

# CLASSIFICAÇÃO ADABOOST PARA DETECÇÃO E CONTAGEM AUTOMÁTICA DE PLAQUETAS

Trabalho de Conclusão de Curso

Engenharia de Computação

Aluna: Débora Natália de Oliveira Nascimento

Orientador: Prof. Dr. Carmelo José Albanez Bastos Filho

Débora Natália de Oliveira Nascimento

*Classificação AdaBoost para detecção e  
contagem automática de plaquetas*

Monografia apresentada para obtenção do Grau  
de Bacharel em Engenharia de Computação  
pela Universidade de Pernambuco

Orientador:

Prof. Dr. Carmelo José Albanez Bastos Filho

GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO  
ESCOLA POLITÉCNICA DE PERNAMBUCO  
UNIVERSIDADE DE PERNAMBUCO

Recife - PE, Brasil

Dezembro de 2011

Escola Politécnica de Pernambuco  
Graduação em Engenharia de Computação  
Projeto de Final de Curso



### MONOGRAFIA DE FINAL DE CURSO

#### Avaliação Final (para o presidente da banca)\*

No dia 20 de Dezembro de 2011, às 14:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente DEBORA NATALIA DE OLIVEIRA NASCIMENTO, orientado pelo professor Carmelo Bastos Filho, sob título Classificação adaboost para detecção e contagem automática de plaquetas, a banca composta pelos professores:

**Rita de Cássia Moura do Nascimento**

**Carmelo Bastos Filho**


Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

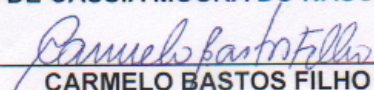
Aprovada       Aprovada com Restrições\*       Reprovada

e foi-lhe atribuída nota: 10.0 ( dez )

\*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 03 dias para entrega da versão final da monografia a contar da data deste documento.

  
\_\_\_\_\_  
RITA DE CÁSSIA MOURA DO NASCIMENTO

  
\_\_\_\_\_  
CARMELO BASTOS FILHO

\* Este documento deverá ser encadernado juntamente com a monografia em versão final.

## *Resumo*

Esse trabalho apresenta um novo método automatizado de realizar a detecção e contagem de plaquetas em imagens sanguíneas. Para o desenvolvimento do sistema de detecção serão utilizadas técnicas de processamento digital de imagens e o algoritmo de classificação *AdaBoost*. O objetivo é construir um sistema de contagem que tenha baixo custo de implementação e manutenção e que não seja sensível à presença de anormalidades como agregados plaquetários, macroplaquetas, plaquetas gigantes, e de fragmentos celulares nas amostras sanguíneas. Para a redução de ruídos nas imagens, assim como acentuação de detalhes antes despercebidos serão utilizadas as técnicas de equalização global e local do histograma da imagem. O processamento será realizado utilizando tanto as imagens em escala de cinza, quanto o canal verde (G) das imagens. O *AdaBoost* será utilizado para a classificação dos elementos na imagem, afim de identificar as características correspondentes às plaquetas, e possibilitar o seu reconhecimento e contagem. Como protótipo foi projetado um sistema simples para detecção de plaquetas e validação da solução proposta. Os testes do sistema apontaram taxas de identificação de plaquetas de 98% de acerto, 2% de falsos negativos e 3,1% de falsos positivos, utilizando imagens em tons de cinza e equalizadas globalmente. Os resultados dos experimentos realizados para análise de desempenho da solução desenvolvida mostraram que nossa proposta é robusta e é capaz de identificar as plaquetas nas imagens com uma boa taxa de acerto, e apresentando baixa taxa de falsos positivos. Também foi possível observar a influência das técnicas utilizadas para pré-processamento das imagens no resultado final do treinamento da solução de detecção.

## *Abstract*

This work presents a new method to perform automated detection and enumeration of platelets in blood images. The detection system will be developed using digital image processing techniques and the classification algorithm AdaBoost. The goal is to build a counting system with low implementation and maintenance costs and this system shouldn't be sensitive to the presence of abnormalities such as: platelet aggregates, giant platelets, and cell fragments present in blood samples. To reduce noise in images, as well as accent detail unnoticed before, will be used techniques to equalize the image histogram locally and globally. The processing will be performed using both gray-scale images, as the green channel of the images. The AdaBoost will be used to classify the elements in the image, in order to identify the characteristics corresponding to platelets, and enable your recognition and counting. As a prototype was designed a simple system to detect platelet and validate the proposed solution. The system tests has shown identification rates of platelets from 98% hit, 2% of false negative and 3.1% of false positive, using globally equalized grayscale images. The results of experiments performed to performance analysis of the developed solution showed that our proposal is robust and is able to identify platelets in images with a good hit rate, and presenting low rate of false positives. It was also possible to observe the influence of the techniques used for pre-processing of images in the final result of the training solution detection.



## *Agradecimentos*

Primeiramente agradeço aos meus pais Célia e Tita, por todo o suporte, carinho, atenção, amor, paciência e a base educacional necessária para que eu chegasse até aqui. Agradeço também ao meu irmão Victor, pela orientação e apoio ao longo do meu curso, dosada com uma pitada de provocação e desdém entre engenheiros eletrônicos e aspirantes à engenheiros da computação. Agradeço às minhas primas Élida e Tamires, e ao meu irmão Vilson, pela paciência com os pratos não lavados e a bagunça aleatória deixada nos momentos de "crise".

Ao professor Carmelo J. A. Bastos-Filho, por ter me dado uma excelente orientação durante os anos em que fui sua aluna de Iniciação Científica, bem como durante esse trabalho. Aos demais professores de computação, todos vocês participaram fundamentalmente na minha formação acadêmica. E à Ana Georgina por toda a ajudinha durante a graduação. À médica Sílvia Monteiro e à Prof. Rita de Cássia Moura do Nascimento pela colaboração durante este trabalho. Aos alunos de iniciação científica Alessandra Maranhão e Henrique Andrade pela ajuda em todo o processo de desenvolvimento deste trabalho.

Aos meus amigos que encontrei durante estes anos de caminhada na POLI, o meu muito obrigado. São tantos nomes a serem citados, mas faço um agradecimento especial à Andréa Santos, Kenelly Almeida, Edgar Almeida, Rodrigo Carneiro, Cristóvão Rufino, Fellipe Férrer e Ricardo Teixeira, que tornaram esse meu caminho muito mais agradável e as madrugadas de projetos e estudos menos sonolentas.

Agradeço também a todos os amigos além do âmbito faculdade por terem todos contribuído de alguma forma para o que sou hoje. Vocês, sem dúvidas, contribuíram com este trabalho.

Por fim agradeço a Deus por ter colocado todas essas pessoas em minha vida, e por ter me dado a sabedoria, saúde e determinação necessárias durante toda a minha caminhada.

À todos vocês, muito obrigada!

*"(...) We don't get a chance to do that many things,  
and every one should be really excellent.*

*Because this is our life.*

*Life is brief, and then you die,*

*You know?*

*And we've all chosen to do this with our lives.*

*So it better be damn good.*

*It better be worth it."*

**Steve Jobs**

# *Sumário*

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Algoritmos</b>	<b>xi</b>
<b>Lista de Abreviaturas, Siglas e Estrangeirismos</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação e Caracterização do Problema . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Organização do Documento . . . . .	3
<b>2 Tópicos sobre Plaquetas</b>	<b>4</b>
2.1 Morfologia das Plaquetas . . . . .	4
2.2 Anomalias Morfológicas e Quantitativas . . . . .	4
2.3 Técnicas de Contagem de Plaquetas . . . . .	6
<b>3 Processamento de Imagens</b>	<b>8</b>
3.1 Introdução . . . . .	8
3.2 Sistema de Processamento de Imagens . . . . .	8
3.3 Imagem Digital . . . . .	9
3.4 Sistemas de amostragem de cor . . . . .	12
3.5 Pré Processamento . . . . .	12
3.5.1 Grayscale (escala de cinza) . . . . .	12



3.5.2	Canais da Imagem . . . . .	13
3.5.3	Equalização do Histograma . . . . .	13
3.6	Segmentação de imagens . . . . .	14
3.6.1	Segmentação por detecção de borda . . . . .	14
3.7	Classificação e Reconhecimento de Padrões . . . . .	15
<b>4</b>	<b>Classificação e Detecção</b>	<b>18</b>
4.1	Método de detecção de faces proposto por Viola e Jones . . . . .	18
4.2	Imagem Integral . . . . .	19
4.3	Características de Haar . . . . .	19
4.4	Algoritmo de Aprendizado . . . . .	20
4.4.1	Classificador Fraco . . . . .	21
4.4.2	<i>AdaBoost</i> . . . . .	21
4.5	Cascata de Classificadores . . . . .	22
4.5.1	Treinamento da Cascata de Classificadores . . . . .	23
<b>5</b>	<b>Solução Proposta, Experimentos e Resultados</b>	<b>25</b>
5.1	Solução Proposta . . . . .	25
5.2	Arranjo Experimental . . . . .	25
5.2.1	Pré-processamento das imagens . . . . .	25
5.2.2	Desenvolvimento do sistema de detecção de plaquetas . . . . .	27
5.3	Análise dos Resultados . . . . .	29
<b>6</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>38</b>
6.1	Conclusão . . . . .	38
6.2	Trabalhos Futuros . . . . .	38
	<b>Referências</b>	<b>40</b>

## *Lista de Figuras*

1	<i>Plaquetas Normais.</i> . . . . .	5
2	Ilustração de anomalias plaquetárias. . . . .	6
3	<i>Modelo de um Sistema de Processamento de Imagens.</i> . . . . .	9
4	<i>Processo de Amostragem.</i> . . . . .	10
5	<i>Processo de quantização.</i> . . . . .	11
6	Exemplo da aplicação do filtro de tons de cinza em uma imagem onde, o lado esquerdo da figura corresponde à imagem original e o lado direito é o resultado da aplicação do filtro. . . . .	13
7	<i>Aplicação dos filtros de detecção de borda e suas respectivas máscaras.</i> . . . .	17
8	Exemplo da imagem de entrada $i$ e suas sub-janelas A, B, C e D. . . . .	20
9	Características básicas de Haar. . . . .	20
10	Representação de uma cascata de detecção. Cada estágio da cascata contém um classificador que irá avaliar o conjunto de entrada. . . . .	23
11	<i>Pré-processamento das imagens utilizando equalização global do histograma.</i> . .	31
12	<i>Pré-processamento das imagens utilizando equalização local do histograma.</i> . .	32
13	<i>Aplicação dos filtros de detecção de borda na imagem em tons de cinza após equalização global do histograma.</i> . . . . .	33
14	<i>Aplicação dos filtros de detecção de borda no canal G da imagem após equalização global do histograma.</i> . . . . .	34
15	<i>Aplicação dos filtros de detecção de borda na imagem em tons de cinza após equalização local do histograma.</i> . . . . .	35
16	<i>Aplicação dos filtros de detecção de borda no canal G da imagem após equalização local do histograma.</i> . . . . .	36
17	Gráfico da distribuição de acertos para cada experimento. . . . .	36

18 Gráfico da distribuição de falsos positivos para cada experimento. . . . . 37

## *Lista de Tabelas*

1	Principais causas de Plaquetopenias e Plaquetoses. . . . .	7
2	Classificação das imagens utilizadas nos experimentos de desenvolvimento do sistema de detecção de plaquetas. . . . .	27
3	Resultados médios e (desvio padrão), para o treinamento do <i>AdaBoost</i> por experimento. . . . .	29
4	Resultados médios e (desvio padrão), para o tempo de treinamento do <i>AdaBoost</i> por experimento. . . . .	29

## *Lista de Algoritmos*

1	Pseudocódigo do <i>AdaBoost</i> . . . . .	22
2	Pseudocódigo do treinamento de uma cascata de classificadores . . . . .	24
3	Pseudocódigo sistema de detecção de plaquetas . . . . .	28

# *Lista de Abreviaturas, Siglas e Estrangeirismos*

**AdaBoost** – *Adaptative Boost.*

**Boosting** – Algoritmo de aprendizagem supervisionada.

**C/C++** – Linguagem de programação.

**Grayscale** – Tons de cinza.

**Pixel** – *Picture Elements.*

**RGB** – Sistema de cores RGB.

**PDI** – Processamento digital de imagens.

**SVM** – Máquina de Vetor de Suporte (do inglês, Support Vector Machine).

**ASH** – *American Society of Hematology.*

# 1 *Introdução*

*“Nada é tão poderoso no mundo como uma ideia  
cuja oportunidade chegou.”*

– Victor Hugo.

Neste trabalho de conclusão de curso, foi utilizada uma técnica de classificação baseada em Classificadores fracos, chamada de *AdaBoost*, em conjunto com técnicas de pré-processamento digital de imagens, para a construção de um sistema de detecção e contagem automática de plaquetas em imagens sanguíneas.

Neste capítulo é apresentada a introdução deste documento, que está organizada em 3 seções. Na Seção 1.1, é apresentada a motivação para a realização deste trabalho bem como o problema abordado pelo mesmo. Em seguida, na Seção 1.2, são apresentados os objetivos gerais e específicos e a hipótese de como o problema poderá ser solucionado. Por fim, na Seção 1.3 é descrita a estrutura do restante do documento.

## 1.1 Motivação e Caracterização do Problema

As plaquetas são células sanguíneas que têm como principal função atuar na coagulação sanguínea em processos hemorrágicos [1].

A contagem de plaquetas é importante para que sejam identificadas alterações plaquetárias, tanto qualitativas, quanto quantitativas. A redução do número de plaquetas, conhecida por plaquetopenia, pode ser ocasionada por doenças, como as viroses em geral, anemias ou mesmo leucemias agudas. Já o aumento da quantidade de plaquetas, pode ser ocasionado em períodos pós-hemorrágicos, doenças infecciosas crônicas, Leucemia Mielóide, dentre outras complicações [2].

Existem métodos manuais e automatizados para realizar a contagem das plaquetas. O método mais utilizado é a contagem manual por microscopia, que também pode servir para



controlar a qualidade de processos de contagem automatizados. Nos casos em que se usa analisadores hematológicos mais modernos, uma triagem das amostras é feita previamente para definir quais delas necessitam passar por contagem manual [1]. Já os processos de contagem automatizados podem utilizar tecnologia de impedância, dispersão de luz, ou ambas [3]. Apesar de produzirem resultados mais precisos, as contagens automatizadas estão sujeitas a falhas. A presença de agregados plaquetários, plaquetas gigantes ou macroplaquetas, por exemplo, pode apontar um falso cenário de plaquetopenia. De forma que a presença de fragmentos de hemácias, por exemplo, pode resultar em um diagnóstico errôneo de aumento do número de plaquetas [1]. É importante salientar também que a aquisição, operação e manutenção de equipamentos mais modernos para realizar a contagem das células implica em um alto custo, sendo inviável principalmente para laboratórios de pequeno porte.

Diante desse contexto, se faz necessária a busca por soluções alternativas, que proporcionem resultados precisos acerca da contagem das células e que apresentem baixo custo, de modo que estas soluções possam ser utilizadas por pequenos laboratórios.

A inteligência computacional possui técnicas capazes de proporcionar um melhor desempenho e maior eficiência no treinamento de sistemas baseados em técnicas de reconhecimento de padrões. Utilizando inteligência computacional é possível construir sistemas capazes de aprender e se adaptar às modificações do ambiente a partir de uma base de conhecimento prévia. Tais características tornam estas técnicas bastante promissoras para solucionar problemas de reconhecimento de padrões, dado que, uma vez treinadas para reconhecer determinado padrão, são capazes classificar corretamente novos exemplos que não estavam em sua base de conhecimento. Sistemas de reconhecimento de padrões utilizando técnicas de inteligência computacional podem ser uma boa alternativa para o desenvolvimento de uma ferramenta de auxílio na identificação e contagem automática de plaquetas a partir de imagens sanguíneas.

## 1.2 **Objetivos**

Neste trabalho propõe-se desenvolver um sistema computacional capaz de auxiliar na identificação e contagem automática de plaquetas. Este sistema deve ter baixo custo de implementação e manutenção, e não deve ser sensível à presença de agregados plaquetários, macroplaquetas, plaquetas gigantes, bem como fragmentos celulares nas amostras sanguíneas. Para a construção do sistema ora proposto, foi utilizada uma técnica de classificação baseada em Classificadores fracos, chamada de *AdaBoost*, em conjunto com técnicas de pré-processamento digital de imagens.

A partir do objetivo maior deste trabalho, pode-se definir outros objetivos específicos, como: (i) realizar um estudo sobre técnicas de realce e extração de características de imagens, (ii) realizar um estudo para aplicação das técnicas de reconhecimento de padrões, analisando os algoritmos adequados e ao mesmo tempo viáveis para o problema. Deste modo, é esperado encontrar as técnicas que, em conjunto, possibilitem a obtenção de resultados precisos, com tempos de processamento e de resposta cada vez menores. Também é importante que sejam definidos comandos simples para configuração e utilização do sistema de contagem, de modo que facilite o manuseio da ferramenta.

## 1.3 Organização do Documento

O trabalho está organizado em 6 capítulos. No Capítulo 2 serão abordados aspectos relativos à morfologia das plaquetas e suas anomalias, assim como as técnicas existentes de contagem de plaquetas. Em seguida, no Capítulo 3, serão apresentados alguns conceitos de processamento digital de imagens relevantes para o desenvolvimento deste trabalho. O Capítulo 4 descreve o mecanismo de classificação e detecção de objetos que foi utilizado. A solução proposta, os arranjos experimentais e os resultados das simulações são apresentados no Capítulo 5. Por fim, no Capítulo 6, serão discutidas as principais contribuições e conclusões desse trabalho, bem como propostas para trabalhos futuros.

## 2 *Tópicos sobre Plaquetas*

*“Aqueles que desejam alcançar o sucesso devem primeiro fazer as perguntas certas.”*

– Aristóteles.

Neste capítulo serão apresentados os aspectos morfológicos das plaquetas, bem como anomalias relacionadas. Também serão mostradas algumas técnicas de contagem manual e automática de plaquetas. O objetivo deste capítulo é promover uma visão geral sobre o cenário do problema abordado, de modo que sejam identificadas as necessidades e mapeados os parâmetros para a construção da solução.

### 2.1 Morfologia das Plaquetas

As plaquetas são células sanguíneas formadas a partir da fragmentação do citoplasma dos megacariócitos da medula óssea. As plaquetas possuem uma estrutura complexa e um elevado conteúdo energético. Apresentam formato que se assemelha a um disco, com diâmetros que variam de 1  $\mu\text{m}$  até 3  $\mu\text{m}$  [1].

As plaquetas possuem pequenos grânulos dispersos pelo citoplasma ou concentrados em seu centro. A parte central do citoplasma, na qual ficam alojados os grânulos é chamada de granulômero. Já a parte mais externa e agranular é conhecida como hialômero [3]. A Figura 1 mostra imagens de plaquetas normais encontradas no sangue humano.

### 2.2 Anomalias Morfológicas e Quantitativas

Em condições normais as plaquetas encontram-se dispostas separadamente umas das outras e não apresentam alterações de tamanho. Entretanto, quando há ocorrência de determinadas complicações como viroses, infecções, hemorragias, as plaquetas podem tender a formar agregados plaquetários ou aumentar de tamanho. Também é comum que hajam alterações na quantidade de plaquetas presentes no sangue.

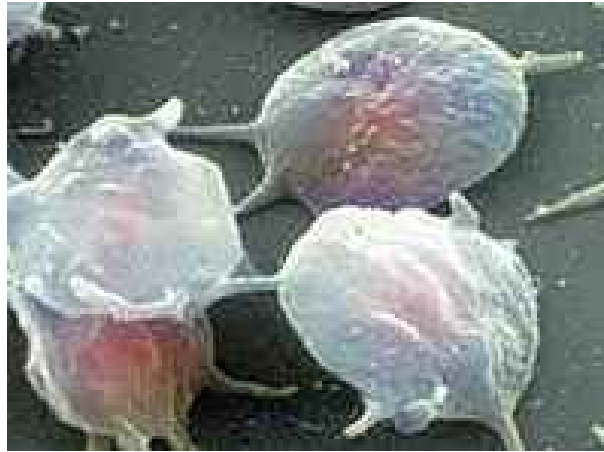
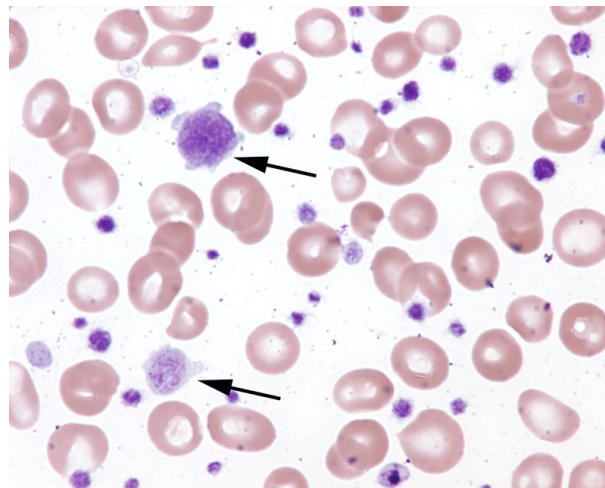


Figura 1: *Plaquetas Normais.*

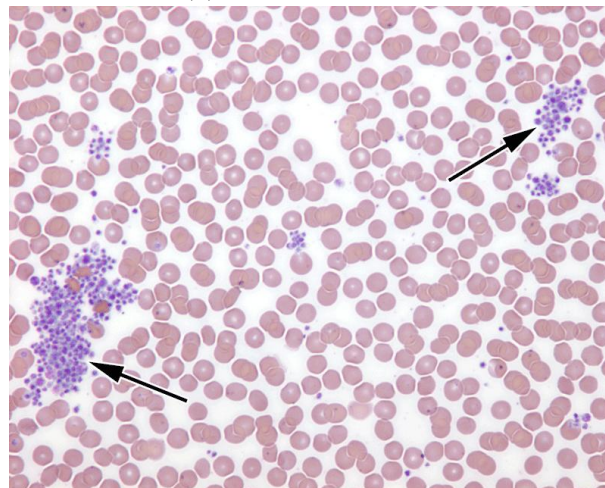
**Anormalidades quanto ao tamanho das plaquetas** – O tamanho de uma plaqueta pode ser analisado comparando-se o diâmetro dela com o diâmetro das hemácias. Plaquetas grandes, aquelas que apresentam diâmetro maior do que  $4\mu$  são denominadas macrotrombócitos. Há também alguns casos em que as plaquetas atingem tamanhos próximos ao das hemácias, chegando a cerca de  $7,5\mu$  de diâmetro e são chamadas de plaquetas gigantes e podem ser visualizadas na Figura 2a. Quando a renovação das plaquetas está acelerada, é comum também que elas apresentem aumento de tamanho. Em casos em que o paciente apresenta diminuição na taxa de plaquetas, por exemplo, e não forem observadas plaquetas aumentadas, é um provável cenário de anomalia na produção de plaquetas, já que a taxa diminuiu, e não há sinais de que as plaquetas estão se renovando rapidamente para retornar à taxa normal. O aumento no tamanho das plaquetas também pode ocorrer em decorrência de doenças como a Leucemia [3].

**Anormalidades quanto à distribuição das plaquetas** – Outro tipo de anomalia que pode acometer as plaquetas é a formação de agregados plaquetários. Tais agregados podem ser formados, por exemplo em punções venosas complicadas. Quando isso ocorre, as plaquetas são estimuladas e podem acabar liberando os grânulos e então se associando a eles, formando assim agregados de plaquetas agranuladas [3]. Agregados plaquetários podem ser observados na Figura 2b.

**Anomalias quantitativas** – Podem também ser observadas anormalidades quanto à contagem das plaquetas. Quando é detectada uma redução no número de plaquetas diz-se que há um caso de plaquetopenia, ou trombocitopenia. Ao passo que os casos onde há elevação na quantidade de plaquetas diz-se que ocorre plaquetose, ou trombocitose [2]. As principais doenças causadoras de anomalias quantitativas podem ser observadas na Tabela 1.



(a) Plaquetas gigantes.



(b) Agregados plaquetários.

Figura 2: Ilustração de anomalias plaquetárias.

## 2.3 Técnicas de Contagem de Plaquetas

A contagem pode ser realizada por instrumentos automatizados e semi-automatizados [3]. Entretanto, as técnicas manuais são usadas em paralelo com técnicas automatizadas. Servem tanto como método de referência, quanto para a averiguação de resultados anômalos encontrados pelo método de contagem automática [1].

Para realizar contagem através do microscópio, é necessário diluir o sangue em uma espécie de câmara, conhecida como hemocitômetro ou utilizar o plasma rico em plaquetas [3]. Porém, a contagem através do hemocitômetro é bastante trabalhosa.

Outra maneira de realizar a contagem é através dos contadores de impedância semi-automatizados. Eles utilizam o plasma rico em plaquetas, e tal como a contagem através do microscópio demanda muito trabalho [3].

Tabela 1: Principais causas de Plaquetopenias e Plaquetoses.

<b>Plaquetopenias</b>	<b>Plaquetoses</b>
Leucemias Agudas	Policitemia Vera
Anemia Aplástica	Leucemia mielóide crônica
Esplenomegalia	Período pós-hemorragico
Grandes hemorragias	Após esplenectomia
Viroses em geral	Doenças infecciosas crônicas
AIDS	Trombocitemia essencial
Púrpuras imunológicas	
Quimioterapia e radioterapia antitumoral	

Fonte: Renato Failace, *Hemograma - Manual de Interpretação*, 2003 [2].

Entretanto, apenas a contagem através de aparelhos totalmente automatizados permite suprir a demanda exigida pela prática médica. Nesses instrumentos a contagem é realizada por técnicas de impedância elétrica ou dispersão de luz [3]. Porém, em casos onde há presença de plaquetas gigantes, o contador pode confundir tais plaquetas com outros elementos do sangue e acabar apontando um falso caso de plaquetopenia. Assim, em decorrência do resultado anômalo da contagem, torna-se necessário refazê-la através do hemocitômetro [3]. Outro ponto negativo desses equipamentos é o alto custo despendido. Já que, além de serem instrumentos caros, ainda necessitam de treinamento específico para manuseio, bem como de uma equipe especializada para a manutenção.

Neste trabalho será proposto um novo método de contagem das plaquetas, capaz de realizar a contagem automática através de imagens sanguíneas, utilizando técnicas de processamento digital de imagens combinadas com técnicas de inteligência computacional.

## 3 *Processamento de Imagens*

*“Uma imagem vale mais que dez mil palavras.”*

– Anônimo.

Neste capítulo serão apresentados os conceitos de processamento de imagens relevantes para a implementação da solução de detecção de plaquetas. Maiores detalhes sobre os tópicos aqui abordados podem ser encontrados em Gonzalez e Woods [4] e Gomes e Velho [5].

### 3.1 Introdução

Os avanços tecnológicos têm proporcionado ao homem ferramentas para a difusão de conhecimento e aquisição de informações. O processamento destes dados (informações), quando representados através de imagens, é denominado Processamento de Imagens. Processar uma imagem consiste em realizar transformações sucessivas através de operações matemáticas (algoritmos), com o intuito de extrair as informações nela presentes. Porém, as particularidades de cada conjunto de imagem e dos sistemas aos quais eles estão associadas tornam o processamento uma tarefa difícil. É necessário ter conhecimento do sistema envolvido, bem como das características da imagem, a fim de utilizar as melhores técnicas que auxiliarão a extrair a informação desejada.

### 3.2 Sistema de Processamento de Imagens

Para a construção de um sistema de processamento de imagens é necessária uma sequência de etapas que proporcionarão a obtenção dos resultados. Cada etapa possui uma função específica e funcionamento encadeado no qual a saída de uma etapa servirá como entrada para a etapa seguinte. A partir da imagem inicialmente utilizada, as etapas que serão executadas fornecerão os dados necessários para a extração da informação desejada. A sequência de



etapas executadas por um sistema de processamento de imagens é apresentada na Figura 3.

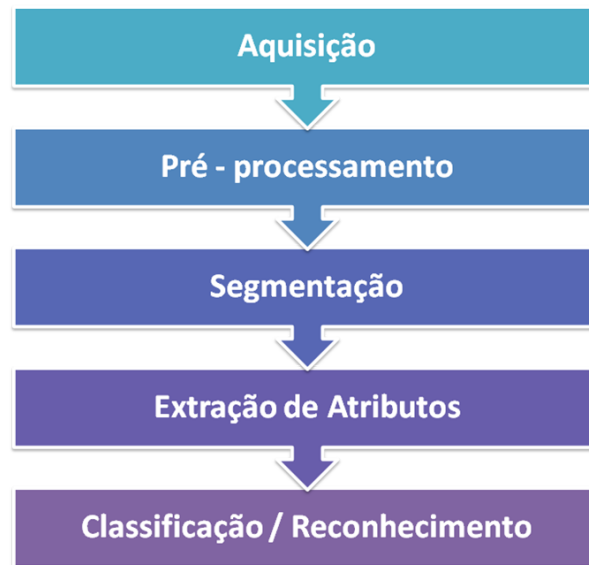


Figura 3: *Modelo de um Sistema de Processamento de Imagens.*

- **Aquisição** – Etapa responsável pela digitalização da informação. Um hardware é utilizado para capturar a imagem e realizar a conversão de sinal analógico para digital. Como exemplos de hardware têm-se: câmeras fotográficas digitais, scanners e filmadoras.
- **Pré-processamento** – Etapa na qual a imagem é tratada para a eliminação de possíveis ruídos e realce de contrastes. Deste modo a imagem será melhorada, aumentando a chance de sucesso nas etapas seguintes.
- **Segmentação** – Durante a segmentação a imagem é dividida em objetos constituintes da imagem que serão analisados através de características do objeto de interesse.
- **Extração de Atributos** – Na extração de atributos é determinado quais atributos poderão ser utilizados na caracterização do objeto de interesse.
- **Classificação / Reconhecimento** – Etapa responsável pela classificação e reconhecimento do objeto de interesse, dados os atributos extraídos na etapa anterior.

### 3.3 Imagem Digital

No mundo real a representação dos sinais pode ser descrita como uma função contínua. Entretanto, o computador só é capaz de armazenar dados em forma de bits, que assumem

"0" ou "1". Como forma de simplificação, 8 bits formam uma palavra, o byte. E cada byte pode assumir  $2^8 = 256$  valores diferentes, variando de 0 a 255.

Deste modo, o computador só manipula números inteiros, ou uma aproximação de um número real, o ponto flutuante. Consequentemente, funções contínuas não podem ser representadas pelo computador. O que se pode fazer é simulá-las através de uma aproximação.

Para que uma função seja representada pelo computador é necessário discretizá-la. O processo de discretização do eixo  $x$  (o domínio) é chamado de Amostragem, como exemplificado na Figura 4. O processo de discretização do eixo  $f(x)$  (o contradomínio) é chamado de Quantização, como exemplificado na Figura 5.

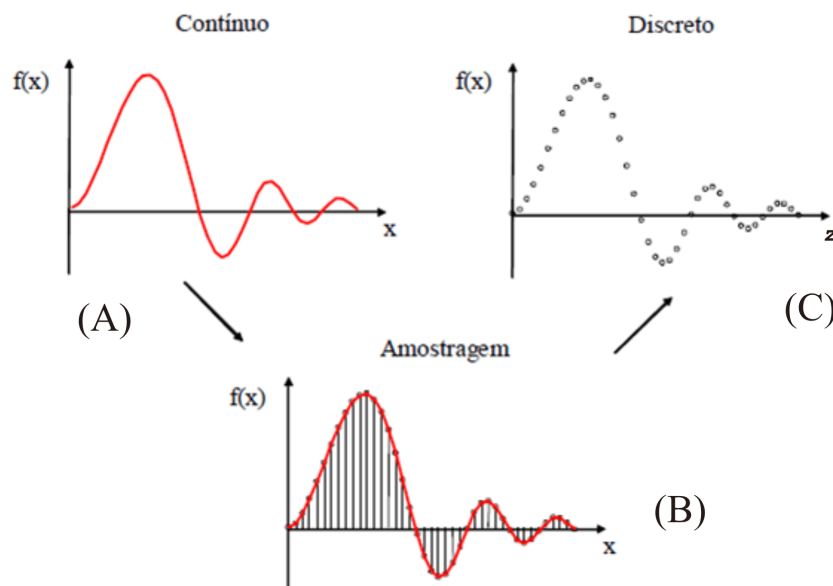
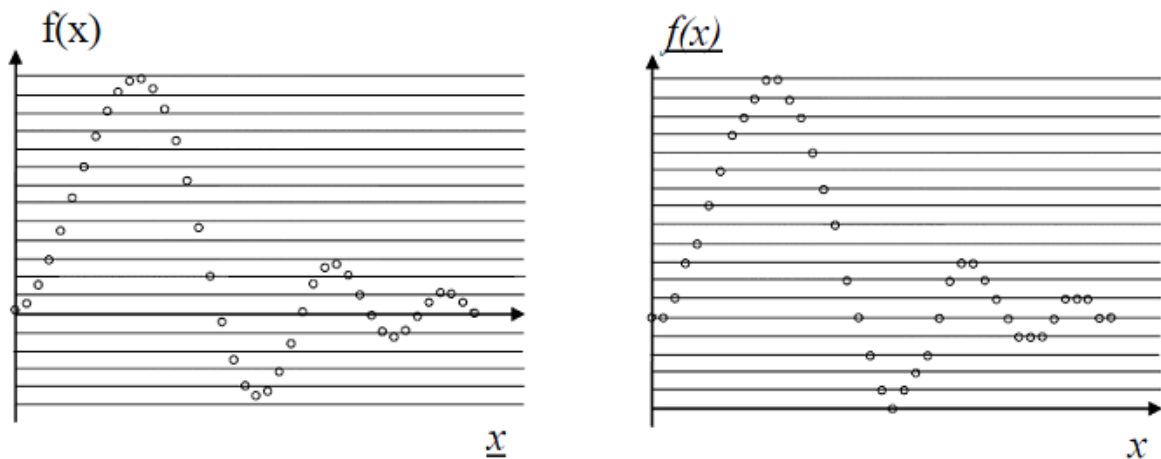


Figura 4: *Processo de Amostragem.*

Uma imagem é o resultado de estímulos luminosos produzidos por um suporte bidimensional. Para representar e manipular imagens no computador, é necessário que sejam definidos modelos matemáticos adequados.

Dado que a imagem, neste estudo, é um sinal bidimensional, pode ser denotada pela função de intensidade luminosa bidimensional  $f(x, y)$ , em que o valor ou amplitude de  $f$  nas coordenadas espaciais  $(x, y)$  dá a intensidade (brilho) da imagem naquele ponto. A natureza básica de  $f(x, y)$  pode ser caracterizada por dois componentes: (1) quantidade de luz que incide na cena em observação (2) a quantidade de luz que é refletida pelos objetos. Tais componentes são denominados iluminação e reflectância, respectivamente. A iluminação é representada por  $i(x, y)$ , a reflectância por  $r(x, y)$ , e  $f(x, y)$  é o produto destas duas componentes conforme



(a) Amostragem do sinal.

(b) Sinal quantizado.

Figura 5: *Processo de quantização.*

Equação 3.1.

$$f(x, y) = i(x, y)r(x, y). \quad (3.1)$$

Para que a função de intensidade  $f(x, y)$  possa ser processada pelo computador, é necessário que seja digitalizada tanto espacialmente (amostragem da imagem), quanto em amplitude (quantização em níveis de cinza).

Uma imagem digital pode ser representada através de uma matriz de números reais  $M \times N$  resultante da amostragem e quantização da função  $f(x, y)$ . Cada elemento da matriz é denominado elemento de imagem. Os elementos da imagem consistem, essencialmente, das coordenadas e da informação de cor de cada *pixel* (abreviação de elemento de figura – do inglês, *picture element*). O número de componentes do *pixel* define o espaço de cores da imagem. No caso de imagens monocromáticas, por exemplo, cada *pixel* possui apenas uma única componente. Pode-se ainda definir o *gamute* de uma imagem digital, que é o conjunto de cores do espaço de cor quantizado da imagem. As imagens monocromáticas cujo *gamute* possui apenas duas cores são chamadas imagens binárias, ou de dois níveis. Já as imagens monocromáticas cujo *gamute* possui mais de duas cores são denominadas imagens em **escala de cinza** (do inglês *grayscale*).

Dado que o espaço de cor de uma imagem tenha uma dimensão  $k$ , cada componente de cor pode ser considerado em separado. Deste modo a imagem pode ser reduzida a  $k$  imagens com escala de cinza. Cada uma destas imagens é, portanto, uma componente da imagem original, e que poder ser manipulada separadamente.

## 3.4 Sistemas de amostragem de cor

No início do século XIX foi proposto pelo físico T. Young um modelo tricromático para o espaço de cor do olho humano. Segundo este modelo, o olho possui apenas três tipos de moléculas fotossensíveis às faixas baixa, média e de alta frequência do espectro visível de cores. Deste modo, conforme a teoria de Young, o espaço de cor do olho humano possui dimensão  $k = 3$ , onde cada dimensão corresponde à uma das faixas, Vermelha (*Red*), Verde (*Green*) e Azul (*Blue*), do espectro de cores. Mais à frente, após a contribuição do físico Helmholtz afim de comprovar a teoria de Young, esta ficou conhecida como teoria de Young-Helmholtz.

Tendo em vista a teoria de Young-Helmholtz, é comum que seja utilizada uma representação tridimensional do espaço de cor cuja base de cores primárias está nas faixas vermelho, verde e azul, e que é conhecido como sistema RGB.

## 3.5 Pré Processamento

Como citado anteriormente na seção 3.2, um dos passos de um sistema de processamento de imagens é o pré-processamento. Ao se fazer processamento de imagens é de suma importância definir as operações de pré-processamento que serão executadas, uma vez que é nesta etapa que são eliminados possíveis ruídos e ressaltadas características da imagem. O sucesso dos processos seguintes dependerá da qualidade das imagens de saída do pré-processamento. Nesta seção serão explicadas as técnicas que foram utilizadas durante o desenvolvimento deste trabalho.

### 3.5.1 Grayscale (escala de cinza)

Uma técnica de pré-processamento de imagens bastante comum é a transformação das imagens coloridas para o domínio das imagens monocromáticas em escala de cinza. Este tipo de imagem possui para cada pixel apenas a informação de intensidade, onde podem ter mais intensidade (mais próxima da cor branca) ou menos intensidade (mais próxima da cor preta). Representando-se assim as cores é minimizada a quantidade de cálculos por pixel, e também é possível que alguns ruídos sejam eliminados. Na Figura 6 é apresentado a aplicação de um filtro de tons de cinza em um imagem colorida.



Figura 6: Exemplo da aplicação do filtro de tons de cinza em uma imagem onde, o lado esquerdo da figura corresponde à imagem original e o lado direito é o resultado da aplicação do filtro.

### 3.5.2 Canais da Imagem

Como citado anteriormente na Seção 3.3, é possível manipular separadamente cada componente de uma imagem cujo espaço de cores tem dimensão maior que 1. No caso das imagens tricromáticas representadas no sistema de cores RGB, apresentado na seção 3.4, cada espectro de cor (vermelho, verde e azul) é um componente da imagem, e pode ser denominado também de canal da imagem. Outra técnica muito comum é separar estes canais da imagem e processá-los individualmente, podendo-se, por exemplo, utilizar dentre os 3 canais, apenas o que realce melhor os detalhes das imagens das quais se que extrair informações.

### 3.5.3 Equalização do Histograma

Um histograma é a representação gráfica da distribuição da intensidade de cada *pixel* em uma imagem. Equalizar o histograma de uma imagem nada mais é do que mapear os valores das intensidades dos pixels entre uma região com pouco contraste e uma com muito contraste, redistribuindo uniformemente a intensidade dos *pixels* ao longo da imagem. Deste modo, é obtida uma imagem equalizada. A equalização do histograma pode ainda ser feita localmente. Neste caso, o mapeamento e redistribuição da intensidade dos *pixels* são feitos levando-se em consideração apenas os vizinhos de cada *pixel* e não a imagem como um todo.

**Vizinhança de um pixel** – Dado um pixel  $p$  de coordenadas  $(x, y)$ , este possui quatro vizinhos horizontais e verticais, cujas coordenadas são:  $(x + 1, y)$ ,  $(x - 1, y)$ ,  $(x, y + 1)$

e  $(x, y - 1)$ . Este conjunto de *pixels* é chamado de vizinhança-de-4 de  $p$ , é representado por  $N_4(p)$  e cada um dos *pixels* está a uma unidade de distância de  $(x, y)$ . Um *pixel* possui ainda vizinhos diagonais, que são denotados por  $N_D(p)$  e ocupam as coordenadas:  $(x + 1, y + 1)$ ,  $(x + 1, y - 1)$ ,  $(x - 1, y + 1)$  e  $(x - 1, y - 1)$ . A união entre os conjuntos  $N_4(p)$  e  $N_D(p)$  é denominada vizinhança-de-8 de  $p$ .

## 3.6 Segmentação de imagens

Após ter sido realizado o pré-processamento das imagens, é necessário que elas sejam subdivididas em suas partes ou objetos constituintes. A segmentação é realizada para que a região da imagem a ser analisada seja restringida, de modo que passem para a etapa de reconhecimento apenas os objetos de interesse da aplicação. Existem diversas técnicas de segmentação de imagens, dentre as quais destaca-se a segmentação por detecção de borda.

### 3.6.1 Segmentação por detecção de borda

A definição de borda de uma imagem pode ser dada como o limite entre duas regiões com variação súbita entre os níveis de intensidade dos pixels. A função dos detectores de borda é encontrar essas variações e quando os pixels estiverem próximos conectá-los formando um contorno que limitará uma região ou objeto. Dentre os filtros de detecção de borda estão Canny, Laplace e Sobel.

- **Canny** – O algoritmo de Canny foi desenvolvido por John Canny em 1986. Ele é um operador gaussiano de primeira derivada que suaviza os ruídos e localiza as bordas com base em três critérios básicos: taxa de erro, localização e resposta [6]. A Figura 7a ilustra a segmentação por detecção de bordas utilizando o filtro de Canny. Na figura 7b é apresentado um exemplo de máscara utilizada pelo filtro de Canny.
- **Laplace** – O Laplaciano  $L(x, y)$  de uma imagem é uma derivada de segunda ordem da função que descreve a intensidade de seus pixels  $f(x, y)$ , definida pela Equação 3.2. Considerando que a imagem é armazenada como uma coleção de pixels discretos, é necessário produzir uma forma discreta da derivada segunda que é representada por um núcleo de convolução (máscara ou kernel). A exigência para a definição do laplaciano de uma imagem digital é que o pixel central seja positivo e os pixels externos sejam negativos como o exemplo da Figura 7d. O Laplaciano responde bem à transições de intensidade, entretanto é muito sensível a ruído e produz bordas duplas. A Figura 7c

ilustra a segmentação por detecção de bordas utilizando o filtro Laplaciano.

$$L(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (3.2)$$

- **Sobel** – Sobel consiste na aplicação de um filtro de convolução na imagem que combina o gradiente com atenuação do ruído, visto que possui um efeito secundário de suavização. As principais vantagens são sua fácil implementação e execução rápida. A Figura 7e ilustra a segmentação por detecção de bordas utilizando o filtro de Sobel. Na figura 7f é apresentado um exemplo de máscara utilizada pelo filtro de Sobel.

### 3.7 Classificação e Reconhecimento de Padrões

O objetivo da etapa de classificação de um sistema de processamento de imagens é distinguir os objetos que foram encontrados durante a segmentação de forma que os parâmetros sejam agrupados de acordo com a semelhança para cada conjunto de pixels delimitado. Tal classificação deve ser ponderada de acordo com uma medida de erro de classificação. A classificação é uma etapa bastante complexa e existem diversos níveis de automatização deste processo. Dentre estes processos, os mais sofisticados utilizam técnicas de inteligência computacional. Das técnicas mais difundidas destacam-se:

- **K-médias** – método de agrupamento não-hierárquico por repartição, que produz k diferentes conjuntos com a maior distinção possível entre eles [7].
- **Árvores de decisão** – método preditivo onde cada ramo da árvore é uma questão de classificação e cada folha é uma partição do conjunto de dados com sua classificação [8].
- **Redes neurais artificiais** – técnica inspirada na estrutura e funcionamento de um neurônio que adquire conhecimento a partir da experiência. É composta por várias unidades de processamento, onde cada unidade corresponde a um neurônio. Os neurônios são conectados por canais de comunicação que estão associados a pesos e realizam operações sobre seus dados locais. O comportamento inteligente de uma rede neural artificial é dado pela interação entre as unidades de processamento da rede [9].
- **Classificadores bayesianos** – método estatístico que classifica um objeto em uma determinada classe, baseando-se na probabilidade deste objeto pertencer a esta classe [10].



- **Máquinas de vetor de suporte (SVMs, do inglês - Support Vector Machines)**
  - nesta técnica a separação entre classes do conjunto de treinamento é feita a partir de um hiperplano que é determinado por alguns pontos, os quais são denominados vetores de suporte [11].

O reconhecimento de padrões exige um conhecimento prévio sobre o objeto a ser reconhecido. Deste modo, para obter sucesso durante o reconhecimento é necessário que se tenha uma base de conhecimento acerca dos objetos a serem reconhecidos. Tal base pode tanto ser implementada diretamente no código através de regras, quanto pode ser aprendida a partir de um conjunto de amostras dos objetos que se deseja reconhecer através do classificador.

Um objeto pode ser definido por diversos atributos, tais como: textura, forma, cor, dimensões, etc. E o reconhecimento de cada um destes atributos tende a facilitar o reconhecimento do objeto como um todo.

As técnicas de reconhecimento de padrões subdividem-se em dois grandes grupos: Estruturais, nas quais os padrões são descritos simbolicamente e a estrutura é o modo como estes padrões se relacionam; O segundo grupo baseia-se em técnicas que utilizam teoria de decisão. No grupo em questão, os padrões são descritos conforme propriedades quantitativas e deve-se decidir se o objeto possui ou não estas propriedades.

Um sistema de reconhecimento de padrões divide-se em dois blocos: Operacional e de treinamento. O bloco Operacional é responsável pela aquisição dos dados, tratamento, extração das características e classificação dos dados. Já o bloco de treinamento é responsável por treinar o sistema para que a etapa de classificação possa reconhecer e decidir se o dado apresentado na entrada possui ou não as características desejadas.



(a) Aplicação do filtro de Canny.

2	4	5	4	2
4	9	12	9	4
5	12	15	12	5
4	9	12	9	4
2	4	5	4	2

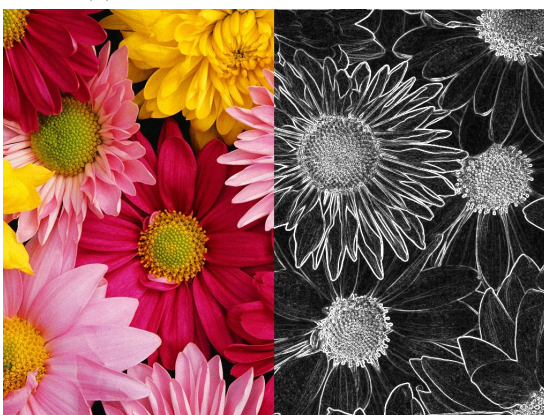
(b) Máscara para o filtro de Canny.



(c) Aplicação do filtro de Laplace.

0	-1	0
-1	4	-1
0	-1	0

(d) Máscara para o filtro de Laplace.



(e) Aplicação do filtro de Sobel.

1	2	1
0	0	0
-1	-2	-1

1	0	-1
2	0	-2
-1	0	-1

(f) Máscaras para o filtro de Sobel.

Figura 7: Aplicação dos filtros de detecção de borda e suas respectivas máscaras.

## 4 *Classificação e Detecção*

*“Um dos aspectos mais interessantes do mundo é que se pode considerar que ele é feito de padrões. Um padrão é essencialmente um arranjo. Ele é caracterizado pela ordem dos elementos que o compõe, no lugar da natureza intrínseca desses elementos.”*

– Norbert Wiener.

Conforme visto na Seção 3.7 do capítulo anterior, existem várias técnicas que podem ser utilizadas para realizar tarefas de classificação e reconhecimento de padrões. Para o desenvolvimento do sistema proposto neste trabalho foram escolhidas as árvores de decisão. Neste caso, cada ramo da árvore é uma questão de classificação e cada folha é uma partição do conjunto de dados com sua classificação. Esta técnica foi eleita por apresentar um bom desempenho para classificação de duas classes e por já existirem propostas utilizando Árvores de Decisão para este fim. Tais quais a proposição inicial feita por Viola e Jones [12] para solucionar o problema de detecção de faces, e uma mais recente foi aplicada por Firmo [13], para auxílio no diagnóstico da Esquistossomose Mansônica no litoral de Pernambuco.

Neste capítulo serão apresentados os conceitos propostos por Viola e Jones [12], que serviram como fundamentos para o desenvolvimento do sistema aqui proposto. O objetivo deste capítulo é fazer uma explanação das principais ferramentas utilizadas na construção da solução de detecção.

### 4.1 Método de detecção de faces proposto por Viola e Jones

Viola e Jones [12] propuseram uma nova metodologia para detecção de faces e objetos que provê uma implementação rápida e robusta. Três contribuições importantes surgiram deste trabalho. São elas:

1. Introdução de uma nova representação de imagens denominada imagem integral;

2. Construção de um classificador simples e eficiente a partir da seleção de características de Haar [14] utilizando o algoritmo de aprendizado do tipo *boosting* [15], denominado *AdaBoost*;
3. Combinação de classificadores fracos para formar uma estrutura de cascata que aumenta a velocidade da classificação dando mais foco a áreas mais promissoras da imagem.

## 4.2 Imagem Integral

Dado uma imagem digital, a imagem integral para um ponto  $(x, y)$  desta imagem é a soma da intensidade de todos os pixels localizados a esquerda e acima de  $(x, y)$ , incluindo-o, conforme mostrado na Equação 4.1. O cálculo a imagem integral pode ser dividido em duas etapas: (1) somatório dos pontos ao longo das linhas (Equação 4.2) e (2) somatório ao longo das colunas (Equação 4.3).

A utilização desta forma de representação permite que qualquer retângulo da imagem seja calculado a partir de uma única interação do algoritmo, facilitando assim o processo de extração de características da imagem. Dado uma imagem  $i(x', y')$ , a integral de uma sub-janela  $D$  de  $i$  pode ser calculada a partir de 4 *arrays*. Tendo como base o exemplo apresentado na Figura 8, o valor da imagem integral para o ponto 1 é a soma dos pixels do retângulo  $A$ . O valor para o ponto 2 é o somatório  $A + B$ , para o ponto 3 é  $A + C$  e por fim, para 4 é  $A + B + C + D$ . Assim, a soma para a sub-janela  $D$  pode ser calculada como sendo  $4 + 1 - (2 + 3)$ .

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (4.1)$$

onde  $ii(x, y)$  é a integral da imagem e  $i(x', y')$  é a imagem original.

$$s(x, y) = s(x, y - 1) + i(x, y), \quad (4.2)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y), \quad (4.3)$$

onde  $s(x, y)$  é o somatório acumulado de uma linha,  $s(x, -1) = 0$ , e  $ii(-1, y) = 0$ .

## 4.3 Características de Haar

O sistema de detecção de faces proposto por Viola e Jones utiliza características simples para classificar as imagens. As características são baseadas nas funções de Haar que tem sido usadas por Papageorgiou [16].

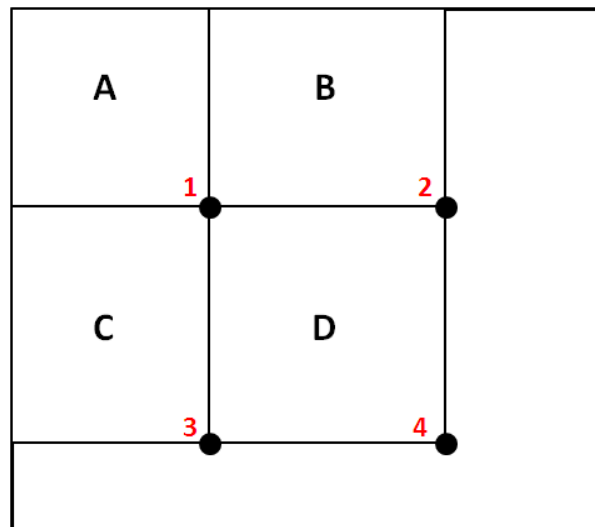


Figura 8: Exemplo da imagem de entrada  $i$  e suas sub-janelas A, B, C e D.

As características correspondem basicamente a retângulos que possuem duas regiões, sendo uma clara e outra escura. O cálculo do valor correspondente à característica é feito subtraindo o somatório da intensidade dos pixels da região escura do total acumulado pela intensidade dos pixels da região clara. Este cálculo pode ser feito rapidamente utilizando a representação de imagem integral apresentada na seção 4.2.

Na Figura 9 são apresentadas as características básicas de Haar que foram utilizadas por Viola e Jones.



Figura 9: Características básicas de Haar.

## 4.4 Algoritmo de Aprendizado

Na proposição de Viola e Jones foi utilizado o algoritmo *AdaBoost* (do inglês *Adaptive Boosting*, proposto por Freund e outros [15], para selecionar as características e treinar o classificador. Nesta seção será explicado o funcionamento do *AdaBoost*, bem como da unidade básica que compõe o classificador, denominada classificador fraco.

### 4.4.1 Classificador Fraco

Um classificador fraco é uma estrutura simples que contém um vetor de características  $f$ , um limiar e uma paridade. Durante o treinamento do classificador fraco deve ser encontrado um limiar que melhor separe o valor de uma característica de exemplos definidos como positivos dos negativos. A paridade é utilizada para indicar se os objetos positivos devem ser classificados abaixo ou acima do limiar [17].

A equação 4.4 define formalmente um classificador fraco.

$$h_j(x) = \begin{cases} 1, & \text{se } p_j f_j(x) < p_j \theta_j; \\ 0, & \text{caso contrário.} \end{cases}, \quad (4.4)$$

onde  $p_j$  é a paridade,  $\theta_j$  é o limiar e  $f_j(x)$  é o valor da característica. Caso os exemplos positivos sejam classificados abaixo do limiar, a polaridade recebe "1", caso contrário recebe "-1". Na proposta de Viola e Jones a característica de um classificador fraco é baseada nas características de Haar apresentadas na Seção 4.3.

Existem algumas propostas para o cálculo do limiar, dentre elas destaca-se a de Carbonetto [17], apresentada na Equação 4.5, e que é baseada na distribuição de valores associados às características dos exemplos positivos e negativos.

$$\theta_j(x) = \frac{1}{2} \left( \frac{1}{|C_0|} \sum_{x \in C_0} f_j(x) + \frac{1}{|C_1|} \sum_{x \in C_1} f_j(x) \right), \quad (4.5)$$

onde  $C_0$  é o conjunto de exemplos negativos,  $C_1$  é o conjunto de exemplos positivos e  $f_j(x)$  é o valor da característica para um dado exemplo.

### 4.4.2 AdaBoost

O *AdaBoost* é um algoritmo de aprendizado supervisionado do tipo *boost*. O *AdaBoost* combina um conjunto de funções simples de classificação, denominadas classificadores fracos para formar um classificador forte.

Um classificador forte é composto de um conjunto de classificadores fracos, associados a pesos que classificam de forma precisa dois conjuntos de dados pré-rotulados, onde as características com pesos maiores são mais significativas para a classificação de exemplos definidos como parte de um certo conjunto. Na Equação 4.6 é formalizada a criação de um

classificador forte através de um algoritmo de *Boosting*.

$$H(x) = \alpha_1 h_1 + \alpha_2 h_2 + \dots + \alpha_n h_n(x), \quad (4.6)$$

onde  $H(x)$  é um classificador forte,  $\alpha_i$  é o peso associado ao classificador  $h_i$ .

Dado uma base de dados de entrada, a função do *AdaBoost* é encontrar o conjunto de características que comporão o classificador forte provendo uma melhor classificação do conjunto de entrada.

No caso do trabalho de Viola e Jones, o conjunto de entrada era composto por imagens pré-rotuladas, das quais algumas eram positivas (faces) e outras negativas (não faces). O *AdaBoost* foi utilizado para encontrar as características que fornecessem uma melhor classificação dentre os elementos face e não-face. Cada classificador fraco é derivado de um conjunto de características básicas de Haar.

O pseudocódigo do *AdaBoost* é descrito no Algoritmo 1.

---

**Algoritmo 1:** Pseudocódigo do *AdaBoost* .

---

- 1 Obtém  $N$  exemplos de imagens  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , onde  $x$  corresponde a matriz de pixels de uma imagem, e  $y = 0$ , ou 1 para exemplos negativos e positivos, respectivamente.
  - 2 Inicializa os pesos  $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$  para  $y_i = 0$  e 1 respectivamente, onde  $m$  é o total de exemplos negativos e  $l$  o total de positivos.
  - 3 **para**  $t=1, \dots, T$  **faça**
  - 4     Normalize os pesos:  $\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^N \omega_{t,j}}$ , onde  $\omega_t$  é uma distribuição de probabilidade.
  - 5     **para cada característica**  $j$  **faça**
  - 6         Treine um classificador  $h_j$  restrito ao uso de uma única característica.
  - 7         Avalie o erro de acordo com:  $\omega_j, \epsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|$ .
  - 8         Escolha o classificador  $h_t$  com o menor erro  $\epsilon_t$ .
  - 9         Atualize os pesos:  $\omega_{t+1,i} = \omega_{t,i} \beta_t^{(1-e_i)}$ , onde  $e_i = 0$ , se o exemplo  $x_i$  for classificado corretamente,  $e_i = 1$  caso contrário, e  $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$ .
  - 10 O classificador forte é definido por:  $H(x) = \begin{cases} 1, & \text{se } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{caso contrário.} \end{cases}$ ,  
onde  $\alpha_t = \log \frac{1}{\beta_t}$ .
- 

## 4.5 Cascata de Classificadores

Um dos problemas da tarefa de classificação é a quantidade de tempo demandada para realizar o treinamento. Visando solucionar este problema, Viola e Jones [12] propuseram um



novo método que utiliza classificadores em uma estrutura de árvores de decisão ou cascata de classificadores.

A estrutura foi descrita por Viola e Jones como sendo um tipo de árvore de decisão degenerativa [18], a qual foi chamada de cascata. A ideia é que os primeiros estágios da cascata sejam mais genéricos, e ao longo da cascata os classificadores fiquem mais específicos. Deste modo, acelera o processo de avaliação dos exemplos, uma vez que grande parte dos exemplos negativos serão eliminados nos primeiros estágios da cascata.

A estrutura de um sistema que utiliza cascata de classificadores pode ser observado na Figura 10.

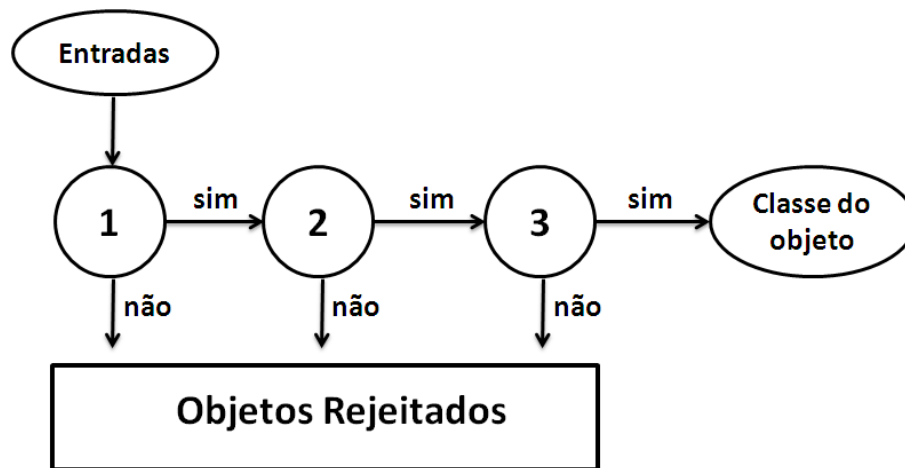


Figura 10: Representação de uma cascata de detecção. Cada estágio da cascata contém um classificador que irá avaliar o conjunto de entrada.

#### 4.5.1 Treinamento da Cascata de Classificadores

O treinamento da cascata de classificadores é realizado visando atingir uma série de metas tanto de detecção quanto de desempenho. O número de estágios da cascata e o número de classificadores de cada estágio deve ser suficiente para que seja atingida uma taxa de detecção e desempenho similares as do *AdaBoost*, ao passo que minimiza o processamento exigido.

Dada uma cascata de classificadores treinada, a taxa de falso positivo é encontrada conforme Equação 4.7.

$$F = \prod_{i=1}^K f_i, \quad (4.7)$$

onde  $F$  é a taxa de falso positivo da cascata de classificadores,  $K$  é o número de classificadores, e  $f_i$  é a taxa de falso positivo de cada classificador. A taxa de detecção é encontrada através



da Equação 4.8.

$$D = \prod_{i=1}^K d_i, \quad (4.8)$$

onde  $F$  é a taxa de detecção da cascata de classificadores,  $K$  é o número de classificadores, e  $d_i$  é a taxa de detecção de cada classificador.

O treinamento de uma cascata de classificadores é realizado conforme descrição do Algoritmo 2.

---

**Algoritmo 2:** Pseudocódigo do treinamento de uma cascata de classificadores .

---

- 1 Seleciona os valores para:  $f$ : a máxima taxa de aceitação de falso positivo por estágio da cascata.  $d$ : a taxa mínima de detecção por estágio aceita.
  - 2  $m$  = conjunto de exemplos negativos.
  - 3  $l$  = conjunto de exemplos positivos.
  - 4  $F_0 = 1, 0$ ;  $D_0 = 1, 0$ ;  $i = 0$ .
  - 5 **enquanto**  $F_i > F_{alvo}$  **faça**
  - 6      $i = i + 1$ .
  - 7      $n_i = 0$ ;  $F_i = F_{i-1}$ .
  - 8     **enquanto**  $F_i > f \cdot F_{i-1}$  **faça**
  - 9         Usa  $l$  e  $m$  para treinar um classificador com  $n_i$  características usando o *AdaBoost*.
  - 10         Avalia a cascata de classificadores no conjunto de validação para determinar  $F_i$  e  $D_i$ .
  - 11         Diminui o limiar para o  $i$ -ésimo classificador até que a cascata atual tenha uma taxa de detecção de pelo menos  $d(D_{i-1})$ , o que também afeta  $F_i$ .
  - 12      $m \leftarrow 0$ .
  - 13     **se**  $F_i > F_{alvo}$  **então**
  - 14         Avalia o detector da cascata atual no conjunto de imagens negativas e coloca qualquer detecção falsa no conjunto  $m$ .
-

## 5 *Solução Proposta, Experimentos e Resultados*

*"Don't worry about a thing, 'Cause every little thing is gonna be all right."*

**Three Little Birds – Bob Marley.**

Neste capítulo será apresentada a solução proposta neste trabalho como método de contagem de plaquetas. Serão também descritos os arranjos experimentais realizados tanto durante a concepção do sistema, quanto para a sua validação, e os resultados alcançados.

### 5.1 Solução Proposta

A construção da solução proposta neste trabalho para detecção e contagem de plaquetas em imagens sanguíneas foi dividida em duas etapas distintas: na primeira etapa é realizado o pré-processamento das imagens. Na segunda etapa é apresentado o sistema desenvolvido para a detecção e contagem das plaquetas a partir das imagens geradas na primeira etapa.

### 5.2 Arranjo Experimental

Nesta seção será detalhado o processo de desenvolvimento das etapas de pré-processamento e desenvolvimento do sistema de detecção.

#### 5.2.1 Pré-processamento das imagens

Para compor o banco de imagens a ser utilizado neste projeto foram coletadas imagens sanguíneas contendo plaquetas e os demais elementos sanguíneos. As imagens foram obtidas a partir do banco de imagens da Sociedade Americana de Hematologia (ASH - do inglês, *American Society of Hematology*) [19].

Após formado o banco de imagens a ser usado no projeto, foram realizados testes com os algoritmos de pré-processamento de imagens, apresentados no Capítulo 3, afim de definir quais teriam maior impacto no realce de detalhes e eliminação de ruídos.

Inicialmente as imagens foram convertidas para escala de cinza e foram também separados os seus canais R (vermelho), G (verde), e B (azul). Após isso, foram realizadas a equalização global e a equalização local separadamente. Na Figura 11 são apresentados os resultados que utilizaram equalização global. Na Figura 12 são apresentados os resultados para as imagens às quais se aplicou a equalização local do histograma.

Conforme pode-se observar nas Figuras 11 e 12, Os detalhes das plaquetas são mais ressaltados com relação ao restante da imagem ao se utilizar a imagem em tons de cinza, ou o canal G da imagem. Com relação ao tipo de equalização utilizado, nota-se que a utilização da equalização local auxilia na eliminação do fundo da imagem, deixando em evidência os objetos de interesse.

Após a equalização do histograma, as imagens foram submetidas à testes utilizando os filtros de detecção de borda *Canny*, *Laplace* e *Sobel*, que foram apresentados na Seção 3.6. Os resultados utilizando as imagens após equalização global do histograma podem ser observados nas Figuras 13 e 14. Nas Figuras 15 e 16 são mostrados os resultados após a equalização local das imagens. Dado que os melhores resultados da etapa anterior foram obtidos para o canal G e para a imagem em tons de cinza, somente serão apresentados nesta etapa os resultados destas duas abordagens.

Ao se analisar as imagens das Figuras 13, 14, 15 e 16, é possível observar que a utilização dos filtros de segmentação de borda apenas acrescentou ruído as imagens. Isto se deve ao fato de a imagem conter outros elementos diferentes de plaquetas em maior quantidade e maior tamanho, como por exemplo, as hemácias.

Com base nas análises realizadas após os testes de pré-processamento das imagens, optou-se por utilizar para o desenvolvimento do sistema apenas as imagens em tons de cinza e o canal verde, e equalização global ou local da imagem. Dado que a utilização dos filtros de segmentação apenas conferiram mais ruído às imagens, a utilização destes foi descartada. Para um melhor entendimento dos experimentos que serão descritos posteriormente, as imagens utilizadas nas próximas etapas serão classificadas conforme a cor, e o método de equalização utilizado. Esta classificação é apresentada na Tabela 2.

Tabela 2: Classificação das imagens utilizadas nos experimentos de desenvolvimento do sistema de detecção de plaquetas.

Tipo	Descrição
1	imagem em tons de cinza com equalização global do histograma
2	imagem em tons de cinza com equalização local do histograma
3	canal verde da imagem com equalização global do histograma
4	canal verde da imagem com equalização local do histograma

### 5.2.2 Desenvolvimento do sistema de detecção de plaquetas

Após a conclusão da Etapa 1, foi necessária a construção do software para realizar a detecção das plaquetas. A concepção desta fase pode ser dividida em 3 atividades: a criação do conjunto de treinamento, treinamento do *AdaBoost* para detecção de plaquetas e o desenvolvimento e teste da aplicação de detecção de plaquetas.

O sistema foi desenvolvido utilizando a linguagem de programação C/C++. As simulações foram realizadas em um computador com 2Gb de memória principal, processador Intel core 2 duo 1,73 GHz e sistema operacional Windows XP.

#### Atividade 1 – Criação do conjunto de treinamento.

O conjunto de treinamento foi montado a partir do banco de imagens coletado na etapa anterior. Inicialmente o banco de imagens foi dividido em dois subgrupos: um grupo de treinamento, composto por 75% do total de imagens; e outro grupo de imagens de teste, formado pelos 25% restantes.

A partir das imagens de treino foram construídos dois grupos de imagens: um grupo de imagens positivas (imagens de plaquetas) e um grupo de imagens negativas (imagens que não são plaquetas).

Para compor o banco de imagens positivas, foram extraídas as plaquetas contidas nas imagens do conjunto de treino. As negativas foram obtidas extraíndo fragmentos das imagens de treino que não continham plaquetas.

#### Atividade 2 – Treinamento do *AdaBoost* para detecção de plaquetas.

Para a construção do sistema de detecção foi realizado inicialmente o treinamento do *AdaBoost* para detecção de plaquetas. O treinamento foi feito conforme o algoritmo original do *AdaBoost* (algoritmo 1), apresentado na Seção 4.4.

Durante o treinamento foram utilizadas para compor os classificadores apenas as quatro características básicas de Haar, apresentadas anteriormente na Figura 9.

Ao fim do treinamento são gerados dois arquivos: o primeiro contendo os classificadores fracos encontrados, e o segundo armazenando os pesos associados a cada classificador. Foram realizados 4 experimentos de treinamento do *AdaBoost*. Para cada um dos experimentos o conjunto de imagens positivas e negativas era composto por imagens de um dos tipos apresentados na Tabela 2. Os experimentos realizados são descritos a seguir:

**Experimento 1** – Treinamento do *AdaBoost* utilizando imagens positivas e negativas em tons de cinza e com equalização global do histograma.

**Experimento 2** – Treinamento do *AdaBoost* utilizando imagens positivas e negativas em tons de cinza e com equalização local do histograma.

**Experimento 3** – Treinamento do *AdaBoost* utilizando o canal verde das imagens positivas e negativas, e equalização global do histograma.

**Experimento 4** – Treinamento do *AdaBoost* utilizando o canal verde das imagens positivas e negativas, e equalização local do histograma.

**Atividade 3**– Desenvolvimento e teste da aplicação de detecção.

Após o treino do *AdaBoost*, foi desenvolvido um sistema simples de detecção de plaquetas. A ideia inicial do sistema é apenas que ele seja funcional e capaz de testar os 4 experimentos propostos para treinamento do *AdaBoost*. O funcionamento do sistema é descrito no Algoritmo 3.

---

**Algoritmo 3:** Pseudocódigo sistema de detecção de plaquetas .

---

```
1 Carrega as informações do treinamento.
2 Carrega  $N$  exemplos de imagens.
3  $nAcertos \leftarrow 0$ 
4  $nFalsosPositivos \leftarrow 0$ 
5 para  $n=1, \dots, N$  faça
6   Executa análise da imagem.
7   se detectado objeto de interesse corretamente então
8      $nAcertos \leftarrow nAcertos + 1$ .
9     Exibe a quantidade atual de acertos.
10  se objeto de interesse foi detectado incorretamente então
11     $nFalsosPositivos \leftarrow nFalsosPositivos + 1$ .
```

---

### 5.3 Análise dos Resultados

O treinamento do *AdaBoost* para detecção de plaquetas foi realizado em cada um dos 4 experimentos descritos anteriormente. Assim, após o treinamento, foi executado o teste no sistema de detecção para avaliar o desempenho dos classificadores gerados em cada um dos experimentos. Foram coletados o número de detecções corretas e o número de falsos positivos. O número máximo de objetos de interesse que podem ser detectados em cada experimento é 36, o número de imagens negativas analisadas é 36 e o ótimo para a quantidade de falsos positivos é zero. Os resultados obtidos para cada experimento são apresentados na Tabela 3. O tempo médio de treinamento do *AdaBoost* para cada experimento é mostrado na tabela 4.

Tabela 3: Resultados médios e (desvio padrão), para o treinamento do *AdaBoost* por experimento.

	Cor	Equalização	Acertos	% Acertos	Falsos Positivos	% Falsos Positivos
1	Tons de cinza	Global	35,57(0,68)	98%	1,13(0,97)	3,1%
2	Tons de cinza	Local	32,4(1,85)	90%	5,27(2,15)	14%
3	Canal Verde	Global	34,2(1,32)	95%	5,3(2,22)	14,7%
4	Canal Verde	Local	33,03(1,79)	92%	4,93(2,05)	13,7%

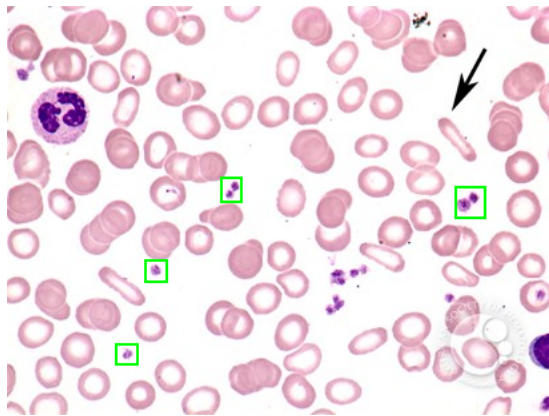
Tabela 4: Resultados médios e (desvio padrão), para o tempo de treinamento do *AdaBoost* por experimento.

	Cor	Equalização	Tempo (s)
1	Tons de cinza	Global	177(6,35)
2	Tons de cinza	Local	180(3,38)
3	Canal Verde	Global	170(1,25)
4	Canal Verde	Local	178(6,92)

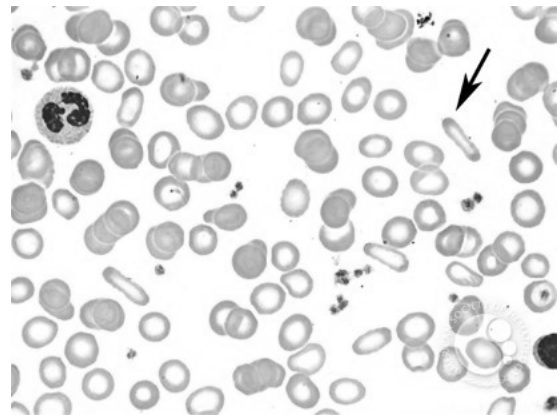
Os gráficos com a distribuição de acertos e falsos positivos dos experimentos são apresentados nas Figuras 17 e 18, respectivamente.

Os dados da tabela mostram que, apesar de a média de acertos entre os 4 experimentos, considerando o desvio padrão, ser bem próxima, os experimentos 1 e 2 conseguem se destacar

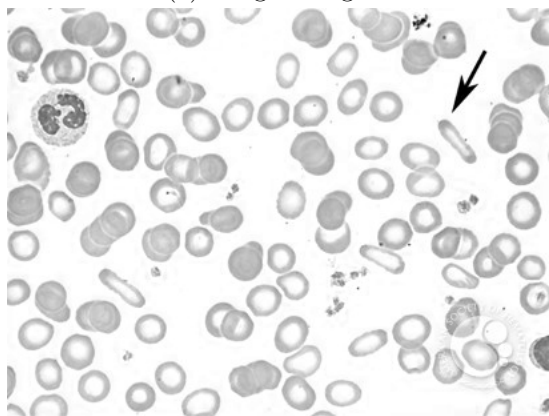
assumindo um percentual de acertos acima de 95%. Entretanto, ao se analisar o número de falsos positivos, o experimento 1 se mostra bem superior aos demais, pois possui uma taxa de falsos positivos inferior a 5%.



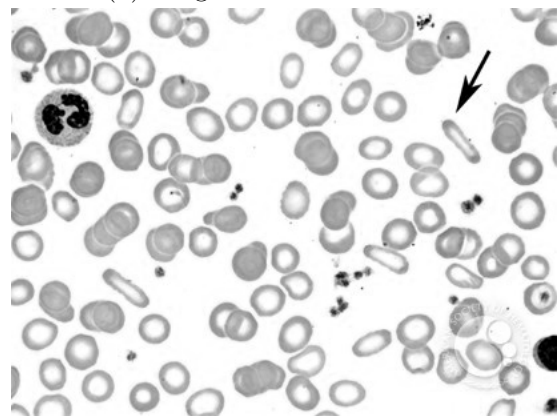
(a) Imagem original.



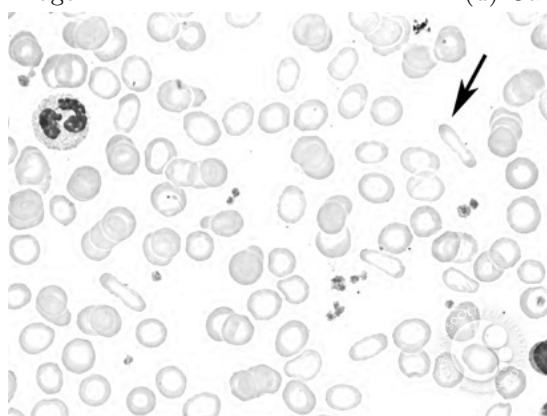
(b) Imagem em escala de cinza.



(c) Canal R da imagem.



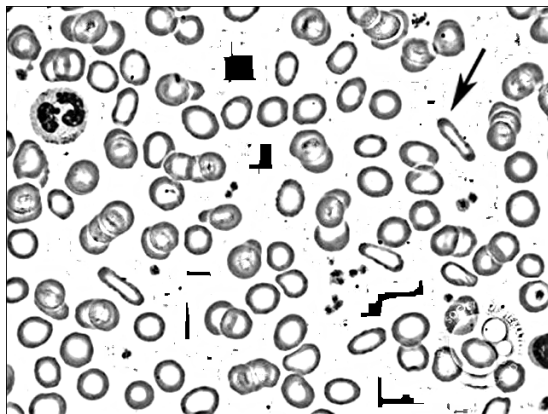
(d) Canal G da imagem.



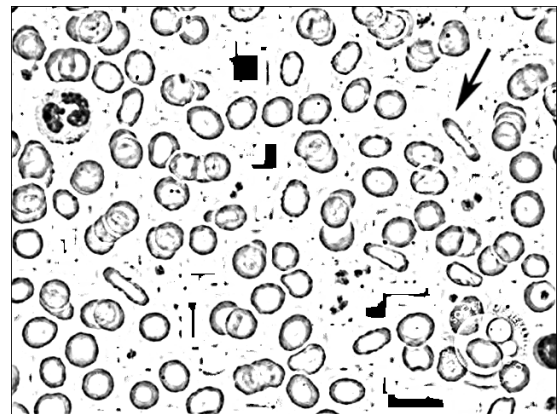
(e) Canal B da imagem.

Figura 11: *Pré-processamento das imagens utilizando equalização global do histograma.*

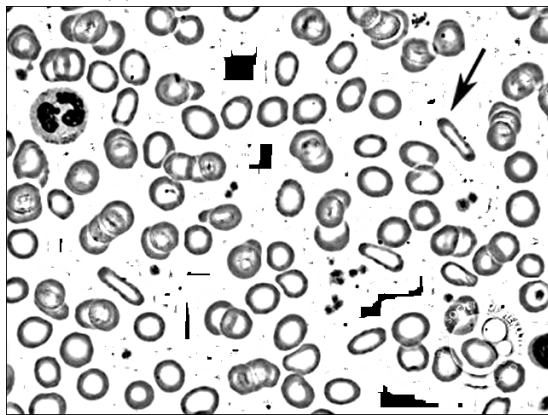




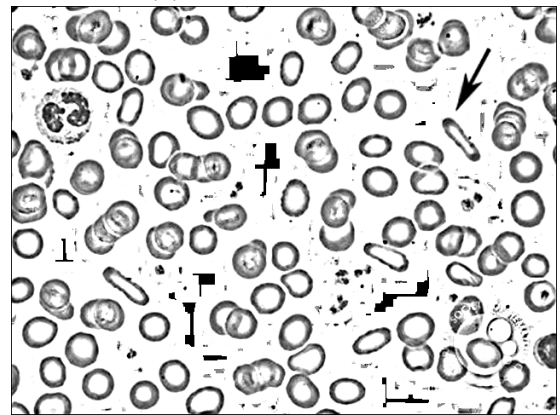
(a) Imagem em escala de cinza.



(b) Canal R da imagem.

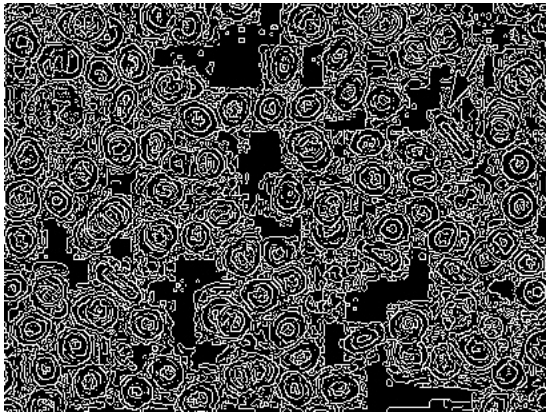


(c) Canal G da imagem.

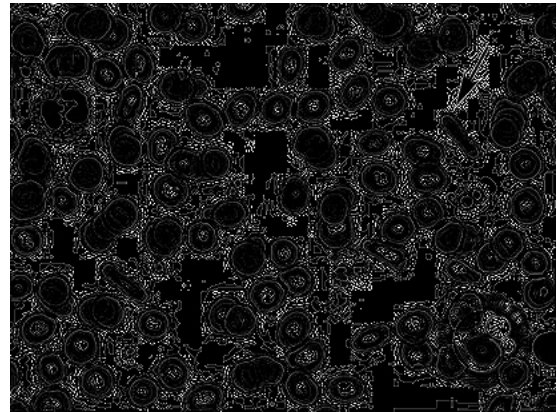


(d) Canal B da imagem.

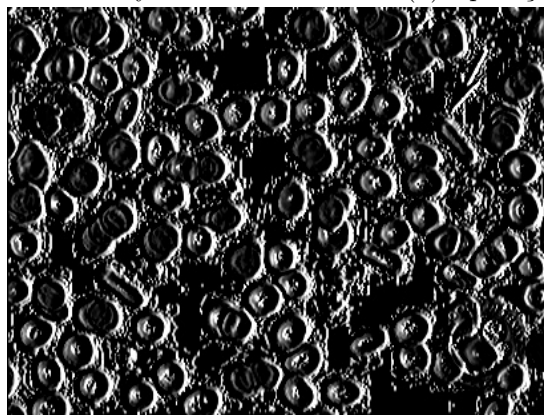
Figura 12: *Pré-processamento das imagens utilizando equalização local do histograma.*



(a) Aplicação do filtro de Canny.

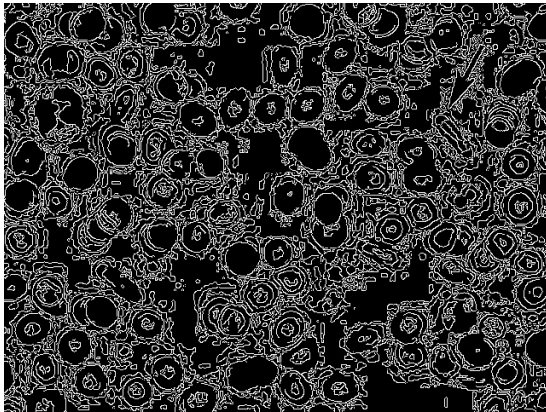


(b) Aplicação do filtro de Laplace.

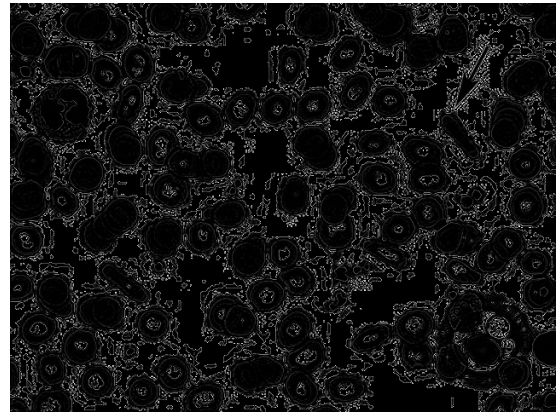


(c) Aplicação do filtro de Sobel.

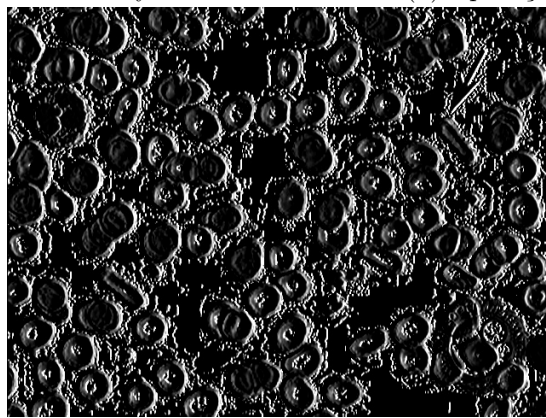
Figura 13: *Aplicação dos filtros de detecção de borda na imagem em tons de cinza após equalização global do histograma.*



(a) Aplicação do filtro de Canny.



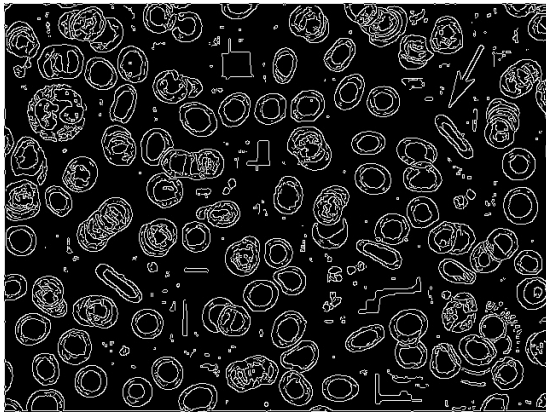
(b) Aplicação do filtro de Laplace.



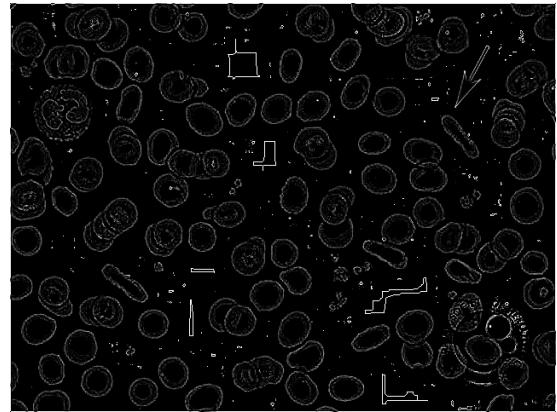
(c) Aplicação do filtro de Sobel.

Figura 14: *Aplicação dos filtros de detecção de borda no canal G da imagem após equalização global do histograma.*

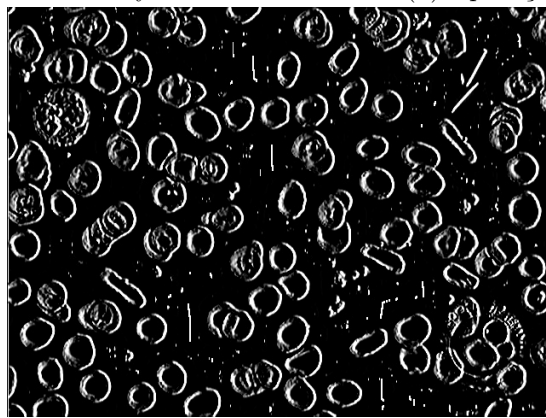




(a) Aplicação do filtro de Canny.

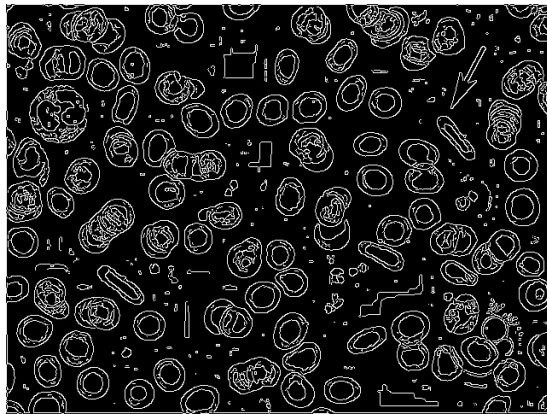


(b) Aplicação do filtro de Laplace.

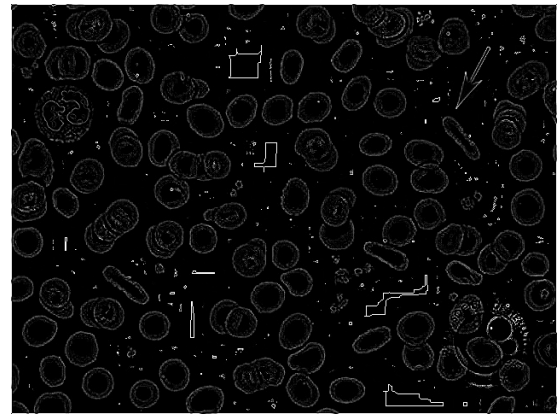


(c) Aplicação do filtro de Sobel.

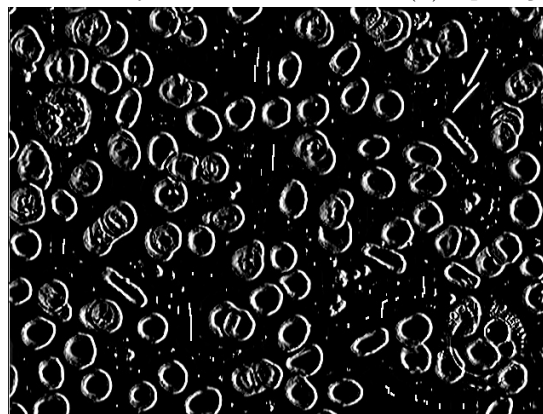
Figura 15: *Aplicação dos filtros de detecção de borda na imagem em tons de cinza após equalização local do histograma.*



(a) Aplicação do filtro de Canny.



(b) Aplicação do filtro de Laplace.



(c) Aplicação do filtro de Sobel.

Figura 16: Aplicação dos filtros de detecção de borda no canal G da imagem após equalização local do histograma.

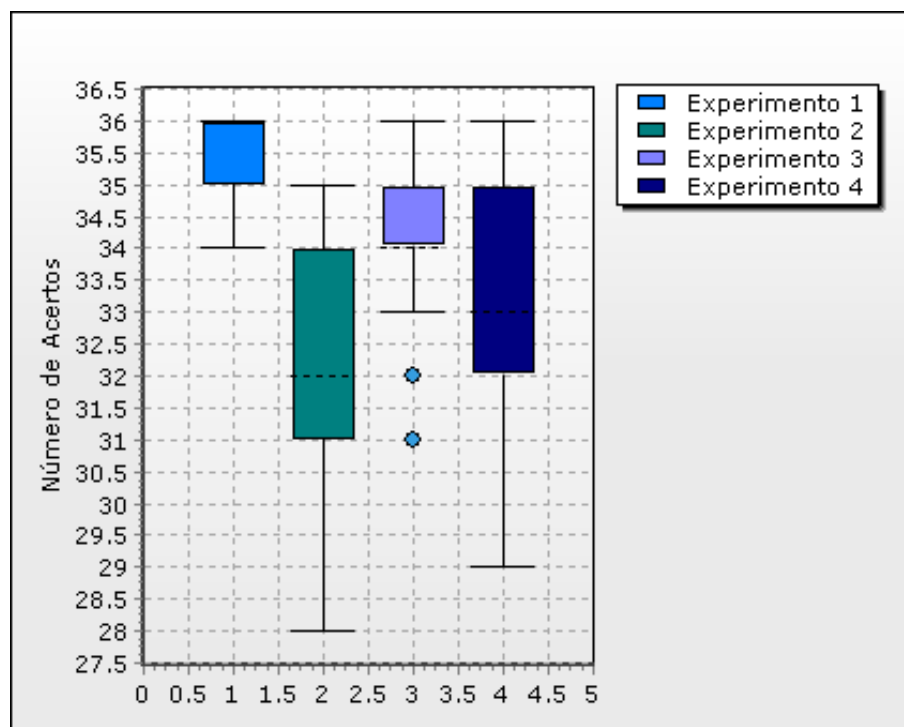


Figura 17: Gráfico da distribuição de acertos para cada experimento.

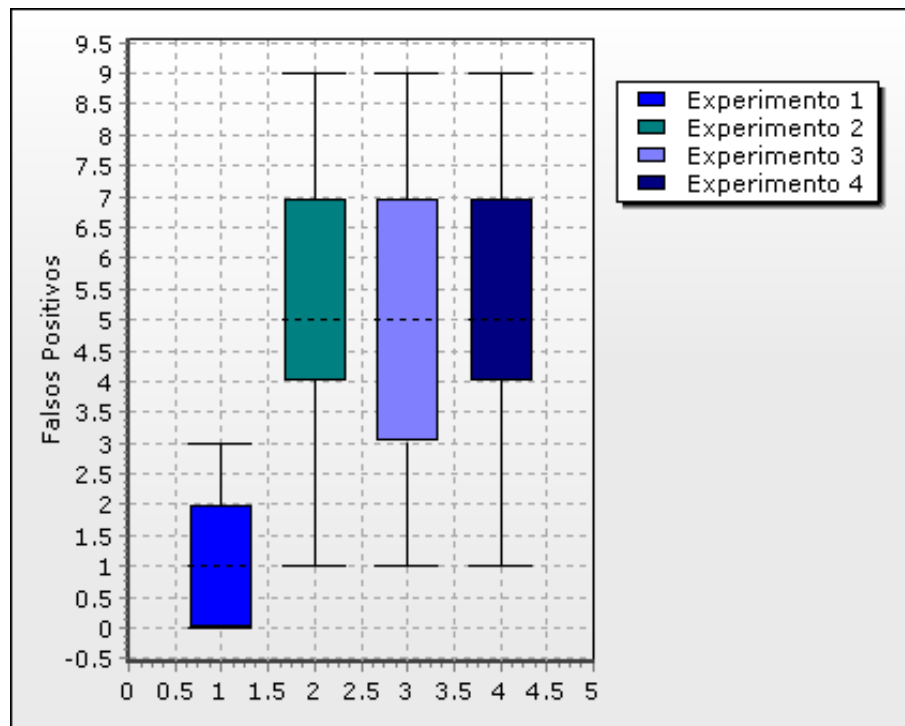


Figura 18: Gráfico da distribuição de falsos positivos para cada experimento.

## 6 *Conclusões e Trabalhos Futuros*

*“A lógica do vento, o caos do pensamento, a paz na solidão. A órbita do tempo, a pausa do retrato, a voz da intuição.”*

**Lenine – Labiata.**

Neste capítulo serão apresentadas as conclusões observadas durante o desenvolvimento deste trabalho, bem como os trabalhos futuros, que poderão ser realizados em complementação deste.

### 6.1 Conclusão

A partir da metodologia e resultados do capítulo anterior, foi possível atestar a viabilidade da solução proposta neste trabalho.

Com o protótipo do sistema de detecção proposto foi possível verificar o desempenho satisfatório nas análises iniciais. A utilização de imagens em tons de cinza, e com o histograma equalizado globalmente, proporcionou resultados mais promissores do que as demais variações analisadas nos experimentos.

Os primeiros resultados com o protótipo da solução indicam que é viável a implementação da solução para a contagem de plaquetas e apontam para a viabilidade de uso desta solução para realizar a contagem dos outros elementos sanguíneos, assim como identificar anomalias nas células. Deste modo, poderá ser desenvolvido um sistema de auxílio no diagnóstico de doenças que afetem tanto a morfologia, quanto a quantidade, das células sanguíneas.

### 6.2 Trabalhos Futuros

Como trabalhos futuros, destacam-se as seguintes ações:

1. Desenvolvimento de uma base de dados positivos e negativos maior, afim de atingir taxas de detecção mais confiáveis e estáveis;
2. Desenvolvimento de características de Haar específicas para o problemas de detecção de plaquetas;
3. Realização de testes estatísticos mais específicos, afim de analisar a influência da cor da imagem, do método de equalização utilizado, e da interação entre estes dois fatores;
4. Treinamento do *AdaBoost* utilizando a cascata de classificadores;
5. Otimização do treinamento utilizando técnicas de Inteligência de Enxames;
6. Construção de um sistema de detecção em tempo real;
7. Treinamento do *AdaBoost* para a detecção dos demais elementos sanguíneos;
8. Desenvolvimento do sistema de detecção de plaquetas e demais elementos sanguíneos aplicando boas práticas de engenharia de software, com ênfase em usabilidade.



## Referências

- [1] COMAR, S. R.; DANCHURA, H. S. M.; SILVA, P. H. Contagem de plaquetas: avaliação de metodologias manuais e aplicação na rotina laboratorial. *Revista Brasileira de Hematologia e Hemoterapia*, v. 31, n. 6, p. 431–436, 2009. ISSN 1516-8484. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-84842009000600011&lng=pt&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-84842009000600011&lng=pt&nrm=iso&tlng=pt)>.
- [2] FAILACE, R. *HEMOGRAMA: MANUAL DE INTERPRETAÇÃO*. 4. ed. Porto Alegre: ARTMED, 2003. 298 p. ISBN 8536301589.
- [3] BAIN, B. J. *Células sanguíneas: um guia prático*. 2. ed. Porto Alegre: ARTMED, 1997. 334 p. ISBN 8573072679.
- [4] GONZALEZ, R. C.; WOODS, R. E. *PROCESSAMENTO DE IMAGENS DIGITAIS*. 1. ed. São Paulo: EDGARD BLUCHER, 2000. 509 p. ISBN 8521202644.
- [5] GOMES, J.; VELHO, L. *Computação Gráfica: Imagem*. 2. ed. Rio de Janeiro: IMPA, 2002. 424 p. ISBN 8524400889.
- [6] CANNY, J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, PAMI-8, n. 6, p. 679–698, 1986. ISSN 01628828. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4767851>>.
- [7] PATIL, R. V.; JONDHALE, K. C. Edge based technique to estimate number of clusters in k-means color image segmentation. In: *2010 3rd International Conference on Computer Science and Information Technology*. IEEE, 2010. p. 117–121. ISBN 978-1-4244-5537-9. Disponível em: <[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=5563647](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5563647)>.
- [8] JIANG, L. et al. Automatic image annotation based on decision tree machine learning. In: *2009 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. IEEE, 2009. p. 170–175. ISBN 978-1-4244-5218-7. Disponível em: <[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=5342168](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5342168)>.
- [9] ZHAI, J.-H.; WANG, X.-Z.; ZHANG, S.-F. Rough-Neural Image Classification using Wavelet Transform. In: *2007 International Conference on Machine Learning and Cybernetics*. IEEE, 2007. p. 3045–3050. ISBN 978-1-4244-0972-3. Disponível em: <[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4370671](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4370671)>.
- [10] DONG, L. et al. Classification of Urinary Sediments Image Based on Bayesian Classifier. In: *2007 International Conference on Mechatronics and Automation*. IEEE, 2007. p. 556–560. ISBN 978-1-4244-0827-6. Disponível em: <[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4303603](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4303603)>.

- [11] HABIB, T. et al. Speeding up Support Vector Machine (SVM) image classification by a kernel series expansion. In: *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008. p. 865–868. ISBN 978-1-4244-1765-0. Disponível em: <[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4711892](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4711892)>.
- [12] VIOLA, P.; JONES, M. J. Robust Real-Time Face Detection. *International Journal of Computer Vision*, Springer, v. 57, n. 2, p. 137–154, 2004. ISSN 09205691. Disponível em: <<http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:VISI.0000013087.49260.fb>>.
- [13] FIRMO, A. C. A. *Classificação Adaboost com Treinamento por Exame de Partículas para Diagnóstico da Esquistossomose Mansônica no Litoral de Pernambuco*. Recife: [s.n.], 2010.
- [14] ACHARYA, T.; RAY, A. K. *Image processing: principles and applications*. [S.l.]: John Wiley and Sons, 2005. 428 p. ISBN 0471719986.
- [15] FREUND, Y.; SCHAPIRE, R. E.; AVENUE, P. A Short Introduction to Boosting. *Society*, Citeseer, v. 14, n. 5, p. 771–780, 1999. ISSN 09128085. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.5148&rep=rep1&type=pdf>>.
- [16] PAPAGEORGIOU, C. P.; OREN, M.; POGGIO, T. A general framework for object detection. *Sixth International Conference on Computer Vision IEEE Cat No98CH36271*, Narosa Publishing House, v. 6, n. January, p. 555–562, 1998. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=710772>>.
- [17] CARBONETTO, P. Robust object detection using boosted learning. *Department of Computer Science, University of British Columbia, Vancouver*, 2002. Disponível em: <<http://people.cs.ubc.ca/pcarbo/cs530project.pdf>>.
- [18] QUINLAN, J. R. Induction of decision trees. *Machine Learning*, Springer, v. 1, n. 1, p. 81–106, 1986. ISSN 1062936X.
- [19] American Society of Hematology. *American Society of Hematology Image Bank*. 2011. Disponível em: <<http://imagebank.hematology.org/>>.