}}$ é uma função de dissimilaridade que combina os valores de dissimilaridade obtidos de cada descriptor $D_i \in \mathcal{D}, i = 1, 2, \dots, k$.

A [Figura 2.1](#) ilustra o uso de um descriptor simples D para computar a dissimilaridade entre duas imagens $\hat{I}_A$ e $\hat{I}_B$. Um algoritmo de extração ϵ_D é primeiro utilizado para computar os vetores de características $\vec{v}_{\hat{I}_A}$ e $\vec{v}_{\hat{I}_B}$ associados às imagens. Depois, a função δ_D é utilizada para obter o valor da dissimilaridade d entre as imagens.

Um fator importante de um descriptor é a sua função de dissimilaridade que deve apresentar algumas propriedades. Sejam A , B e C instâncias de formas de uma coleção S , é desejado que a função de dissimilaridade δ_D satisfaça as seguintes condições ([Veltkamp e Latecki, 2006](#)):

Figura 2.1: O uso de um descriptor simples D para computar a similaridade entre duas imagens. Adaptado de [Torres e Falcão \(2006\)](#)

- (i) **Não negatividade:** $\delta_D(A, B) \geq 0$.
- (ii) **Identidade:** $\delta_D(A, A) = 0$ para todas as formas $A \in S$.
- (iii) **Unicidade:** $\delta_D(A, B) = 0$ se e somente se $A = B$.
- (iv) **Desigualdade triangular forte:** $\delta_D(A, B) + \delta_D(A, C) \geq \delta_D(B, C)$.

As condições (ii) e (iv) implicam na não negatividade (i). Uma função de distância que satisfaça as condições (ii), (iii) e (iv) é chamada de *métrica*, enquanto uma em que apenas as condições (ii) e (iv) são satisfeitas pode ser chamada de *semi-métrica*, ou *(dis)similaridade*.

A condição (iv), também chamada de desigualdade ultramétrica, é um caso especial da:

- (v) **Desigualdade triangular:** $\delta_D(A, B) + \delta_D(B, C) \geq \delta_D(A, C)$.

Algumas funções de dissimilaridade apresentam a característica de *reconhecimento parcial*, ou seja, podem reconhecer um objeto mesmo na ausência de suas partes e nessa condição, geralmente a desigualdade triangular não é satisfeita. Citando o exemplo de [Veltkamp e Latecki \(2006\)](#): a distância da forma de um homem para um centauro é pequena, a distância de um centauro para um cavalo é pequena, mas a distância de um homem para um cavalo é grande, logo a afirmação $d(\text{homem}, \text{centauro}) + d(\text{centauro}, \text{cavalo}) \geq d(\text{homem}, \text{cavalo})$ não é verdadeira. Define-se então a condição (vi) para uma constante $c \geq 1$:

- (vi) **Desigualdade triangular relaxada:** $c(\delta_D(A, B) + \delta_D(B, C)) \geq \delta_D(A, C)$.

2.2 Descritores de forma por contorno

Nesta seção serão revisados alguns descritores de forma por contorno. Descritores de forma por contorno consideram apenas as informações da borda do objeto na sua descrição. Naturalmente, tendem a ser muito sensíveis a ruídos do contorno.

2.2.1 Beam Angle Statistics (BAS)

O descriptor BAS ([Arica e Vural, 2003](#)) é construído com o conceito de raios (*beams*) emitidos a partir de um ponto de contorno. Um raio é definido como um conjunto de segmentos de retas que conectam um ponto de borda aos demais pontos vizinhos. Para cada ponto, o ângulo entre um par de raios é calculado, extraindo a estrutura topológica do contorno, para que então um vetor de características seja criado utilizando os momentos de terceira ordem de todos os ângulos no conjunto de vizinhanças.

A similaridade entre dois momentos BAS é mensurada pelo algoritmo OCS (*Optimal Correspondent Subsequence*) (Wang e Pavlidis, 1990), tendo complexidade $O(N^2)$, em que N é o tamanho do vetor de características, mas como as formas devem ser alinhadas, um casamento circular é realizado tornando a função $O(N^3)$.

2.2.2 *Hough Transform Statistics* (HTS) e *HTS-neighborhood* (HTSn)

Os descritores HTS e HTSn (Souza e Marana, 2014) usam estatísticas do espaço *Hough* para caracterizar formas de objetos ou regiões. Suas funções de distância possuem complexidade $O(N^2)$ para o casamento de formas não-alinhadas, e $O(N)$ quando previamente alinhadas.

2.2.3 *Inner-Distance Shape Context* (IDSC-DP)

O descritor *Inner-Distance Shape Context* (IDSC-DP) (Ling e Jacobs, 2007) é uma extensão do popular método *Shape Context* (Belongie et al., 2002), porém, ao invés de utilizar a distância euclidiana entre dois pontos no contorno, é considerado o caminho mais curto que não ultrapassa os limites da borda da forma, ou seja, a distância interna pela forma.

A função de distância do método IDSC-DP, emprega um algoritmo baseado em programação dinâmica, com complexidade $O(N^2)$ para encontrar a melhor correspondência entre duas formas pré-alinhadas, caso contrário, a função é $O(N^3)$, tornando-o não ótimo para grandes coleções. Porém, Ling e Jacobs (2007) demonstram que um casamento circular completo não é necessário, pois em testes realizados foi possível obter resultados semelhantes com apenas K rotações, suficientes para alinhar as formas e obter um descritor invariante a rotação, assim, sua complexidade final é $O(KN^2)$.

2.2.4 Descritor de saliências de forma com aspectos globais (SSD+GF)

O descritor de saliências de forma com aspectos globais (*Shape Saliency Descriptor with Global Features* – SSD) (Pedrosa et al., 2013) utiliza pontos de saliência de contorno, analisados de uma perspectiva multiescala, para caracterizar formas. Seus vetores de características podem ser de tamanhos diferentes, e seu algoritmo de casamento utiliza atributos locais e globais extraídos do contorno da forma (*global features* – GF). Para calcular a distância entre a melhor correspondência de duas formas, a minimização de uma função de energia é aplicada.

2.2.5 *Multi-scale Convexity Concavity* (MCC)

No descritor MCC (Adamek e O'Connor, 2004), convexidades e concavidades dos contornos são representadas em diferentes níveis de escala com matrizes bidimensionais, e por meio de um algoritmo baseado em programação dinâmica, a dissimilaridade é calculada. O algoritmo deve ser repetido N vezes com diferentes pontos representando o início do contorno.

O descritor utiliza um algoritmo de programação dinâmica baseado no conceito de *dynamic time warping* (DTW), que possui complexidade $O(N^2)$, em que N é o número de amostras no contorno. A complexidade final do MCC é $O(N^3)$, já que N rotações são necessárias para se obter um casamento circular de formas invariante a rotação.

2.2.6 *Height Functions (HF)*

O descritor *Height Functions* (Wang *et al.*, 2012) é uma representação compacta, baseada em características extraídas de pontos de contorno amostrados ao redor do objeto. O método utiliza um conceito de “função altura” para descrever pontos salientes ao longo do contorno. Apesar de apresentar bons resultados nos experimentos, é computacionalmente custoso devido a sua função de distância, baseada também em programação dinâmica.

2.2.7 *Complex Network Spectrum Signature (CNSS)*

O trabalho Oliveira *et al.* (2015) propõe um método baseado em redes complexas para caracterização de formas (CNSS). O descritor é bastante compacto e apresenta bons resultados, porém, foi validado apenas para análise de grafos gerados a partir de pontos amostrados do contorno externo de uma forma.

2.3 Descritores de forma por região

Descritores de forma baseados em informações de região, compõem o vetor de características considerando a intensidade do interior da forma.

2.3.1 *Multi-layer Eigenvector Shape Descriptor (MLEV)*

O *Multi-layer Eigenvector Shape Descriptor* (MLEV) (Kim e Kim, 2000) é um descritor de região invariante às transformações de rotação, escala e translação. O vetor de características do MLEV é construído a partir informações obtidas de sub-regiões da imagem, a partir de um processo de subdivisão recursiva por quadrantes (Figura 2.2). A divisão de uma região é realizada tomando como base o seu centro de massa, então, o quadrante é definido alinhado às direções principais da região.

Figura 2.2: Exemplo de subdivisão da forma pelo descritor MLEV. (a) Primeira subdivisão; (b) Segunda subdivisão.

As regiões são representadas por uma estrutura hierárquica de *quadtrees* na qual os nós correspondem às sub-regiões da forma. O processo de subdivisão é realizado duas vezes gerando 21 regiões (Figura 2.3). Então, quatro atributos invariantes a translação, rotação e escala são extraídos de cada nó da *quadtrees* e dois para o nó raiz. O vetor de características é formado pelos 82 atributos extraídos da *quadtrees* e a distância de similaridade é calculada pela soma das diferenças absolutas entre os vetores.

Figura 2.3: Representação quadtree com as sucessivas subdivisões da forma pelo descritor MLEV.

Capítulo 3

Fundamentos de *Tensor Scale*

Neste capítulo são apresentados os conceitos básicos relacionados à escala em tarefas de processamento de imagens e visão computacional, bem como uma introdução da *tensor scale* (TS) e seus algoritmos de cálculo. Adicionalmente, os trabalhos prévios de descritores relacionados ao conceito de TS são detalhados para posterior embasamento dos novos métodos.

3.1 Conceitos de Escala

A percepção visual humana é capaz de identificar o seu ambiente a partir de estruturas que são simultaneamente observadas em diferentes resoluções. Tais resoluções variam em níveis, que por sua vez variam das resoluções amplas às resoluções mais finas. Por exemplo, pode-se identificar uma árvore pelo seu formato cilíndrico quando observada de muito longe, mesmo sendo composta por galhos e outras partes, e quando observada de mais perto, suas folhas tornam-se evidentes e pode-se observar sua textura em uma escala mais fina (Lindeberg, 1993). Com isso, certos detalhes relevantes em imagens existem apenas em um intervalo restrito de resoluções espaciais e, portanto, existe uma dependência entre a estrutura da imagem e o seu nível de resolução, sendo importante para o estudo da percepção (Koenderink, 1984).

Define-se escala como uma resolução, ou ainda, como um intervalo de resoluções necessárias para garantir uma representação suficiente e compacta de um objeto (Koenderink, 1984; Saha, 2005). Esse conceito desempenha um papel importante na determinação do compromisso entre suavização de ruídos e percepção/detecção de estruturas (Saha, 2005).

O trabalho de Marr (1982) aponta diversos benefícios da abordagem multiescala para análise de imagens, que parte do princípio de representar e analisar uma imagem em diferentes resoluções. Tal abordagem se demonstrou fundamental em casos em que não existem informações preliminares sobre a escala de observação apropriada (Dorini e Leite, 2010). Os trabalhos de Koenderink (1984) e Witkin (1984) formulam matematicamente o conceito de escala, na forma da teoria espaço-escala.

A escala auxilia na divisão de tarefas complexas de visão computacional em uma hierarquia de processamentos, que progridem gradualmente do macro às microestruturas, ou o inverso (Xu *et al.*, 2012). Geralmente a representação de uma imagem em uma escala específica é obtida pela convolução de um núcleo Gaussiano cujo desvio padrão está relacionado à escala escolhida.

Embora a teoria espaço-escala tenha se demonstrado bastante relevante em uma gama enorme de aplicações, os estudos para formalizar uma “escala local” foram motivados por duas questões observadas nas representações espaço-escala: (1) ausência de um mecanismo para unificar informações

de diferentes escalas, e (2) ausência de um esquema para identificar a escala ótima em cada ponto individual de uma imagem. Essas questões são de bastante impacto em tarefas de visão computacional e processamento de imagens. Por exemplo, o conhecimento de uma escala local de uma imagem permitiria otimizar o tamanho de uma vizinhança em diferentes situações, levando a escolha de uma vizinhança pequena em regiões muito detalhadas ou próximas a contornos, e vizinhanças maiores em regiões homogêneas ou no interior de um objeto.

Nessa direção, diferentes formulações para escala local foram propostas. [Tabb e Ahuja \(1997\)](#) representam a escala local por meio de dois parâmetros, um que controla uma função de distância espacial e outro uma função de distância do critério de homogeneidade, parâmetros os quais são determinados pela variação da presença das bordas das estruturas, sendo este método otimizado para aplicações de segmentação e não generaliza para outras aplicações ([Madabhushi et al., 2006](#)). [Pizer et al. \(1998\)](#) definem um conceito de escala ótima local, selecionada a partir de uma representação espaço-escala na qual os valores de uma medida de eixo medial em diferentes escalas são maximizados. [Elder e Zucker \(1998\)](#) descrevem escala local como a menor escala que fornece uma medida de gradiente acima de um limiar estatisticamente confiável. *Wavelets* e Funções Elementares de Gabor (GEF – *Gabor Elementary Functions*) proveem informações relacionadas à escala local ([Bovik et al., 1990](#); [Mallat, 1989](#)). Mais especificamente, *wavelets* fornecem informações sobre as frequências locais e foram extensivamente estudadas em aplicações para compressão de imagens. As funções GEF são filtros que podem ser otimamente localizados nos domínios do espaço e da frequência e, utilizando uma GEF assimétrica, é possível obter a orientação de padrões locais, sendo populares na análise de texturas ([Bovik et al., 1990](#); [Saha, 2005](#)).

Entretanto, essas abordagens de escala local proveem apenas uma inferência estatística sobre a frequência e orientação local de uma região de tamanho pré-determinado, logo não é óbvio como esses métodos podem ser utilizados para determinar o tamanho, orientação e anisotropia de uma estrutura formada localmente pela continuidade da sua homogeneidade.

Assim, os trabalhos de [Saha et al. \(2000, 2001\)](#) introduzem o conceito de escala local morfométrica utilizando um modelo baseado em bolas, denominado *ball scale*, para determinar o tamanho da estrutura local em um determinado ponto da imagem. Este modelo é generalizável para n -dimensões, ou seja, círculo em duas dimensões e esfera em três dimensões. Portanto, para o caso 2D, a medida *ball scale* de um ponto é definida como o raio do maior círculo circunscrito em uma região homogênea centrado num ponto ([Figura 3.1](#)).

Figura 3.1: Representação da escala ball scale no ponto p em uma forma irregular (plano euclidiano).

Porém, ambas as teorias de espaço-escala e escala local utilizam implicitamente um modelo isotrópico de escalas, enquanto a maioria das estruturas no mundo real são anisotrópicas (Xu *et al.*, 2012), justificando o conceito de uma “escala tensorial”, ou seja, uma representação unificada e simultânea de atributos da estrutura local.

Portanto, o modelo batizado de *tensor scale*, apresentado nos trabalhos de Saha (2003, 2005), foi introduzido como a evolução natural do modelo isotrópico de escala *ball scale*, sendo um modelo mais rico, que resolve a maior limitação do modelo esférico, que é a falta de informações sobre a orientação e anisotropia da estrutura local (Saha, 2003, 2005).

Já o trabalho de Madabhushi *et al.* (2006) desenvolve o modelo de uma *escala generalizada*, na qual a escala local de um ponto é representada como o maior conjunto de pontos conectados ao ponto em consideração, que satisfazem algum critério de homogeneidade, sem assumir nenhuma forma como bola ou elipse, ou restrição de anisotropia, como impõem os modelos *ball* ou *tensor scale* (Souza *et al.*, 2008).

3.2 Conceitos de *Tensor Scale*

Formalmente, define-se o modelo de escala local morfométrica *tensor scale* (TS) como a representação paramétrica da maior elipse, ou elipsoide em um volume 3D, contida em uma região homogênea centrada em um ponto p . Uma região homogênea é caracterizada por algum critério pré-definido de homogeneidade, que varia de acordo com a aplicação. Por exemplo, em imagens tons-de-cinza pode-se adotar a contiguidade espacial dos valores de intensidade e, para o caso de imagens binárias, a transição entre os rótulos dos objetos. O modelo de *tensor scale* pode ser calculado para todos os pontos de uma imagem, fornecendo, portanto, uma representação unificada e simultânea de três fatores, conforme representados pela Figura 3.2:

- Orientação do semieixo maior da elipse: $\theta_{\mathbf{t}_1(p)}$;
- Anisotropia: $\sqrt{1 - \frac{\|\mathbf{t}_2(p)\|^2}{\|\mathbf{t}_1(p)\|^2}}$;
- Espessura: $\|\mathbf{t}_2(p)\|$,

em que $\|\mathbf{t}_1(p)\|$ e $\|\mathbf{t}_2(p)\|$ correspondem aos comprimentos dos dois semieixos da elipse centrada em p e $\|\mathbf{t}_1(p)\| \geq \|\mathbf{t}_2(p)\|$.

A espessura assume um valor que é proporcional ao tamanho da estrutura local e depende da escala (tamanho) da imagem. Já a anisotropia é uma razão invariante, assumindo um valor no intervalo $[0, 1]$, representando a dessemelhança da elipse de um círculo. A orientação θ do semieixo maior das elipses, originalmente em Saha (2003, 2005), está contida no intervalo $[0, \pi]$. Neste trabalho, o intervalo utilizado será estendido para $[0, 2\pi]$ (Seção 4.1).

3.2.1 Visualização da Tensor Scale

Para a visualização de imagens *tensor scale*, este trabalho adota a mesma convenção de representação definida por Saha (2003), a partir do mapeamento para um espaço de cores *Hue-Saturation-Intensity* (HSI). Assim, os três atributos da *tensor scale*, que são, orientação, anisotropia e espessura,

Figura 3.2: Representação da tensor scale no ponto p .

são respectivamente codificados para matiz (*hue*), saturação (*saturation*) e intensidade (*intensity*) do espaço HSI. O valor da espessura $\|t_2(p)\|$ é normalizado pela maior espessura da imagem.

No modelo original da *tensor scale* (Saha, 2005), a orientação θ de uma elipse está no intervalo $[0^\circ, 180^\circ]$, ou seja, quando $\theta > 180^\circ$ utiliza-se o valor $\theta - 180^\circ$. Assim, devido a simetria axial das elipses, mesmo quando possuem direções opostas, como 45° e 225° , são idênticas e são atribuídas ao mesmo valor de matiz, como observado na Figura 3.3a. Na Seção 4.1, como a orientação das elipses será estendida para 360° , o valor do matiz do modelo HSI poderá ser atribuído diretamente a orientação da elipse, sem nenhuma conversão (Figura 3.3b).

Figura 3.3: Mapeamento de valores matiz (H) do modelo HSI para a orientação das elipses da *TS*.

3.3 Algoritmos para cálculo da *Tensor Scale*

Diversos algoritmos para cálculo da *tensor scale* foram propostos na literatura, os quais estão detalhados nesta seção.

3.3.1 Algoritmo de Saha

No algoritmo proposto por Saha (2003, 2005), para obter o *TS* de um determinado pixel, calcula-se a maior elipse contida na mesma região homogênea traçando linhas de amostragem de 0 a 179 graus ao redor do mesmo, seguido dos seguintes passos:

- (i) obtenção do perfil de intensidade em cada linha amostrada (Figura 3.4a);
- (ii) localização de dois pontos ótimos de borda em cada linha amostrada (Figura 3.4b);
- (iii) reposicionamento dos pontos de borda de forma que fiquem equidistantes do pixel, seguindo o eixo de simetria da elipse (Figura 3.4c);
- (iv) cálculo da elipse que melhor se ajusta aos pontos de borda reposicionados usando Análise de Componentes Principais (PCA) (Figura 3.4d).

Figura 3.4: (a) Representação de doze pares ($m = 12$) de linhas de amostragem traçadas de 0 a 179 graus com comprimento determinado por um limiar; (b) Pontos de borda localizados em ambas as direções (setas vermelha e amarela). (c) Reposicionamento dos pontos mais distantes. (d) Elipse melhor ajustada aos pontos.

Estes passos são executados para todos os pixels da imagem até que todas as elipses sejam calculadas (Saha e Udupa, 2003; Saha, 2005; Saha *et al.*, 2003).

O cálculo da elipse melhor ajustada aos pontos de borda é realizado em dois sub-passos:

- (i) Determinação da orientação θ da elipse;
- (ii) Cálculo dos comprimentos dos semieixos.

Para implementar o [sub-passo \(i\)](#) são utilizados os autovetores e autovalores da matriz de covariância do PCA, que é calculada considerando todos os pontos reposicionados. Portanto, o ângulo da elipse, é o formado entre o autovetor associado ao maior autovalor com o eixo principal. Já [sub-passo \(ii\)](#) é realizado pela minimização de uma função de erro. Entretanto, as elipses calculadas por esse método podem não estar inteiramente contidas na mesma região homogênea ([Figura 3.5](#)), já que os valores dos comprimentos dos semieixos a e b podem não ser válidos em algumas situações ([Saha, 2005](#)).

Figura 3.5: Exemplo de elipse não contida no objeto. (a) Linhas de amostragem traçadas em ponto localizado no centro do objeto. (b) Pontos de borda considerados. (c) Elipse calculada pelo algoritmo de [Saha \(2005\)](#) e não está contida no objeto.

3.3.2 Algoritmo de Miranda *et al.*

O principal problema da *tensor scale*, como originalmente proposta por [Saha \(2003, 2005\)](#), é o seu alto custo computacional, tornando-a impraticável em três ou mais dimensões ([Xu *et al.*, 2012](#)). Uma abordagem levemente otimizada para o cálculo da *tensor scale* foi apresentada por [Miranda *et al.* \(2005\)](#), com três mudanças principais.

A primeira mudança se encontra na fase de localização dos pixels de borda sobre as linhas de amostragem traçadas ao redor do pixel sob análise. Ao invés de se percorrer cada linha de amostragem individualmente, propõe-se percorrer aos pares e de forma simultânea ([Figura 3.6](#)). Dessa forma, a etapa de reposicionamento dos pontos não é mais necessária ([Figura 3.4c](#)).

A segunda mudança é a localização das bordas a partir de simples limiares conectados, e a terceira melhoria, corresponde a fase de encontrar a melhor elipse ajustada aos pontos.

Ao invés de fazer uso de PCA, é proposta uma função que permite encontrar o ângulo de inclinação da elipse, encontrando o valor γ que minimiza a função g ([Equação 3.1](#)), na qual (x_i, y_i) são coordenadas relativas dos pontos de borda em cada um dos m pares de linha de amostragem em relação ao pixel central $p = (x_p, y_p)$ da elipse, e $(x_{i_\gamma}, y_{i_\gamma})$ são as novas coordenadas dos pontos após aplicar uma rotação pelo ângulo γ ([Miranda *et al.*, 2005](#)).

Figura 3.6: Pares de linhas de amostragem são percorridos simultaneamente nas direções opostas. Otimização proposta por [Miranda et al. \(2005\)](#).

$$g(\gamma) = \sum_{i=1,2,\dots,2m} [x_{i_\gamma}^2 - y_{i_\gamma}^2], \quad (3.1)$$

$$\text{em que } x_{i_\gamma} = x_i \cos \gamma - y_i \sin \gamma \quad (3.2)$$

$$y_{i_\gamma} = x_i \sin \gamma + y_i \cos \gamma \quad (3.3)$$

3.3.3 Algoritmo de Andaló *et al.*

Para imagens binárias (ou com valores provenientes de rótulos fixos), em [Andaló et al. \(2007\)](#) foi proposta uma implementação de TS simplificada, que explora os círculos obtidos pela Transformada de Distância Euclidiana – *Euclidean Distance Transform* (EDT). A EDT é calculada em tempo linear usando a Transformada Imagem-Floresta – *Image Foresting Transform* (IFT) ([Falcão et al., 2004](#)), um método baseado em grafos para a criação de operadores em imagens baseados em conexidade.

Algoritmo de *Tensor Scale* via Transformada Imagem-Floresta

Para localizar os pontos de borda mais próximos do pixel em cada linha de amostragem ([Figura 3.7a](#)), calcula-se a EDT da imagem usando a IFT. Em uma fila de prioridade, são inseridos como sementes de custo unitário para a IFT, todos os pixels com algum vizinho de rótulo diferente. Ao rodar a IFT com uma função de custo de distância euclidiana, é obtido um mapa de custos ([Figura 3.7b](#)), no qual a intensidade de cada pixel representa a distância euclidiana até o ponto mais próximo pertencente à um rótulo distinto ([Falcão et al., 2004](#)).

Para obter os pontos de borda em cada linha de amostragem ao redor de um determinado pixel ([Figura 3.7a](#)), os valores da EDT são utilizados para otimizar o processo de busca por meio de uma sequência de saltos, pois, ao invés de se realizar uma busca linear exaustiva, todos os pixels subscritos no círculo de raio dado pela EDT são descartados ([Figuras 3.7c, 3.7d e 3.7e](#)). Esta abordagem pode ser estendida para imagens multidimensionais, além de eliminar a restrição de tamanho máximo das elipses ([Miranda et al., 2005; Saha, 2005](#)). Vale observar a semelhança entre esta estratégia de busca do ponto de borda com a técnica introduzida por [Hart \(1996\)](#) como *sphere tracing*, no contexto de computação gráfica como uma forma eficiente de renderização de superfícies implícitas ([Figura 3.8](#)).

Figura 3.7: Perseguição de bordas com auxílio da EDT. (a) Linhas de amostragem traçadas para localização das bordas na direção da seta. (b) Mapa de custos obtido com a EDT. (c) Primeiro passo da perseguição de borda com círculo de raio igual ao valor da EDT no ponto. (d) Segundo passo. (e) Borda encontrada em apenas seis passos.

Figura 3.8: Processo de busca da intersecção de raio com superfície com sphere tracing.

Ao considerar imagens de tamanhos distintos obtidas após uma operação de mudança de escala, as distâncias a serem percorridas serão proporcionais ao fator de escala. Da mesma forma, os valores da EDT também variam na mesma proporção preservando o número de passos necessários para chegar ao próximo ponto de borda. Neste sentido, a complexidade do algoritmo é linear em relação ao número de pixels. Vale notar também que a localização de pontos de borda para pixels distintos é independente, podendo, portanto, ser calculada de forma paralela.

Nesta dissertação, foi adotada uma abordagem na qual o valor da EDT representa a espessura $b = \|\mathbf{t}_2(p)\|$. Assim, $a = \|\mathbf{t}_1(p)\|$ é a única variável restante, o que leva a equações simplificadas, estando a elipse final contida na mesma região homogênea de forma mais consistente.

Portanto, pode-se encontrar a a partir da [Equação 3.4](#):

$$a = \sqrt{\frac{C \cdot b^2}{b^2 \cdot A - B}} \quad (3.4)$$

em que A , B e C correspondem a:

$$A = \sum u_i^2, \quad B = \sum u_i^2 v_i^2, \quad C = \sum u_i^4, \quad (3.5)$$

e (u_i, v_i) são as coordenadas relativas dos pontos de borda reposicionados em relação ao pixel central e já rotacionadas em $-\theta$, tal que o maior semieixo da elipse fique alinhado com o eixo horizontal. Para garantir que todos os valores a sejam válidos, os pontos com $v_i^2 \geq b^2$ devem ser ignorados durante o cálculo da Equação 3.4, já que os mesmos correspondem a pixels fora da região de possíveis pontos de borda das elipses com espessura b . Para garantir que no mínimo um ponto de borda está em uma região válida, uma linha de amostragem adicional é traçada na direção dada pelo ângulo θ .

Para os propósitos do descritor de forma para CBIR, pode-se assumir que θ é simplesmente a direção ortogonal da linha ligando o pixel central ao seu pixel mais próximo de rótulo distinto (raiz do caminho) dado pela EDT. Nos experimentos foi adotada essa simples e eficiente variação, considerando também uma única linha de amostragem na direção de θ para estimar a .

3.4 Trabalhos prévios com *tensor scale* para descrição de formas

Diferentes trabalhos para descrição de formas com *tensor scale* já foram propostos, os quais são detalhados nesta seção.

3.4.1 Descritor *Tensor Scale* (TSD)

O descritor de imagens em tons de cinza TSD (Miranda *et al.*, 2005) baseia-se na observação de que objetos distintos geralmente apresentam diferentes padrões de distribuição das orientações de *tensor scale*. Assim, utilizando o algoritmo apresentado na Subseção 3.3.2, calcula-se o TS em cada pixel da imagem. A Figura 3.9 ilustra o resultado do algoritmo em uma imagem em tons de cinza. Então, um histograma das orientações locais de 180 posições é construído, considerando apenas as elipses com anisotropia alta e espessura dentro de um certo intervalo de valores relativamente altos. Essa restrição é aplicada para filtrar elipses com baixa anisotropia ou muito espessas, pois geralmente não possuem uma orientação bem definida.

O histograma do TSD deve ser analisado de forma circular, pois as posições 0 e 179 do histograma são influenciadas por elipses com orientações próximas. Assim, o efeito de uma rotação por um ângulo θ equivale a deslocar por θ posições no histograma discretizado, tornando a análise insensível à rotação.

Para o casamento dos histogramas TSD, primeiramente deve-se normalizar seus valores para que o processo se torne insensível à escala, depois, a distância entre os histogramas é encontrada pelo menor erro absoluto dentre as rotações de uma correspondência circular.

3.4.2 Descritor baseado em saliências de contorno por *Tensor Scale* (TSCS)

Define-se como saliências de um objeto, pontos de alta curvatura ao longo do contorno, ou vértices ao longo do contorno com descontinuidade na derivada primeira (Andaló *et al.*, 2007). A sua detecção é importante em algumas aplicações de processamento de imagens, como registro, aproximação poligonal, análise de movimento e descrição de formas (Pedrosa *et al.*, 2011; Torres e Falcão, 2007; Torres *et al.*, 2003). A utilização de saliências para descrição de formas é baseada no fato apontado

Figura 3.9: Exemplos de TS calculado pelo método apresentado por Miranda et al. (2005). (a) Imagem de entrada. (b) Versões de imagem TS calculadas para imagem original e sua rotação a 15° . A codificação HSI das orientações é representada pelo círculo do centro. Extraído de Miranda et al. (2005).

por Attneave (1954), que afirmou que a informação visual é altamente redundante no espaço e no tempo (Costa e Cesar Jr., 2009). Em um experimento demonstrou que um tipo de redundância está relacionado a segmentos de reta ao longo do contorno de uma forma, significando que os pontos de alta curvatura concentram mais informação do que outros pontos na forma. Por esta razão, pode-se concluir que a curvatura, por concentrar muita informação da forma, é uma importante característica para a identificação de aspectos geométricos (Andaló, 2007).

Considerando esses aspectos, o descritor por contorno *Tensor Scale Contour Saliences* (TSCS) proposto em Andaló et al. (2010) utiliza saliências de contorno para descrever formas e constrói um método para detectá-las utilizando as orientações de *tensor scale*.

O método é iniciado calculando o TS para todos os pixels do interior do objeto (Figura 3.10a). Em um mapa de valores M , para todo ponto p de contorno, é atribuído o valor da orientação da elipse com a máxima anisotropia contida na região Voronoi discreta (zona de influência) associada àquele ponto. Como mais de um ponto pode ser atribuído à mesma zona de influência, pontos sem orientação definida recebem o mesmo valor de orientação de pontos vizinhos. As zonas de influência são eficientemente obtidas por meio de uma propagação de rótulos utilizando uma IFT com distância Euclidiana (Falcão et al., 2004).

Para que as saliências ao longo do contorno sejam detectadas, um processamento nesse mapa de orientações é realizado, comparando a diferença entre o ângulo de cada ponto $p \in M$ com os vizinhos $p - 1$ e $p + 1$, e um limiar é utilizado para filtrar apenas as respostas de diferenças angulares acima de um determinado limiar (Figura 3.11).

O vetor de características é construindo com os valores de saliência (Figura 3.10b) e as suas posições relativas à um ponto de referência, sendo que o tamanho do vetor é variável e depende do número de saliências detectadas. A função de distância do TSCS é um algoritmo heurístico que registra os vetores a partir de um ponto de referência, levando em consideração o tamanho dos vetores e os valores de saliência, e é de complexidade $O(N^2)$.

Figura 3.10: *Descriptor TSCS. (a) Imagem TS calculada para o interior de um objeto. (b) Pontos de saliência detectados pelo método TSCS. Extraído de Andaló et al. (2010)*

Figura 3.11: *Descriptor TSCS: Limiar aplicado para detectar pontos de saliência a partir das diferenças angulares. Adaptado de Andaló et al. (2010).*

3.4.3 Descritor *Tensor Scale* por zonas de influência (TSDIZ)

O descritor por contorno baseado em *tensor scale* por zonas de influência (*Tensor Scale Descriptor with Influence Zones* – TSDIZ) (Andaló et al., 2010), utiliza o conceito, assim como no descritor TSCS, de mapear todos os pontos de contorno de uma forma à alguma informação de sua zona de influência. O contorno primeiramente é subdividido em um número predefinido de segmentos de mesmo comprimento. Então, um vetor de características é construído mapeando para cada segmento, a média angular (Mardia e Jupp, 1999) das orientações das elipses da zona de influência ponderada pelas suas respectivas anisotropias.

Como o vetor de características do TSDIZ depende da ordem com que os pontos são processados, para se calcular a similaridade entre dois vetores, é necessário realizar um alinhamento para que o ponto inicial dos contornos se coincidam. Com os vetores alinhados, a função de similaridade encontra para um par de vetores de características, a diferença mínima de rotação das orientações entre os segmentos.

Figura 3.12: *Descriptor TSDIZ. (a) Contorno de objeto com 40 pontos amostrados. (b) Zonas de influência calculadas para os pontos amostrados. Extraído de Andaló et al. (2010).*

A função de distância do TSDIZ é de complexidade $O(CN^2)$, em que C geralmente é constante em 180 e N o número de pontos no contorno do objeto.

Capítulo 4

Novos descritores de forma baseados em *Tensor Scale*

A primeira contribuição deste trabalho, é a extensão do algoritmo de cálculo da TS, originalmente proposto por [Andaló et al. \(2010\)](#), para que as orientações das elipses estejam no intervalo $[0, 2\pi]$. Sob este modelo mais rico da TS, são propostos dois novos descritores de forma, com maior poder de discriminação em relação aos descritores anteriores baseados em TS para tarefas de recuperação de imagens por conteúdo: *Tensor Scale* baseado em setores (TSS) e outro baseado em histograma de bandas circulares (TSB).

4.1 Extensão da *Tensor Scale* para 360°

Nesta seção, o modelo tradicional de *tensor scale* será estendido para um modelo que considera a orientação das elipses até 2π . Esta extensão aumenta a capacidade de discriminação dos descritores a serem apresentados, já que a orientação das elipses capta informações de polaridade das transições de intensidades de imagens binárias.

Considerando uma imagem binária, temos uma imagem de rótulos. Teoricamente, a imagem pode ter n rótulos distintos. Utilizaremos a notação $L(p)$ para representar o valor do rótulo no pixel p .

As elipses são diferenciadas comparando o valor do rótulo $L(p)$ no seu pixel central p com o rótulo $L(r)$ do pixel mais próximo r de rótulo distinto ([Figura 4.1](#)). Considera-se o vetor $\vec{v}_{rp}$ como a diferença entre os vetores $\vec{p}$ e $\vec{r}$:

$$\vec{v}_{rp} = \vec{p} - \vec{r} = (p_x, p_y) - (r_x, r_y) \quad (4.1)$$

Se $L(p) > L(r)$, então a orientação da elipse é obtida do ângulo formado pelo vetor $\vec{v}_{rp}$ rotacionado 90° no sentido horário ([Figura 4.1a](#)), caso contrário ele é rotacionado 90° no sentido anti-horário ([Figura 4.1b](#)).

A [Figura 4.2](#) mostra alguns exemplos de TS codificados no espaço de cores do HSI, conforme convencionado na [Subseção 3.2.1](#). O TS do método de [Andaló et al. \(2010\)](#) considera as orientações somente $[0, \pi]$, dessa forma elipses com direções opostas recebem a mesma cor/orientação ([Figura 4.2a](#)). Ao utilizar o procedimento ilustrado na [Figura 4.1](#), é possível estender as orientações

Figura 4.1: O TS com 360 graus é obtido comparando o valor do rótulo $L(p)$ no pixel central p com o rótulo $L(r)$ do pixel mais próximo r , com rótulos distintos. Nesta figura, temos: (a) $\theta = 270^\circ$ ($L(p) > L(r)$) e (b) $\theta = 90^\circ$ ($L(p) \leq L(r)$).

a 360° , tal que direções opostas são atribuídas à valores distintos (Figura 4.2b).

O uso da *tensor scale* estendida a 360° resulta em padrões de orientação mais complexos, tais como as descontinuidades nos pixels próximos ao eixo medial da forma.

Figura 4.2: Comparação da TS com 180 e 360 graus, calculada dentro e fora da forma. (a) Imagem TS calculada pelo algoritmo de Andaló et al. (2010) codificada no espaço de cores HSI. (b) Mesma imagem com a TS estendida a 360 graus, como definido neste trabalho.

4.2 TSS: Descritor por setores circulares

No primeiro descritor projetado neste trabalho, o *Tensor Scale Sector* (TSS), considera-se uma região circular ao redor do objeto, que é subdividida em setores de bandas concêntricas (Figura 4.3). Então, as elipses de *tensor scale* são calculadas apenas para os pixels dentro de todos esses setores.

A Figura 4.3a ilustra um exemplo de uma região circular usando quatro bandas e vinte setores por banda. Os raios das bandas são uniformemente espaçados em intervalos regulares.

As elipses que estão mais próximas do contorno circular externo do que do objeto são desconsideradas, pois não apresentam informações relevantes sobre a forma sob análise (Figura 4.2).

Para a construção do vetor de características deste descritor, apenas uma grade é considerada na imagem, diferentemente de outras abordagens tais como *Shape Contexts* (Belongie *et al.*, 2002), que estabelece uma grade para cada ponto de interesse amostrado ao longo do contorno.

Figura 4.3: Setores circulares do TSS e o mapeamento no vetor de características. (a) Extração do vetor de características do TSS. (b) Exemplo de região circular em um caso simples com três bandas ($B = 3$ e oito setores por banda ($S = 8$)). Cada elipse está contida numa banda radial, de índice $b \in [0, B - 1]$, e com índice de setor de $s \in [0, S - 1]$ em uma determinada banda. A posição j no vetor de características é dada por $j = b + s \times B$.

Os atributos das elipses de cada setor são usados como características para compor um vetor de tamanho fixo, considerando a orientação θ de cada elipse convertida à um ângulo γ mensurado de forma relativa ao seu próprio setor, como ilustrado na Figura 4.4.

O uso de ângulos relativos γ simplifica o cálculo da função de distância do TSS e possibilita a invariância a rotação do descritor TSB, conforme detalhado na Subseção 4.2.1.

De forma a sumarizar e obter a direção local dominante que representa os ângulos relativos γ_i contidos em cada setor, poderia ser utilizada a direção média, conforme Equação 4.2 (Mardia e Jupp, 1999).

$$\bar{\gamma} = \tan^{-1} \left[\frac{\sum \sin(\gamma_i)}{\sum \cos(\gamma_i)} \right] \quad (4.2)$$

Porém, neste trabalho, foi utilizada uma abordagem mais elaborada. Cada ângulo é tratado

Figura 4.4: Para cada elipse no TSS e TSB, são considerados ângulos γ relativos ao seu próprio setor, ao invés do ângulo θ

como um vetor $(v_x, v_y) = (w_i \cos(\gamma_i), w_i \sin(\gamma_i))$ para que a sua magnitude seja também explorada. Um peso w_i é atribuído a cada elipse, sendo obtido pelo valor da anisotropia ao quadrado. O intuito é penalizar as elipses com baixa anisotropia, já que tendem a ser círculos (simétricas) e, portanto, não possuem uma orientação muito bem definida.

Assim, a orientação relativa dominante local $(\bar{v}_x, \bar{v}_y)$ dos setores é obtida com as [Equações 4.3](#) e [4.4](#), em que γ_i e w_i são os valores das orientações relativas e o quadrado das anisotropias de todas as n elipses inscritas no setor.

$$\bar{v}_x = \frac{1}{n} \sum_{i=1}^n w_i \cos(\gamma_i) \quad (4.3)$$

$$\bar{v}_y = \frac{1}{n} \sum_{i=1}^n w_i \sin(\gamma_i) \quad (4.4)$$

Para captar ângulos relativos com orientações opostas, extrai-se também a soma dos valores absolutos das respostas ([Equações 4.5](#) e [4.6](#)). Assim, haverá para cada setor um vetor com quatro características $\mathbf{v} = \{v_1, v_2, v_3, v_4\}$, em que $v_1 = \bar{v}_x$ e $v_2 = \bar{v}_y$ ([Figura 4.5](#)). Estas características foram inspiradas no vetor de quatro-dimensões usado no método de detecção de pontos chave do descritor SURF ([Bay et al., 2008](#)).

$$v_3 = \frac{1}{n} \sum_{i=1}^n |w_i \cos(\gamma_i)| \quad (4.5)$$

$$v_4 = \frac{1}{n} \sum_{i=1}^n |w_i \sin(\gamma_i)| \quad (4.6)$$

As características extraídas dos setores com a mesma região radial são concatenadas para compor o vetor de características do TSS ([Figura 4.3b](#)). Na etapa de casamento, para tratar imagens refletidas,

Figura 4.5: (a-f) Características $\mathbf{v} = \{v_1, v_2, v_3, v_4\}$ de setores com diferentes padrões de orientação. (a-b) Elipses com orientação vertical na mesma direção. (c) Elipses com orientação vertical com direções opostas. (d-e) Elipses na horizontal com a mesma orientação. (f) Elipses na horizontal com direções opostas.

deve-se considerar um novo vetor de características com os sinais invertidos nas componentes x das Equações 4.3, 4.4, 4.5 e 4.6.

Há dois parâmetros que precisam ser definidos, o centro do objeto e o raio da região circular. Estes parâmetros podem ser personalizados para diferentes aplicações. Nos experimentos apresentados no Capítulo 5, o centro de gravidade (centroide) da forma foi utilizado como ponto central, e o raio como três vezes o tamanho do centroide (Dryden e Mardia, 1998, p. 24; Costa e Cesar Jr., 2009, p. 287).

O espaçamento entre os raios das bandas circulares pode também ser obtido com uma subdivisão igualmente espaçada no espaço log-polar (Traver e Pla, 2003). Em experimentos com um ponto cego central de 10% do raio máximo, obteve-se resultados similares.

4.2.1 Função de dissimilaridade TSS

A dissimilaridade entre dois vetores de características TSS é calculada com um casamento circular, necessário para obter a invariância rotacional, dada pelo Algoritmo 1, sendo a distância ℓ_2 entre setores correspondentes (Linhas 12–13), que empiricamente demonstrou resultados superiores entre outras métricas de distância. Vale ressaltar que sem o uso dos ângulos relativos γ , durante o casamento circular (Linha 5), os ângulos θ deveriam ser rotacionados pelo mesmo deslocamento

circular, antes da comparação dos vetores fv_1 e fv_2 (Linhas 7–14). Com isso, seria necessário aplicar uma matriz de rotação, que é evitada utilizando os ângulos γ para compensar a diferença, pois os mesmos já são invariantes ao deslocamento circular dos setores.

Algoritmo 1. – FUNÇÃO DE DISSIMILARIDADE TSS

ENTRADA: Vetores de características fv_1 e fv_2

SAÍDA: Distância

```

1.  $S \leftarrow$  número de setores por banda
2.  $B \leftarrow$  número de bandas
3.  $N \leftarrow S \cdot B$ 
4.  $d_{min} \leftarrow \infty$ 
5. Para  $k \leftarrow 0$  Até  $S - 1$  Faça
6.    $offset \leftarrow B \cdot k$ 
7.    $d \leftarrow 0.0$ 
8.   Para  $i \leftarrow 0$  Até  $N - 1$  Faça
9.      $j \leftarrow (i + offset) \% N$ 
10.     $(a_1, a_2, a_3, a_4) \leftarrow fv_1[i]$ 
11.     $(b_1, b_2, b_3, b_4) \leftarrow fv_2[j]$ 
12.     $(d_1, d_2, d_3, d_4) \leftarrow (b_1 - a_1, b_2 - a_2, b_3 - a_3, b_4 - a_4)$ 
13.     $d \leftarrow d + \sqrt{(d_1)^2 + (d_2)^2 + (d_3)^2 + (d_4)^2}$ 
14.     $i \leftarrow i + 1$ 
15.   Se  $d < d_{min}$  Então  $d_{min} \leftarrow d$ 
16.    $k \leftarrow k + 1$ 
17. Retorne  $d_{min}/N$ 

```

4.3 TSB: Descritor baseado em histogramas por banda

O segundo descritor de formas proposto neste trabalho, o *Tensor Scale Band* (TSB), tem por objetivo ser um descritor com função de distância eficiente, e para isso, sacrifica informações do deslocamento angular dos setores.

Diferentemente do TSS, as informações espaciais do TSB estão incorporadas apenas a partir do deslocamento radial, obtido por um histograma normalizado de 60 posições por banda, capturando a distribuição angular relativa de γ . Pelo fato da área das bandas radiais cobrir boa parte da imagem, se demonstrou apropriado utilizar histogramas de 60 posições com a distribuição angular ao invés de apenas um vetor $\mathbf{v}$ de quatro dimensões, como no TSS. Os histogramas são então concatenados para compor um vetor de características de tamanho fixo, que é naturalmente invariante a rotação devido ao uso dos ângulos relativos γ de cada setor.

Para que o método seja menos sensível às posições do histograma com valores muito altos, utiliza-se as raízes dos valores após uma normalização. Assim, a diferença entre o maior e o menor valor das posições do histograma ficam menores que na escala original. Pelo mesmo motivo, foi observado que a métrica de distância ℓ_1 resultou em resultados melhores quando comparados com a métrica ℓ_2 .

Nas Figuras 4.7, 4.8 e 4.9 são mostrados alguns exemplos do funcionamento do descritor TSB. É possível observar que as características do TSB são robustas a deformações na imagem (Figura 4.7), invariantes a rotação (Figura 4.9), e capazes de discriminar objetos com buracos (Figura 4.8) devido

Figura 4.6: Ilustração de bandas concêntricas e setores circulares. No TSB, um histograma é construído para cada banda (a) e não utiliza informações do deslocamento angular (b) para a comparação entre seus histogramas.

às informações de polaridade providas pelas elipses de ângulo estendido da TS com 360° . A função de distância do TSB exige apenas o cálculo de uma distância ℓ_1 entre dois vetores de características, sendo assim, muito eficiente.

No TSB, o número S de setores por banda interfere apenas no cálculo do ângulo relativo (Figura 4.4), mas não aumenta o tamanho do vetor de características. Isso permite que seja utilizado um número S maior de setores por banda no TSB.

Figura 4.7: Vetores de características TSB, exemplificando robustez a deformações na forma.

Figura 4.8: Capacidade de distinção entre formas com buracos com histogramas TSB.

Figura 4.9: Vetores de características TSB, exemplificando invariância a rotação.

Figura 4.10: Atributos globais considerados no descritor composto: (a) Retângulo envolvente alinhado a orientação da forma. (b) Fecho convexo. (c) Elipse com os mesmos momentos estatísticos de segunda ordem da forma.

4.4 Descritor Composto com Características Geométricas Globais

Considerando o fato que ambos os descritores TSS e TSB descrevem uma forma resumizando apenas características locais sobre a TS de regiões delimitadas, setores no TSS e bandas no TSB, este trabalho investigou a hipótese de, ao combinar descritores com características geométricas globais extraídas da forma, poderia aumentar o poder de discriminação original. Esta estratégia foi explorada também nos trabalhos de Adamek e O'Connor (2004), e Pedrosa *et al.* (2016, 2013) para aumentar a distância entre formas dissimilares e aproximar formas que são visualmente similares.

Assim, neste trabalho foram utilizadas as três seguintes características globais sobre a forma para a construção do descritor composto: razão entre os lados do retângulo envolvente (*aspect ratio*) alinhado à orientação da forma (Figura 4.10a), excentricidade, e solidez. A solidez é definida pela razão entre a área da forma e seu fecho convexo, que por sua vez é definido pelo menor polígono convexo que contém todos os pontos pertencentes ao objeto (Figura 4.10b). Já a excentricidade, é definida pela razão entre os semieixos da elipse que possui os mesmos momentos estatísticos de segunda ordem que a forma, centrada no centro de massa do objeto (Figura 4.10c). Vale ressaltar que estas características globais também apresentam as propriedades de invariância a rotação, escala e espelhamento.

A função de dissimilaridade δ_{GF} do descritor composto é obtida a partir de uma combinação convexa das características geométricas globais com a função de distância do descritor original (TSS ou TSB), combinadas por meio da operação *log-sum-exp* (Boyd e Vandenberghe, 2004). Assim, a nova função de dissimilaridade é dada por:

$$\delta_{GF}(A, B) = \delta_D(A, B) + \log \sum_{k=1}^m \exp(\alpha_k \cdot (\varphi_k^A - \varphi_k^B)) \quad (4.7)$$

em que δ_D é a função de dissimilaridade a ser estendida; $\varphi_k^S = (\varphi_1, \dots, \varphi_m)$ o conjunto de m características geométricas da forma S ; e α_k são os coeficientes da combinação convexa, com $\alpha_k > 0$ e $\sum_{k=1}^m \alpha_k = 1$.

Nos experimentos apresentados no Capítulo 5, os valores dos parâmetros α foram encontrados por meio de uma validação cruzada. Vale ressaltar que os valores ótimos são específicos às aplicações, mas os experimentos realizados demonstram que diferentes configurações de parâmetros desempenham razoavelmente bem, o que reforça a generalidade dessa abordagem.

4.5 Análise de Complexidade

No contexto de CBIR, o custo computacional consiste em duas etapas: (i) extração do vetor de características e (ii) cálculo da dissimilaridade por uma função de distância. Em sistemas CBIR que tratam grandes coleções, a segunda etapa é a mais importante, já que em uma consulta a distância deve ser determinada contra cada objeto da coleção a partir do objeto de consulta, enquanto os vetores de características estão geralmente pré-calculados e armazenados de antemão em alguma estrutura de dados. Portanto, para a realização de uma nova consulta, somente o vetor que descreve a imagem de consulta é extraído, e assim, o desempenho de consultas em sistemas CBIR depende diretamente da eficiência da função de distância do descritor utilizado.

A função de distância do descritor TSS apresentada no [Algoritmo 1](#) requer o cálculo da distância entre cada setor correspondente de ambos os vetores de características. O tamanho do vetor de características do TSS é proporcional ao número total de setores $N = S \cdot B$, em que S é o número de setores por banda e B o número de bandas. A complexidade final para calcular a distância entre dois vetores é $O(S \cdot N)$. Quando o espelhamento é desejado, deve-se executar uma segunda iteração sobre o mesmo vetor refletido, e nessa situação o casamento é $O(2 \cdot S \cdot N) = O(S \cdot N)$.

Já no TSB, o vetor de características possui tamanho $N = 60 \cdot B$, em que B é o número de bandas, já que o histograma possui 60 posições por banda radial. Sua função de distância necessita apenas do cálculo da distância ℓ_1 entre dois vetores de características, que possui complexidade $O(N)$, o que se torna $O(60 \cdot B) = O(B)$, ao considerar como constante o número de posições. A utilização da métrica ℓ_1 possibilita o uso de estruturas de indexação eficientes.

Ambas as funções de distância são invariantes à rotação e não necessitam que as formas estejam previamente alinhadas.

A [Tabela 4.1](#) apresenta uma comparação das complexidades do TSB e TSS com outros descritores avaliados nos experimentos, que serão apresentados no [Capítulo 5](#).

Tabela 4.1: Complexidade computacional das funções de distância de diferentes descritores, dispostos em ordem alfabética. Os descritores propostos neste trabalho estão em **negrito**.

Descritor	Complexidade	Referência	Tipo
BAS-60	$O(N^3)$	Arica e Vural (2003)	Contorno
CNSS	$O(N)$	Oliveira <i>et al.</i> (2015)	Contorno
Height Functions	$O(N^3)$	Wang <i>et al.</i> (2012)	Contorno
HTS	$O(N^2)$	Souza e Marana (2014)	Contorno
HTSn	$O(N^2)$	Souza e Marana (2014)	Contorno
IDSC+DP	$O(KN^2)$	Ling e Jacobs (2007)	Contorno
MCC	$O(N^3)$	Adamek e O'Connor (2004)	Contorno
Path Similarity	$O(N^3)$	Bai e Latecki (2008)	Esqueleto
SC+DP (Shape Contexts)	$O(KN^2)$	Ling e Jacobs (2007)	Contorno
SS	$O(N^2)$	Torres e Falcão (2007)	Contorno
SSD+GF	$O(N^2)$	Pedrosa <i>et al.</i> (2013)	Contorno
TSB	$O(N)$	Freitas <i>et al.</i> (2016)	Região
TSCS	$O(N^2)$	Andaló <i>et al.</i> (2010)	Contorno
TSD	$O(N^2)$	Miranda <i>et al.</i> (2005)	Região
TSDIZ	$O(CN^2)$	Andaló <i>et al.</i> (2010)	Contorno
TSS-41	$O(SN)$	Freitas <i>et al.</i> (2016)	Região

Capítulo 5

Resultados Experimentais

Neste capítulo serão detalhados os experimentos realizados com diferentes bases de imagens e descritores de forma, com o fim de avaliar as propriedades e a eficácia da metodologia proposta neste trabalho em diferentes situações.

5.1 Medidas de avaliação

Diferentes medidas para a avaliação dos descritores propostos foram utilizadas nos experimentos. Cada medida avalia um aspecto diferente dos descritores e uma análise combinada das medidas orienta a eficácia das metodologias propostas.

5.1.1 Precisão vs. Revocação

Para a avaliação da eficácia de sistemas CBIR, as curvas de Precisão vs. Revocação (*Precision vs. Recall* – $P \times R$) são utilizadas na maioria dos trabalhos e se tornaram também muito comuns na avaliação de descritores de forma. Define-se precisão (*precision*) como a fração de imagens recuperadas que são relevantes à consulta (Equação 5.1), e de forma análoga, a revocação (*recall*) como a fração de imagens relevantes que foram recuperadas (Equação 5.2) (Torres e Falcão, 2008).

$$\text{Precisão} = \frac{|\{\text{imagens relevantes}\} \cap \{\text{imagens recuperadas}\}|}{|\{\text{imagens recuperadas}\}|} \quad (5.1)$$

$$\text{Revocação} = \frac{|\{\text{imagens relevantes}\} \cap \{\text{imagens recuperadas}\}|}{|\{\text{imagens relevantes}\}|} \quad (5.2)$$

Os valores de precisão e revocação também podem ser definidos considerando apenas os primeiros k resultados retornados por uma consulta, desde que os resultados estejam ordenados pela sua relevância, ou seja, pelos valores da dissimilaridade (Büttcher *et al.*, 2010) (Equações 5.3 e 5.4).

$$\text{Precisão}@k = \frac{|\{\text{imagens relevantes}\} \cap \{k \text{ imagens recuperadas}\}|}{|\{k \text{ imagens recuperadas}\}|} \quad (5.3)$$

$$\text{Revocação}@k = \frac{|\{\text{imagens relevantes}\} \cap \{k \text{ imagens recuperadas}\}|}{|\{\text{imagens relevantes}\}|} \quad (5.4)$$

Excepcionalmente nas equações 5.1, 5.2, 5.3, 5.4 e 5.5, a notação $|\cdot|$ é utilizada para denotar a cardinalidade de um conjunto.

As curvas $P \times R$, indicam o compromisso entre as duas medidas e, portanto, quanto mais próximas do topo do gráfico, melhores são os resultados (Torres e Falcão, 2008), pois capturam o fato que é desejável maximizar o número de acertos enquanto é minimizado o número de falsos positivos.

5.1.2 Medida *Bullseye Precision* (BEP)

Comum na avaliação de descritores especificamente para a base *MPEG-7 CE-Shape-1* (Subseção 5.2.1), a medida *bullseye precision* (BEP), também foi adotada nos experimentos deste trabalho. Genericamente, é definida para uma única imagem de consulta, conforme a Equação 5.5, em que N_q é o número de amostras da classe de uma imagem de consulta q (Kim e Kim, 2000):

$$BEP(q) = \frac{|\{\text{objetos relevantes}\} \cap \{2N_q \text{ mais similares}\}|}{N_q} \quad (5.5)$$

Assim, é possível obter a medida BEP para uma classe ou para toda a base pela média das acurácias de recuperação de todas as imagens de interesse.

Para o caso da base MPEG-7 ($N_q = 20$), portanto, é a contagem de acertos nos 40 resultados mais similares, classificados em ordem crescente pela dissimilaridade, considerando todas as imagens da base como consulta. O resultado então é normalizado pelo maior número possível de acertos ($28.000 = 20 \times 1.400$). A medida BEP para a base MPEG-7 é equivalente ao valor de Revocação considerando as primeiras 40 imagens recuperadas, ou seja, o valor de Revocação@40 (Equação 5.4).

5.1.3 Taxa de acerto via classificador 1-NN

A avaliação de métodos de reconhecimento de formas pode ser também realizada por meio de um experimento com um classificador por k vizinhos mais próximos (k -NN) (Wang *et al.*, 2012). Com o procedimento de *leave-one-out*, cada uma das N imagens da base sob avaliação são usadas para treinamento. O classificador é construído comparando uma imagem de teste contra todas as outras $N - 1$ imagens com a função de distância do descritor. O rótulo da imagem com a menor distância diferente da imagem de teste é considerado como o resultado da classificação, no caso quando $k = 1$.

A acurácia do classificador 1-NN é calculada como a fração entre o número de objetos corretamente classificados e o tamanho N do conjunto de teste, ou seja, para um conjunto de testes $\mathbf{x}$ é o número em que as estimações do classificador $\hat{f}$ concordam com os rótulos y , isto é:

$$\frac{1}{N} \sum_{i=1}^N \begin{cases} 1, & \text{se } \hat{f}(\mathbf{x}_i) = y_i; \\ 0, & \text{caso contrário.} \end{cases} \quad (5.6)$$

Esta medida é equivalente ao valor de precisão na posição 1 (Precisão@1) (Equação 5.3).

5.2 Experimentos

Os descritores TSS e TSB foram comparados com outros descritores populares, sendo avaliados com as seguintes bases de imagens:

- *MPEG-7 CE-Shape-1* (parte B) (Kim e Kim, 2000),
- Base de dígitos MNIST (LeCun e Cortes, 1998).

Experimentos com diversas configurações foram realizados. Para o descritor TSS, os melhores resultados foram obtidos com a configuração de 4 bandas e 20 setores por banda. A última banda foi descartada já que as suas elipses são mais influenciadas pelo contorno circular externo da região usada pelo descritor, do que propriamente pela forma do objeto, sem adicionar informações relevantes. Adicionalmente, como na banda central os setores possuem áreas muito pequenas, eles foram agrupados de forma que apenas uma característica média seja extraída, ao invés de considerar seus setores separadamente. Dessa forma, restam 40 características de setores das duas bandas remanescentes e uma da banda central. Esse descritor foi batizado de TSS-41, se referindo a $B + B + 1$, em que B denota o número de setores por banda ($B = 20$).

No caso do TSB, os experimentos foram realizados com 5 bandas e 32 setores por banda, e assim como no TSS, a última banda também foi descartada, resultando em 4 histogramas de $60 \times 4 = 240$ posições para compor o vetor de características.

5.2.1 Análise de Formas da base *MPEG-7 Core Experiment Shape-1*

A coleção de imagens *MPEG-7 CE-Shape-1* (Kim e Kim, 2000) é dividida em três partes para a execução de experimentos com diferentes objetivos:

Parte A: experimentos de robustez a escala (A1) e rotação (A2);

Parte B: avaliação de desempenho em experimentos de recuperação por similaridade;

Parte C: robustez a mudanças causadas por transformações não-rígidas.

Considera-se que as condições propostas pela Parte A são úteis em qualquer descritor de forma. A Parte B contém 1.400 formas distribuídas em 70 classes, nas quais cada classe contém 20 formas com diversas deformações rígidas e não-rígidas, ruído e mudança de ângulo de visualização (Veltkamp e Latecki, 2006). A Figura 5.1 ilustra alguns exemplos de imagens desta base. É comum na avaliação de descritores de forma com esta coleção considerar apenas a Parte B, já que inclui todos os casos das partes A1, A2 e C, portanto o mesmo critério foi adotado neste trabalho.

Deve-se destacar que a classificação de algumas classes desta coleção não é uma tarefa trivial, devido ao fato que muitas instâncias são apenas mudanças no ponto de vista do mesmo objeto e diferem bastante de outras instâncias da mesma classe, ou ainda, são mais próximas visualmente a imagens de outras classes, como ilustrado na Figura 5.2.

Na Figura 5.3 alguns exemplos de busca por similaridade são ilustrados utilizando o descritor TSS. Como é possível observar, o TSS foi capaz de retornar imagens similares logo nas primeiras posições. Para a classe **key** (linha 4 da Figura 5.3), alguns resultados de outras classes são observados devido à falta de características discriminantes entre eles.

Figura 5.1: Amostra de imagens da base MPEG-7

A Figura 5.4 mostra as curvas $P \times R$ obtidas para todos os descritores avaliados. Com exceção das curvas do TSS, TSB, HTS, HTSn, BAS, TSD e TSDIZ, as curvas dos outros métodos foram extraídas dos artigos originais (Adamek e O'Connor, 2004; Ling e Jacobs, 2007; Pedrosa *et al.*, 2013; Souza e Marana, 2014) por meio de uma cuidadosa inspeção visual. Os resultados dos métodos HTS e HTSn foram gentilmente cedidos pelos autores Souza e Marana.

As distâncias mútuas obtidas com o TSS para todos pares de formas da coleção são mostradas na Figura 5.5. As distâncias foram normalizadas entre 0 e 1, portanto, números menores representam maior similaridade.

A tabelas 5.1 e 5.2 comparam os resultados com outros métodos avaliados nos experimentos, com as medidas BEP e a taxa de acertos via classificador 1-NN respectivamente. No caso dos descritores TSS+GF e TSB+GF, os valores do parâmetro α foram encontrados por validação cruzada.

5.2.2 Análise de Formas da base de dígitos MNIST

O conjunto de teste da base de dígitos MNIST consiste em 10.000 imagens de dígitos manuscritos, originalmente coletadas pelo instituto norte-americano NIST (*National Institute of Standards and Technology*), modificadas e pré-processadas para uso em experimentos de aprendizagem computacional. O uso dessa base é popular para a avaliação de classificadores (LeCun e Cortes, 1998), já que possui

Tabela 5.1: Taxa de acertos do classificador 1-NN observada em diferentes descritores (MPEG-7 CE-Shape-1).

Descritor	1-NN (%)
Height Functions	98.86
TSS	98.00
TSB	94.71

Figura 5.2: Exemplos de variações entre imagens da coleção MPEG-7 CE-Shape-1, com difícil distinção. (a) e (b) apresentam instâncias de uma mesma classe (horse e octopus). (c) Instâncias das classes deer e horse visualmente similares.

um grande número de amostras já classificadas e rotuladas. Na [Figura 5.6](#) é possível observar uma pequena amostra de imagens de diferentes dígitos.

As imagens originais possuem dimensão de 28×28 pixels em tons de cinza (normalizados entre 0 e 1) e para a condução deste experimento, um pré-processamento foi realizado: as imagens foram redimensionadas para 280×280 com interpolação bilinear, e depois limiarizadas em 0.5. Como pode-se verificar na [Figura 5.7](#), houve um aumento na resolução, permitindo o cálculo das elipses de TS para serem utilizadas pelos descritores sem interferência ocasionados pela discretização das imagens com pouca resolução.

Como as imagens da base MNIST estão quase alinhadas, não havendo muita variação de rotação, no experimento com o descritor TSS, a função de distância utilizada considera apenas um intervalo de busca de $S/4$ ([Algoritmo 1, Linha 5](#)), o que corresponde a uma rotação de $-\pi/4$ a $\pi/4$, para encontrar a melhor correspondência circular entre os setores. Esse processo também evita que outros erros sejam ocasionados por um casamento circular completo em S , como por exemplo classificar os números 6 e 9 à mesma classe.

A [Figura 5.8](#) mostra as curvas de Precisão vs. Revocação obtidas pelos métodos avaliados.

Consulta	1	2	3	4	5	6	7	8	9
0.00	0.09	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.12
0.00	0.10	0.16	0.16	0.17	0.18	0.19	0.23	0.23	0.24
0.00	0.04	0.05	0.07	0.07	0.08	0.09	0.10	0.10	0.14
0.00	0.08	0.08	0.10	0.11	0.11	0.12	0.12	0.12	0.12
0.00	0.04	0.04	0.08	0.14	0.15	0.16	0.16	0.17	0.17
0.00	0.09	0.14	0.17	0.17	0.18	0.19	0.19	0.19	0.20

Figura 5.3: Exemplos de consulta por similaridade com TSS, mostrando a imagem de consulta (primeira coluna) e os primeiros nove resultados com suas distâncias (MPEG-7 CE-Shape-1). Erros de classificação estão indicados em **vermelho**.

Tabela 5.2: Valores BEP obtidos com diferentes descritores (MPEG-7 CE-Shape-1).

Descritor	BEP (%)	Complexidade	Tipo
Height Functions	89.66	$O(N^3)$	Contorno
CNSS	89.47	$O(N)$	Contorno
IDSC+DP	88.11	$O(KN^2)$	Contorno
TSS-41+GF	87.22	$O(SN)$	Região
TSS-41	86.20	$O(SN)$	Região
MCC	84.93	$O(N^3)$	Contorno
BAS-60	76.78	$O(N^3)$	Contorno
SC+DP (Shape Contexts)	76.51	$O(KN^2)$	Contorno
TSB+GF	75.20	$O(N)$	Região
Path Similarity	75.16	$O(N^3)$	Esqueleto
TSB	74.01	$O(N)$	Região
SSD+GF	71.00	$O(N^2)$	Contorno
TSDIZ	69.44	$O(CN^2)$	Contorno
HTSn	59.97	$O(N^2)$	Contorno
HTS	56.01	$O(N^2)$	Contorno
TSD	46.08	$O(N^2)$	Região

Figura 5.4: Curvas de Precisão vs. Revocação comparando TSS e TSB com diferentes descritores (MPEG-7).

Figura 5.5: Matriz representando a distância mútua entre cada par de imagens com os descritores TSS e TSB nas 70 classes da base MPEG-7 CE-Shape-1.

Figura 5.6: Amostra de imagens da base MNIST com pré-processamento para possibilitar o cálculo da TS.

Figura 5.7: Pré-processamento com interpolação bilinear e limiarização aplicado na base MNIST.

Figura 5.8: *Curvas Precisão vs. Revocação (base MNIST).*

Capítulo 6

Conclusão

Neste trabalho foram apresentados dois novos descritores de forma baseados no conceito de *tensor scale* (TS), com resultados que superam os métodos anteriores existentes também baseados em TS.

Os resultados apresentados demonstram que o descritor TSS desempenha melhor que o descritor TSB, já que inclui mais informações espaciais sobre a forma, ao considerar cada setor individualmente ao invés agrupá-los em bandas como no TSB. Portanto, há um compromisso entre a quantidade de informação espacial utilizada e o custo computacional da função de dissimilaridade do descritor resultante. Ainda, as características geométricas globais ao serem combinadas aos descritores são capazes de aumentar a capacidade de discriminação, orientando assim que são complementares às características dos descritores originais.

Os descritores TSS e TSB naturalmente resolvem alguns problemas típicos de métodos baseados em contornos, por serem abordagens de região. Por exemplo, métodos de contorno são geralmente mais sensíveis a deformações que podem mudar o comprimento relativo das partes do mesmo, geralmente ocasionados por erros de segmentação ou ruídos, pois pontos amostrados ao longo do contorno dificilmente terão boas correspondências. Os descritores TSS e TSB também exploram informações de buracos do interior da forma, tornando-os menos sensíveis às partes que se tocam, como por exemplo, é observado em diversas imagens da coleção MNIST nas quais alguns números apresentam um contorno fechado ([Figura 6.1](#)).

Os experimentos demonstram que o TSS atinge resultados comparáveis aos métodos estado da arte, que geralmente dependem de algoritmos custosos de otimização de correspondências, enquanto o TSB com sua função de distância com custo computacional linear demonstra ser uma solução mais eficiente para sistemas CBIR em grandes coleções. Além disso, os vetores de características são compactos, possuindo dimensões muito menores que o número de pixels da imagem, fator que também

Figura 6.1: *Imagens da mesma classe com variações da forma em partes que se tocam.*

contribui para que os descritores sejam empregados em grandes bases. Ambos os descritores podem ser utilizados em conjunto com técnicas de reclassificação (Bai *et al.*, 2010; Pedronette e Torres, 2010, 2012; Yang *et al.*, 2009) para melhorar os resultados de consultas CBIR e ainda, em abordagens com classificadores combinados.

Este trabalho resultou no artigo “*TSS & TSB: Tensor Scale Descriptors within Circular Sectors for Fast Shape Retrieval*” (Freitas *et al.*, 2016), publicado na revista *Pattern Recognition Letters*, na edição especial “*Efficient Shape Representation, Matching, Ranking, and its Applications*”.

6.1 Trabalhos Futuros

Como trabalho futuro pretende-se elaborar novos descritores baseados em *tensor scale*, combinando informações do esqueleto, região e contorno das formas. O objetivo de tal descritor seria aumentar a capacidade de discriminação em bases de objetos articulados, nas quais abordagens baseadas somente em região ou contorno tendem a ser limitadas. Outro aspecto a ser explorado é a extensão e validação dos métodos apresentados nesta dissertação em bases de formas 3D, como no caso da base Chang *et al.* (2015), da competição de SHREC¹.

Por um outro lado, os descritores de forma que foram o pilar de sistemas de reconhecimento de objetos (Dickinson e Pizlo, 2013), em sistemas de visão computacional, têm cedido espaço para abordagens que extraem padrões diretamente de imagens tons de cinza ou coloridas, sem necessidade de segmentação prévia, tais como *Bag-of-Visual-Words* (Csurka *et al.*, 2004; Sivic e Zisserman, 2003, 2009) ou por aprendizagem computacional com redes neurais profundas (*deep learning*), apresentando resultados promissores. Em vista dessa tendência, uma linha de trabalho futuro corresponde a estender os métodos desenvolvidos para atuarem diretamente sobre imagens coloridas e tons-de-cinza.

¹Shape Retrieval Contest: <http://www.shrec.net/> (Acesso em: 24 out. 2017)

Referências Bibliográficas

- Adamek e O'Connor (2004)** T. Adamek e N. E. O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):742–753. ISSN 1051-8215. DOI:10.1109/TCSVT.2004.826776. Citado na pág. 8, 32, 34, 38
- Andaló et al. (2010)** F. Andaló, P. A. V. Miranda, R. S. Torres, e A. X. Falcão. Shape feature extraction and description based on tensor scale. *Pattern Recognition*, 43(1):26–36. ISSN 0031-3203. DOI:10.1016/j.patcog.2009.06.012. Citado na pág. 3, 20, 21, 22, 23, 24, 34
- Andaló et al. (2007)** F. A. Andaló, P. A. V. Miranda, R. S. Torres, e A. X. Falcão. A new shape descriptor based on Tensor Scale. *Mathematical Morphology and its Applications to Signal and Image Processing (International Symposium on Mathematical Morphology)*, páginas 141–152. Citado na pág. 3, 17
- Andaló et al. (2007)** F. A. Andaló, P. A. V. Miranda, R. S. Torres, e A.X. Falcão. Detecting Contour Saliences using Tensor Scale. Em *IEEE International Conference on Image Processing (ICIP)*, volume 6, páginas 349–352. IEEE. ISBN 978-1-42-441436-9. DOI:10.1109/ICIP.2007.4379593. Citado na pág. 3, 19
- Andaló (2007)** Fernanda A. Andaló. Descritores de forma baseados em Tensor Scale. Dissertação de Mestrado, Instituto de Computação, Universidade de Campinas (IC-UNICAMP), Campinas, SP, Brasil. Citado na pág. 20
- Arica e Vural (2003)** N. Arica e F. T. Y. Vural. BAS: A Perceptual Shape Descriptor based on the Beam Angle Statistics. *Pattern Recognition*, 24(9-10):1627–1639. DOI:10.1016/S0167-8655(03)00002-3. Citado na pág. 2, 3, 7, 34
- Attneave (1954)** Fred Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193. DOI:10.1037/h0054663. Citado na pág. 20
- Bai e Latecki (2008)** Xiang Bai e Longjin J. Latecki. Path Similarity Skeleton Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1282–1292. ISSN 0162-8828. DOI:10.1109/TPAMI.2007.70769. Citado na pág. 2, 34
- Bai et al. (2010)** Xiang Bai, Xingwei Yang, Longin Jan Latecki, Wenyu Liu, e Zhuowen Tu. Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):861–874. ISSN 0162-8828. DOI:10.1109/TPAMI.2009.85. Citado na pág. 3, 46
- Bai et al. (2014)** Xiang Bai, Cong Rao, e Xinggang Wang. Shape Vocabulary: A Robust and Efficient Shape Representation for Shape Matching. *IEEE Transactions on Image Processing*, 23(9):3935–49. ISSN 1057-7149. DOI:10.1109/TIP.2014.2336542. Citado na pág. 3
- Bai et al. (2016)** Xiang Bai, Michael Donoser, Hairong Liu, e Longin Jan Latecki. Efficient shape representation, matching, ranking, and its applications. *Pattern Recognition Letters*, 83:241–242. ISSN 0167-8655. DOI:10.1016/j.patrec.2016.08.007. Citado na pág. 2

- Bay et al. (2008)** Herbert Bay, Andreas Ess, Tinne Tuytelaars, e Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359. ISSN 1077-3142. DOI:10.1016/j.cviu.2007.09.014. Citado na pág. 26
- Belongie et al. (2002)** S. Belongie, J. Malik, e J. Puzicha. Shape Matching and Object Recognition using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4): 509–522. ISSN 0162-8828. DOI:10.1109/34.993558. Citado na pág. 8, 25
- Bookstein (1978)** F.L. Bookstein. *The Measurement of Biological Shape and Shape Change*. Lecture Notes in Biomathematics. Springer Berlin Heidelberg. ISBN 978-3-642-93093-5. Citado na pág. 1
- Bovik et al. (1990)** Alan Conrad Bovik, Marianna Clark, e Wilson S. Geisler. Multichannel Texture Analysis Using Localized Spatial Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):55–73. ISSN 0162-8828. DOI:10.1109/34.41384. Citado na pág. 12
- Boyd e Vandenberghe (2004)** S.P. Boyd e L. Vandenberghe. *Convex Optimization*. Cambridge University Press. ISBN 978-0-521-83378-3. Citado na pág. 32
- Büttcher et al. (2010)** S. Büttcher, C.L.A. Clarke, e G.V. Cormack. *Information Retrieval: Implementing and Evaluating Search Engines*. MIT Press. ISBN 978-0-26-202651-2. Citado na pág. 35
- Chang et al. (2015)** Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, e Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. *CoRR*, abs/1512.03012. URL <http://arxiv.org/abs/1512.03012>. Citado na pág. 46
- Costa e Cesar Jr. (2009)** Luciano da Fona Costa e Roberto Marcondes Cesar Jr. *Shape Classification and Analysis: Theory and Practice, Second Edition*. Image Processing Series. CRC Press. ISBN 978-0-849-37940-6. DOI:10.1201/9780849379406. Citado na pág. 1, 5, 20, 27
- Csurka et al. (2004)** Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, e Cédric Bray. Visual categorization with bags of keypoints. *Workshop on Statistical Learning in Computer Vision, ECCV*, páginas 1–22. Citado na pág. 3, 46
- Di Ruberto (2004)** Cecilia Di Ruberto. Recognition of shapes by attributed skeletal graphs. *Pattern Recognition*, 37(1):21–31. ISSN 0031-3203. DOI:10.1016/j.patcog.2003.07.004. Citado na pág. 3
- Dickinson e Pizlo (2013)** S. Dickinson e Z. Pizlo. *Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective*. Advances in Computer Vision and Pattern Recognition. Springer London. ISBN 978-1-447-15195-1. Citado na pág. 46
- Diesendruck e Bloom (2015)** Gil Diesendruck e Paul Bloom. How specific is the shape bias? *Child development*, 74(1):168–178. ISSN 0009-3920. DOI:10.1111/1467-8624.00528. Citado na pág. 1
- Dorini e Leite (2010)** Leyza Baldo Dorini e Neucimar Jerônimo Leite. Teoria espaço-escala: abordagem morfológica. *Revista de Informática Teórica e Aplicada (RITA)*, 17(1):115–136. Citado na pág. 11
- Dryden e Mardia (1998)** I. L. Dryden e K. V. Mardia. *Statistical Shape Analysis*. Wiley Series in Probability & Statistics. Wiley. ISBN 978-0-47-195816-1. Citado na pág. 1, 27
- Dryden e Mardia (2016)** I.L. Dryden e K.V. Mardia. *Statistical Shape Analysis: With Applications in R*. Wiley Series in Probability and Statistics. Wiley. ISBN 978-0-470-69962-1. Citado na pág. 1

- Elder e Zucker (1998)** James H. Elder e Steven W. Zucker. Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7): 699–716. ISSN 0162-8828. DOI:10.1109/34.689301. Citado na pág. 12
- Falcão et al. (2004)** A.X. Falcão, J. Stolfi, e R. de Alencar Lotufo. The Image Foresting Transform: Theory, Algorithms, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29. ISSN 0162-8828. DOI:10.1109/TPAMI.2004.1261076. Citado na pág. 17, 20
- Freitas et al. (2016)** Anderson M. Freitas, Ricardo da S. Torres, e Paulo A.V. Miranda. TSS & TSB: Tensor scale descriptors within circular sectors for fast shape retrieval. *Pattern Recognition Letters*, 83, Part 3:303–311. ISSN 0167-8655. DOI:10.1016/j.patrec.2016.06.005. Special issue on “Efficient Shape Representation, Matching, Ranking, and its Applications”. Citado na pág. 34, 46
- Geiger et al. (2003)** D. Geiger, Tyng-Luh Liu, e R.V. Kohn. Representation and self-similarity of shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):86–99. ISSN 0162-8828. DOI:10.1109/TPAMI.2003.1159948. Citado na pág. 3
- Hart (1996)** John C. Hart. Sphere tracing: a geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer*, 12(10):527–545. ISSN 1432-2315. DOI:10.1007/s003710050084. Citado na pág. 17
- Jain e Vailaya (1998)** Anil K. Jain e Aditya Vailaya. Shape-Based Retrieval: A Case Study with Trademark Image Databases. *Pattern Recognition*, 31(9):1369–1390. ISSN 00313203. DOI:10.1016/S0031-3203(97)00131-3. Citado na pág. 2
- Karu e Jain (1996)** K. Karu e A. Jain. Fingerprint Classification. *Pattern Recognition*, 29(3): 389–404. DOI:10.1016/0031-3203(95)00106-9. Citado na pág. 2
- Kendall (1977)** D. G. Kendall. The Diffusion of Shape. *Advances in Applied Probability*, 9(3):428. ISSN 0001-8678. DOI:10.2307/1426091. Citado na pág. 1
- Kendall et al. (2009)** D.G. Kendall, D. Barden, T.K. Carne, e H. Le. *Shape and Shape Theory*. Wiley Series in Probability and Statistics. Wiley. ISBN 978-0-470-31784-6. Citado na pág. 1
- Kim e Kim (2000)** H. Kim e J. Kim. Region-based shape descriptor invariant to rotation, scale and translation. *Signal Processing: Image Communication*, 16(1-2):87–93. DOI:10.1016/S0923-5965(00)00018-7. Citado na pág. 2, 9, 36, 37
- Kimia (2002)** Benjamin B. Kimia. *Content-Based Retrieval of Images Based on Shape*, páginas 345–372. Wiley Interscience. ISBN 978-0-471-46407-5. Citado na pág. 5
- Koenderink (1984)** Jan J. Koenderink. The Structure of Images. *Biological Cybernetics*, 50(5): 363–370. ISSN 0340-1200. DOI:10.1007/BF00336961. Citado na pág. 11
- Korn et al. (1998)** P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, e Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *IEEE Transactions on Knowledge and Data Engineering*, 10(6):889–904. ISSN 1041-4347. DOI:10.1109/69.738356. Citado na pág. 2
- Landau et al. (1988)** Barbara Landau, Linda B. Smith, e Susan S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, 3(3):299–321. ISSN 08852014. DOI:10.1016/0885-2014(88)90014-7. Citado na pág. 1
- Latecki et al. (2003)** Longin Jan Latecki, Rolf Lakämper, e Diedrich Wolter. Shape Similarity and Visual Parts. Em *International Conference on Discrete Geometry for Computer Imagery (DGCI)*, volume 2886 of *Lecture Notes in Computer Science*, páginas 34–51. Springer. ISBN 978-3-54-039966-7. DOI:10.1007/978-3-540-39966-7_3. Citado na pág. 3, 5

- LeCun e Cortes (1998)** Yann LeCun e Corinna Cortes. The MNIST database of handwritten digits. URL <http://yann.lecun.com/exdb/mnist/>. Citado na pág. 37, 38
- Lindeberg (1993)** Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Springer. ISBN 978-1-44-195139-7. DOI:10.1007/978-1-4757-6465-9. Citado na pág. 11
- Ling e Jacobs (2007)** H. Ling e D.W. Jacobs. Shape Classification Using the Inner-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299. ISSN 0162-8828. DOI:10.1109/TPAMI.2007.41. Citado na pág. 2, 8, 34, 38
- Loncaric (1998)** Sven Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001. ISSN 0031-3203. DOI:10.1016/S0031-2023(97)00122-2. Citado na pág. 5
- Madabhushi et al. (2006)** Anant Madabhushi, Jayaram K. Udupa, e Andre Souza. Generalized scale: Theory, algorithms, and application to image inhomogeneity correction. *Computer Vision and Image Understanding*, 101(2):100–121. ISSN 1077-3142. DOI:10.1016/j.cviu.2005.07.010. Citado na pág. 12, 13
- Mallat (1989)** S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693. ISSN 0162-8828. DOI:10.1109/34.192463. Citado na pág. 12
- Mardia e Jupp (1999)** K. V. Mardia e P. E. Jupp. *Directional Statistics*. John Wiley & Sons. Citado na pág. 21, 25
- Marques e Furht (2012)** O. Marques e B. Furht. *Content-Based Image and Video Retrieval*. Multimedia Systems and Applications. Springer US. ISBN 978-1-461-50987-5. Citado na pág. 5
- Marr (1982)** David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press. ISBN 978-0-26-251462-0. DOI:10.7551/mitpress/9780262514620.001.0001. Citado na pág. 11
- Miranda et al. (2005)** P. A. V. Miranda, R. S. Torres, e A. X. Falcão. TSD: A Shape Descriptor Based on a Distribution of Tensor Scale Local Orientation. *XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, páginas 139–146. ISSN 1530-1834. DOI:10.1109/SIBGRAPI.2005.51. Citado na pág. 3, 16, 17, 19, 20, 34
- Ng et al. (2014)** Bernard Ng, Matthew Toews, Stanley Durrleman, e Yonggang Shi. *Shape Analysis for Brain Structures*, páginas 3–49. Springer International Publishing. ISBN 978-3-319-03813-1. DOI:10.1007/978-3-319-03813-1_1. Citado na pág. 1
- Oliveira et al. (2015)** Alessandro Bof Oliveira, Patrícia Rodrigues Silva, e Dante Augusto Couto Barone. A novel 2D shape signature method based on complex network spectrum. *Pattern Recognition Letters*, 63:43–49. ISSN 0167-8655. DOI:10.1016/j.patrec.2015.05.018. Citado na pág. 4, 9, 34
- Pedronette e Torres (2010)** Daniel Carlos Guimarães Pedronette e Ricardo da S. Torres. Exploiting Contextual Information for Image Re-ranking. Em *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 15th Iberoamerican Congress on Pattern Recognition (CIARP)*, volume 6419 of *Lecture Notes in Computer Science*, páginas 541–548. Springer. ISBN 978-3-64-216687-7. DOI:10.1007/978-3-642-16687-7_71. Citado na pág. 3, 46
- Pedronette e Torres (2012)** Daniel Carlos Guimarães Pedronette e Ricardo da S. Torres. Exploiting pairwise recommendation and clustering strategies for image re-ranking. *Information Sciences*, 207:19–34. ISSN 0020-0255. DOI:10.1016/j.ins.2012.04.032. Citado na pág. 3, 46

- Pedrosa et al. (2016)** Glauco V. Pedrosa, Agma J M Traina, e Celia A Z Barcelos. Retrieving 2D shapes by similarity based on bag of salience points. *Multimedia Tools and Applications*, páginas 1–15. ISSN 1573-7721. DOI:10.1007/s11042-016-4046-2. Citado na pág. 32
- Pedrosa et al. (2011)** G.V. Pedrosa, C. Barcelos, e M.A. Batista. A shape contour descriptor based on salience points. Em *IEEE International Conference on Image Processing (ICIP)*, páginas 1049–1052. DOI:10.1109/ICIP.2011.6115604. Citado na pág. 19
- Pedrosa et al. (2013)** G.V. Pedrosa, C. Barcelos, e M.A. Batista. Image feature descriptor based on shape salience points. *Neurocomputing*, 120:156–163. ISSN 0925-2312. DOI:10.1016/j.neucom.2012.07.055. Citado na pág. 8, 32, 34, 38
- Pizer et al. (1998)** Stephen M Pizer, David Eberly, Daniel S. Fritsch, e Bryan S Morse. Zoom-Invariant Vision of Figural Shape: The Mathematics of Cores. *Computer Vision and Image Understanding*, 69(1):55–71. ISSN 1077-3142. DOI:10.1006/cviu.1997.0563. Citado na pág. 12
- Saha e Udupa (2003)** P. K. Saha e J. K. Udupa. Tensor scale-based fuzzy connectedness image segmentation. Em *SPIE Medical Imaging 2003: Image Processing*, volume 5032, páginas 1580–1590. DOI:10.1117/12.481882. Citado na pág. 3, 15
- Saha (2003)** Punam K. Saha. Novel theory and methods for tensor scale: a local morphometric parameter. *Medical Imaging 2003: Image Processing*, 5032:743–753. DOI:10.1117/12.480645. Citado na pág. 2, 3, 4, 13, 14, 16
- Saha (2005)** Punam K. Saha. Tensor scale: A local morphometric parameter with applications to computer vision and image processing. *Computer Vision and Image Understanding*, 99(3):384–413. ISSN 1077-3142. DOI:10.1016/j.cviu.2005.03.003. Citado na pág. 2, 3, 4, 11, 12, 13, 14, 15, 16, 17
- Saha e Udupa (2001)** Punam K. Saha e Jayaram K. Udupa. Scale-based diffusive image filtering preserving boundary sharpness and fine structures. *IEEE Transactions on Medical Imaging*, 20(11):1140–1155. ISSN 0278-0062. DOI:10.1109/42.963817. Citado na pág. 12
- Saha e Wehrli (2004)** Punam K. Saha e Felix W. Wehrli. A robust method for measuring trabecular bone orientation anisotropy at in vivo resolution using tensor scale. *Pattern Recognition*, 37(9):1935–1944. ISSN 0031-3203. DOI:10.1016/j.patcog.2003.12.019. Citado na pág. 3
- Saha e Xu (2010)** Punam K. Saha e Ziyue Xu. An Analytic Approach to Tensor Scale with an Efficient Algorithm and Applications to Image Filtering. Em *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, páginas 429–434. IEEE. ISBN 978-1-42-448816-2. DOI:10.1109/DICTA.2010.79. Citado na pág. 3
- Saha et al. (2000)** Punam K. Saha, Jayaram K. Udupa, e Dewey Odhner. Scale-Based Fuzzy Connected Image Segmentation: Theory, Algorithms, and Validation. *Computer Vision and Image Understanding*, 77(2):145–174. ISSN 1077-3142. DOI:10.1006/cviu.1999.0813. Citado na pág. 12
- Saha et al. (2003)** Punam K. Saha, Hui Zhang, Jayaram K. Udupa, e James C. Gee. Tensor scale-based image registration. Em *SPIE Medical Imaging 2003: Image Processing*, volume 5032, páginas 314–324. DOI:10.1117/12.481433. Citado na pág. 3, 15
- Sebastian et al. (2004)** T.B. Sebastian, P.N. Klein, e B.B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571. ISSN 0162-8828. DOI:10.1109/TPAMI.2004.1273924. Citado na pág. 3
- Sebastian e Kimia (2005)** Thomas B. Sebastian e Benjamin B. Kimia. Curves vs. skeletons in object recognition. *Signal Processing*, 85(2):247–263. ISSN 0165-1684. DOI:10.1016/j.sigpro.2004.10.016. Citado na pág. 3

- Shekar et al. (2015)** B. H. Shekar, Bharathi Pilar, e Josef Kittler. An unification of Inner Distance Shape Context and Local Binary Pattern for Shape Representation and Classification. *Proceedings of the 2nd International Conference on Perception and Machine Intelligence - PerMin '15*, páginas 46–55. DOI:10.1145/2708463.2709062. Citado na pág. 3
- Sivic e Zisserman (2003)** Josef Sivic e Andrew Zisserman. Video Google: a text retrieval approach to object matching in videos. Em *IEEE International Conference on Computer Vision (ICCV)*, volume 2, páginas 1470–1477. IEEE. ISBN 0-769-51950-4. DOI:10.1109/ICCV.2003.1238663. Citado na pág. 3, 46
- Sivic e Zisserman (2009)** Josef Sivic e Andrew Zisserman. Efficient Visual Search of Videos Cast as Text Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4): 591–606. ISSN 0162-8828. DOI:10.1109/TPAMI.2008.111. Citado na pág. 3, 46
- Souza et al. (2005)** Andre Souza, Jayaram K. Udupa, e Punam K. Saha. Volume rendering in the presence of partial volume effects. *IEEE Transactions on Medical Imaging*, 24(2):223–235. ISSN 0278-0062. DOI:10.1109/TMI.2004.840295. Citado na pág. 3
- Souza et al. (2008)** Andre Souza, Jayaram K. Udupa, e Anant Madabhushi. Image filtering via generalized scale. *Medical Image Analysis*, 12(2):87–98. ISSN 1361-8415. DOI:10.1016/j.media.2007.07.007. Citado na pág. 13
- Souza e Marana (2014)** Gustavo B. Souza e A. Nilceu Marana. HTS and HTSn: New shape descriptors based on Hough Transform Statistics. *Computer Vision and Image Understanding*, 127:43–56. ISSN 1077-3142. DOI:10.1016/j.cviu.2014.06.010. Citado na pág. 2, 8, 34, 38
- Tabb e Ahuja (1997)** Mark Tabb e Narendra Ahuja. Multiscale image segmentation by integrated edge and region detection. *IEEE Transactions on Image Processing*, 6(5):642–655. ISSN 1057-7149. DOI:10.1109/83.568922. Citado na pág. 12
- Torres e Falcão (2007)** R. da S. Torres e A.X. Falcão. Contour salience descriptors for effective image retrieval and analysis. *Image and Vision Computing*, 25(1):3–13. ISSN 0262-8856. DOI:10.1016/j.imavis.2005.12.010. Citado na pág. 19, 34
- Torres et al. (2006)** Ricardo da S. Torres, Claudia Bauzer Medeiros, Marcos André Gonçalves, e Edward A. Fox. A Digital Library Framework for Biodiversity Information Systems. *International Journal on Digital Libraries*, 6(1):3–17. ISSN 1432-5012. DOI:10.1007/s00799-005-0124-1. Citado na pág. 2
- Torres e Falcão (2006)** Ricardo S. Torres e Alexandre X. Falcão. Content-Based Image Retrieval: Theory and Applications. *Revista de Informática Teórica e Aplicada (RITA)*, 13(2):161–185. Citado na pág. 2, 6, 7
- Torres e Falcão (2008)** Ricardo S. Torres e Alexandre X. Falcão. Recuperação de Imagens Baseada em Conteúdo. Em Aparecido Nilceu Marana e José Remo Ferreira Brega, editores, *Técnicas e Ferramentas de Processamento de Imagens Digitais e Aplicações em Realidade Virtual e Misturada*, capítulo 4, páginas 111–132. Editora Canal 6. ISBN 978-8-599-72846-8. Citado na pág. 6, 35, 36
- Torres et al. (2003)** R.S. Torres, E.M. Picado, A.X. Falcão, e L.F. Costa. Effective Image Retrieval by Shape Salience. Em *XVI Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, páginas 167–174. IEEE. DOI:10.1109/SIBGRA.2003.1241005. Citado na pág. 19
- Traver e Pla (2003)** V. Javier Traver e Filiberto Pla. Designing the Lattice for Log-Polar Images. Em *International Conference on Discrete Geometry for Computer Imagery (DGCI)*, volume 2886 of *Lecture Notes in Computer Science*, páginas 164–173. Springer. ISBN 978-3-54-039966-7. DOI:10.1007/978-3-540-39966-7_15. Citado na pág. 27

- Veltkamp e Latecki (2006)** Remco C. Veltkamp e Longin Jan Latecki. Properties and Performance of Shape Similarity Measures. páginas 47–56. Springer. ISBN 978-3-54-034415-5. DOI:10.1007/3-540-34416-0_6. Citado na pág. 6, 7, 37
- Wang et al. (2012)** Junwei Wang, Xiang Bai, Xinge You, Wenyu Liu, e Longin Jan Latecki. Shape matching and classification using height functions. *Pattern Recognition Letters*, 33(2):134–143. ISSN 0167-8655. DOI:10.1016/j.patrec.2011.09.042. Citado na pág. 2, 3, 9, 34, 36
- Wang e Pavlidis (1990)** Y.P. Wang e T. Pavlidis. Optimal correspondence of string subsequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1080–1087. ISSN 0162-8828. DOI:10.1109/34.61707. Citado na pág. 8
- Wei et al. (2009)** C.H. Wei, Y. Li, W.Y. Chau, e C.T. Li. Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42(3): 386–394. ISSN 0031-3203. DOI:10.1016/j.patcog.2008.08.019. Citado na pág. 2
- Witkin (1984)** A. Witkin. Scale-space filtering: A new approach to multi-scale description. Em *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 9, páginas 150–153. IEEE. DOI:10.1109/ICASSP.1984.1172729. Citado na pág. 11
- Xu et al. (2009)** Ziyue Xu, Milan Sonka, e Punam K. Saha. Recent improvements in tensor scale computation and its applications to medical imaging. Em *SPIE Medical Imaging 2009: Image Processing*, volume 7259. DOI:10.1117/12.811360. Citado na pág. 3
- Xu et al. (2011)** Ziyue Xu, Milan Sonka, e Punam K Saha. Improved tensor scale computation with application to medical image interpolation. *Computerized Medical Imaging and Graphics*, 35 (1):64–80. ISSN 1879-0771. DOI:10.1016/j.compmedimag.2010.09.007. Citado na pág. 3
- Xu et al. (2012)** Ziyue Xu, P. K. Saha, e S. Dasgupta. Tensor scale: An analytic approach with efficient computation and applications. *Computer Vision and Image Understanding*, 116(10): 1060–1075. ISSN 1077-3142. DOI:10.1016/j.cviu.2012.05.006. Citado na pág. 3, 11, 13, 16
- Yang et al. (2009)** Xingwei Yang, S. Koknar-Tezel, e L.J. Latecki. Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. Em *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 357–364. ISBN 978-1-42-443992-8. DOI:10.1109/CVPR.2009.5206844. Citado na pág. 3, 46
- Zelditch et al. (2012)** M.L. Zelditch, D.L. Swiderski, e H.D. Sheets. *Geometric Morphometrics for Biologists: A Primer*. Elsevier Science. ISBN 978-0-123-86904-3. Citado na pág. 1
- Zhang e Lu (2004)** Dengsheng Zhang e Guojun Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19. ISSN 0031-3203. DOI:10.1016/j.patcog.2003.07.008. Citado na pág. 5
- Ziezold (1977)** Herbert Ziezold. *On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces*, páginas 591–602. Springer Netherlands, Dordrecht. ISBN 978-94-010-9910-3. DOI:10.1007/978-94-010-9910-3_63. Citado na pág. 1

Índice Remissivo

Bases de imagens

MNIST, [38](#)

MPEG-7, [37](#)

Características Geométricas Globais, [32](#)

CBIR, *veja Content-Based Image Retrieval*

Content-Based Image Retrieval, [1](#)

Descritor

definição, [6](#)

BAS, [7](#)

HTS, [8](#)

HTSn, [8](#)

IDSC-DP, [8](#)

MCC, [8](#)

MLEV, [9](#)

SSD+GF, [8](#)

TSB, [28](#)

TSB+GF, [32](#)

TSCS, [19](#)

TSD, [19](#)

TSDIZ, [21](#)

TSS, [25](#)

TSS+GF, [32](#)

Descritores

por contorno, [7](#)

por região, [9](#)

Dissimilaridade

definição, [6](#)

TSS, [27](#)

Distância, *veja* dissimilaridade

EDT

transformada, *veja* transformada de distância

Euclidiana

Escala, [11](#)

Euclidiana

distancia, [53](#)

transformada, *veja* transformada de distância

Euclidiana

Medidas

BEP, [36](#)

Bullseye, *veja* BEP

Precisão, [35](#)

Precisão vs. Revocação, [35](#)

Revocação, [35](#)

Taxa de acerto 1-NN, [36](#)