Aplicação de Técnicas de Visão Computacional e Aprendizado de Máquina para a Detecção de Exsudatos Duros em Imagens de Fundo de Olho

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Tiago José de Carvalho e aprovada pela Banca Examinadora.

Campinas, 13 de julho de 2010.

Siome K. Goldenstein (Orientador)

Jacques Wainer (Co-orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Fabiana Bezerra Müller – CRB8 / 6162

Carvalho, Tiago José de
C253a Aplicação de técnicas de visão computacional e aprendizado de máquina para a detecção de exsudatos duros em imagens de fundo de olho/Tiago José de Carvalho-- Campinas, [S.P. : s.n.], 2010.

Orientadores : Siome Klein Goldenstein ; Jacques Wainer.
Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Computação.

1.Visão por computador. 2.Processamento de imagens. 3.Aprendizado do computador - Técnicas . I. Goldenstein, Siome Klein. II. Wainer, Jacques. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Título em inglês: Application of techniques of computer vision and machine learning for detection of hard exudates in images of eye fundus

Palavras-chave em inglês (Keywords): 1. Computer vision. 2. Image processing. 3. Machine learning – Technique.

Área de concentração: Visão computacional

Titulação: Mestre em Ciência da Computação

Banca examinadora: Prof. Dr. Siome Klein Goldenstein (IC – UNICAMP)
Prof. Dr. Eduardo Alves do Valle Júnior (IC – UNICAMP)
Prof. Dr. Maurício Marengoni (Fac. de Comp. e Inform. – Mackenzie)

Data da defesa: 13/07/2010

Programa de Pós-Graduação: Mestrado em Ciência da Computação

TERMO DE APROVAÇÃO

Dissertação Defendida e Aprovada em 13 de julho de 2010, pela Banca examinadora composta pelos Professores Doutores:

Prof. Dr. Mauricio Marengoni
Faculdade de Computação e Informática / Mackenzie

Prof. Dr. Eduardo Alves do Valle Junior
IC / Unicamp

Prof. Dr. Siome Klein Goldenstein
IC / UNICAMP

Instituto de Computação
Universidade Estadual de Campinas

Aplicação de Técnicas de Visão Computacional e Aprendizado de Máquina para a Detecção de Exsudatos Duros em Imagens de Fundo de Olho

Tiago José de Carvalho

Julho de 2010

Banca Examinadora:

- Siome K. Goldenstein (Orientador)
- Eduardo Alves do Valle Junior - Instituto de Computação - UNICAMP
- Mauricio Marengoni - Faculdade de Computação e Informática - Mackenzie
- Ricardo Torres (Suplente Interno) - Instituto de Computação - UNICAMP
- Léo Pini Magalhães (Suplente Externo) - Faculdade de Engenharia Elétrica e de Computação - UNICAMP

Resumo

O desenvolvimento de métodos computacionais capazes de auxiliar especialistas de diversas áreas na realização de suas tarefas é foco de diversos estudos. Na área da saúde, o diagnóstico precoce de doenças é muito importante para a melhoria da qualidade de vida dos pacientes. Para oftalmologistas que tratam de pacientes com diabetes, um método confiável para a detecção de anomalias em imagens de fundo de olho é importante para um diagnóstico precoce evitando o aparecimento de complicações na retina. Tais complicações podem causar até cegueira. *Exsudatos duros* é uma das anomalias mais comuns encontradas na retina, sendo sua detecção o foco de vários tipos de abordagens na literatura. Esta dissertação apresenta uma nova e eficiente abordagem para detecção de exsudatos duros em imagens de fundo de olho. Esta abordagem utiliza técnicas de visão computacional e inteligência artificial, como descritores locais, dicionários visuais, agrupamentos e classificação de padrões para detectar exsudatos nas imagens.

Abstract

The computational methods development can helps specialists of several areas in your works is focus of many studies. In health area the premature diagnosis of diseases is very important to improve the patient's life quality. To ophthalmologists who treat patients with diabetics, a reliable method to anomalies detects in eye fundus images is important to a premature diagnosis, avoiding appear of retina complications. Such complications can cause blindness. Hard Exsudates is one of more common anomalies found at retina, being your detection is the focus of many kinds of approaches in literature. This master's thesis presents a new and efficient approach for detection of exsudates at eye fundus images. This approach uses computer vision and artificial inteligence techniques like visiual dictionaries, clustering and pattern recognition to detect hard exsudates in images

Agradecimentos

Desde o momento de seu nascimento, o ser humano vive uma jornada cheia de altos e baixos, problemas e soluções, desafios e superações. Mas nunca estamos sozinhos. Sempre há um pai, uma mãe, um irmão, um amigo, ou se não temos nenhum deles ao nosso lado, ainda assim temos uma força maior, um Deus infinito em bondade que nos ajuda e nos acolhe em qualquer que seja a situação. E é a esse Deus que venho em primeiro lugar agradecer. Agradecer pela força necessária para vencer noites em claro, pela determinação de levantar a cabeça a cada derrota, a cada fracasso, e seguir em frente, nunca desistindo diante dos obstáculos. Agradecer pela sabedoria necessária para realizar meu trabalho da melhor forma possível e acima de tudo agradecer pela felicidade de estar completando com êxito mais esta fase na minha vida.

E o que seria de mim, sem todas as pessoas que me amam e me ajudam a cada dia? Venho agradecer a meus pais, que sempre estão ao meu lado, seja com uma palavra de incentivo, ou para ouvir minhas muitas lamentações nos momentos de raiva. A minha irmã que sempre está com seu jeito doce e meigo próxima a mim amenizando um pouco a dor que sinto de estar longe de casa. A meus tios Regina e Magela, que sempre me acolheram como um filho em sua casa nesses dois anos em que estou longe de casa. A meus colegas de laboratório, e porque não dizer, amigos de laboratório, que por muitas vezes me deram uma pequena luz nos momentos em que nenhuma ideia aflorava de minha mente. A meu orientador Siome e aos professores Anderson e Jacques que me deram todo o suporte para que esse trabalho fosse finalizado com sucesso.

São tantas as pessoas a que devo um agradecimento, que todas as páginas desta dissertação não seriam suficientes. Por isso de uma maneira muito geral, eu queria agradecer a todos que um dia fizeram parte de minha vida, e aos que hoje ainda fazem, pois, é através de vocês, através das experiências que vivi ao lado de vocês, é que construí meu caráter, me tornando o homem que sou hoje e vencendo mais essa batalha.

A todos vocês o meu muito obrigado!

Sumário

Resumo **v**

Abstract **vi**

Agradecimentos **vii**

1 Preliminares **1**
1.1 Introdução 1
1.2 O Problema 2
1.3 Objetivos 2
1.4 Contribuições 3

2 Retinopatia Diabética **4**
2.1 A Doença 4
2.2 Anomalias Causadas pela RD 7

3 O Estado da Arte na Detecção de Anomalias provenientes da Retinopatia Diabética **10**
3.1 Detecção de Exsudatos 10
3.2 Doentes × Não Doentes 13

4 Revisão Bibliográfica **16**
4.1 Imagem 16
4.2 Pontos Característicos em Imagens 16
4.2.1 Detector de Hessian 17
4.3 Descritores de Imagens 18
4.4 Descritores Locais 18
4.4.1 SIFT 19
4.4.2 SURF 21
4.5 Aprendizado de Máquina 22

4.5.1 Reconhecimento de Padrões . . . 23
4.5.2 Agrupamento . . . 23
4.5.3 K-Médias . . . 26
4.5.3.1 Organização dos Dados . . . 26
4.5.3.2 Seleção dos Centróides Iniciais . . . 27
4.5.3.3 Partição dos Dados . . . 27
4.5.3.4 Atualização dos Centróides . . . 27
4.5.3.5 Critério de Convergência . . . 28
4.5.4 Classificadores . . . 28
4.5.4.1 Validação Cruzada . . . 29
4.5.4.2 SVM . . . 29
4.6 Curvas ROC . . . 30

5 Dicionários Visuais **33**
5.1 Sacola de Palavras Visuais . . . 34
5.1.1 Construção das Sacolas . . . 35

6 Metodologia Cientifica **38**
6.1 Base de Dados . . . 38
6.2 Geração de Dobras . . . 40
6.3 Pré-processamento . . . 41
6.4 Extração e Descrição de Regiões da Imagem . . . 41
6.4.1 Extração e Descrição Utilizando SIFT . . . 41
6.4.2 Extração e Descrição Utilizando SURF . . . 42
6.5 Construção dos Dicionários Visuais . . . 43
6.5.1 Conjunto de Treinamento . . . 43
6.5.2 Seleção das Palavras Visuais . . . 43
6.5.2.1 Por Agrupamento . . . 44
6.5.2.2 Por Seleção Aleatória . . . 44
6.5.2.3 Por Seleção de Região . . . 44
6.5.2.4 Por Seleção de Manual . . . 46
6.6 Composição das Sacolas de Palavras (Histogramas Visuais) . . . 47
6.7 Treinamento e Classificação . . . 48
6.7.1 SVM – Treinamento e Classificação . . . 50

7 Experimentos e Resultados **52**
7.1 Agrupamento × Seleção Aleatória . . . 52
7.2 SURF × SIFT . . . 54
7.3 25 × 50 ×100 × 500 × 1000 Palavras Visuais . . . 57

7.4 Seleção de Regiões 59
7.5 Seleção Manual 66
7.6 Comparação de Resultados 67

8 Conclusões e Trabalhos Futuros **70**

Bibliografia **72**

Lista de Tabelas

7.1 Experimento 1 - base de dados 53
7.2 Experimento 1 - dicionário visual 53
7.3 Experimento 1 - classificador 53
7.4 Experimento 2 - base de dados 56
7.5 Experimento 2 - dicionário visual 56
7.6 Experimento 2 - classificador 56
7.7 Experimento 3 - base de dados 59
7.8 Experimento 3 - dicionário visual 59
7.9 Experimento 3 - classificador 59
7.10 Experimento 4 - base de dados 63
7.11 Experimento 4 - dicionário visual 63
7.12 Experimento 4 - classificador 63
7.13 Experimento 5 - base de dados 66
7.14 Experimento 5 - dicionário visual 66
7.15 Experimento 4 - classificador 66
7.16 Resultados de alguns métodos para detecção de exsudatos. 69

Lista de Figuras

2.1 Projeções da OMS para o número de pessoas com diabetes no mundo retirados. 5
2.2 Exemplo de imagem de fundo de olho saudável com as principais estruturas da retina destacadas. 6
2.3 Exemplo microaneurismas em imagem de fundo de olho. 7
2.4 Exemplo exsudato duro. 8
2.5 Exemplo de um dos tipos de hemorragia intra-retinal. 9
2.6 Exemplo de neovascularização. 9

3.1 Exemplos de imagens com e sem exsudato: à esquerda uma imagem de fundo de olho de uma retina normal. Já a imagem da direita apresenta exsudatos duros no local indicado pelo círculo. 11

4.1 Ponto característico encontrado na imagem de uma mesma cena vista de diferentes pontos de observação. 17
4.2 Exemplo do funcionamento de um descritor de imagens. Adaptado de [64]. 19
4.3 Pontos característicos extraídos com o algoritmo SIFT. 20
4.4 Pontos característicos extraídos com o algoritmo SURF. 22
4.5 Exemplo de uma das possíveis configurações para sistemas de reconhecimento de padrões. 24
4.6 Exemplo de um agrupamento ideal: após o agrupamento os elementos de um conjunto são apenas similares entre si e diferentes dos elementos de qualquer outro conjunto. 25
4.7 Exemplos de curvas ROC. Na imagem da esquerda a curva em vermelho é sempre melhor do que a curva em azul. Já na imagem da direita, a curva em azul se sobrepõe a curva em vermelho em uma parte do gráfico, indicando que nessa parte a curva em azul representa os melhores resultados. 32

5.1 Exemplo de palavras visuais. Imagem reproduzida de [71] com autorização do autor. 34
5.2 Geração das sacolas de palavras. 37

6.1 Exemplos de imagens contidas na base de dados. 39
6.2 Fluxograma de seleção de palavras visuais utilizando agrupamento. 45
6.3 Fluxograma de seleção de palavras visuais utilizando seleção aleatória. . . . 46
6.4 Fluxograma de seleção de palavras visuais utilizando seleção baseada em regiões marcadas. 47
6.5 Fluxograma de seleção de palavras visuais utilizando seleção manual das palavras por um especilista. 48

7.1 Curva ROC média das abordagens com e sem agrupamento. 54
7.2 Curvas de desvio padrão da abordagem com agrupamento. 55
7.3 Curvas de desvio padrão da abordagem sem agrupamento. 55
7.4 Curva ROC média com a utilização do SIFT e SURF na extração dos PCs. 57
7.5 Curvas de desvio padrão da abordagem com o SIFT. 58
7.6 Curvas de desvio padrão da abordagem com o SURF. 58
7.7 Curvas representando testes com diferentes números de palavras visuais. . 60
7.8 Curvas de desvio padrão utilizando 25 palavras visuais. 60
7.9 Curvas de desvio padrão utilizando 50 palavras visuais. 61
7.10 Curvas de desvio padrão utilizando 100 palavras visuais. 61
7.11 Curvas de desvio padrão utilizando 500 palavras visuais. 62
7.12 Curvas de desvio padrão utilizando 1000 palavras visuais. 62
7.13 Curvas representando testes com e sem a seleção de regiões, juntamente com uma curva representando um teste com agrupamento. 64
7.14 Curvas de desvio padrão não utilizando seleção de regiões. 64
7.15 Curvas de desvio padrão utilizando seleção de regiões. 65
7.16 Curvas de desvio padrão utilizando agrupamento. 65
7.17 Curvas representando testes com e sem a seleção manual de palavras, juntamente com curvas representando a seleção por regiões e a seleção aleatória. 67
7.18 Curvas de desvio padrão utilizando seleção manual. 68
7.19 Curvas de desvio padrão utilizando seleção de regiões. 68
7.20 Curvas de desvio padrão utilizando seleção aleatória. 69

Capítulo 1

Preliminares

1.1 Introdução

Atualmente, os cuidados com a saúde tem sido foco da atenção de governantes, principalmente, em países desenvolvidos. Nesses países a diabetes torna-se a cada dia um problema ainda maior. Ela é uma doença crônica, causada pelo aumento da taxa de glicose no sangue e que causa danos aos vasos sanguíneos. Seu aumento está intimamente ligado à obesidade. Dados da Organização Mundial de Saúde estimam que se não forem tomadas providências urgentes, em 2030 o número de pessoas portadoras de diabetes no mundo deve superar os 300 milhões.

Uma das complicações provenientes do *diabetes mellitus* é a retinopatia diabética, que se caracteriza por afetar os vasos sanguíneos da retina. Isto pode causar o aparecimento de inchaços ou ainda permitir o vazamento de fluidos, comprometendo as estruturas da retina que precisam estar em boas condições para uma visão de qualidade.

A *retinopatia diabética* é uma doença que apresenta diversos estágios, sendo a evolução de um estágio para outro muito rápida. Isso torna necessário que além do empenho na prevenção da doença diminuindo os fatores de risco como a obesidade, seja realizado um acompanhamento de pacientes com diabetes os quais possuem alto risco de apresentar tal doença. O diagnóstico precoce possibilita que um tratamento seja iniciado nos primeiros estágios evitando grandes danos à retina. O diagnóstico é realizado por especialistas através de uma análise de imagens de fundo de olho.

Um dos fatores que impossibilitam a implantação de um programa eficiente de acompanhamento é o baixo número de médicos especialistas comparado com o aumento do número de casos da doença.

Um sistema computacional capaz de identificar a presença de uma anomalia em imagens de fundo de olho seria de grande ajuda para os especialistas. No entanto, existem diversos tipos de anomalias e todas as abordagens para tentar automatizar esse tipo detec-

ção são desenvolvidas de forma muito específicas, o que na maioria das vezes, impossibilita que um método aplicado para uma anomalia seja estendido para outra.

Exsudato duro é um, dentre os diversos tipos de anomalias provenientes da retinopatia diabética. Aparecem logo nos primeiros estágios da doença e são caracterizados por seu brilho intenso nas imagens. Sua identificação é o foco de boa parte dos esforços para detecção de anomalias em imagens de fundo de olho. Operadores morfológicos e abordagens baseadas na intensidade de brilho são a maneira mais comum de identificar os exsudatos.

Esta dissertação propõe um novo tipo de abordagem para a identificação de exsudatos em imagens de fundo de olho utilizando dicionários visuais e classificadores para determinar a presença ou não desta anomalia.

A dissertação é organizada de forma que o capítulo dois apresenta formalmente conceitos ligados à doença. O capítulo três faz uma apresentação do estado da arte na detecção de exsudatos, juntamente com a detecção de anomalias gerais na retina. Uma revisão bibliográfica de métodos da literatura de visão computacional e inteligência artificial é o tema do capítulo quarto. O capítulo cinco especifica o conceito de dicionários visuais, no qual este trabalho é baseado. Os capítulos seis e sete apresentam respectivamente a metodologia utilizada na pesquisa, bem como os principais experimentos realizados com seus resultados. Por fim o capítulo oito apresenta as conclusões e propostas de trabalhos futuros geradas a partir deste trabalho.

1.2 O Problema

O número de casos de retinopatia diabética vem crescendo muito, impulsionados pelo aumento no número de pessoas com *diabetes mellitus*. Torna-se necessário a construção de um método computacional confiável que auxilie especialistas a identificar de forma rápida anomalias ocorrentes na retina, possibilitando um tratamento precoce e evitando danos significativos à acuidade visual dos pacientes.

1.3 Objetivos

O objetivo deste trabalho se baseia na construção de um método capaz de identificar anomalias de exsudato duro em imagens de fundo de olho utilizando dicionários visuais e técnicas de aprendizado de máquina.

1.4 Contribuições

Ao contrário da maioria dos métodos já propostos, o método proposto nessa dissertação é capaz de detectar exsudatos em imagens de fundo de olho, sem qualquer melhoramento nas imagens, ou ainda sem que o método procure características que são específicas de apenas um tipo de anomalia. No estado da arte atual, exsudatos são identificados através de seu brilho. Porém, o método proposto nesse trabalho, não se concentra nesse tipo de abordagem, utilizando dicionários visuais. Outra contribuição importante foi a constatação de que para esse tipo de problema, o uso de agrupamento na seleção das palavras visuais não realizou um melhoramento significativo nos resultados, contradizendo grande parte da literatura de dicionários visuais. Isso mostra que a seleção aleatória de palavras visuais também funciona.

Capítulo 2

Retinopatia Diabética

A diabetes vem se tonando mais e mais comum em no mundo todo, contando com diversos fatores agravantes (*e.g.* obesidade). Um paciente portador de diabetes, pode apresentar diversas complicações em seu quadro clínico, sendo a retinopatia diabética uma dessas complicações. Neste capítulo são apresentados diversos conceitos, dados e características relacionados à retinopatia diabética.

2.1 A Doença

A *diabetes mellitus* **(DM)** é o nome atribuído a uma doença sistêmica, crônica e que apresenta risco de vida. Ocorre quando o pâncreas não produz insulina suficiente ou quando o corpo não consegue processá-la de forma adequada, fazendo aumentar dessa forma o nível de glicose no sangue. Com o passar do tempo causa danos aos vasos sanguíneos podendo afetar os olhos e o sistema nervoso, além de rins, coração e outros órgãos [1].

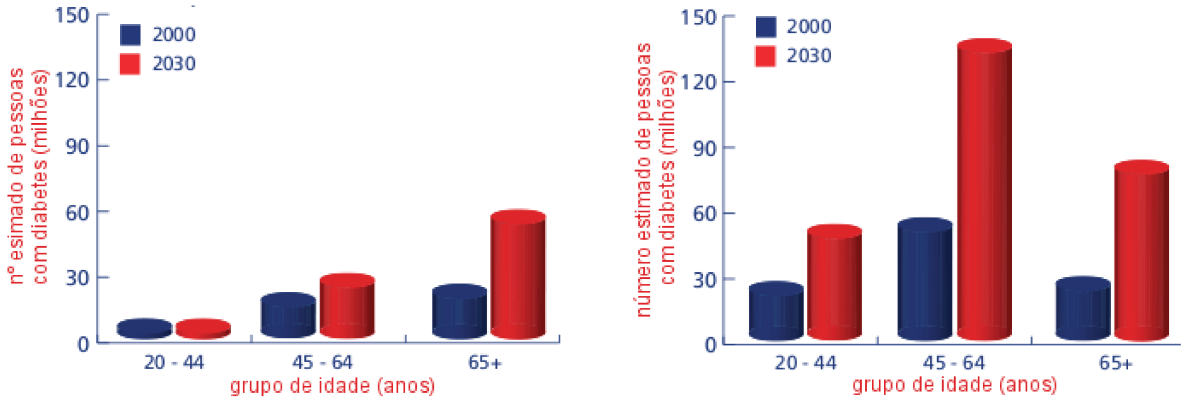
Vem crescendo de forma rápida em todo o mundo e segundo dados fornecidos pela Organização Mundial de Saúde[1] **(OMS)** 135 milhões de pessoas tem DM no mundo todo e esse número deve subir para 300 milhões em 2025 [90]. Os gráficos das Figuras[2] 2.1(a) e 2.1(b) mostram uma projeção fornecida pela OMS para 2030 do número de pessoas com DM em países desenvolvidos e em países em desenvolvimento respectivamente.

Dados fornecidos pelo Programa de Diabetes da OMS indicam que a diabetes causa cerca de 5% de todas as mortes no mundo por ano e, se nenhuma ação urgente for tomada, esse número será aumentado em 50% nos próximos 10 anos.

A retinopatia diabética **(RD)** é um dos tipos de complicações provocados pela DM no organismo. Se não tratada a tempo e de forma correta pode eventualmente causar a

[1] http://www.who.int/diabetes/en/index.html

[2] Figuras retiradas de http://www.who.int/diabetes/facts/en/ em 05 de agosto de 2010



(a) Pessoas com DM - países desenvolvidos (b) Pessoas com DM - países em desenvolvimento

Figura 2.1: Projeções da OMS para o número de pessoas com diabetes no mundo retirados.

perda de visão do indivíduo. Sozinha ela cega aproximadamente 25 mil pessoas com DM por ano nos E.U.A. e é a principal causa de cegueira da população em idade de trabalho nos E.U.A. e na Europa [54].

Estes danos ocorrem quando o nível de glicose no sangue está alto, causando danos em pequenos vasos sanguíneos da retina. Em determinadas pessoas, esses vasos sanguíneos podem inchar e vazar fluidos. Em outras, um novo vaso sanguíneo cresce na superfície da retina. Como a retina é um tecido muito sensível do olho, uma boa visão depende de sua estrutura estar saudável.

A RD é afetada por dois fatores principais, idade e tempo da doença, sendo que mais de 90% dos pacientes com diabetes dependentes de insulina a mais de 20 anos, desenvolvem esse tipo de anomalia [14].

Segundo o Instituto Nacional dos Olhos dos Estados Unidos[2] **(INO)** a RD apresenta quatro estágios evolutivos sendo

- *retinopatia diabética não proliferativa leve*: é o primeiro estágio da doença apresentando pequenos *microaneurismas* (pequenos inchaços) próximos aos vasos mais finos;
- *retinopatia diabética não proliferativa moderada*: é o segundo estágio da doença, no qual vasos responsáveis por irrigar a retina são bloqueados;
- *retinopatia diabética não proliferativa severa*: o terceiro estágio da doença, caracterizado pelo bloqueio de um número muito maior de vasos, comprometendo dessa

[2]http://www.nei.nih.gov/health/diabetic/retinopathy.asp

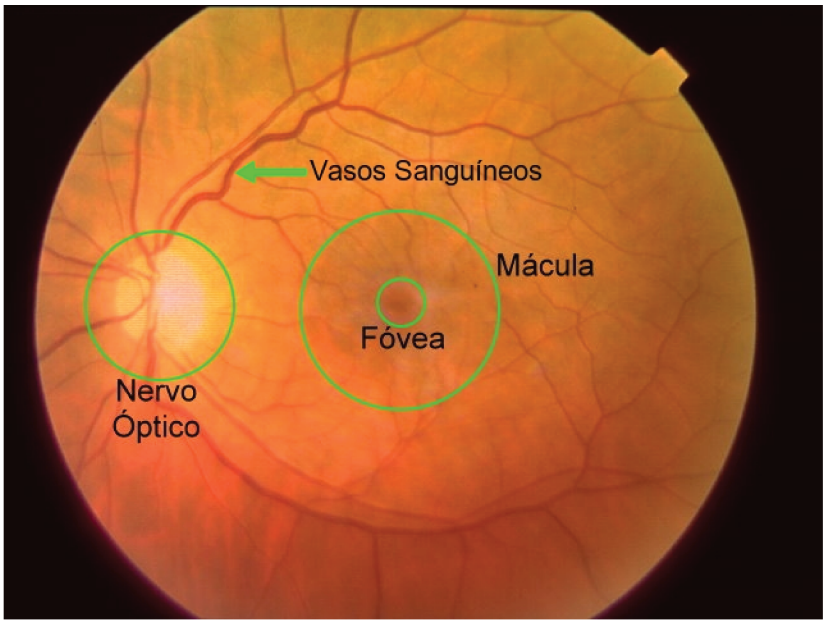


Figura 2.2: Exemplo de imagem de fundo de olho saudável com as principais estruturas da retina destacadas.

forma a irrigação da retina. Devido a essa deficiência de irrigação algumas áreas da retina enviam sinais para o corpo para que o mesmo faça crescer novos vasos;

- *retinopatia proliferativa*: é o estágio mais avançado da doença. Devido ao sinal enviado ao corpo no terceiro estágio da doença, começam a surgir novos vasos sanguíneos. Sozinhos eles não causam danos à retina, no entanto, esses novos vasos são anormais e frágeis e por isso eles podem se romper causando uma perda significativa de visão chegando até mesmo a causar cegueira.

O grande problema da retinopatia diabética é que ela não apresenta sintomas nos primeiros estágios da doença, o que dificulta um diagnóstico precoce. Por isso ela está na lista de prioridades da Organização Mundial de Saúde de condições dos olhos que devem ser acompanhadas e tratadas.

Em pacientes com diabetes, a realização de exames de fundo de olho periodicamente é essencial para desse modo prevenir e tratar a retinopatia diabética de forma correta, uma vez que especialistas utilizam imagens de fundo de olho **(IFO)** para realização de um diagnóstico.

A Figura 2.2 exibe um exemplo de uma imagem de fundo de olho **(IFO)** com as principais regiões da retina destacadas.

2.2 Anomalias Causadas pela RD

Existem diversos tipos de anomalias causadas na retina provenientes da RD. Correa e Eagle Jr. [16] definem várias dessas anomalias das quais algumas serão apresentadas aqui. É importante notar que existem muitos tipos de anomalias provenientes da RD que podem aparecer em imagens de fundo de olho como

- *microaneurismas*: como vistos na Figura 2.3, microaneurismas retinianos aparecem como gomos de uva ou dilatações fusiformes nos capilares. Não se sabe exatamente como estes microaneurismas se formam. Vários microaneurismas parecem bem celulares sugerindo que a proliferação endotelial nos capilares possa estar envolvida em sua formação;

Figura 2.3: Exemplo microaneurismas em imagem de fundo de olho.

- *edema retiniano e exsudatos duros*: essas anomalias aparecem no segundo estágio da retinopatia diabética e refletem a quebra da barreira hemato-retiniana, a qual é uma das lesões funcionais mais precoces em olhos diabéticos. A RD ativa diversos processos fisiológicos culminando na incompetência da barreira hemato-retiniana, o que permite o acesso de fluido rico em lipídeos e proteína ao parênquima retiniano, causando edema e exsudação. A Figura 2.4 exibe um exemplo de exsudato duro;

Figura 2.4: Exemplo exsudato duro.

- *hemorragias*: tais anomalias possuem formatos variados, dependendo da localização do extravasamento de sangue na retina. Um dos exemplos de hemorragia pode ser visto na Figura 2.5;

- *neovascularização*: o processo de neovascularização se inicia na retina onde é constatada a presença de anormalidades microvasculares intra-retinianas. No entanto os novos vasos formados são frágeis e crescem desordenadamente, crescendo sobre a superfície interna da retina e prejudicando a mesma. Um exemplo de neovascularização é visto na Figura 2.6.

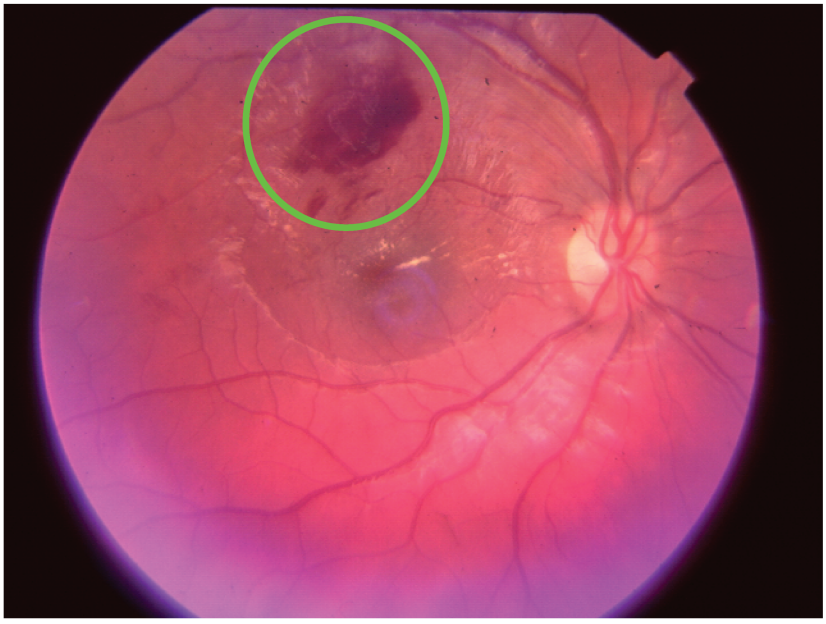


Figura 2.5: Exemplo de um dos tipos de hemorragia intra-retinal.

Figura 2.6: Exemplo de neovascularização.

Capítulo 3

O Estado da Arte na Detecção de Anomalias provenientes da Retinopatia Diabética

A maioria dos danos causados à visão devido a RD, ocorrem devido ao fato que na maioria das vezes a doença é descoberta já em um estágio avançado. Para que o tratamento das anomalias provenientes da RD seja eficiente, é preciso que o mesmo seja ministrado nos primeiros estágios da doença. Assim, uma detecção precoce através de exames regulares se torna de grande importância. Para diminuir o custo de tais exames, a tecnologia deve ser utilizada para uma análise das IFOs através de métodos de processamento de imagem, visão computacional, inteligência artificial, dentre outros [24]. Este capítulo apresenta uma visão geral dos principais métodos existentes na literatura para a detecção de anomalias.

3.1 Detecção de Exsudatos

O exsudato é uma das anomalias provenientes da RD, sendo que sua detecção é fonte de diversos estudos. Em imagens de fundo de olho ele pode ser caracterizado por uma região amarela, de brilho intenso e dimensionalidade variada. A dimensão dessa região varia de acordo com o estágio da doença em que o paciente se encontra. A diferença entre duas imagens, com e sem exsudato, pode ser observada na Figura 3.1.

Ele é uma grande preocupação uma vez que pode gerar uma grande perda de visão quando ocorre na região central da mácula. Assim, esse tipo de lesão precisa ser regularmente monitorada e tratada de forma correta para evitar danos na acuidade visual do paciente.

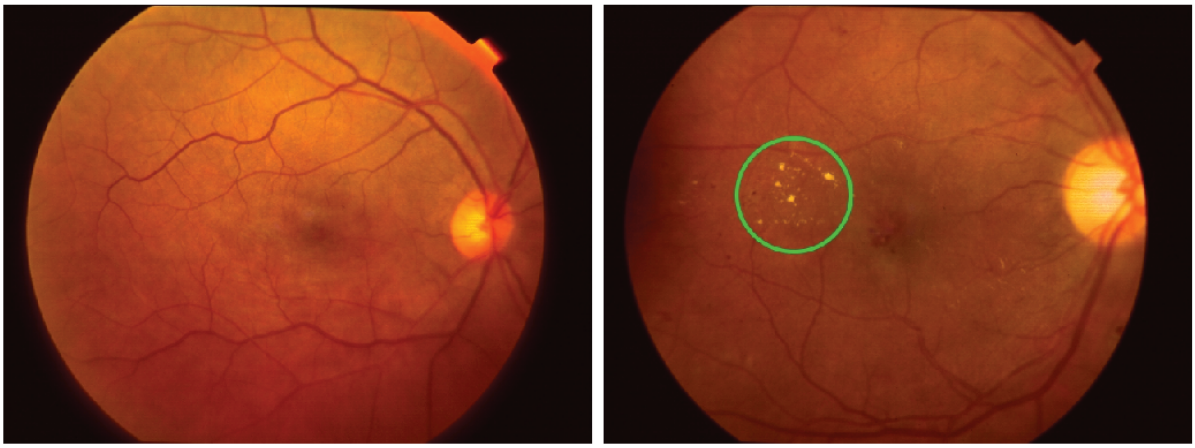


Figura 3.1: Exemplos de imagens com e sem exsudato: à esquerda uma imagem de fundo de olho de uma retina normal. Já a imagem da direita apresenta exsudatos duros no local indicado pelo círculo.

Welfer *et.al.* [99] foca seu método para a detecção de exsudato na busca de possíveis sinais de *edema diabético macular* **(DME)**, que é também um tipo de exsudação da retina.

O método trabalha no canal L do sistema de cores LUV, utilizando operações morfológicas [32] na detecção. Começa detectando exsudatos de forma grosseira através do uso de aberturas e fechamentos morfológicos utilizando uma elipsoide como elemento estruturante. Utiliza também técnicas como a detecção da região mínima [2, 20], a transformação H-máxima [81, 20] ou ainda a transformada de *Watershed* [72].

Esta primeira etapa, detecta os pontos candidatos a exsudato, que são então refinados utilizando o canal G do sistema de cores RGB da imagem original, operações morfológicas baseadas em um elemento estruturante em forma de diamante e segmentação, resultando em uma imagem dos exsudatos contidos na imagem original.

Já Sopharak *et.al.* [83] realiza um pré-processamento mudando o espaço de cores das imagens de RGB para HSI. Ele trabalha utilizando a banda I (intensidade) aplicando um filtro média para diminuir o ruído, e logo após realiza um melhoramento no contraste através da aplicação de **CLAHE** [32]. Isso porque lesões provenientes de exsudato e a região do disco óptico exibem valores de alta intensidade nesse canal, e o melhoramento de contraste associa a eles valores de grande intensidade.

A seguir utiliza-se de operações morfológicas para remover vasos de alto contraste e bordas, removendo também a região correspondente ao disco óptico através do método contido em [83]. Após o uso de outras operações morfológicas e uma limiarização na imagem uma imagem final contendo os exsudatos é produzida.

Em outro de seus trabalhos, Sopharak *et.al.* [82], propõe uma nova forma de detecção para exsudatos, utilizando um agrupamento do tipo "fuzzy" com o algoritmo K-médias. O *agrupamento "fuzzy" por K-médias* **(AFM)**, é um tipo de agrupamento no qual uma mesma amostra pode pertencer a dois ou mais grupos graus diferentes de adesão para cada grupo. A descrição do método AFM é detalhada em [82, 69].

A primeira parte do método proposto por Sopharak *et.al.* conta com uma seleção empírica de características auxiliada por uma pesquisa junto a oftalmologistas, a qual resulta na seleção de quatro características de entrada: o valor de intensidade depois de um pré-processamento, o desvio padrão da intensidade, a matiz e o número de pixels de borda da imagem.

AFM é então aplicado sobre as imagens obtidas na etapa anterior, com um número de grupos determinado empiricamente. O grupo importante resultante da aplicação do AFM é o grupo que contém a maior parte da informação contida na imagem original, porém, sem as áreas de exsudato, o qual é chamado de *primeiro grupo*.

Os pixeis de exsudatos são obtidos subtraindo o primeiro grupo da imagem original dando origem a uma imagem $\hat{I}_1$. Refinando a seleção, o primeiro grupo é também utilizado como marcador, $\hat{I}_2$, enquanto a imagem original de intensidades é usada como uma máscara, $\hat{I}_3$. De posse desses componentes, uma reconstrução morfológica por dilatação e então aplicada em $\hat{I}_2$ utilizando a máscara $\hat{I}_3$.

O resultado final contendo os exsudatos é obtido aplicando uma operação de limiarização sobre a diferença entre a imagem original e a imagem reconstruída.

Um tipo diferente de abordagem é proposto por Garcia *et.al.* [52] na qual classificadores são utilizadas para a detecção de exsudatos. O método começa com a normalização das imagens para com isso aumentar o contraste entre os exsudatos e o fundo da imagem. Essa normalização é baseada em um método proposto em [50] que estima a imagem original sem distorções através de estimativas de luminosidade e desvios de contraste.

Tenta então localizar e segmentar as regiões de exsudato presentes na imagem, para que possam ser classificadas pelos classificadores na próxima etapa do método. Essa tentativa de localização se baseia em propriedades locais dos exsudatos juntamente com uma combinação de metódos de histogramas globais e adaptativos. No entanto, o disco óptico pode ser considerado erroneamente como uma das regiões candidatas a exsudato. Para evitar tal problema, o método utiliza uma combinação de morfologia matemática, detecção do máximo regional [40], transformada de Hough [21] e uma aproximação circular para o disco óptico, para com isso realizar sua remoção.

Das regiões candidatas é extraído um conjunto de características para que os classificadores possam identificar quais entre as regiões são realmente exsudatos. As características propostas por Garcia são escolhidas com base nas características que especialistas utilizam para distinguir exsudatos em uma IFO e são definidas como a média dos valores RGB

dentro da região região, o desvio padrão dos valores RGB dentro da região, a média dos valores RGB fora da região, o desvio padrão dos valores RGB fora da região, os valores médios do centroide da região, o tamanho da região, a compacidade da região, e robustez da borda da região.

Por fim, o método utiliza-se de alguns classificadores contidos na literatura de aprendizado de máquina para separar as regiões candidatas em exsudato ou não exsudato.

Fleming [28] apresenta uma abordagem para detecção de exsudatos baseada em operações morfológicas multi-escala e probabilidades. Seu método parte de um pré-processamento que filtra a imagem no canal G, com o intuito de reduzir o ruído, seguido de uma identificação e exclusão do disco óptico da forma descrita em [27] dando origem a uma nova imagem $\hat{\mathrm{I}}_1$.

Os candidatos a exsudato são obtidos da construção em cinco diferentes escalas, de uma imagem $\hat{\mathrm{I}}_{2S}$ (onde S representa a escala da imagem) resultante da diferença entre $\hat{\mathrm{I}}_{1S}$ e o máximo dentre aberturas morfológicas realizadas por múltiplos elementos estruturantes lineares. Logo cada uma das imagens $\hat{\mathrm{I}}_{2S}$ representa um melhoramento do brilho dos pontos da imagem original em uma escala S.

Para cada uma das escalas, é então aplicado uma limiarização dinâmica descrita em detalhes em [28], a qual elimina dos candidatos regiões com brilho característico devido a reflectância da retina. Assim, as regiões candidatas a exsudato são escolhidas tomando-se os máximos regionais em $\hat{\mathrm{I}}_1$ nos quais para alguma escala S o máximo regional é maior que um limiar.

Candidatos são avaliados através da construção de uma região de crescimento baseada em [26], a qual tem os próprios candidatos como ponto inicial. Essa região de crescimento evita vasos e outras lesões, e é usada para avaliar a luminosidade média e o contraste do fundo da imagem de um candidato.

Das regiões de crescimento, são extraídas características como luminosidade normalizada, desvio padrão normalizado, número de pixeis da região, gradiente normalizado das bordas calculado como a magnitude média do gradiente de $\hat{\mathrm{I}}_1$ ao longo da região, dentre outras.

Por fim, tais características foram classificadas com o SVM utilizando o tipo de "kernel" RBF, determinando se a região era um exsudato, um "drusen" ou ainda uma parte do fundo da imagem.

3.2 Doentes × Não Doentes

Ao contrário do tipo de abordagem descrita na sessão anterior, a qual procura por um tipo específico de anomalia, existem pesquisas que se preocupam com a detecção de anomalias de modo geral, sem necessariamente identificar de qual tipo se trata. Essa abordagem se

preocupa apenas em identificar se uma determinada imagem contém ou não algum tipo de anomalia.

Focado nessa abordagem, Nayak *et.al.* [62] propuseram um método que tem como passo inicial um pré-processamento das imagens corrigindo diversos tipos de variação através da aplicação de um histograma de equalização adaptativo.

O próximo passo no método proposto por Nayak é extrair características relacionadas aos vasos sanguíneos, exsudato e textura. Para os vasos sanguíneos são utilizadas uma sequência de aberturas morfológicas com um elemento estruturante do tipo diamante no canal G da imagem para obter o traçado dos vasos. As características referentes a essas estruturas são representadas por seus perímetro e a área.

Os exsudatos, são detectados com uma adaptação do método contido em [97]. Primeiro, elimina-se os vasos sanguíneos através de fechamentos morfológicos utilizando um elemento estruturante em forma de octógono. Logo após, é calculada uma imagem de variância, que consiste em todas as regiões de brilho com variação nos valores de intensidade. Uma limiarização é então aplicada nessa imagem, resultando em uma imagem binária com os brilhos dos objetos e as bordas. Mais uma sequência de operações morfológicas é aplicada, seguida de uma remoção do disco óptico. Essa remoção se baseia no componente de intensidade das imagens de fundo de olho como descrito em [77]. Finalmente a área ocupada pelos exsudatos é computada pelo somatório de todos os pixeis brancos.

A textura da imagem é determinada baseada na medida da quantidade de variação da intensidade na imagem, utilizando matrizes de co-ocorrência [86].

Como última parte do método, a classificação é realizada através de uma rede neural, a qual utiliza as quatro características citadas anteriormente para classificar as imagens como em não retinopatia, retinopatia não proliferativa ou ainda retinopatia proliferativa.

Assim como Nayak. o método proposto por Yun *et.al.* [102] começa através de uma etapa de pré-processamento. Essa etapa conta com a utilização de histogramas de equalização, operadores morfológicos, e uma binarização. O histograma de equalização serve para aumentar o domínio dinâmico do histograma da imagem. O próximo passo é utilizar operadores morfológicos para suavizar o fundo, fazendo com que as veias, hemorragias e microaneurismas possam ser vistas claramente. Assim uma sequência de operações de abertura morfológica utilizando alternadamente operadores morfológicos em forma de diamante, e em forma de disco é realizada, sendo que, entre cada uma das operações é realizado um ajuste de intensidades na imagem. Ao fim da sequência de aberturas morfológicas, o perímetro e a área de estruturas características podem ser facilmente extraídas.

Os pixels do perímetro dos objetos são obtidos de uma imagem binária da forma descrita em [34]. A outra característica, a área, é determinada por uma limiarização da imagem transformando o fundo em preto e as características em branco. Isso é diferente do método de Otsu, utilizado em outras abordagens, uma vez que tal método escolhe o limiar

para converter uma imagem em tons de cinza para uma binária minimizando a variância entre as classes de pixeis pretos e brancos. Para finalizar a etapa de pré-processamento, um limiar tem que ser cuidadosamente escolhido para binarizar a imagem.

O próximo passo do método é a extração de seis características denominadas perímetro da camada vermelha **(PCVM)**, perímetro da camada verde **(PCVD)**, perímetro da camada azul **(PCAZ)**, área da camada vermelha **(ACVM)**, área da camada verde **(ACVD)** e área da camada azul **(ACAZ)**. Todas são baseadas nas imagens pré-processadas através das operações morfológicas. Perímetro refere-se ao número de pixeis presente na periferia das veias. Já a área é igual ao número de pixeis pretos presentes no interior das veias.

Por fim as imagens são classificadas por uma rede neural utilizando "backpropagation", a qual atribui às imagens um dos quatro estágios da RD: sem RD, RD moderada, RD severa e RD proliferativa.

O método apresentado por Usher [94] é baseado num processamento das imagens realizado em quatro etapas, seguido de uma classificação das amostras, a qual é gerada em duas etapas.

Como ponto de partida é realizado um pré-processamento, no qual é realizado um melhoramento adaptativo local do contraste na banda de intensidade para melhorar o contraste e normalizar a intensidade.

Identifica-se então as estruturas básicas da retina. O disco óptico, através da construção de uma imagem de variância, seguida pela identificação de picos de intensidade da imagem [77].

Já os vasos sanguíneos foram determinados utilizando um filtro Gaussiano multi-resolução com 16 orientações como descrito em [37].

Lesões brilhantes como os exsudatos são então extraídos utilizando uma combinação de regiões de crescimento recursivo [26] e limiarização adaptativa de intensidade [84]. Já as lesões escuras foram extraídas de um modo similar mas com o uso adicional de um operador especialmente desenvolvido para o melhoramento de arestas, denominado *operador Moat* [76].

Para cada uma das partes candidatas identificadas são então quantificados numericamente o tamanho, a forma, a matiz e a intensidade. Para cada uma das regiões candidatas, um especialistas atribui ainda o tipo de lesão que a região contém. Isso é usado como entrada para uma rede neural, a qual classifica cada imagem como normal ou anormal de acordo com a presença ou não de lesões.

Estes são alguns dos métodos propostos na literatura até hoje, relacionados à detecção de anomalias provenientes de RD. Nos próximos capítulos são apresentados alguns dos conceitos em que o método aqui proposto se baseia, bem como toda a metodologia científica aplicada no seu desenvolvimento.

Capítulo 4

Revisão Bibliográfica

O método proposto neste trabalho para a detecção de anomalias em IFOs faz uso de conceitos básicos da literatura de visão computacional, inteligência artificial e processamento de imagens. Este capítulo apresenta uma revisão bibliográfica destes conceitos.

4.1 Imagem

Existem diversas formas de se interpretar uma imagem, no entanto, a forma discreta parece bem apropriada ao contexto deste trabalho. Torres e Falcão [17] descrevem uma *imagem* $\hat{\mathbf{I}}$ de uma forma discreta, como um par ordenado $(\mathrm{D}_I, \overrightarrow{I})$ onde:

- D_I é um conjunto finito de pixeis (pontos em $\mathbb{N}^2$, isto é, $\mathrm{D}_I \subset \mathbb{N}^2$), e
- $\overrightarrow{I} : \mathrm{D}_I \rightarrow \mathbb{R}^n$ é uma função que associa para cada pixel p em D_I um vetor $\overrightarrow{I}(p) \in \mathbb{R}^n$ (*e.g.* $\overrightarrow{I}(p) \in \mathbb{R}^3$ quando uma cor no sistema RGB é associada a um pixel).

4.2 Pontos Característicos em Imagens

Uma situação muito comum em problemas de visão computacional pode ser descrita da seguinte forma: encontrar de forma independente e sem conhecimento *à priori* um mesmo ponto, presente em duas imagens da mesma cena, cena esta que é vista de pontos de observação diferentes. Tais pontos são chamados de *pontos característicos* **(PC)** da imagem e possuem aplicações nos mais variados campos: recuperação de imagens em grandes bases de dados, reconhecimento baseado em modelo, recuperação de objetos em vídeo, dentre outros [56, 74, 91].

Pontos característicos têm como principais propriedades serem invariantes a diferentes pontos de observação e operações geométricas simples tais como rotação e escala, como visto na Figura 4.1.

Ao redor do ponto característicos pode ser definida uma região, sendo esta de tamanho variável, denominada uma *região de interesse*.

Figura 4.1: Ponto característico encontrado na imagem de uma mesma cena vista de diferentes pontos de observação.

Existem diversos detectores de pontos citados na literatura, porém, um dos mais citados é o detector de pontos de *Hessian* [56].

4.2.1 Detector de Hessian

O *detector de Hessian* representa uma classe de detectores denominada detectores de "blobs" capaz de identificar esse tipo de regiões em imagens. "Blobs" indicam uma região que difere (por ser mais claro ou escuro) da região ao seu redor. A segunda matriz 2×2 obtida da expansão de Taylor da função de intensidade I(x) de uma imagem $\hat{I}$ é a matriz Hessiana:

$$\mathcal{H}\left(\mathrm{x}, \sigma\right) = \begin{bmatrix} I_{xx}\left(\mathrm{x}, \sigma_D\right) & I_{xy}\left(\mathrm{x}, \sigma_D\right) \\ I_{xy}\left(\mathrm{x}, \sigma_D\right) & I_{yy}\left(\mathrm{x}, \sigma_D\right) \end{bmatrix} \tag{4.1}$$

sendo que I_{xx}, etc. representam a convolução de uma imagem $\hat{\mathrm{I}}$ no ponto x com uma derivada de segunda ordem de uma Gaussiana [92].

Baseados no determinante e no traço dessa matriz, muitos filtros interessantes foram desenvolvidos, sendo que o último e mais frequentemente referenciado é o Laplaciano. Além disso, o máximo local de ambas as medidas (determinante e traço) podem ser usados para se detectar estruturas de "blobs" em uma imagem [4] identificando dessa forma os pontos característicos.

4.3 Descritores de Imagens

Descritores de imagens são uma forma utilizada para se caracterizar uma imagem a partir de informações de baixo nível como cor, textura ou forma de objetos contidos na imagem [17, 65, 85].

Segundo Torres e Falcão [17] o descritor de uma imagem é composto por duas partes:

1. Um algoritmo de extração capaz de codificar as características da imagem em um vetor de características.

2. Uma medida de similaridade para comparar duas imagens baseada em uma métrica de distância, representa o grau de similaridade entre os vetores de características das imagens.

Formalmente um descritor D de uma imagem $\hat{I}$ pode ser definido como um par ordenado (ϵ_D,δ_D) [64]

- $\epsilon_D : \left\{\hat{\mathrm{I}}\right\} \rightarrow \mathbb{R}^n$ é uma função a qual extrai o vetor de características $\overrightarrow{v_{\hat{\mathrm{I}}}}$ de uma imagem $\hat{\mathrm{I}}$;

- $\delta_D : \mathbb{R}^n$ x $\mathbb{R}^n \rightarrow \mathbb{R}$ é uma função de similaridade que calcula o grau de similaridade entre os vetores de características de duas imagens, similaridade essa que é inversamente proporcional à distância dos vetores $\overrightarrow{v_{\hat{\mathrm{I}}}}$ das imagens.

Um *vetor de características* $\overrightarrow{v_{\hat{\mathrm{I}}}}$ de uma imagem $\hat{\mathrm{I}}$ é um conjunto de pontos $\overrightarrow{v_{\hat{\mathrm{I}}}} = (v_1, v_2, ..., v_d)$, onde $v_i \in \mathbb{R}$ e d representa a dimensão do vetor.

A representação esquemática de um descritor de imagens pode ser visto na Figura 4.2 Descritores de imagens podem ainda ser divididos em dois grandes grupos: globais e locais.

De uma forma bem resumida, os descritores globais são caracterizados por levarem em consideração informações de padrão global como cor, forma, e textura, sem no entanto, focar sua análise em pontos específicos. Exemplos dessa classe de descritores podem ser encontrados em [88, 87, 68, 75, 86, 31, 61, 103, 60, 13, 43, 7, 42].

Já os descritores locais são descritos de forma mais completa na próxima seção.

4.4 Descritores Locais

Descritores locais computados a partir de regiões de interesse têm sido muito empregados nos mais diversos tipos de problemas [58]. Áreas como reconhecimento de objetos [25], montagem de panoramas [10], recuperação de imagens [55] dentre outras, se alguns dos exemplos nos quais os descritores locais podem ser empregados.

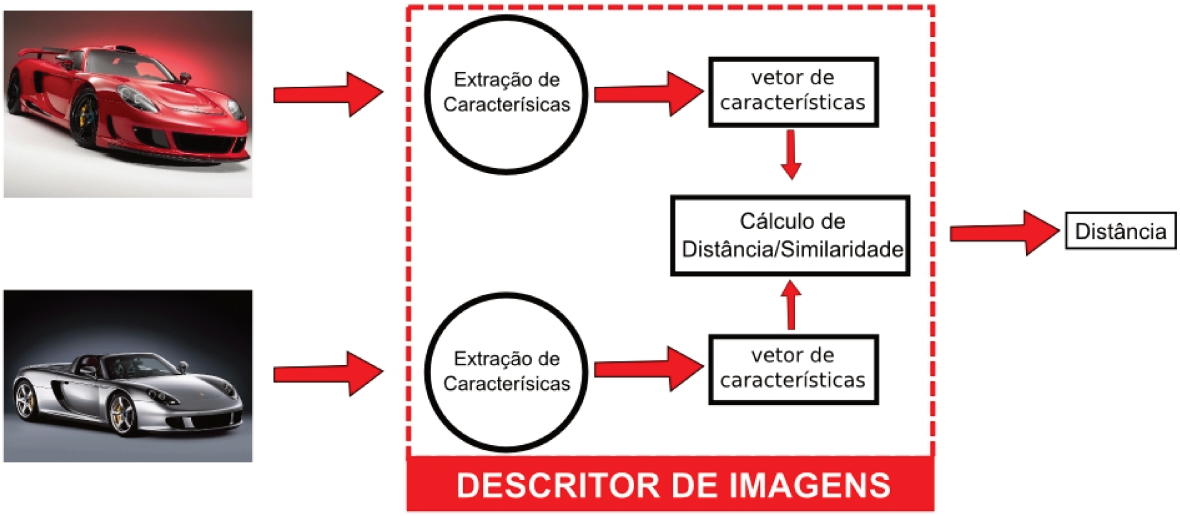


Figura 4.2: Exemplo do funcionamento de um descritor de imagens. Adaptado de [64].

Dado um detector de regiões de interesse, deve-se escolher o descritor mais apropriado para a região, sendo que a escolha do descritor depende do detector de regiões. Há um grande número de possíveis descritores para essas regiões, os quais, enfatizam diferentes propriedades da imagem como intensidade dos pixels, cor, textura, arestas, etc.

Enquanto o descritor global gera um único vetor de características d-dimensional para representar uma imagem (gerando uma representação de $1 \times d$ dimensões), um descritor local irá extrair s pontos característicos da imagem, e gerar um vetor d-dimensional para cada ponto (gerando uma representação de $s \times d$ dimensões).

4.4.1 SIFT

O descritor denominado *Scale Invariant Features Transform* **(SIFT)** [49] é um dos descritores locais mais citados na literatura de visão computacional. Seu grande emprego se dá devido a sua propriedade de ser fortemente invariante à mudanças de escala e orientação, fato que faz com que seja empregado nas mais diversas áreas [67, 80, 57].

Seu algoritmo possui quatro etapas básicas [49, 18]:

- *detecção de extremos no espaço-escala*: nesta etapa o algoritmo busca pontos candidatos, os quais devem ser invariantes à mudanças de escala da imagem. Isso é realizado buscando por características estáveis em todas as possíveis escalas, usando uma função contínua de escala denominada *espaço escala* [100], o qual pode ser implementado utilizando uma função de *diferença de Gaussianas* **(DoG)** [49]

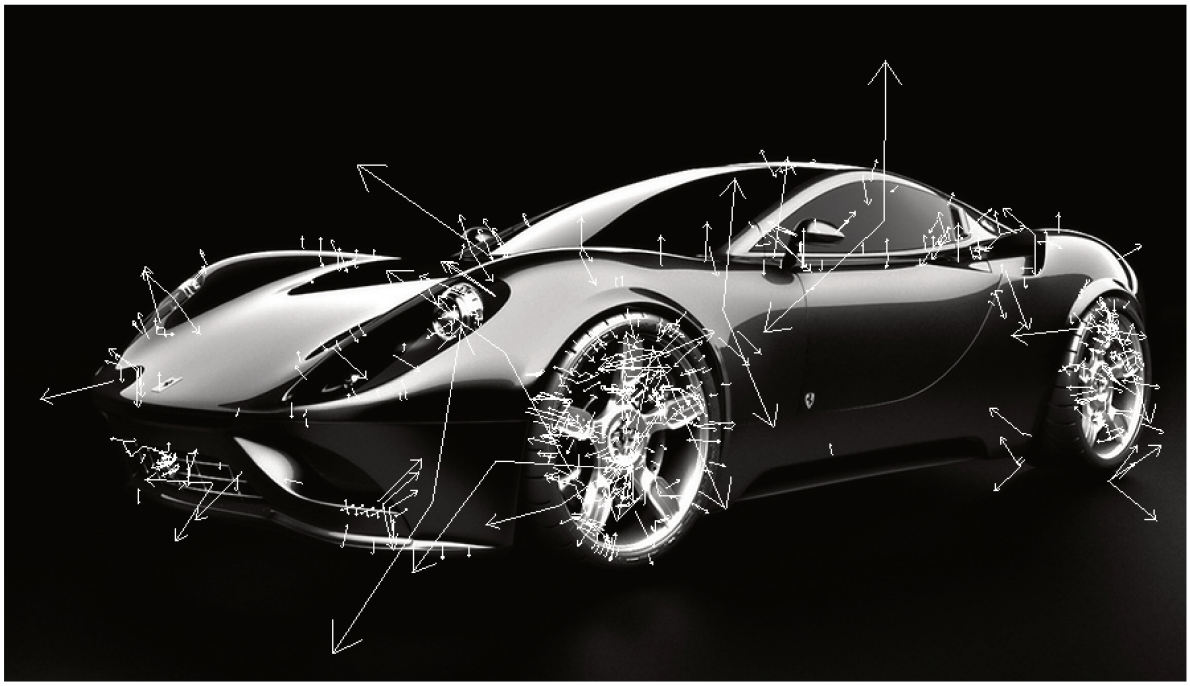


Figura 4.3: Pontos característicos extraídos com o algoritmo SIFT.

- *localização dos pontos característicos*: para cada ponto candidato sua localização e escala são refinadas utilizando uma aproximação baseada na expansão de Taylor [22] (até os termos quadrados) da função de espaço . Pontos característicos são selecionados com base em medidas de sua estabilidade, fazendo com que pontos de baixo contraste e de borda sejam eliminados.

- *associação de orientação*: a cada PC é associada uma ou mais orientações baseada nas direções dos gradientes locais da imagem. Todas as futuras operações são então realizadas sobre a região do PC em sua escala e orientação. Isso confere ao descritor invariância a essas transformações.

- *descritor do ponto característico*: os gradientes da imagem local são medidos numa região de tamanho determinado ao redor de cada PC. Tais medidas são ponderadas por uma janela gaussiana, transformando a representação de modo que ela se torne invariante a distorções locais de forma e mudanças de iluminação.

A Figura 4.3 mostra os PCs localizados pelo SIFT em uma imagem.

4.4.2 SURF

Assim como o SIFT, o descritor denominado *Speed-Up Robust Features* **(SURF)** é um tipo de descritor local muito utilizado para questões de visão computacional tais como registro de imagens, calibração de câmeras, reconhecimento de objetos dentre outros [3, 35].

Assim como outros descritores locais, o SURF também busca características como a invariância a rotação e escala. Pode ser descrito em três etapas:

- *detecção dos PCs*: os pontos são detectados baseados numa aproximação do detector de Hessian, a qual utiliza *filtros caixa* [32]. Aliado ao uso de *imagens integrais* [96], tal aproximação reduz muito o custo computacional;

- *representação do espaço escala*: diferente do espaço escala do do algoritmo SIFT, o qual é construído através da função DoG, o espaço escala utilizado pelo SURF é construído através da aplicação de filtros caixa aumentados iterativamente e aplicados sobre a imagem original. Isso ocorre sem que a velocidade de aplicação do filtro seja influência por seu tamanho, graças ao uso das imagens integrais;

- *localização dos PCs*: para localizar os pontos característicos em uma imagem cobrindo todas as possíveis escalas, uma supressão não máxima é aplicada ao redor de uma vizinhança utilizando uma variante do método proposto por [63]. A máxima do determinante da matriz Hessiana é então interpolada no espaço escala da imagem com o método proposto por [11];

- *associação de orientação*: a orientação dos PCs no SURF se baseia nos valores de resposta gerados pela aplicação das waveletes de Haar [51] sobre o ponto nas direções x e y. Tais valores são então pesados através de uma Gaussiana centrada no PC e representados como pontos ao longo das ordenadas (para respostas na direção y) e das abcissas (para respostas na direção x). Assim, a orientação dominante é estimada calculando o somatório de todas as respostas dentro de uma janela deslizante de orientação e tamanho $\frac{\pi}{3}$;

- *descrição dos pontos característicos*: para descrição do PC, o primeiro passo é o cálculo dos valores de respostas das waveletes de Haar nas direções horizontal e vertical em uma região ao redor do PC, levando em consideração que os termos horizontal e vertical são definidos em relação a orientação do PC. Para aumentar a robustez do descritor os valores de resposta recebem pesos através de uma Gaussiana centrada no ponto característico. Os descritores são então compostos baseados no valor dos somatórios destas respostas nas direções horizontal e vertical, bem como no valor absoluto desses somatórios, levando em consideração uma região de tamanho fixo ao redor do PC.

Figura 4.4: Pontos característicos extraídos com o algoritmo SURF.

A Figura 4.4 mostra os pontos característicos localizados pelo SURF em uma imagem.

4.5 Aprendizado de Máquina

O termo aprendizado de máquina define de forma ampla a classe de algoritmos que conseguem melhorar sua performance através do ganho de algum tipo de experiência. A questão de aprendizado é definida formalmente por Mitchell [59] da seguinte forma,

Definição: um algoritmo é dito *aprender* através da experiência **E** atuando sobre uma classe de problema **T** e medidas de performance **P**, se essa performance **P** em relação ao problema **T**, melhora com a experiência **E**.

Uma sub-área de aprendizado de máquina de especial importância que vem crescendo muito rapidamente é denominada *reconhecimento de padrões*.

4.5.1 Reconhecimento de Padrões

O emprego de máquinas capazes de reconhecer padrões é foco de muitos estudos na atualidade uma vez que esse tipo de reconhecimento se torna cada vez mais necessário no nosso dia a dia, fazendo parte de áreas como identificação de digitais, identificação de sequencias de DNA, detecção anomalias em imagens médicas e diversas outras [69].

Formalmente, um sistema de reconhecimento de padrões, no contexto de aprendizado de máquina, é responsável por associar classes (normalmente em forma de rótulos) a objetos. *Classe* é o nome dado a um conjunto de objetos com as mesmas características. *Objeto* é o nome dado a um conjunto de medidas chamadas de *características* ou *atributos* [45].

Pode-se separar os métodos de reconhecimento de padrões em

- *métodos não supervisionados*: nesse tipo de método, o algoritmo não tem nenhuma informação prévia sobre as classes a que os objetos pertencem. Um exemplo são alguns métodos de agrupamento;

- *métodos supervisionados*: nesse tipo de método, o algoritmo deve passar por uma etapa conhecida como *treinamento*, na qual o classificador escolhido aprende um determinado padrão para o tipo de dados com o qual ele está lidando, baseado em uma parte dos dados, denominada *conjunto de treinamento*.

Uma visão geral de uma das possíveis configurações para sistemas de reconhecimento de padrões pode ser visto na Figura 4.5.

Um método de reconhecimento de padrões está fortemente ligado ao sistema de extração de características utilizado para representar os objetos. Quanto melhor é o método utilizado para se extrair as características e dessa forma representar os objetos, mais trivial pode ser a forma abordada para o reconhecimento de padrões. No entanto, representações pobres dos objetos podem exigir um reconhecimento de padrões mais robustos.

4.5.2 Agrupamento

Um *agrupamento* é um tipo de classificação imposta a um conjunto finito de objetos que busca separar os elementos por grupos onde todos os elementos de um mesmo grupo possuam características parecidas. Cada elemento desse conjunto é denominado um *ponto* (também chamado padrão). A distância entre um par de pontos, que indica o grau de similaridade entre esse par, pode ser medida através de uma métrica de distância [19](*e.g.* a distância Euclidiana) [39, 89, 45]. De maneira mais intuitiva, os pontos pertencentes a um grupo deveriam ser similares entre si e não similares aos membros de um outro grupo como visto na Figura 4.6.

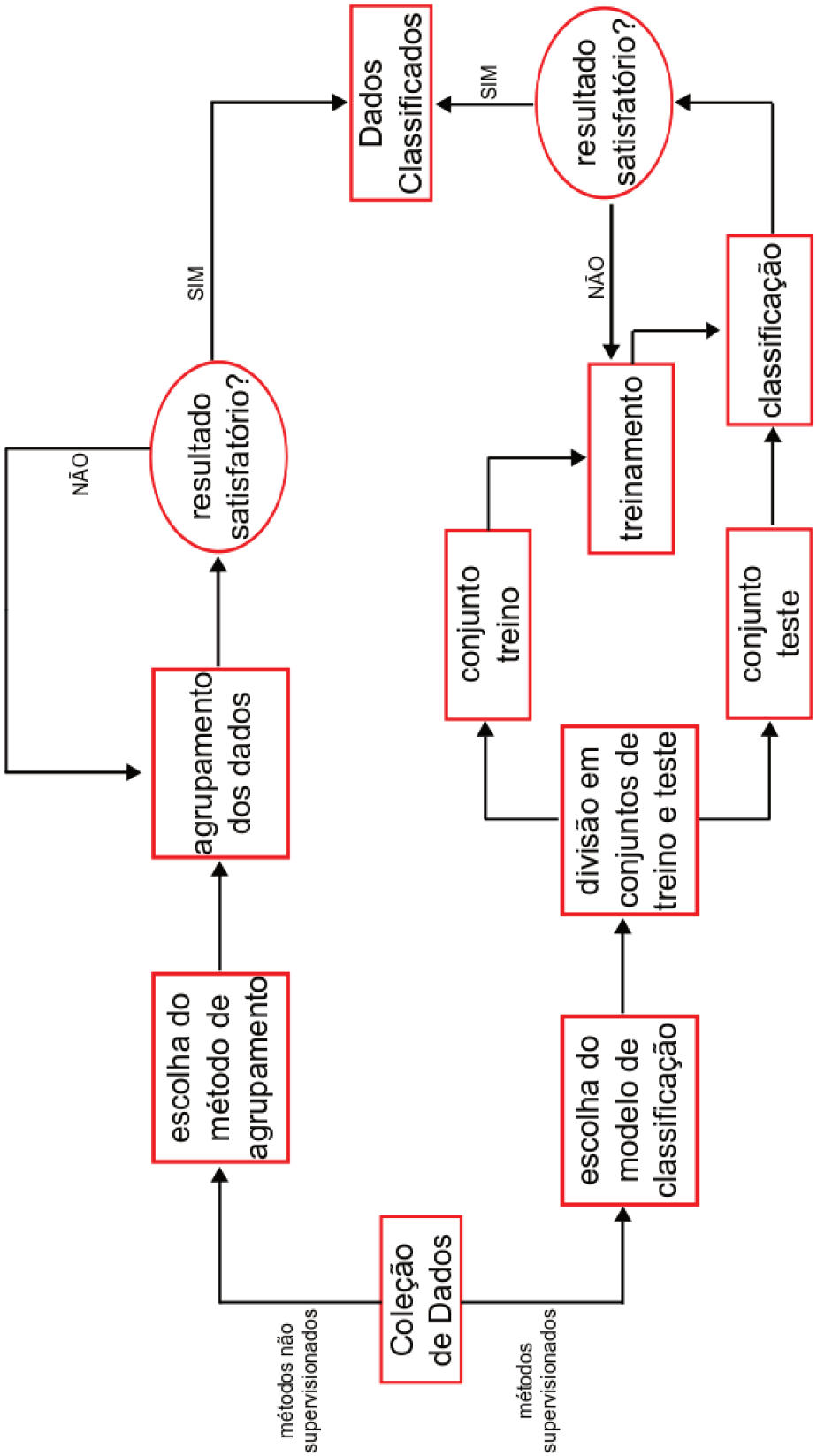


Figura 4.5: Exemplo de uma das possíveis configurações para sistemas de reconhecimento de padrões.

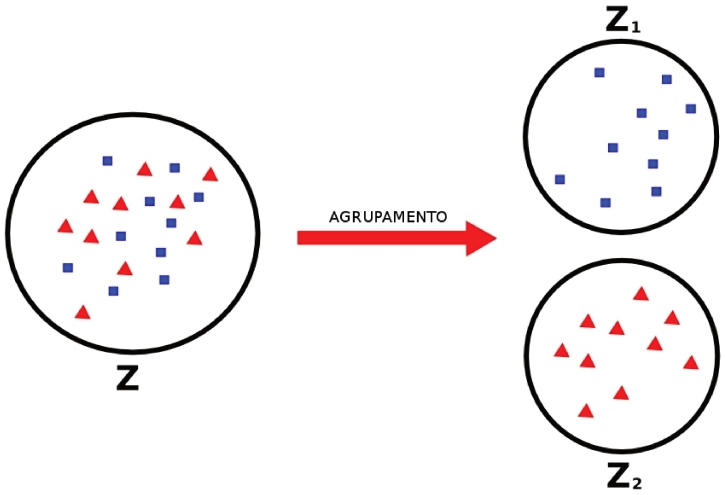


Figura 4.6: Exemplo de um agrupamento ideal: após o agrupamento os elementos de um conjunto são apenas similares entre si e diferentes dos elementos de qualquer outro conjunto.

Um algoritmo para agrupamento, segundo Kuncheva [45], opera em um conjunto de dados não classificados $\mathbf{Z}$ e produz uma partição denotada por $\mathbf{P} = \left(\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(c)}\right)$, onde $\mathbf{Z}^{(i)} \subseteq \mathbf{Z}$ e

$$\mathbf{Z}^{(i)} \cap \mathbf{Z}^{(j)} = \emptyset, \qquad i, j = 1, \ldots, c, i \neq j \tag{4.2}$$

$$\bigcup_{i=1}^{c} \mathbf{Z}^{(i)} = \mathbf{Z} \tag{4.3}$$

Um importante conceito dentro do escopo de agrupamento é o de *agrupamento baseado em protótipos*. Nesse tipo de método, os pontos de um conjunto são mais similares ao protótipo que define o grupo do que aos protótipos que definem os outros grupos. Para dados com atributos contínuos o protótipo do grupo é geralmente o *centroide* (*i.e.* a média de todos os pontos no grupo). Quando um centroide não é totalmente significativo, ou ainda quando os dados tem atributos categóricos, o protótipo é geralmente o *medóide* (*i.e.* o ponto mais representativo do grupo) [89].

É importante não se confundir um método de agrupamento com um algoritmo de agrupamento. Um *método* de agrupamento especifica a estratégia geral para se agrupar os pontos. Um *algoritmo* de agrupamento por outro lado, é uma sequência de etapas que pode ser utilizada em um programa de computador que implementa o método e incorpora várias *heurísticas*.

Os métodos de agrupamento podem ainda ser divididos em agrupamento hierárquico e o agrupamento particional [89].

4.5.3 K-Médias

O algoritmo intitulado *K-Médias* é um algoritmo de agrupamento particional baseado em protótipos que tenta dividir o conjunto de pontos iniciais em um número específico (K) de grupos. Ele define um protótipo em termos de um centróide, o qual é geralmente a média dos pontos pertencentes ao grupo.

De forma geral, na execução do algoritmo k-médias é necessário se realizar seis passos básicos:

1. Organizar os dados em uma matriz de m linhas por n colunas. (pré-processamento).
2. Selecionar os K centroides iniciais dos grupos.
3. Gerar uma partição associando cada ponto ao grupo cuja distância do ponto ao centróide do grupo seja a menor dentre todos os grupos.
4. Calcular o valor do novo centróide de cada grupo.
5. Repetir os passos 3 e 4 até que o valor ótimo de uma função critério seja encontrado.
6. Ajustar o número de grupos agrupando ou repartindo os grupos existentes.

De forma mais clara, o k-médias pode ser escrito como o algoritmo 1.

```
1: selecione K pontos, dentre o conjunto inicial de pontos, como centróides iniciais.
2: repeat
3:    gerar K grupo associando cada ponto ao centróide mais próximo
4:    recalcular o centroide de cada grupo
5: until centroides não mudarem ou atingir um número limite de iterações
```

Algoritmo 1: algoritmo básico do k-médias

4.5.3.1 Organização dos Dados

Dado um conjunto **Z** formado de m pontos onde cada ponto possui dimensão igual a n, para se dividir tal conjunto em K grupos, é necessário se organizar os pontos de **Z** em forma de uma matriz de pontos. Tal matriz será composta por m linhas, correspondendo ao número total de pontos do conjunto, e n colunas, de modo que cada coluna da matriz represente uma dimensão dos dados.

4.5.3.2 Seleção dos Centróides Iniciais

Os primeiros centroides utilizados no algoritmo serão considerados as sementes iniciais para a geração da primeira partição dos dados. Assim, os K pontos podem ser selecionados das seguintes formas:

1. Selecionando-se as K primeiras linhas da matriz de pontos.
2. Selecionando-se K linhas da matriz de pontos de forma aleatória.

Sementes iniciais diferentes podem gerar partições diferentes para o conjunto de dados, o que permite grupos finais diferentes, uma vez que o algoritmo é baseado no erro quadrado podendo convergir para um mínimo local. Isso ocorre principalmente se os grupos não estiverem bem separados [39].

4.5.3.3 Partição dos Dados

Para se particionar um conjunto de dados, é necessário se associar cada um dos pontos pertencentes a esse conjunto a um dos K centroides existentes. Para isso é necessário se definir uma medida que quantifique a noção de próximo entre o ponto analisado e cada um dos centroides. A distância Euclidiana (também chamada de distância L2) é uma das mais usadas no algoritmo K-médias utilizando pontos no espaço Euclidiano [89]. Entretanto, existem outros tipos de medidas de proximidade que são apropriadas para outros tipos de dados. Um exemplo de outra medida de proximidade utilizada no algoritmo K-médias é a distância de Mahalanobis [101].

Definida uma medida de similaridade, é calculada a proximidade do ponto até cada um dos centroides dos grupos. O ponto é atribuído então ao grupo cuja distância entre o centróide e o ponto é a menor dentre todos os centroides analisados. Quando tal análise é realizada uma vez para cada um dos pontos, tem-se um *ciclo*.

4.5.3.4 Atualização dos Centróides

Segundo Kuncheva [45], a atualização dos centroides pode ser realizada em dois momentos:

- no momento em que um novo ponto é adicionado ao grupo;
- ao final de cada ciclo;

Independente do momento dessa atualização, ela depende da medida de proximidade dos dados e do objetivo do agrupamento. Esse objetivo é expresso por uma função denominada *função objetivo* a qual depende da proximidade entre o ponto e os demais ou entre o ponto e o centróide do grupo (*e.g.* minimizando a distância quadrada de cada ponto para o centróide mais próximo).

A função objetivo mede a qualidade do agrupamento. No caso mais comum, o que inclui o uso da distância Euclidiana como medida de proximidade, a função de similaridade utilizada é o *somatório dos erros quadrados* **(SSE)** definida formalmente como

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{x \in C_i} \mathrm{D}\left(c_i, \mathrm{x}\right)^2, \tag{4.4}$$

onde D é a distância Euclidiana, C_i corresponde ao i-ésimo grupo e c_i representa o centróide do mesmo.

Tan *et.al.* [89] afirma que o centróide que minimiza o SSE é a média. Isso implica que o novo centróide de C_i seja calculado como

$$c_i = \frac{1}{e_i} \sum_{\mathrm{x} \in C_i} \mathrm{x}, \tag{4.5}$$

sendo e_i o número total de elementos do grupo i.

4.5.3.5 Critério de Convergência

O critério de convergência é a condição que determina o final da execução do algoritmo. No K-médias, cada um dos pontos possui um rótulo associado ao grupo a que o ponto pertence. Assim, o critério adotado na maioria das vezes para o K-médias é a não modificação desses rótulos. Em outras palavras, se o rótulo dos pontos permanece imutável por um determinado número de iterações, o critério de convergência foi alcançado e o algoritmo termina.

No entanto, este critério pode levar muito tempo para ocorrer, exigindo desta forma o consumo de um tempo muito grande na execução do algoritmo. Uma alternativa é determinar que ao invés do rótulo de todos os pontos ter que permanecer imutável para que o algoritmo termine, apenas uma determinada porcentagem satisfaça essa condição. Ou ainda, que um número máximo de iterações delimite o algoritmo.

4.5.4 Classificadores

Existem diversas formas para se definir um classificador. No domínio estatístico matemático, um classificador Θ é definido como uma função [45]

$$\Theta : \mathbb{S}^d \to \Omega, \tag{4.6}$$

onde $\mathbb{S}^d$ denota um espaço de características d-dimensional e Ω denota o conjunto de rótulos das c classes do problema $\Omega = \{\omega_1, ..., \omega_c\}$.

Tomando como referência o trabalho de Rocha [70], pode-se dizer que no campo da Inteligência Artificial, um classificador é um tipo de motor de inferência que implementa estratégias eficientes para computar relações de classificação entre pares de conceitos ou para computar relações entre um conceito e um conjunto de instâncias.

No contexto de classificação de dados, existem alguns termos importantes que precisam ser definidos antes de qualquer coisa[44]:

- precisão: a precisão de um classificador é a probabilidade de ele classificar corretamente uma amostra selecionada aleatoriamente.
- algoritmo de indução: é o algoritmo que constrói o classificador a partir de um conjunto de dados conhecido.
- "holdout": também chamado de estimação das amostras de teste, consiste em particionar os dados em dois sub-conjuntos mutuamente exclusivos chamados conjunto de treino e conjunto de teste, ou conjunto "holdout". Quando trata-se de uma estimação das amostras de teste aleatória, o método de partição é repetido varias vezes e a precisão estimada é derivada da média do número de vezes de execução. Por sua vez, o desvio padrão pode ser estimado com base na média da precisão.

4.5.4.1 Validação Cruzada

A técnica denominada validação cruzada, é utilizada dentro do contexto de aprendizado de máquina com o intuito de validar um algoritmo de indução para ser eficaz em qualquer conjunto de teste.

Uma validação cruzada de k dobras, consiste em dividir o conjunto de dados aleatoriamente em k sub-conjuntos mutuamente exclusivos (as dobras) aproximadamente do mesmo tamanho.

O algoritmo de indução é treinado e testado k vezes: cada uma das vezes utiliza k-1 conjuntos para treino e um conjunto para teste[44]. O resultado final do algoritmo é obtido através da média dos resultados destas k execuções.

Segundo Breiman e Spector [9], a escolha de dividir o conjunto de dados em cinco dobras gera bons resultados, superando a escolha mais comum de 10 dobras.

4.5.4.2 SVM

Maquinas de vetores de suporte ou, **SVM** como é mais conhecido, é um método sistemático, reproduzível, e baseado na teoria estatística de aprendizado. Seu treinamento

envolve o uso de uma função de custo convexo na qual não há falsos mínimos locais. Ele utiliza o conceito de substituição de "kernels" sendo referenciado de maneira ampla como um método de kernel [5].

De maneira informal, pode-se definir que o objetivo do SVM é encontrar um hiperplano capaz de separar linearmente as amostras de um problema com a maior margem de separação possível [38].

No entanto, há casos em que um hiperplano gerado por funções discriminantes lineares simples (como uma reta) não funcionam bem no espaço de características original. Nesses casos é necessário se expandir o espaço das amostras para uma dimensão maior. Com isso algoritmos de classificação linear podem ser aplicados no espaço expandido gerando funções não lineares no espaço de características original. Para realizar esse tipo de mapeamento o SVM, utiliza dois artifícios

- o primeiro é o uso de uma *margem de maximização*, C, evitando o aparecimento de "overfiting" no mapeamento dos dados;
- o segundo é a utilização de "kernels" (Ke), os quais realizam o mapeamento de pontos para um novo espaço de forma prática como pode ser visto no trabalho de Bennett e Campbel [5].

Dados dois pontos representados por seus vetores u e v dois dos "kernels", **Ke**, mais comuns na literatura são:

- *linear* $\mathbf{Ke}(u, v)$: $(u.v + 1)^d$;
- *RBF* $\mathbf{Ke}(u, v)$: $exp\left(-\frac{||u-v||^2}{2\sigma}\right)$

onde d e σ são parâmetros do "kernel".

4.6 Curvas ROC

A curva ROC foi desenvolvida no contexto de detecção de sinais eletrônicos e problemas com radares, durante a Segunda Guerra Mundial [104], e hoje em dia é utilizada em áreas como medicina, economia, computação dentre outras.

A curva ROC original é baseada em duas medidas de probabilidade denominadas *sensitividade* (S_E) e *especificidade* (E_S) e apresenta através de um grafico o equilíbrio de diferentes limiares entre obter uma maior taxa de verdadeiros positivos (sensitividade) com um custo adicional de falsos positivos (1 − especificidade)[30].

A S_E é uma medida de probabilidade baseada em valores de verdadeiro positivo **(VP)** e falsos negativos **(FN)**. Verdadeiros positivos indicam quando um resultado indica positivo dado que o resultado esperado deva ser realmente positivo. Já os falsos negativos indicam quando um resultado deveria indicar positivo e no entanto, ele indica negativo. A Equação 4.7 mostra o calculo da sensitividade.

$$S_E = \frac{VP}{VP + FN} \tag{4.7}$$

Já a E_S é uma medida de probabilidade baseada em valores de verdadeiro negativo **(VN)** e falsos positivos **(FP)**. Verdadeiros negativos indicam quando um resultado indica negativo dado que o resultado esperado deve ser realmente negativo. Já os falsos positivos indicam quando um resultado deveria indicar negativo no entanto ele indica positivo. A Equação 4.8 mostra o calculo da especificidade.

$$E_S = \frac{VN}{VN + FP} \tag{4.8}$$

S_E e E_S não são calculadas sobre os mesmos dados, ou seja, S_E é calculada sobre um conjunto de dados que deveriam conter algum tipo de anomalia (*e.g.* no contexto médico no qual curvas ROC são muito utilizadas esse conjunto é usualmente representado por indivíduos doentes de uma população). Já a E_S trata do conjunto que deveria estar livre de qualquer anomalias (*e.g.* indivíduos saudáveis pertencentes a uma população) [53].

Todas as curvas começam no canto inferior esquerdo, representando um limiar no qual todos os casos sao classificados como negativos, e terminam no canto superior direito representando o limiar no qual todos os casos são classificados como positivo. As melhores curvas são aquelas mais próximas ao canto superior esquerdo do gráfico (se uma curva está posicionada sobre outra em um mesmo ponto no eixo horizontal, a curva situada mais acima é melhor, detectando mais verdadeiros positivos enquanto gera a mesma porcentagem de falsos positivos exibida na curva situada abaixo como pode ser visto na Figura[1] 4.7).

Uma das características mais interessantes das curvas ROC é que elas permitem a comparação de dois modelos independente do limiar utilizado para cada modelo. Quando a curva de um modelo A está completamente sobre a curva de outro modelo B, isso indica de forma clara que o modelo A é melhor que o modelo B independente do limiar utilizado. Mas se duas curvas possuem uma interseção, a escolha do melhor modelo depende do limiar utilizado.

[1]Figura retirada de http://commons.wikimedia.org/ em 05 de agosto de 2010

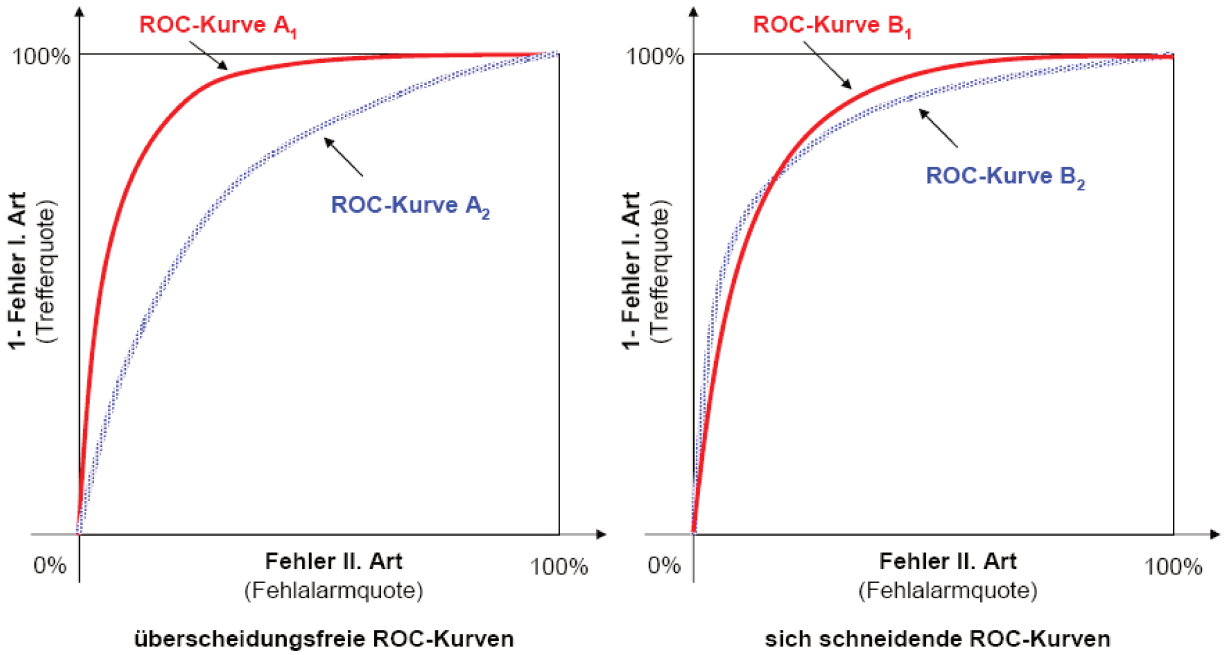


Figura 4.7: Exemplos de curvas ROC. Na imagem da esquerda a curva em vermelho é sempre melhor do que a curva em azul. Já na imagem da direita, a curva em azul se sobrepõe a curva em vermelho em uma parte do gráfico, indicando que nessa parte a curva em azul representa os melhores resultados.

Capítulo 5

Dicionários Visuais

Apesar de serem uma boa opção para descrever imagens de forma mais detalhada, o uso de descritores locais pode se tornar uma grande dificuldade no momento da classificação das amostras. Na forma usual como são utilizados, os classificadores não conseguem trabalhar bem com representações geradas por descritores locais. Isso porque tais descritores não geram o mesmo número de pontos para todas as imagens, fazendo com que os classificadores tenham dificuldades para trabalhar com este tipo de descritores.

Devido ao fato de descritores locais possuírem funções de distância complexas o classificador SVM, por exemplo, precisa de "kernels" muito específicos, como os descritos no trabalho de Eichhorn [23], que consigam lidar com descritores locais. Outros classificadores baseados em distâncias explícitas como o algoritmo *K Vizinhos Mais Próximos* **(KNN)** [45], utilizam-se de medidas de distância especiais como a *Earth Mover's Distance* **(EMD)** [73] para tratar este tipo de descritores. No entanto, estas alternativas empregadas pelos classificadores para lidar com dados provenientes de descritores locais podem possuir um alto custo computacional, chegando até a impossibilitar uma classificação em tempo real.

Uma possível solução para esse tipo de dificuldade é o uso de *dicionários visuais*. Eles são baseados em representações utilizadas para processamento de texto, nas quais um documento é visto apenas como uma coleção de palavras, sendo que a ordem em que as palavras aparecem no texto não é considerada [29, 6]. As palavras que representam o texto pertencem a um vocabulário fixo determinado de acordo com a classe de cada texto (e.g. textos relacionados a esportes possuem um vocabulário diferente de textos relacionados a política).

Analogamente, dicionários visuais são uma forma de representação robusta para imagens onde cada imagem é vista apenas como uma coleção de regiões nas quais a informação espacial da região não importa. A única informação considerada é a aparência da região [79].

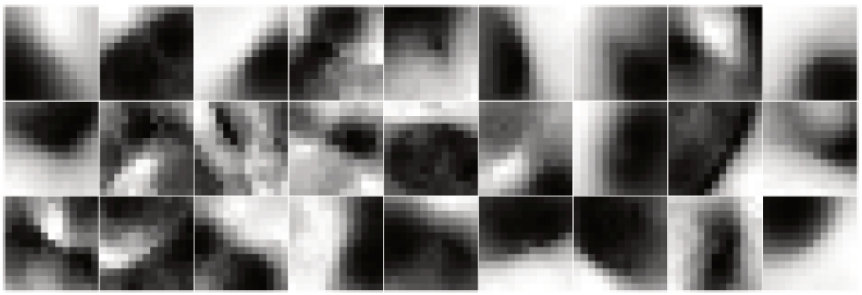


Figura 5.1: Exemplo de palavras visuais. Imagem reproduzida de [71] com autorização do autor.

O objetivo da construção de um dicionário visual é aprender, a partir de um conjunto de treinamento, o modelo gerador [93] capaz de selecionar as r regiões mais representativas do problema, de forma que o número de regiões selecionadas deve ser grande o suficiente para distinguir mudanças relevantes nas imagens, mas não tão grande a ponto de distinguir variações irrelevantes como ruído [48, 15]. Tais regiões compõem um dicionário visual, o qual pode ser visto como uma representação de um espaço de Hilbert [33] $\mathbf{H}$ d-dimensional, sendo que cada uma das regiões representada ao mesmo tempo uma *palavra visual* [95] no dicionário e uma dimensão no espaço. Um exemplo de palavras visuais pode ser visto na Figura 5.1.

De posse de um dicionário visual, a imagem é então descrita baseada no conjunto das palavras visuais que a mesma contém, ou seja, mapeando pontos contidos em um espaço de entrada qualquer n-dimensional Φ (representado inicialmente por todas as regiões que a imagem contém) para o de Hilbert $\mathbf{H}$ (representado pelas palavras do dicionário visual).

O maior desafio dessa técnica é a construção de um dicionário que possa representar de forma robusta as imagens do domínio do problema.

5.1 Sacola de Palavras Visuais

Através do emprego de dicionários visuais é possível se adaptar técnicas antes utilizadas para o processamento de textos como LDA [6] e pLSA [36], e utilizá-las em problemas de categorização e segmentação de imagens [46].

Assim, uma imagem é considerada como um simples conjunto de palavras visuais. Denominada *sacola de palavras visuais* [95], essa técnica simplifica a maneira de descrever imagens, considerando-as simplesmente como um histograma de palavras visuais, ignorando informações como a localização espacial das palavras.

5.1.1 Construção das Sacolas

Após definidos os conceitos de dicionários visuais, palavra visual, e sacolas de palavras visuais se faz necessário um detalhamento do processo de construção de uma sacola de palavras.

A primeira parte do processo se dá com a extração e descrição de regiões da imagem. Uma região pode ser representada por pontos característicos ou pequenas partes de tamanho fixo da imagem denominadas "patches". No caso da representação via pontos característicos, algoritmos importantes da literatura como o SIFT ou o SURF podem ser utilizados. No caso da utilização de "patches", é necessária a extração de partes de tamanho fixo da imagem, e descrição através de descritores de imagens conhecidos na literatura. Ao fim dessa etapa cada imagem será descrita como um conjunto de r regiões sendo cada região descrita por um vetor de características com d dimensões.

De posse das regiões já descritas, é necessário realizar a escolha das p palavras visuais para compor o dicionário, observando três tópicos importantes na literatura:

1. dicionários visuais são tipicamente criados a partir de um conjunto de imagens de treinamento [41];

2. a escolha do número de palavras visuais que irão compor o dicionário varia de acordo com o problema tratado, podendo chegar à 1.2M [66]. Há uma grande variação na literatura sobre o número correto de palavras;

3. a escolha das palavras é geralmente realizada através do uso de métodos de agrupamento. O k-médias é o algoritmo mais utilizado na literatura para esse tipo de seleção de palavras [41, 47, 78]. Nesse caso, ao fim da execução do algoritmo, os centroides dos p grupos serão as palavras visuais do dicionário.

Definidas as palavras visuais, o dicionário visual que representa o problema está completo. Resta agora gerar os histogramas para cada uma das imagens.

Dado que cada uma das palavras visuais representa o centro de um bin do histograma, o histograma pode ser gerado de acordo com o Algoritmo 2.

O processo completo pode ser observado na Figura 5.2

Require: conjunto das c regiões da imagem.
conjunto das n palavras visuais do dicionário.
zerar a frequência das n posições do histograma.
for cada uma das c regiões da imagem **do**
  for cada i-ésima palavra do dicionário visual **do**
    calcula a distância entre a região e a palavra.
    armazena a distância calculada e o indice i da palavra.
  end for
  determina a menor distância dentre as n distâncias calculadas.
  obtém o índice i da menor distância.
  adiciona 1 na frequência da posição i do histograma.
end for

Algoritmo 2: geração da sacola de palavras de uma imagem

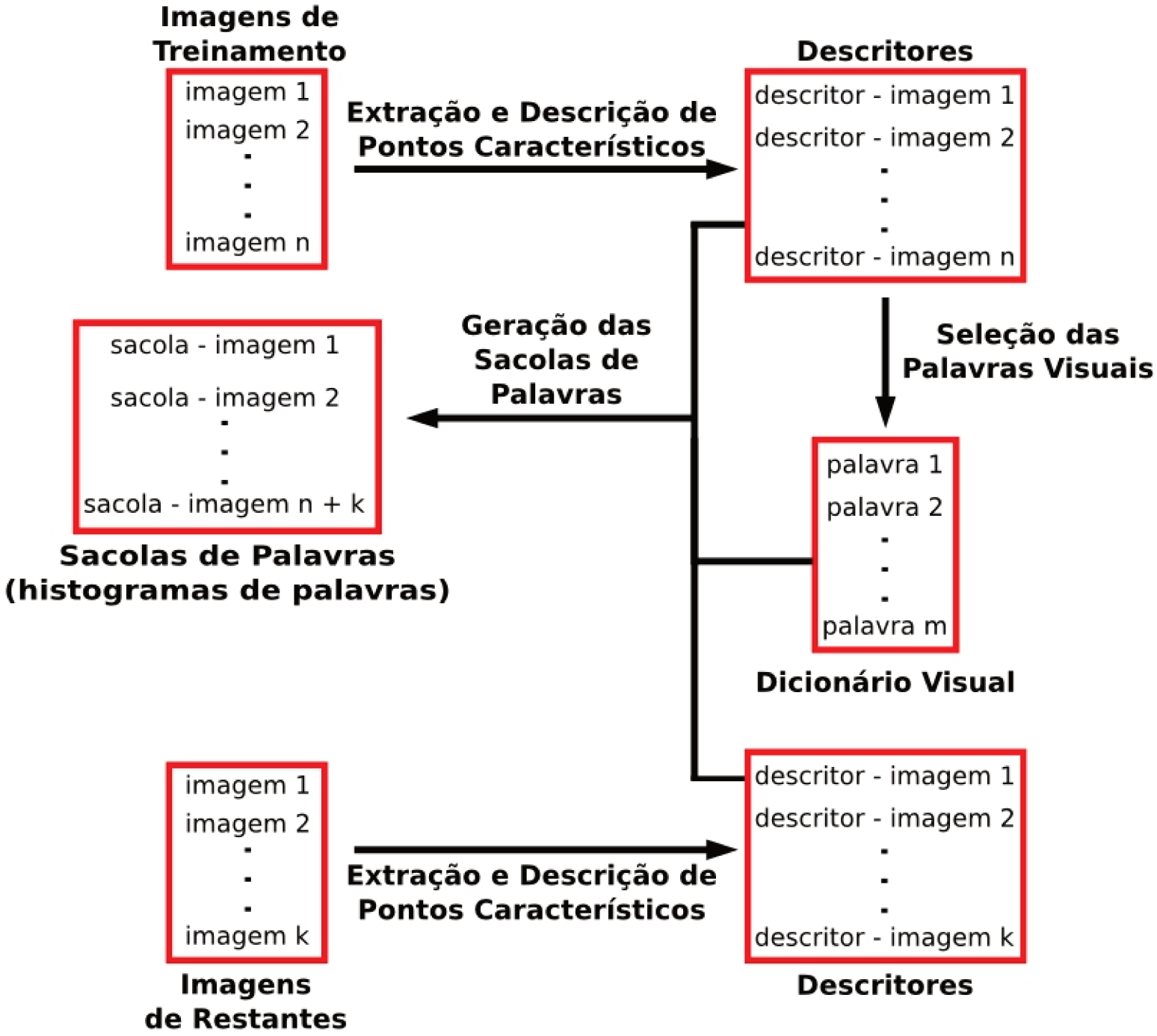


Figura 5.2: Geração das sacolas de palavras.

Capítulo 6

Metodologia Cientifica

O uso de métodos computacionais empregados na resolução de problemas reais tem se tornado cada vez mais comum. No entanto, ao se desenvolver, ou mesmo empregar um método cientifico é necessário que os passos seguidos até a obtenção dos resultados finais sejam bem descritos para que tal método possa ser compreendido e empregado de maneira correta por qualquer estudioso da área. Esse capítulo apresenta uma descrição detalhada de toda a metodologia utilizada nesta dissertação. Todos os parâmetros utilizados nesse trabalho são escolhidos empiricamente baseados em experimentos realizados.

6.1 Base de Dados

As imagens, também chamadas de *amostras*, que compõem a base foram obtidas junto a Universidade Federal de São Paulo (UNIFESP). São um total de 8072 imagens de fundo de olho, sendo que 7696 são do tipo TIF com dimensões de 640×480. As demais imagens são do tipo JPG, com dimensões variadas (*e.g.* 816×499, 899×715, 963×710, 966×711, dentre outros).

Cada uma dessas imagens foi analisada por um ou mais médicos especialistas da própria UNIFESP, os quais constataram se a referida imagem possuía ou não alguma anomalia. Diversas imagens da base foram constatadas pelos médicos como não apresentando as condições mínimas para se fazer um diagnóstico sendo descartadas da base, restando um total de 2307 imagens, que compuseram a base de dados deste trabalho.

Uma amostra das imagens contidas na base podem ser vistas na Figura 6.1. Pode-se observar, que as imagens variam muito em diversos aspectos, tais como, cor, contraste, iluminação, alinhamento quanto às estruturas da retina e dimensões.

Das 2307 pertencentes à base, 687 foram classificadas pelos médicos como sendo imagens de retinas saudáveis (normais). As demais apresentavam algum tipo de anomalia (doentes). Segundo os médicos, entre as imagens doentes foram encontradas imagens de

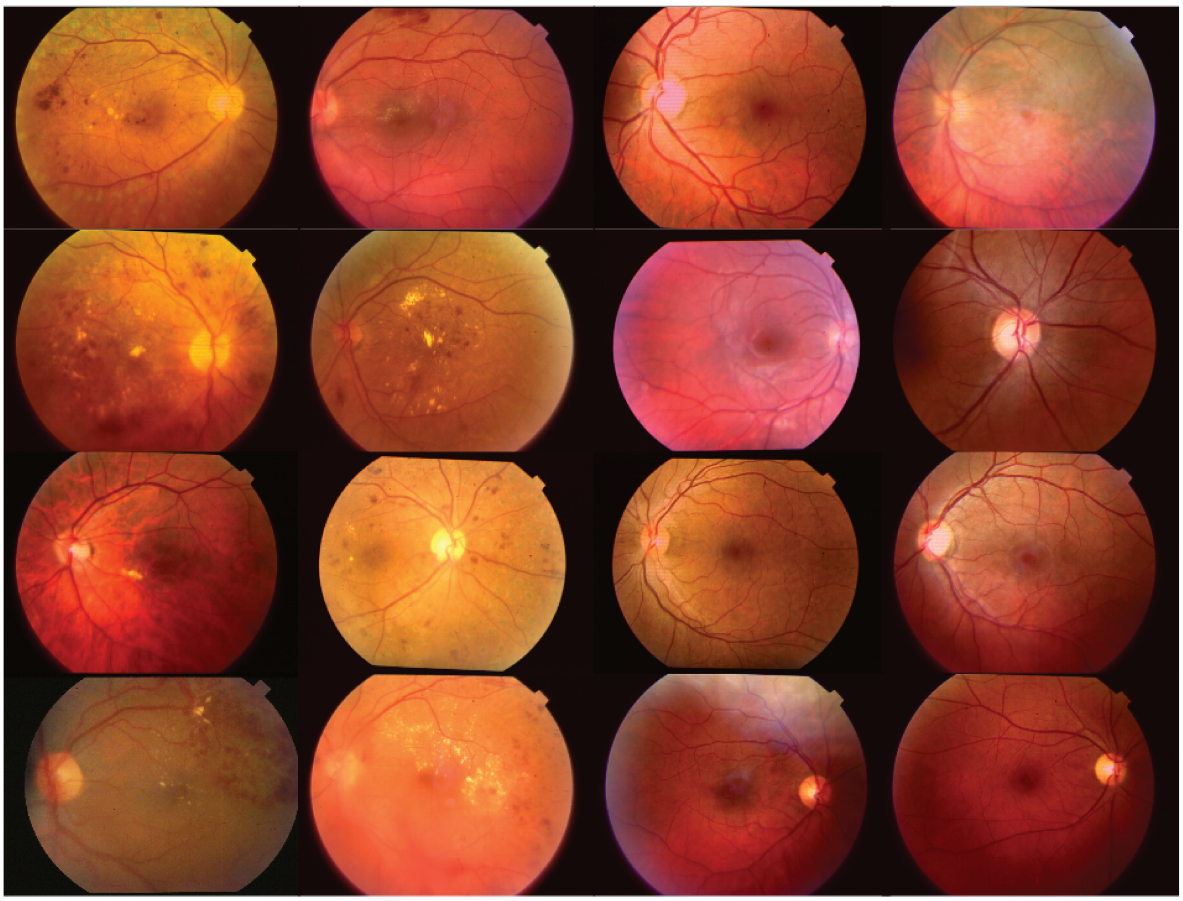


Figura 6.1: Exemplos de imagens contidas na base de dados.

mais de 100 tipos de anomalias.

As classes mais comuns de anomalias na base são o exsudato duro (264 imagens), hemorragia profunda (170 imagens), hemorragia superficial (120 imagens) e microaneurismas (79 imagens).

Uma *classe* nesse trabalho representa o tipo de diagnóstico com o qual a imagem está associada. Se uma imagem é classificada como doente, ela pode apresentar mais de uma classe.

Como o método proposto foi testado apenas em cima da detecção de exsudato, para um entendimento mais fácil, a metodologia sera descrita sobre duas classes: normais e exsudato.

6.2 Geração de Dobras

Como a geração de dicionários visuais, bem como a classificação dos dados, necessita de uma parte dos dados para seu treinamento, optamos por dividir a base de dados em b dobras, dos quais, l dobras ($l < b$) serão utilizados no treinamento para a construção do dicionário visual, e $b - l$ dobras serão utilizados nos testes. Esta parte é necessária também para a utilização da validação cruzada dos testes, validando assim os resultados dos experimentos.

O método proposto trabalha com cinco dobras para cada classe.

As dobras foram compostas de forma que cada uma das imagens de uma classe era atribuída de forma aleatória a um das dobras da classe. Essa atribuição leva em conta também o número de imagens em cada dobra da classe, de modo que nenhuma dobra possua mais imagens que outra dobra da mesma classe.

Assim, podemos descrever a separação das imagens em dobras para uma classe, segundo o Algoritmo 3

Require: conjunto **Z** de imagens de uma classe.
inteiro a referente ao número de imagens contidas em **Z**.
inteiro l referente ao número de dobras

1: zerar os contadores que representam o total de imagens de cada dobra.
2: **for** cada imagen $\hat{\mathrm{I}}_i \in$ **Z do**
3: escolher aleatoriamente j de modo que $1 \leq j \leq l$.
4: **if** numero de imagens do "fold" $j \leq \frac{a}{l} + 1$ **then**
5: $\hat{\mathrm{I}}_i$ é adicionada ao "fold" j.
6: **else**
7: **while** $\hat{\mathrm{I}}_i$ não for adicionada a nenhum "fold" **do**
8: escolher aleatoriamente j de modo que $1 \leq j \leq l$.
9: **if** numero de imagens do "fold" $j \leq \frac{a}{l} + 1$ **then**
10: $\hat{\mathrm{I}}_i$ é adicionada ao "fold" j.
11: **end if**
12: **end while**
13: **end if**
14: **end for**

Algoritmo 3: algoritmo para divisão das imagens de uma determinada classe da base de dados em dobras

6.3 Pré-processamento

Ao contrário do que ocorre na maior parte dos métodos propostos na literatura até o momento, os quais sempre realizam correções (contraste, brilho, luminosidade, etc.) na etapa de pré-processamento, o método proposto nesse trabalho não faz uso de tais técnicas para melhorar as características das imagens.

Na etapa de pré-processamento é realizada apenas uma conversão das imagens coloridas para imagens em tons de cinza. Isso porque na próxima etapa, a etapa de extração e descrição de pontos característicos, os algoritmos utilizados precisam trabalhar com esse tipo de imagens. Para realizar essa conversão é utilizada uma biblioteca de código aberto denominada "Intel Open Source Computer Vision Library" **(OpenCV)** [8] implementada em linguagem C/C++.

Dada uma imagem colorida qualquer $\hat{\mathbf{I}}$, simplesmente converte-se para uma nova imagem em tons de cinza $\hat{\mathbf{I}}$**'** preservando as mesmas dimensões da imagem original.

6.4 Extração e Descrição de Regiões da Imagem

Uma vez realizada a etapa de pré-processamento, a próxima etapa do método é a extração e descrição das regiões das imagens.

Como visto no capítulo anterior, as regiões de uma imagem para a construção de um dicionário visual, podem ser obtidas de diversas formas, como "patches" da imagem e pontos característicos.

O método aqui proposto utiliza pontos característicos para representar as regiões da imagem, extraídos com a ajuda de dois algoritmos muito utilizados na literatura de visão computacional: SIFT e SURF.

Uma vez alcançada uma configuração capaz de gerar um número de pontos satisfatório e um vetor de características robusto o suficiente para cada imagem do problema, todos os testes são realizados tomando como base os pontos característicos extraídos uma única vez, os quais são armazenados em disco. Em outras palavras, ao invés de se extrair os pontos característicos de uma imagem $\hat{\mathrm{I}}$ a cada novo cenário de teste, extraiu-se os pontos característicos uma única vez e sempre que necessário tais pontos eram reutilizados.

6.4.1 Extração e Descrição Utilizando SIFT

Na extração de pontos característicos utilizando o SIFT, foi utilizada uma implementação em linguagem C disponibilizada pelo próprio autor do método[1]. Assim para cada uma das

[1]http://www.cs.ubc.ca/ lowe/keypoints/

a imagens válidas contidas na base de dados, foi gerado um arquivo texto de pontos característicos. O arquivo de pontos característicos contém informações relativas ao número de pontos encontrados na imagem, da dimensão do descritor utilizado para descrever cada ponto e para cada um dos pontos extraídos, são armazenadas as informações referentes à localização da linha e coluna do ponto, com precisão de sub-pixel, escala e orientação (em radianos de -π até π) respectivamente. Por fim, cada um dos pontos característicos conta com um vetor d-dimensional o qual abriga em cada uma das dimensões um inteiro variando de 0 a 255.

Nos experimentos realizados o vetor de características possui a dimensão 128 originalmente proposta no SIFT.

Na construção do primeiro oitavo do espaço escala todas as imagens tiveram sua dimensão duplicada aumentando em quatro vezes o número de pontos característicos encontrados, porém, reduzindo em quatro vezes o tempo de computação do algoritmo. Cada um dois oitavos também foi dividido em três escalas. O σ inicial da Gaussiana utilizada em cada oitavo, possui valor de 1,6. Outro importante parâmetro na configuração foi o comprimento da janela gaussiana de pesos, o qual foi utilizado com o valor de 1,5.

Como resultado, essa configuração extraiu uma média de 764 pontos característicos por imagem nas classes normal e exsudato duro.

6.4.2 Extração e Descrição Utilizando SURF

Por sua vez, na extração de pontos característicos utilizando o algoritmo SURF, utilizou-se a implementação também em linguagem C disponibilizada na internet pelo autor[2]. Assim como no caso do SIFT, para cada uma das a imagens da base de dados foi gerado utilizando o SURF um arquivo texto com os pontos característicos extraídos da imagem. No arquivo de pontos gerado pelo SURF, são armazenadas informações como dimensão do vetor de características, o número de pontos extraídos da imagem, bem como informações referentes a cada um dos pontos característicos (posição x do ponto, posição y do ponto, três entradas da matriz de segundo momento, sinal do laplaciano, e o descritor do ponto).

O SURF foi proposto para gerar inicialmente um vetor de características de 64 dimensões. No entanto, ele possui uma versão estendida, na qual o vetor possui 128 dimensões. Apesar do tempo para calcular o vetor ser um pouco maior, o método aqui proposto utiliza essa representação mais robusta.

Para aumentar o número de pontos característicos localizados nas imagens com o SURF, é necessário se dobrar a dimensão da imagem na iteração inicial, além de escolher um bom limiar para o valor de respostas dos "blobs". Um limiar de 3500 apresentou um bom resultado no número de pontos extraídos das imagens.

[2]http://www.vision.ee.ethz.ch/ surf/download.html

Com tais configurações o SURF apresentou um número médio de 984 pontos característicos por imagem nas classes normal e exsudato duro.

6.5 Construção dos Dicionários Visuais

Uma vez que foram extraídos os pontos característicos de todas as imagens, o próximo passo do método é a determinação do dicionário visual para representar as classes do problema.

Essa etapa por sua vez pode ser divida nas sub-etapas que se seguem.

6.5.1 Conjunto de Treinamento

Como dito anteriormente, os dicionários visuais são construídos a partir de uma parte dos dados denominado *conjunto de treinamento*. Para a composição da matriz que representará o conjunto de treinamento de cada classe do problema, realiza-se o seguinte procedimento: dada a classe, para cada uma das imagens do conjunto de treinamento, são selecionados todos os vetores de características pertencentes à imagem. Esses vetores de características são então inseridos em uma matriz M_T, onde T corresponde à classe de anomalia, de forma que tal matriz possua m linhas por d colunas. Baseado nessa afirmação, o número de colunas d representa o número de dimensões dos vetores de características das imagens e m pode ser entendido como

$$m = \sum_{i=1}^{l} \sum_{j=1}^{a_i} s_{\hat{\mathrm{I}}_j}, \tag{6.1}$$

onde l é igual ao número de dobras do conjunto de treinamento, a_i é o número total de imagens contidas no "fold" i e $s_{\hat{\mathrm{I}}_j}$ representa o número de vetores de características pertencentes a imagem $\hat{\mathrm{I}}_j$.

Assim, para cada uma das classes de anomalias (normal e exsudato), é gerada uma matriz M_T, que no fim desta etapa, são unidas em uma única matriz M, representando o conjunto de treinamento.

6.5.2 Seleção das Palavras Visuais

O grande desafio na construção de um dicionário visual é determinar quais são as palavras visuais mais representativas para as classes do problema. No método proposto, cada uma das palavras visuais é representada por um vetor de características d dimensional, sendo n igual ao número de colunas da matriz de entrada.

Como visto no Capítulo 5, uma vez determinado o número de palavras visuais do dicionário, nosso método as seleciona de quatro maneiras distintas: por agrupamento, de forma aleatória, através da seleção de regiões de interesse, através da seleção manual das palavras realizada por um especialista.

6.5.2.1 Por Agrupamento

Nessa forma de seleção de palavras visuais, é utilizado o algoritmo k-médias para realizar o agrupamento. O algoritmo foi implementado em linguagem C, e tem como entrada a matriz M gerada na etapa anterior e o número p de grupos.

Na implementação utilizada, os centroides iniciais são obtidos selecionando-se p linhas de forma aleatória na matriz M de entrada, sendo que, com exceção dessas linhas, todas as demais recebem inicialmente o rótulo de um mesmo grupo.

Outro ponto a ser observado é a métrica de distância utilizada. Apesar de ser uma das métricas de distância mais simples da literatura, a distância Euclidiana utilizada na implementação do k-médias produz bons resultados. Métricas mais sofisticadas como a de Mahalanobis [101] são muito custosas computacionalmente, o que dificulta sua implantação.

Para evitar que o algoritmo entre em um "loop" infinito, sua execução termina quando um dos dois critérios abaixo é alcançado

- o rótulo de 90% ou mais das linhas da matriz de entrada permanece inalterado por duas iterações;
- o algoritmo realiza 200 iterações.

Ao fim da execução do algoritmo os centroides dos p grupos são então escolhidos como sendo as palavras visuais.

Um fluxograma com todo o processo de seleção das palavras visuais utilizando agrupamento pode ser visto na Figura 6.2

6.5.2.2 Por Seleção Aleatória

A forma aleatória seleciona p linhas aleatórias da matriz M de entrada e escolhe esses pontos como sendo as palavras visuais.

Um fluxograma com todo o processo de seleção das palavras visuais utilizando seleção aleatória pode ser visto na Figura 6.3

6.5.2.3 Por Seleção de Região

Essa é uma forma de seleção semi-supervisionada das palavras visuais, utilizada no método proposto. Sua primeira diferença em relação às formas de seleção anteriores, é que ao invés

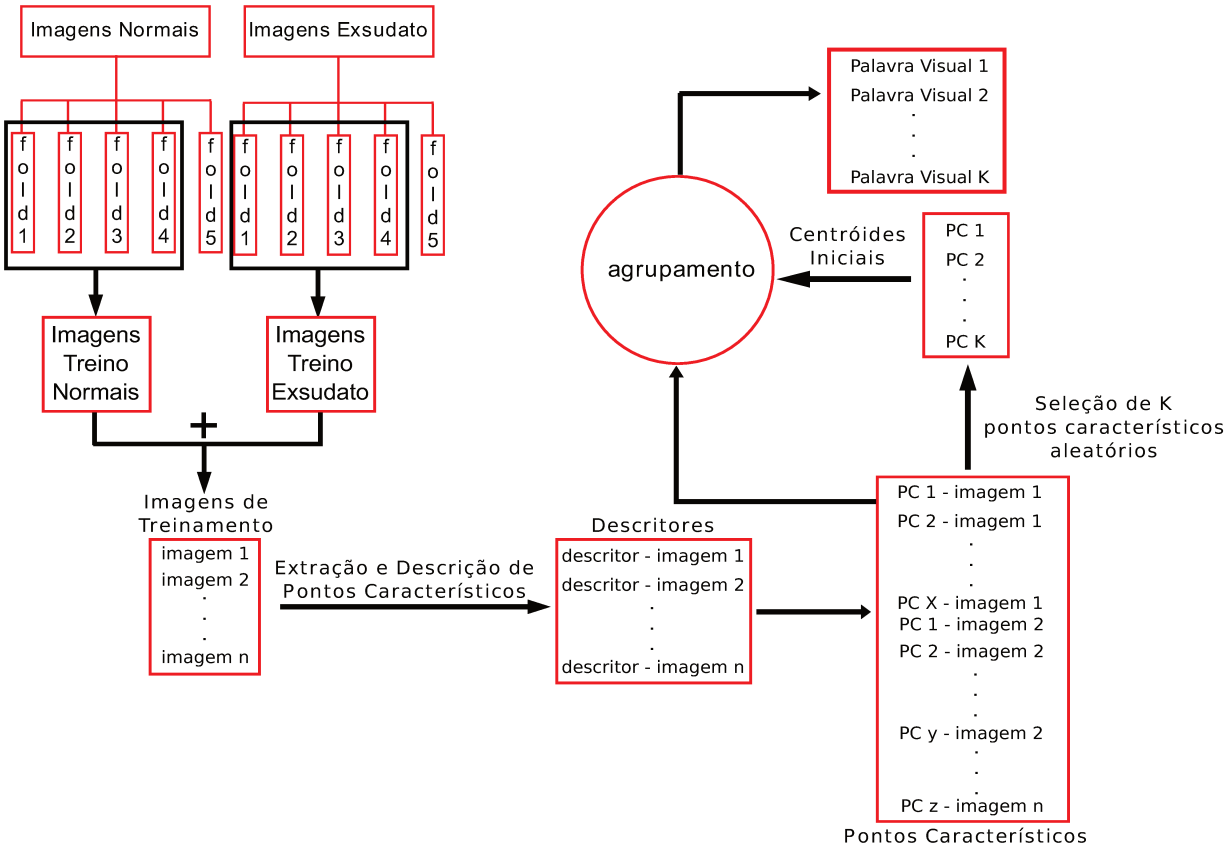


Figura 6.2: Fluxograma de seleção de palavras visuais utilizando agrupamento.

de contar com uma única matriz M de entrada, essa forma de seleção de palavras tem como entrada c matrizes do tipo M_T, onde c representa a quantidade de classes do problema.

Para a matriz representante das imagens normais, seleciona-se aleatoriamente $\frac{p}{c}$ linhas aleatoriamente.

No entanto, para a seleção das demais palavras visuais, é necessária a adição de uma nova etapa no processo.

Para cada uma das imagens que não pertencem à classe normal, a área ao redor da anomalia deve ser selecionada. À partir dessa área, são novamente extraídos pontos característicos para a imagem. Assim, ao fim dessa etapa, todas as imagens pertencentes a uma classe T qualquer (excluindo-se a classe de imagens normais), irão possuir dois arquivos de pontos característicos, um que indica os pontos característicos de toda a imagem, e outro que indica os pontos característicos extraídos apenas na área próxima a anomalia. Um novo conjunto de treinamento M'_T para a classe T é então formado, baseado apenas nos arquivo que contêm os pontos extraídos na área da anomalia. São

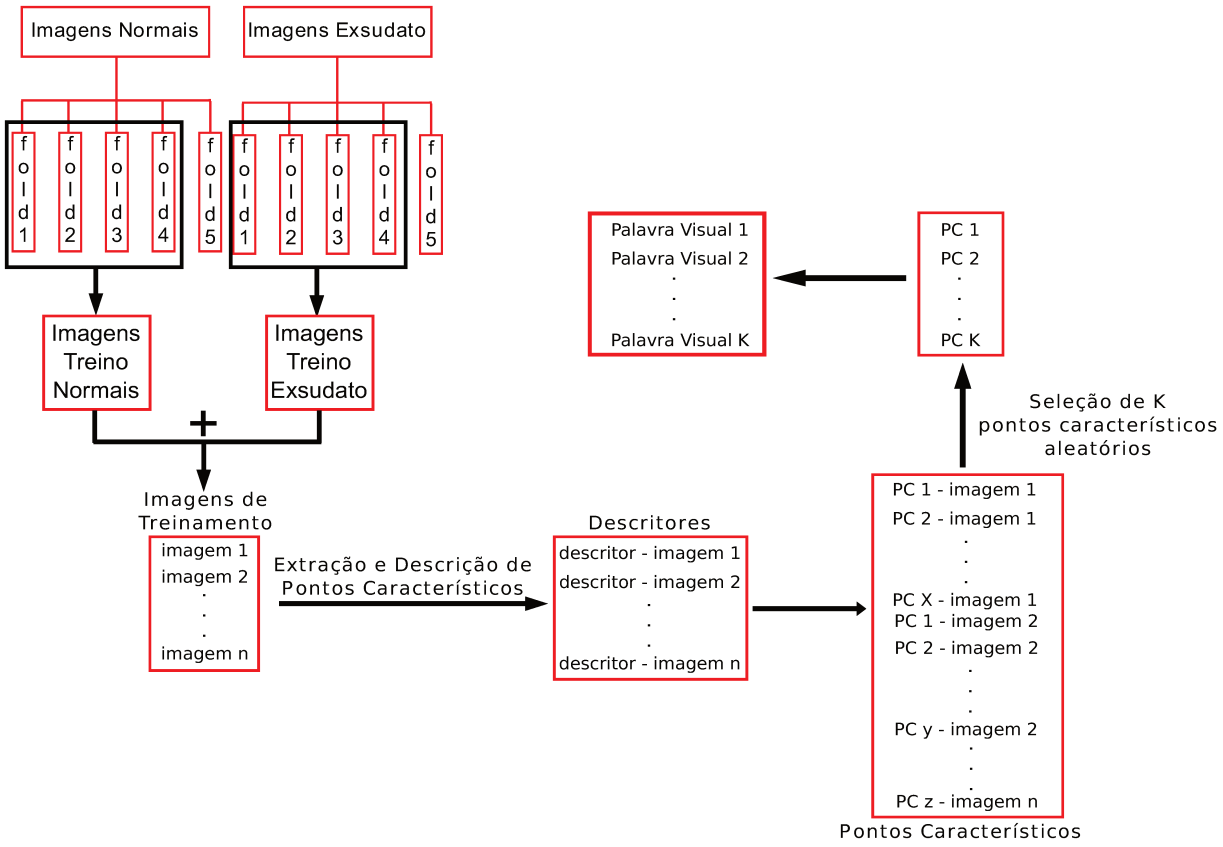


Figura 6.3: Fluxograma de seleção de palavras visuais utilizando seleção aleatória.

então extraídas $\frac{p}{c}$ palavras visuais à partir da nova matriz M'_T.

Por fim todas as palavras visuais extraídas nessa etapa são então unidas compondo dessa forma o dicionário visual do problema.

Um fluxograma com todo o processo de seleção das palavras visuais utilizando regiões marcadas pode ser visto na Figura 6.4

6.5.2.4 Por Seleção de Manual

A segunda forma de seleção semi-supervisionada das palavras visuais, utilizada no método proposto depende de um especialista capaz de identificar os pontos de exsudato dentre os pontos característicos extraídos nas imagens de treinamento.

Após a extração dos pontos característicos nas imagens de treinamento, um especialista seleciona manualmente, dentre os pontos característicos extraídos, $\frac{p}{2}$ pontos que ele tenha certeza serem pontos que representam regiões normais e $\frac{p}{2}$ pontos que ele tem certeza que

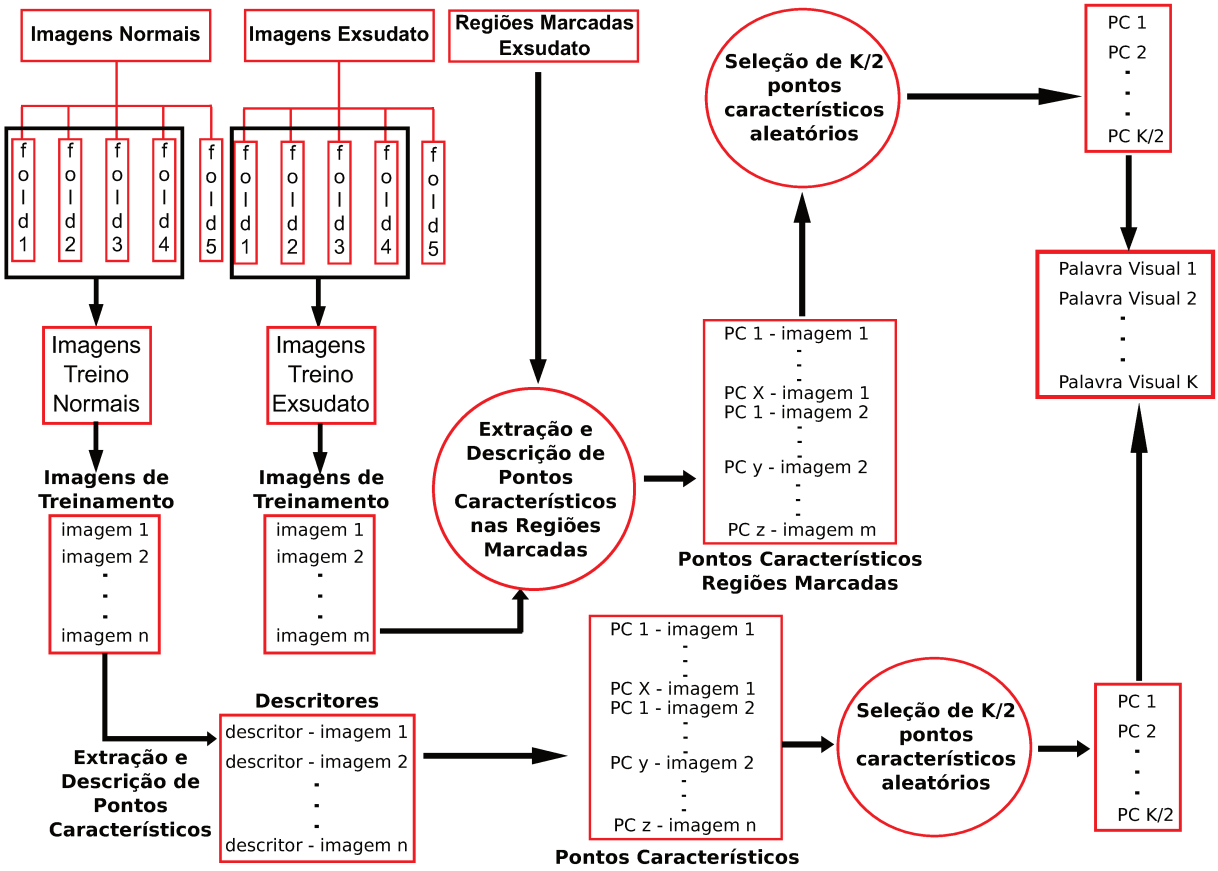


Figura 6.4: Fluxograma de seleção de palavras visuais utilizando seleção baseada em regiões marcadas.

representam pontos de regiões de doença.

Esses pontos selecionados manualmente serão os pontos que irão compor o dicionário visual.

Um fluxograma com todo o processo de seleção das palavras visuais utilizando seleção manual pode ser visto na Figura 6.5

6.6 Composição das Sacolas de Palavras (Histogramas Visuais)

Uma vez determinado o dicionário visual do problema como descrito na seção anterior, é necessário então se construir o histograma visual que passará a descrever cada imagem.

De posse do arquivo contendo todos os x pontos característicos extraídos a partir de Î (não apenas os pontos da região de doença, mas sim, os pontos localizados e extraídos

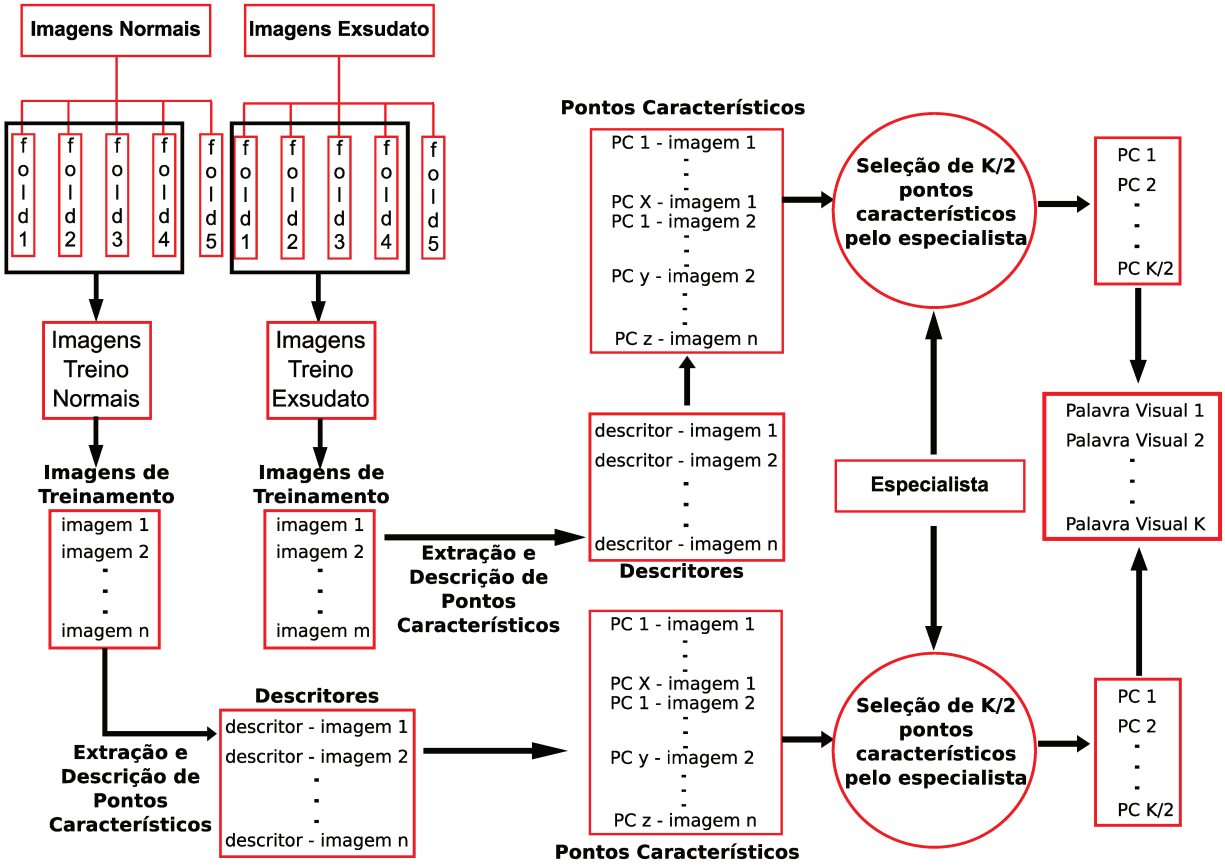


Figura 6.5: Fluxograma de seleção de palavras visuais utilizando seleção manual das palavras por um especilista.

na imagem toda), e do dicionário visual contendo as p palavras visuais selecionadas como descrito na seção anterior, a composição do histograma visual que passará a representar $\hat{I}$ é descrita pelo Algoritmo 4:

6.7 Treinamento e Classificação

A classificação dos dados envolve duas etapas: treinamento do classificador, e a classificação.

Na etapa de treinamento, utiliza-se a mesma distribuição de dobras vista na Seção 6.2. No entanto, a partir desse ponto, os vetores de características utilizados tanto no treinamento quanto na classificação dos dados são os histogramas, gerados a partir dos dicionários visuais, e não mais os descritores gerados na extração de regiões da imagem.

É composta então uma matriz de histogramas para ser utilizada no treinamento, onde

Require: conjunto **PC** contendo os pontos característicos de uma imagem.
conjunto **PV** contendo as palavras visuais.
número p de palavras visuais.

construir um vetor **V** de p posições onde cada posição representa uma palavra visual.
construir um vetor temporário **T** de p posições onde cada posição representa uma palavra visual.
for cada ponto característico i contido em **PC do**
  for cada palavra visual j contida em **PV do**
    calcular a distancia Euclidiana entre i e j.
    armazenar no vetor **T** na posição referente à palavra visual j o resultado do calculo de distância.
  end for
  selecionar o índice a (referente à palavra visual) da posição com o menor valor em **T**.
  somar um no valor contido na posição a de **V**.
end for
return V

Algoritmo 4: Composição da sacola de palavras de uma imagem

as linhas representam o número de imagens total contidas no conjunto de treinamento, e as colunas correspondem ao número de palavras visuais do dicionário, uma vez que a dimensão de cada histograma visual é igual ao número de palavras visuais do dicionário. Essa matriz possui ainda uma coluna adicional, a qual indica o rótulo da classe que cada linha pertence.

É necessário também se criar um conjunto de testes, o qual é composto pelos histogramas de todas imagens do "fold" não utilizado no treinamento. Assim o classificador é treinado com quatro dobras e testado em uma.

Para garantir a veracidade dos resultados, foi utilizado o método conhecido como *validação cruzada*, na qual são executados um número de testes total igual ao número de dobras para cada um dos cenários de testes. A cada teste um "fold" diferente é escolhido como o conjunto de testes, utilizando os demais para o treinamento.

Os testes são realizados com uma configuração padrão do classificador para todos os testes. Depois de definidos empiricamente os parâmetros que serão utilizados, todos os testes utilizam a mesma configuração do classificador. O único parâmetro variável durante a execução dos testes é o peso de cada uma das classes.

Isso acontece para uma melhor representação dos resultados através de curvas ROC, as quais são muito empregadas na exibição de dados referentes a detecção de anomalias, como é o caso do problema aqui apresentado.

Ao fim dos testes em todas as dobras, são calculados média e desvio padrão dos

resultados obtidos e assim, os mesmos são representados em forma de curva para exibir os resultados.

6.7.1 SVM – Treinamento e Classificação

Para o uso do SVM foi utilizada a biblioteca LIBSVM [12] a qual é implementada em C/C++. Para a utilização da biblioteca, o primeiro passo é o ajuste do formato dos dados. Uma vez que os dados encontram-se no formato da LIBSVM, o primeiro passo para o treinamento do classificador é o escalamento dos dados. Os primeiros a serem escalados são os dados pertencentes ao conjunto de treinamento. Baseado no problema, e na documentação da biblioteca, os dados do treinamento são escalados no intervalo fechado de [-1:1]. Já os dados de teste são escalados baseados num domínio definido pela própria biblioteca através dos dados de treinamento.

À partir dos dados já escalados, o próximo passo é o treinamento do classificador. A LIBSVM utiliza diversos parâmetros no treinamento do SVM, no entanto, devido ao tipo de "kernel" utilizado neste método, tais parâmetros pouco influenciam na classificação dos dados.

Foram realizados testes empíricos sobre varias configurações do classificador para diversas formas de dados, de forma que a melhor configuração, a qual foi utilizada. Em todos os testes, foi obtida com os parâmetros indicados abaixo:

- *tipo de "kernel"*: linear. Apesar de ser o mais simples foi o que apresentou os melhores resultados na classificação;
- *custo de penalidade C* (para aumentar a margem de separação dos dados): 0,5, o que dentro do contexto dos demais parâmetros, apresentou os melhores resultados;
- γ: determina o parâmetro gama em uma função de "kernel". Foi utilizado com o valor de 0,5, no entanto, devido ao uso do tipo linear para o "kernel" sua variação não altera os resultados da classificação;
- *tipo do SVM*: foi utilizado o tipo de abordagem *C-SVM*, o qual é a forma mais simples dentro do contexto do SVM para tratar problemas de classificação binária;
- ϵ: define a tolerância do critério de terminação do método. Foi utilizado com o valor de 0,2.
- w_1 e w_2: define o peso de cada uma das classes no classificador. Este parâmetro varia de 0 a 2, sendo que o valor de w_1 é complementar ao valor de w_2, somando o valor de 2 no total.

É importante ressaltar que essa variação nos pesos não foi realizada de forma linear, uma vez que há regiões no intervalo de peso das classes em que uma pequena variação do peso causa uma grande variação no número de acertos do classificador. Assim como há outras regiões com o comportamento inverso. Esse ajuste de peso é feito na etapa de treinamento, uma vez que para mudar os pesos de uma classe é necessário re-treinar o classificador.

Ao fim do treinamento, é gerado o arquivo do *modelo*, o qual tem seus parâmetros aprendidos segundo os dados do conjunto de treinamento e que é utilizado pelo SVM para, dada uma nova amostra de entrada, classificá-la dentre uma das duas classes do problema.

Capítulo 7

Experimentos e Resultados

Toda a investigação cientifica precisa de um bom embasamento teórico, mas são seus resultados que comprovam todas as teorias e promessas oferecidas pelo método. Nesse capítulo serão apresentados os principais resultados relacionados ao método proposto nesse trabalho bem como uma discussão de cada um dos resultados aqui apresentados.

7.1 Agrupamento × Seleção Aleatória

Segundo a maior parte da literatura existente em torno de dicionários visuais, para que a seleção de palavras visuais seja de boa qualidade e consiga bons resultados, é necessário a utilização de métodos de agrupamento (*e.g.* o algoritmo k-médias) sobre o conjunto de treinamento, para que assim as palavras mais representativas do problema sejam selecionadas. No entanto, quando se trabalha um volume muito grande de dados, tais algoritmos podem tornar o processo extremamente lento, tornando inviável sua aplicação.

O primeiro experimento realizado buscou a verificação da informação acima. Como o conjunto de treinamento possui um volume considerável de dados (cerca de 800 mil pontos característicos, com vetores de características de 128 dimensões para cada ponto), a utilização de um agrupamento de 500 palavras já é muito custoso. Existem ainda abordagens na literatura que sugerem um número muito maior de palavras, cerca de 5000, ou mais.

Isso torna inviável a aplicação de algoritmos de agrupamento em conjunto de treinamento com grande volume de dados buscando um número alto de palavras visuais devido ao fato de que algoritmos desse tipo podem levar até mais de 600 iterações para convergir, como foi constatado em alguns experimentos realizados. Isso faz com que o alto custo computacional seja um grande problema na utilização de agrupamentos.

Baseado nisso, o primeiro teste do método aplica a hipótese de uma seleção de palavras aleatórias, ao invés de contar com o uso do agrupamento.

O cenário deste teste tem as seguintes características: o conjunto de treinamento é formado tomando quatro das cinco dobras de cada uma das classes, somando aproximadamente 700 mil pontos no conjunto de treinamento quando os PC são extraídos com o SURF, utilizando os parâmetros exatamente como descritos na Seção 6.4.2.

Os dicionários visuais foram formados utilizando 500 palavras visuais. Na abordagem utilizando agrupamento, foi utilizado o algoritmo k-médias, com 500 grupos, utilizando a seleção de centroides iniciais de forma aleatória, selecionando 500 pontos entre todos os pontos pertencentes ao conjunto de treinamento, e tomando como palavras visuais os centroides dos grupos gerados quando o algoritmo termina. Já na abordagem sem o agrupamento, as palavras visuais são determinadas como sendo os centroides iniciais (*i.e.* os 500 centroides iniciais escolhidos aleatoriamente para o algoritmo de agrupamento).

Uma vez escolhidas as palavras visuais, seguem-se os passos para a iniciação dos testes: geração das sacolas de palavras; composição do conjunto de treinamento. Para o classificador, baseado nas sacolas de palavras e utilizando as mesmas dobras utilizadas para a seleção das palavras visuais; treinamento do SVM de acordo com os parâmetros indicados na seção 6.7.1 e por fim os testes das amostras contidas no conjunto de testes.

Os parâmetros e características mais relevantes utilizados para composição da base de dados, bem como aqueles utilizados na composição das sacolas de palavras, e na classificação das amostras podem ser observados respectivamente nas Tabelas 7.1, 7.2, 7.3.

Tabela 7.1: Experimento 1 - base de dados

amostras normais	687
amostras doentes	264
dobras normais	5
dobras doentes	5
imagens por dobras normais	~ 136
imagens por dobras doentes	~ 53

Tabela 7.2: Experimento 1 - dicionário visual

dobras treino normal	4
dobras treino exsudato	4
algoritmo extração de PCs	SURF
PC conjunto de treinamento	$\sim 700\,mil$
palavras visuais utilizadas	500
algoritmo agrupamento	k-médias
máximo iterações k-médias	200
tempo médio seleção palavras com k-médias	82 hrs
tempo médio de seleção aleatória palavras	40 seg

Tabela 7.3: Experimento 1 - classificador

dobras treino normal	4
dobras treino exsudato	4
dobras teste normal	1
dobras teste exsudato	1
classificador	SVM
tipo de "kernel"	linear
tipo SVM	C-SVM
penalidade do classificador (C)	0.5
validação cruzada	sim

Assim como foi explicado na Seção 6.7.1, as curvas resultantes são obtidas através da variação dos pesos das duas classes (normais e doentes), fato que ocorre no momento de

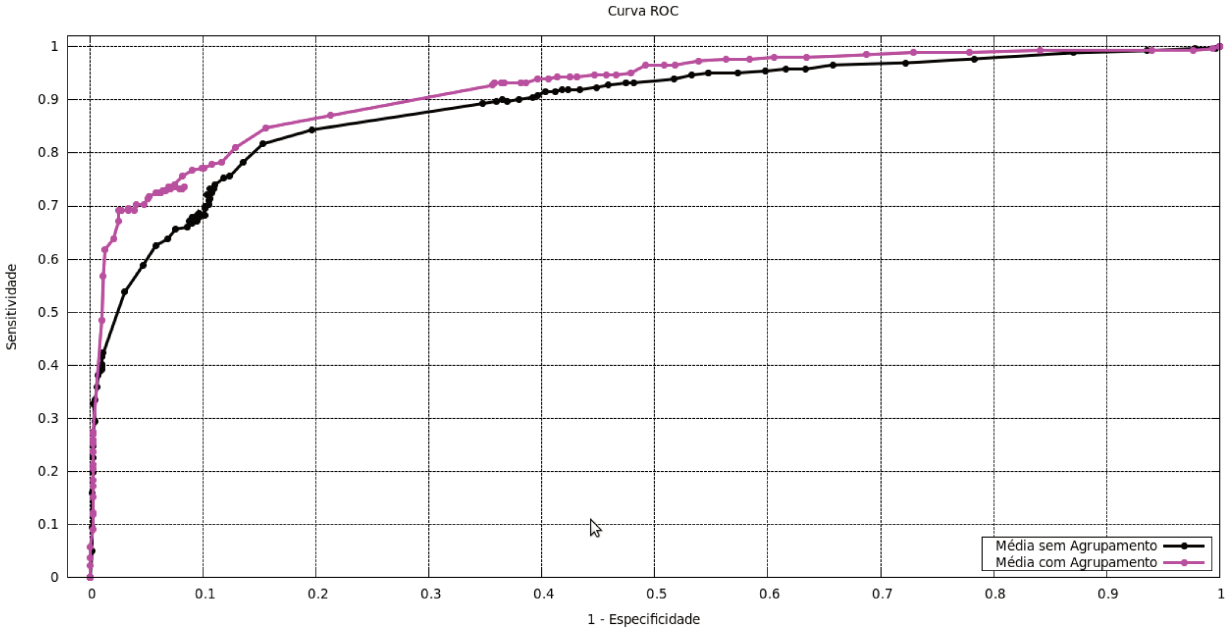


Figura 7.1: Curva ROC média das abordagens com e sem agrupamento.

treinamento do classificador.

No gráfico visto na Figura 7.1 são apresentados em forma de curva ROC os resultados obtidos no cenário descrito acima. É importante observar, que as curvas que exibem os resultados das duas abordagens (com e sem agrupamento) são na verdade a curva média obtida dentre a curva resultante das cinco dobras, utilizados para a validação cruzada, de cada uma das abordagens.

Os gráficos das figuras 7.2 e 7.3 exibem respectivamente as curvas do desvio padrão das abordagens com e sem agrupamento respectivamente.

Ao fim dos testes neste cenário, a principal constatação a ser feita é que o custo/benefício não é compensatório na utilização do agrupamento dos dados. Isso porque no problema aqui apresentado, o alto custo computacional em relação ao tempo de execução, não compensa o pequeno aumento no número de acertos do método.

7.2 SURF × SIFT

Dado que dicionários visuais constroem sua representação das imagens baseados em regiões que sejam suficientemente representativas para o problema, é necessário que as regiões candidatas a serem uma palavra visual possuam características que as tornem robustas e invariantes a pequenas modificações (*e.g.* a adição de ruído).

Como visto na Seção 4.4, os descritores locais, SIFT e SURF, são capazes de extrair

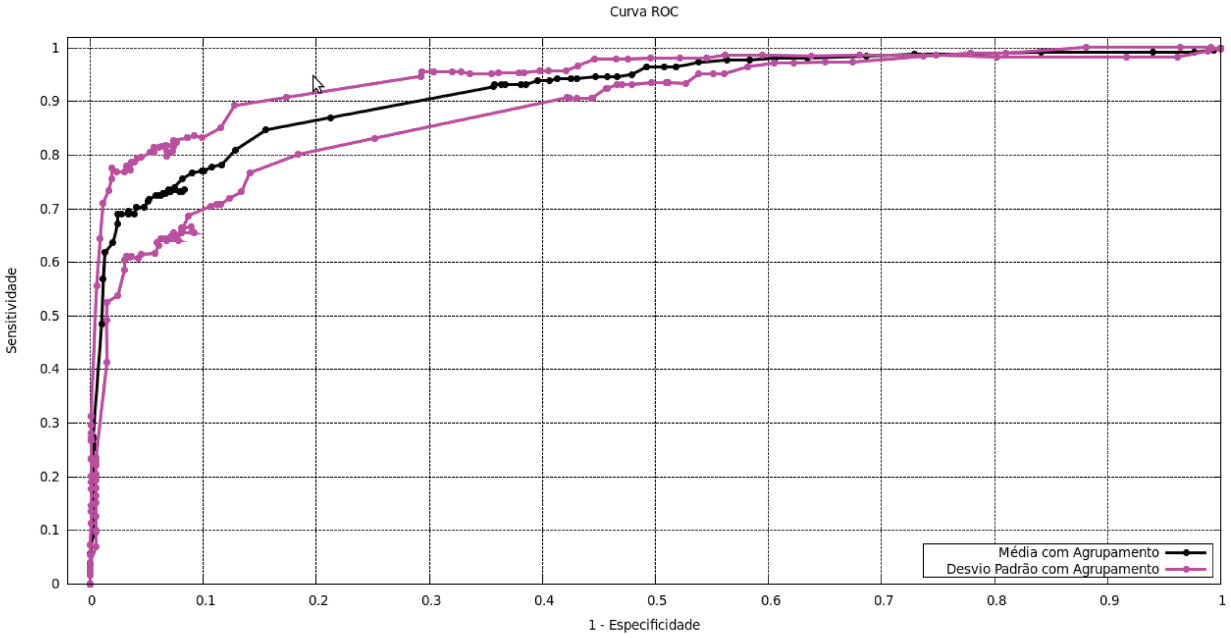


Figura 7.2: Curvas de desvio padrão da abordagem com agrupamento.

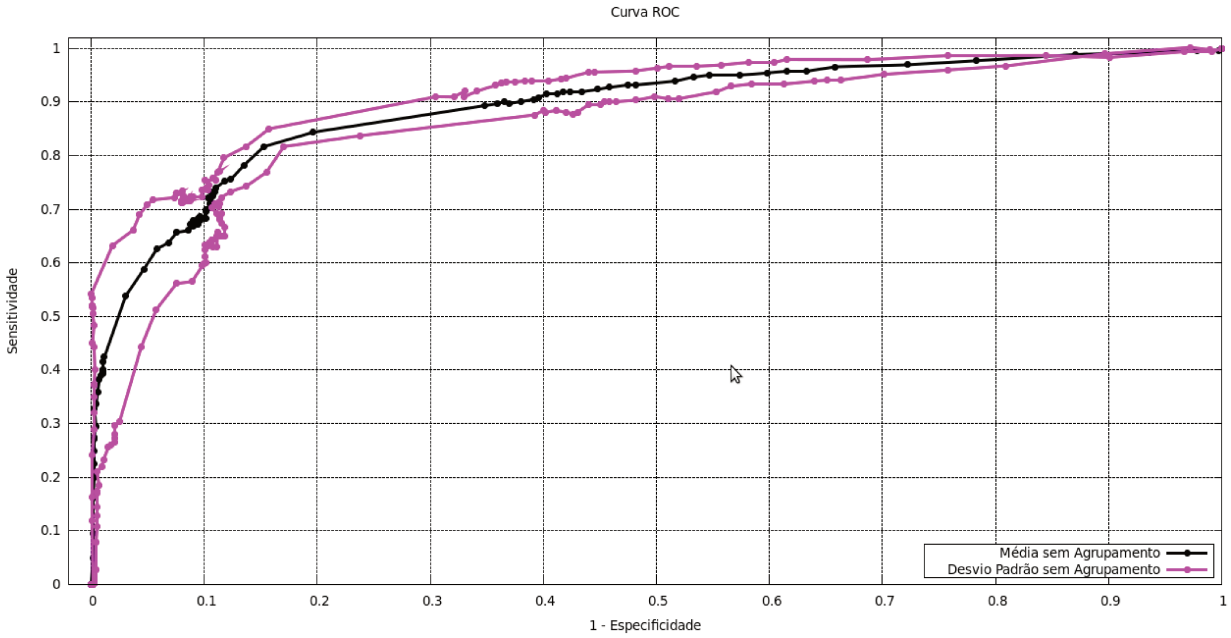


Figura 7.3: Curvas de desvio padrão da abordagem sem agrupamento.

PCs, que possuem esses tipos de características. Desta forma, é possível que ao invés de se considerar "patches" da imagem para representar suas regiões, utilize-se PCs.

Baseado nisso, as palavras que compõem o dicionário visual serão representadas pelos descritores dos PCs extraídos com os algoritmos acima citados.

Este experimento visa determinar qual dos dois descritores apresentados possui uma melhor representação das imagens para o tipo de problema aqui apresentado.

Os parâmetros de configuração para extração dos PCs no SIFT e no SURF são apresentados respectivamente nas Seções 6.4.1 e 6.4.2. Ambos os algoritmos geram como representação para cada um dos PCs vetores de 128 dimensões.

O segundo experimento segue os mesmos passos do primeiro. Compõe o conjunto de treinamento tomando quatro dobras. Para o teste com o SIFT, o conjunto de treinamento possui um total de aproximadamente 600 mil pontos. Já para o teste com o SURF, o conjunto de treinamento possui um pouco mais, aproximadamente 700 mil pontos.

São utilizadas 500 palavras visuais, as quais são escolhidas aleatoriamente a partir do conjunto de treinamento.

Por fim são construídas as sacolas de palavras, é composto o conjunto de treinamento do classificador, o mesmo é treinado e as amostras do conjunto de teste são classificadas, utilizando validação cruzada nas cinco dobras.

Os parâmetros e características mais relevantes utilizados para composição da base de dados, bem como aqueles utilizados na composição das sacolas de palavras, e na classificação das amostras podem ser observados respectivamente nas Tabelas 7.4, 7.5, 7.6.

Tabela 7.4: Experimento 2 - base de dados

amostras normais	687
amostras doentes	264
dobras normais	5
dobras doentes	5
imagens por dobras normais	~ 136
imagens por dobras doentes	~ 53

Tabela 7.5: Experimento 2 - dicionário visual

dobras treino normal	4
dobras treino exsudato	4
algoritmos de extração de PCs	SIFT e SURF
PC conjunto de treinamento SIFT	$\sim 600\,mil$
PC conjunto de treinamento SURF	$\sim 700\,mil$
palavras visuais utilizadas	500
seleção palavras visuais	aleatória

Tabela 7.6: Experimento 2 - classificador

dobras treino normal	4
dobras treino exsudato	4
dobras teste normal	1
dobras teste exsudato	1
classificador	SVM
tipo de "kernel"	linear
tipo SVM	C-SVM
penalidade do classificador (C)	0.5
validação cruzada	sim

As curvas resultantes dos testes utilizando os parâmetros do experimento acima des-

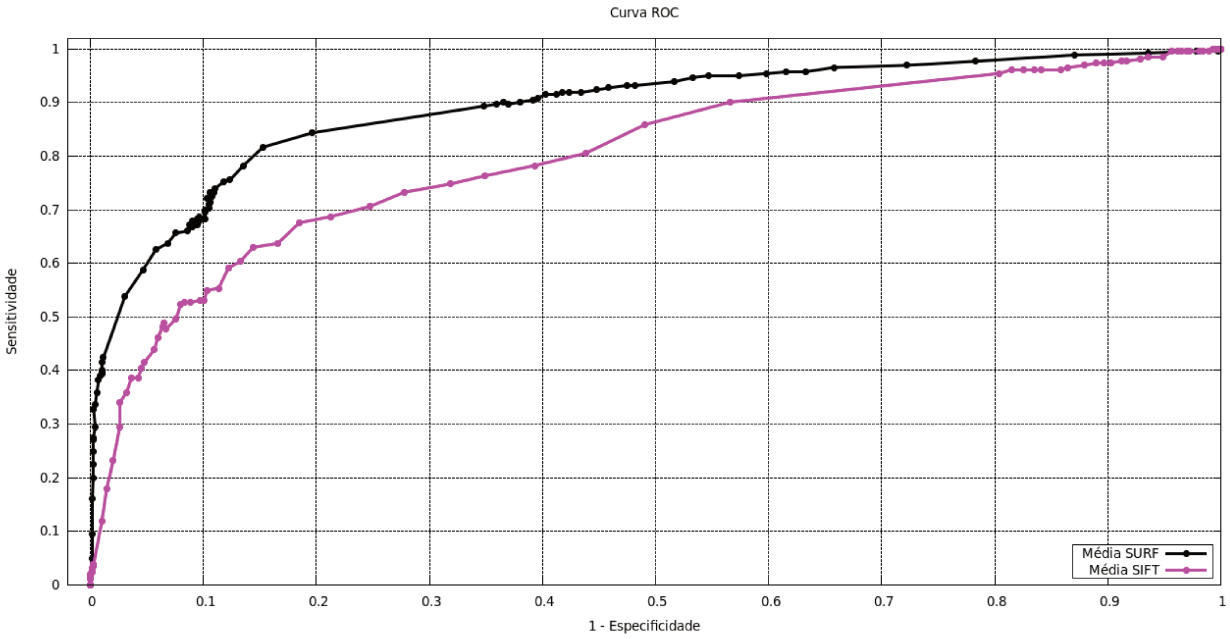


Figura 7.4: Curva ROC média com a utilização do SIFT e SURF na extração dos PCs.

crito, podem ser vistas no gráfico da Figura 7.4. Assim como no primeiro experimento, é importante observar, que as curvas que exibem os resultados das duas abordagens de extração de PCs (SIFT e SURF), são na verdade a curva média obtida entre a curva resultante das cinco dobras, utilizados para a validação cruzada.

São apresentados também, nos gráficos das figuras 7.5 e 7.6 o desvio padrão das abordagens utilizando o SIFT e o SURF para extrair os PCs.

Os testes neste cenário mostram que no problema abordado neste trabalho, a utilização do SURF para a extração dos PCs além de gerar um número maior de acertos gera um desvio padrão menor.

7.3 25 × 50 ×100 × 500 × 1000 Palavras Visuais

A composição do terceiro cenário de testes é realizada em torno da constatação na literatura de que alguns tipos de aplicações, exigem um alto número de palavras no seu dicionário. Essa afirmação, implica que uma vez aumentado o número de palavras visuais utilizadas na construção do dicionário, a porcentagem no número de acertos na classificação também deveria ser maior.

O cenário do terceiro experimento leva em consideração os resultados obtidos nos experimentos um e dois para investigar essa afirmação.

Inicialmente, o SURF é utilizado para a extração dos PCs. A escolha das palavras é

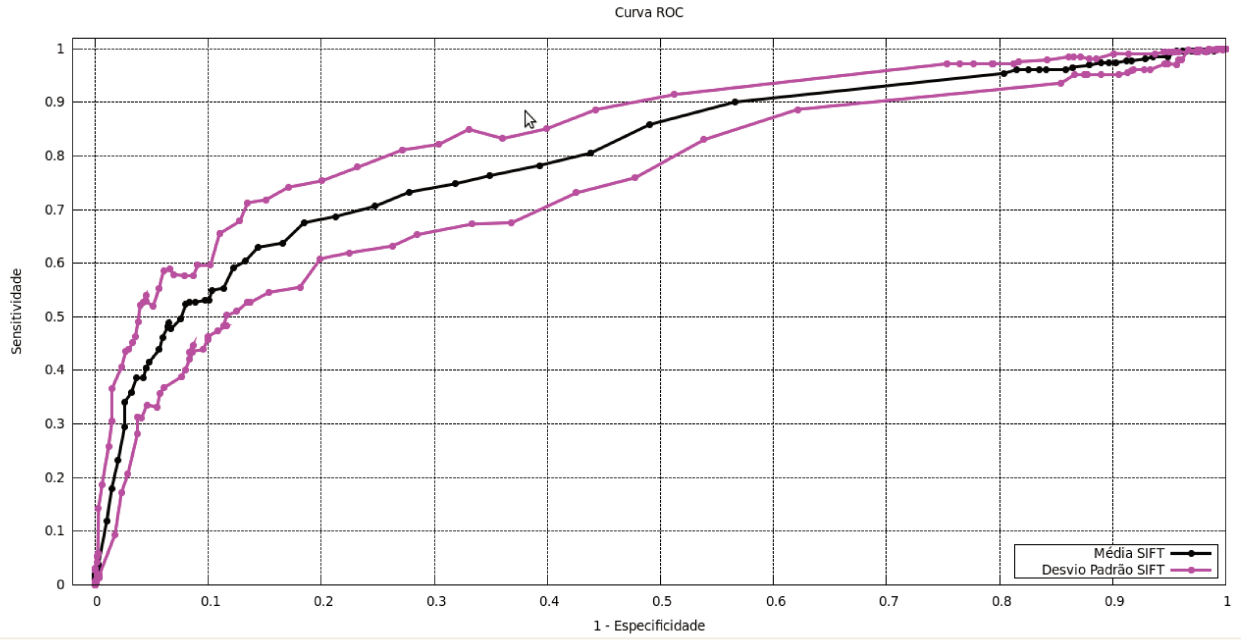


Figura 7.5: Curvas de desvio padrão da abordagem com o SIFT.

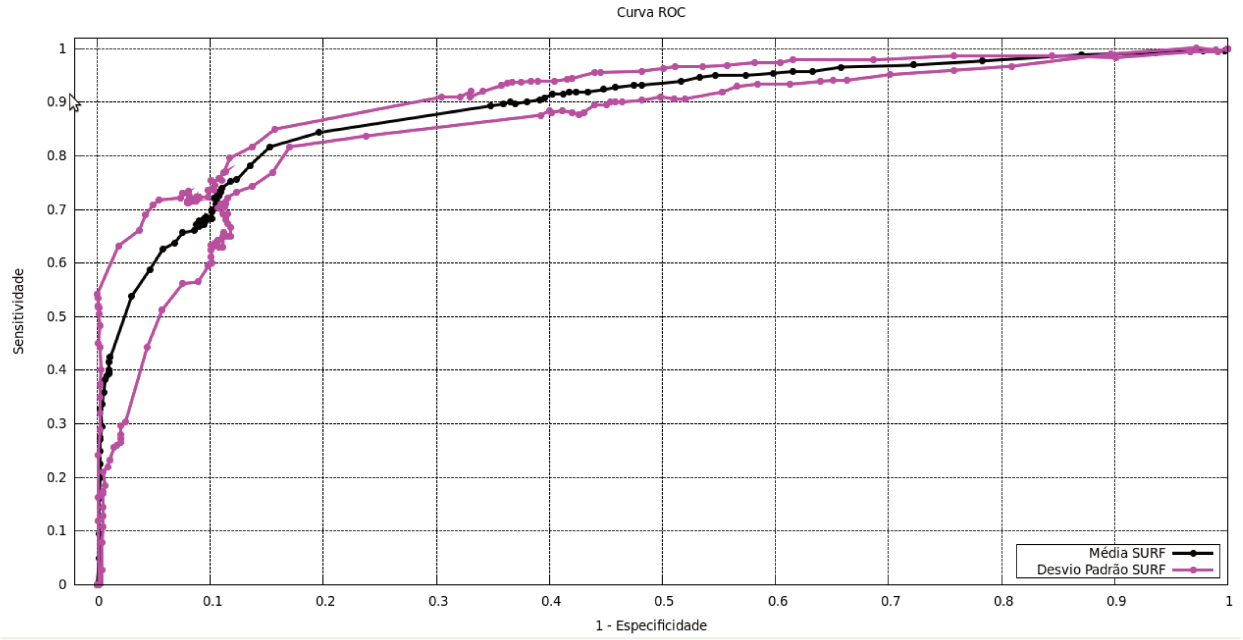


Figura 7.6: Curvas de desvio padrão da abordagem com o SURF.

feita de forma aleatória, realizando testes com 25, 50, 100, 500 e 1000 palavras. Os demais procedimentos como composição das sacolas de palavras, e a classificação dos dados, são idênticos aos realizados nos experimentos um e dois.

As Tabelas 7.7, 7.8, 7.9 apresentam os parâmetros e características deste cenário.

Tabela 7.7: Experimento 3 - base de dados

amostras normais	687
amostras doentes	264
dobras normais	5
dobras doentes	5
imagens por dobras normais	~ 136
imagens por dobras doentes	~ 53

Tabela 7.8: Experimento 3 - dicionário visual

dobras treino normal	4
dobras treino exsudato	4
algoritmos de extração de PCs	SURF
PC conjunto de treinamento SURF	$\sim 700\,mil$
palavras visuais utilizadas	25, 50, 100, 500, 1000
seleção palavras visuais	aleatória

Tabela 7.9: Experimento 3 - classificador

dobras treino normal	4
dobras treino exsudato	4
dobras teste normal	1
dobras teste exsudato	1
classificador	SVM
tipo de "kernel"	linear
tipo SVM	C-SVM
penalidade do classificador (C)	0.5
validação cruzada	sim

As curvas médias mostrando os resultados com diferentes números de palavras visuais, bem como as curvas de desvio padrão dos experimentos com 25, 50, 100, 500 e 1000 palavras visuais podem ser vistas nos gráficos das Figuras 7.7, 7.8, 7.9, 7.10, 7.11, 7.12.

Este cenário de testes, permite afirmar que para o tipo de problema aqui tratado, o número de palavras visuais não influencia significativamente na porcentagem de acertos.

7.4 Seleção de Regiões

O quarto cenário de testes é um pouco diferente dos demais e para entende-lo é necessário uma rápida explicação de como as imagens utilizadas no treinamento são rotuladas. Uma equipe médica da UNIFESP é a responsável por essa atribuição de rótulos, diagnosticando a anomalia presente em cada uma das imagens. No entanto, o médico responsável pelo diagnóstico (e consequentemente pelo rótulo) da imagem não indica a localização da anomalia na imagem. Ele simplesmente indica que a imagem apresenta uma determinada anomalia, sem qualquer informação adicional.

O cenário do quarto teste utiliza uma informação adicional para a seleção das palavras visuais. Devido ao fato de exsudatos serem lesões que uma pessoa pode facilmente identificar em uma imagem, para cada uma das imagens contendo exsudato, foi indicada a região aproximada onde a doença se localizava.

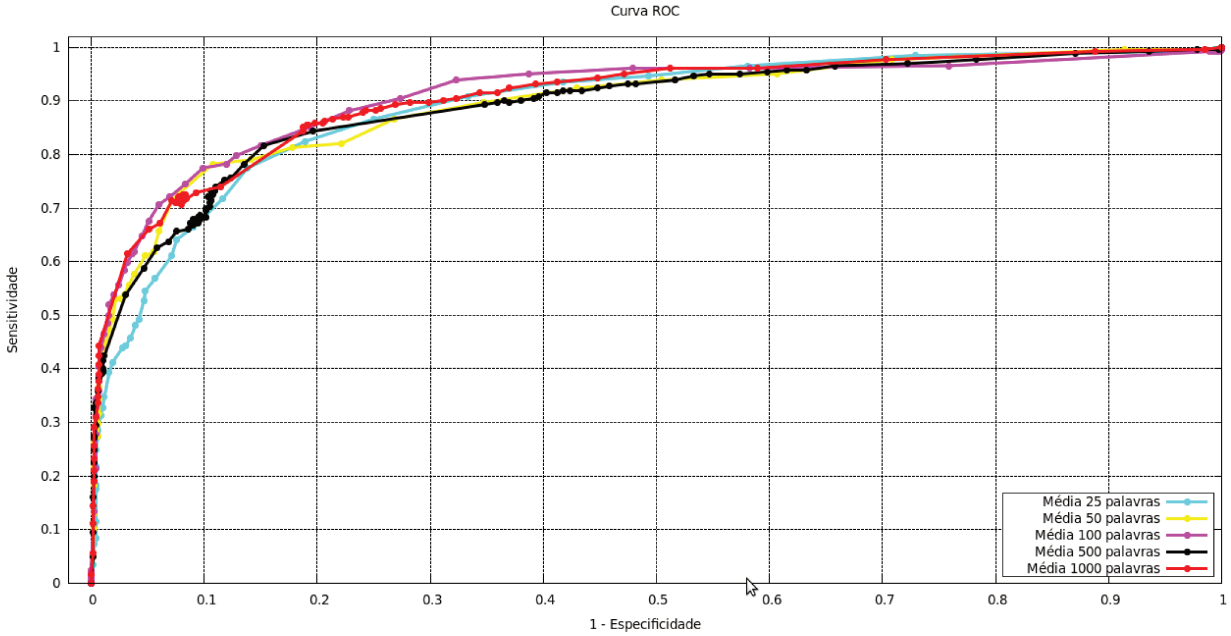


Figura 7.7: Curvas representando testes com diferentes números de palavras visuais.

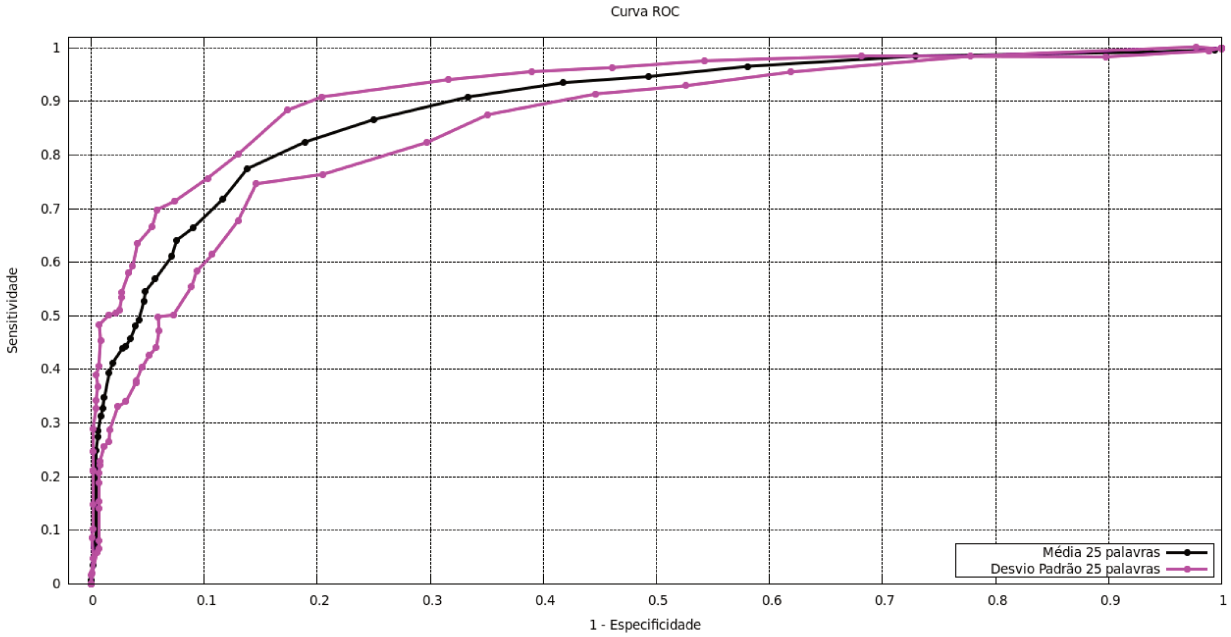


Figura 7.8: Curvas de desvio padrão utilizando 25 palavras visuais.

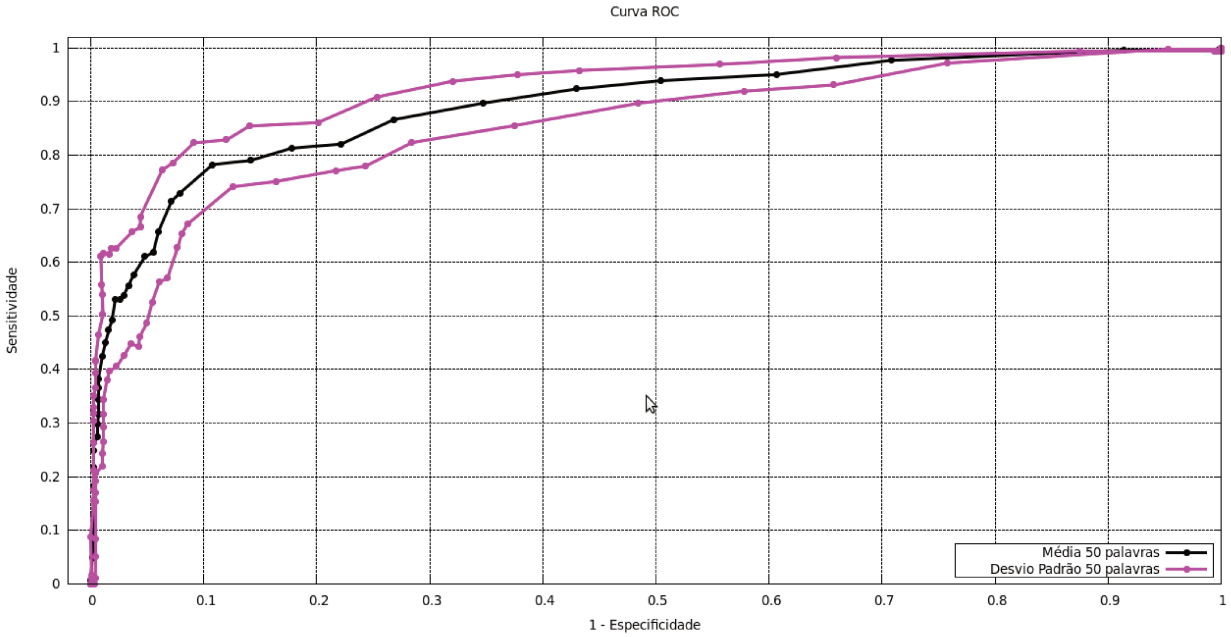


Figura 7.9: Curvas de desvio padrão utilizando 50 palavras visuais.

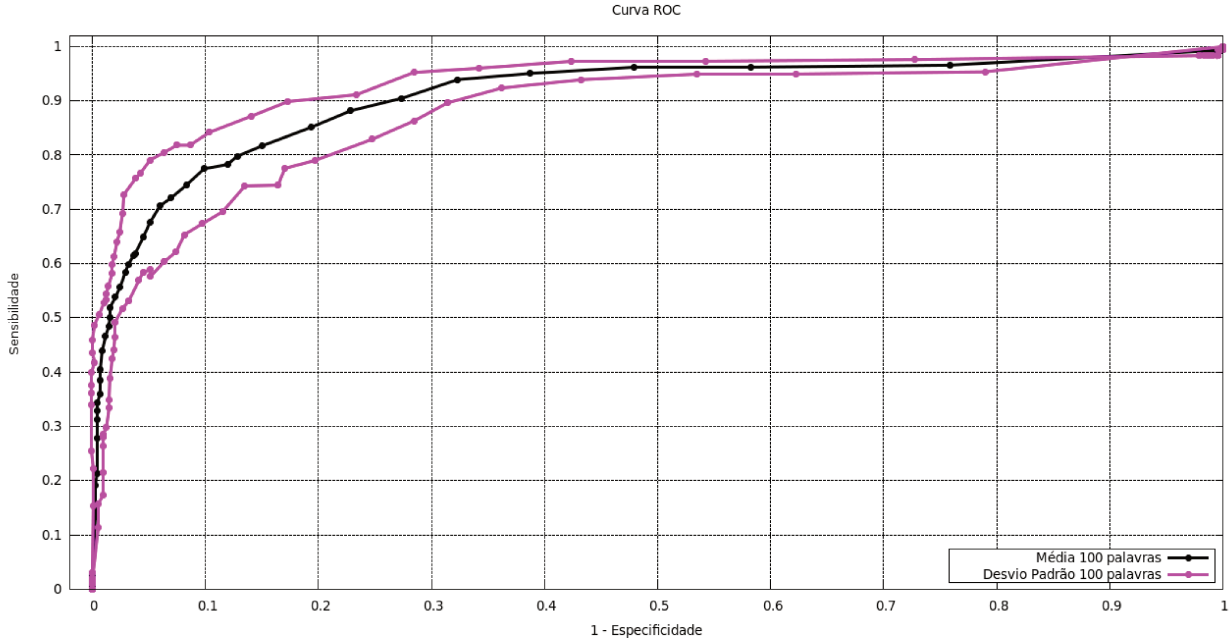


Figura 7.10: Curvas de desvio padrão utilizando 100 palavras visuais.

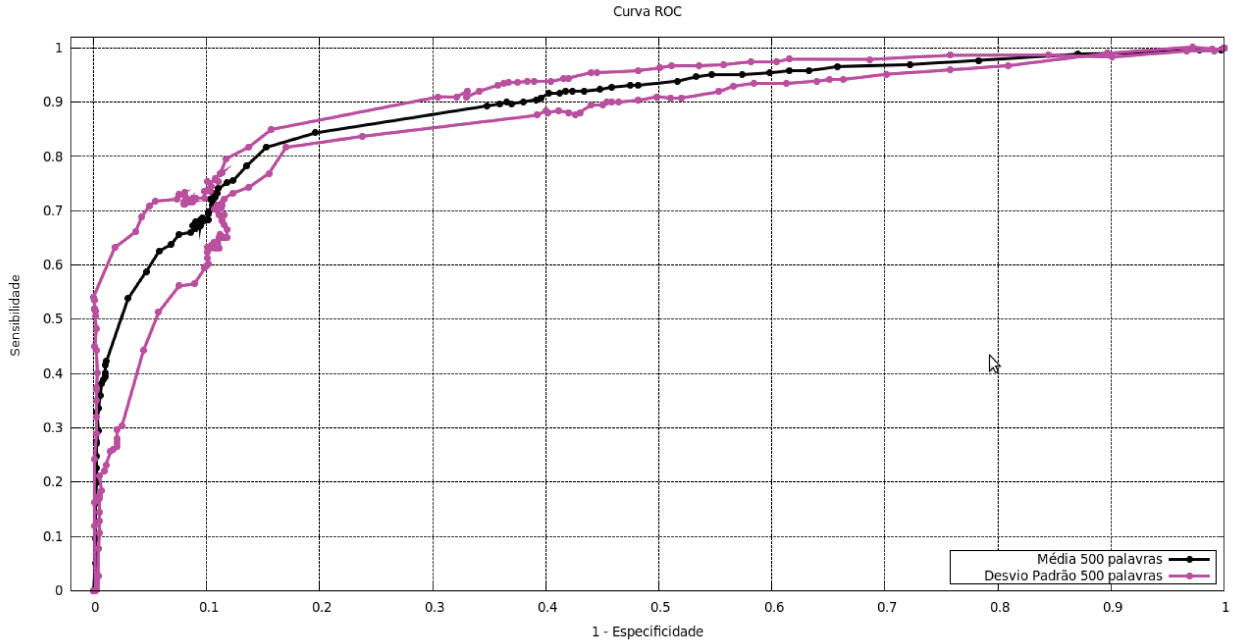


Figura 7.11: Curvas de desvio padrão utilizando 500 palavras visuais.

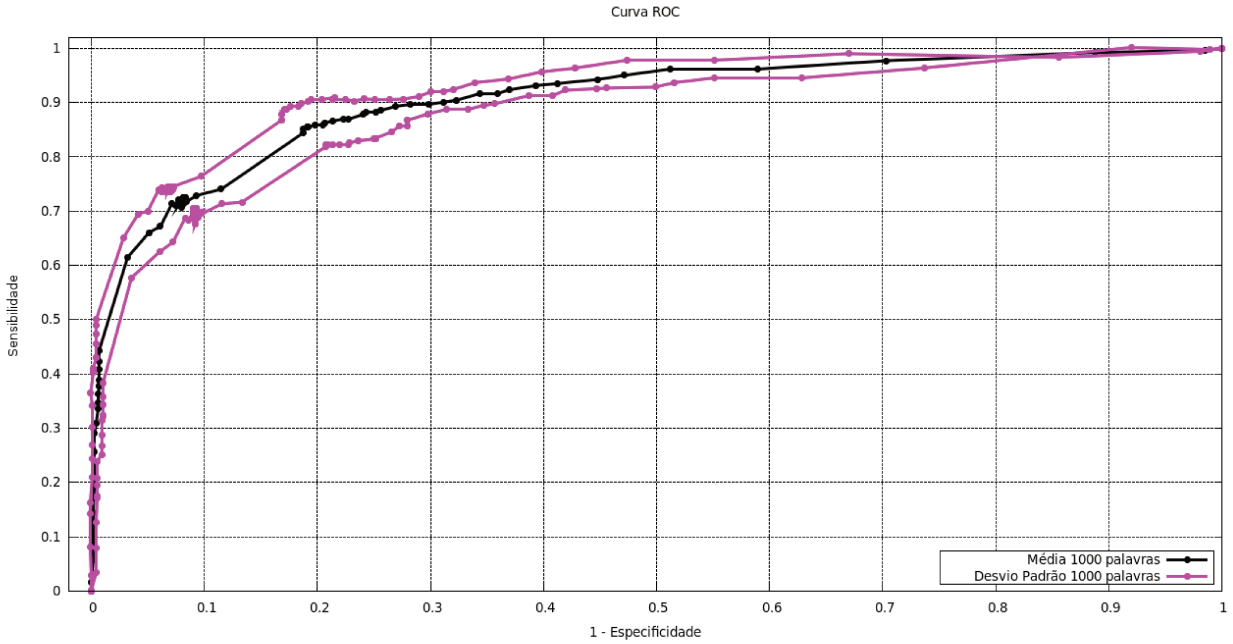


Figura 7.12: Curvas de desvio padrão utilizando 1000 palavras visuais.

Assim, neste cenário, a formação do dicionário visual, é um pouco diferente. Foram utilizadas 500 palavras visuais. Destas, 250 foram extraídas aleatoriamente apenas dos pontos pertencentes a imagens normais. Já as 250 palavras restantes, além de serem retiradas das imagens com anomalia, foram retiradas também de forma aleatória só que nas regiões marcadas por um usuário como sendo a região onde a anomalia estava contida.

Essa é a principal característica que diferencia este cenário dos demais vistos até agora. As demais características, como o algoritmo para extração dos PCs, uso tipo de classificador, etc. podem ser vistos nas Tabelas 7.13, 7.14, 7.15 e seguem o mesmo padrão dos testes anteriores.

Tabela 7.10: Experimento 4 - base de dados

amostras normais	687
amostras doentes	264
dobras normais	5
dobras doentes	5
imagens por dobras normais	~ 136
imagens por dobras doentes	~ 53

Tabela 7.11: Experimento 4 - dicionário visual

dobras treino normal	4
dobras treino exsudato	4
algoritmos de extração de PCs	SURF
PC conjunto de treinamento SURF	$\sim 700\,mil$
palavras visuais utilizadas	500 (250 imagens normais + 250 regioes exsudato)
seleção palavras visuais	aleatória dentre as regiões marcadas

Tabela 7.12: Experimento 4 - classificador

dobras treino normal	4
dobras treino exsudato	4
dobras teste normal	1
dobras teste exsudato	1
classificador	SVM
tipo de "kernel"	linear
tipo SVM	C-SVM
penalidade do classificador (C)	0.5
validação cruzada	sim

Os resultados deste cenário são apresentados no gráfico da Figura 7.13. Nele estão presentes as curvas médias provindas de um teste utilizando a seleção de regiões, outro teste utilizando apenas a seleção de palavras visuais de forma aleatória como feito nos cenários anteriores, e ainda uma curva representando os resultados de um teste utilizando o agrupamento.

O desvio padrão dos três tipos de seleção das palavras visuais são apresentados nos gráficos das Figuras 7.14, 7.15, 7.16.

O resultado mais interessante observado ao fim desse experimento é que a seleção das palavras visuais utilizando seleção de regiões consegue praticamente o mesmo resultado da seleção utilizando o agrupamento, no entanto com um desvio padrão menor.

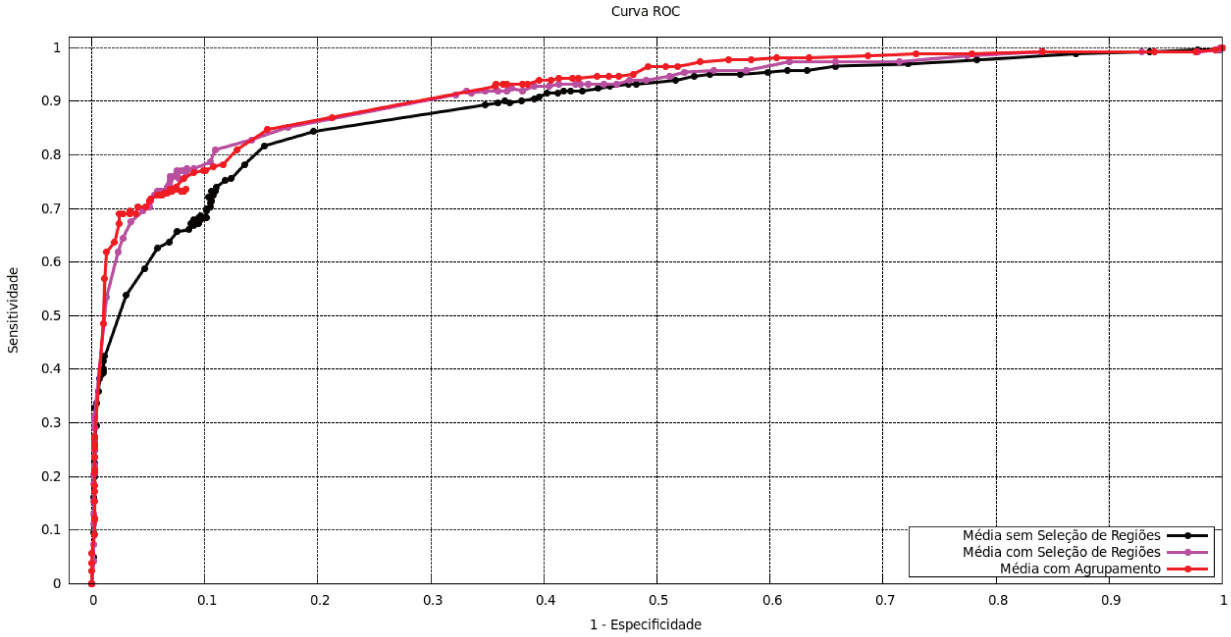


Figura 7.13: Curvas representando testes com e sem a seleção de regiões, juntamente com uma curva representando um teste com agrupamento.

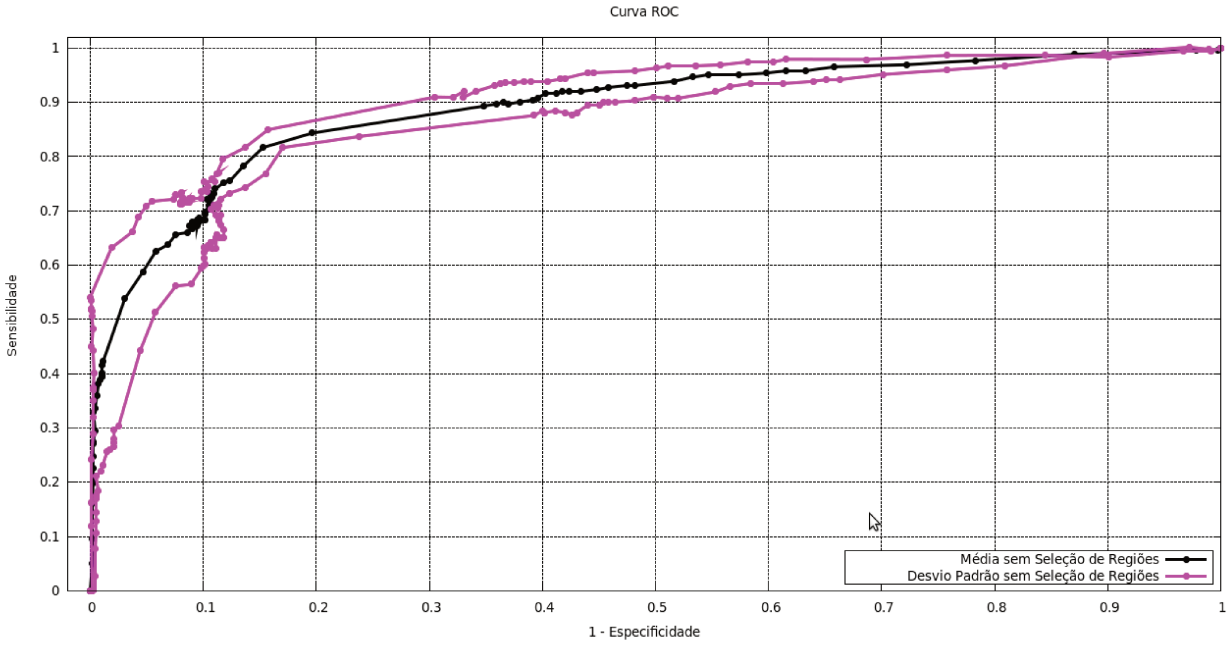


Figura 7.14: Curvas de desvio padrão não utilizando seleção de regiões.

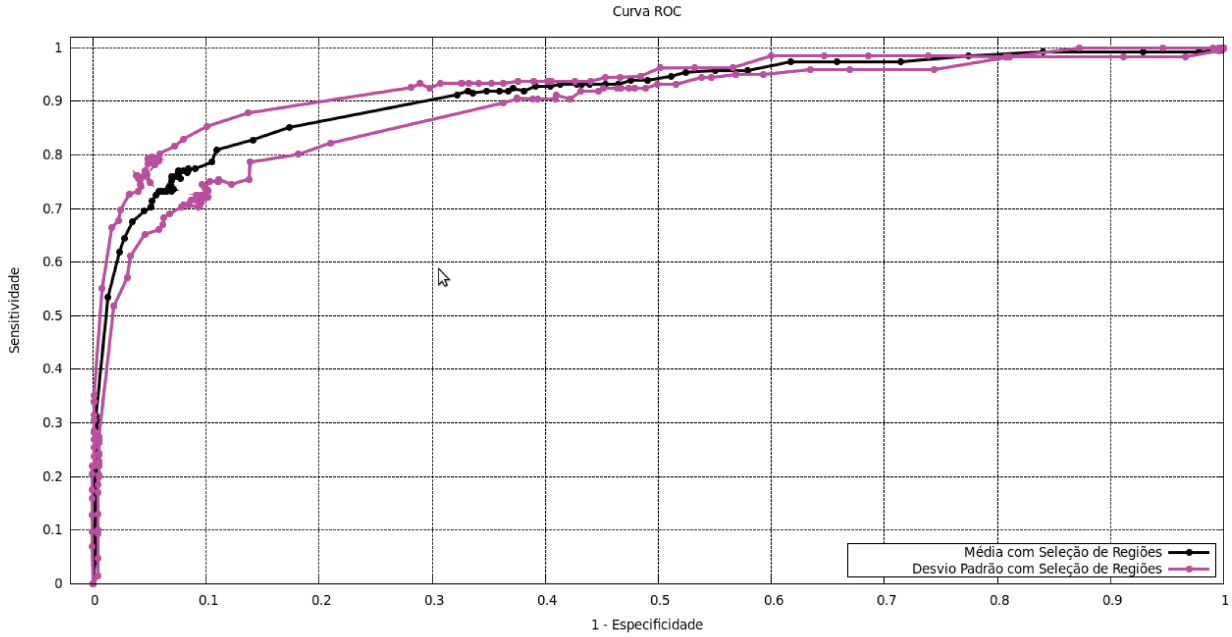


Figura 7.15: Curvas de desvio padrão utilizando seleção de regiões.

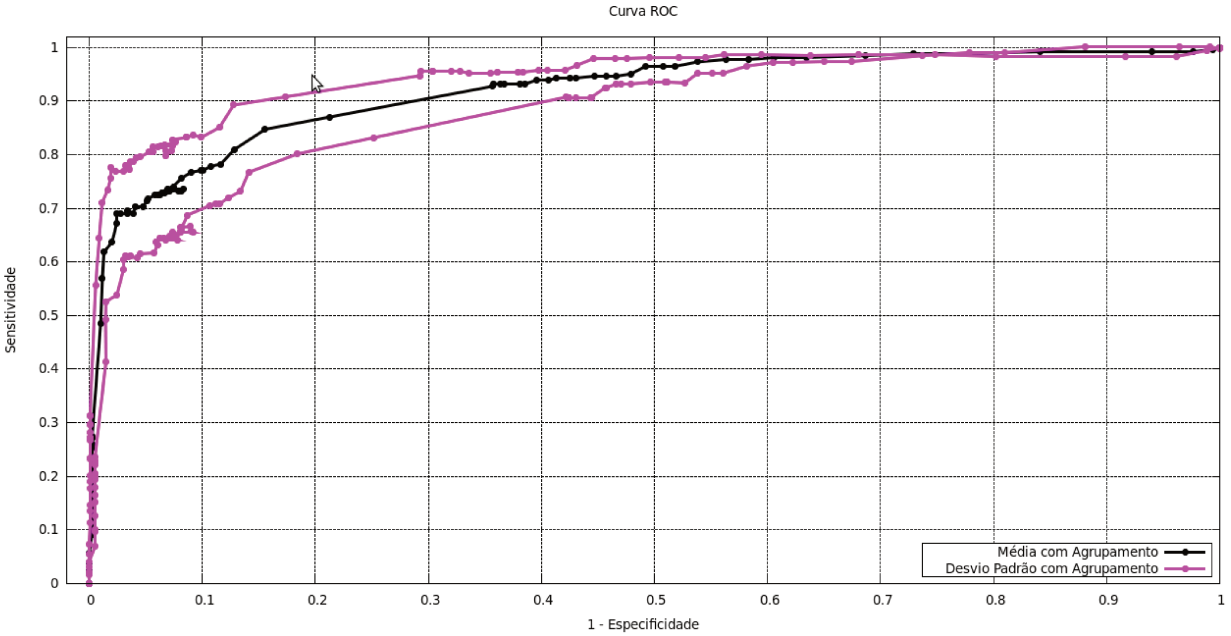


Figura 7.16: Curvas de desvio padrão utilizando agrupamento.

7.5 Seleção Manual

O quinto e último cenário de testes busca mostrar que quanto melhor é a seleção das palavras que compõem um dicionário visual melhores são os resultados na classificação das amostras.

O processo de formação do dicionário visual neste experimento começa com a extração dos PCs através do algoritmo SURF. No entanto, ao invés de gerar apenas um arquivo texto com os PCs, pra cada uma das imagens é gerada uma segunda imagem mostrando cada um dos PCs extraídos. Baseado nestas imagens, um especialista seleciona dentre os PCs extraídos, pertencentes ao conjunto de treinamento, 50 PCs que estejam visivelmente localizados em regiões de exsudato e 50 PCs que estejam visivelmente localizados em regiões que não pertençam a regiões de doença (PCs normais).

Logo, o dicionário visual deste experimento é formado por 100 palavras visuais (50 PCs de exsudato e 50 PCs normais).

Essa é a principal característica que diferencia este cenário dos demais. As demais características, como o algoritmo para extração dos PCs, uso tipo de classificador, etc. podem ser vistos nas Tabelas 7.13, 7.14, 7.15 e seguem o mesmo padrão dos testes anteriores.

Tabela 7.13: Experimento 5 - base de dados

amostras normais	687
amostras doentes	264
dobras normais	5
dobras doentes	5
imagens por dobras normais	~ 136
imagens por dobras doentes	~ 53

Tabela 7.14: Experimento 5 - dicionário visual

dobras treino normal	4
dobras treino exsudato	4
algoritmos de extração de PCs	SURF
PC conjunto de treinamento SURF	$\sim 700\, mil$
palavras visuais utilizadas	100
seleção palavras visuais	manual feita por especialista, através de regiões marcadas e aleatória

Tabela 7.15: Experimento 4 - classificador

dobras treino normal	4
dobras treino exsudato	4
dobras teste normal	1
dobras teste exsudato	1
classificador	SVM
tipo de “kernel”	linear
tipo SVM	C-SVM
penalidade do classificador (C)	0.5
validação cruzada	sim

Os resultados deste cenário são apresentados no gráfico da Figura 7.17. Nele estão presentes as curvas médias provindas de um teste utilizando a seleção de palavras visuais de forma manual e é comparado com o melhor resultado obtido até o presente momento utilizando a seleção de regiões.

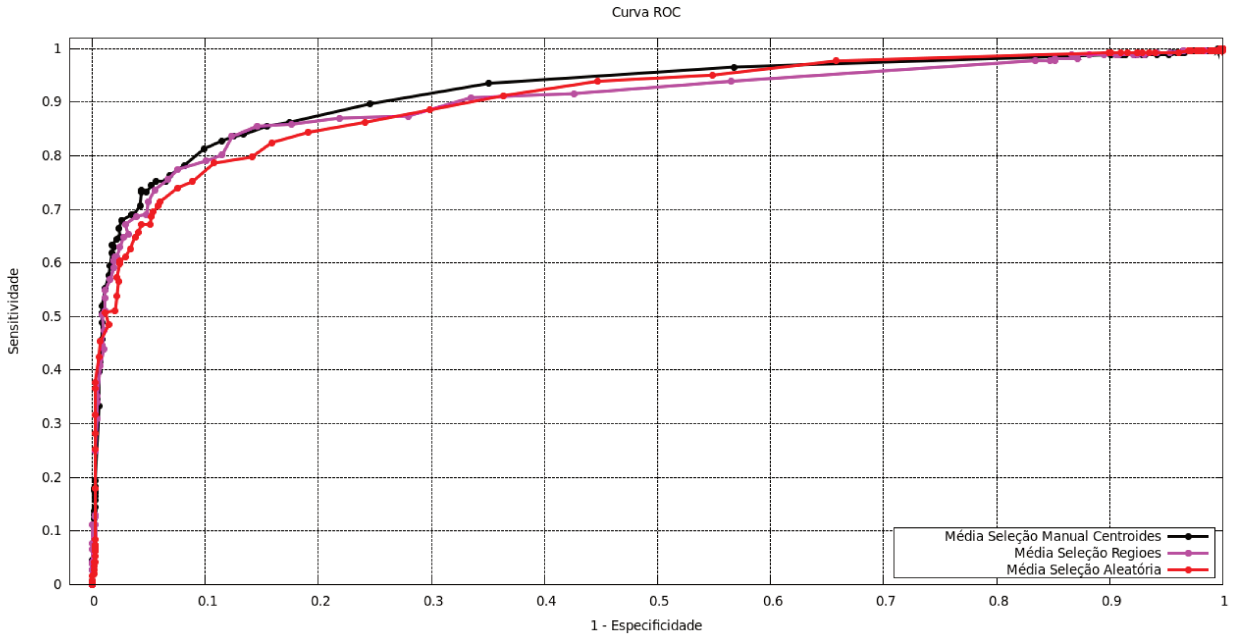


Figura 7.17: Curvas representando testes com e sem a seleção manual de palavras, juntamente com curvas representando a seleção por regiões e a seleção aleatória.

O desvio padrão dos três tipos de seleção de palavras visuais deste cenário são apresentados nos gráficos das Figuras 7.18, 7.19, 7.20.

7.6 Comparação de Resultados

A Tabela 7.6 exibe os resultados de métodos propostos para a detecção de exsudatos na literatura.

Num método de detecção de anomalias voltado para a medicina, é mais importante dar prioridade para uma alta taxa de sensitividade, uma vez que um falso negativo pode comprometer até mesmo a vida de um paciente.

Assim, o método apresentado neste trabalho busca maximizar a sensibilidade mantendo uma taxa de especifidade aceitável. Seu melhor resultado também é apresentado na Tabela 7.6.

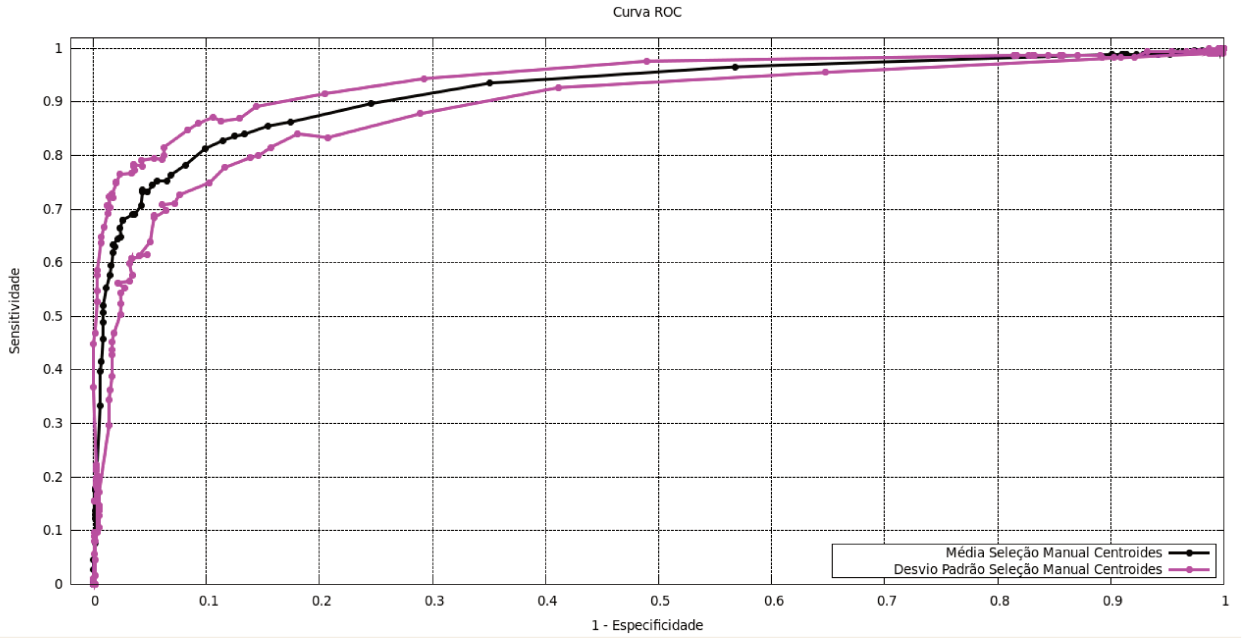


Figura 7.18: Curvas de desvio padrão utilizando seleção manual.

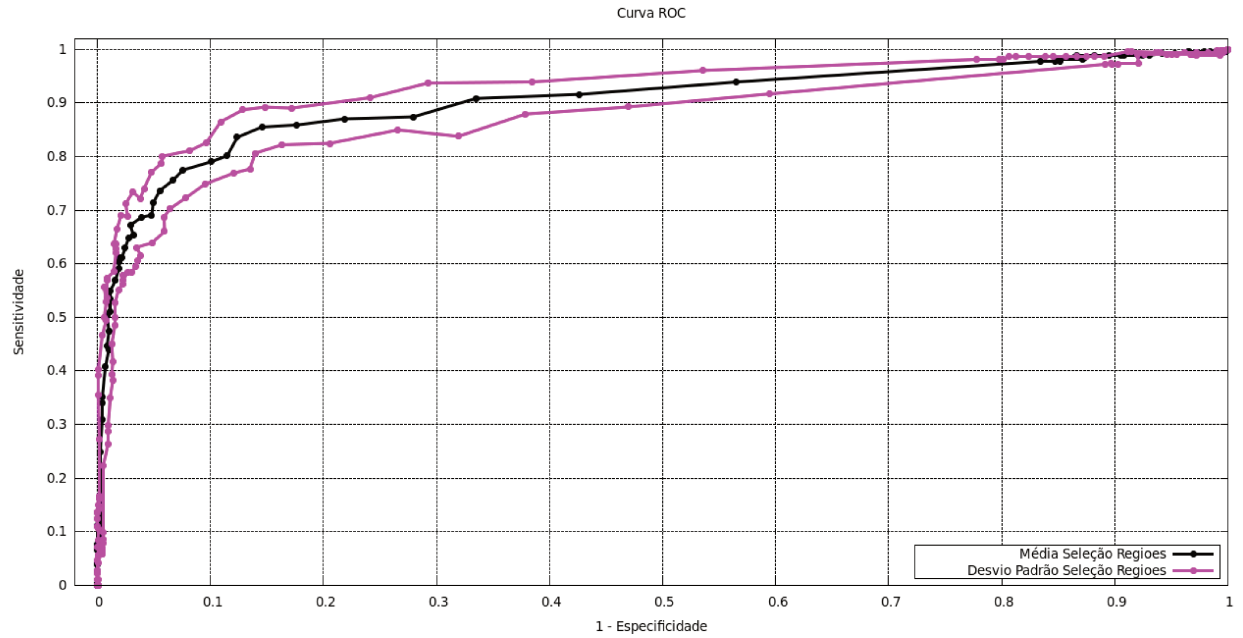


Figura 7.19: Curvas de desvio padrão utilizando seleção de regiões.

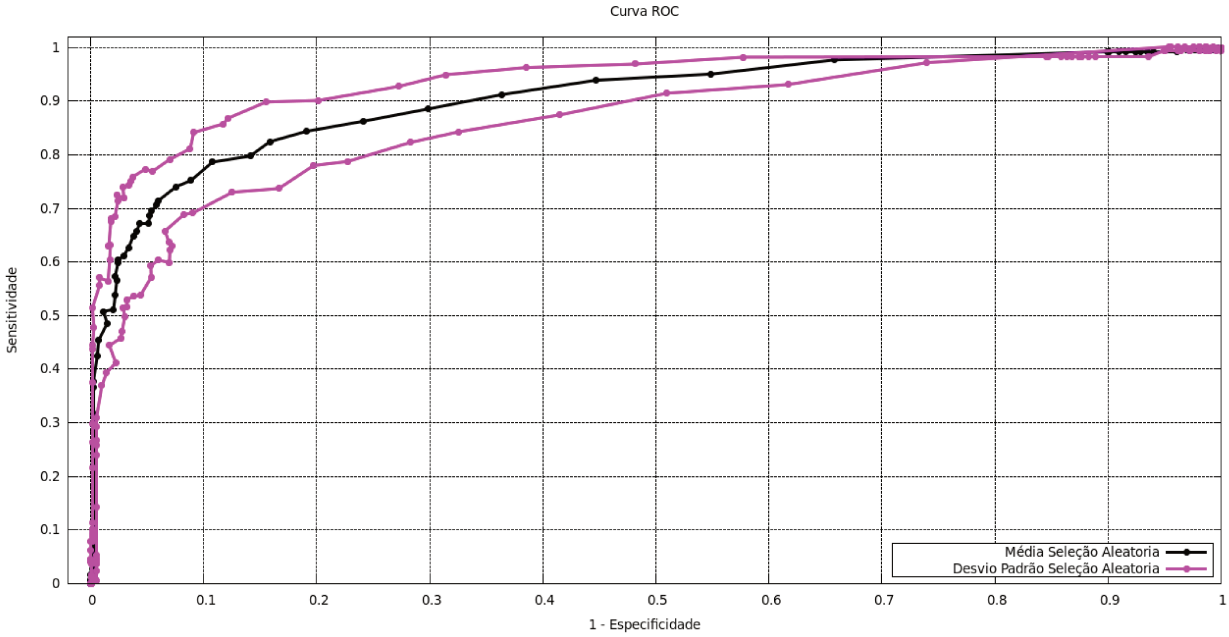


Figura 7.20: Curvas de desvio padrão utilizando seleção aleatória.

Técnica	**Sensibilidade**	**Especificidade**	**Base de Dados**	**Tipo de Abordagem**
Método Apresentado	92%	70%	951	Dicionários Visuais e SVM
Sopharak *et.al.* [83]	80%	99,5%	60	Operadores Morfológicos
Yun *et.al.* [102]	80%	99,5%	124	Operadores Morfológicos e Redes Neurais
Welfer *et.al.* [99]	70.5%	98,8%	89	Operadores Morfológicos
Wang *et.al.* [98]	100%	70%	154	Ajustes de bilho seguidos de métodos de classificação estatísticos
Garcia *et.al.* [52]	88%	84%	117	Redes Neurais e SVM sobre "patches" da imagem
Sopharak *et.al.* [82]	87,3%	99,3%	60	Operadores Morfológicos e Agrupamentos com C-Médias "Fuzzy"
Fleming *et.al.* [82]	95%	86,6%	13219	Operadores Morfológicos e Decomposição Multi-Escala

Tabela 7.16: Resultados de alguns métodos para detecção de exsudatos.

Capítulo 8

Conclusões e Trabalhos Futuros

O método proposto nesta dissertação, apresentou resultados que permitem algumas conclusões interessantes a seu respeito. Primeiro, diferentes tipos de descrições das regiões das imagens geram diferentes taxas de acerto. Devido ao fato das anomalias serem pequenas modificações nas imagens, suas regiões precisam conseguir expressar essas pequenas modificações. A técnica de dicionários visuais utilizando PCs como regiões da imagem se adapta bem a essa necessidade.

Ao longo dos experimentos, é observado que quanto melhor é a escolha das palavras visuais que formam o dicionário visual e dessa forma representar o problema, melhores os resultados. Uma prova disto é o experimento utilizando a escolha manual das palavras do dicionário. Tal experimento conseguiu os melhores resultados produzidos dentre todos os cenários, levando em consideração a quantidade de acertos.

Neste problema o aumento do número de palavras não gera melhorias significativas nos resultados de classificação. A utilização de agrupamento também não produziu uma elevação no número de acertos que justifique o grande tempo utilizado no seu processamento.

Mas, muita coisa ainda precisa ser investigada a respeito do problema. Há diversos pontos que apresentam-se promissores para futuras pesquisas. Perguntas como, o uso de agrupamentos, não mais em todo o conjunto de treinamento, mas sobre as regiões poderia tornar a escolha das palavras melhor, fazendo com que os resultados fossem elevados de forma a compensar o alto custo computacional do agrupamento?

Ou ainda, a utilização de outros tipos de classificadores poderia produzir um acerto maior? Estender esse método para a detecção não de um, mas sim de qualquer tipo de anomalia é realmente viável?

Como proposta para trabalhos futuros, é sugerido a tentativa de expansão do método para detectar qualquer tipo de anomalia presente na retina, utilizando para isso dicionários visuais com outros tipos de descrição para as regiões das imagens, bem como o emprego

de outros classificadores, ou até mesmo, um método composto por múltiplos descritores e classificadores atuando de forma conjunta para identificação de anomalias nas imagens de fundo de olho.

Referências Bibliográficas

[1] K. G. Alberti and P. Z. Zimmet. Definition, diagnosis and classification of diabetes mellitus and its complications, part 1: diagnosis and classification of diabetes mellitus provisional report of a who consultation. *Diabetic Medicine*, 15(7):539–553, 1998.

[2] P. Geissler. B. Jahne, H. Haussecker, editor. *Handbook of Computer Vision and Applications*. Academic Press, London, 1999.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, pages 404–417, 2006.

[4] P. R. Beaudet. Rotationally invariant image operators. In *International Joint Conference on Pattern Recognition*, pages 579–583, 1978.

[5] Kristin P. Bennett and Colin Campbell. Support vector machines: Hype or hallelujah? *Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining*, 2(2):1–13, 2000.

[6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[7] H. Blum. Biological shape and visual science. *Journal of Theoretical Biology*, 38(2):205â287, 1973.

[8] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008.

[9] Leo Breiman and Philip Spector. Submodel selection and evaluation in regression. the x-random case. *International Statistical Review*, 60(3):291–319, 1992.

[10] M. Brown and D. G. Lowe. Recognising panoramas. In *IEEE International Conference on Computer Vision*, page 1218. IEEE Computer Society, 2003.

[11] Matthew Brown and David Lowe. Invariant features from interest point groups. In *British Machine Vision Conference*, pages 656–665, 2002.

[12] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001.

[13] G. Chuang and C.C. Kuo. Wavelet descriptor of planar curves: Theory and applications. In *IEEE Transaction on Image Processing*, volume 5, 1996.

[14] Ronald Pitts Crick and Peeng Tee Khaw. *A Textbook of Clinical Ophthalmology*. World Scientific Publishing Company; 3 edition, 2003.

[15] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision*, 2004.

[16] Zelia Maria da Silva Correa and Ralph Eagle Jr. Aspectos patológicos da retinopatia diabética. *Revista de Oftalmologia de São Paulo*, 68(3):410–414, 2005.

[17] Ricardo da Silva Torres and Alexandre Xavier Falc ao. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, 13(2):161–185, 2006.

[18] Jurandy Gomes de Almeida Júnior. Recuperação de imagens por cor utilizando análise de distribuição discreta de características. Master's thesis, Universidade Estadual de Campinas, 2007.

[19] Hygino H. Domingues. *Espaços Métricos e Introdução a Topologia*. Atual Editora, 1982.

[20] E.R. Dougherty and R.A. Lotufo. Hands-on morphological image processing. In *SPIE PRESS*, 2003.

[21] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[22] Dean G. Duffy. *Advanced Engineering Mathematics with MATLAB*. CRC Press, 1997.

[23] Jan Eichhorn. *Applications of Kernel Machines to Structured Data*. PhD thesis, Technical University of Berlin, 2006.

[24] Oliver Faust, Rajendra Acharya U., E. Y. K. Ng, Kwan-Hoong Ng, and Jasjit S. Suri. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review. *Journal of Medicine System*, 1:1, 2010.

[25] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool. Simultaneous object recognition and segmentation by image exploration. In *European Conference on Computer Vision*, 2004.

[26] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Transactions Medical Imaging*, 25:1223–1232, 2006.

[27] Alan D Fleming, Keith A Goatman, Sam Philip, John A Olson, and Peter F Sharp. Automatic detection of retinal anatomy to assist diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(2):331–345, 2007.

[28] Alan D. Fleming, Sam Philip, Keith A. Goatman, Graeme J. Williams, John A. Olson, and Peter F. Sharp. Automated detection of exudates for diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(24):7385–7396, 2007.

[29] Tomaso Poggio Florian Wolf and Pawan Sinha. Human document classification using bags of words. Technical report, MIT, 2006.

[30] James Fogarty, Ryan S. Baker, and Scott E. Hudson. Case studies in the use of roc curve analysis for sensor-based estimates in human computer interaction. In *Graphics Interface*, pages 129–136. Canadian Human-Computer Communications Society, 2005.

[31] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[32] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice Hall; 2 edition, 2002.

[33] Michiel Hazewinkel, editor. *Encyclopaedia of Mathematics*. Springer, 1995.

[34] Henk J.A.M. Heijmans. *Morphological Image Operators*. Academic Press, 1994.

[35] Tinne Tuytelaars Luc Van Gool Herbert Bay, Andreas Ess. Speeded-up robust features (surf). *Computer Vision Image Understand*, 110(3):346–359, 2008.

[36] Thomas Hofmann. Probabilistic latent semantic indexing. In *ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, New York, NY, USA, 1999. ACM.

[37] Adam Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19:203–210, 2000.

[38] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, National Taiwan University, 2003.

[39] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[40] Michel Grabisch Jean Figue and Marie Pierre Charbonnel. A method for still image interpretation relying on a multi-algorithms fusion scheme. application to human face characterization. *Fuzzy Sets and Systems*, 103:317–337, 1999.

[41] Frederic Jurie and Bill Triggs. Creating efficient codebooks for visual recognition. In *IEEE International Conference on Computer Vision*, volume 1, 2005.

[42] S. J. Dickinson K. Siddiqi, A. Shokoufandeh and S. W. Zucker. Shock graphs and shape matching. In *IEEE International Conference on Computer Vision*, pages 222–229., 1998.

[43] A. Khotanzad and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):489–497, 1990.

[44] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Inteligence*, 1995.

[45] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[46] Diane Larlus and Frederic Jurie. Latent mixture vocabularies for object categorization and segmentation. *Image Vision Computing*, 27(5):523–534, 2009.

[47] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, 2006.

[48] Fei-Fei Li and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 524–531, 2005.

[49] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[50] E. Grisan M. Foracchia and A. Ruggeri. Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, 9(3):179–190, 2005.

[51] Stephane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2008.

[52] Maria I. Lopez Daniel Abasolo Roberto Hornero Maria Garcia, Clara I. Sanchez. Neural network based detection of hard exudates in retinal images. *Computer Methods and Programs in Biomedicine*, 93:9–19, 2009.

[53] Edson Zangiacomi Martinez, Francisco Louzada-Neto, and Basílio de Bragança Pereira. A curva roc para testes diagnósticos. *Cadernos Saúde Coletiva (Rio de Janeiro)*, 11(1):7–31, 2003.

[54] Maria S.A. Suttorp-Schulten Max A. Viergever Stephen R. Russell Bram van Ginneken Michale D. Abràmoff, Meindert Niemeijer. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care*, 31(2):193–198, 2008.

[55] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *IEEE International Conference on Computer Vision*, pages 525–531, 2001.

[56] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[57] Krystian Mikolajczyk, Bastian Leibe, and Bernt Schiele. Local features for object class recognition. In *IEEE International Conference on Computer Vision*, pages 1792–1799, Washington, DC, USA, 2005. IEEE Computer Society.

[58] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[59] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[60] Farzin Mokhtarian, Sadegh Abbasi, and Josef Kittler. Efficient and robust retrieval by shape content through curvature scale space. In *International Workshop on Image Databases and Multimedia Search*, pages 35–42, 1996.

[61] Javier A. Montoya-Zegarra, Neucimar J. Leite, and Ricardo da S. Torres. Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 121–128. IEEE Computer Society, 2007.

[62] Jagadish Nayak, P. Subbanna Bhat, Rajendra Acharya U, C. M. Lim, and Manjunath Kagathi. Automated identification of diabetic retinopathy stages using digital fundus images. *Journal of Medical Systems*, 32(2):107–115, 2008.

[63] Alexander Neubeck and Luc Van Gool. Efficient non-maximum suppression. In *International Conference on Pattern Recognition*, August 2006.

[64] Otávio Augusto Bizetto Penatti. Estudo comparativo de descritores para recuperação de imagens por conteúdo na web. Master's thesis, Universidade Estadual de Campinas (UNICAMP), 2009.

[65] Otávio Augusto Bizetto Penatti and Ricardo da Silva Torres. Color descriptors for web image retrieval: A comparative study. In *Brazilian Symposium on Computer Graphics And Image Processing*, pages 163–170. IEEE Computer Society, 2008.

[66] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[67] Yimin Zhang Qi Zhang, Yurong Chen and Yinlong Xu. Sift implementation and optimization for multi-core systems. In *IEEE International Symposium in Parallel and Distributed Processing*, 2008.

[68] K. Shanmugam R. M. Haralick and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.

[69] P. E. Hart R. O. Duda and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience; 2 edition, 2000.

[70] Anderson Rocha. Randomização progressiva para esteganálise. Master's thesis, Universidade Estadual de Campinas, 2006.

[71] Anderson Rocha, Daniel C. Hauagge, Jacques Wainer, and Siome Goldenstein. Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture*, 70(1):96–104, 2010.

[72] Jos B. T. M. Roerdink and Arnold Meijster. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41(1-2):187–228, 2000.

[73] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision.*, 40(2):99–121, 2000.

[74] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[75] William Robson Schwartz and Hélio Pedrini. Método para classificação de imagens baseada em matrizes de co-ocorrência utilizando características de textura. In *Colóquio Brasileiro de Ciências Geodésicas*, 2003.

[76] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine*, 19(2):105–112, 2002.

[77] Chanjira Sinthanayothin, James F. Boyce, Helen L. Cook, and Thomas H. Williamson. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophtalmology*, 83:902–910, 1999.

[78] Josef Sivic, Bryan Russell, Alexei A. Efros, Andrew Zisserman, and Bill Freeman. Discovering objects and their location in images. In *International Conference on Computer Vision*, 2005.

[79] Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. In *9th IEEE International Conference on Computer Vision*, pages 1470–1477, 2003.

[80] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM Special Interest Group on Computer Graphics*, pages 835–846, New York, NY, USA, 2006. ACM.

[81] Pierre Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., 2003.

[82] Akara Sopharak, Bunyarit Uyyanonvara, and Sarah Barman. Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy c-means clustering. *Sensors*, 9(3):2148–2161, 2009.

[83] Akara Sopharak, Bunyarit Uyyanonvara, Sarah Barman, and Thomas H. Williamson. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 32:8, 2008.

[84] Akara Sopharak, Bunyarit Uyyanonvara, Sarah Barman, and Thomas H. Williamson. Comparative analysis of automatic exsudate detection between machine learning and traditional approaches. *IEICE Transactions on Information and Systems*, 92:2264–2271, 2009.

[85] Renato O. Stehling, Mario A. Nascimento, and Alexandre X. Falc ao. Cell histograms versus color histograms for image representation and retrieval. *Knowledge and Information System*, 5(3):315–336, September 2003.

[86] Tatiane Stein. Avaliação de descritores de textura para segmentação de imagens. Master's thesis, Universidade Federal do Paraná, 2005.

[87] Markus Stricker and Markus Orengo. Similarity of color images. In *Storage and Retrieval of Image and Video Databases III*, volume 2, pages 381–392, 1995.

[88] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[89] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley Longman Publishing, 2005.

[90] Kenneth W. Tobin, E. Chaum, V. Priya Govindasamy, and Thomas P. Karnowski. Detection of anatomic structures in human retinal imagery. *IEEE Transactions on Medical Imaging*, 26(12):1729–1739, 2007.

[91] Tinne Tuytelaars and Luc J. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *International Conference on Visual Information and Information Systems*, pages 493–500, London, UK, 1999. Springer-Verlag.

[92] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant features detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 07 2007.

[93] Ilkay Ulusoy and Christopher M. Bishop. Generative versus discriminative methods for object recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 2005.

[94] D. Usher, M. Dumskyj, M. Himaga, T.H. Williamson, S. Nussey, and J. Boyce. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabetic medicine : a journal of the British Diabetic Association*, 21(1):84–90, 2004.

[95] Eduardo Valle and Matthieu Cord. Advanced techniques in cbir: Local descriptors, visual dictionaries and bags of features. *Tutorials of the Brazilian Symposium on Computer Graphics and Image Processing*, 0:72–78, 2009.

[96] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[97] T. Walter, J. C. Klein, P. Massin, and A. Erginay. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Transactions on Medical Imaging*, 21(10):1236–1243, 2002.

[98] Huan Wang, Wynne Hsu, Kheng Guan Goh, and Mong Li Lee. An effective approach to detect lesions in color retinal images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[99] Daniel Welfer, Jacob Scharcanski, and Diane Ruschel Marinho. A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images. *Computerized Medical Imaging and Graphics*, 34(3):228–235, 2010.

[100] A. P. Witkin. Scale-space filtering. In *International Joint Conferences on Artificial Intelligence*, volume 2, pages 1019–1022., 1983.

[101] Shiming Xiang, Feiping Nie, and Changshui Zhang. Learning a mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 41(12):3600–3612, 2008.

[102] Wong Li Yun, U. Rajendra Acharya, Y. V. Venkatesh, Caroline Chee, Lim Choo Min, and E. Y. K. Ng. Identification of different stages of diabetic retinopathy using retinal optical images. *Information Sciences: an International Journal*, 178:106–121, 2008.

[103] J. A. M. Zegarra, J. P. Papa, N. J. Leite, R. da S. Torres, and A. X. Falc ao. Learning how to extract rotation-invariant and scale-invariant features from texture images. *Eurasip Journal on Advances in Signal Processing*, 2008:15 pages, 2008.

[104] M.H. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.