

UNIOESTE – Universidade Estadual do Oeste do Paraná

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

Colegiado de Ciência da Computação

Curso de Bacharelado em Ciência da Computação

**Comparação entre dois Métodos de Segmentação de
Imagens de Eletroforese**

Marco Antonio da Rosa

CASCABEL

2011

MARCO ANTONIO DA ROSA

**COMPARAÇÃO ENTRE DOIS MÉTODOS DE SEGMENTAÇÃO DE
IMAGENS DE ELETROFORESE**

Monografia apresentada como requisito parcial
para obtenção do grau de Bacharel em Ciência
da Computação, do Centro de Ciências Exatas
e Tecnológicas da Universidade Estadual do
Oeste do Paraná - Campus de Cascavel

Orientador: Prof. André Luiz Brun

CASCADEL

2011

MARCO ANTONIO DA ROSA

**COMPARAÇÃO ENTRE DOIS MÉTODOS DE SEGMENTAÇÃO DE
IMAGENS DE ELETROFORESE**

Monografia apresentada como requisito parcial para obtenção do Título de *Bacharel em Ciência da Computação*,
pela Universidade Estadual do Oeste do Paraná, Campus de Cascavel, aprovada pela Comissão formada pelos
professores:

Prof. André Luiz Brun (Orientador)
Colegiado de Ciência da Computação,
UNIOESTE

Prof. Greicy Kiel (Co-Orientadora)
Departamento de Biologia,
Faculdade Assis Gurgacz (FAG)

Prof. Adair Santa Catarina
Colegiado de Ciência da Computação,
UNIOESTE

Cascavel, 03 de novembro de 2011.

DEDICATÓRIA

Aos meus pais e ao meu irmão.

AGRADECIMENTOS

Agradeço a minha mãe, Catarina Starke da Rosa, por todo o amor, apoio, incentivo e confiança depositada em mim, durante toda minha vida e, principalmente nesta etapa que está prestes a terminar.

Agradeço a toda minha família pelo apoio enquanto cursava esta graduação.

Ao meu orientador neste Trabalho de Conclusão de Curso, André Luiz Brun, pela paciência, apoio e dedicação durante a elaboração deste trabalho.

A minha professora do ensino fundamental, Ivanir Tremea Furlan, na pequena escola rural da Linha Limoeiro, que apesar de todas as dificuldades, passou-me uma base sólida de conhecimento, fazendo com que eu conseguisse chegar até aqui.

Aos meus amigos Alessandro R. Franco, Daniel B. Cassanelli, Fernando L. Grando, Fernando Martins, Pedro P. Finamore, Rafael A. de Oliveira e Rafael F. C. Frank, por todas as alegrias que passamos juntos durante o curso e pelo apoio que recebi quando necessitei.

Lista de Figuras

1.1	Imagem de um gel de eletroforese	2
2.1	Ilustração do funcionamento da eletroforese em gel	11
2.2	Exemplo de uma imagem de eletroforese	12
3.1	Representação matricial da imagem	15
3.2	Etapas de um sistema de Processamento de Imagens	17
3.3	Imagem e seu respectivo histograma	19
3.4	Imagens em tons de cinza e seus respectivos histogramas	20
3.5	Uma máscara 3x3 com coeficientes (pesos) arbitrários	21
3.6	Filtragem no domínio espacial com uma máscara 3x3	22
3.7	Duas máscaras do filtro da média, de tamanhos 3x3 e 5x5 respectivamente	23
3.8	Exemplos de aplicação dos filtros de suavização para remoção de ruído	24
3.9	Exemplo de um histograma bimodal	25
3.10	Três histogramas com diferentes características	28
3.11	Aplicação de histogramas de projecção em uma imagem	28
4.1	Etapas constituintes do sistema	30
4.2	Exemplo de uma imagem de eletroforese	30
4.3	Imagem resultante da operação de limiarização	32
4.4	(a) Histograma de projecção vertical; (b) Trilhas detectadas	34
4.5	Algoritmo Iterativo para suavização da projecção	35
4.6	(a) Imagem original; (b) Máximos locais; (c) Resultado da aplicação do filtro iterativo	35
4.7	Algoritmo 1	36
4.8	Algoritmo 2	36
4.9	Algoritmo 7	37
4.10	Matriz de similaridade	38
4.11	Algoritmo para construção do dendograma	38
4.12	Dendograma gerado com base na matriz de similaridade da figura 4.6	39
5.1	Tela principal da ferramenta	41
5.2	Menus com todas as funcionalidades da ferramenta	42

5.3	Tela principal do sistema com uma imagem carregada	42
5.4	Janela apresentada ao usuário para a escolha da característica da imagem	43
5.5	Janela apresentando o resultado da binarização da imagem	43
5.6	Janela mostrando as trilhas detectadas	43
5.7	Dendograma	44

Sumário

Lista de Figuras	vi
Sumário	viii
Resumo	x
Introdução	1
Biologia Molecular	4
2.1 Marcadores Moleculares.....	5
2.1.1 RFLP (Polimorfismo de Comprimento de Fragmentos de DNA).....	7
2.1.2 RAPD (DNA Polimórfico Amplificado ao Acaso).....	7
2.1.3 AFLP (Polimorfismo de Tamanho do Fragmento Amplificado).....	8
2.1.4 DGGE (Eletroforese em Gel de Gradiente Desnaturante).....	8
2.2 Extração do DNA.....	9
2.3 Reação da Polimerase em Cadeia – PCR	10
2.4 Eletroforese em Gel.....	10
Processamento de Imagens	14
3.1 Imagens Digitais.....	15
3.2 Etapas Fundamentais em Processamento de Imagens.....	16
3.3 Realce de Imagens.....	17
3.3.1 Histograma	18
3.3.2 Filtragem no Domínio Espacial.....	20
3.4 Segmentação de Imagens.....	24
3.4.1 Limiarização.....	25
3.4.2 Limiarização Otsu	26
3.4.3 Histogramas de Projeção.....	28
Metodologia	29
4.1 Aquisição das Imagens	30
4.2 Pré-Processamento	30

4.3 Segmentação	33
4.3.1 Segmentação utilizando histogramas de projeção	33
4.3.2 Segmentação utilizando o algoritmo iterativo.....	34
4.4 Análise	37
4.5 Geração do Dendograma	38
Resultados e Discussão	40
5.1 A Ferramenta Construída	40
5.2 Características dos Testes Realizados	44
5.3 Resultados dos Testes	46
Conclusões	49
Referências Bibliográficas	51

Resumo

O processo de análise de imagens de géis de eletroforese é um processo oneroso e minucioso, o que torna tal tarefa complexa dada à subjetividade da percepção humana. Uma alternativa para reduzir erros humanos e tornar o processo de interpretação mais rápido e eficaz é a aplicação de análise digital das imagens. Este trabalho buscou realizar uma comparação quanto à eficácia de dois métodos de segmentação das trilhas destas imagens, utilizando para tanto a construção de uma ferramenta computacional multiplataforma capaz de analisar imagens digitais de eletroforese. Esta ferramenta identifica as trilhas que compõe a imagem e realiza a construção do dendograma representando a similaridade entre as amostras em estudo, cujos materiais genéticos foram identificados na imagem. O sistema obtido mostrou-se bastante rápido e eficiente, propiciando a análise das imagens com alta precisão. Após a realização de testes envolvendo um conjunto de 52 imagens heterogêneas, verificou-se que o método que aplica os histogramas de projeção apresentou maior acurácia (87,47%) na detecção das trilhas, em comparação ao algoritmo iterativo (66,95%). Quanto ao consumo de tempo, os métodos mostraram-se equivalentes (gastando menos de 1 segundo para processar cada imagem, o que é perfeitamente aceitável para o propósito da aplicação).

Palavras-chave: Processamento de Imagens, Visão Computacional, Imagens de Eletroforese, Dendogramas.

Capítulo 1

Introdução

O desenvolvimento de técnicas para o estudo do DNA propiciou um enorme e acelerado conhecimento da biosistemática, evolução e genética de animais e plantas. A grande sensibilidade dessas técnicas, associada à descoberta de regiões com alto grau de variabilidade, tem permitido avanços significativos nos estudos de genética de populações, biogeografia e polimorfismo em populações humanas, para as quais têm sido desenvolvidas aplicações na medicina forense, na determinação de paternidade e no diagnóstico de doenças hereditárias [Arias e Infante-Malachias, 2001].

Foi o surgimento dos marcadores moleculares que possibilitou o estudo em nível de DNA. O principal avanço que essas técnicas trazem é a possibilidade de acessar diretamente o genótipo de um indivíduo, evitando assim, a expressão do fenótipo e a influência do ambiente sobre este [Milach, 1998].

O princípio básico envolvido na obtenção e detecção dos marcadores moleculares reside no uso de eletroforese em gel, a qual consiste na separação eletroquímica de moléculas. Esta por sua vez, é baseada na migração diferencial de moléculas com cargas e tamanhos diferentes, quando submetidas a um campo elétrico.

Um exemplo de imagem de gel de eletroforese é apresentado na figura 1.1. Os corredores verticais são chamados de trilhas ou canaletas (cada trilha representa uma amostra). O material contido nas trilhas, que são as posições nas quais as moléculas param, conforme seu tamanho é chamado de banda.

A análise do resultado da eletroforese é feita comparando-se as trilhas entre si. No entanto, esta comparação é geralmente um processo complexo devido a subjetividade da percepção visual humana. Baseados no mesmo material duas pessoas podem chegar a conclusões diferentes [Machado *et al.*, 1997; Paulino *et al.*, 2007].

Atualmente os pesquisadores analisam as imagens de eletroforese visualmente, e constroem uma matriz de presença (também chamada de matriz de similaridade). Esta matriz

serve como entrada para softwares estatísticos, tais como o GENES [Cruz, 2008], o STATISTICA [StatSoft, 2011], o NTSYS [Rohlf, 1998], o GGMOL [Cruz e Shuster, 2008], o SPSS [IBM Corporation, 2010] e o SAS [SAS Institute Inc., 1999], os quais realizam então as análises de agrupamento e de dispersão dos dados [Faleiro, 2007].

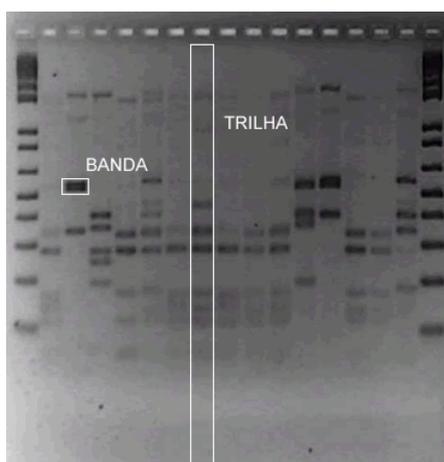


Figura 1.1: Imagem de um gel de eletroforese.
Fonte: Paulino *et al.* (2007).

O processo de eletroforese é bastante aplicado em sequenciamento de DNA e em estudos de variação qualitativa e quantitativa da separação de proteínas ou ácidos nucleicos obtidos de diferentes fontes. Pesquisadores utilizam eletroforese para obter informações das substâncias em estudo, bem como uma comparação da composição de amostras, ou quantificação do montante e propriedades dos diferentes elementos presentes em uma coleção de amostras [Ye *et al.*, 1999].

De acordo com Ye *et al.* [1999], a análise digital de imagens de eletroforese emerge como uma importante aplicação para reduzir erros humanos e melhorar a velocidade da avaliação dos dados. Também, segundo Machado *et al.* [1997], a análise automática do padrão das bandas de uma trilha poderia permitir a avaliação de muitos parâmetros que geralmente são ignorados por analistas humanos.

No entanto, tarefas simples como identificar as trilhas em uma imagem de eletroforese, facilmente feito por especialistas humanos, surge como um problema que pode ser difícil de automatizar [Machado *et al.*, 1997].

Uma abordagem para identificação das trilhas utilizando técnicas de Processamento de Imagens Digitais (PID) foi proposta por Brun e Kiel [2010]. Nesta abordagem são aplicados histogramas de projeção na segmentação de imagens de eletroforese. Outra abordagem para

resolução do problema foi proposta por Machado *et al.* [1997], a qual utiliza um algoritmo iterativo para identificar a posição das trilhas.

O objetivo deste trabalho foi realizar uma comparação entre os dois métodos de segmentação de imagens de eletroforese citados anteriormente. Para tanto, foi desenvolvida uma ferramenta para construção de dendogramas a partir da análise de imagens de eletroforese, com a qual foi possível verificar o desempenho e a acurácia de cada um dos métodos de segmentação.

Tendo em vista que a análise manual de imagens de eletroforese é um processo subjetivo e tedioso para os pesquisadores. A análise automatizada de imagens de eletroforese é interessante porque reduz erros humanos e torna o processo de interpretação mais rápido, podendo inclusive ser utilizada para estudar grandes bancos de dados de amostras.

Portanto, uma ferramenta capaz de realizar a análise automatizada de forma precisa e eficiente seria de grande valia para as áreas de Biologia Molecular e Genética, pois poderia reduzir o tempo das pesquisas e também os custos destas.

No capítulo 2 as técnicas da biologia molecular denominadas marcadores moleculares, utilizadas para o estudo do DNA, são descritas detalhadamente. Também são apresentadas algumas das suas principais aplicações. O processo de análise utilizando estas tecnologias também será explicado.

No capítulo 3 são apresentadas as técnicas e algoritmos de processamento de imagens utilizadas na ferramenta desenvolvida para realizar a análise das imagens.

No capítulo 4 é apresentada a metodologia utilizada para o desenvolvimento da ferramenta proposta. São detalhados os métodos e algoritmos utilizados na ferramenta, bem como as decisões de projeto adotadas para cada técnica utilizada.

No capítulo 5 é apresentado o software desenvolvido, além dos testes realizados visando a comparação dos dois métodos de segmentação e os resultados obtidos.

No último são apresentadas as conclusões e também as perspectivas para trabalhos futuros visando a melhoria dos resultados da ferramenta construída.

Capítulo 2

Biologia Molecular

A Biologia Molecular é o estudo da Biologia em nível molecular, com especial foco no estudo da estrutura e função do material genético. Mais concretamente, a Biologia Molecular investiga as interações entre os diversos sistemas celulares, incluindo a relação entre DNA, RNA e síntese proteica. É um campo de estudo muito amplo, que abrange outras áreas da Biologia e da Química, em especial Genética e Bioquímica. A maior parte do trabalho realizado no campo da Biologia Molecular está relacionado com a obtenção, identificação e caracterização de genes [Schaefer, 2006].

Desde os primórdios da Biologia evolutiva, a caracterização e a explicação da variabilidade genética (também chamada de biodiversidade molecular) intra e inter populacional tem sido um dos principais desafios [Solferini e Selivon, 2001]. Tradicionalmente esta variabilidade tem sido estudada com base no fenótipo (aparência de um indivíduo) utilizando-se técnicas citológicas e morfológicas, porém as interações do genótipo com o ambiente podem mascarar o fenótipo, tornando esta estratégia nem sempre bem sucedida. Com a Biologia Molecular foi possível estudar a variabilidade com base no genótipo, o qual independe do ambiente e geralmente não muda durante o ciclo de vida de um indivíduo [Arias e Infante-Malachias, 2001; Milach, 1998].

Uma das técnicas da Biologia Molecular que possibilitou a detecção da variabilidade existente diretamente ao nível de DNA são os marcadores moleculares [Arias e Infante-Malachias, 2001; Assunção, 2004]. Tais técnicas impulsionaram e revolucionaram a caracterização genética de populações após 1970, permitindo um grau de precisão impraticável até aquele momento. Diferenças na sequência genética agora podem ser diretamente observadas e descritas com um grau de precisão antes impraticável [Assunção, 2004].

Na próxima seção será apresentado o conceito de marcadores moleculares, bem como as principais tecnologias para obtenção destes. Nas seções seguintes serão descritas as etapas necessárias para a obtenção de marcadores genéticos-moleculares. Estas etapas são a extração do DNA, a amplificação via reação em cadeia de polimerase (PCR), separação por eletroforese, fotodocumentação e análise dos marcadores gerados [Ferreira e Grattapaglia, 1998].

2.1 Marcadores Moleculares

Marcadores moleculares são características de DNA, herdadas geneticamente, que permitem diferenciar dois ou mais indivíduos [Milach, 1998]. Mais concretamente, um marcador pode ser definido como todo e qualquer fenótipo proveniente de um gene expresso, como no caso de isoenzimas ou de um segmento específico de DNA, cuja sequência e função podem ou não ser conhecidas e possuem comportamento de acordo com as leis de Mendel [Ferreira e Grattapaglia, 1998].

Com a introdução dessas técnicas da biologia molecular os estudos de identificação, caracterização e mapeamento genético estão sendo realizados com maior segurança, rapidez e eficiência. A utilização de marcadores moleculares está possibilitando a avaliação da variabilidade genética existente dentro e entre espécies distintas [Bored, Neto e Carvalho, 1997].

Os marcadores moleculares permitem compreender e organizar a variabilidade genética de forma única, acessando a variabilidade de DNA, que não é influenciado pelo ambiente como são, por exemplo, os caracteres morfológicos e fenotípicos em geral [Milach, 1998].

Segundo Sartoretto e Farias [2010], o uso destas ferramentas representa um grande avanço na avaliação de populações, as quais poderão ser utilizadas no monitoramento da variabilidade genética, na identificação de indivíduos ou famílias divergentes e construção de mapas genéticos.

O grande interesse pela utilização de marcadores de DNA se dá por estes apresentarem várias características desejáveis, como herança mendeliana simples, ausência de efeitos epistáticos e por serem utilizados em qualquer fase do ciclo biológico. Isso permite acelerar o processo de recombinação e seleção dos indivíduos desejados, reduzindo assim o tempo necessário para a conclusão do estudo em andamento e melhorando a confiabilidade dos resultados [Ferreira e Grattapaglia, 1998].

Dentre as diversas aplicações dos marcadores moleculares, pode-se citar as seguintes [Passaglia e Zaha, 2003]:

- **Estudo da evolução molecular:** A análise de sequência de nucleotídeos fornece os dados mais informativos a respeito da história evolutiva das espécies.
- **Diagnóstico de doenças genéticas:** Doenças como a anemia falciforme, B-talassemias, distrofia muscular de Duchenne, síndrome de Lesch-Nyham, fenilcetonúria, fibrose cística, doença de Tay-Sachs e doença de Gaucher, têm sido diagnosticadas com o auxílio dos Marcadores Moleculares. Para algumas doenças em que a predisposição genética é bastante alta, tais como doenças cardiovasculares e doenças auto-imunes, o uso dos Marcadores Moleculares também tem sido de grande utilidade.
- **Medicina forense:** As sequências VNTR (*Variable Numbers of Tandem Repeats*) têm sido analisadas por Marcadores Moleculares. Essas sequências nucleotídicas curtas apresentam um padrão que varia de pessoa para pessoa, a menos que sejam geneticamente idênticas, como no caso de gêmeos univitelinos ou monozigóticos. Este tipo de análise é uma alternativa simples e rápida para utilização em medicina forense.
- **Diagnóstico de doenças infecciosas:** A utilização de marcadores moleculares tem aumentado a rapidez, a sensibilidade e a especificidade no diagnóstico de doenças infecciosas. Tuberculose, infecções por *Chlamydia Trachomatis*, toxoplasmose, infecção pelo vírus HPV, meningites virais, hepatite B, hepatite C e infecção pelo vírus HIV são algumas das patologias nas quais os marcadores moleculares facilitam o diagnóstico [Filho e Remualdo, 2007].
- **Melhoramento de plantas:** Com um papel fundamental no desenvolvimento da agricultura, gerando novas variedades em espécies de interesse agrônômico que tenham maior produtividade e qualidade, o melhoramento de plantas é auxiliado pelo uso de marcadores moleculares pelo fato de possibilitar a análise direta do genótipo de um indivíduo. Entre as principais aplicações destes em programas de melhoramento de plantas estão: o monitoramento e organização da variabilidade genética, a seleção assistida por marcadores moleculares e a proteção de cultivares [Milach, 1998].

Existem vários tipos de marcadores moleculares, que se diferenciam pela tecnologia utilizada para revelar a variabilidade em nível molecular. E assim variam quanto à capacidade de detectar diferenças entre indivíduos, custos, facilidade de uso e repetibilidade. Os

principais tipos podem ser classificados em dois grupos, conforme a metodologia utilizada para identificá-los: hibridização ou amplificação do DNA. Entre os identificadores por hibridização os principais são os marcadores RFLP (Restriction Fragment Length Polymorphism). Já aqueles revelados por amplificação incluem os marcadores do tipo: RAPD (Random Amplified Polymorphic), AFLP (Amplified Fragment Length Polymorphism DNA) e DGGE (Desnaturing Gradient Gel Electrophoresis) [Milach, 1998]. Estas quatro tecnologias são descritas brevemente a seguir.

2.1.1 RFLP (Polimorfismo de Comprimento de Fragmentos de DNA)

A técnica RFLP foi uma das primeiras a ser utilizada para a produção de marcadores de moleculares. Esta técnica envolve a clivagem de moléculas de DNA por enzimas de restrição, a separação por eletroforese dos fragmentos gerados e sua visualização em forma de bandas observadas por meio da hibridização desses fragmentos com sondas de sequências homólogas, marcadas com radioatividade ou fluorescência [Arias e Infante-Malachias, 2001; Faleiro, 2007].

A RFLP tem como principais vantagens a co-dominância dos marcadores e a alta reprodutibilidade destes. Porém, exige amostras de DNA com alta qualidade e em grandes quantidades para proceder a análise, necessita de sondas específicas para a hibridização e trata-se de uma técnica mais trabalhosa, demorada e de custo elevado quando comparada a outras técnicas [Faleiro, 2007; Lopes *et al.*, 2002].

2.1.2 RAPD (DNA Polimórfico Amplificado ao Acaso)

A técnica RAPD trouxe uma verdadeira “democratização” da análise de polimorfismo molecular, ao permitir a realização de estudos de análise genética em espécies anteriormente não contempladas [Sartoretto e Farias, 2010].

Marcadores RAPD são fragmentos de DNA amplificados via PCR (Reação da Polimerase em Cadeia) utilizando oligonucleotídeos iniciadores curtos (geralmente 21 nucleotídeos) de sequência aleatória. A obtenção desses marcadores envolve a extração e a amplificação do DNA, eletroforese em gel, a coloração com brometo de etídio e fotodocumentação sob luz ultravioleta [Faleiro, 2007].

Esta técnica tem como principais vantagens a facilidade e a rapidez para obtenção dos marcadores, a necessidade de quantidades mínimas de DNA e a universalização das análises. Além disso, é possível trabalhar com qualquer espécie, tendo em vista que não é necessário o conhecimento prévio de dados de sequência de DNA para a construção dos primers utilizados.

As principais desvantagens são a dominância destes marcadores, não diferenciando assim os locos em heterozigose dos locos em homozigose, e a baixa reprodutibilidade das marcas por causa da sensibilidade da técnica às condições experimentais, principalmente quando estas não estão devidamente padronizadas [Faleiro, 2007].

2.1.3 AFLP (Polimorfismo de Tamanho do Fragmento Amplificado)

A técnica AFLP é uma combinação de RFLP e PCR, envolvendo as seguintes etapas para obtenção dos marcadores: extração do DNA, digestão com enzimas de restrição, ligação de adaptadores específicos, ampliação seletiva dos fragmentos via PCR e separação dos fragmentos por eletroforese em gel de alta resolução [Faleiro, 2007].

Entre as principais vantagens desta técnica estão a geração de um grande número de polimorfismos em cada reação, permitindo assim maior rapidez no mapeamento genético, e também a não obrigatoriedade de conhecimento prévio sobre as sequências de DNA para a construção dos primers utilizados. As desvantagens desta técnica são a expressão genética dominante dos marcadores, o custo elevado e as várias etapas e reagentes necessários para a obtenção dos marcadores [Faleiro, 2007].

2.1.4 DGGE (Eletroforese em Gel de Gradiente Desnaturante)

A técnica da DGGE permite analisar produtos da PCR de acordo com suas sequências de pares de bases e não com diferenças no tamanho dos produtos. Esta técnica utiliza géis contendo um gradiente linear de desnaturantes, no qual moléculas de DNA com o mesmo tamanho mas com sequências diferentes de nucleotídeos, apresentam taxas de migração diferentes [Reis Junior *et al.*, 2002].

A DGGE pode ser utilizada para análise de qualquer grupo de genes amplificados por PCR e, assim sendo, esta metodologia é uma poderosa ferramenta para estudos ecológicos [Reis Junior *et al.*, 2002]. As principais vantagens desta técnica são a baixa quantidade de DNA necessária para a análise e a co-dominância dos marcadores. As desvantagens são a necessidade do conhecimento prévio da sequência para a construção dos primers e a dificuldade de reprodutibilidade exigindo que as condições da eletroforese sejam bem padronizadas [Faleiro, 2007].

2.2 Extração do DNA

Como mencionado anteriormente, a extração do DNA é a primeira etapa no processo de desenvolvimento de marcadores moleculares. O isolamento de DNA é uma etapa importante na análise da estrutura e organização do genoma. Independente do tipo de estudo molecular, as preparações de DNA devem produzir amostras puras suficientes para não inibir os tratamentos enzimáticos ou causar interferências nos padrões de migração em gel de eletroforese [Romano e Brasileiro, 1999].

As considerações seguintes são importantes na obtenção de DNA de boa qualidade e devem ser atendidas independentemente do método utilizado [Romano e Brasileiro, 1999]:

1. No caso dos vegetais, bactérias e fungos as paredes celulares devem ser rompidas com o objetivo de liberar os constituintes celulares. Essa etapa é realizada geralmente pelo congelamento do tecido em nitrogênio líquido e posterior quebra mecânica, com o auxílio de um pilão e de um almofariz, no caso de extração em larga escala. A extração em pequena escala pode ser realizada somente na presença de tampão de extração, sem a adição de nitrogênio líquido, utilizando-se um pequeno bastão de vidro e um tubo de microcentrífuga para a maceração.
2. As membranas celulares devem ser rompidas para liberação do DNA. Essa etapa é realizada pela ação de um detergente como SDS (dodecil sulfato de sódio) ou CTAB (brometo de cetiltrimetilamônio).
3. Para evitar a ação de DNAsas, que podem degradar o DNA, os tampões de extração possuem pH por volta de 8,0, enquanto o pH ótimo para ação de DNAsas endógenas fica por volta de 7,0. Outro expediente empregado é a adição de EDTA (ácido etileno diamono tetracético) no tampão de extração.
4. Ácidos nucleicos devem ser separados das proteínas. Para tanto, realiza-se uma ou mais extrações com fenol e/ou clorofórmio, que desnaturam as proteínas tornando-as insolúveis à fase aquosa, onde se encontram os ácidos nucleicos.
5. O DNA deve ser protegido da ação de compostos fenólicos, que oxidam o DNA irreversivelmente. A contaminação por compostos fenólicos pode ser evidenciada pela coloração do DNA que tende a ficar marrom. Para evitar o efeito oxidativo dos polifenóis, deve ser adicionado ao tampão de extração agentes anti-oxidantes, como PVP (polivinilpirrolidona), BSA (albumina de soro bovino) ou b-mercaptoetanol.

6. Os ácidos nucleicos devem ser separados de polissacarídeos, os quais inibem a ação de enzimas de restrição e tornam a amostra de DNA excessivamente viscosa, interferindo na migração do DNA em corridas eletroforéticas. O detergente CTAB é utilizado com essa finalidade, já que polissacarídeos e ácidos nucleicos possuem solubilidade diferenciada na presença desse detergente. Polissacarídeos também podem ser removidos pelo emprego de gradiente de cloreto de cério (CsCl).

2.3 Reação da Polimerase em Cadeia – PCR

A classe de marcadores identificados por amplificação do DNA baseia-se na reação da polimerase em cadeia (PCR, do inglês, *Polymerase Chain Reaction*). A PCR é uma técnica relativamente simples, e consiste na síntese enzimática de milhares ou milhões de cópias de um segmento específico de DNA na presença da enzima de DNA polimerase. Todo o procedimento é realizado *in vitro*, gerando DNA em quantidade suficiente para análises posteriores de forma bastante rápida. A técnica é extremamente sensível, possibilitando a ampliação de DNA a partir de uma quantidade mínima de amostra [Passaglia e Zaha, 2003].

De acordo com Mاتيoli e Passos-Bueno [2001], esta técnica aumentou muito a eficiência da detecção de polimorfismos no nível do DNA, resultando na redução do tempo de execução dos experimentos, diminuindo seu custo e também a complexidade.

As etapas básicas da PCR envolvem a desnaturação do DNA, anelamento dos iniciadores (primers) e extensão das cadeias de DNA (polimerização) que estão sendo amplificadas. Esses primers são sintetizados em laboratório de maneira que sua sequência de nucleotídeos seja complementar a sequências específicas que flanqueiam a reação alvo. Repetindo-se várias vezes as etapas de desnaturação, anelamento e polimerização (geralmente 20 a 30 ciclos) permite-se a amplificação do DNA.

2.4 Eletroforese em Gel

O princípio básico envolvido na obtenção e detecção de cada classe de marcador molecular reside no uso de eletroforese em gel. A análise de DNA por eletroforese é uma das técnicas fundamentais nos laboratórios de pesquisa e de diagnóstico [Naoum, 1990].

O princípio básico da eletroforese está no fato da molécula de DNA possuir carga negativa em valores de pH neutro ou alcalino e conseqüentemente, quando aplicado ou imerso em uma matriz de gel (que pode ser de agarose ou poliacrilamida) submetida a um campo elétrico, migra em direção ao pólo positivo (cátodo). O gel forma uma malha fina por onde as

moléculas de DNA migram de acordo com a carga e o tamanho dos fragmentos. Fragmentos menores migram mais facilmente por essa malha do que os fragmentos maiores, por isso em um dado momento da eletroforese moléculas de tamanhos distintos se encontram em diferentes pontos do gel [Arias e Infante-Malachias, 2001; Corrêa e Possik, 2011]. A figura 2.1 ilustra o sistema de eletroforese em gel.

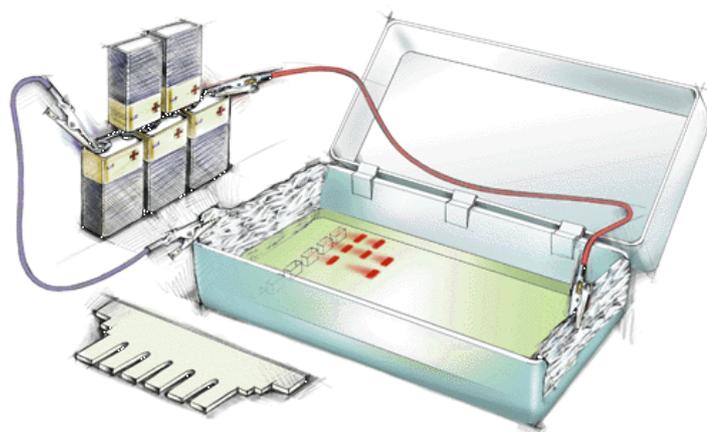


Figura 2.1: Ilustração do funcionamento da eletroforese em gel.
Fonte: <http://edu.glogster.com/media/5/18/37/24/18372460.gif>

Para a execução da eletroforese em gel, as amostras de DNA são aplicadas em cavidades no gel, feitas durante a sua polimerização. Aplica-se então a diferença de potencial e a corrida eletroforética inicia-se, após um período de tempo a corrente elétrica é interrompida.

O passo seguinte é a visualização dos fragmentos, o que geralmente ocorre por métodos químicos. O brometo de etídeo é um dos agentes utilizados, o qual se intercala com as moléculas de DNA e se torna fluorescente quando exposto à luz ultravioleta. Porém, é um agente mutagênico, e por isso deve ser manuseado com muito cuidado. A coloração por prata também tem sido empregada e, apesar de envolver várias etapas, é um método mais sensível, possibilitando a detecção de pequenas quantidades de DNA. Um terceiro método para a visualização das bandas no gel consiste na utilização de material radioativo. Fósforo ou enxofre radioativos são utilizados para marcar as moléculas de DNA antes da corrida eletroforética, e a detecção dos fragmentos é feita pela exposição do gel a um filme de raio-X. Esse tipo de marcação é muito eficaz na detecção de fragmentos pequenos [Arias e Infante-Malachias, 2001].

Por último o gel é fotografado para que possa ser realizada a análise estatística dos marcadores gerados. A figura 2.2 mostra o padrão de fragmentos resultado da eletroforese em uma pesquisa para identificação de espécies de formigas.

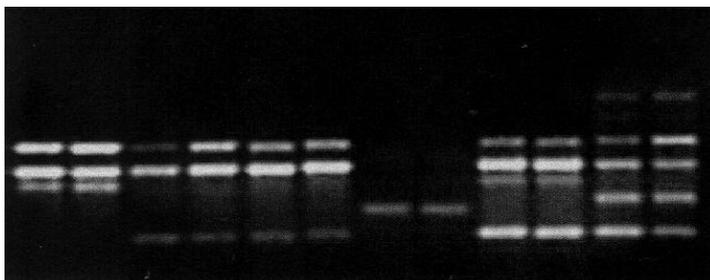


Figura 2.2: Exemplo de uma imagem de eletroforese.
Fonte: Grutzmacher *et al.* (2007).

Quanto ao tipo de substrato (agarose ou poliacrilamida) usado para a preparação do gel depende da técnica aplicada e do tamanho dos fragmentos de DNA que se pretende separar, devido à diferença no tamanho dos poros desses géis. O gel de agarose é indicado para a separação de fragmentos grandes que variam de 300 a 20000 pb (pares de bases), e o gel de poliacrilamida para separação de fragmentos pequenos, de 10 a 100 pb [Arias e Infante-Malachias, 2001].

Embora o gel de poliacrilamida seja mais efetivo para separar fragmentos pequenos, a poliacrilamida em solução, isto é, quando não polimerizada, é neurotóxica. Assim, a preparação do gel exige certa cautela. Além disso, a preparação é mais custosa, requer a mistura de componentes adicionais à poliacrilamida para ajudar na polimerização, e esta ainda é mais demorada. Já o gel de agarose é feito apenas através da mistura de um tampão e agarose, não apresentando toxicidade, e ainda sua polimerização é mais rápida [Corrêa e Possik, 2011].

Segundo Solferini e Selivon [2001], a eletroforese em gel tem como uma de suas vantagens a possibilidade de se amostrar vários indivíduos por vez e também ter diversos locos analisados para cada um.

O princípio de análise é o mesmo para todos os tipos de marcadores moleculares; marcadores comuns entre as amostras significam semelhança genética e marcadores não-comuns significam diferenças genéticas [Ferreira e Grattapaglia, 1998]. O primeiro passo para a análise é a codificação dos fragmentos moleculares em dados binários, onde normalmente são codificados como 1 (um) na presença do marcador e 0 (zero) na ausência do marcador.

Em seguida os dados codificados são utilizados para a estimativa de índices de similaridade ou de distância genética entre cada par de amostras. Com base nos índices é estabelecida uma matriz de similaridade ou de distâncias entre os acessos a qual vai servir de base para as análises de agrupamento e de dispersão das amostras. As análises de agrupamento normalmente são baseadas em métodos hierárquicos que podem utilizar diferentes critérios de agrupamento, como por exemplo: vizinho mais próximo (*Single Linkage*), vizinho mais distante (*Complete Linkage*) e baseado na média das distâncias (*Unweighted Pair-Group Method Using Arithmetic Average*) [Faleiro, 2007].

Capítulo 3

Processamento de Imagens

Os métodos de processamento de imagens digitais possuem dois campos principais de aplicação: melhoria de informação visual para a interpretação humana e o processamento de dados de cenas para percepção automática através de máquinas [Gonzalez e Woods, 2000].

A área de processamento de imagens vem crescendo vigorosamente desde a década de 60. Técnicas da mesma são utilizadas para resolver uma variedade de problemas, nos quais é necessário melhorar a informação visual para a análise e interpretação humanas. Dentre as áreas que são apoiadas pelo processamento de imagens estão a pesquisa espacial, medicina, geografia, arqueologia, física, astronomia, biologia, medicina nuclear, apoio à lei, defesa e aplicações comerciais [Gonzalez e Woods, 2000].

O outro campo de aplicação de técnicas de processamento de imagens digitais consiste na solução de problemas relacionados à percepção por máquina ou visão computacional (VC), onde o objetivo é a extração de informações de uma imagem e a identificação e classificação de objetos presentes nessa imagem, de forma adequada para o processamento computacional. Estas informações geralmente são diferentes das características utilizadas pelo ser humano na interpretação do conteúdo de uma imagem [Conci, Azevedo e Leta, 2008; Gonzalez e Woods, 2000].

A percepção por máquina é frequentemente utilizada nas mais diversas áreas, como por exemplo, reconhecimento militar, inspeção de peças em linhas de montagem, processamento automático de impressões digitais, análise de resultados de raio-X, amostras de sangue em tela, reconhecimento automático de caracteres, processamento de imagens aéreas e de satélites para previsão do tempo e monitoração do plantio [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

3.1 Imagens Digitais

A palavra imagem tem origem do grego *imago*, que significa representação visual de um objeto. Uma imagem pode ser adquirida (fotos, filmes, cenas, etc.) ou gerada pelo ser humano (pinturas, desenhos, esculturas, etc.). Desta forma, uma imagem consiste em qualquer forma visual de expressão de uma ideia [Conci, Azevedo e Leta, 2008].

Segundo [Conci, Azevedo e Leta, 2008], uma imagem digital é um formato de representação que possibilita sua utilização por computadores. Para o processamento a imagem é digitalizada, ou seja, é convertida do domínio contínuo (real) para uma representação discreta (digital).

Uma imagem pode ser descrita por uma função bidimensional $f(x,y)$ de intensidade luminosa, onde x e y denotam as coordenadas espaciais e o valor de f em qualquer ponto representa o brilho (ou nível de cinza) da imagem naquele ponto. Uma representação digital, portanto, é uma imagem $f(x,y)$ discretizada tanto em coordenadas espaciais quanto em brilho [(Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008)].

Uma imagem digital pode ser representada por uma matriz, como ilustrado na figura 3.1, na qual os índices de linha e coluna referenciam a intensidade ou o brilho de um intervalo contínuo amostrado no ponto correspondente da cena. Um ponto (x,y) de uma imagem digitalizada, é chamado de pixel, como abreviação de “*picture element*” ou “elementos da imagem” [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

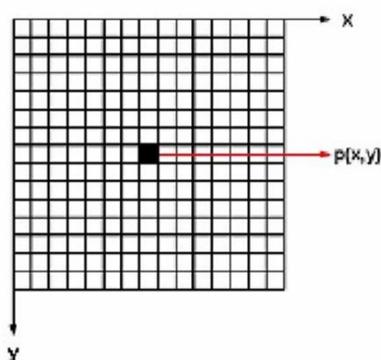


Figura 3.1: Representação matricial da imagem.

3.2 Etapas Fundamentais em Processamento de Imagens

Para resolver um problema utilizando processamento de imagens é necessário aplicar algumas etapas fundamentais [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. A seguir estão descritas cada uma destas. A figura 3.2 ilustra como elas relacionam-se entre si.

- **Aquisição:** Tem como objetivo adquirir uma imagem digitalizada. Isto é feito com a captura da imagem por meio de um sensor de imageamento e a digitalização do sinal produzido pelo sensor. Os principais dispositivos para aquisição de imagens são câmeras fotográficas, tomógrafos médicos, satélites e scanners.
- **Pré-Processamento:** Esta etapa visa melhorar a qualidade da imagem, que pode apresentar imperfeições ou degradações decorrentes da aquisição, de forma a aumentar as chances de sucesso nos processos seguintes. Tipicamente são aplicadas técnicas de realce de imagens para atenuação de ruído, correção de contraste ou brilho e suavização de determinadas propriedades da imagem.
- **Segmentação:** A segmentação é o processo que subdivide uma imagem em partes ou objetos que a constituem. Quando se deseja a segmentação automática esta é uma das tarefas mais críticas no processamento de imagens digitais, pois algoritmos de segmentação fracos ou erráticos quase sempre acarretam em falha no processamento.
- **Representação e descrição:** Esta etapa tem como objetivo extrair características das imagens resultantes da segmentação através de descritores que permitam caracterizar com precisão os elementos desejados na cena. Estes descritores devem ser representados por uma estrutura de dados adequada ao algoritmo de reconhecimento. A entrada desta etapa ainda é uma imagem, mas a saída é um conjunto de dados correspondentes àquela imagem.
- **Reconhecimento e Interpretação:** Esta última etapa envolve o reconhecimento e a interpretação dos componentes de uma imagem. Reconhecimento é o processo que atribui um rótulo a um objeto, baseado nas características fornecidas pelo seu descritor. O processo de interpretação consiste em atribuir um significado a um conjunto de objetos reconhecidos.

O conhecimento sobre o domínio do problema está codificado em um sistema de processamento de imagens na forma de uma base de conhecimento. O tamanho e a complexidade desta base de conhecimento depende de cada aplicação.

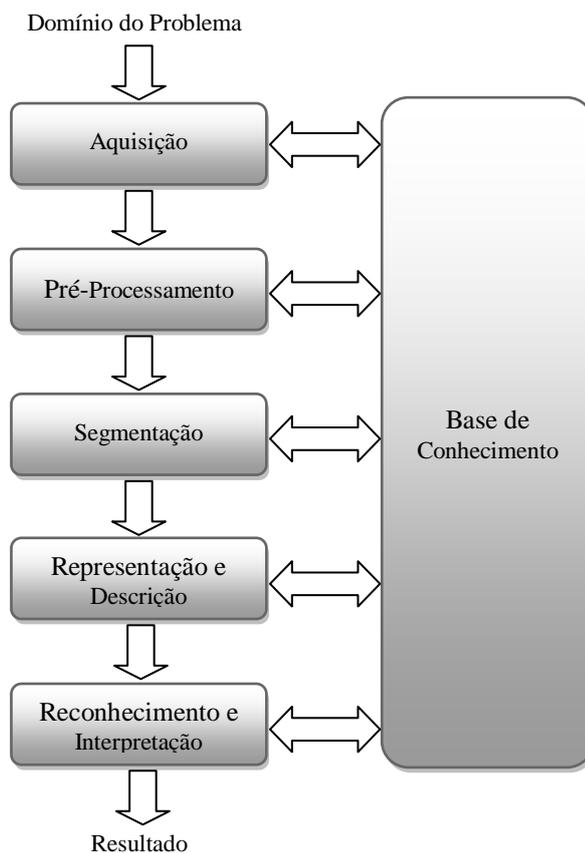


Figura 3.2: Etapas de um sistema de Processamento de Imagens.
 Fonte: Pedrini e Schwartz (2008).

A base de conhecimento é utilizada para guiar a operação de cada módulo, e também auxilia em operações de realimentação entre eles.

É importante salientar que, nem todas as aplicações de processamento de imagens requerem todas estas etapas. Várias aplicações práticas são desempenhadas sem uma base de conhecimento explícita e, em alguns casos, nem mesmo todos os outros módulos são necessários. Por exemplo, o realce de imagens para interpretação visual humana raramente vai além da etapa de pré-processamento. A etapa de reconhecimento e interpretação geralmente está presente apenas em aplicações de análise de imagens, onde o objetivo é a extração automática (ou parcialmente automática) de informação a partir de uma imagem [Gonzalez e Woods, 2000].

3.3 Realce de Imagens

O principal objetivo das técnicas de realce de imagens é acentuar ou melhorar a aparência de determinadas características da imagem, tornando-a mais adequada para uma aplicação específica [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. Estas técnicas são bastante

dependentes da aplicação em questão, ou seja, um método pode ser apropriado para uma aplicação, mas pode não ser a melhor abordagem para outra [Gonzalez e Woods, 2000].

Segundo Pedrini e Schwartz [2008], é útil aplicar as técnicas de realce em imagens que sofreram degradação ou perda de qualidade em decorrência da introdução de ruído, perda do contraste, borramento, distorção causada pelo equipamento de aquisição ou condições inadequadas de iluminação.

De acordo com Gonzalez e Woods [2000], não há uma teoria geral para realce de imagens. Quando uma imagem é processada para interpretação visual, a avaliação da qualidade da imagem resultante da aplicação de um método particular é feita visualmente por um observador. Sendo este um processo altamente subjetivo, torna-se difícil elaborar uma definição do que seria uma “boa imagem”, através da qual se possa comparar o desempenho dos algoritmos utilizados. No caso do processamento de imagens para percepção por computador, esta avaliação é um pouco mais fácil: dada uma aplicação, o melhor método de realce seria aquele que produzisse os melhores resultados para o reconhecimento por máquina. Entretanto, mesmo nas situações em que um critério de desempenho possa ser definido, usualmente será necessário realizar diversas tentativas antes de poder estabelecer uma abordagem adequada de processamento de imagens.

3.3.1 Histograma

O histograma de uma imagem corresponde à distribuição dos níveis de cinza da imagem, ou seja, quantos pixels na imagem possuem determinado nível de cinza. O histograma geralmente é representado por um gráfico de barras, que indica o número (ou o percentual) de pixels na imagem para cada nível de cinza [Conci, Azevedo e Leta, 2008; Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. A figura 3.3 mostra uma imagem e seu respectivo histograma.

A probabilidade de ocorrência de cada nível de cinza em uma imagem pode ser expressa como:

$$p(r_k) = \frac{n_k}{n} \quad (3.1)$$

onde:

- r_k é o k -ésimo nível de cinza;
- n_k = número de pixels na imagem com esse nível de cinza;
- n = número total de pixels na imagem;

- $k = [0, L - 1]$, onde L é o número de níveis de cinza da imagem digitalizada;
- $p(r_k)$ = probabilidade do k -ésimo nível de cinza.

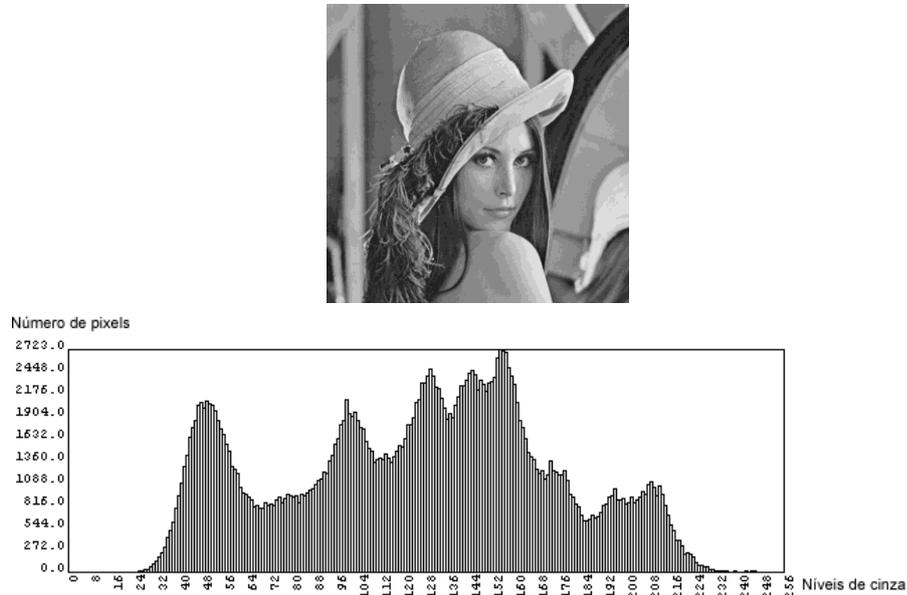


Figura 3.3: Imagem e seu respectivo histograma.

O histograma de uma imagem pode ser utilizado para avaliar a qualidade desta quanto ao nível de contraste e quanto a sua luminosidade média, ou seja, se a imagem é predominantemente clara ou escura [Conci, Azevedo e Leta, 2008; Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. A figura 3.4 mostra a relação entre imagem e histograma de quatro tipos básicos de imagens. A figura 3.4(a) possui um histograma com níveis de cinza bem distribuídos ao longo da escala, correspondendo a uma imagem com alto contraste. O histograma da figura 3.4(b) tem uma forma estreita que indica uma escala dinâmica pequena e, portanto, corresponde a uma imagem com baixo contraste. O histograma da figura 3.4(c) mostra que os níveis de cinza estão concentrados em direção à extremidade clara do intervalo de níveis de cinza, o que corresponde a uma imagem com características predominante claras. Exatamente o oposto se verifica para a figura 3.4(d), sendo esta uma imagem predominantemente escura.

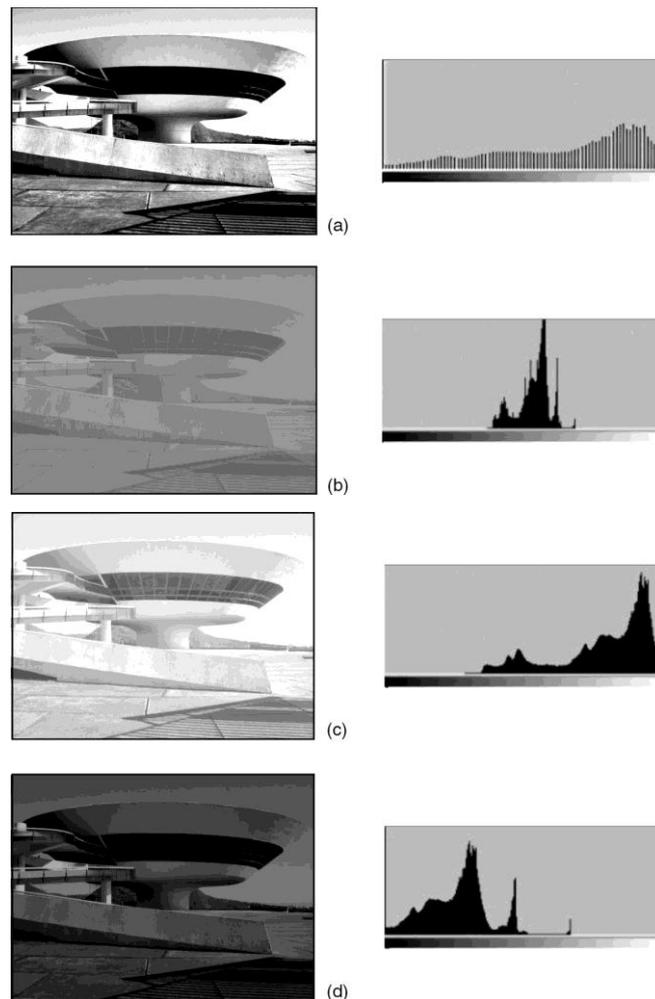


Figura 3.4: Imagens em tons de cinza e seus respectivos histogramas. (a) Imagem de alto contraste. (b) Imagem de baixo contraste. (c) Imagem clara. (d) Imagem escura.
 Fonte: Conci, Azevedo e Leta (2008).

De acordo com Conci, Azevedo e Leta [2008], o conceito de histograma também é aplicado a imagens coloridas. Neste caso, o histograma pode ser representado pelo número de tons presentes na imagem, se for uma imagem com paleta de cores. A imagem também pode ser decomposta e, para cada componente, é calculado o histograma correspondente de cor. No caso de uma imagem no formato RGB, por exemplo, deve-se calcular um histograma separado para cada uma das bandas R, G e B.

3.3.2 Filtragem no Domínio Espacial

O domínio espacial refere-se ao próprio plano da imagem, ou seja, ao conjunto de pixels que compõe a imagem. Sendo assim, filtros no domínio espacial são procedimentos que operam diretamente sobre os pixels da imagem [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Segundo Gonzalez e Woods [2000] e Pedrini e Schwartz [2008], os filtros normalmente são classificados em três categorias: passa-baixas, passa-altas e passa-faixa. A seguir são apresentadas as características de cada categoria:

- **Passa-baixas:** Um filtro passa-baixa atenua ou elimina os componentes de alta-frequência e deixa as frequências baixas inalteradas, ou seja, “deixa passar” as baixas frequências. Os componentes de alta frequência estão relacionados com informações de detalhes da imagem, como bordas e outros detalhes finos. O efeito resultante da filtragem passa-baixa é o borramento da imagem.
- **Passa-altas:** Um filtro passa-alta atenua ou elimina os componentes de baixa frequência e não altera as frequências altas. Os componentes de baixa frequência são responsáveis pelas características que variam lentamente em uma imagem, tais como o contraste total e a intensidade média. Sendo assim, o efeito resultante da filtragem passa-alta é uma redução destas características, correspondendo a uma aparente agudização das bordas e outros detalhes finos.
- **Passa-faixa:** Um filtro passa-faixa realça intervalos selecionados de frequências entre altas e baixas frequências. Estes filtros geralmente são usados para restauração de imagens, e raramente são interessantes para realce de imagens.

No domínio espacial, o processo de filtragem normalmente é realizado por meio de matrizes denominadas máscaras, as quais são aplicadas sobre a imagem. A cada posição da máscara está associado um valor numérico, o qual é chamado de peso ou coeficiente. A aplicação da máscara com centro na coordenada (x,y) da imagem, consiste na substituição do valor no pixel na posição (x,y) por um novo valor, o qual é resultado da soma dos produtos entre os coeficientes da máscara e as intensidades dos pixels sob a máscara [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. A figura 3.5 mostra uma máscara genérica de tamanho 3×3 .

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

Figura 3.5: Uma máscara 3×3 com coeficientes (pesos) arbitrários.

Denotando os níveis de cinza dos pixels na imagem sob a máscara por z_1, z_2, \dots, z_9 , a resposta da máscara é

$$R = w_1 * z_1 + w_2 * z_2 + \dots + w_9 * z_9 \quad (3.2)$$

Considerando que o centro da máscara está numa posição (x,y) na imagem, o nível de cinza do pixel nesta posição será substituído por R . Este processo é repetido em todas as posições de pixels da imagem, e recebe o nome de convolução. A figura 3.6 ilustra esta operação.

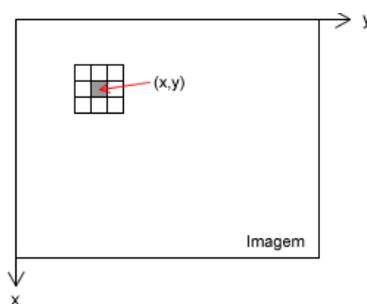


Figura 3.6: Filtragem no domínio espacial com uma máscara 3x3.
Fonte: Gonzalez e Woods (2000).

Os valores de R são armazenados em uma nova imagem, ao invés de mudar os valores de pixel na imagem original. Isto evita que os pixels já alterados influenciem no resultado da aplicação da máscara, ou seja, cada pixel da imagem resultante depende apenas dos pixels da imagem original [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Embora outras formas de vizinhança, tal como aproximações de círculos, sejam as vezes utilizadas, as matrizes retangulares e quadradas são muito mais comuns devido à menor complexidade de implementação [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

É importante ressaltar que são os valores dos coeficientes da máscara que determinam a natureza do processo, ou seja, determinam se o resultado do processo de convolução será, por exemplo, o aguçamento ou o borramento da imagem [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Filtros de suavização

Os filtros passa-baixas atenuam as altas frequências, as quais correspondem às transições abruptas na imagem, por isso o efeito destes filtros é o suavização da imagem. Estes filtros são utilizados para borramento e redução de ruídos em imagens. O borramento é utilizado no

pré-processamento para remoção de pequenos detalhes de uma imagem antes da remoção de grandes objetos, e também para a conexão de pequenas discontinuidades em linhas e curvas [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Um exemplo de filtro de suavização é o filtro da média, no qual cada pixel é substituído pelo valor médio de seus vizinhos. A máscara deste filtro possui todos os coeficientes iguais a 1, e o resultado da convolução é dividido por um fator de normalização que é, geralmente, igual à soma dos coeficientes da máscara [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008]. Duas máscaras, com tamanhos diferentes, deste filtro são mostradas na figura 3.7.

$$\frac{1}{9} \times \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \quad \frac{1}{25} \times \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

Figura 3.7: Duas máscaras do filtro da média, de tamanhos 3x3 e 5x5 respectivamente.

A aplicação do filtro da média em uma imagem não altera os níveis de cinza de regiões homogêneas, isto é, regiões com níveis de cinza constantes [Pedrini e Schwartz, 2008].

Uma desvantagem do filtro da média é que ele suprime detalhes finos e bordas da imagem. Se o objetivo for alcançar a redução de ruído sem borrar muito a imagem, uma alternativa consiste na utilização do filtro da mediana [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

O filtro da mediana consiste em substituir o valor de cada pixel pela mediana das intensidades dos pixels na vizinhança do pixel. Para uma vizinhança de $n \times n$ pixels, sendo n um número ímpar, a mediana das intensidades ordenadas em ordem crescente ou decrescente será o elemento localizado na posição $(n^2 + 1)/2$ [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Quando aplicado em imagens com ruído impulsivo do tipo sal-e-pimenta, o filtro da mediana é bastante eficiente na redução do ruído. Isto ocorre porque as intensidades dos pixels que diferem significativamente de seus vizinhos (valores altos ou baixos), em uma dada vizinhança, são descartados pelo filtro. Além disso, o filtro da mediana não introduz na imagem resultante valores de intensidades diferentes daqueles presentes na imagem original, e por não afetar muito as bordas, este filtro pode ser aplicado iterativamente em uma imagem [Pedrini e Schwartz, 2008].

A figura 3.8 apresenta o resultado da filtragem utilizando o filtro da média e o filtro da mediana, com máscaras de tamanho 3x3, em uma imagem corrompida por ruído impulsivo do tipo sal-e-pimenta. A superioridade, neste caso, do filtro da mediana em relação ao filtro da média é evidente. Também é possível perceber, a perda de definição da imagem devido ao borramento em ambos os resultados.

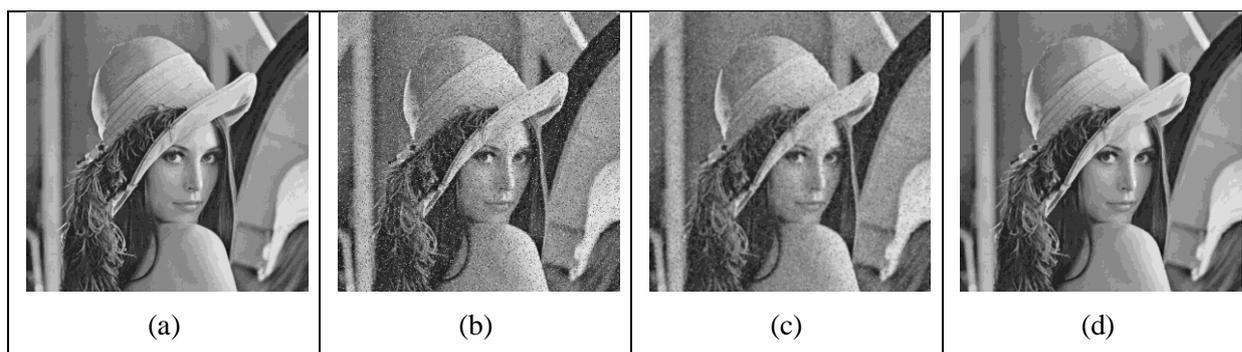


Figura 3.8: Exemplos de aplicação dos filtros de suavização para remoção de ruído. (a) Imagem Original; (b) Imagem corrompida por ruído impulsivo do tipo sal-e-pimenta; (c) Resultado da filtragem com filtro da média 3x3; (d) Resultado da filtragem com filtro da mediana 3x3.

Quanto maior for o tamanho da máscara utilizada para a filtragem, maior será o efeito de borramento e conseqüentemente a perda de definição da imagem [Conci, Azevedo e Leta, 2008; Gonzalez e Woods, 2000].

O filtro da média é considerado um filtro linear, pois o valor resultante de cada pixel é a combinação linear dos valores dos pixels em uma vizinhança local do pixel em questão na imagem original. Já o filtro da mediana é classificado como um filtro não-linear, porque sua operação baseia-se diretamente nos valores dos pixels na vizinhança considerada, não utilizando assim coeficientes na máscara. Sendo assim, nos filtros não-lineares a máscara serve apenas para definir a vizinhança utilizada para computar o valor de cada pixel.

3.4 Segmentação de Imagens

Para analisar uma imagem é necessário extrair, ou isolar, os objetos de interesse do restante da imagem. A segmentação subdivide uma imagem em suas partes ou objetos constituintes, por exemplo, separando os objetos do fundo. Portanto, a segmentação da imagem geralmente é o primeiro passo para se proceder a sua análise [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Segundo Gonzalez e Woods [2000], a segmentação automática das imagens é geralmente uma das tarefas mais difíceis no processamento de imagens. Essa etapa determina o eventual

sucesso ou fracasso na análise; quando realizada de forma efetiva quase sempre garante sucesso no reconhecimento. Para fazer uma segmentação automática robusta, quanto a variações das intensidades das imagens, o algoritmo deve selecionar o valor de corte baseado nas intensidades de cinza presentes na imagem, geralmente analisando o histograma [Conci, Azevedo e Leta, 2008].

As abordagens para segmentação de imagens monocromáticas são geralmente baseadas em duas propriedades básicas dos valores de níveis de cinza: descontinuidade e similaridade. Na primeira categoria, os métodos visam particionar a imagem com base em mudanças bruscas nos níveis de cinza, que caracterizam linhas e bordas na imagem. Na segunda categoria, os métodos procuram agrupar os pixels da imagem que apresentam valores similares para um determinado conjunto de características [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

3.4.1 Limiarização

A limiarização é uma das mais importantes abordagens para a segmentação de imagens. A técnica consiste na classificação dos pixels de uma imagem de acordo com a especificação de um ou mais limiares, e é bastante útil nos casos em que os objetos de interesse contrastam com o fundo [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

A figura 3.9 mostra o histograma de uma imagem onde os objetos a serem segmentados apresentam uma tonalidade bem diferente do fundo da imagem. Neste caso, os níveis de cinza dos pixels do objeto e do fundo estão separados em dois grupos dominantes. Um histograma com tal distribuição é chamado bimodal [Pedrini e Schwartz, 2008].

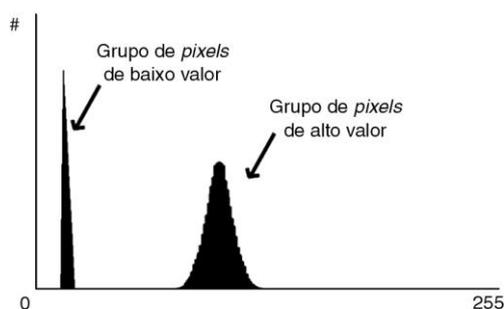


Figura 3.9: Exemplo de um histograma bimodal.
Fonte: Conci, Azevedo e Leta (2008).

Para extrair os objetos do fundo, basta escolher um limiar T que separe os dois grupos. Em seguida, cada pixel da imagem será rotulado como um ponto do objeto se o valor do seu tom

de cinza for maior que T ou como um ponto pertencente ao fundo caso contrário [Gonzalez e Woods, 2000; Pedrini e Schwartz, 2008].

Uma imagem limiarizada $g(x, y)$ pode ser definida como:

$$g(x, y) = \begin{cases} 0, & \text{se } f(x, y) \leq T \\ 1, & \text{se } f(x, y) > T \end{cases} \quad (3.3)$$

onde:

- $f(x, y)$ representa o tom de cinza correspondente a um pixel de entrada;
- T é o limiar escolhido;
- $g(x, y)$ representa a imagem limiarizada resultante.

Os pixels rotulados como 1 (ou qualquer outro nível de cinza conveniente) correspondem aos objetos, enquanto que aqueles rotulados como 0 correspondem ao fundo da imagem. Nesse caso, a limiarização é denominada binarização, devido à imagem resultante possuir apenas dois valores de intensidade [Pedrini e Schwartz, 2008].

3.4.2 Limiarização Otsu

O método de Otsu [1979], é uma abordagem para a seleção automática de um limiar na binarização de imagens em escala de cinza. Este método tem por objetivo a seleção não supervisionada (sem a intervenção do usuário) de um limiar adequado para separar os objetos do fundo da imagem na etapa de segmentação.

Como entrada é requerido apenas o histograma da imagem que se deseja binarizar, não sendo necessário nenhum outro conhecimento prévio sobre a imagem em questão.

O método leva em consideração que o histograma de uma imagem é composto por duas classes, os pixels do objeto e os pixels do fundo. Cada classe possui suas próprias características, ou seja, sua média e desvio padrão. A escolha do limiar ótimo é baseada na análise de discriminante, buscando maximizar a separabilidade das classes resultantes.

A operação de limiarização é considerada como sendo o particionamento dos pixels de uma imagem de L níveis de cinza em duas classes, C_0 e C_1 , que representam o objeto e o fundo, ou vice-versa, sendo que esta partição se dará no nível de cinza k , dessa forma C_0 denota os pixels com níveis $[1, 2, \dots, k]$, e C_1 denota os pixels $[k + 1, k + 2, \dots, L]$. Seja σ_B^2 a

variância entre as classes e σ_T^2 a variância total, um limiar ótimo pode ser obtido encontrando o limiar k que maximize a razão $n(k)$ da variância entre as classes em relação à variância total.

$$n(k) = \frac{\sigma_B^2}{\sigma_T^2} \quad (3.4)$$

Sendo que:

$$\sigma_T^2 = \sum_{i=1}^L (i - \mu_T)^2 p_i \quad (3.5)$$

$$\mu_T = \sum_{i=1}^L i p_i \quad (3.6)$$

$$\sigma_B^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \quad (3.7)$$

$$\omega_0 = \sum_{i=1}^k p_i \quad (3.8)$$

$$\mu_0 = \frac{\mu(k)}{\omega_0} \quad (3.9)$$

$$\mu(k) = \sum_{i=1}^k i p_i \quad (3.10)$$

$$\omega_1 = 1 - \omega_0 \quad (3.11)$$

$$\mu_1 = \frac{\mu_T - \mu(k)}{\omega_1} \quad (3.12)$$

$$p_i = \frac{n_i}{N} \quad (3.13)$$

onde n_i é o número de pixels com nível de cinza i , N é o número total de pixels da imagem, p_i representa a probabilidade de ocorrência do nível de cinza i na imagem.

Para encontrar o limiar ótimo k^* calcula-se então a razão $n(k)$ para todos os valores possíveis de k (ou seja, para todos os níveis de cinza da imagem $[1, 2, \dots, L]$). Então k^* é definido como

$$k^* = \arg \max n(k) \quad (3.14)$$

Este método funciona tanto para o caso ideal, no qual o histograma possui um vale profundo entre dois picos (Figura 3.10-a), representando objetos e fundo respectivamente, e o limiar ótimo pode ser encontrado no fundo deste vale, quanto para os casos em que o vale é plano e extenso, imbuído de ruído (figura 3.10-b), ou quando os dois picos possuem alturas muito diferentes (figura 3.10-c), não sendo possível detectar um vale. Sendo o segundo caso muito comum na maioria das imagens reais.

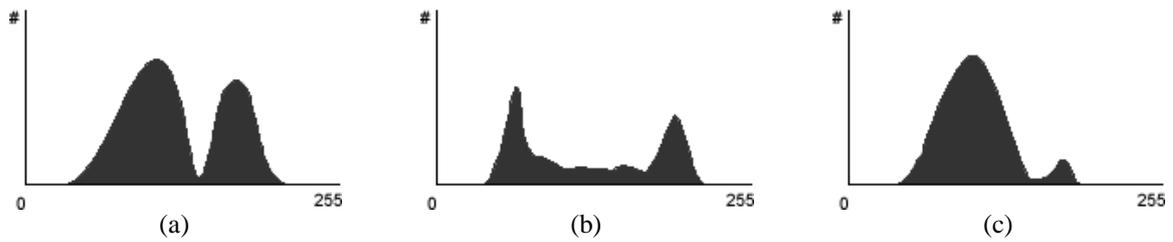


Figura 3.10: Três histogramas com diferentes características. (a) Vale profundo entre dois picos; (b) Vale plano e extenso; (c) Picos com alturas bem distintas.

3.4.3 Histogramas de Projeção

Uma forma interessante de realizar a segmentação de imagens digitais é através dos histogramas de projeção. Esta técnica faz uso da projeção dos pixels da imagem sobre cada uma das duas dimensões existentes (horizontal e vertical) [Silva e Thomé, 2003].

Para uma imagem binária, a projeção horizontal, responsável pela segmentação das linhas, é computada realizando-se a contagem dos pixels com tom de cinza igual a 1 (geralmente tomado como a cor do objeto a ser segmentado) em cada linha da imagem (Equação 3.15), e a projeção vertical é computada realizando-se a mesma contagem em cada coluna da imagem (Equação 3.16) [Pedrini e Schwartz, 2008].

$$HP_h(y) = \sum_{x=1}^M f(x, y) \quad (3.15)$$

$$HP_{vh}(x) = \sum_{y=1}^N f(x, y) \quad (3.16)$$

onde: x representa a linha e y a coluna, M representa a largura e N a altura da imagem.

Uma das aplicações dos histogramas de projeção é na segmentação de caracteres. Silva e Thomé [2003] utilizaram esta técnica para realizar o reconhecimento de caracteres manuscritos – dígitos e letras – provenientes de formulários de concursos. Bockholt [2008] utilizou-a no seu trabalho sobre a recuperação de documentos baseada em conteúdo, e Marcon [2006] utilizou-a no reconhecimento de palavras em textos digitais. Um exemplo desta aplicação é apresentado na Figura 3.11, onde abaixo da imagem está o histograma de projeção vertical e ao lado a projeção horizontal.



Figura 3.11: Aplicação de histogramas de projeção em uma imagem, vertical (abaixo) e horizontal (à direita).

Capítulo 4

Metodologia

Propõe-se aqui um sistema de processamento e análise de imagens baseado nas etapas descritas na seção 3.2. Neste capítulo, são apresentados os métodos utilizados na ferramenta, bem como as decisões de projeto adotadas para cada técnica utilizada.

Para que o sistema processe uma imagem de entrada e reconheça nesta as informações necessárias para a aplicação, essa abordagem deve ser seguida respeitando suas etapas, e mantendo a interligação entre os módulos. Tal estrutura de projeto é composta pelos seguintes passos:

- Aquisição das Imagens
- Pré-processamento
- Segmentação
- Análise
- Geração do Dendograma

A Figura 4.1 mostra o esquema de execução das etapas que a ferramenta utiliza, desde a entrada da imagem a ser analisada, até o resultado final, ou seja, um dendograma representando a similaridade genética entre as amostras (trilhas) presentes na imagem.

As etapas apresentadas em tal figura serão descritas detalhadamente nas próximas seções, informando as decisões de projeto, bem como as técnicas adotadas no desenvolvimento. As seções estão organizadas seguindo a ordem de execução apresentada na mesma.

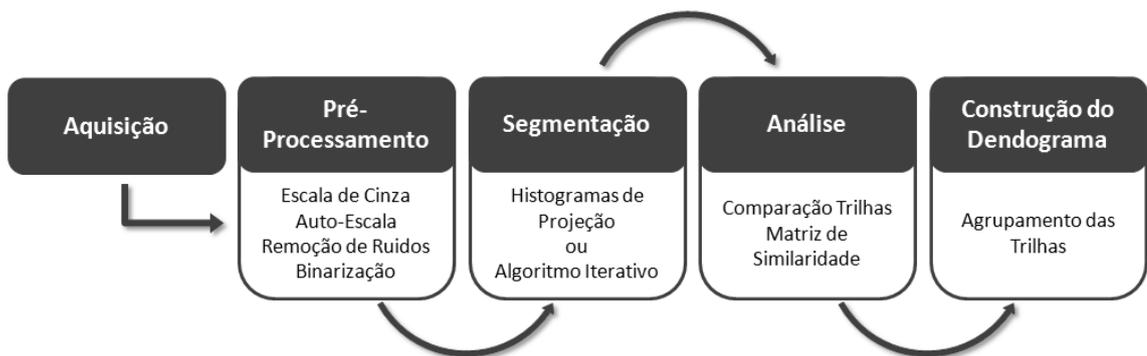


Figura 4.1: Etapas constituintes do sistema.

4.1 Aquisição das Imagens

A etapa de aquisição de imagens não faz parte do escopo do trabalho, ficando a cargo do usuário do sistema obter a imagem digitalizada a ser analisada da maneira que desejar (seja por meio de scanner, câmera fotográfica, etc.).

Para os testes realizados, a fim de verificar a eficiência e a acurácia do sistema, foram utilizadas imagens provenientes da internet, de trabalhos já realizados e com pesquisadores que atuam nas áreas de Biologia Molecular e Genética.

A figura 4.2 mostra uma imagem de eletroforese, a qual será utilizada como imagem de entrada para o sistema para demonstrar as operações realizadas em cada uma das etapas descritas a seguir.

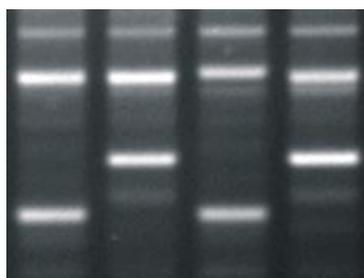


Figura 4.2: Exemplo de uma imagem de eletroforese.
Fonte: Lopes *et al*, (2002).

4.2 Pré-Processamento

Como visto na sessão 3.2, a etapa de pré-processamento tem por objetivo melhorar a qualidade da imagem de entrada, que pode apresentar imperfeições ou degradações

decorrentes da aquisição ou do próprio meio, de forma a aumentar as chances de sucesso nos processos seguintes.

A primeira operação efetuada nesta etapa é a quantização da imagem para 256 tons de cinza. Isto é realizado com intuito de reduzir o número de cores da imagem (de 2^{24} para 2^8), para que operações envolvendo o processamento de histogramas sejam simplificadas. Partindo da afirmação que uma cor é composta de três canais de cores primárias R (vermelho), G (verde) e B (azul), a quantização de uma cor em um ponto (x, y) para seu tom de cinza correspondente é realizada pela Equação (4.1):

$$g(x, y) = 0.114R(x, y) + 0.587G(x, y) + 0.299B(x, y) \quad (4.1)$$

sendo as funções R , G e B correspondentes à intensidade dos canais do pixel (x, y) .

O segundo passo desta etapa é a remoção de ruídos, no qual pode ser utilizado tanto o filtro da média quanto o filtro da mediana, como descrito a seguir:

- Filtro da Mediana: aplicado com finalidade de remover ruídos representados por pontos isolados na imagem. Como valor padrão para o método foi fixado um tamanho de máscara de 3×3 pixels, isso porque um filtro desse tamanho, além de remover pequenos ruídos, preserva as bandas estreitas.
- Filtro da Média: diferentemente do filtro da mediana que remove os ruídos, a aplicação do filtro da média é recomendada quando se deseja suprimir os ruídos em uma imagem, atribuindo assim à imagem um efeito de “borramento”. O valor padrão do tamanho da máscara também é 3×3 pixels.

Um terceiro filtro para remoção de ruídos também foi implementado, sendo este uma combinação do filtro da média com o filtro da mediana. O tamanho da máscara também foi fixado em 3×3 , onde os valores dos níveis de cinza dos nove pixels sob a máscara são ordenados em ordem crescente e então são descartados os dois menores e os dois maiores valores, sobrando os cinco valores centrais. O valor do pixel é então substituído pela média destes cinco valores restantes.

Este filtro tem como vantagens não funcionar apenas para ruído representado por pequenos pontos, como no caso do filtro da mediana, e também não apresenta um borramento tão acentuado como no caso do filtro da média.

Terminada a remoção de ruídos é construído o histograma da imagem. O próximo passo visa aumentar o contraste da imagem, pois em imagens que apresentam baixo contraste

devido à luminosidade mal controlada, pode não ocorrer diferenciação entre o objeto e o fundo. Para tanto é aplicada a função de auto-escala, a qual expande automaticamente a escala de tons de cinza da imagem para que ocupe todo o intervalo possível. Levando em conta que o objetivo é obter uma imagem com 256 (com valores entre 0 e 255) tons de cinza, a função de auto-escala foi implementada utilizando-se a Equação (4.2), que calcula o nível de cinza resultante g para cada pixel f .

$$g = \frac{255}{f_{max} - f_{min}} (f - f_{min}) \quad (4.2)$$

onde f_{max} e f_{min} são, respectivamente, os níveis de cinza máximo e mínimo presentes na imagem original. A escolha da auto-escala para aumento do contraste se deu pelo fato desta não distorcer histogramas bimodais, característica esta que é útil no passo seguinte.

O último passo desta etapa é a limiarização da imagem, que tem como objetivo criar uma dicotomia da imagem, onde os *pixels* pretos representam o fundo da imagem e os *pixels* brancos indicam a presença de material genético.

A escolha do valor do limiar é realizada de forma automática, utilizando para tal o método de Otsu descrito na subseção 3.4.2. Logo após é realizada a limiarização da imagem, conforme descrito na subseção 3.4.1, utilizando o valor de limiar selecionado automaticamente. A imagem resultante é então apresentada ao usuário que, por sua vez, tem a opção de escolher outro valor de limiar caso o resultado da limiarização automática não seja satisfatório. A figura 4.3 demonstra o resultado da aplicação da etapa de pré-processamento na imagem da figura 4.2.

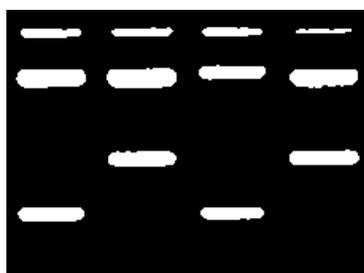


Figura 4.3: Imagem resultante da operação de limiarização.

4.3 Segmentação

O objetivo da etapa de segmentação é extrair os objetos de interesse do restante da imagem. No caso deste trabalho, o objetivo é identificar e separar as trilhas do restante da imagem.

Nesta etapa foram utilizadas duas abordagens diferentes. A primeira com base no método proposto por Brun e Kiel [2010], que utiliza histogramas de projeção (descritos na subseção 3.4.3) para a segmentação das trilhas. A segunda abordagem foi baseada no algoritmo iterativo proposto por Machado *et al.* [1997]. A seguir é descrito detalhadamente o funcionamento de cada uma destas abordagens.

4.3.1 Segmentação utilizando histogramas de projeção

Este método leva em consideração o fato de que as trilhas são geralmente corredores verticais no gel. Para segmentar a imagem aplica-se então o histograma de projeção vertical que, como visto na subseção 3.4.3, consiste na contagem dos pixels brancos (pixels estes que representam o material genético) em cada coluna da imagem.

A figura 4.4-a demonstra o resultado da aplicação do histograma de projeção vertical na imagem da figura 4.2. Observando o histograma de projeção gerado, pode-se perceber que este é formado por “picos”, que correspondem às trilhas, e “vales” que por sua vez correspondem às regiões sem material genético entre as trilhas. Desta forma identificando os pontos iniciais e finais de cada “pico” tem-se a localização de cada trilha. Para isto, basta identificar os pontos de descontinuidade na projeção.

A presença de ruído na imagem pode resultar na contagem dos pixels nas regiões de vale maior que zero, por isso, é necessário estabelecer um limiar para fazer o descarte destas colunas. O valor deste limiar depende da quantidade de material genético e também da quantidade de ruído presente na imagem, características estas que geralmente variam de uma imagem para outra. Desta forma não é possível estabelecer um valor de limiar fixo para todas as imagens. Tendo este problema em vista, foi criada uma forma de calcular este valor automaticamente, utilizando para tanto o próprio histograma de projeção da imagem. Para isto, primeiramente é encontrada a média de todos os valores do histograma de projeção, e esta é multiplicada pela constante 0,22. O valor desta constante foi definido empiricamente, através da análise de várias imagens com características distintas, onde se concluiu que valores menores podem fazer com que ruído não seja descartado e valores maiores podem

fazer com que trilhas com pouco material genético sejam descartadas ou regiões de vale não sejam detectadas.

Na figura 4.4-b é apresentada a imagem original com as trilhas detectadas com a utilização do método dos histogramas de projeção para a segmentação.

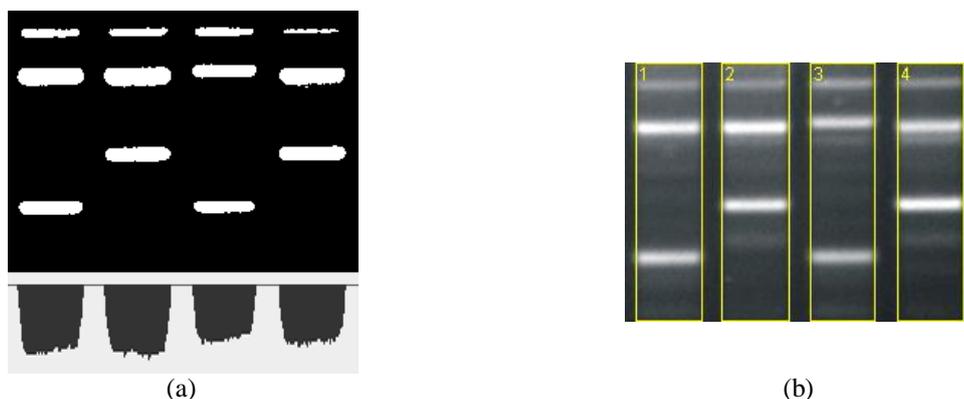


Figura 4.4: (a) Histograma de projeção vertical; (b) Trilhas detectadas.

4.3.2 Segmentação utilizando o algoritmo iterativo

Este método também parte do princípio de que as trilhas são geralmente corredores verticais no gel. Além disso considera que a distância entre todas as trilhas da imagem é a mesma, assim como a largura. Um problema surge quando alguma das cavidades no gel não recebe material genético, o que resulta na falta de uma trilha, interrompendo assim a regularidade do padrão do gel.

A partir da imagem pré-processada é realizada a projeção vertical da imagem em escala de cinza, que consiste na soma dos valores dos tons de cinza de cada coluna. O princípio é o mesmo dos histogramas de projeção descritos na subseção 3.4.3, no entanto neste capítulo será denotado como projeção para diferenciar dos histogramas de projeção computados utilizando-se imagens binárias.

A projeção possui muitos máximos locais que podem indicar a presença de trilhas na imagem. Para detectar todas as trilhas corretamente é necessário reduzir o número de máximos locais (gerados pelas altas-frequências) presentes na projeção. Para isto suaviza-se a projeção aplicando o filtro da média móvel de largura três.

Em geral uma única filtragem (ou suavização da projeção) não é suficiente para reduzir o número de máximos locais ao desejado, isto é, não é o suficiente para detectar as trilhas corretamente. Sendo assim, o algoritmo mostrado na figura 4.5 é utilizado para decidir quantas vezes aplicar o filtro de suavização.

1. *Filtre a projeção com o filtro especificado.*
2. *Encontre os máximos locais.*
3. *Ordene as distâncias entre cada par de máximos locais vizinhos e encontre a mediana.*
4. *Se existir alguma distância que não seja um múltiplo inteiro dessa mediana*
 Volte ao passo 1
 Senão
 Finalizar.

Figura 4.5: Algoritmo Iterativo para suavização da projeção.

Este algoritmo é baseado no fato de que as cavidades no gel feitas por um pente, onde são depositadas as amostras, são equidistantes. Dessa forma, se alguma das trilhas não recebeu material genético, a distância entre as trilhas anterior e posterior será o dobro do valor da mediana. No entanto, essa relação inteira entre as distâncias e a mediana deve ter alguma tolerância.

O funcionamento do algoritmo iterativo é ilustrado na figura 4.6. A imagem original é mostrada na figura 4.6-a, enquanto que a figura 4.6-b apresenta a projeção da imagem e os máximos locais encontrados inicialmente. O resultado da aplicação do algoritmo é mostrado na figura 4.6-c, onde pode ser observada a projeção suavizada e os centros das trilhas localizadas.

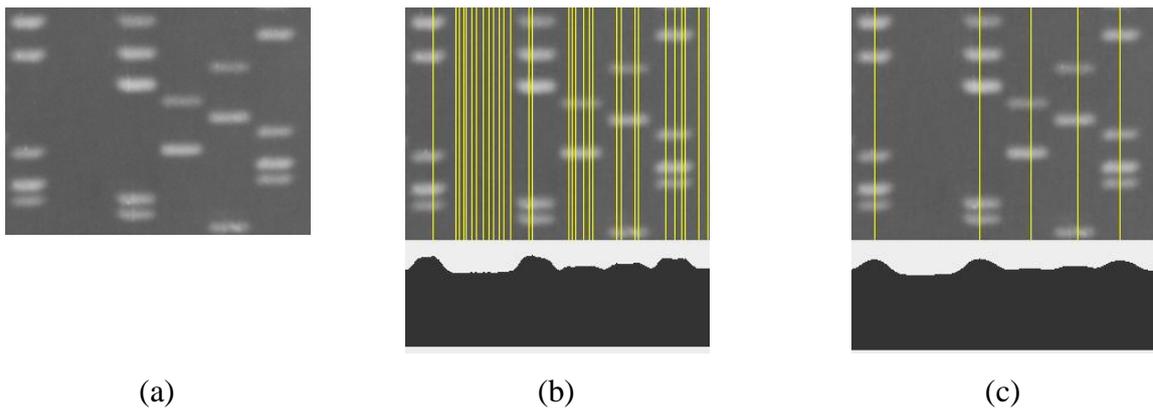


Figura 4.6: (a) Imagem original; (b) Máximos locais; (c) Resultado da aplicação do filtro iterativo.

Algumas vezes a posição correta do centro da trilha na imagem não corresponde exatamente à posição do máximo detectado no processo de filtragem. Isso pode ocorrer tanto por efeito do filtro quanto pela aparência do gel, que pode apresentar diferentes concentrações de material, levando a projeções não simétricas em relação ao centro da trilha. Três algoritmos

alternativos foram propostos por Machado *et al.* [1997] para detectar a posição correta do centro da trilha.

A partir da projeção suavizada f , encontra-se seus máximos e mínimos na forma

$$a_0, b_1, a_1, b_2, \dots, b_n, a_n \quad (5.2)$$

onde b_i é a coordenada x do i -ésimo máximo, que é precedido e sucedido por dois mínimos respectivamente denotados por a_{i-1} e a_i . Na figura 4.7 é mostrado o primeiro algoritmo para calcular o centro de cada trilha com base nesta projeção.

1. *Encontre os valores dos n máximos b e $n + 1$ mínimos a .*
2. *Para i de 1 até n faça*
 - *Calcule a área entre dois mínimos consecutivos $A_i = \sum_{j=a_{i-1}}^{a_i} f_j$*
 - *Calcule o primeiro momento entre dois mínimos consecutivos $m_i = \sum_{j=a_{i-1}}^{a_i} j f_j$*
 - $c_i = \frac{m_i}{A_i}$ *(c_i será a coordenada x da trilha i)*

Figura 4.7: Algoritmo 1

Este algoritmo tem a vantagem de considerar toda a distribuição do material ao invés de somente o máximo. Porém, em imagens com trilhas sem material, a posição calculada pode cair fora da canaleta, em virtude da área da trilha vazia também ser computada. Outro algoritmo para calcular a posição correta das trilhas é apresentado na figura 4.8.

1. *Encontre os valores dos n máximos b e $n + 1$ mínimos a .*
2. *Para i de 1 até n faça*

$$c_i = ((a_{i-1} + a_i)/2 + b_i)/2$$

Figura 4.8: Algoritmo 2

Este algoritmo tem a vantagem de levar em conta a simetria entre os dois mínimos e o máximo de cada trilha. Possui também a mesma desvantagem do anterior no caso de trilhas sem material.

O algoritmo 3, mostrado na figura 4.9, visa melhorar a identificação de trilhas em imagens que apresentam ausência de material em algumas delas.

Este algoritmo calcula o centro da área compreendida simetricamente entre os máximos. Com isso, elimina-se o problema de detecção da posição das trilhas vizinhas a uma trilha sem material genético.

1. *Encontre os valores dos n máximos b e $n+1$ mínimos de a .*
2. *Para i de 1 até n faça*
 - *Encontre o mínimo d_i mais próximo de b_i fazendo $d_i = \min(b_i - a_{i-1}, a_i - b_i)$.*
 - *Calcule a área entre o segmento que vai de $(b_i - d_i, f(b_i - d_i))$ até $(b_i + d_i, f(b_i + d_i))$ e chame-a de A_i .*
 - *Calcule o primeiro momento da área considerada no passo anterior e chame-a de m_i .*
 - $c_i = \frac{m_i}{A_i}$

Figura 4.9: Algoritmo 3

4.4 Análise

Após a segmentação das trilhas é realizada a comparação entre as canaletas localizadas. Cada trilha é comparada com todas as outras utilizando a comparação simples de imagens. A comparação é feita entre duas trilhas de cada vez, onde cada pixel (x, y) da primeira trilha é comparado com o pixel (x, y) da segunda trilha: caso os dois pixels possuam o mesmo valor o índice de similaridade entre essas duas trilhas é acrescido em uma unidade.

Para efetuar esta comparação é importante que todas as trilhas possuam a mesma largura, portanto, é necessário encontrar a trilha com a menor largura dentre as trilhas detectadas e então “podar” as laterais das trilhas maiores para que todas tenham a mesma largura da trilha mais estreita.

O resultado desta etapa é a geração de uma matriz de similaridade, a qual possui o número de pixels coincidentes para cada par de trilhas. A figura 4.10 apresentada a matriz de similaridade gerada com base nas trilhas da imagem da figura 4.2 localizadas na etapa anterior. A partir da matriz de similaridade é construído então o dendograma.

	1	2	3
2	6461		
3	6825	6258	
4	6371	7040	6220

Figura 4.10: Matriz de similaridade.

4.5 Geração do Dendograma

O dendograma é uma representação gráfica em forma de árvore que expressa a estrutura dos agrupamentos. Nos métodos de agrupamentos de dados hierárquicos aglomerativos, o dendograma representa a ordem em que os dados foram agrupados [Vale, 2005].

A construção do dendograma é realizada utilizando-se o algoritmo apresentado na figura 4.11:

1. *Selecionar na matriz os dois elementos com maior similaridade.*
2. *Se os dois elementos ainda não foram marcados*
Criar um novo grupo com estes dois elementos.
Senão
Se apenas um dos dois elementos já foi marcado
Vincular o elemento não marcado ao grupo do outro elemento.
Senão
Se os dois elementos não estão no mesmo grupo
Juntar os dois grupos que contém cada um dos elementos.
Senão
Descartar o valor de similaridade.

Figura 4.11: Algoritmo para construção do dendograma.

Este algoritmo é aplicado iterativamente para todos os elementos da matriz de similaridade. A figura 4.12 apresenta o dendograma resultante da matriz da figura 4.10, representado assim a similaridade entre os elementos da imagem de eletroforese da figura 4.2.

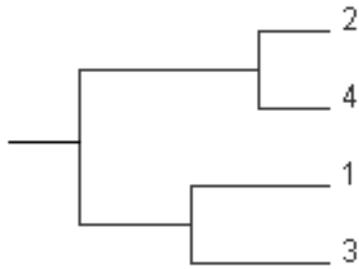


Figura 4.12: Dendograma gerado com base na matriz de similaridade da figura 4.6.

Analisando o dendograma construído observa-se que os elementos 2 e 4 possuem a maior semelhança (conforme mostrado na matriz de similaridade). Os elementos 1 e 3 apresentam a segunda maior similaridade e, por fim, os dois grupamentos separados constituem um único grupamento.

Interessante destacar que quanto maior a similaridade entre os elementos e grupamentos, mais próxima será sua junção. Este fato pode ser notado comparando-se a junção dos elementos 2 e 4 com a ligação entre 1 e 3.

Capítulo 5

Resultados e Discussão

5.1 A Ferramenta Construída

A ferramenta foi implementada na linguagem de programação JAVA utilizando a IDE de desenvolvimento Netbeans 6.9.1. A escolha desta linguagem deve-se ao fato de ser uma linguagem moderna baseada no paradigma de orientação a objetos e portátil (independência de plataforma), possibilitando a execução independente de sistema operacional, necessitando apenas da instalação da Máquina Virtual Java (JVM).

Foi utilizada também para a implementação a Interface de Programação de Aplicações JAI (Java Advanced Imaging) visto que esta oferece uma série de recursos de processamento de imagens já implementados. Esta API também facilita a manipulação de imagens em aplicações Java por possuir um formato interno padrão para as imagens carregadas em memória, independente do formato do arquivo da imagem em disco. Além disso, os métodos implementados possuem alto desempenho. Os métodos utilizados que são implementados pela API são:

- Codificadores e decodificadores para os formatos de arquivos BMP, JPEG, PNG e TIFF;
- Quantização para tons de Cinza;
- Geração do Histograma;
- Inversão da Imagem;
- Limiarização Global;
- Filtro da Média;
- Filtro da Mediana;

Os outros métodos, por seu uso não ser tão comum, não estão presentes na API e, portanto foram implementados. São eles:

- Histogramas de Projeção;
- Auto-Escala;
- Filtro combinando Média e Mediana;
- Método de limiarização Otsu.

Com o intuito de apresentar a ferramenta construída, serão apresentadas imagens colhidas do sistema em execução. A tela principal é apresentada na figura 5.1.

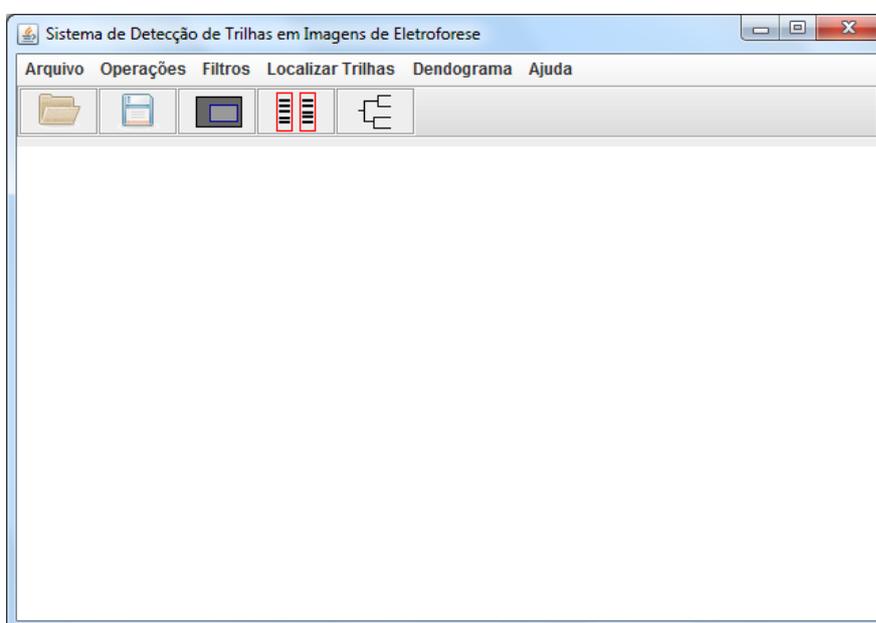


Figura 5.1: Tela principal da ferramenta.

A tela principal possui uma barra de menus que permite o acesso a todas as funcionalidades da ferramenta, sendo elas: opções de abrir e salvar uma imagem (Figura 5.2-a), operações de pré-processamento (Figura 5.2-b), filtros para remoção de ruído (Figura 5.2-c), localização das trilhas utilizando o método dos histogramas de projeção ou o método do algoritmo iterativo (Figura 5.2-d) e geração do histograma (Figura 5.2-e). Esta tela também possui uma barra com botões, em forma de ícones, que são uma forma rápida de acessar as principais funções do sistema, facilitando assim a utilização da ferramenta. A função de cada botão é respectivamente (da esquerda para a direita): abrir uma imagem, salvar uma imagem, selecionar uma região de interesse na imagem para ser analisada, detectar as trilhas da

imagem automaticamente (utilizando o método dos histogramas de projeção) e gerar o dendograma com base nas trilhas detectadas.

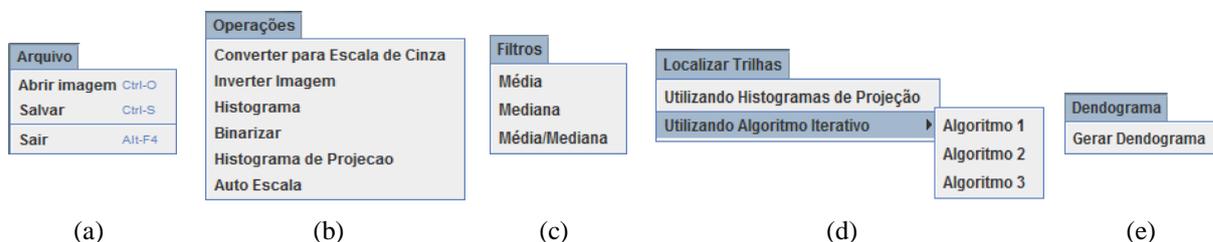


Figura 5.2: Menus com todas as funcionalidades da ferramenta.

A figura 5.3 apresenta a tela principal do sistema com uma imagem de eletroforese carregada, a qual será utilizada para demonstrar sequencialmente o funcionamento do sistema. O usuário pode utilizar as operações de pré-processamento combinadas da forma que desejar para então aplicar um dos dois métodos de detecção das trilhas, ou pode simplesmente clicar no botão “Localizar trilhas automaticamente”, o qual está destacado na figura 5.3, para que o sistema realize o pré-processamento e a detecção automaticamente.

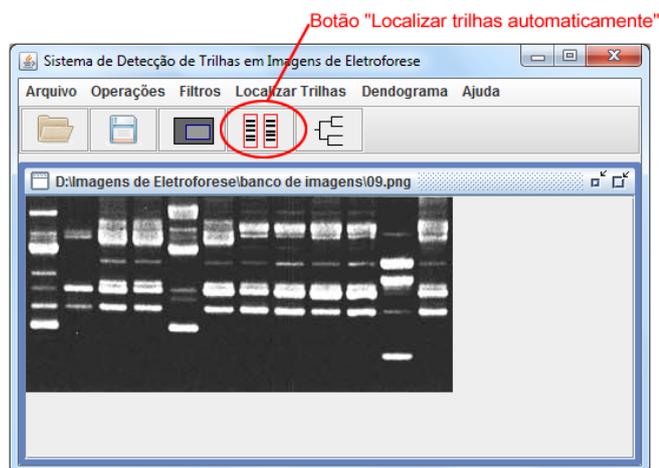


Figura 5.3: Tela principal do sistema com uma imagem carregada.

Quando o usuário optar por fazer a localização das trilhas automaticamente o sistema solicitará que ele informe a característica da imagem (Figura 5.4), ou seja, se a mesma apresenta material claro sobre fundo escuro ou vice-versa. Caso o usuário informe que a imagem apresenta material escuro sobre um fundo claro, esta será invertida antes de iniciar o pré-processamento.

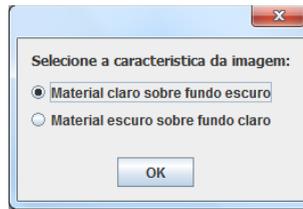


Figura 5.4: Janela apresentada ao usuário para a escolha da característica da imagem.

Após a seleção da característica da imagem o sistema realiza automaticamente as seguintes operações sobre ela: quantização para tons de cinza, inversão da imagem (caso necessário), auto-escala do histograma e suavização utilizando o filtro média/mediana.

Terminada a fase de pré-processamento o sistema realiza a binarização automática da imagem utilizando o método de Otsu e o resultado é apresentado ao usuário (Figura 5.5), oferecendo a opção de selecionar outro valor de limiar, através do controle deslizante, que melhor reflita a separação entre o material genético e o fundo da imagem.

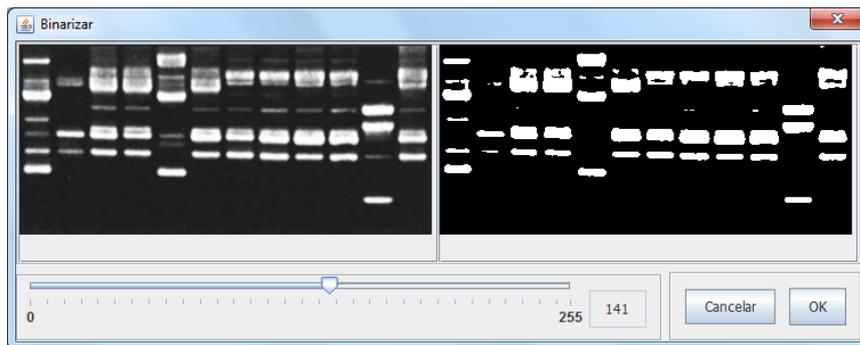


Figura 5.5: Janela apresentando o resultado da binarização da imagem.

Depois de o usuário clicar no botão "OK" o sistema realiza a segmentação das trilhas utilizando o método dos histogramas de projeção. O resultado é apresentado em uma nova janela, com as trilhas desenhadas sobre a imagem original como mostrado na figura 5.6.

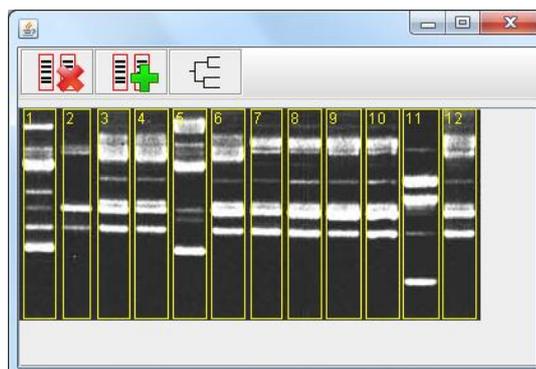


Figura 5.6: Janela mostrando as trilhas detectadas.

Nesta janela o usuário tem a opção de remover trilhas que eventualmente foram detectadas incorretamente. Para tanto, basta clicar sobre a trilha que se deseja excluir e em seguida clicar no primeiro botão da barra de ferramentas. Também é oferecida ao usuário a opção de adicionar trilhas que não tenham sido detectadas. Para isto, o usuário clica sobre a imagem na posição onde deseja adicionar a nova trilha e arrasta o mouse formando um retângulo que corresponda à largura da trilha; na sequência basta clicar no segundo botão da barra de ferramentas.

Finalizada a detecção das trilhas, o usuário tem a opção de gerar o dendograma. Clicando no botão para geração do mesmo (terceiro botão da barra de ferramentas) o sistema realiza a sua construção e exibe o resultado ao usuário. A figura 5.7 apresenta o dendograma gerado que representa a similaridade entre as trilhas detectadas na imagem de eletroforese da Figura 5.6.

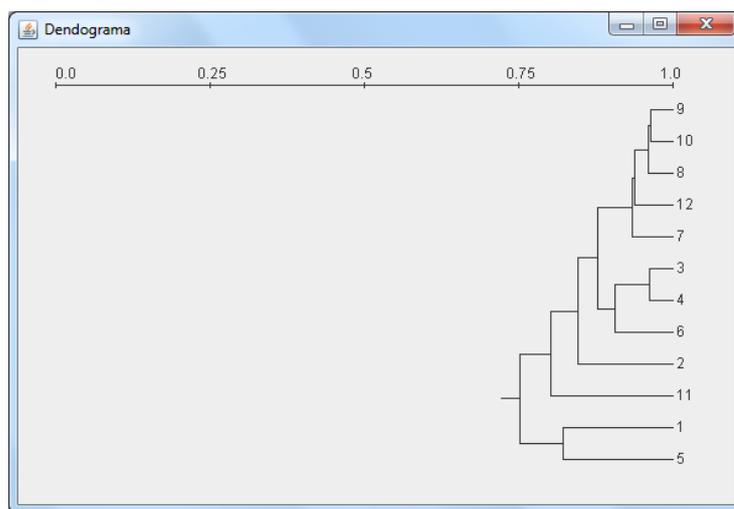


Figura 5.7: Dendograma.

5.2 Características dos Testes Realizados

Utilizando a ferramenta desenvolvida, foram realizados testes com o objetivo de comparar os dois métodos de segmentação de imagens de eletroforese implementados, que são segmentação utilizando histogramas de projeção e segmentação utilizando o algoritmo iterativo, os quais foram apresentados nas subseções 4.3.1 e 4.3.2 respectivamente.

Os testes foram realizados utilizando um conjunto de 52 imagens provenientes da internet, de trabalhos já realizados e de várias pesquisas diferentes fornecidas por pesquisadores que

atuam nas áreas de Biologia Molecular e Genética. Estas imagens possuíam tamanhos variados, além de diferirem na qualidade e no número de trilhas, constituindo assim um conjunto de imagens bastante heterogêneo.

Como forma de mensurar os resultados foi contabilizado o número de trilhas detectadas corretamente em cada imagem. Foram consideradas como trilhas detectadas corretamente apenas aquelas nas quais os limites iniciais e finais estavam bem posicionados, não considerando trilhas detectadas parcialmente.

Com o objetivo de verificar se as operações de pré-processamento aplicadas visando melhorar a qualidade da imagem e, desta forma, aumentar a chances de sucesso na detecção das trilhas, realmente trazem benefícios, foi realizada uma bateria de testes com as imagens sem pré-processamento e outra com pré-processamento. As operações que visam melhorar a qualidade da imagem são a auto-escala e a remoção de ruídos, sendo assim, na primeira bateria elas não foram aplicadas nas imagens.

Na etapa de segmentação, os testes foram realizados com os parâmetros automáticos em todas as imagens, sem alteração manual para se adequar melhor às características de cada imagem, sendo assim a única entrada manual era para informar ao sistema se a imagem possuía material claro sobre fundo escuro ou vice-versa.

No caso do método que utiliza histogramas de projeção, o valor do limiar aplicado na binarização foi sempre aquele calculado pelo método Otsu. No método do algoritmo iterativo, existe o parâmetro que define a tolerância da relação inteira entre as distâncias e a mediana. O valor desta tolerância depende da largura das trilhas, e como existia diferença nesta largura de uma imagem para outra foi necessário criar uma forma de calcular este valor de forma automática para cada imagem. Para tal utiliza-se o próprio valor da mediana, que o algoritmo acredita ser a largura aproximada de todas as trilhas, multiplicando este valor pela constante 0,18. O valor desta constante foi definido empiricamente através de testes com várias imagens que possuíam diferentes características, onde chegou-se a conclusão que este valor apresentava os melhores resultados para a maioria das imagens.

Como algumas das imagens utilizadas nos testes possuíam trilhas sem material genético, no método do algoritmo iterativo foi utilizado a 3ª abordagem para detectar o centro de cada trilha e também as coordenadas inicial e final de cada uma, visto que este algoritmo contorna o problema da trilha “vazia”.

Para a execução dos testes foi utilizado um computador com processador Intel® Core™2 Duo 2.4GHz e 4 Gigabytes de memória RAM. O sistema operacional utilizado foi o Windows 7 Home Premium e a Máquina Virtual Java versão 6.

5.3 Resultados dos Testes

Nesta seção são apresentados os resultados obtidos através dos testes realizados. A tabela 5.1 mostra a quantidade de trilhas presentes em cada imagem, a quantidade de trilhas detectadas corretamente com a aplicação de cada um dos métodos de segmentação, com pré-processamento e também sem pré-processamento, além da quantidade de trilhas detectadas incorretamente. Os cabeçalhos NTDC e NTDI são as abreviaturas de número de trilhas detectadas corretamente e número de trilhas detectadas incorretamente, respectivamente.

Tabela 5.1: Resultado dos testes comparativos.

id da imagem	Nº de Trilhas	Método Histogramas de Projeção				Método Algoritmo Iterativo			
		Sem pré-processamento		Com pré-processamento		Sem pré-processamento		Com pré-processamento	
		NTDC	NTDI	NTDC	NTDI	NTDC	NTDI	NTDC	NTDI
1	5	5	0	5	0	4	1	4	1
2	5	5	0	5	0	5	0	5	0
3	7	7	0	7	0	2	0	2	0
4	6	6	0	6	0	5	2	5	1
5	8	8	0	8	1	0	16	0	16
6	8	8	0	8	0	7	2	7	2
7	12	10	0	10	0	12	0	12	0
8	8	0	0	0	1	0	2	1	2
9	12	12	0	12	0	12	0	12	0
10	6	6	0	6	1	5	2	5	2
11	17	17	0	17	0	17	0	17	0
12	7	7	0	7	0	5	4	7	0
13	7	7	0	7	0	6	2	6	2
14	9	7	1	6	1	9	0	9	0
15	8	8	0	8	0	0	16	7	2
16	4	4	0	4	0	0	8	0	8
17	7	7	0	7	0	6	0	6	0
18	7	7	0	7	0	5	5	5	5
19	7	7	0	7	0	6	0	6	0
20	16	12	4	12	3	0	5	0	5
21	10	9	0	9	0	10	0	10	0
22	6	5	6	5	2	5	2	5	2

23	35	4	1	4	1	35	0	35	0
24	13	12	1	11	2	0	4	0	4
25	8	8	0	8	0	2	1	2	1
26	13	11	3	11	3	0	2	0	2
27	5	5	0	5	0	5	3	5	3
28	7	7	0	7	0	0	14	0	14
29	5	5	0	5	0	0	10	0	10
30	10	10	0	10	0	9	2	9	2
31	10	10	0	10	0	10	0	10	0
32	10	10	0	10	0	10	0	10	0
33	13	13	0	13	0	13	0	13	0
34	8	8	0	8	0	7	2	7	2
35	17	17	1	17	0	17	0	17	1
36	6	5	3	6	1	0	2	5	2
37	4	4	0	4	0	3	2	3	2
38	10	9	2	10	0	1	4	1	4
39	7	7	0	7	0	5	0	5	0
40	5	5	0	5	0	5	0	5	0
41	8	8	0	8	0	8	0	8	0
42	8	6	1	8	0	8	0	8	0
43	6	6	0	6	0	1	10	1	10
44	7	7	0	7	0	5	0	5	0
45	12	10	4	10	1	0	3	0	3
46	17	15	1	15	2	0	2	0	2
47	12	12	0	12	0	12	0	12	0
48	5	5	0	5	0	4	2	4	2
49	5	5	0	5	0	5	0	5	0
50	7	7	0	7	1	6	2	6	2
51	4	4	0	4	0	3	2	3	2
52	4	3	0	4	0	0	8	0	8
Total:	463	402	28	405	20	295	142	310	124
		86,83%		87,47%		63,71%		66,95%	

Analisando os resultados obtidos pode-se observar que o método dos histogramas de projeção se mostrou mais eficaz. Nas imagens sem pré-processamento este método detectou corretamente 402 trilhas de um total de 463, representando 86,83% de acerto, e apenas 28 trilhas foram detectadas incorretamente. Nas imagens com pré-processamento este método detectou corretamente 405 trilhas, representando 87,47% de acerto, e apenas 20 trilhas foram detectadas incorretamente. Embora não muito significativa, houve uma melhora nos resultados das imagens com pré-processamento.

O método do algoritmo iterativo para as imagens sem pré-processamento, detectou corretamente 295 trilhas de um total de 463, representando 63,71% de acerto, e 142 trilhas foram detectadas incorretamente. Nas imagens com pré-processamento este método detectou corretamente 310 trilhas, representando 66,95% de acerto, e 124 trilhas foram detectadas incorretamente. Pode-se observar que houve uma melhora significativa tanto no número de acertos quando no número de erros para as imagens que passaram pela etapa de pré-processamento.

Ao calcular a porcentagem de trilhas detectadas corretamente em cada imagem individualmente e extrair a média aritmética simples destes valores, obteve-se os valores de 92,17% para imagens sem pré-processamento e 93,28 % com o pré-processamento, para o método dos histogramas de projeção, com desvio padrão de 19,2% para ambos. Para o método do algoritmo iterativo estas médias foram de 62,19% e 66,26%, com desvio padrão de 41,7% e 39,9%, respectivamente. Estes valores de desvio padrão mais altos para o segundo método mostram que o método dos histogramas de projeção foi mais estável. Indica também que o método apresentado por Machado *et al.* (1997) consegue detectar eficientemente as trilhas em algumas imagens e em outras, esta eficiência cai consideravelmente.

Para os dois métodos de segmentação, o resultado final foi sempre apresentado ao usuário em menos de 1 segundo, que é um tempo perfeitamente aceitável para o propósito desta aplicação. Isto mostra que ambos os métodos possuem consumo de processamento aceitável.

Analisando as imagens que não obtiveram bons resultados nos testes utilizando o método dos histogramas de projeção (imagens 8, 14, 20, 22 e 23), foi constatado que todas elas apresentavam iluminação muito irregular, dificultando assim o processo de limiarização, e/ou as bandas de uma trilha invadiam a região de outra trilha, não sendo possível determinar corretamente a divisa entre as duas trilhas.

Quanto ao método do algoritmo iterativo, pode-se constatar através dos testes que é muito difícil estabelecer um valor de tolerância para o critério de parada do algoritmo, quando aplicado em imagens onde a largura das trilhas não é a mesma em todas elas.

O método dos histogramas de projeção também foi capaz de determinar com maior precisão as coordenadas inicial e final de cada trilha, fato este que faz com que a comparação entre as trilhas e a geração do dendograma sejam mais precisas.

Capítulo 7

Conclusões

Com a realização deste trabalho verificou-se que a computação pode ser uma importante aliada ao estudo biológico, inclusive à genética de populações. O desenvolvimento de ferramentas adequadas propicia não só agilidade nos processos comparativos, tornando o processo menos oneroso, mas também minimizam o risco de erros de interpretação. Além disso, a ferramenta permite a avaliação de muitos parâmetros que geralmente são ignorados por analistas humanos.

Na comparação entre os dois métodos de detecção de trilhas, pode-se constatar que o método dos histogramas de projeção mostrou ser mais vantajoso em relação ao método do algoritmo iterativo nos testes realizados. O primeiro método conseguiu identificar um número significativamente maior de trilhas corretamente e detectou poucas trilhas de forma incorreta. Verificou-se que o principal desafio do segundo método é estimar o valor da tolerância para a relação inteira entre as distâncias e a mediana.

A utilização do método de Otsu para seleção automática do valor do limiar na operação de binarização das imagens de eletroforese apresentou bons resultados. Na maioria das imagens o limiar calculado pelo método era aquele que realmente resultava na melhor separação entre o fundo da imagem e o material genético.

O software construído mostrou-se bastante eficiente em termos de consumo de tempo para a detecção e interpretação das trilhas, bem como na construção do dendograma das mesmas.

Visando a obtenção de resultados melhores poderiam ser estudadas e implementadas, futuramente, técnicas de limiarização local das imagens ao invés da limiarização global, buscando melhorar o desempenho em imagens com iluminação irregular. Outras técnicas para remoção de ruído também poderiam ser estudadas, como por exemplo, as operações de abertura e fechamento da morfologia matemática. Além disso, poderia ser estudada e testada

uma forma de descartar automaticamente as trilhas detectadas incorretamente com base na média das larguras.

Alguma forma para tentar identificar automaticamente o tipo da imagem, ou seja, se a mesma apresenta material genético claro sobre fundo escuro ou material escuro sobre fundo claro, também poderia ser estudada futuramente. Como as imagens de eletroforese geralmente possuem uma área pequena ocupada pelo material genético em relação à área ocupada pelo fundo, uma idéia seria a aplicação do método de Otsu e a verificação de qual das duas classes resultantes possui o menor número de pixels.

Referências Bibliográficas

ARIAS, M. C.; INFANTE-MALACHIAS, M. E. *Biologia Molecular e Evolução*, capítulo: RFLP: O emprego de enzimas de restrição para detecção de polimorfismos do DNA, p. 143-152, Holos Editora, Ribeirão Preto, 1. ed., 2001.

ASSUNÇÃO, N. S. *Divergência genética em Cedrela fissilis Vell. obtida por marcadores RAPD*. Monografia (Monografia de Graduação) – Universidade Federal de Lavras, Lavras – MG, Junho 2004.

BOCKHOLT T. C. *Recuperação de documentos baseada em conteúdo*. Monografia (Monografia de Graduação) – Escola Politécnica de Pernambuco - Universidade de Pernambuco, Recife, Dezembro 2008.

BORED, F.; NETO, J. F. B.; CARVALHO, F. I. F. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. *Ciência Rural*, Santa Maria, v. 27, n. 3, p. 513-520, 1997.

BRUN, A. L.; KIEL, G. Ferramenta computacional para determinação da similaridade de espécies em imagens de eletroforese. In: *II Simpósio de Computação Aplicada (SCA)*. Passo Fundo: [s.n], 2010.

CONCI, A.; AZEVEDO, E.; LETA, F. R. *Computação Gráfica [v.2]*. 1. ed. Rio de Janeiro: Elsevier, 2008.

CORRÊA, E. M.; POSSIK P. A. *A análise de DNA por eletroforese*. Capítulo do livro *Eletroforeses: Hemoglobinopatias, Proteínas Séricas, Lipoproteínas e DNA*, que está sendo editado pelo Grupo Editorial Nacional (GEN), com lançamento previsto para 2011.

CRUZ, C. D. . *Programa Genes - Diversidade Genética*. 1. ed. Viçosa, MG: Editora UFV, 2008. v. 1. 278 p.

CRUZ, C. D.; SCHUSTER I. *Programa GQMOL – Genética quantitativa e molecular*. Viçosa: Universidade Federal de Viçosa, 2008. Consultado na internet: <http://www.ufv.br/dbg/gqmol/gqmol.htm>, em 05/08/2011.

FALEIRO, F. G. *Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos*. 1. ed. Planaltina, DF: Embrapa Cerrados, 2007.

FERREIRA, M. E.; GRATTAPAGLIA, D. *Introdução ao uso de marcadores moleculares em análise genética*. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998.

FILHO, A. M.; REMUALDO, V. *Biologia Molecular Aplicada ao Diagnóstico de Doenças Infecciosas. Prática Hospitalar*, Ano IX, n. 53, p. 33-35, 2007.

GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. 1. ed. São Paulo: Edgard Blüncher Ltda, 2000.

GRUTZMACHER D. D. *et al.* Variabilidade genética interespecífica em formigas cortadeiras do gênero *Acromyrmex* que ocorrem no Estado do Rio Grande do Sul. *Ciência Rural*, Santa Maria, v. 37, n. 4, p. 921-927, jul-ago, 2007.

IBM CORPORATION. *SPSS Advanced Statistics*, 2010. Consultado na internet: <http://www-01.ibm.com/software/analytics/spss/products/statistics/advanced-statistics/>, em 05/08/2011.

LOPES, R. Marcadores moleculares dominantes (Rapl e Aflp). *Biotecnologia Ciência e Desenvolvimento*. Brasília, Ano V, Número 29, p. 56-60, nov/dez, 2002.

MACHADO, A. M. C. *et al.* An iterative algorithm for segmenting lanes in gel electrophoresis images. In: *XI Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens – SIBGRAPI*. Campos do Jordão: [s.n.], 1997.

MARCON, M. *Reconhecimento de palavras em textos digitais*. Monografia (Monografia de Graduação) – Universidade Estadual do Oeste do Paraná, Cascavel , Dezembro 2006.

MATIOLI, S. R.; PASSOS-BUENO, M. R. S. *Biologia Molecular e Evolução*, capítulo: Métodos baseados em PCR para análise de polimorfismos de ácidos nucleicos, p. 153-161, Holos Editora, Ribeirão Preto, 1. ed., 2001.

MILACH, S. C. K. Marcadores de DNA. *Biotecnologia Ciência e Desenvolvimento*. Brasília, Ano II, Número 9, p. 40-43, jul/ago, 1999.

NAOUM, P. C. *Eletroforese – Técnicas e Diagnósticos*. 2. ed. São Paulo: Livraria Santos Editora, 1990.

OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, n. 1, p. 62-66, 1979.

PASSAGLIA, L. M. P.; ZAHA, A. *Biologia Molecular Básica*, capítulo: Técnicas de Biologia Molecular, p. 379–413, Mercado Aberto, Porto Alegre, 3. ed. , 2003.

PAULINO, A. A. *et al.*. Detecção de canaletas em imagens de géis de eletroforese usando filtro de kalman. In: *XXX Congresso Nacional de Matemática Aplicada e Computacional – CNMAC*. Florianópolis: [s.n.], 2007.

PEDRINI, H.; SCHWARTZ, W. R. *Análise de imagens digitais: Princípios, algoritmos e aplicações*. 1.ed. São Paulo: Thonson Learning, 2008.

REIS JUNIOR F. B et al. *Uso de ferramentas moleculares em estudos da diversidade de microrganismos do solo*. Planaltina: Embrapa Cerrados, Dezembro, 2002. Documentos 51.

ROHLF, F. J. *NTSYSpc – Numerical taxonomy and multivariate analysis system*. v 2.0. New York: 1998.

ROMANO, E.; BRASILEIRO, A. C. M. Extração de DNA de plantas. *Biotecnologia Ciência e Desenvolvimento*. Brasília, Ano I, Número 5, p. 14-17, Abril, 1998.

SARTORETTO, L. M.; FARIAS, P. C. Diversidade genética e técnicas biotecnológicas. *Unoesc & Ciências – ACET*, Joaçaba, v. 1, n. 2, p. 155-162, jul./dez. 2010.

SAS INSTITUTE INC. *SAS/STAT User's Guide*. v. 8. Cary, NC: 1999.

SCHAEFER, R. *Técnicas em Biologia Molecular*. Concórdia: Embrapa, Dezembro, 2006. Relatório Técnico 116.

SILVA, E.; THOMÉ, A. C. G. Reconhecimento de caracteres manuscritos utilizando time de redes neurais. *Anais do Congresso da Sociedade Brasileira de Computação*, SBC, [S.l.], 2003.

SOLFERINI, V. N.; SELIVON, D. *Biologia Molecular e Evolução*, capítulo: Polimorfismo de Isoenzimas, p. 137-142, Holos Editora, Ribeirão Preto, 1. ed., 2001.

STATSOFT. *Electronic Statistics Textbook*. Tulsa, OK, 2011. Consultado na Internet: <http://www.statsoft.com/textbook/>, em 05/08/2011.

VALE, M. N. *Agrupamentos de dados: Avaliação de métodos e desenvolvimento de aplicativo para análise de grupos*. Dissertação (Dissertação de Mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Agosto 2005.

YE, X. et al. *A recent development in image analysis of electrophoresis gels*. *Vision Interface '99, Trois-Rivières*, Canadá, VI, p. 432-438, Maio, 1999.