



**UNIVERSIDADE FEDERAL DO CEARÁ
ENGENHARIA DE COMPUTAÇÃO**

ISAAC NEWTON MELO MACHADO

**UM NOVO MÉTODO DE RECONHECIMENTO DE FORMAS BASEADO
EM ESTATÍSTICAS DE DISTÂNCIAS EUCLIDIANAS**

**SOBRAL - CE
JUNHO - 2018**

ISAAC NEWTON MELO MACHADO

Um Novo Método de Reconhecimento de Formas Baseado em Estatísticas de Distâncias Euclidianas

Trabalho de Conclusão de Curso submetido à coordenação do curso de Engenharia de Computação da Universidade Federal do Ceará, como requisito para a obtenção do grau de Engenheiro de Computação.

Orientador:

Jarbas Joaci de Mesquita Sá Junior

Sobral - CE

Junho - 2018

RESUMO

A análise de formas é um campo importante da visão computacional, e ao longo dos anos, muitos métodos têm sido elaborados para caracterização de formas. Para contribuir a esta linha de pesquisa, o presente trabalho propõe um novo descritor de forma baseado em estatísticas de distâncias euclidianas. Para isso, a técnica calcula as distâncias euclidianas entre pontos do contorno da forma e armazena-as em uma matriz, que por sua vez é subdividida em submatrizes. Para cada submatriz, o algoritmo calcula valores como média, desvio padrão e distância mínima para cada coluna. Finalmente, os atributos são calculados a partir desses dados, a fim de construir a assinatura da forma.

ABSTRACT

Shape analysis is an essential branch of computer vision and until now several techniques have been suggested for shape characterization. To contribute to this field, this work introduces a novel shape descriptor based on statistics of Euclidean distances. The method consists in computing Euclidean distances between all points of the shape's contour and filling up a matrix, which is subdivided into submatrices. For every submatrix, the algorithm calculates values such as mean, standard deviation and minimum distance in each column. Lastly, a shape signature is built from these statistics.

AGRADECIMENTOS

Em primeiro lugar, agradeço a minha mãe, Marilene de Carvalho Melo, por sua força em propiciar tudo que foi necessário para minha formação.

Agradeço à Universidade Federal do Ceará pelo apoio financeiro com a manutenção da bolsa do Programa de Educação Tutorial durante o período do curso.

Ao meu orientador Jarbas Joaci de Mesquita Sá Junior, por sua paciência, orientação e ensinamentos que pude adquirir ao longo de quase três anos.

A alguns professores do curso, que transmitiram uma parcela de seu conhecimento com amizade, atenção e solidariedade. Uma salva especial a Iális Cavalcante de Paula Júnior, que tanto contribui ao desenvolvimento dos alunos e do curso.

Aos alunos Francisco Evangelista Nobre Filho e Antonio Márcio Albuquerque Almeida, pela participação na publicação do artigo e neste trabalho.

Por fim, a todos meus amigos durante a graduação, sem eles eu não teria conseguido. Aos amigos da minha turma, Magdiel Campelo Alves de Sousa, Eloise de Carvalho Rodrigues, Syllas Rangel Carneiro Magalhães, Isaac Ben Judá Castelo Branco de Senna Dourado e Laercio Santana Rodrigues, meus agradecimentos.

A minha mãe, Marilene Melo, e minha avó, Maria do Rosário.

*"Se cheguei até aqui foi porque me
apoiiei no ombro dos gigantes."
(Isaac Newton)*

Sumário

1	Introdução	5
1.1	Introdução	5
1.2	Trabalhos Relacionados	5
1.3	Motivação e Objetivos	6
1.4	Considerações	7
2	Fundamentação Teórica	8
2.1	Visão Computacional	8
2.2	Sistema de Visão Computacional	9
2.2.1	Aquisição	9
2.2.2	Processamento	9
2.2.3	Segmentação	10
2.2.4	Extração de Características	11
2.2.5	Reconhecimento de Padrões	13
3	Método Proposto	17
4	Metodologia	19
5	Avaliação de Parâmetros e Resultados	22
5.1	Avaliação de parâmetros	22
5.2	Comparação dos resultados	23
5.3	Análise de complexidade	24
6	Conclusões e Perspectivas	25
6.1	Conclusões	25
6.2	Perspectivas	25
	Referências Bibliográficas	26

Lista de Figuras

2.1	Passos fundamentais para o processamento digital de imagens [3].	8
2.2	Da esquerda para a direita, os datasets MPEG7, Robotic, Box e Surface [21].	9
4.1	Exemplo das classes do banco de imagens <i>generic shapes</i> [12,13].	19
4.2	Imagens do banco <i>generic shapes</i> com oclusão e rotação [12,13].	19
4.3	Exemplo das classes do banco de imagens <i>fish</i> [14].	20
4.4	Fluxograma da etapas do sistema de visão computacional utilizado como metodologia.	20

Lista de Tabelas

5.1	Acurácia para alguns valores T	22
5.2	Avaliação dos descritores. Os resultados dos métodos de comparação foram obtidos nos trabalhos [2, 18].	23

LISTA DE TRABALHOS PUBLICADOS

- I. N. M. MACHADO; F. E. NOBRE FILHO; J. J. M. SA JUNIOR *A novel method for shape analysis based on statistics of Euclidean distances*, **WVC2017**, NATAL, RN, Sociedade Brasileira de Computação, (2017).

Capítulo 1

Introdução

1.1 Introdução

Os objetos possuem características inerentes a sua aparência e definição, sendo estas a cor, textura e forma, onde a última dependente apenas da existência do objeto, ou seja, ela é invariante em relação a alterações na iluminação e textura, por exemplo. Referindo-se a reconhecimento visual, a forma é um atributo que permite a identificação dos objetos de forma simples e objetiva.

Dada a necessidade da identificação de objetos, diversos métodos de análise foram propostos. Na área de visão computacional, os problemas envolvendo classificação e segmentação de formas são antigos e ao passar do tempo as soluções estão cada vez mais robustas e discriminativas. Elas se dividem em três grupos: baseados em contorno, região e esqueleto [1]. Basicamente, suas diferenças se encontram no local de extração de características [2]. Os métodos baseados em contorno consideram a imagem como um conjunto de pontos ordenados e conectados por segmentos de reta. Nesses métodos, como esperado, pontos não interligados podem gerar alterações nos resultados. Como exemplos dessa categoria, podemos citar os descritores de Fourier [3, 5] e *Curvature Scale Space* (CSS) [6]. Já os métodos baseados em região, que são amplamente utilizados em formas genéricas, utilizam agrupamentos de *pixels* semelhantes para gerar descritores [7]. Os momentos de Zernike [8, 9] são um exemplo. Por fim, as técnicas baseadas em esqueleto geralmente modelam as formas por meio de grafos ou dimensões fractais, utilizando as informações dos eixos médios do objeto. O método de “similaridade de caminho” é um exemplo dessa categoria [10, 11]. As assinaturas desses métodos são adequadas para imagens que possuem oclusão ou articulação [1, 2].

1.2 Trabalhos Relacionados

A proposta para a análise de formas, presente em imagens, é apresentada em diversos trabalhos e, ao longo dos anos, muitos métodos foram propostos para a caracterização da forma. Entre esses métodos, podemos citar o proposto no trabalho de Backes e Bruno [1].

Esse trabalho se caracteriza pela aplicação de uma abordagem complexa baseada em rede neural, *Small World*, que produz uma descrição da forma de contorno, presente em uma imagem. Para analisar os dados de contornos extraídos pela rede neural, foi utilizado um classificador linear chamado de *Linear Discriminant Analysis* (LDA). A ideia ao aplicar LDA é encontrar uma projeção dos dados onde a variação entre as classes é maximizada em comparação com a variação dentro das classes, que é minimizada. A avaliação é feita usando análise estatística no descritor proposto pelos autores usando o LDA em um esquema de validação cruzada *leave-one-out*. Ao longo dos experimentos, no intuito de avaliar o resultado do projeto, foram utilizados bancos de dados de imagens de contorno (*generic shapes* e *fish*), nas quais foram obtidos os resultados de 95,96% e 99,38%, respectivamente.

Outro estudo relacionado, o qual serve de base para este trabalho, está descrito no artigo *Shape classification using line segment* [2] dos autores Sá Junior e Backes. De forma semelhante ao experimento anteriormente citado, utilizando para avaliação o LDA em combinação ao *leave-one-out*, o artigo propôs um descritor que utiliza o contorno da forma, no qual a assinatura é obtida a partir da média e do desvio padrão da distância entre os pontos do contorno. No mesmo, foi atingida uma acurácia de 96,96% e 99,40% nos bancos de imagens *generic shapes* e *fish*, respectivamente.

1.3 Motivação e Objetivos

A extração de características é a base da recuperação de informação visual e as características de forma podem reconhecer muitos objetos e pertencerem a muitas aplicações, como por exemplo a recuperação e indexação imagens, aplicações na biomedicina, indústria, no agronegócio, em veículos autônomos, etc. Para isso, os algoritmos responsáveis pela extração de características devem ter alta capacidade de discriminação, baixo custo computacional e, levando em conta as formas, esses descritores devem ser invariantes a escala, rotação, translação, oclusão, iluminação ou qualquer outro ruído que possa prejudicar a recuperação de informação. Logo, visando o aperfeiçoamento dos sistemas de visão computacional, o desenvolvimento de descritores cada vez mais robustos possibilitará o surgimento de mais aplicações, assim como aumentará a confiabilidade das existentes.

Considerando a introdução na Seção 1.1 e a motivação supracitada, o presente trabalho propõe um descritor de formas que utiliza características estatísticas (média, desvio padrão e distância mínima) das distâncias relativas dos pontos de contorno entre si. Dessa forma, as propriedades invariantes a escala, rotação e translação são mantidas. Além disso, objetiva-se que o método apresente alta capacidade discriminativa, e haja vista a complexidade de alguns algoritmos, seja de simples entendimento. Para tal fim, o descritor proposto será reproduzido e comparado aos demais resultados encontrados nos artigos apresentados na Seção 1.2.

1.4 Considerações

Este trabalho está dividido do seguinte modo: o Capítulo 2 traz uma fundamentação teórica da área de visão computacional, com destaque a Subseção 2.2.4 de descritores, o Capítulo 3 descreve o método proposto, o Capítulo 4 apresenta a metodologia do experimento, O Capítulo 5 exhibe os resultados obtidos e, por fim, o Capítulo 6 a apresenta a conclusão do trabalho e futuros aprimoramentos.

Capítulo 2

Fundamentação Teórica

2.1 Visão Computacional

A visão computacional é o processo de modelagem e replicação da visão humana usando software e hardware. Em ciências da computação, a visão computacional desenvolve técnicas matemáticas para recuperar a forma tridimensional e aparência dos objetos em imagens [19]. Essas técnicas fazem parte da base de conhecimento do processamento de imagem, a qual é dividida em duas amplas categorias: técnicas nas quais a entrada e saída são imagens e técnicas nas quais a imagem de entrada gera atributos como saída [3].

Na primeira categoria, geralmente denominada de pré-processamento, encontramos as etapas de aquisição, filtragem e realce, restauração, processamento de imagens coloridas, *wavelets* e multirresolução, compressão e processamento morfológico. A segunda, a qual converte a imagem em informação, é a parte de visão computacional propriamente dita e é subdividida em três etapas: segmentação, descrição e reconhecimento de objetos.

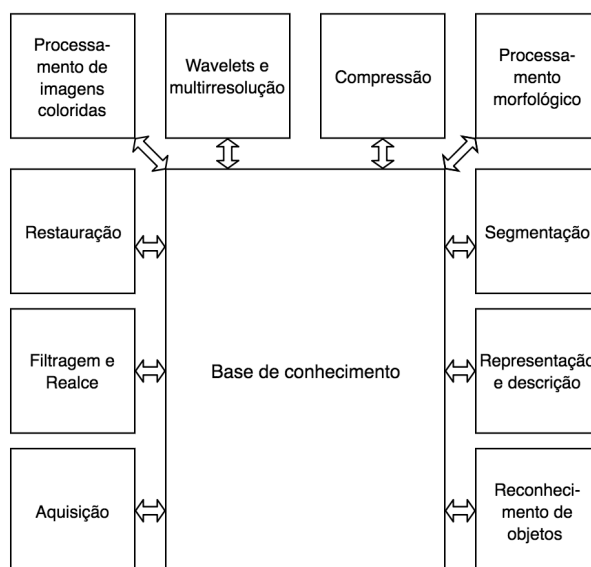


Figura 2.1: Passos fundamentais para o processamento digital de imagens [3].

Todos essas etapas, ver Figura 2.1, são fundamentais ao processamento de imagens com

a finalidade de recuperar informações e, geralmente, são executados de forma sequencial. Entretanto, a exemplo o presente trabalho, nem todos os procedimentos utilizam todos esses passos ou etapas, por isso vamos definir o sistema de visão computacional de forma mais simples e curta, sendo este dividido em : aquisição, processamento de imagem, segmentação, extração de características e reconhecimento de padrões. Essa sistemática será detalhada, em seguida, e usada como base da metodologia do atual trabalho.

2.2 Sistema de Visão Computacional

2.2.1 Aquisição

A aquisição é responsável pela captação da imagens, ou seja tenta simular a função dos olhos [4]. Ela pode ser tão simples quanto receber uma imagem que já esteja em formato digital ou , em geral, ser obtida a partir de *scanners*, filmadoras, máquinas fotográficas, *smartphones*, etc.

O processo de aquisição pode ser custoso e demandar muito tempo, além disso, pode se fazer necessário uma preocupação maior com a etapa subsequente de processamento, pois o processo de aquisição é suscetível a ruídos. Por esse motivo, alguns trabalhos da literatura utilizam banco de imagens adquiridos e até mesmo processados, ou seja, já com seus ruídos reduzidos. Esses bancos de imagens são chamados de *benchmark datasets*, pois em adição a facilidade de obtê-los os autores podem comparar a eficiência de seus métodos utilizando do mesmo conjunto de entrada. Segue alguns exemplos desses bancos de imagens na Figura 2.2.



Figura 2.2: Da esquerda para a direita, os datasets MPEG7, Robotic, Box e Surface [21].

2.2.2 Processamento

O processamento de imagem ou comumente chamado "pré-processamento" é a etapa que compreende desde operações básicas como rotação, escalonamento e recorte a processos mais robustos como filtragem, transformadas, equalizações e operações morfológicas com o objetivo de remover os ruídos e realçar as características das imagens. Esses métodos de processamento podem ser divididos em três áreas: operações básicas, filtragem e morfologia matemática.

Operações Básicas

Neste grupo de operações destacamos os métodos simples que agem no domínio do espaço (rotação, escalonamento, recorte e translação) e que atuam sobre a intensidade dos pixel das imagem (operações ponto a ponto, contraste e equalização de histograma).

Filtragem

A filtragem espacial se fundamenta em uma operação de convolução de uma máscara (*mask*, *kernel* ou *template*) e da imagem digital considerada. A máscara é um arranjo matricial de dimensões inferiores às da imagem a ser filtrada e, em geral, quadrada, cujos valores são definidos como fatores de ponderação (pesos) a serem aplicados sobre *pixels* da imagem. A operação é executada progressivamente sobre toda a imagem, coluna a coluna e linha a linha.

Dentre os filtros mais comuns utilizados em processamento digital de imagens encontram-se os da média, da mediana e da moda [3], todos destinados à suavização da imagem. Esses filtros atenuam variações abruptas nos níveis de cinza da imagem, o que possibilita sua aplicação à redução de ruído de origens diversas.

Morfologia Matemática

A morfologia matemática concentra seu esforço no estudo de estruturas geométricas presentes em uma imagem através de ferramentas matemáticas. A partir do elemento estruturante, a morfologia matemática objetiva extrair informações relativas à geometria e à topologia de conjuntos desconhecidos de uma imagem. Com esse objetivo e utilizando as operações básicas de dilatação, erosão, abertura e fechamento a mesma elabora algoritmos capazes de extrair fronteiras, preencher regiões, extrair componentes conectados e sobre a forma causar afinamento, espessamento ou esqueletização.

2.2.3 Segmentação

É um processo que se caracteriza com a subdivisão de uma imagem em regiões ou objetos que a compõem. Esta etapa de subdivisão pode ser implementada para a segmentação de cores, texturas, pontos, linhas, descontinuidades, bordas, entre outros detalhes nas imagens. Esse processo varia conforme o problema. Tratando-se da amostra de parte do conteúdo de uma imagem, como por exemplo: um objeto, um rosto, parte do corpo, dependendo dessas aplicações, é interessante utilizar técnicas para remoção de conteúdos dispensáveis presentes na imagem, destacando assim somente o objeto ou região de interesse. Uma das mais comuns abordagens é a divisão de uma imagem com base nas mudanças bruscas de intensidade, como as bordas [3]. Além disso, podemos citar a detecção de pontos e linhas, limiarização e segmentação baseado em região.

2.2.4 Extração de Características

A extração de características junto do reconhecimento de padrões são as principais etapas do processo de visão computacional. Nesse passo, a imagem segmentada passará por um algoritmo que visa extrair alguma informação quantitativa, onde seja básica para discriminação entre classes de objetos [3] e de baixa dimensionalidade, para que o processo seguinte de reconhecimento tenha uma melhor precisão e performance, respectivamente.

O conjunto dessas informações obtidas é denominada de vetor de atributos ou assinatura da imagem e o algoritmos capazes de extrair essas características são chamados de descritores. Quanto a esses descritores, eles podem ser divididos em dois tipos: descritores de formas e descritores de textura e cor.

Descritores de Forma

Como explicado anteriormente, é desejável que a assinatura dos objetos sejam discriminativas. Em se tratando de forma, por exemplo, esses atributos podem ser expressados de várias maneiras, como pelo cálculo da área, contorno, quantidade de buracos, simetria, diâmetro etc. Assim, muitos desses objetos podem ser classificados por esses descritores utilizando dessas ou demais individualidades.

Em geral, esses descritores de forma são classificados quanto as informações que recuperam, sendo esses descritores: geométricos, os quais extraem informações como área, perímetro, compacidade, simetria etc; de Fourier e Wavelet, descritores que extraem informação das transformadas da imagem; de complexidade, que traz dados relacionados a irregularidade da ocupação do espaço e distribuição dos *pixels*; e de contorno, que utilizam apenas os *pixels* do contorno da forma.

Segue o exemplo de alguns descritores de forma encontrados na literatura.

- **Descritores de Fourier** - Os descritores de Fourier são formados pelos coeficientes da transformada discreta de Fourier (DFT). O descritor de Fourier tem como característica expressar as informações globais da curvatura do objeto extraído da imagem. A transformada discreta de Fourier basea-se no teorema de Fourier que tem como princípio que toda função contínua pode ser decomposta em uma função de frequências de senos e cossenos. Os dados de entrada de uma DFT são sequências de números discretos ou contínuos [3].
- **Momentos de Zernike** [9] - Os momentos de Zernike são baseados em polinômios ortogonais sobre um disco na imagem. Tais momentos são invariantes a rotação e podem também ser invariantes em relação às transformações de escala e translação se a imagem de entrada passar por uma manipulação prévia baseada na teoria dos momentos geométricos.
- **Momentos de Hu** - Esse descritor considera os valores individuais dos *pixels* como uma função densidade de probabilidade, dessa forma é possível calcular os momentos

estatísticos de uma imagem. Sendo estes momentos, características que descrevem a distribuição espacial dos pontos. A equação a seguir mostra como calcular esses momentos.

$$\eta_{ij} = \sum_x \sum_y x^i y^j I(x, y) \quad (2.1)$$

Em que, η_{ij} representa o i -ésimo momento para o eixo x e j -ésimo momento para o eixo y ; x e y são índices correspondentes a cada pixel; e $I(x, y)$ indica o valor, em escala de cinza, do pixel correspondente.

Para formar a assinatura da imagem, Ming-Kuei Hu, propôs em *Visual pattern recognition by moment invariants* [7] um conjunto de equações que, a partir dos momentos de uma imagem, obtêm valores que são invariantes em relação à rotação, escala e translação. Estas equações são conhecidas como momentos de Hu ou momentos invariantes de uma imagem.

- **Descritores de Curvatura** [16] - O método de aproximação poligonal baseado em curvatura combina as técnicas de detecção de canto e aproximação poligonal para detectar os pontos dominantes e em seguida fazer a correspondência de formas. Esse método de detecção consiste em três procedimentos: (1) extrair os pontos de quebra que não estão em uma linha reta, (2) detectar os cantos potenciais e (3) realizar a aproximação poligonal particionando as curvas entre dois cantos potenciais consecutivos.
- **Multiscale Fractal Dimension** [17] - Esta técnica combina dois descritores de forma, dimensão fractal multiescala e saliências de contorno, usando uma abordagem baseada em gráficos (*image foresting transform*). Para isso, localiza as saliências do contorno a partir da relação entre o contorno e o esqueleto. Por fim, o descritor de saliência de contorno consiste em um vetor, com localização e valor de saliência ao longo do contorno, e um algoritmo de correspondência.
- **Curvature Scale Space (CSS)** [6] - Os métodos que usam Curvature Scale Space consistem em transformar o contorno em uma representação em forma de arco que caracteriza a inflexão de cada ponto da forma quanto a sua suavização. Os máximos dessa transformada são usados para representar a assinatura forma.

Descritores de Textura e Cor

Textura é uma propriedade presente em praticamente todas as estruturas, como nuvens, vegetação, paredes, cabelo e outros. Não obstante, existe outra gama de métodos, descritores de textura, os quais obtêm as assinaturas da imagem a partir da distribuição do níveis de cinza da imagem.

Além da textura, para as imagens coloridas, ou seja, presentes no domínio RGB, há descritores baseados em cor. Esses algoritmos são relativamente robustos por apresentar independência ao tamanho e rotação da imagem. Outra vantagem desses descritores é

que as cores são facilmente associadas a descrições textuais (nome da cor), facilitando a utilização em muitos sistemas de busca e recuperação.

São exemplos de descritores de textura e baseados em cor encontrados na literatura:

- **Matrizes de Co-ocorrência** - Das matrizes de co-ocorrência são computados valores numéricos que facilitam o entendimento da informação e reduzem a dimensionalidade dos dados. As características, então, podem ser extraídas desta matriz de várias maneiras, as quais dependem do aspecto de interesse. Em *Textural Features for Image Classification* [22], por exemplo, Haralick propõe 14 medidas de textura que servem de assinatura da mesma. São elas: momento angular, contraste, correlação, variância, momento de diferença, inverso, média da soma, variância da soma, entropia da soma, entropia, variância da diferença, entropia da diferença, medidas de informação da correlação, coeficiente de correlação máximo. Esse é um descritor de textura
- **Color Bitmap** - O descritor de cor baseado no artigo *Image Bitmap Feature* [23] calcula o vetor de atributos através das cores dos *pixels* no espectro RGB da imagem de forma global e por regiões fixas. Em seguida, a média e desvio padrão são calculados para cada espaço R, G e B separadamente para todos os *pixels* da imagem. A imagem é então dividida em blocos sem sobreposição e a média dos valores de cada bloco é calculada formando o vetor desejado.
- **Border/Interior pixel Classification (BIC)** [24] - Outro descritor de cor utilizado na literatura é o BIC, descritor de cor que classifica os *pixels* da imagem em *pixels* de borda e interior. Os pixels que possuem a mesma cor dos seus vizinhos são agrupados como interior e os outros são agrupados como borda. Então, são calculados dois histogramas de cor, sendo um para os *pixels* de borda e outro para os *pixels* de interior e, assim, extraídos dos mesmos os atributos.

2.2.5 Reconhecimento de Padrões

O reconhecimento de estruturas em imagens está ligado a conceitos de aprendizagens a partir de um treinamento dos padrões de amostras, são exemplos: a forma do objeto, o tipo de gesto feito, a classificação do reconhecimento facial da pessoa. Existem muitos métodos que têm como utilização a classificação do reconhecimento, como: redes neurais, máquina de vetor de suporte, *K-means*, *Linear Bayesiana* e classificadores *Naive Bayes* [20]. Ainda mais, esses classificadores podem ser divididos quanto a sua forma de aprendizagem em dois grupos: classificadores supervisionados e não supervisionados.

Classificadores Supervisionados

São classificadores que necessitam de um conjunto de dados com os respectivos rótulos. Baseado no treinamento com esses dados, os métodos podem tomar decisões acerca das

classes, existindo assim uma relação entre entrada e saída. Em seguida, alguns exemplos desses algoritmos.

- **Naive Bayes** - Naive Bayes é um classificador probabilístico baseado no Teorema de Bayes, que mostra como determinar a probabilidade de um evento condicional através da probabilidade inversa.
- **Support Vector Machine** - A máquina de vetores de suporte (SVM) desenvolve um modelo tomando as entradas de treinamento, mapeando-as no espaço multidimensional e utilizando regressão para encontrar um hiperplano que melhor separa as classes dessas entradas. Em outras palavras, uma SVM encontra uma linha de separação, hiperplano, que busca maximizar a distância entre os pontos mais próximos em relação a cada uma das classes.
- **Linear discriminant analysis** - O *LDA* é um classificador amplamente utilizado e consiste em estimar um sub-espço linear, de forma que a variância entre as classes seja significativamente maior quando comparada à variância dentro das classes [15]. O classificador realiza esse processo definindo uma função densidade $f_i(x), i = 1, 2, \dots, z$, em que z é o número de classes que um banco de imagens possui e x é a amostra. Com esses dados, o classificador LDA obtém o valor da função densidade para cada classe i e classifica a amostra x para k , em que k é uma classe que representa o maior valor de $f_i(x)$. É importante ressaltar que essa função é definida através dos valores de média da classe i e da matriz de covariância de todas as classes, como mostra a Equação 2.2.

$$f_i(x) = -\frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x - \mu_i)' \Sigma^{-1} (x - \mu_i), \quad (2.2)$$

Classificadores Não Supervisionados

A aprendizagem não supervisionada, aborda problemas com pouca ou nenhuma ideia da informação que os resultados devem apresentar e pode automaticamente encontrar padrões e relações em um conjunto de dados. Esses dados do conjunto de amostras nas quais não se sabe sua relação com as classes, no final do processo se agrupam. Esses agrupamentos, ou *clusters*, podem ser hierárquicos, onde cada grupo pode formar subgrupos, ou não hierárquicos, formando apenas uma partição dos dados, bem como, podem ter sobreposição ou não. Algoritmos com sobreposição indicam que uma única amostra pode pertencer a duas ou mais classes [4]. Segue exemplos desses métodos.

- **Agrupamentos hierárquicos** - Nessas classes de algoritmos podemos citar AGNES e DIANA, que, respectivamente, partem do nada e do todo fazendo inclusões ou exclusões dos parâmetros semelhantes formando novos grupos. Além deles há também o algoritmo de *Dendogram*, que usa métricas (mais próximo, mais distante e média) entres os vizinhos para criar seus grupos.

- ***K-means*** - O algoritmo *K-means* não é hierárquico e é um dos mais utilizados na literatura. Ele usa de medidas de distâncias para calcular as amostras mais próximas a partir dos dados e dos centroides de K sementes. No final do processo, as amostras são rotuladas com a classes da semente mais próxima.
- ***Expectation-maximization*** - O algoritmo de expectativa de maximização (EM) é um método iterativo para encontrar estimativas de máxima verossimilhança nos dados em modelos estatísticos.
- **Modelo Mistura de Gaussianas** - O modelo mistura é um modelo probabilístico para representar a presença de sub-populações dentro de uma população geral. Formalmente, um modelo mistura corresponde à distribuição mistura que representa a probabilidade de distribuição de observações na população geral de dados.

Redes Neurais Artificiais

As redes neurais artificiais (RNAs) formam um importante conjunto de algoritmos de classificação que contêm alto poder discriminativo e podem ser supervisionados e não supervisionados. Além disso, apresentam vantagens como robustez, resistência a falhas e são computacionalmente paralelizáveis.

O modelo computacional de uma RNA é inspirado no sistema nervoso humano e é capaz de realizar o aprendizado de máquina. Esse sistema computacional geralmente é apresentado como neurônios (unidades de cálculo) interconectados, que podem computar valores de entradas, simulando o comportamento de redes neurais biológicas.

Como métodos de classificação baseados em redes neurais artificiais podemos citar não supervisionados: *self-organizing map* (SOM), *adaptive resonance theory* (ART), Hamming Network, Max Net, Neocognitron e redes neurais convolucionais; e os supervisionados: perceptron multicamadas, Adaline, Madaline e redes de base radial;

Estratégias de Validação

- ***Hold-out*** - Este método consiste em dividir o conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação).
- ***K-fold*** - O método de validação cruzada denominado *K-fold* consiste em dividir o conjunto total de dados em K subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os $K - 1$ restantes são utilizados para estimação dos parâmetros e calcula-se a acurácia do modelo. Este processo é realizado K vezes alternando de forma circular o subconjunto de teste.
- ***Leave-one-out*** - A validação cruzada *leave-one-out* é uma técnica que busca avaliar a confiabilidade da generalização de um modelo. De forma simples, ela estima o erro do nosso classificador. Esse processo é realizado utilizando uma amostra x como

teste, enquanto as demais são utilizadas para um conjunto de treinamento. Esse processo continua até que todas as amostras x passem pelo processo de validação. Ao final, a taxa de acerto média é a medida de desempenho do classificador.

Capítulo 3

Método Proposto

Para obter o vetor de características almejado, o processo parte da segmentação da forma, onde a região do contorno da imagem é separada e a partir da mesma formado um conjunto $C_N = \{c_1, c_2, \dots, c_N\}$ de todos os pontos do contorno da forma, em que a coordenada $c = (x, y) \in R^2$ representa o ponto. Em seguida, definimos D como a matriz extraída das distâncias euclidianas dos pontos em C_N entre si. Assim, cada linha w da matriz D representa a distância do ponto c para todos os outros pontos do conjunto C_N .

Com a finalidade de tornar o método invariante à rotação e escala, as distâncias das linhas de $D_{N \times N}$ são ordenadas e divididas por N , sendo este o número de pontos do contorno, garantindo assim uma invariabilidade na geração das características e uma proporção quanto ao tamanho, respectivamente. A propriedade de invariância à translação é respeitada, pois são tomados como referência os próprios pontos do contorno.

Após a transformação acima descrita, a matriz resultante é subdividida, de forma a obter sub-matrizes, com o intuito de extrair um maior número de características e fornecer um conjunto mais abrangente de estatísticas acerca dos valores das distâncias. Para isso, definimos pela Equação 3.1 um conjunto H_k de sub-matrizes que contém número igual de linhas, N , e suas colunas definidas por $N(t_k)$.

$$H_k = \{S_{N \times N(t_1)}^1, S_{N \times N(t_2)}^2, \dots, S_{N \times N(t_k)}^k\} \quad (3.1)$$

Desta forma, o conjunto de sub-matrizes depende diretamente do conjunto de parâmetros T_k , definido na Equação 3.2, ordenados de forma crescente e composto pelos coeficientes t_i , com $i \in \mathbb{N}^*$.

$$T_k = \{0, t_1, t_2, \dots, t_k\} \quad | \quad 0 < t_1 < t_2 < \dots < t_k \leq 1 \quad (3.2)$$

Esses coeficientes definem o intervalo de colunas da matriz resultante $D_{N \times N}$ pertencente a cada sub-matriz $S_{N \times N(t_i)}^i$, dada por $D[\alpha, \beta]$, com α e β definidos na Equação 3.3.

$$\alpha = \{1, 2, \dots, N\} \quad \beta = \{\lfloor N \cdot (t_{i-1}) \rfloor + 1, \dots, \lfloor N \cdot (t_i) \rfloor\} \quad (3.3)$$

Tais intervalos dizem respeito às colunas que variam de $\lfloor N \cdot (t_{i-1}) \rfloor + 1$ até $\lfloor N \cdot (t_i) \rfloor$. Portanto, T_k formará os argumentos do descritor proposto, que variam de forma a obter um melhor arranjo para a aplicação.

Uma vez a etapa passada concluída, resulta-se um conjunto H de sub-matrizes. De cada umas dessas sub-matrizes, extraem-se três vetores, que caracterizam a média ($\vec{\tau}_i$), desvio padrão ($\vec{\Psi}_i$) e mínima distância ($\vec{\Upsilon}_i$) de cada coluna j . Em outras palavras, cada sub-matrizes é reduzida a três vetores de três estatísticas diferentes. Esses vetores são definidos, respectivamente, nas Equações 3.4, 3.5 e 3.6.

$$\begin{aligned} \vec{\tau}_i &= [A_1 \quad A_2 \quad A_3 \quad \dots \quad A_{N(t_i)}] \quad \text{em que} \\ A_j &= \frac{1}{N} \sum_{w=1}^N S_{w,j}^i \end{aligned} \quad (3.4)$$

$$\begin{aligned} \vec{\Psi}_i &= [B_1 \quad B_2 \quad B_3 \quad \dots \quad B_{N(t_i)}] \quad \text{em que} \\ B_j &= \sqrt{\frac{1}{N} \sum_{w=1}^N (S_{w,j}^i - A_j)^2} \end{aligned} \quad (3.5)$$

$$\begin{aligned} \vec{\Upsilon}_i &= [C_1 \quad C_2 \quad C_3 \quad \dots \quad C_{N(t_i)}] \quad \text{em que} \\ C_j &= \min_{\forall w \in \alpha} S_{w,j}^i \end{aligned} \quad (3.6)$$

A partir desse ponto, a dimensão dos dados é reduzida novamente com a criação do vetores $\vec{\phi}_1, \vec{\phi}_2, \dots, \vec{\phi}_k$, cada um dos mesmos representa a porção final de características de cada submatriz $S_{N \times N(t_i)}^i$ do conjunto H . Desse modo, é definido $\vec{\phi}_i$, um vetor de seis características da sub-matriz de índice i , onde essas são obtidas a partir da média μ e desvio padrão σ dos vetores supracitados nas Equações 3.4, 3.5 e 3.6.

$$\vec{\phi}_i = [\mu(\vec{\tau}_i) \quad \sigma(\vec{\tau}_i) \quad \mu(\vec{\Psi}_i) \quad \sigma(\vec{\Psi}_i) \quad \mu(\vec{\Upsilon}_i) \quad \sigma(\vec{\Upsilon}_i)]. \quad (3.7)$$

Por fim, para chegar ao vetor de características final $\vec{\Omega}$ da forma, concatenamos na equação 3.8 todos os vetores $\vec{\phi}_i$, reunindo as características de todas as sub-matrizes S^i em um vetor final.

$$\vec{\Omega} = [\vec{\phi}_1 \quad \vec{\phi}_2 \quad \dots \quad \vec{\phi}_k] \quad (3.8)$$

Capítulo 4

Metodologia

Neste capítulo é apresentado o experimento realizado, com o objetivo de analisar a eficiência do método descrito no Capítulo 3. Vale ressaltar que o presente experimento toma como base a mesma metodologia dos experimentos utilizados nos trabalhos apresentados na Seção 1.2.

Nesse experimento, são classificados dois bancos de imagens: (i) *generic shapes* e (ii) *fish*. O primeiro é composto por 9 classes com 11 imagens cada, o que nos fornece um total de 99 imagens [12, 13]. Cada classe tem uma forma única e com diferenças significativas com relação às outras, como pode ser visto na Figura 4.1. Estão incluídas nessas diferenças oclusão, rotação e partes que faltam (Figura 4.2).

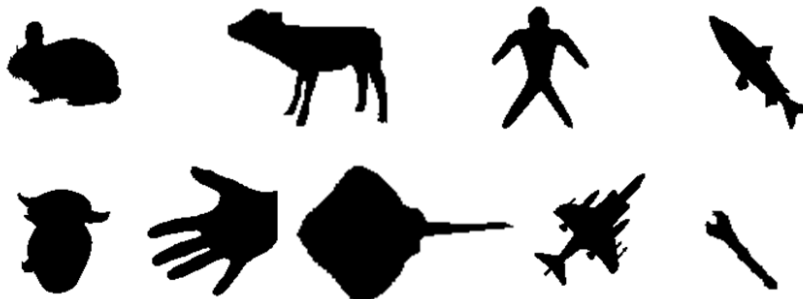


Figura 4.1: Exemplo das classes do banco de imagens *generic shapes* [12, 13].



Figura 4.2: Imagens do banco *generic shapes* com oclusão e rotação [12, 13].

O segundo banco de imagens apresenta formas de peixes [14]. Ele é dividido em 1.100 classes distintas, cada uma possuindo 10 imagens, o que totaliza 11.000 [18]. Cada imagem possui escala e rotação diferentes, o que permite a realização de um teste mais abrangente e preciso das propriedades de invariância a rotação e escala. Exemplos do banco de imagens (ii) são dados na Figura 4.3.



Figura 4.3: Exemplo das classes do banco de imagens *fish* [14].

Tal experimento de visão computacional foi realizado com o auxílio do *software* MATLAB e de acordo com o explicado na Seção 2.1, entretanto, as etapas de aquisição e pré-processamento foram dispensadas pela natureza dos bancos de imagens utilizados, os quais já se encontram binarizados. Assim o mesmo foi dividido em 3 etapas: **(I)** segmentação, **(II)** utilização do descritor proposto e **(III)** validação, conforme a Figura 4.4.

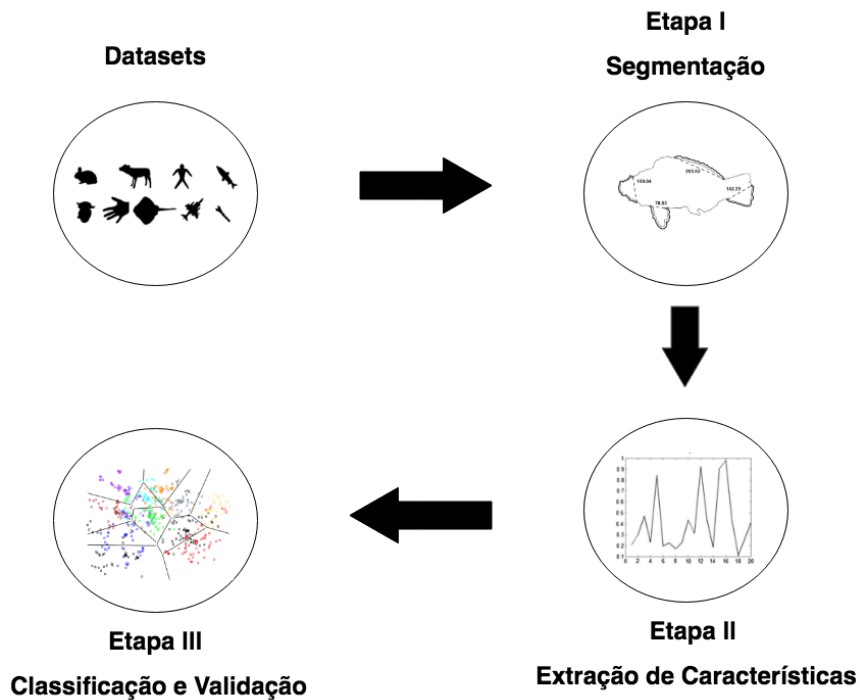


Figura 4.4: Fluxograma da etapas do sistema de visão computacional utilizado como metodologia.

Etapa(I) Na primeira parte, uma segmentação morfológica foi realizada em cada uma das imagens a fim de obter o contorno.

Etapa(II) Em seguida, a partir das imagens segmentadas, o método descrito no Capítulo 3 é executado. os contornos passam pelo descritor proposto, no qual a matriz D de distâncias é gerada e em consequência, ao final das operações, o vetor de características $\vec{\Omega}$ também.

Etapa(III) Quando as características são extraídas, a etapa de validação se inicia. A validação consiste em classificar de forma supervisionada e validar as informações contidas em cada $\vec{\Omega}$ das imagens. Para o presente trabalho foram utilizadas duas técnicas: *Linear Discriminant Analysis* (LDA) e validação cruzada *leave - one - out*, ambas citadas na Seção 2.2.5.

A fim de obter resultados aprimorados, na Seção 5.1 mostra-se a avaliação de parâmetros e em seguida, na Seção 5.2, apresenta-se uma tabela de comparação dos resultados do experimento com outros trabalhos de descritores tradicionais explicados anteriormente no atual trabalho, como: descritores de Fourier [3, 5], momentos de Zernike [8, 9], descritores de Curvatura [16], *Multiscale Fractal Dimension* [17] e *Curvature Scale Space* (CSS) [6].

Capítulo 5

Avaliação de Parâmetros e Resultados

5.1 Avaliação de parâmetros

Como explicado no Capítulo 2, o método proposto depende dos parâmetros da Equação 3.2, dessa forma, para atingir o melhor vetor de características $\vec{\Omega}$, faz-se necessário encontrar um arranjo de T que maximize a discriminação entre as classes de imagens. Além disso, no experimento, é necessário definir a mesma configuração de parâmetros para as duas bases de imagens com o intuito de encontrar um resultado que pode ser usado para comparações futuras. As taxas de acerto do experimento definido na Capítulo 4 foram obtidas variando-se T , valores expostos na Tabela 5.1.

TABELA 5.1: Acurácia para alguns valores T.

Parâmetro T	Nº Desc.	Acurácia(%)	
		<i>Generic shapes</i>	<i>Fish</i>
{0.5, 1}	12	97.97	98.00
{0.33, 0.66, 1}	18	100	98.49
{0.25, 0.5, 0.75, 1}	24	100	98.77
{0.2, 0.4, 0.6, 0.8, 1}	30	100	98.87
{0.1, 0.2, 0.4, 0.6, 0.8, 1}	36	98.98	99.11
{0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1}	42	97.97	99.15
{0.1, 0.2, ..., 0.5, 0.6, 0.8, 1}	48	97.97	99.17
{0.05, 0.1, 0.2, ..., 0.8}	54	100	99.22
{0.05, 0.1, 0.15, ..., 0.5}	54	93.93	98.85
{0.1, 0.2, ..., 0.8, 1}	60	96.96	99.24
{0.05, 0.1, ..., 0.4, 0.5, 0.75, 1}	66	98.98	99.30
{0.05, 0.1, ..., 0.5, 0.75, 1}	72	97.97	99.31
{0.05, 0.1, 0.15, ..., 0.7}	84	71.71	99.11

De acordo com os experimentos foi observado que configurações de T que contêm cardinalidade elevada ou intervalos pequenos (0.01 - 0.04) têm suas taxas de acerto muito baixas, evidenciando a dificuldade em diferenciar valores de características similares ou em grande quantidade. Portanto, esses dados foram retirados da tabela.

Um fato verificado diz respeito à primeira metade da matriz, percebe-se que ao adicionar valores de intervalos abaixo de 0.5 a T reproduz melhores resultados, permitindo uma classificação mais precisa das imagens. Isso mostra que segmentos de reta menores qualificam mais as características do contorno. Essa dedução é inferida da consequência da ordenação das linhas da matriz em D .

Outra característica relevante, é a acurácia se manter superior a 98% quanto a banco de imagens *fish*, independente da variação do número de características. De uma forma geral, manter intervalos uniformes produz bons resultados, mostrando a eficácia das características obtidas das estatísticas de reta do contorno.

5.2 Comparação dos resultados

Dado a análise dos resultados da seção anterior e de acordo com a Tabela 5.1 foi eleito o vetor de tamanho médio $T = \{0.05, 0.1, 0.2, \dots, 0.8\}$ como melhor resultado, servindo como métrica de comparação dos resultados. Para esse arranjo, é possível distinguir as classes com uma acurácia de 99.22% na base de imagens *fish* e também manter eficiência de 100% de acertos quando aplicada na base de imagens *generic shapes*.

Em seguida a definir os parâmetros que permitem que o proposto descritor tenha um melhor desempenho, foi realizada a comparação de resultados com alguns descritores da literatura, citados na a Seção 2.2.4. Os resultados desses outros descritores foram obtidos nos trabalhos [2, 18].

TABELA 5.2: Avaliação dos descritores. Os resultados dos métodos de comparação foram obtidos nos trabalhos [2, 18].

Descritores	Acurácia(%)	
	<i>Generic shapes</i>	<i>Fish</i>
Descritor proposto	100	99.22
<i>Fourier</i>	83.84	99.07
<i>Zernike</i>	91.92	12.23
<i>Curvature</i>	76.77	97.55
<i>M. S. Fractal Dimension</i>	87.88	37.32
<i>Curvature Scale Space (CSS)</i>	71.72	43.52

A Tabela 5.2 mostra que o descritor proposto atinge uma taxa de acerto elevada, que é significativamente maior que os resultados do descritores tradicionais encontrados na literatura.

A acurácia encontrada para o banco *generic shapes* se mostra superior às alcançadas em outros métodos. Apesar de o resultado escolhido ter um grande número de descritores em comparação aos outros métodos, a quantidade de classificações corretas se mantém relativamente alta para escolhas de parâmetros que geram menos descritores (12, 18 e 24), mantendo as taxas de acertos próximas à máxima possível. Isso indica que o método proposto é eficiente para identificar contornos com deformações ou oclusões.

Quanto aos resultados no segundo banco, *fish*, o qual apresenta imagens rotacionadas e escalonadas, é visto que o descritor anteriormente proposto é capaz para extrair características mantendo as propriedades de invariância a rotação e escala. Com exceção dos descritores de *Fourier* e do *Curvature*, nos quais os ganhos são de 0.15% e 1.67%, respectivamente, o método proposto nesse trabalho mostra um resultado, no mínimo, 50% superior quando comparado aos demais. Então, demonstra-se que o método também é eficaz para distinguir formas similares.

5.3 Análise de complexidade

Quanto à complexidade do algoritmo, necessitamos calcular as distâncias de todos os pontos do contorno entre si, gerando N^2 distâncias. Assim, caracteriza-se um algoritmo $O(N^2)$. Essa complexidade é uma desvantagem em comparação aos outros métodos, como *Zernike* e *Fourier* que possuem complexidade $O(N \log N)$.

Capítulo 6

Conclusões e Perspectivas

6.1 Conclusões

No presente trabalho foi apresentado um descritor de formas que utiliza as distâncias euclidianas dos pontos contorno de simples entendimento e implementação. O descritor foi testado em dois bancos de imagens diferentes e obteve resultados superiores a outros descritores tradicionais. Podemos concluir pelo experimento nos bancos de imagens que o método proposto é eficiente em classificar imagens com oclusão, articulação, partes faltantes, e é invariante a rotação, escala e translação.

Além disso foi exposta uma visão geral sobre a literatura de descritores e suas utilidades, bem como uma visão geral e teórica de um experimento de visão computacional.

6.2 Perspectivas

Por fim, como trabalhos futuros, métodos semelhantes que reduzam a complexidade da assinatura proposta, trabalhos com otimização ou seleção de características, devido ao alto número parâmetros e de descritores, e trabalhos que utilizem apenas uma porção dos segmentos de reta, são exemplos de trabalhos que podem vir a melhorar os resultados aqui expostos e fazer parte da literatura de descritores.

Referências Bibliográficas

- [1] A. R. Backes and O. M. Bruno, Shape classification using complex network and Multi-scale Fractal Dimension, *Pattern Recognition Letters* 31(1): 44-51, 2010.
- [2] J. J. M. Sá Junior and A. R. Backes. Shape classification using line segment statistics. *Information Sciences*, 305: 349–356, 2015.
- [3] R. Gonzalez and R. E. Woods. *Processamento de Imagens Digitais*. Editora Edgard Blucher, São Paulo, 2000.
- [4] J. J. M. Sá Junior and A. R. Backes. *Introdução à Visão Computacional Usando Matlab*. Editora Alta Books, Rio de Janeiro, 2016.
- [5] W. P. Lee, M. A. Osman, A. Z. Talib, K. Yahya, J. C. Burie, J. M. Ogier, and J. Mennesson, Recognition of fish based on generalized color Fourier descriptor, In *Science and Information Conference (SAI)*, London, pages 680–686, 2015. Disponível em: <http://ieeexplore.ieee.org/document/7237215/>
- [6] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, 2003.
- [7] M.K. Hu, Visual pattern recognition by moment invariants, *IEEE Trans. Inform. Theory* 8(2): 179–187, 1962.
- [8] N. D. Mustaffa Kamal and N. Jalil, Fastener recognition using combination of colour features with shape-based features and Zernike moments, In *IEEE Student Conference on Research and Development (SCOReD)*, Kuala Lumpur, pages 1–5, 2016. Disponível em: <http://ieeexplore.ieee.org/document/7810064/>
- [9] M. Zhenjiang, Zernike moment-based image shape analysis and its application, *Pattern Recogn. Lett* 21 (2): 169–177, 2000.
- [10] X. Bai, X. Yang, D. Yu, and L. J. Latecki, Skeleton-based shape classification using path similarity, *International Journal of Pattern Recognition and Artificial Intelligence*, 22(4): 733–746, 2008.
- [11] X. Bai and L. J. Latecki, Path similarity skeleton graph matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(7): 1282–1292, 2008.

- [12] D. Sharvit, J. Chan, H. Tek, and B. B. Kimia, Symmetry-based indexing of image databases, *Journal of Visual Communication and Image Representation*, 9(4): 366–380, 1998.
- [13] T. B. Sebastian, P. N. Klein, and B. B. Kimia, Recognition of shapes by editing their shock graphs, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(5): 550–571, 2004.
- [14] S. Abbasi, F. Mokhtarian, and J. Kittler, *Search for similar shapes in the SQUID system: shape queries using image databases*, 2005. Disponível em: <http://www.ee.surrey.ac.uk/CVSSP/demos/css/demo.html>.
- [15] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed., Academic Press, 1990.
- [16] W. Y. Wu and M. J. J. Wang, Detecting the dominant points by the curvature-based polygonal approximation, *Graph. Models Image Process.* 55(2): 79–88, 1993.
- [17] R. da S. Torres, A. X. Falcão and L. da F. Costa, A graph-based approach for multiscale shape analysis, *Pattern Recognition* 37(6): 1163–1174, 2003.
- [18] A. R. Backes, D. Casanova, and O. M. Bruno, A complex network-based approach for boundary shape analysis, *Pattern Recognition*, 42(1): 54–67, 2009.
- [19] R. Szeliski, Computer Vision - Algorithms and Applications, *Springer London Dordrecht Heidelberg New York*, 3–5, 2011.
- [20] Russell, S.J. and Norvig, P., Artificial Intelligence: A Modern Approach, *Prentice Hall Series in Artificial*, 2010.
- [21] C.V. Jawahar and Shiguang Shan, Computer Vision, *ACCV 2014 Workshops: Singapore*, 2014.
- [22] R.M. Haralick, K. Shanmungan and I. Dinstein, Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n. 6, November (1973).
- [23] Tzu-Chuen Lu and Chin-Chen Chang, Color image retrieval technique based on color features and image bitmap, *Information Processing & Management*, Volume 43, Issue 2, 2007, Pages 461-472, ISSN 0306-4573.
- [24] R. O. Stehling, M. A. Nascimento and A. X. Falcao, A compact and efficient image retrieval approach based on border/interior pixel classification, *In CIKM 02: Proceedings of the eleventh international conference on Information and knowledge management*, 2002