

**SiMCaPe - SISTEMA PARA A MONTAGEM DE
CARIÓTIPO DE PEIXE BASEADO EM
CONJUNTOS DIFUSOS**

Por

Elaine Ribeiro de Faria

DISSERTAÇÃO APRESENTADA À
UNIVERSIDADE FEDERAL DE UBERLÂNDIA, MINAS GERAIS,
COMO PARTE DOS REQUISITOS EXIGIDOS
PARA OBTENÇÃO DO TÍTULO DE MESTRE
EM CIÊNCIA DA COMPUTAÇÃO

FEVEREIRO DE 2006

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “**SiMCaPe - Sistema para a Montagem de Cariótipo de Peixe Baseado em Conjuntos Difusos**” por **Elaine Ribeiro de Faria** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 23 de Fevereiro de 2006

Orientador:

Prof^a. Dr^a. Denise Guliato
Universidade Federal de Uberlândia UFU/MG

Banca Examinadora:

Prof. Dr. Alcimar Barbosa Soares
Universidade Federal de Uberlândia UFU/MG

Prof. Dr. Luciano da Fontoura Costa
Universidade de São Paulo USP/São Carlos-SP

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Data: Fevereiro, 2006

Autor: **Elaine Ribeiro de Faria**
Título: **SiMCaPe - Sistema para a Montagem de Cariótipo de Peixe
Baseado em Conjuntos Difusos**
Faculdade: **Faculdade de Computação**
Grau: **Mestrado**

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão de cópias deste documento para propósitos exclusivamente acadêmicos, desde que o autor seja devidamente informado.

Autor

O AUTOR RESERVA PARA SI QUALQUER OUTRO DIREITO DE PUBLICAÇÃO DESTE DOCUMENTO, NÃO PODENDO O MESMO SER IMPRESSO OU REPRODUZIDO, SEJA NA TOTALIDADE OU EM PARTES, SEM A PERMISSÃO ESCRITA DO AUTOR.

Dedicatória

Aos meu pais Manoel e Olga e aos meus irmãos Eliana e Arley.

Ao meu namorado José Gustavo.

Agradecimentos

Quero agradecer, primeiramente à Deus, que me ajudou a vencer todos os obstáculos, me deu sabedoria, paz no coração, perseverança e calma. Ele ouviu minhas preces e permitiu que eu chegasse neste momento tão sonhado.

À Denise Guliato, fica minha gratidão por ser este exemplo de pessoa e profissional. Obrigada pela paciência na leitura dos meus textos, dedicação e auxílio no projeto e tantos conselhos. Você é um exemplo de pessoa, pelo seu caráter, inteligência, honestidade, visão política e personalidade. Mas ao mesmo tempo é simples, humilde, alegre e acolhedora.

Agradeço aos meus pais, Manoel e Olga, e meus irmãos, Eliana e Arley, que estiveram sempre ao meu lado me apoiando e me dando forças. Obrigada por terem me ajudado a chegar até uma universidade federal para cursar a graduação e o mestrado. Obrigado por entenderem as minhas demoradas idas pra casa. Vocês são muito importantes pra mim.

Agradeço à Sandra e Alessandra que me apresentaram o problema que motivou minha dissertação e me deram todo apoio necessário. Agradeço também aos alunos Roberto, Robson e Carolina que auxiliaram na execução dos testes.

Ao Jean Carlo, agradeço pelo auxílio no desenvolvimento do projeto. Sempre muito dedicado, alegre, otimista e companheiro. Esteve sempre do meu lado, nunca reclamou nem desistiu e aguentou firme as pressões. Obrigada pela ajuda.

Obrigada aos amigos do LCC-UFU pelo carinho e companheirismo. É ótimo trabalhar com pessoas tão alegres, sábias, humildes, pessoas que se doam, que compartilham. Vocês são inesquecíveis.

Agradeço à CAPES - Conselho de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa oferecida, que permitiu o desenvolvimento deste trabalho.

E finalmente, ao meu amor, José Gustavo, que soube compreender a minha ausência e os meus dias de stress. Que me deu tranquilidade nos momentos mais difíceis, apoiando minhas decisões. Você é minha referência, sempre centrado, calmo, sereno e alegre. Obrigada por me acompanhar nesta caminhada e por ser tão compreensivo.

"O caminho da Sabedoria é não ter medo de errar."
(Paulo Coelho)

Resumo

O estudo do cariótipo de diversos organismos tem como objetivo observar variações numéricas, detectar características e estudar a evolução das espécies. A montagem manual do cariótipo é uma tarefa repetitiva, subjetiva e consome tempo. Este projeto descreve o SiMCaPe, um sistema interativo para apoiar na montagem do cariótipo de peixe, a partir de uma fotomicrografia digitalizada de uma célula de peixe em metáfase. O SiMCaPe possui 5 módulos: pré-processamento e segmentação, rotação, localização do centrômero, classificação e pareamento. Para pré-processar e segmentar os cromossomos da imagem foram propostos métodos baseados em conjuntos difusos. Para rotacionar os cromossomos para a posição vertical foi utilizado a orientação do eixo central do esqueleto do cromossomo. Para localizar o centrômero foi utilizado um vetor de projeção horizontal do cromossomo. Para classificar os cromossomos nas classes metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico foi utilizado um sistema de inferência difusa, que combina regras incluindo descritores de características como índice centromérico e razão entre braço maior e braço menor. O pareamento foi feito com base no tamanho dos cromossomos. Os métodos propostos foram aplicados a 5 espécies de peixe: *Astyanax altiparanae*, *Astyanax eigenmanniorum*, *Astyanax scabripinnis*, *Hoplias malabaricus* e *Hoplias lacerdade*. Para testes foram utilizadas 40 imagens, contendo em média 49,2 cromossomos por imagem, sendo que no processo de segmentação cerca de 64,05 % dos cromossomos por imagem foram corretamente separados, no processo de rotação cerca de 69,5% dos cromossomos por imagem foram corretamente rotacionados. Os resultados obtidos pelo sistema SiM-CaPe para processo de localização do centrômero e classificação foram comparado com os resultados obtidos por dois grupos de citogeneticistas. Comparando os resultados do sistema com o primeiro grupo de citogeneticistas, o índice de concordância na localização do centrômero foi de 71,95 % e para a classificação foi de 78,52%. Comparando os resultados do sistema com o segundo grupo de citogeneticistas, o índice de concordância na localização do centrômero foi de 72,83 % e para a classificação foi de 81,16%.

Palavras chave: cariótipo, segmentação de cromossomos, localização do centrômero, classificação de cromossomos, sistema de inferência difusa e pareamento de cromossomos.

Abstract

The karyotype study of several organisms has as its objective to observe numerical variations, for detecting specific characteristics and to study the evolution of the species. Manual assembling of karyotype is a repetitive, subjective and time consuming task. This project describes the SiMCaPe, an interactive system for assisting the assembling of fish karyotype from digitized photomicrography of fish cells in metaphase. The SiMCaPe has 5 modules: pre-processing and segmentation, rotation, centromere location, classification and pairing. To pre-process and segment the chromosomes, methods based on fuzzy sets were used. To rotate the chromosome to its vertical position, the orientation of the central axis of the chromosome skeleton was used. To locate the centromere, a horizontal projection vector of the chromosome was used. To classify the chromosomes in metacentric, submetacentric, subtelocentric and acrocentric classes, a fuzzy inference system, which combines rules including feature descriptors as the centromeric index and the ratio between the longer and shorter arms, was used. The pairing of the chromosomes was made according to the chromosomes size. The proposed methods were applied to 5 species of fish: *Astyanax altiparanae*, *Astyanax eigenmanniorum*, *Astyanax scabripinnis*, *Hoplias malabaricus* and *Hoplias lacerdade*. The tests were carried out using 40 images, which contained on average 49.2 chromosomes for each image. In the segmentation process, on average 64.05% of the chromosomes in each image were separated correctly, in the rotation process, on average 69.5% of the chromosomes in each image were rotated correctly. The results obtained by the SiMCaPe system for the centromere location and classification process were compared with the results obtained by the two groups of cytogeneticists. Comparing the results of the system with the first group of cytogeneticists, the level of agreement between the two for the centromere location was 71.95% and for the classification 78.52%. Comparing the results of the system with the second group, the level of agreement between the two for the centromere location was 78.83% and for the classification 81.16%.

Keywords: karyotype, chromosome segmentation, centromere location, chromosome classification, inference fuzzy system and pairing of chromosomes.

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Objetivo	2
1.3	Organização da dissertação	4
2	O Cariótipo de Peixe e os Principais Sistemas para Análise Citogenética	6
2.1	Introdução	6
2.2	Importância do estudo do cariótipo de peixe	7
2.3	O Cromossomo	7
2.4	Cariótipo	9
2.5	Processo manual para a montagem do cariótipo de peixe	11
2.6	Diferenças entre o cariótipo humano e o de peixe	13
2.7	Sistemas para análise e montagem de cariótipos	14
2.8	Considerações Finais do Capítulo	27
3	Segmentação dos Cromossomos e Extração de Características	29
3.1	Introdução	29
3.2	Visão Geral da Fase de Segmentação e Extração de Características	29
3.3	Pré-processamento da imagem	30
3.3.1	Aumento de contraste	31
3.3.2	Realce da imagem baseada em conjuntos difusos	31
3.3.3	Limiarização e rotulação da imagem	33
3.4	Segmentação da imagem baseada no contorno	34

3.4.1	Separação dos cromossomos que se tocam	36
3.4.2	Segmentação dos cromossomos	36
3.5	Rotação dos cromossomos	38
3.5.1	Método de Holt	40
3.5.2	Método de <i>Staircase Removal</i>	41
3.6	Extração das características do cromossomo	41
3.6.1	Localização do centrômero utilizando um vetor de projeção	42
3.6.2	Cálculo do índice centromérico e razão entre braços	42
3.7	Considerações Finais do Capítulo	44
4	Classificação e Pareamento	45
4.1	Introdução	45
4.2	Visão geral das etapas de classificação e pareamento	45
4.3	Classificação	46
4.3.1	Sistema de Inferência Difusa: visão geral	47
4.3.2	Sistema de Inferência Difusa usado na Classificação de Cromos- mos de Peixe	51
4.4	Pareamento	58
4.5	Considerações Finais do Capítulo	59
5	O Sistema para Montagem de Cariótipos	61
5.1	Introdução	61
5.2	Recursos utilizados no desenvolvimento do SiMCaPe	62
5.3	Pré-processamento e segmentação dos cromossomos	63
5.4	Separação dos cromossomos que se sobrepõem	65
5.5	Rotação de cromossomos	66
5.6	Pareamento de cromossomos	67
5.7	Escolha do melhor cariótipo	68
5.8	Modelo da Base de Dados proposta para o SiMCaPe	70

5.9	Considerações Finais do Capítulo	71
6	Avaliação e Discussão dos Resultados	72
6.1	Metodologia para realização dos experimentos	72
6.2	Avaliação dos resultados	73
6.3	Considerações Finais do Capítulo	76
7	Conclusão	78

Lista de Figuras

1.1	Visão Geral do Sistema	3
2.1	Estrutura do cromossomo	8
2.2	Classes de cromossomo (a) cromossomo metacêntrico (b) cromossomo sub- metacêntrico (c) cromossomo subtelocêntrico (d) cromossomo acrocêntrico	9
2.3	Cariótipo de Peixe (a) Imagem de uma fotomicrografia de uma célula de peixe em metáfase (b) Resultado final do processo de cariotipagem (cari- ograma)	10
3.1	Visão geral das etapas de segmentação e extração de características	31
3.2	Resultado obtido pela aplicação da função quadrática expressa na Equação 4.2 (a) imagem original (b) imagem com aumento de contraste	32
3.3	Resultado da aplicação do procedimento de realce baseado em conjuntos difusos	33
3.4	Resultado da operação lógica AND entre a imagem binária e a imagem original	34
3.5	Resultado da aplicação da função de pertinência difusa definida na Equação 3.3	35
3.6	Resultado da aplicação da limiarização	36
3.7	Operação lógica AND (a) imagem resultante do processo de filtragem (b) imagem resultante do pré-processamento (c) imagem resultante da oper- ação lógica AND entre as imagens (a) e (b).	37
3.8	Resultado do contorno sobreposto à imagem original	38

3.9	Resultado do processo de rotação (a) imagem binária com a região entre as cromátides preenchida e o esqueleto sobreposto (b) imagem binária rotacionado de acordo com o seu esqueleto (c) imagem original rotacionada de acordo com o seu esqueleto	39
3.10	Resultado do processo de rotação para uma imagem que possui esqueleto com ramificações (a) imagem binária com a região entre as cromátides preenchida e o esqueleto sobreposto (b) imagem binária com o esqueleto sem ramificações (c) imagem binária rotacionado de acordo com o seu esqueleto (d) imagem original rotacionada de acordo com o seu esqueleto	39
3.11	Vizinhança de um pixel C, usada no algoritmo de Holt	41
3.12	Máscaras utilizadas no processo de <i>staircase removal</i>	41
3.13	Cálculo do vetor de projeção (a) imagem do cromossomo depois que a região entre as cromátides foi preenchida (b) vetor de projeção usado para escolher a posição do centrômero	43
3.14	Resultado da localização do centrômero (a) cromossomo metacêntrico; (b) cromossomo submetacêntrico e (c) cromossomo subtelocêntrico	43
4.1	Visão geral das etapas de classificação e pareamento	46
4.2	Módulos do sistema de inferência fuzzy	48
4.3	Módulo Motor de Inferência Difusa	50
4.4	Números difusos que representam os estados lingüísticos da variável lingüística CI	52
4.5	Números difusos que representam os estados lingüísticos da variável lingüística AR	52
4.6	Números difusos que representam os estados lingüísticos da variável lingüística CL_i	53
4.7	Imagem de um cromossomo de peixe, com seu centrômero localizado e que possui $CI = 31,7$ e $AR = 2,15$	55
4.8	Conclusão do motor de inferência difusa para a regra 2 da variável lingüística $CL_{metacentrico}$	55

4.9	Conclusão do motor de inferência difusa para as regras da classe <i>CLmetacentrico</i>	
	(a) Resultado da implicação para cada regra da classe <i>CLmetacentrico</i> e	
	(b) Resultado da união para a classe <i>CLmetacentrico</i>	56
4.10	Método do centróide aplicado ao resultado da agregação 4.9 (b), resultando	
	no valor 6,2.	58
4.11	Cariótipo (a) imagem de cromossomos de peixe em metáfase (b) resultado	
	do processo de classificação dos cromossomos (c) resultado do processo de	
	pareamento	60
5.1	Escolha do limite superior para execução da função de pertinência	64
5.2	Cromossomos sobrepostos duplicados que serão separados interativamente	65
5.3	Cromossomos sobrepostos que foram separados interativamente	66
5.4	Rotação manual de um cromossomo	67
5.5	Rotação manual de múltiplos cromossomos	67
5.6	Resultado do pareamento	68
5.7	Tela de escolha do melhor cariótipo	69
5.8	Modelo do Banco de Dados	70

Lista de Tabelas

2.1	Resumo dos principais sistemas existentes para montagem de cariótipo . . .	26
-----	--	----

Capítulo 1

Introdução

1.1 Introdução

A análise de cromossomos é uma tarefa essencial para a detecção de algumas doenças, detecção de células anormais e variações numéricas. O processo usado para visualização e classificação dos cromossomos é chamado de cariotipagem. Os cromossomos são classificados em quatro classes: metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico, dependendo da posição do centrômero. Então em cada uma destas classes os cromossomos são pareados de acordo com sua similaridade e dispostos em ordem decrescente de tamanho, resultando no chamado cariótipo.

O processo manual para obter o cariótipo é muito repetitivo, consome tempo e requer um profissional experiente, o que torna o processo caro. Uma vez feita a aquisição da imagem digital dos cromossomos, o processo manual requer que o citogeneticista recorte cada cromossomo da imagem e visualmente determine a posição do centrômero (ponto onde as cromátides se juntam). Depois de detectar o centromero, os braços dos cromossomos são medidos. De acordo com a relação entre o tamanho dos braços, os cromossomos são classificados em uma das quatro classes, mencionadas acima. Os cromossomos de cada classe são pareados de acordo com sua similaridade e o cariótipo é apresentado com os pares de cromossomo em ordem decrescente de tamanho.

Para classificar os cromossomos automaticamente, a maioria dos sistemas apresentados na literatura são baseados em redes neurais, que usam o índice centromérico, tamanho

do cromossomo e o padrão de banda como entrada para o sistema e são aplicados à cromossomos humanos. Entretanto os métodos usados para montar o cariótipo humano não podem ser aplicados para cariótipo de peixe devido ao fato de que os peixes apresentam três principais particularidades: Primeira, o cromossomo de peixe não apresenta a característica padrão de banda. Como a maioria dos sistemas existentes usam padrão de banda como entrada para o processo de classificação, não é possível utilizar tais sistemas para a classificação de cromossomos de peixe. Segunda, nos cromossomos humanos é possível identificar 46 cromossomos. Os peixes não têm um número constante de cromossomos. Algumas espécies apresentam 46 cromossomos, outras 48 ou 50. Além disso, os peixes de uma mesma espécie podem apresentar números de cromossomos diferentes, como por exemplo, a espécie *Astyanax scabripinnis*. Por último, o número e as características dos pares de cromossomos de peixe não são previamente conhecidos como nos humanos. O cariótipo humano possui 46 cromossomos, divididos em 22 pares de cromossomos e 1 par relacionado ao sexo, totalizando 23 pares de cromossomos. As características de cada par de cromossomos humanos são bem conhecidas, enquanto que nos peixes o número de pares de cromossomos e suas características não são conhecidas até que o cariótipo seja montado.

1.2 Objetivo

Considerando a falta de sistemas aplicados a montagem de cariótipo de peixe e as diferenças associadas à solução deste problema, o objetivo deste trabalho é desenvolver um sistema computadorizado e interativo para auxiliar na montagem do cariótipo de peixe, tendo como entrada uma fotomicrografia digitalizada do cromossomo de peixe no estágio de metáfase (estágio de divisão celular).

O sistema proposto está dividido em 4 módulos principais: módulo para segmentação [1], módulo para extração de características [2], módulo para classificação e módulo para pareamento. Um esquema geral do sistema é mostrado na Figura 1.1:

O módulo para segmentação do cromossomo inclui dois sub-módulos:

- Pré-processamento: tem por objetivo pré-processar a imagem a fim de eliminar os

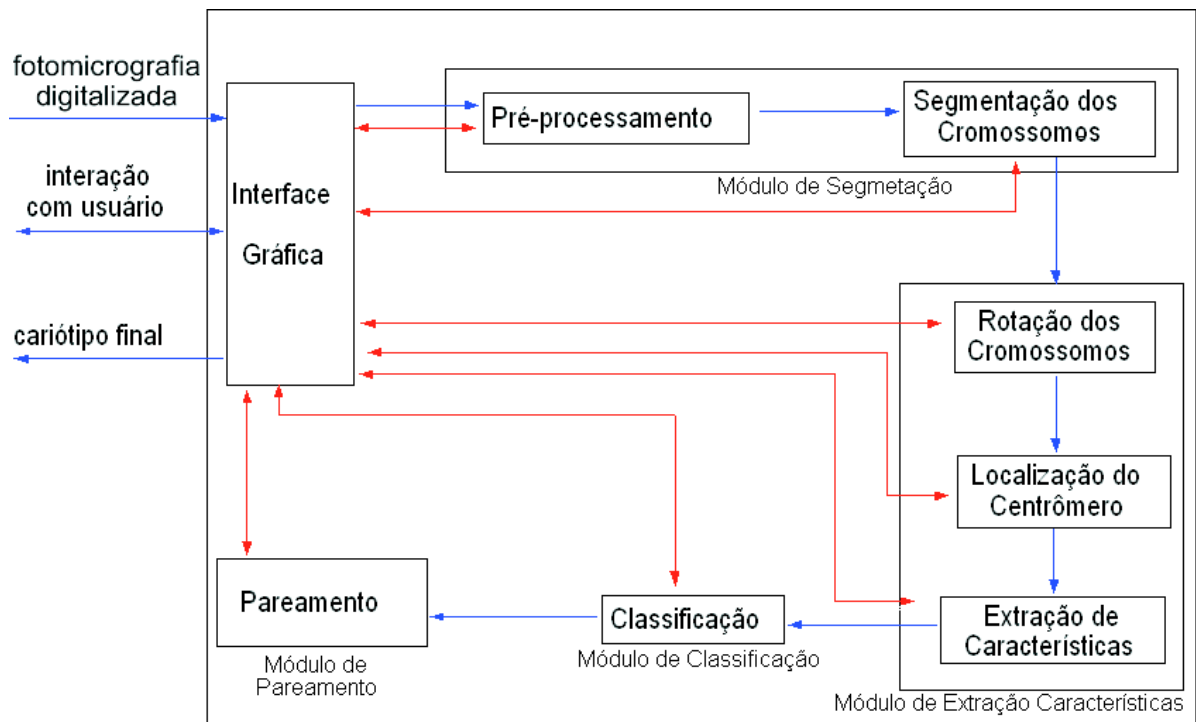


Figura 1.1: Visão Geral do Sistema

ruídos da imagem, aumentar o contraste entre os cromossomos e o fundo e eliminar o fundo da imagem;

- Segmentação: tem por objetivo segmentar os cromossomos da imagem, separando inclusive os cromossomos que se tocam, ou seja, aqueles cromossomos que estão muito próximos e que não existe um região de descontinuidade entre eles;

O módulo para extração de características inclui 3 sub-módulos:

- Rotação: tem por objetivo rotacionar os cromossomos para a posição vertical a fim de facilitar a visualização por parte do citogeneticista e facilitar a localização do centrômero;
- Localização do centrômero: tem por objetivo localizar a posição do centrômero, que é o ponto onde as cromátides (filamentos que forma o cromossomo) se juntam. O centrômero divide cada cromátide em dois braços: braço maior e braço menor.

- Extração de características: tem por objetivo extrair características como perímetro, tamanho do cromossomo, índice centromérico (CI) e razão entre braços (AR). O CI e AR são medidas extraídas a partir do tamanho dos braços do cromossomo e são usadas no processo de classificação;

O módulo para classificação tem por objetivo utilizar as características do cromossomo para classificá-lo em um das 4 classes conhecidas: metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico;

O módulo para pareamento tem por objetivo parear os cromossomos de acordo com sua similaridade, montando assim o cariótipo final da espécie.

O sistema também:

- disponibiliza uma interface gráfica que permite separar interativamente os cromossomos que se encontram sobrepostos;
- disponibiliza um banco de dados para que os resultados do processo de cariotipagem sejam armazenados para posteriores consultas ou alterações;
- disponibiliza recursos de interação para que a cada etapa do sistema, o usuário possa alterar interativamente os resultados propostos (veja linhas vermelhas na Figura 1.1);
- disponibiliza uma interface gráfica que facilita, ao usuário, a escolha do cariótipo que melhor representa uma população de peixe em estudo.

1.3 Organização da dissertação

A dissertação está organizada da seguinte maneira. O capítulo 2 está dedicado à uma revisão sobre os conceitos citogenéticos relacionados a este trabalho bem como uma revisão dos principais trabalhos desenvolvidos sobre automação de cariotipagem. O capítulo 3 apresenta os métodos propostos para pré-processamento e segmentação de cromossomos de peixe, utilizando conjuntos difusos, e apresenta o método de extração de características do cromossomo. O capítulo 4 apresenta os métodos propostos para classificação e pareamento

de cromossomos de peixe. O capítulo 5 mostra o sistema SiMCaPe, focando nos recursos de interface disponíveis. O capítulo 6 apresenta uma discussão dos resultados obtidos pelo sistema. O capítulo 7 apresenta as conclusões finais e os trabalhos futuros.

Capítulo 2

O Cariótipo de Peixe e os Principais Sistemas para Análise Citogenética

2.1 Introdução

A análise e classificação citogenética é uma tarefa comumente realizada nos laboratórios. O estudo do cariótipo de diversos seres vivos tem se tornado uma tarefa importante em vários campos de estudo, visando observar variações numéricas, detectar características a partir de sua estrutura e estudar a evolução das espécies.

A montagem do cariótipo, é uma tarefa importante para o estudo do cromossomo, pois dispõe os cromossomos distribuídos em classes e organizados em pares de acordo com sua similaridade, facilitando a análise do citogeneticista. No entanto, esta é uma tarefa manual, repetitiva, subjetiva e sujeita a erros. Vários sistemas existentes atualmente têm como objetivo automatizar ou semi-automatizar o processo de montagem de cariótipos, tornando o processo mais rápido e com menores chances de erro. A maioria destes sistemas são aplicados a cariótipo humano e pouca atenção tem sido dada à montagem semi-automática de cariótipo de peixes.

Neste capítulo, será apresentada a importância do estudo do cariótipo dos peixes e a estrutura biológica do cromossomo. Em seguida será mostrado o que é um cariótipo, como representá-lo e o processo manual para obtê-lo. Também serão apresentadas as principais diferenças entre o cariótipo dos peixes e dos humanos. Finalmente serão apresentados

os principais sistemas existentes para a montagem de cariótipos humanos e de peixe, e diante das limitações destes sistemas, será apresentado o escopo deste trabalho.

2.2 Importância do estudo do cariótipo de peixe

Segundo Carvalho et al. [3] os peixes formam um grande grupo de organismos atualmente, cerca de 24.618 espécies. Este número representa cerca de 50% dos vertebrados vivos. Do subfilo dos vertebrados, eles são os animais mais primitivos e constituem a base da evolução dos anfíbios, répteis, aves e mamíferos, por isso ocupam uma posição central na evolução desses animais [4]. Para Denton [5], os peixes constituem um dos grupos mais favoráveis para estudos genéticos e citogenéticos. Para Kossing [6], os peixes são um grupo importante para estudos evolutivos.

Segundo Artoni et al. [7], a maior aplicabilidade da citogenética de peixes, sem dúvida alguma se refere aos estudos evolutivos. Os dados da citogenética também têm sido associados a dados da morfologia, biogeografia, comportamento e genética molecular para se chegar mais próximo à história evolutiva real dos organismos.

Como pode ser visto a citogenética é uma ciência que traz resultados importantes principalmente para os estudos evolutivos, sendo que esta ainda apresenta muitos problemas a serem resolvidos. Como esta ciência se mostra bastante receptiva ao uso de novas tecnologias, como por exemplo, os sistemas computacionais, muito ainda pode ser feito pela associação de cientistas da computação e biólogos citogeneticistas.

2.3 O Cromossomo

O cromossomo é uma estrutura microscópica em forma de filamento que contém o material genético, o DNA. Os cromossomos situam-se no interior do núcleo das células nos organismos eucariotes. Os cromossomos são facilmente estudados em seu estado mais condensado, que ocorre durante a metáfase da mitose ou da meiose [8]. Vale ressaltar que mitose e meiose são processos de divisão celular.

Cada cromossomo metafásico é constituído por duas subunidades paralelas chamadas

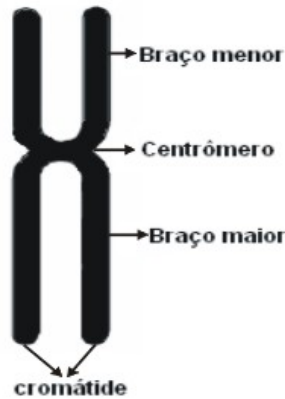


Figura 2.1: Estrutura do cromossomo

de cromátides. As cromátides são ligadas entre si por uma região chamada constrição primária ou centrômero. O centrômero divide cada cromátide em dois segmentos, os chamados braços cromossômicos [9]. Considerando um cromossomo corretamente rotacionado na posição vertical, o braço localizado na parte superior à posição do centrômero, é chamado de braço menor (p) e o localizado na parte inferior é chamado braço maior (q). A Figura 2.1 mostra a estrutura do cromossomo.

De acordo com a posição do centrômero, os cromossomos são classificados em quatro classes: metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico. Os cromossomos metacêntricos possuem o centrômero localizado na sua parte central ou quase central, dividindo os braços do cromossomo em partes iguais ou aproximadamente iguais. Já nos cromossomos submetacêntricos, o centrômero está deslocado para a parte superior e divide o cromossomo em dois braços de tamanhos diferentes. Os cromossomos subtelocêntricos possuem um centrômero próximo da sua extremidade, formando dois braços com tamanhos bem diferentes, um muito grande e outro muito pequeno. Por último, os cromossomos acrocêntricos são aqueles que possuem o centrômero em uma das suas extremidade, sendo que praticamente não existe o braço menor. A Figura 2.2, ilustra cada classe de cromossomo.

A classificação de um cromossomo, em geral, é definida pelas medidas dadas pela: a razão entre os braços (AR), Equação 4.1, e pelo o índice centromérico (CI), Equação

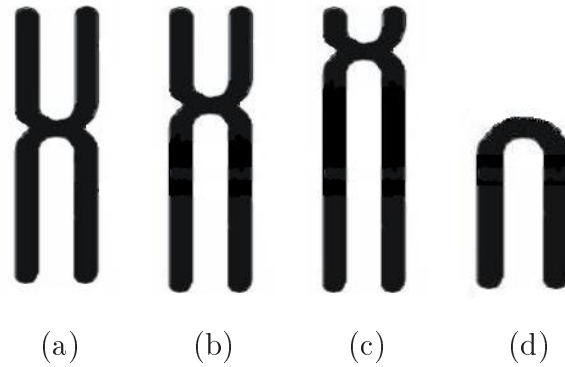


Figura 2.2: Classes de cromossomo (a) cromossomo metacêntrico (b) cromossomo submetacêntrico (c) cromossomo subtelocêntrico (d) cromossomo acrocêntrico

2.2. A razão entre os braços é a divisão entre o tamanho do braço maior do cromossomo pelo braço menor. O índice centromérico é a razão entre o tamanho do braço menor pelo tamanho total do cromossomo, multiplicado por 100.

$$AR = \frac{\text{comprimento}(q)}{\text{comprimento}(p)} \quad (2.1)$$

$$CI = \frac{\text{comprimento}(p)}{\text{tamanho do cromossomo}} * 100 \quad (2.2)$$

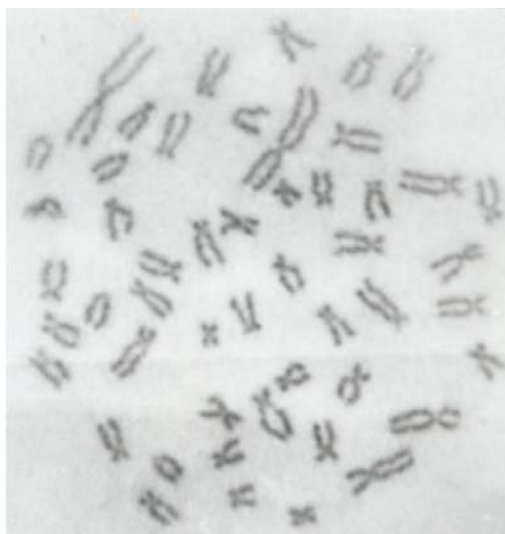
2.4 Cariótipo

A palavra karyotype (cariótipo em português) é derivada da palavra Grega "karyon" que significa núcleo e "typos" que significa imprimir. Segundo Guerra [9], o cariótipo é a descrição das características do conjunto cromossômico de uma espécie.

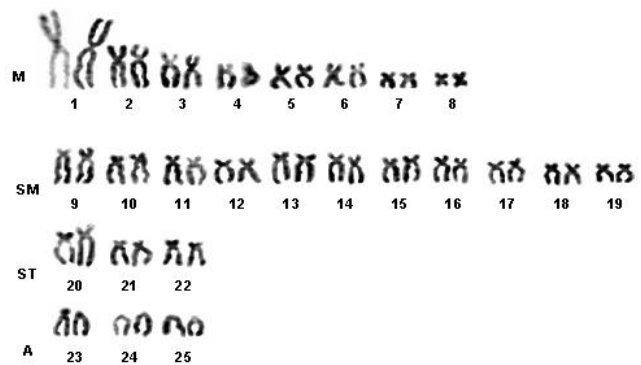
O estudo do cariótipo é de grande importância para analisar variações entre indivíduos da mesma espécie ou para comparar espécies diferentes [10] e [4]. Também é muito importante para auxiliar nos estudos evolutivos. Além disso visa o diagnóstico de anomalias genéticas relacionadas ao número ou à morfologia dos cromossomos. Um exemplo, é a leucemia promielocítica, detectada por um prolongamento na extremidade de um cromossomo [11] e outro é a Síndrome de *Down* ou trissomia do cromossomo 21, detectada pela

ocorrência de 3 cromossomos no par 21, ou seja, a existência de um cromossomo a mais do que o normal.

O processo de montagem de um cariótipo, conhecida como cariotipagem é obtido a partir de uma fotomicrografia de uma célula em metáfase, veja Figura 2.3 (a). Várias características são obtidas de cada cromossomo (exemplo: CI, AR, padrão de banda, área e perímetro) com o objetivo de classificá-los. Os cromossomos de cada classe são pareados de acordo com sua similaridade, montando assim o cariótipo final do indivíduo. O cariótipo pode ser representado por um cariograma, onde os cromossomos são segmentados e os cromossomos homólogos são emparelhados, ou por um idiograma, que é uma representação esquemática, utilizando valores médios da posição do centrômero e do tamanho do cromossomo. Normalmente, o estudo do cariótipo é feito analisando o cariograma, veja Figura 2.3 (b).



(a)



(b)

Figura 2.3: Cariótipo de Peixe (a) Imagem de uma fotomicrografia de uma célula de peixe em metáfase (b) Resultado final do processo de cariotipagem (cariograma)

Para montar o cariograma, três características são levadas em consideração. A primeira delas é o centrômero. Dependendo da posição do centrômero, o cromossomo pode ser classificado em uma das quatro classes, mencionadas na seção 2.3. A segunda é o tamanho dos cromossomos. O tamanho varia muito de acordo com a espécie. Algumas espécies

apresentam cromossomos maiores e algumas apresentam cromossomos bem pequenos, o que dificulta a análise cromossômica. O tamanho é uma característica importante ao montar o cariótipo final, visto que os cromossomos são dispostos em ordem decrescente de tamanho. A terceira característica é o número de cromossomos. Em geral, o número de cromossomos dos indivíduos de uma mesma espécie é um número constante. Mas cada espécie pode apresentar um número de cromossomos diferente. No entanto existem algumas exceções, como é o caso dos peixes da espécie *Astyanax scabripinnis*, onde indivíduos da mesma espécie apresentam números de cromossomos diferentes.

Além das três características mencionadas acima a característica que se refere às colorações das bandas cromossômicas é muito útil no processo de classificação e pareamento cromossômico. Em cromossomos humanos, observa-se uma boa qualidade das bandas cromossômicas, sendo que nos peixes, a qualidade é bem inferior, não sendo útil para o propósito de classificação.

2.5 Processo manual para a montagem do cariótipo de peixe

Como foi mencionado anteriormente, o estudo do cariótipo é uma tarefa importante para caracterizar os seres vivos, estudar a sua evolução e detectar anormalidades e variações numéricas de cromossomos. No entanto o processo manual para montagem de cariótipos é um processo repetitivo, tedioso e sujeito a erros. A maior parte do processo é bastante subjetivo e necessita de um profissional experiente, o que torna este processo caro e dispendioso.

Para a montagem manual do cariótipo de peixe é necessário obter uma fotomicrografia de uma célula no estágio de metáfase. A imagem adquirida, em geral, apresenta ruído e baixo contraste. Assim, muitas vezes, o citogeneticista usa *software* de processamento digital de imagem para eliminar ruídos, aumentar o contraste entre os cromossomos e o fundo e eliminar alguns artefatos presentes na imagem, como por exemplo parte do núcleo metafásico que se encontra condensado.

Em seguida, cada cromossomo é recortado da imagem para ser analisado individual-

mente. A posição do centrômero de cada cromossomo é marcada pelo citogeneticista. No caso dos cromossomos acrocêntricos, dos cromossomos muito tortos e no caso de sobreposição de cromossomos, a experiência do citogeneticista na determinação da localização do centrômero é muito importante.

Depois de localizar o centrômero, usando a fotomicrografia que mostra os cromossomos ampliados, o tamanho dos braços do cromossomo é medido usando um paquímetro. Para tanto, vários segmentos de reta são traçados ligando o centrômero até a extremidade inferior e superior do cromossomo, de forma a representar a curvatura dos braços menor e maior do cromossomo. O tamanho de cada braço do cromossomo é então obtido pelo somatório do tamanho destes segmentos.

Com base no tamanho dos braços do cromossomo, é então calculada a razão entre os braços (AR), conforme mostra a Equação 4.1. De acordo com o valor do AR cada cromossomo é classificado em uma das 4 possíveis classes, obedecendo às seguintes regras, estabelecidas por Levan [12] e revistas por Guerra [9]. Se o AR (relação entre os tamanhos dos braços) está no intervalo $[1; 1, 7]$ o cromossomo é classificado como metacêntrico. Se AR está no intervalo $]1, 7; 3, 0]$ o cromossomo é submetacêntrico. Se AR está no intervalo $]3; 7, 0]$ o cromossomo é subtelocêntrico. E se o AR é maior que 7,0 o cromossomo é acrocêntrico.

Depois de classificados, os cromossomos da mesma classe são então ordenados de forma decrescente com relação ao seu tamanho e pareados de acordo com sua similaridade. Para o pareamento são analisadas características como o tamanho do cromossomo e a posição do centrômero. Ao final deste processo obtém-se o cariótipo.

Como este processo é muito subjetivo, muitas vezes, depois de montar o cariótipo, o citogeneticista faz modificações no pareamento, alterando os pares de cromossomos ou até mesmo trocando a classificação de um cromossomo.

Quando uma determinada população está sendo estudada, ao final da análise de vários cariótipos, é necessário encontrar o cariótipo que melhor represente a população em estudo. Neste caso o citogeneticista avalia os vários cariótipos já montados, e escolhe o melhor deles. O cariótipo escolhido é aquele que mantém as mesmas características estruturais encontradas nos demais cariótipos, que apresenta uma imagem de boa qualidade,

centrômeros bem localizados e que possui pareamentos consistentes.

Em geral, o processo manual para montagem de cariótipos de todos os seres vivos é semelhante, visto que todos se baseiam no cariótipo humano. No entanto, o cariótipo de cada espécie pode apresentar algumas particularidades, o que pode alterar alguns dos passos do processo de montagem de cariótipos. Na próxima seção serão apresentadas as principais diferenças entre o cariótipo humano e o cariótipo de peixes.

2.6 Diferenças entre o cariótipo humano e o de peixe

Como será visto na próxima seção, existem muitos sistemas para auxiliar o citogeneticista na montagem de cariótipos humanos. No entanto, estes sistemas não podem ser usados na montagem de cariótipos de peixe, pois os peixes apresentam algumas particularidades que serão discutidas nesta seção.

A principal diferença entre o cariótipo de peixes e humanos, está no número de cromossomos. Nos seres humanos, o número de cromossomos em uma célula somática é 46 cromossomos (célula diplóide). Nos peixes o número de cromossomos varia de acordo com a espécie, e em alguns casos também é possível que uma mesma espécie de peixe apresente números cromossômicos diferentes. Um exemplo é a espécie *Astyanax scabripinnis*, na qual foram detectados cariótipos com 46 [13], 48 [14] e 50 [10] cromossomos.

Outra diferença está na qualidade da imagem e na forma dos cromossomos. A imagem de peixe, em geral é mais ruidosa e apresenta menor contraste. As técnicas criadas para a aquisição da célula em metáfase em humanos foram adaptadas para peixe. No entanto, nos peixes essa técnica não produz tão bons resultados sendo que em muitos casos o material químico utilizado para obter a célula em metáfase proporciona baixo contraste e ruídos na imagem. Além disso, a qualidade do padrão de banda G é ruim, visto que o arranjo do DNA do peixe é diferente do humano. Nos humanos, existe uma melhor compartimentalização do DNA, portanto a qualidade do padrão de banda G é bem superior, sendo que esta característica é freqüentemente usada para auxiliar nos processos de classificação e pareamento. Os cromossomos de peixes, normalmente são pequenos e numerosos, por isso, existe uma dificuldade grande no seu pareamento [4]. Além disso, os

braços dos cromossomos estão disjuntos, apresentando assim esqueletos mais ramificados do que o dos cromossomos humanos.

O processo de classificação e pareamento também difere entre humanos e peixes. Em humanos, o processo de classificação além de caracterizar os cromossomos como metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico, também separa os cromossomos em 7 grupos, o chamado grupos de Denver [15]. Para classificar nestes 7 grupos (grupo A, B, C, D, E, F e G) características como o tamanho do cromossomo, a posição do centrômero e o padrão de banda são utilizados. O processo de pareamento considera o conhecimento de que existem 22 pares de cromossomos e 1 par de cromossomos relacionado ao sexo, totalizando 23 pares de cromossomos.

Ao contrário dos humanos, os peixes não possuem um padrão, previamente conhecido, de ordenação dos pares cromossômicos. Também não é possível identificar previamente o número de cromossomos do cariótipo. Para obter o pareamento dos cromossomos de peixe, inicialmente cada cromossomo é classificado em uma das 4 classes citadas anteriormente (metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico) e os cromossomos de cada classe são pareados de acordo com sua similaridade. Assim, a cada população, um cariótipo diferente pode ser encontrado.

Outra particularidade do cariótipo de peixe é que algumas espécies de peixe apresentam cromossomos extras chamados de cromossomos supranumerários, também chamados de cromossomos acessórios ou cromossomos B. Estes cromossomos são diferenciados em relação aos outros cromossomos que compõem o cariótipo, que são chamados cromossomos A, pois não são pareados com nenhum outro. O cromossomo B necessita ser previamente detectado pelo citogeneticista para que ele não participe do processo de pareamento.

2.7 Sistemas para análise e montagem de cariótipos

Há vários anos cientistas da computação em conjunto com biólogos citogeneticistas estão tentando construir sistemas que possam, a partir de uma imagem de uma célula em metáfase, auxiliar no processo de montagem de um cariótipo, com o mínimo de intervenção humana.

Uma das primeiras tentativas de automatizar o processo de classificação dos cromossomos humanos foi feita por Ledley and Ruddle [16]. Eles propuseram o FIDACSYS, um sistema computadorizado que desempenha as seguintes tarefas: eliminação do fundo da imagem, contagem do número de cromossomos, medição do comprimento dos cromossomos e de sua área, classificação e pareamento. Inicialmente o sistema propõe um valor de nível de cinza que será usado como um limiar que separa os cromossomos do fundo da imagem, gerando em seguida o contorno dos cromossomos. O contorno do cromossomo é representado por um conjunto de segmentos de curva. Com base nestes segmentos, o sistema aplica uma técnica denominada reconhecimento de padrões sintaticamente dirigido e classifica o cromossomo. Depois disso os cromossomos homólogos são comparados de acordo com a área, o tamanho e a relação do comprimento dos braços. Embora muito rudimentar, este sistema foi muito importante para alavancar as pesquisas na área de automação da montagem de cariótipos.

Van Vliet et al. propuseram um dos primeiros sistemas comerciais para montagem de cariótipos humanos, o sistema Athena [17]. Este sistema semi-automático propõe técnicas para segmentar a imagem, separar os cromossomos que se tocam e classificá-los baseado no seu comprimento, índice centromérico e padrão de banda. A separação dos cromossomos sobrepostos é feita manualmente. A segmentação dos cromossomos, proposta pelo Athena, é baseada na combinação das técnicas de *shadding correction*, *threshold* e operações binárias. A técnica do *shadding correction* possibilita a remoção de artefatos causados por iluminação inadequada ou devido à sensibilidade da câmera. A classificação dos cromossomos é baseada no método Bayesiano que calcula a probabilidade com que cada cromossomo pertence a uma determinada par. O sistema Athena requer que seja fornecida uma base de dados treinada, para servir como modelo para o processo de classificação. Os principais problemas identificados no sistema Athena estão relacionados ao procedimento de separação dos cromossomos que se tocam e à dificuldade em localizar o centrômero e classificar determinados cromossomos. Além disso, em cada fase do processo de montagem do cariótipo, o usuário deve configurar parâmetros relacionados à cada procedimento a ser executado. Este fato torna a montagem do cariótipo uma tarefa difícil para os citogeneticistas, os quais não estão familiarizados com os conceitos e termos usados

em sistemas que incorporam técnicas de processamento digital de imagens.

Jennings and Graham [18] fizeram um estudo sobre o uso de redes neurais para a classificação de cromossomos humanos, usando como entrada o padrão de banda do cromossomo. Este trabalho não tratou dos problemas de segmentação e separação de cromossomos sobrepostos. Duas arquiteturas diferentes foram usadas: *Multi-layer perceptron* (MLP) e *Kohonen self-organizing*. Os testes usaram cerca de metade dos dados para treinamento da rede e a outra metade para teste do desempenho da rede. O desempenho da rede foi medido considerando o número de cromossomos classificados incorretamente no treinamento e na fase de teste. A arquitetura MLP, usou o algoritmo de *backpropagation* para o treinamento, tendo apresentado os melhores resultados, especialmente quando características como índice centromérico e tamanho do cromossomo foram incluídas como entrada da rede. Cromossomos muito tortos e cromossomos sobrepostos não foram usados nos testes com ambas as redes neurais propostas.

Detcheva [19] desenvolveu o sistema CAIS (*Chromosome Analysis Intelligent System*) que tem por objetivo realizar a análise e a classificação cromossômica e a descoberta de anomalias no cariótipo. O sistema é composto das seguintes tarefas: pré-processamento, segmentação dos cromossomos, extração das características e classificação. As etapas de pré-processamento e segmentação utilizaram filtros de suavização, limiarização e detecção do contorno. A separação dos cromossomos que se tocam, a remoção de algum tipo de artefato e a separação dos cromossomos sobrepostos são feitas manualmente. Ao final de cada processo, o usuário pode intervir para fazer correções nos resultados propostos pelo sistema. A proposta da classificação, apresentanda por Detcheva em [20], é incorporar o conhecimento e a experiência do citogeneticista na análise e classificação dos cromossomos, baseando-se no comprimento do cromossomo e no índice centromérico. No caso de ambiguidade na classificação, características como área e padrão de banda são também utilizadas. A classificação é baseada em regras heurísticas definidas pelo citogeneticista e representadas por procedimentos. Estes procedimentos têm como função ordenar os cromossomos pelo tamanho, dividi-los nos grupos de Denver e classificá-los adequadamente de acordo com as regras previamente definidas.

Lerner et al. em [21] propuseram um estudo para examinar o uso de redes neurais

e o uso de características baseadas no MAT (*Medial Axis Transformation*), tais como características extraídas do padrão de banda, índice centromérico e comprimento do cromossomo, para a classificação de cromossomos humanos. O sistema não propõe métodos de segmentação e separação de cromossomos sobrepostos. Duas propostas para o cálculo do eixo médio (MAT) foram apresentadas. A primeira, chamada "esqueleto", visa encontrar um eixo médio preliminar usando o método de propagação de fogo no campo [22]. A segunda é baseada na técnica de *piecewise linear* (PWL). Neste trabalho, foi usado uma rede neural com duas camadas e o algoritmo *backpropagation* para o treinamento da rede. Os resultados mostram que a proposta do "esqueleto", para encontrar o eixo médio, encontrou melhores resultados. A proposta do esqueleto também apresentou melhores resultados no processo de classificação, alcançando uma média de acerto de 98%.

Todesco [23] propôs o uso de uma rede neural artificial com uma função de base radial para classificar os cromossomos humanos nos 23 pares conhecidos, testando diferentes arquiteturas de rede neural. A classificação foi dividida em duas fases. Na primeira fase utilizou-se o comprimento do cromossomo, o índice centromérico e a área como parâmetro para a rede, classificando os cromossomos em 7 grupos, o chamado grupos de Denver. Na segunda fase, utilizou-se as saídas da rede do primeiro estágio, como entrada para segunda rede, que além destas entradas, usou mais 20 entradas extraídas do padrão de banda do cromossomo. Para testes, foram usados três bases de dados conhecidas: Copenhagem, Philadelphia e Edinburg. Os resultados obtidos para o processo de classificação foram comparados com o de outros trabalhos publicados, que utilizaram a mesma base de dados, sendo que este alcançou melhores resultados.

Keller et al. [24] desenvolveram um sistema de inferência difusa para classificar os cromossomos humanos 16, 17 e 18, que pertencem ao grupo de Denver E. O método de segmentação dos cromossomos proposto baseia-se em uma limiarização local seguida de uma rotulação de componentes conectadas. Mais tarde, Keller and Sjahputera [25] melhoram o sistema a fim de otimizar as regras de inferência difusa. O classificador de inferência difusa permite que um cromossomo seja classificado em mais de uma par de cromossomo, atribuindo, a cada cromossomo, graus que estabelecem a sua pertinência a cada par de cromossomos possível. As características usadas como entrada para o sistema

de inferência foram índice centromérico, densidade média normalizada do nível de cinza (GRAY) e seis características do padrão de banda (P1P, P2P, P3P, P1D, P2D, D1P). Os autores também desenvolveram uma rede neural para comparar os resultados. O sistema de inferência difusa alcançou uma média de acerto de 88.3% e a rede neural alcançou uma média de 86.7% de acerto. Os dados utilizados para testes não continham cromossomos sobrepostos. Além disso, os cromossomos utilizados para testes eram de indivíduos normais, sendo que as imagens foram manualmente processadas por dois citogeneticistas, eliminando assim casos de cromossomos que se tocam e possíveis ruídos que permanecem quando o processo de segmentação é feito automaticamente pelo sistema.

Agam and Dinstein [26] propuseram um sistema para separar cromossomos que se tocam e cromossomos sobrepostos baseando-se no contorno dos mesmos. Neste trabalho não é tratado o problema de classificação. Inicialmente, a imagem é pré-processada com o objetivo remover os ruídos. Em seguida, a imagem é segmentada em duas regiões, utilizando uma limiarização global. Em seguida um método para seguir contorno, com conectividade 4, é utilizado para encontrar o contorno dos objetos, detectando os pontos de interesse no contorno do cromossomo. Usando os pontos de interesse do contorno de cada cromossomo, são traçadas linhas ligando estes pontos, para que seja escolhida, em seguida, a melhor linha que separa os cromossomos. Inicialmente são feitas todas as possíveis combinações de ligação dos pontos. Uma série de hipóteses é verificada e a linha que melhor segmenta o objeto é escolhida. A separação dos cromossomos falha quando existe uma grande área de contato entre cromossomos tortos e quando existe grandes agrupamentos contendo pequenos cromossomos.

Lerner em [27] propôs uma rede neural para automatizar a análise de imagens de cromossomos, usando características como tamanho, índice centromérico e 15 características extraídas do padrão de banda. Para a separação dos cromossomos que se tocam foi proposto um método chamado CPOOS (*classification-driven partially occluded object segmentation*). Este método consiste de três principais estágios. No primeiro é realizado o agrupamento dos *pixels* da imagem usando o método *k-means*. No segundo, os agrupamentos de cromossomos que se tocam são identificados porque seu tamanho é maior que a média e porque o sistema não consegue classificar este agrupamento em nenhuma

dos pares de cromossomos conhecidos. Então o contorno 8-conectado do agrupamento é encontrado. Os r pontos mais côncavos do contorno são selecionados e são sugeridos como pontos de corte, para que seja possível desenhar uma linha separando os cromossomos que se tocam. No último estágio, cada par de pontos de corte cria uma linha separando os cromossomos. Quando uma linha conectando um ponto de corte (i) a um ponto de corte (j) é hipotetizada, usando uma rede neural MLP, os dois segmentos criados por esta linha são cromossomos que se tocam. O método CPOOS, para separação dos cromossomos que se tocam, alcançou uma média de acerto de 90% na segmentação. No entanto, o método assume que existe somente um agrupamento de cromossomos que se tocam por imagem e que cada agrupamento é formado por apenas dois cromossomos. Para a classificação de cromossomos foi proposta uma rede neural MLP, usando duas estratégias. Na primeira, os cromossomos são primeiramente classificados nos 7 grupos de Denver e depois classificados dentro de cada grupo, usando as mesmas características como entrada para as duas rede. Na segunda estratégia, o classificador de grupo (grupo de Denver) utiliza somente as características geométricas (CI e tamanho do cromossomo) e o classificador de tipo utiliza a saída do primeiro classificador e mais 15 características do padrão de banda. O desempenho da rede é a média dos experimentos usando as duas estratégias, sendo que a segunda é utilizada como um método de validação cruzada. Os resultados mostraram um acerto de 83,6%.

Popescu et al. [28] desenvolveram o AKS (*Automated Karyotyping System*) que tem por objetivo classificar os cromossomos humanos com um mínimo de intervenção humana, processando automaticamente imagens que contém sobreposição de cromossomos. O pré-processamento da imagem consiste da aplicação de uma limiarização, remoção de ruídos e rotulação de componentes conectadas. A separação dos cromossomos que se tocam foi feita baseada na técnica do caminho mais claro (*pale path*). Nesta técnica admite-se que existe um caminho separando os cromossomos que se tocam, ou seja, existe uma região de fronteira. No entanto esta técnica falha em casos onde os cromossomos estão muito próximos. Visando resolver este problema, os autores propuseram a técnica CSSG (*Cross Section Sequence Graphs*) para separar os cromossomos que estão muito próximos ou sobrepostos, através da extração do eixo do cromossomo. A classificação utiliza o

comprimento, área e 18 características baseadas na WDD (Weighted Density Distribution) [29]. Além disso, a classificação utiliza redes neurais *multilayer feedforward*, a fim de encontrar uma associação de combinações de primitiva, que formam o esqueleto, com um dos cromossomos dos 23 pares de cromossomos. A técnica do CSSG, utilizada no sistema AKS, falha quando o cromossomo apresenta uma região mais clara. Também, em alguns casos de sobreposição de cromossomos, o esqueleto obtido não é capaz de representar dois cromossomos, causando problemas na separação.

Visando resolver o problema de sobreposição Graham et al. [30],[31] desenvolveram dois trabalhos. No primeiro trabalho [30], os autores propuseram um sistema para a resolver a sobreposição de cromossomos, que leva em consideração a classificação, usando o padrão de banda do cromossomo. Quando há uma sobreposição de cromossomos, o padrão de banda na região da sobreposição é obscurecido. No entanto, quatro pequenos segmentos dos cromossomos tornam-se visíveis. Cada um destes segmentos pode unificar com um dos outros três, formando seis possíveis cromossomos completos. Assim, a classificação é desenvolvida unificando as seções não obscurecidas do padrão de banda com um conjunto de modelos previamente selecionado, chamado PCM (*Partial Chromosome Models*). O PCM consiste de um conjunto de modelos que correspondem a segmentos do cromossomo, que possuem padrões de banda com uma boa discriminação entre os pares de cromossomo. Os segmentos candidados são comparados com os modelos usando uma função de classificação quadrática.

No segundo trabalho, Graham et al. [31] propuseram um método para resolver o problema de sobreposição, que leva em consideração a classificação e a forma do cromossomo. O método usado na classificação é o mesmo descrito no trabalho anterior. O modelo usado para encontrar a forma do cromossomo é baseado no PDM (*Point Distribution Models*), que gera um vetor representando a forma do cromossomo. Este vetor é unificado com um conjunto de dados treinados. Então, usando as evidências da classificação e da forma do cromossomo, é possível resolver a incerteza presente na região da sobreposição com mais confiança, do que usando somente uma das evidências. A limitação do método está no fato de que ele trabalha com agrupamentos contendo somente dois cromossomos sobrepostos.

Cho em [32] propõe um estudo sobre a capacidade de uma rede neural ser usada

como um classificador de cromossomos humanos. Foram fornecidas três características do cromossomo para a rede neural: comprimento, índice centromérico e padrão de banda. O comprimento, o índice centromérico e as características do padrão de banda foram encontrados usando o eixo médio (MAT). Para encontrar o eixo médio, um novo método, o qual modifica o algoritmo de afinamento proposto por Gonzales em [33], foi apresentado. A rede neural proposta possui duas camadas e usa o *backpropagation* como algoritmo de treinamento. A classificação com erro mínimo foi obtida usando 27 nós na camada de entrada (um para comprimento, um para índice centromérico e 25 para a densidade) e 24 nós na camada escondida. Um dos problemas do sistema é que o algoritmo de localização do centrômero não pode ser aplicado a cromossomos do grupo D e G, porque o braço menor destes cromossomos é muito pequeno. Outro problema é que o algoritmo usado para treinamento da rede exige muito tempo de processamento.

Pantaleão [11] propôs o SCACC (Sistema Computadorizado para Análise e Classificação Citogenética) que tem por objetivo reconhecer automaticamente os cromossomos humanos e analisar a leucemia pro-mielocítica aguda. Após ser feita a aquisição da imagem, o usuário pode aplicar algumas funções de pré-processamento como compensação de brilho, contraste, saturação e matiz. Em seguida o sistema aplica um filtro de suavização baseada na mediana de tamanho 3x3 e uma operação de limiarização, baseada no histograma da imagem. A detecção das bordas dos objetos é feita baseada no operador Laplaciano 3x3, permitindo assim a segmentação dos cromossomos. Para o reconhecimento automático dos cromossomos foi utilizado o enfoque lógico-combinatório, baseando-se no conceito de analogia. Características extraídas do padrão de banda, a área e a compacidade foram usadas no processo de classificação dos cromossomos. Para a detecção da leucemia pro-mielocítica foi desenvolvido um novo método de prolongamento das extremidades do eixo médio dos cromossomos, baseado na técnica de extrapolação pelos mínimos quadrados. O índice de acerto do sistema SCAAC, no processo de classificação, é baixo, quando comparado com outros sistemas. Problemas como a sobreposição de cromossomos e a localização do centrômero não são tratados no trabalho. Além disso, a rotina de rotação dos cromossomos apresenta limitações, visto que a rotação é feita baseando-se somente nos valores máximos e mínimos nos eixos x e y.

Foram propostos outros sistemas de menor abrangência, que não incluem separação de cromossomos que se tocam ou sobrepostos, nem detecção do centrômero ou classificação de forma automática. Dentre estes podemos citar o Micro Measure [34], que tem como objetivo auxiliar os citogeneticistas no processo de medição do tamanho dos cromossomos. No entanto, este sistema não possui métodos de pré-processamento e segmentação automáticos. Usando o sistema, o usuário marca interativamente as extremidades do cromossomo e alguns pontos que representam a curvatura do braço do cromossomo, bem como a posição do centrômero. Com base nestas informações o sistema calcula o índice centromérico e o tamanho do cromossomo, possibilitando que cada cromossomo seja rotulado. Apesar do sistema não apresentar processamentos automáticos, este já apresenta um avanço em relação ao processo manual de medição do tamanho do cromossomo e cálculo do índice centromérico.

Com o objetivo de auxiliar o pesquisador na elaboração de sistemas mais precisos de análise da constituição cromossômica de peixe, Bressane-Neto et al. [35] propuseram um sistema computacional gerador de cariótipos de peixe. Na fase inicial o usuário deve selecionar a foto e calibrar o padrão de cores a ser utilizado. Em seguida, o sistema faz a segmentação dos cromossomos, utilizando uma função recursiva. Após a segmentação, o usuário, interativamente, faz a rotação dos cromossomos e a classificação segundo a relação dos braços. Para encontrar a relação dos braços, o usuário deve previamente identificar a posição do centrômero e das extremidades dos braços, identificando pontos necessários para contemplar a curvatura do cromossomo. Para a montagem do cariótipo final os cromossomos são posicionados aos pares de acordo com o seu tamanho e morfologia. A principal restrição do sistema é que tarefas como localização do centrômero, cálculo do tamanho dos braços e a classificação dos cromossomos, são feitas interativamente, pelo usuário. Além disso o sistema não trata problemas de sobreposição ou de separação dos cromossomo que se tocam.

O sistema comercializado pela Applied Imaging, o Cytovision [36], pode ser usado para montar o cariótipo de diferentes espécies. Mas este sistema requer como entrada um *template* definido pelo usuário. Assim, o Cytovision não pode ser usado para montar o cariótipo de peixe porque não é possível montar este *template* previamente, visto que o

número de cromossomos e o número de pares de cromossomos não é conhecido à priori.

A Tabela 2.1 resume os sistemas de apoio à montagem de cariótipo e destaca o método de segmentação utilizado, se o sistema possui (ou não) procedimento automático para separação de cromossomos sobrepostos ou que se tocam, o método de classificação e o conjunto de características utilizado e a sua aplicação.

Sistema	Método Segmentação	Separação dos Cromossomos Sobrepostos/ que se tocam	Método de Classificação	Características usadas na Classificação	Aplicação
Ledley et al. [16]	Limiarização e contorno	Não possui	Reconhecimento de padrões sintaticamente dirigido	Área, comprimento e relação entre os braços	Humanos
Van Vliet et al. [17]	Shadding Correction, Threshold, operações binarias	Manual	Método bayesiano. Usa uma base de dados treinada	Padrão de banda, CI e comprimento	Humanos
Jennings and Graham [18]	Não possui	Não possui	Rede Neural: Multi-layer perceptron e Kohonen self-organizing	Padrão de banda, CI e comprimento	Humanos
Detcheva [20],[19]	Filtro, limiarização e contorno	Manual	Regras heurísticas definidas pelo citogeneticista	Padrão de banda, CI, área e comprimento	Humanos
Lerner et al. [21]	Não possui	Não possui	Rede Neural MLP	Características baseadas no MAT: CI, padrão de banda e comprimento	Humanos
Todesco [23]	Não possui	Não possui	Rede Neural com função de base radial	CI, comprimento, área e 20 características extraídas do padrão de banda	Humanos

Sistema	Método Segmentação	Separação dos Cromossomos Sobrepostos/ que se tocam	Método de Classificação	Características usadas na Classificação	Aplicação
Keller et al. [24], [25]	Limiarização local e detecção de objetos utilizando componentes conectados	Não possui	Sistema de Inferência Difusa, sendo que este foi comparado com uma rede neural	CI, densidade média normalizada do nível de cinza (GRAY) e seis características do padrão de banda	Humanos
Agam e Dinstein [26]	Limiarização global e método seguidor de contorno	Traçar linhas ligando pontos de interesse do contorno e escolher a melhor linha que separa os cromossomos	Não possui	Não possui	Humanos
Lerner [27]	Separar os cromossomos que se tocam baseado na técnica CPOOS	Não existe	Rede Neural MLP utilizando duas estratégias	CI, comprimento e 15 características do padrão de banda	Humanos
Popescu [28]	Limiarização, remoção de ruído e rotulação de regiões conectadas	Técnica do caminho mais claro e CSSG	Rede Neural multilayer feedforward	Comprimento, área e 18 características baseadas na WDD (Weighted Density Distribution)	Humanos

Sistema	Método Segmentação	Separação dos Cromossomos Sobrepostos/que se tocam	Método de Classificação	Características usadas na Classificação	Aplicação
Graham [30],[31]	Não possui	Classificação usando a técnica do PCM e a forma do cromossomo (PDM)	Utiliza base de dados do PCM	Padrão de banda	Humanos
Cho [32]	Não possui	Não possui	Rede Neural usando o back-propagation	Comprimento, CI e 25 características da padrão de banda	Humanos
Pantaleão [11]	Filtro mediano, limiarização, operador laplaciano	Não possui	Enfoque lógico-combinatório	Área, compactidade, Características extraídas do padrão de banda	Humanos
Micro Measure [34]	Não possui	Não possui	Não possui	Não possui	Qualquer organismo
Bressane-Neto [35]	Algoritmo recursivo	Não possui	Manual	Manual	Peixe

Tabela 2.1: Resumo dos principais sistemas existentes para montagem de cariótipo

2.8 Considerações Finais do Capítulo

A montagem de cariótipos é uma tarefa importante na rotina dos laboratórios de citogenética. Vários cientistas têm tentado automatizar este processo de modo a torná-lo mais rápido e menos sujeito a erro.

Visando atender a necessidade dos laboratórios de citogenética, que trabalha no estudo do cariótipo de peixe, propomos o SiMCaPe - Sistema para Montagem de Cariótipos de Peixe. Este sistema tem como objetivo tornar a montagem de cariótipos de peixe uma tarefa mais rápida, fácil e com menores chances de erros. Ao final de cada etapa do sistema, o usuário pode alterar manualmente os resultados propostos pelo sistema, caso julgue necessário.

Apesar de existirem muitos sistemas que montam o cariótipo humano, estes não podem ser aplicados para peixe por três razões principais: i) a maioria destes sistemas utilizam padrão de banda como uma das características utilizadas no processo de classificação. Como os peixes não apresentam esta característica, torna-se inviável a utilização destes sistemas; ii) os peixes não possuem um número fixo de cromossomos. O número de cromossomos varia de acordo com espécie de peixe, e existem alguns casos, onde indivíduos da mesma espécie apresentam número de cromossomos diferentes. No humanos, existe um número fixo de cromossomos que é 46. iii) nos humanos existem 46 cromossomos divididos em 22 pares e 1 par relacionadas ao sexo. Portanto, um indivíduo normal possui 23 pares de cromossomos, sendo que as características de cada par são previamente conhecidas. Nos peixes, como o número de cromossomos varia, não existe um número fixo de pares de cromossomos. Além disso as características de cada par não são previamente conhecidas.

Para montar o cariótipo de peixe utilizando o SiMCaPe, o processo é dividido em quatro etapas principais: pré-processamento da imagem e segmentação dos cromossomos, extração das características dos cromossomos, classificação dos cromossomos em uma das quatro classes pré-definidas e finalmente pareamento. Além destas funcionalidades, o sistema possibilita o armazenamento do cariótipo em um Banco de Dados, para futuras consultas ou alterações e disponibiliza uma interface gráfica que permite escolher o melhor cariótipo, que representa uma população em estudo. Os próximos capítulos apresentarão

em detalhes cada uma das etapas do sistema sendo proposto bem como os resultados encontrados.

Capítulo 3

Segmentação dos Cromossomos e Extração de Características

3.1 Introdução

As primeiras tarefas do processo de montagem de cariótipos diz respeito à segmentação da imagem e extração das características dos cromossomos. Em seguida, o processo consiste em classificar os cromossomos e pareá-los, montando assim o cariótipo final.

A segmentação da imagem é uma das tarefas mais difíceis da montagem do cariótipo, devido ao fato de que as imagens não apresentam boa qualidade e os cromossomos às vezes estão muito próximos uns dos outros ou mesmo sobrepostos.

Este capítulo tem como objetivo apresentar o processo de segmentação da imagem, utilizando conjuntos difusos, e o processo de extração das características de cada cromossomo, que estão integrados ao SiMCaPe.

3.2 Visão Geral da Fase de Segmentação e Extração de Características

Para iniciar o processo de montagem do cariótipo de peixe, o sistema SiMCaPe deve receber como entrada uma fotomicrografia digitalizada de uma célula em metáfase. A

fotomicrografia pode estar no formato JPG, BMP, GIF ou TIF. Após receber a imagem, o primeiro passo do sistema consiste em pré-processar a imagem, a fim de melhorar o contraste entre cromossomos e fundo, e eliminar ruídos.

Em seguida inicia-se o processo de segmentação da imagem, cujo objetivo é separar os cromossomos do fundo da imagem. No entanto alguns cromossomos não podem ser separados nesta etapa porque estão se tocando ou estão sobrepostos. Um procedimento é aplicado à imagem a fim de separar os cromossomos que se tocam. No caso de cromossomos sobrepostos, o SiMCaPe dispõe de uma interface gráfica para que o citogeneticista os separe interativamente.

Depois de segmentados, o próximo passo consiste em rotacionar os cromossomos para a posição vertical a fim de facilitar a localização do centrômero. Para a rotação dos cromossomos, primeiramente calcula-se o esqueleto do cromossomo, em seguida o eixo principal do esqueleto é detectado e então o cromossomo é rotacionado.

Após a rotação, um procedimento para localização do centrômero é executado. Com base na posição do centrômero, o tamanho dos braços dos cromossomos são medidos e as variáveis AR (razão entre os braços) e CI (índice centromérico) são calculadas, veja seção 2.3. Além destas, outras características são também calculadas como perímetro, tamanho e número de cromossomos.

A Figura 3.1 mostra uma visão geral da fase de segmentação e extração de características. As próximas seções detalham cada uma das etapas envolvidas nesta fase do processamento.

3.3 Pré-processamento da imagem

A imagem da célula de peixe no estágio da metáfase, em geral, é uma imagem bastante ruidosa e com baixo contraste. O fundo da imagem aparece como uma região mais clara e os cromossomos aparecem como regiões mais escuras. É comum a ocorrência de alguns artefatos na imagem, que são regiões escuras da imagem, que não são cromossomos, e que devem ser eliminados.

A aplicação direta de um método de limiarização para pré-processar a imagem resulta

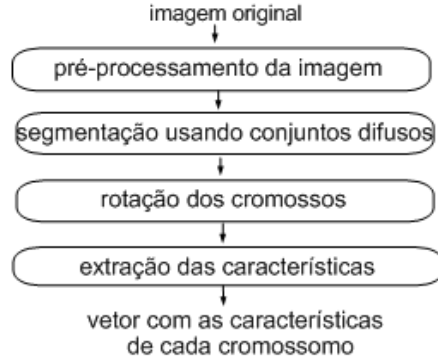


Figura 3.1: Visão geral das etapas de segmentação e extração de características

em perda de partes do cromossomo ou agregação de cromossomos que estão próximos, passando a representar uma só região.

O objetivo da etapa de pré-processamento é melhorar o contraste da imagem de entrada e eliminar os ruídos do fundo. Para realizar esta etapa, os seguintes passos são executados: aumento do contraste usando uma função quadrática, realce da imagem baseada em conjuntos difusos e limiarização e rotulação das regiões 8-conectadas.

3.3.1 Aumento de contraste

Inicialmente é aplicada à imagem original uma função que aumenta o contraste entre os cromossomos e o fundo (*background*), veja Equação 3.1. O objetivo desta função é tornar os *pixels* que pertencem aos cromossomos ainda mais escuros. Os *pixels* do fundo também se tornam mais escuros, mas numa proporção menor. O resultado da aplicação desta função é mostrado na Figura 3.2.

$$p_j = \frac{(p_j)^2}{255}, \text{ onde } p_j \text{ é o } j^{th} \text{ pixel da imagem} \quad (3.1)$$

3.3.2 Realce da imagem baseada em conjuntos difusos

A fim de auxiliar na remoção do fundo da imagem, realçando os cromossomos, uma função de pertinência difusa é aplicada. O objetivo desta função é atribuir um grau de pertinência difusa a cada um dos *pixels* da imagem. Este grau de pertinência é um valor real que varia

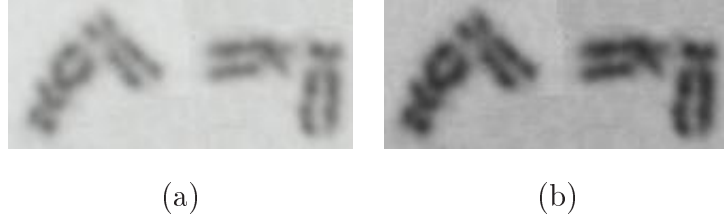


Figura 3.2: Resultado obtido pela aplicação da função quadrática expressa na Equação 4.2 (a) imagem original (b) imagem com aumento de contraste

no intervalo entre 0 e 1. Se o grau de pertinência de um *pixel* é igual a 1 significa que, com certeza, ele pertence a um cromossomo, se é igual a zero, com certeza este *pixel* não pertence ao cromossomo. Valores intermediários indicam o grau com que o *pixel* pertence ao cromossomo.

Para aplicar a função de pertinência difusa, o intervalo em níveis de cinza, que representa a região do cromossomo, deve ser estabelecido. Todos os *pixels* da imagem que estiverem entre o limite inferior e o limite superior do intervalo previamente definido pertencerão ao cromossomo com nível de pertinência 1. Os *pixels* que tiverem valor maior que o limite superior ou menor que o limite inferior, pertencerão ao cromossomo com um nível de pertinência difusa. A função de pertinência difusa com as características citadas acima é dada pela função mostrada na Equação 3.2.

$$\mu_1(p_j) = \begin{cases} 1, & \text{se } L_i \leq p_j \leq L_s \\ \frac{1}{1 + \beta|L_s - p_j|}, & \text{se } p_j > L_s \\ \frac{1}{1 + \beta|L_i - p_j|}, & \text{se } p_j < L_i \end{cases} \quad (3.2)$$

Esta função é baseada no trabalho de Guliato et al. [37] e definida como $\mu_1 : I \rightarrow [0, 1]$, onde I é a imagem, p_j é o j^{th} *pixel* da imagem I , L_i and L_s determinam o intervalo que caracteriza o centro do cromossomo em níveis de cinza e β determina a abertura da função de pertinência: valores altos resultam em uma função com comportamento severo e valores baixos resultam em uma função com comportamento permissivo.

O limite inferior L_i usado na equação 3.2 é zero, já que a região central dos cromossomos, na imagem com aumento de contraste, é mais escura que o fundo. O limite



Figura 3.3: Resultado da aplicação do procedimento de realce baseado em conjuntos difusos

superior L_s é calculado pelo sistema, mas pode ser alterado pelo citogeneticista caso seja necessário. Para encontrar o valor do limite superior, primeiramente o histograma da imagem realçada é calculado. O histograma da imagem com alto contraste, obtida no passo anterior, em geral, apresenta dois picos principais que indicam o fundo da imagem e os cromossomos. A partir destes picos, é possível encontrar um limiar, que é o menor ponto entre estes dois picos, usando o algoritmo de Kittler e Yan [38]. Este valor é usado como o limite superior L_s . Como nem todas as imagens apresentam claramente estes dois picos e o vale, em alguns casos o valor sugerido pelo sistema como limite superior pode não ser apropriado. Neste caso, o usuário pode alterar este valor interativamente.

Ao final do processo de realce, todos os *pixels* da imagem de alto contraste que tiverem grau de pertinência 1, receberão o nível de cinza correspondente ao da imagem original. Os demais *pixels* serão representados por valores de níveis de cinza inversamente proporcionais ao seu grau de pertinência difusa (quanto menor o grau de pertinência, mais alto o nível de cinza do *pixel*). A imagem resultante apresenta fundo próximo de branco e os cromossomos são representados de acordo com os níveis de cinza da imagem original. O resultado deste procedimento é mostrado na Figura 3.3.

3.3.3 Limiarização e rotulação da imagem

A Figura 3.3 mostra que a separação entre cromossomos e fundo é mais evidente, no entanto, o fundo ainda não é totalmente branco, visto que na função de pertinência difusa os *pixels* com baixo grau de pertinência receberam níveis de cinza alto, mas não possuem ainda o valor 255 (branco).

Para eliminar o fundo, inicialmente é aplicada sobre a imagem realçada uma limia-



Figura 3.4: Resultado da operação lógica AND entre a imagem binária e a imagem original

rização convencional a fim de obter uma imagem binária e separar definitivamente os cromossomos do fundo. O valor do limiar é obtido automaticamente pelo sistema utilizando o histograma da image realçada e o algoritmo de Kittler e Yan [38], sendo que o valor do limiar pode ser alterado pelo citogeneticista caso necessário.

Após obter uma imagem binária representada pelos valores 0 para o fundo e 1 para o cromossomo, o passo seguinte consiste em rotular as regiões 8-conectadas, que representam os cromossomos, e extrair as medidas estatísticas: média, desvio padrão e valor do menor nível de cinza. Estas medidas estatísticas serão importantes para o processo de separação dos cromossomos que se tocam. Note que cada região rotulada representa um cromossomo, mas devido ao fato de alguns cromossomos estarem muito próximos ou sobrepostos, uma região pode representar mais de um cromossomo. A separação dos cromossomos que se tocam será tratada na seção 3.4.1

Por último, é feito uma operação lógica AND, entre a imagem binária e a imagem original. A nova imagem terá o valor 255 onde na imagem binária for 0; e terá o valor do *pixel* da imagem original onde na imagem binária for 1. O resultado final é uma imagem original com o fundo representado pela cor branca (sem artefatos), mostrada na Figura 3.4.

3.4 Segmentação da imagem baseada no contorno

Após a obtenção da imagem pré-processada, onde os cromossomos encontram-se realçados e o fundo branco, inicia-se o processo de segmentação da imagem. O objetivo desta fase é segmentar os cromossomos baseado no contorno. Métodos para segmentar automaticamente cromossomos sobrepostos não são tratados neste trabalho. No entanto, o SiMCaPe

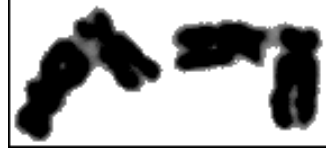


Figura 3.5: Resultado da aplicação da função de pertinência difusa definida na Equação 3.3

disponibiliza uma interface gráfica que permite ao citogeneticista separar os cromossomos sobrepostos.

Para o processo de segmentação, cada cromossomo é primeiramente representado por um conjunto difuso que preserva a informação da região de transição entre cromossomos e o fundo. A função de pertinência difusa, mostrada na Equação 3.3 é aplicada a cada região 8-conectada (possivelmente um cromossomo), que foi rotulada na etapa de pré-processamento.

$$\mu_2(p_j) = \begin{cases} 1, & \text{se } mvR_k \leq p_j \leq \mu - \frac{\sigma}{2} \\ \frac{1}{1 + \beta |((\mu - \frac{\sigma}{2}) - p_j)|}, & \text{senão} \end{cases} \quad (3.3)$$

Nesta equação, mvR_k , μ and σ são o valor do menor nível de cinza, a média e o desvio padrão da região 8-conectada, respectivamente. A variável β é a abertura da função de pertinência difusa.

Neste processo, os pixels dentro e ao redor de cada cromossomo serão representados de acordo com seu nível de similaridade com respeito à característica da região central do cromossomo dada por $\mu - \frac{\sigma}{2}$, onde μ e σ são a média e o desvio padrão do cromossomo que está sendo segmentado. Assim, ao final do processo, os *pixels* com grau de pertinência 1 serão representados pelo nível de cinza 0 (preto) e os *pixels* com grau de pertinência 0 serão representados pelo nível de cinza 255 (branco). Os demais graus de pertinência serão representados por níveis de cinza inversamente proporcionais. A Figura 3.5 mostra o resultado deste processo. Note que o fundo é branco e as regiões de transição entre cromossomos e o fundo são claramente representadas.



Figura 3.6: Resultado da aplicação da limiarização

3.4.1 Separação dos cromossomos que se tocam

Dois cromossomos se tocam quando suas fronteiras possuem uma região em comum. Como a região de fronteira, representada por conjuntos difusos e mostrada na Figura 3.5, é mais clara que a parte central do cromossomo (condição garantida pelo processos anteriores), uma limiarização global pode ser executada para separar os cromossomos que se tocam, preservando a região central dos cromossomos.

Uma vez realizados vários testes, o valor do limiar global foi fixado no nível de cinza 50. A imagem resultante após a aplicação da limiarização é mostrada na Figura 3.6. A fim de eliminar os artefatos indesejados produzidos pelo processo de limiarização, um filtro mediano é aplicado na imagem.

Se no mínimo dois cromossomos são separados, todos os cromossomos devem ser rotulados novamente usando a imagem binária, resultante do aplicação do filtro. Durante a rotulação, um *pixel* semente é encontrado para cada região rotulada que será posteriormente utilizado para segmentar os cromossomos.

Em alguns casos, os cromossomos que se tocam não são separados pelo processo de segmentação. Para este caso, o SiMCPe disponibiliza uma interface gráfica que permite que o usuário separe os cromossomos que se tocam, bem como separe os cromossomos sobrepostos.

3.4.2 Segmentação dos cromossomos

O resultado do processo anterior de separação dos cromossomos que se tocam, produz uma imagem binária, onde os cromossomos são representados por *pixels* preto e o fundo por *pixels* branco. Agora, uma operação lógica AND é executada, entre a imagem resultante do processo de filtragem, Figura 3.7 (a) e a imagem obtida na etapa de pré-processamento,

seção 3.3.3, Figura 3.7 (b). Este procedimento elimina a área comum entre cromossomos adjacentes produzindo uma imagem em níveis de cinza que não possui cromossomos que se tocam, veja Figura 3.7 (c).

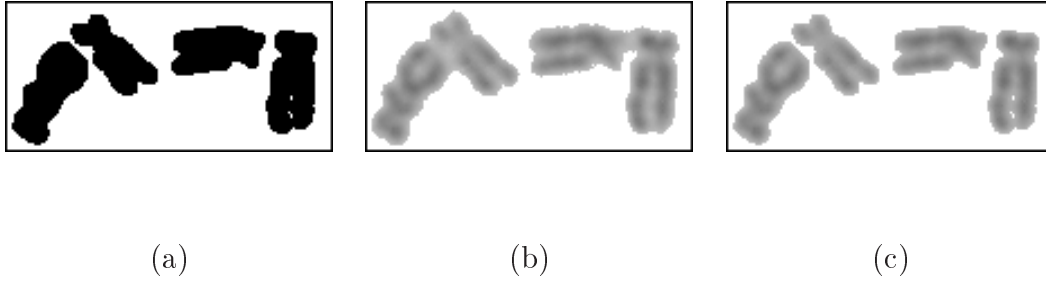


Figura 3.7: Operação lógica AND (a) imagem resultante do processo de filtragem (b) imagem resultante do pré-processamento (c) imagem resultante da operação lógica AND entre as imagens (a) e (b).

Agora, usando o *pixel* semente obtido anteriormente, seção 3.4.1, um algoritmo de crescimento de região é aplicado na imagem em níveis de cinza, para detectar o contorno de todos os cromossomos. O algoritmo parte de um *pixel* semente, obtido na etapa de rotulação, e analisa a vizinhança 8-conectada do *pixel*. Caso o valor do *pixel* vizinho seja menor que um limiar ele é colocado como pertencente à região e a sua vizinhança será analisada a seguir, gerando assim uma recursividade. Quando o valor do *pixel* vizinho for maior que um determinado limiar, o algoritmo pára, e marca este *pixel* pertencente ao contorno. De acordo com testes, o nível de cinza 50 para o limiar tem proporcionado bons resultados.

Em seguida, um pós-processamento é feito para eliminar alguns artefatos e garantir que o contorno tenha 1-*pixel* de largura. O contorno do cromossomo é usado para obter seu perímetro, que será usado para calcular os descritores de características CI e AR, bem como para obter os limites da região retangular de cada cromossomo, que são os valores máximos e mínimos nos eixos X e Y (minX, minY, maxX, maxY). A Figura 3.8, mostra o resultado do contorno sobreposto à imagem original.



Figura 3.8: Resultado do contorno sobreposto à imagem original

3.5 Rotação dos cromossomos

Antes de realizar a extração de características de cada cromossomo é necessário rotacioná-los para a posição vertical. O SiMCaPe disponibiliza um método para rotação automática dos cromossomos, que baseia-se na orientação do eixo central do seu esqueleto.

O processo de rotação utiliza como entrada a imagem binária obtida pelo processo de limiarização, apresentado na seção 3.4.1. Como os braços do cromossomos, em geral, estão disjuntos, antes de rotacionar os cromossomos é necessário fazer um preenchimento da região entre as cromátides, a fim de tornar o cromossomo uma região convexa. O preenchimento é feito na direção horizontal, vertical e diagonal. O objetivo deste preenchimento é tornar o cromossomo o mais convexo possível de forma que o seu esqueleto possua o mínimo de ramificações possíveis.

Após o preenchimento da região entre as cromátides, o esqueleto do cromossomo é calculado. O algoritmo usado para encontrar o esqueleto foi o de Holt [39], sendo que depois foi aplicado o algoritmo de *staircase removal* [39] a fim de obter um esqueleto com 1-*pixel* de largura. Na próxima seção será explicado em detalhes o método de Holt e o *staircase removal*.

Após obter o esqueleto, este pode apresentar algumas ramificações, o que dificulta a escolha do eixo principal a ser usado na rotação. Assim, um pós-processamento é feito no esqueleto a fim de eliminar ramificações extras e manter somente o eixo principal.

Depois de obter um cromossomo sem ramificações extras, uma reta é traçada ligando as extremidades do esqueleto. O ângulo de inclinação desta reta é o ângulo utilizado para realizar a rotação do cromossomo. A Figura 3.9 mostra os resultados do processo de esqueletonização e rotação para um cromossomo que possui esqueleto sem ramificações. A Figura 3.10 mostra os resultados do processo de rotação para um cromossomo que possui

o esqueleto com ramificações.

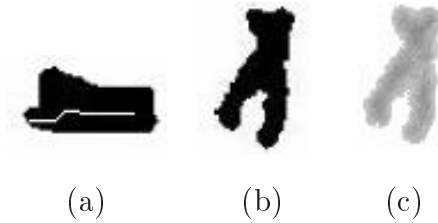


Figura 3.9: Resultado do processo de rotação (a) imagem binária com a região entre as cromátides preenchida e o esqueleto sobreposto (b) imagem binária rotacionado de acordo com o seu esqueleto (c) imagem original rotacionada de acordo com o seu esqueleto

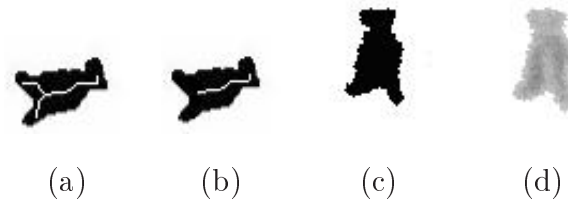


Figura 3.10: Resultado do processo de rotação para uma imagem que possui esqueleto com ramificações (a) imagem binária com a região entre as cromátides preenchida e o esqueleto sobreposto (b) imagem binária com o esqueleto sem ramificações (c) imagem binária rotacionado de acordo com o seu esqueleto (d) imagem original rotacionada de acordo com o seu esqueleto

Muitas vezes, não é possível eliminar as ramificações extras e encontrar o eixo principal. Nestes casos, os cromossomos são rotacionados incorretamente, e portanto, necessitam ser rotacionados com intervenção do usuário. Para estes casos, uma interface gráfica é disponível que permite ao citogeneticista rotacionar o cromossomo interativamente, informando ao sistema o valor do ângulo a ser utilizado.

3.5.1 Método de Holt

O processo de esqueletonização ou esqueletização consiste em diminuir o número de *pixels* que representam uma imagem, preservando sua forma original. Vários algoritmos de esqueletonização têm sido proposto, dentre eles podem citar: Algoritmo de Hilditch [40], Algoritmo de Stentiford [41], Algoritmo de Zhang-Suen [42] e Algoritmo de Holt [39].

Neste trabalho, os esqueletos produzidos pelos 4 algoritmos citados anteriormente foram comparados. Levando-se em consideração que os esqueletos obtidos foram semelhantes e que o algoritmo de Holt é mais rápido que os outros 3 algoritmos citados e não envolve iterações, este algoritmo foi escolhido.

O algoritmo de Holt utiliza duas expressões lógicas para eliminar ou não um pixel. Se o resultado das expressões lógicas for falso, o *pixel* deve ser eliminado, caso contrário, ele não deve ser eliminado. As expressões lógicas são mostradas nas Equações 3.4 e 3.5.

Nas equações 3.4 e 3.5, $v()$ representa o valor do *pixel* e retorna verdadeiro se o *pixel* pertence ao cromossomo e falso caso contrário, $edge()$ indica verdadeiro se o *pixel* estiver na borda do objeto e falso caso contrário. Um *pixel* pertence à borda se sua conectividade for 1 e se possui de 2 a 6 vizinhos pretos. A conectividade indica o número de transições de um ponto preto para um branco no sentido horário da vizinhança 8-conectada de um *pixel*. As letras N, S, L e O representam a vizinhança 3 x 3 do *pixel* central C, conforme mostra a Figura 3.11

$$v(C) \wedge (\neg edge(C) \vee (v(L) \wedge v(S) \wedge (v(N) \vee v(O)))) \quad (3.4)$$

$$v(C) \wedge (\neg edge(C) \vee (v(O) \wedge v(N) \wedge (v(S) \vee v(L)))) \quad (3.5)$$

As duas expressões lógicas, Equação 3.4 e 3.5, foram combinadas por Holt, gerando a expressão mostrada na Equação 3.6.

$$v(C) \wedge (\neg edge(C) \vee (edge(L) \wedge v(N) \wedge v(S)) \vee (edge(S) \wedge v(O) \wedge v(L)) \vee (edge(L) \wedge edge(SE) \wedge edge(S))) \quad (3.6)$$

NO	N	NE
O	C	L
SO	S	SE

Figura 3.11: Vizinhança de um pixel C, usada no algoritmo de Holt

0	1	x
1	1	x
x	x	0

máscara 1

x	1	0
x	1	1
0	x	x

máscara 2

0	x	x
x	1	1
x	1	0

máscara 3

x	x	0
1	1	x
0	1	x

máscara 4

Figura 3.12: Máscaras utilizadas no processo de *staircase removal*

3.5.2 Método de *Staircase Removal*

A fim de melhorar o resultado da Equação 3.6, Holt propôs o método do *staircase removal*. O método do staircase usa a propriedade: metade dos *pixels* que apresentam uma forma semelhante a uma escada podem ser removidos sem afetar o formato ou a conectividade do objeto. O ponto central das máscaras, mostradas na Figura 3.12, podem ser apagados se um dos valores X for 0 (branco).

3.6 Extração das características do cromossomo

Depois de separar os cromossomos e rotacioná-los, inicia-se a fase de extração das características do cromossomo. Estas características serão usadas nos processos de classificação e pareamento dos cromossomos.

A primeira característica extraída é a largura e altura do cromossomo. Em seguida é calculado o perímetro através da contagem do número de *pixels* que pertencem ao contorno do cromossomo.

No entanto, a característica mais importante extraída do cromossomo é a posição do centrômero, pois a partir dela será calculado o índice centromérico e a razão entre braços,

que serão utilizadas como parâmetros de entrada para o processo de classificação. A próxima seção mostra em detalhes o algoritmo para localização do centrômero.

3.6.1 Localização do centrômero utilizando um vetor de projeção

A localização do centrômero é feita com base em um vetor de projeção horizontal. Este vetor de projeção é obtido analisando cada linha da imagem e somando o número de *pixels* que pertencem ao cromossomo, ao longo desta linha. O uso do vetor de projeção, com este propósito, foi primeiramente proposto por Moradi et al. [43], que aplicou este método para localizar o centrômero em cromossomos humanos. O vetor de projeção, como o proposto por Moradi et al. [43] não funciona bem quando aplicado diretamente a cromossomos de peixe, visto que nos peixes as cromátides estão separadas.

Para localizar o centrômero do cromossomos de peixe, usando a proposta do vetor de projeção, inicialmente cada imagem binária rotacionada, deve ter a região entre as cromátides preenchidas novamente, na direção horizontal. Este processo torna cada cromossomo uma região convexa, preservando sua morfologia externa. Depois de analisar cada linha da imagem binária do cromossomo, o vetor de projeção resultante pode ser visto como uma *skyline* que possui dois ou mais picos e alguns vales, como mostra a Figura 3.13.

A posição do centrômero é o ponto mais baixo localizado na maior bacia. Para definir a maior bacia da *skyline*, um algoritmo baseado na transformada de *watershed* foi usado [44]. Este algoritmo é bem sucedido quando os cromossomos são metacêntricos, submetacêntricos e subtelocêntricos. Para a localização automática do centrômero dos cromossomos acrocêntricos, o método proposto não apresenta bons resultados. Neste caso, o citogeneticista deve marcar interativamente a posição do centrômero. A Figura 3.14 mostra três diferentes classes de cromossomos com a posição do centrômero sobreposta à imagem.

3.6.2 Cálculo do índice centromérico e razão entre braços

A partir da posição do centrômero, duas importantes medidas podem ser extraídas dos cromossomos: o índice centromérico (CI) e a razão entre braços (AR). Estas duas medidas

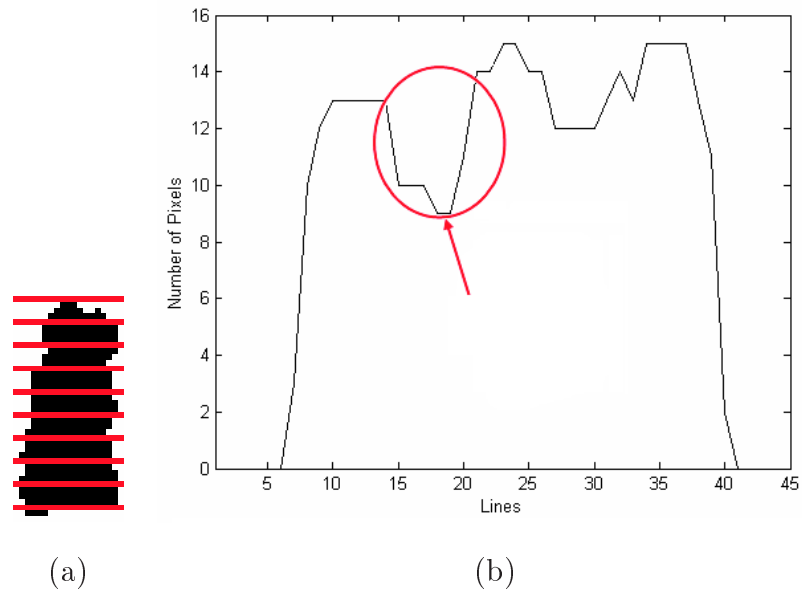


Figura 3.13: Cálculo do vetor de projeção (a) imagem do cromossomo depois que a região entre as cromátides foi preenchida (b) vetor de projeção usado para escolher a posição do centrômero

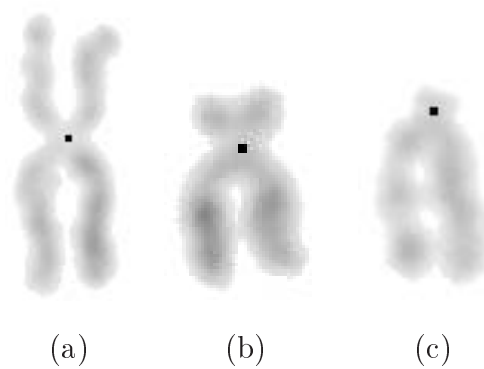


Figura 3.14: Resultado da localização do centrômero (a) cromossomo metacêntrico; (b) cromossomo submetacêntrico e (c) cromossomo subtelocêntrico

serão utilizadas nos processos de classificação e pareamento dos cromossomos.

Formalmente o cálculo do índice centromérico é dado pela Equação 2.2 e o cálculo da razão entre braços é dado pela Equação 2.1, apresentadas na seção 2.3. Como o processo para medir os braços do cromossomo é bastante subjetivo o sistema SiMCPe propõe calcular o CI e o AR da seguinte forma: i) Para o CI, o *comprimento(p)* foi calculado usando o perímetro da parte superior à posição do centrômero, incluindo os dois braços do cromossomo, e o *tamanho do cromossomo* foi calculado usando o perímetro total do cromossomo ii) Para o AR, o *comprimento(q)* foi calculado usando o perímetro do maior braço maior do cromossomo e o *comprimento(p)* foi calculado usando o perímetro do maior braço menor do cromossomo.

3.7 Considerações Finais do Capítulo

Este capítulo apresentou em detalhes o processo de segmentação e extração de características. Para o processo de segmentação, foram utilizados métodos baseados em conjuntos difusos que permitem representar a região de transição entre cromossomos e foi utilizado também o contorno dos cromossomos. Para a extração de características dos cromossomos, estes foram previamente rotacionados utilizando a orientação do eixo principal do seu esqueleto. Para a extração da posição do centrômero foi utilizado um vetor de projeção horizontal do cromossomo. A partir da posição do centrômero, as medidas CI e AR foram calculadas.

Agora que os cromossomos estão separados e suas características foram extraídas, o próximo passo consiste em classificar os cromossomos e pareá-los, montando assim o cariótipo final da espécie. O próximo capítulo apresentará os métodos utilizados para a classificação e pareamento de cromossomos de peixe. Também será mostrado o processo para a escolha do melhor cariótipo, de uma dada população em estudo.

Capítulo 4

Classificação e Pareamento

4.1 Introdução

A partir do momento que as características foram extraídas de cada cromossomo, inicia-se a fase de reconhecimento, que é composta das tarefas de classificação e pareamento.

As tarefas de classificação e pareamento são bastante subjetivas, sendo que o citogeneticista necessita ter experiência anterior para conseguir executá-las. A função do SiMCaPe, nesta etapa, é transformar o conhecimento do citogeneticista em regras que permitam classificar os cromossomos e pareá-los de acordo com a sua similaridade. Além disso, ao final do estudo de uma dada população é necessário escolher o melhor cariótipo, ou seja, um cariótipo que está associado a uma metáfase e que melhor representa aquela população. Então torna-se necessário o desenvolvimento de uma interface que auxilie o citogeneticista a tomar esta decisão.

O objetivo deste capítulo é apresentar o desenvolvimento da etapa de classificação utilizando um sistema de inferência difusa e da etapa de pareamento.

4.2 Visão geral das etapas de classificação e pareamento

Uma vez extraídas as características de cada cromossomo, torna-se necessário agrupar estas informações a fim de que o cromossomo possa ser classificado. Para a classificação, em geral, os citogeneticistas têm utilizado as medidas razão entre braços (AR) e índice

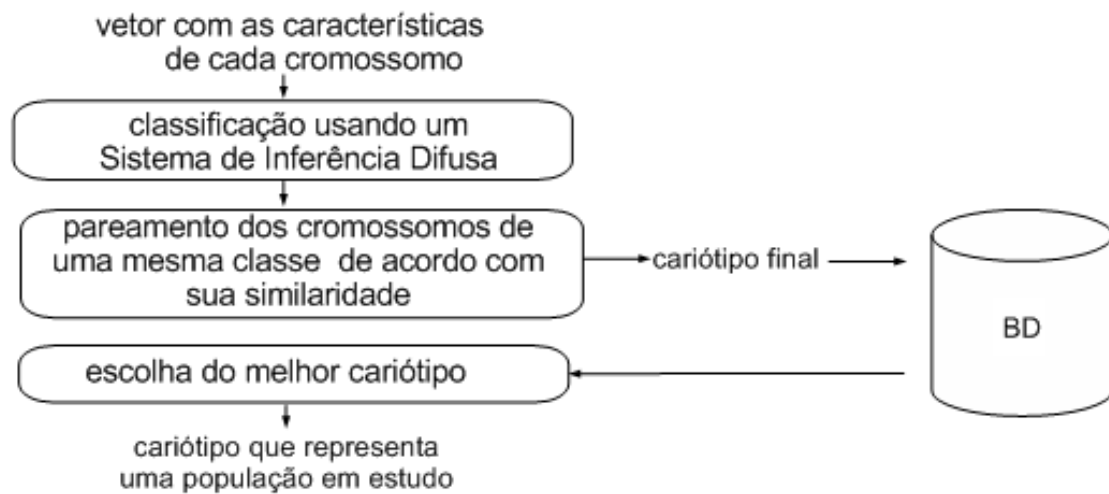


Figura 4.1: Visão geral das etapas de classificação e pareamento

centromérico (CI). A partir destas medidas, o SiMCaPe disponibiliza um sistema de inferência difusa, que utiliza o método de Mamdani, a fim de classificar cada cromossomo em uma das 4 classes conhecidas: metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico.

Após serem classificados, os cromossomos da mesma classe devem ser pareados de acordo com a sua similaridade. Uma característica importante no pareamento é o tamanho do cromossomo. Depois de pareados, obtêm-se o cariótipo final da espécie. Este cariótipo pode ser salvo em um banco de dados para posteriores consultas ou alterações.

Depois de montar o cariótipo de vários indivíduos de uma mesma população, o citogeneticista deve montar o cariótipo que represente esta população sendo estudada. Para isso o SiMCaPe disponibiliza uma interface gráfica onde o citogeneticista interativamente escolhe o melhor cariótipo.

Uma visão geral das etapas de classificação e pareamento é mostrado na Figura 4.1.

4.3 Classificação

O processo de classificação proposto neste trabalho é baseado no sistema de inferência difusa de Mamdani [45] e utiliza como entrada as variáveis CI e AR e retorna como saída

quatro valores que representam o grau de pertinência difusa com que um dado cromossomo pertence a cada uma das classes: metacêntrico, submetacêntrico, subtelocêntrico e acrocêntrico. O cromossomo será atribuído à classe que apresenta o maior grau de pertinência difusa.

4.3.1 Sistema de Inferência Difusa: visão geral

Um dos primeiros sistemas de controle usando a teoria difusa foi desenvolvido por Mamdani e Assilam [45], cujo objetivo era controlar uma máquina a vapor, a fim de automatizar o trabalho de um operador humano. A proposta sugeria que o trabalho do operador poderia ser feito por um sistema, usando a lógica difusa, transformando a experiência do operador em um conjunto de regras. O trabalho de Mamdani e Assilam apresentou bons resultados e hoje, os sistemas de inferência difusa têm sido aplicados em muitas áreas: [46],[47], [48], [49] e [50].

Em geral, os sistemas de inferência difusa são sistemas especialistas, que empregam uma base de conhecimento, expressa em termos de regras de inferência difusa, e um motor de inferência para resolver um dado problema. Este sistemas mapeiam duas ou mais variáveis lingüísticas de entrada, representadas no domínio da aplicação, em uma ou mais saídas que representam graus de pertinência difusa. As regras do sistema de inferência difusa são regras do tipo *if-then*.

Um sistema de inferência difusa é composto de 4 módulos: fuzificação das entradas, uma base de regras difusas, um motor de inferência difusa e um módulo de defuzificação. A Figura 4.2 ilustra os módulos do sistema. A seguir serão detalhados cada um destes módulos.

É importante notar que o primeiro passo do desenvolvimento de um sistema de inferência difusa é identificar as variáveis lingüísticas de entrada e de saída, que são relevantes para o sistema, e o domínio destas variáveis. Para cada variável de entrada e de saída deve ser selecionado um conjunto de estados lingüísticos, que são expressos por números difusos apropriados. Estes números difusos representam rótulos lingüísticos, como por exemplo: *pequeno*, *médio* e *grande*. Assim é obtido uma quantização difusa das variáveis. Depois de definida as variáveis de entrada e seus estados lingüísticos, o conjunto de regras

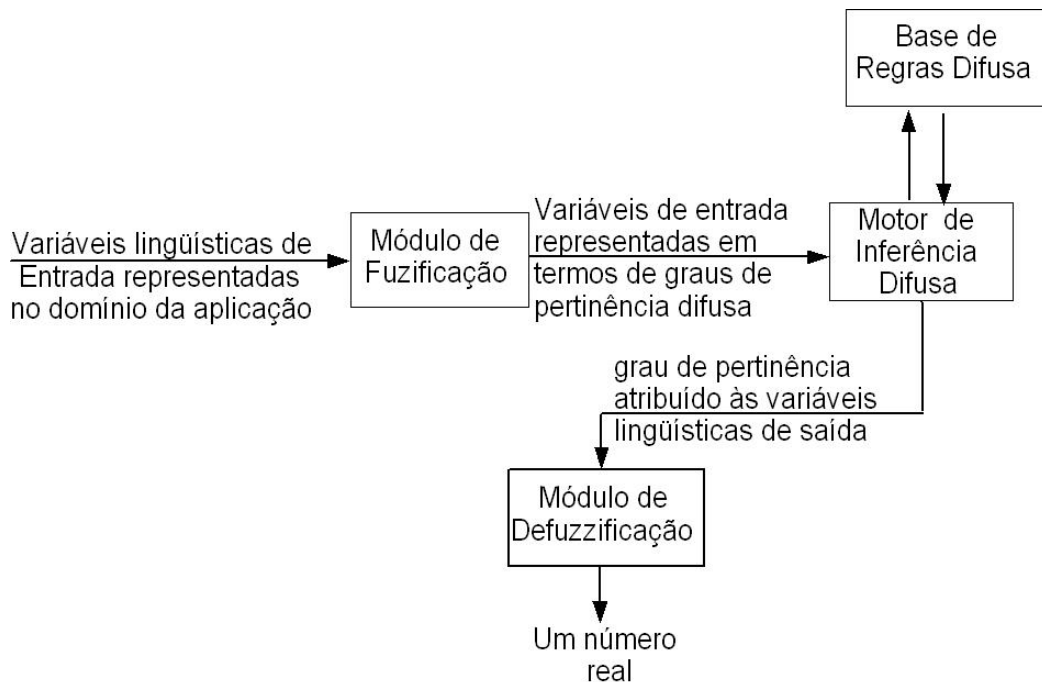


Figura 4.2: Módulos do sistema de inferência fuzzy

if-then pode ser modelado, em geral, usando a experiência do especialista.

Fuzzificação das Entradas

No módulo de fuzzificação as medidas obtidas para cada variável de entrada são convertidas em conjuntos difusos apropriados a fim de expressar as incertezas presentes nestas medidas. As medidas fuzzificadas são usadas pelo motor de inferência difusa para avaliar as regras difusas.

Assim, uma função de fuzzificação é introduzida para cada variável de entrada com o objetivo de interpretar as medidas destas variáveis, expressas por números reais, em uma aproximação difusa mais realística.

Em alguns sistemas de inferência difusa, as variáveis de entrada não são fuzzificadas, isto é, as medidas das variáveis de entrada são empregadas diretamente no processo de inferência como fatos.

Base de Regras Difusas

Neste módulo o conhecimento pertinente ao problema é formulado em termos de um conjunto de regras de inferência difusa. Estas regras podem ser obtidas pela experiência do especialista ou obtidas por métodos de aprendizagem, como por exemplo redes neurais.

O esquema geral das regras difusas, é ilustrado na Equação 4.1. O antecedente de cada regra é conectado pelos operadores AND ou OR. Na equação, $\mathcal{X}$, $\mathcal{Y}$ e $\mathcal{Z}$ são variáveis lingüísticas, cujos valores estão contidos nos domínios X , Y e Z respectivamente e A , B , C são números difusos [51] em X , Y , Z respectivamente, que representam predicados difusos (estados lingüísticos) como por exemplo: *muito frio*, *frio*, *morno* e *quente*.

$$\text{IF } \mathcal{X} \text{ is } A \text{ AND } \mathcal{Y} \text{ is } B \text{ THEN } \mathcal{Z} \text{ is } C \quad (4.1)$$

Se em um dado sistema existirem duas variáveis de entrada e cada variável for modelada por quatro estados lingüísticos, então o número total de possíveis regras não conflitantes é $4^2 = 16$. No entanto, na prática, somente um sub-conjunto destas possíveis regras é suficiente para produzir um desempenho aceitável do sistema de inferência difusa.

Motor de Inferência Difusa

Neste módulo as variáveis de entrada do sistema de inferência difusa, que foram fuzificadas, são combinadas apropriadamente com as regras de informação difusa para fazer inferências com respeito às variáveis de saída. O resultado deste processo é um conjunto difuso, que será então convertido em um número real pelo processo de defuzificação.

O método usualmente utilizado pelo motor de inferência difusa é o método da interpolação [51]. O método de interpolação é constituído de 2 passos. No primeiro passo é calculado a intersecção entre um dado fato e o antecedente de cada regra *if-then* do conjunto de regras difusas. O fato representa a medida da variável lingüística de entrada e é representado por um conjunto difuso. O antecedente da regra, em geral, utiliza dois termos e as funções $t - norm$ e $t - conorm$ para conectá-los. Então neste passo ocorrer a intersecção entre o fato e cada termo de cada regra difusa. Em seguida ocorre a aplicação das funções $t - norm$ e $t - conorm$ para resolver o antecedente da regra. No segundo

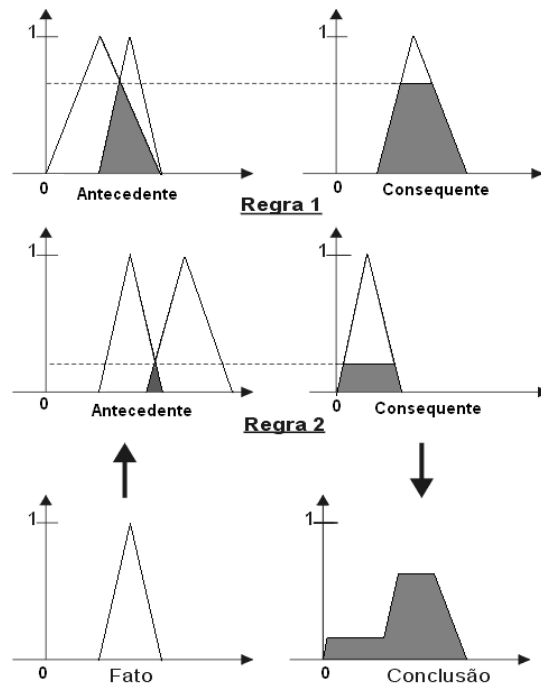


Figura 4.3: Módulo Motor de Inferência Difusa

passo o cálculo da conclusão do motor de inferência difusa é obtido pelo truncamento de cada conjunto difuso de saída das regras difusas, usando o valor obtido pelo primeiro passo. A união destes conjuntos difusos truncados produzem um único conjunto difuso representando cada variável de saída. A Figura 4.3 mostra um esquema do módulo Motor de Inferência Difusa usando o método de interpolação.

Defuzificação

O objetivo do módulo de defuzificação é converter cada conclusão obtida pelo motor de inferência difusa, que é expresso por um conjunto difuso, em um número real. Alguns métodos têm sido usados para o processo de defuzificação, como exemplo, podemos citar: centróide (centro de área), centro de máxima e média de máxima [51].

4.3.2 Sistema de Inferência Difusa usado na Classificação de Cromossomos de Peixe

Embora os sistemas de inferência difusa sejam freqüentemente usados em aplicações na engenharia, medicina e agricultura, poucos trabalhos têm sido desenvolvidos para a área de classificação de cromossomos [24] e [25].

O sistema de inferência difusa proposto para o SiMCaPe, é baseado no método de Mamdani e utiliza duas variáveis lingüísticas de entrada: índice centromérico (CI) e razão entre braços (AR). A variável lingüística de saída do sistema é o CL_i , que representa a classe do cromossomo, sendo que i assume o valor *metacêntrico*, *submetacêntrico*, *subtelocêntrico* ou *acrocêntrico*.

Os módulos do sistema de inferência difusa para classificação de cromossomos serão visto em detalhes nas próximas seções.

Representação das Variáveis de Entrada e Fuzificação

As duas variáveis de entrada do sistema de inferência difusa, AR e CI, são representadas por números reais e foram calculadas de acordo com as equações 2.1 e 2.2 da seção 2.3. O domínio da variável lingüística CI é o conjunto dos números reais contidos no intervalo entre 0 e 50. Como a variável CI representa a razão entre o braço menor e o comprimento total do cromossomo multiplicado por 100, temos que o braço menor do cromossomo pode ser no máximo a metade (0.5) do tamanho total do cromossomo, ou seja, a variável CI nunca assumirá valor maior que 50. O domínio da variável lingüística AR é o conjunto dos números reais no intervalo entre 1 e 10. Quando a razão entre o braço maior e o braço menor do cromossomo for maior que 10 a variável AR recebe o valor 10. Cada variável lingüística é representada pelo conjunto de estados lingüísticos: *{pequeno, médio, grande e muito grande}*.

O sistema também utiliza uma variável lingüística de saída, que indica a classificação do cromossomo. A variável lingüística de saída é representada pelo conjunto de estados lingüístico: *{pequena, média e grande}*.

A Figura 4.4 mostra os quatro números difusos, que representam cada um dos estados

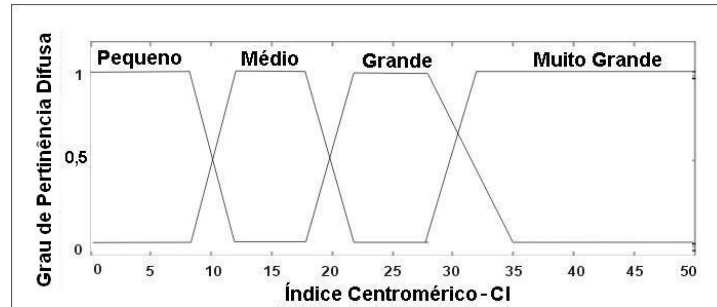


Figura 4.4: Números difusos que representam os estados lingüísticos da variável lingüística CI

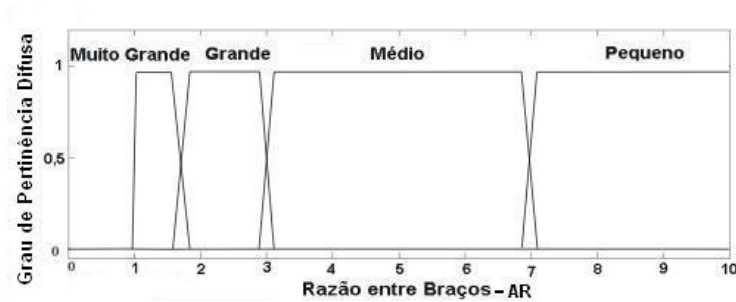


Figura 4.5: Números difusos que representam os estados lingüísticos da variável lingüística AR

do conjunto lingüístico da variável lingüística de entrada CI. A Figura 4.5 mostra os quatro números difusos, que representam cada um dos estados do conjunto lingüístico da variável lingüística de entrada AR. A Figura 4.6 mostra os três números difusos, que representam cada um dos estados do conjunto lingüístico da variável lingüística de saída CL_i .

No sistema SiMCaPe as variáveis de entrada não foram fuzificadas, ou seja, os valores das variáveis de entrada foram empregados diretamente no processo de inferência.

Base de Regras Difusas

Como existem duas variáveis de entrada e cada uma delas é modelada por quatro estados lingüísticos, o número de regras de inferência difusa não conflitantes é 16. No entanto somente o subconjunto de regras relevantes foi utilizado pelo SiMCaPe. Como existem

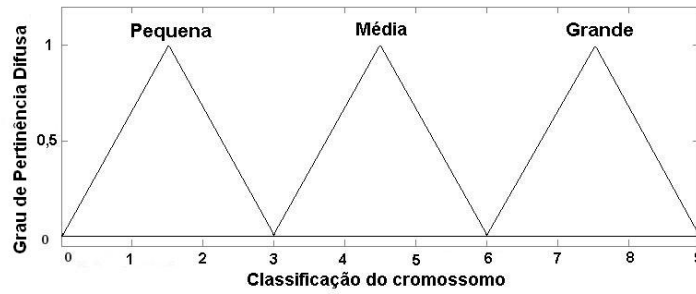


Figura 4.6: Números difusos que representam os estados lingüísticos da variável lingüística CL_i

quatro possíveis classificações para o cromossomo, foram utilizadas quatro conjuntos de regras de inferência difusa, um para cada classe do cromossomo. Cada conjunto de regras difusas será avaliado separadamente pelos módulos motor de inferência difusa e defuzificação. Ao final o sistema produz quatro resultados, que indicam o grau de pertinência do cromossomo a cada uma das quatro classes. Abaixo é mostrado cada conjunto de regras.

a) o conjunto de regras difusas para classificar o cromossomo na classe metacêntrico ($CL_{metacentrico}$)

regra 1: IF CI is muito grande AND AR is muito grande then $CL_{metacentrico}$ is grande

regra 2: IF CI is muito grande AND AR is grande then $CL_{metacentrico}$ is grande

regra 3: IF CI is grande AND AR is muito grande then $CL_{metacentrico}$ is grande

regra 4: IF CI is grande AND AR is grande then $CL_{metacentrico}$ is média

regra 5: IF CI is médio OR AR is médio then $CL_{metacentrico}$ is pequena

regra 6: IF CI is pequeno OR AR is pequeno then $CL_{metacentrico}$ is pequena

b) o conjunto de regras difusas para classificar o cromossomo na classe submetacêntrico ($CL_{submetacentrico}$)

regra 1: IF CI is grande AND AR is grande then $CL_{submetacentrico}$ is grande

regra 2: IF CI is médio AND AR is grande then $CL_{submetacentrico}$ is grande

regra 3: IF CI is grande AND AR is médio then $CL_{submetacentrico}$ is grande

regra 4: IF CI is grande AND AR is muito grande then $CL_{submetacentrico}$ is grande

regra 5: IF CI is muito grande AND AR is grande then *CLsubmetacentrico* is grande

regra 6: IF CI is médio AND AR is médio then *CLsubmetacentrico* is média

regra 7: IF CI is pequeno OR AR is pequeno then *CLsubmetacentrico* is pequena

c) o conjunto de regras difusas para classificar o cromossomo na classe subtelocêntrico (*CLsubtelocentrico*)

regra 1: IF CI is médio AND AR is médio then *CLsubtelocentrico* is grande

regra 2: IF CI is grande AND AR is médio then *CLsubtelocentrico* is grande

regra 3: IF CI is médio AND AR is grande then *CLsubtelocentrico* is grande

regra 4: IF CI is médio AND AR is pequeno then *CLsubtelocentrico* is grande

regra 5: IF CI is pequeno AND AR is médio then *CLsubtelocentrico* is grande

regra 6: IF CI is grande AND AR is grande then *CLsubtelocentrico* is média

regra 7: IF CI is muito grande OR AR is muito grande then *CLsubtelocentrico* is pequena

d) o conjunto de regras difusas para classificar o cromossomo na classe acrocêntrico (*CLacrocentrico*)

regra 1: IF CI is pequeno AND AR is pequeno then *CLacrocentrico* is grande

regra 2: IF CI is pequeno AND AR is médio then *CLacrocentrico* is grande

regra 3: IF CI is médio AND AR is pequeno then *CLacrocentrico* is grande

regra 4: IF CI is médio AND AR is médio then *CLacrocentrico* is média

regra 5: IF CI is muito grande AND AR is muito grande then *CLacrocentrico* is pequena

regra 6: IF CI is grande AND AR is grande then *CLacrocentrico* is pequena

Motor de Inferência Difusa

O motor de inferência difusa foi executado utilizando o método de interpolação, citado na seção 4.3.1. O primeiro passo do método de interpolação é resolver o antecedente de cada regra de cada conjunto de regras. Para isso, o operador lógico AND foi modelado usando uma função de mínimo e o operador lógico OR foi modelado usando uma função de máximo.

O próximo passo é executar conclusão do motor de inferência difusa através da impli-



Figura 4.7: Imagem de um cromossomo de peixe, com seu centrômero localizado e que possui $CI = 31,7$ e $AR = 2,15$

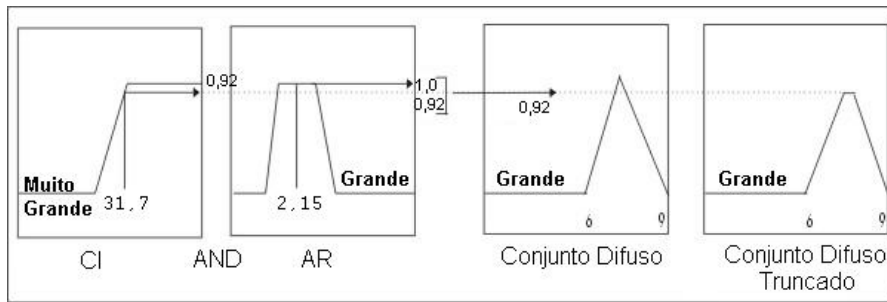
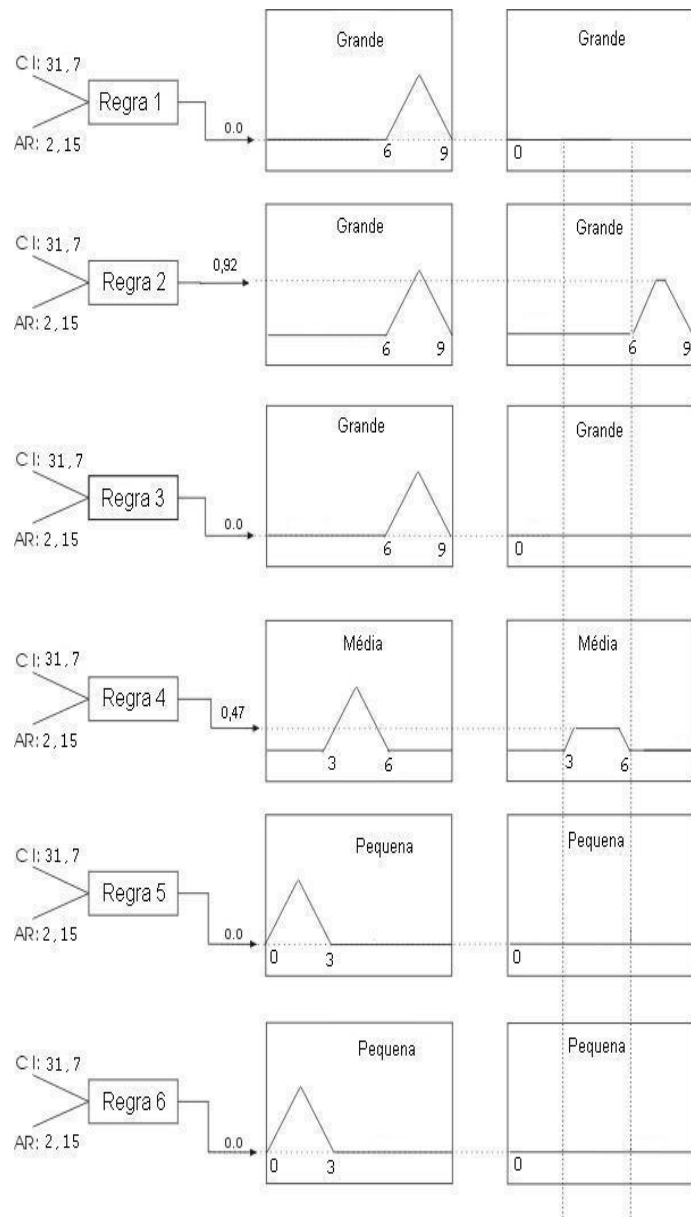


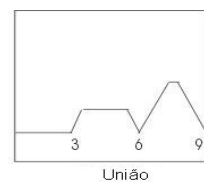
Figura 4.8: Conclusão do motor de inferência difusa para a regra 2 da variável lingüística *CLmetacentrico*.

cação THEN presente em cada regra de cada conjunto de regras, que irá produzir como saída um conjunto difuso truncado. A Figura 4.7 mostra a imagem de um cromossomo, a ser classificado, que possui $CI = 31,7$ e $AR = 2,15$. A Figura 4.8 ilustra o conjunto difuso truncado resultante do método de interpolação aplicado na regra número 2 da variável lingüística *CLmetacentrico*, para classificar o cromossomo mostrado na Figura 4.7.

Em seguida é executada a união dos conjuntos difusos truncados obtidos pela aplicação do método de interpolação para cada uma das regras para a *CLmetacentrico*, produzindo como resultado um único conjunto difuso, que será defuzificado. A união foi modelada pelo operador de máximo. A Figura 4.9 (a) mostra o resultado do método de implicação para cada regra da variável lingüística *CLmetacentrico*, para o cromossomo da Figura 4.7. A Figura 4.9 (b) mostra a conclusão do módulo de inferência difusa para a variável lingüística *CLmetacentrico*. Como existem quatro conjuntos de regras difusas, um para cada classe de cromossomos, então o método de união é aplicado quatro vezes, um para cada conjunto de regras, produzindo quatro conjuntos difusos



(a)



(b)

Figura 4.9: Conclusão do motor de inferência difusa para as regras da classe *CLmetacentrico* (a) Resultado da implicação para cada regra da classe *CLmetacentrico* e (b) Resultado da união para a classe *CLmetacentrico*

Defuzificação

O método proposto para defuzificar a saída do módulo de motor de inferência difusa é o centro de área, também chamado de método de centro de gravidade ou método do centróide [51]. Neste método o valor defuzificado $D(C)$ é definido como o valor dentro do intervalo da variável CLi , para o qual a área abaixo do gráfico da função de pertinência difusa C é dividida em duas áreas iguais.

$$D = \frac{\int C(z)zdz}{\int C(z)dz} \quad (4.2)$$

Para o caso discreto, no qual a variável C é definida em um universo finito $z_1, z_2, \dots, z_n$, a fórmula é dada por:

$$D = \frac{\sum_{k=1}^n C(z_k)z_k}{\sum_{k=1}^n C(z_k)} \quad (4.3)$$

Como o sistema de inferência difusa produz quatro valores difusos, após o processo de defuzificação, um para cada classe de cromossomo, o sistema deverá escolher um deles como a classe a qual o cromossomo sendo analisado pertence. O maior dos valores de defuzificação é escolhido, sendo que a sua respectiva classe é a classe a que o cromossomo pertence. Caso o citogeneticista queira, ele poderá alterar a classificação do cromossomo, sugerida pelo sistema.

O resultado da defuzificação para o conjunto *CLmetacentrico* exemplo é aplicado ao da Figura 4.9 (b) e é mostrado na Figura 4.10. Considerando o cromossomo mostrado na Figura 4.7, o resultado da defuzificação para os conjuntos *CLmetacentrico*, *CLsubmetacentrico*, *CLsubtelocentrico* e *CLacrocentrico* é 6,2; 7,5; 2,8 e 1,5 respectivamente. O sistema classifica o cromossomo exemplo como pertencente à classe submetacêntrico.

Caso de empate

Nos casos de empate no processo de classificação, o citogeneticista poderá escolher uma das classes à qual o cromossomo pertence, ou deixar que o sistema faça a escolha usando um procedimento de classificação baseado no processo manual executado pelo citogeneticista. O processo manual de classificação leva em consideração somente a variável AR

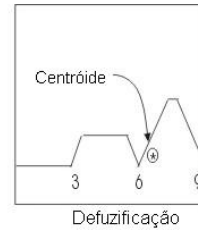


Figura 4.10: Método do centróide aplicado ao resultado da agregação 4.9 (b), resultando no valor 6,2.

e utiliza as seguintes regras de classificação. Se o valor do AR estiver no intervalo $[1; 1, 7]$ o cromossomo é classificado como metacêntrico; se o CI estiver no intervalo $]1, 7; 3]$, o cromossomo é classificado como submetacêntrico; se o CI estiver no intervalo $[3, 0; 7]$ o cromossomo é classificado como subtelocêntrico e se o CI for maior que 7, o cromossomo é classificado como acrocêntrico.

4.4 Pareamento

Após executada a etapa de classificação, os cromossomos são pareados de acordo com sua similaridade. O pareamento é feito entre os cromossomos de uma mesma classe.

Inicialmente, os cromossomos de cada classe são ordenados segundo o seu tamanho, determinado pelo seu eixo longitudinal. Os cromossomos são então dispostos em ordem decrescente de tamanho, organizado em pares.

Antes de realizar o pareamento, o citogeneticista deve selecionar os chamados cromossomos B. Estes cromossomos, já mencionados na seção 2.6, se caracterizam por não serem pareados com nenhum outro cromossomo. Depois que o citogeneticista seleciona estes cromossomos, eles são então excluídos do processo de pareamento e aparecem no final da classe, após o último par de cromossomos, ou após o último cromossomo que não possui par. Em peixes, cada classe pode conter um número par ou ímpar de cromossomos, então um cromossomo que não possui par, não é necessariamente um cromossomo B.

Ao final do processo de pareamento, o cariótipo final de um peixe é obtido. Este cariótipo pode ser interativamente alterado pelo citogeneticista, caso deseje modificar os

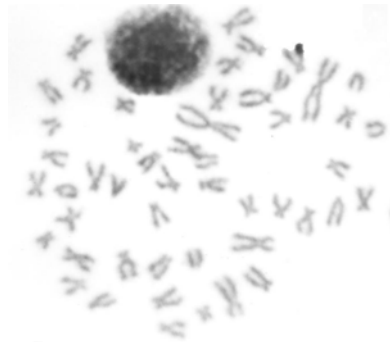
pares de cromossomos ou alterar a classe do cromossomo. O cariótipo pode ser salvo em um banco de dados para posteriores consultas ou alterações. A Figura 4.11 mostra uma imagem de cromossomos de peixe em metáfase, o resultado do processo de classificação e do processo de pareamento, que resulta no cariótipo do peixe.

Após estudar vários exemplares de uma mesma população de peixe e montar o cariótipo de todos eles, o citogeneticista deve escolher o cariótipo que melhor representa a população em estudo. A fim de auxiliar na escolha do melhor cariótipo, o SiMCaPe disponibiliza uma interface gráfica, que será apresentada no próximo capítulo.

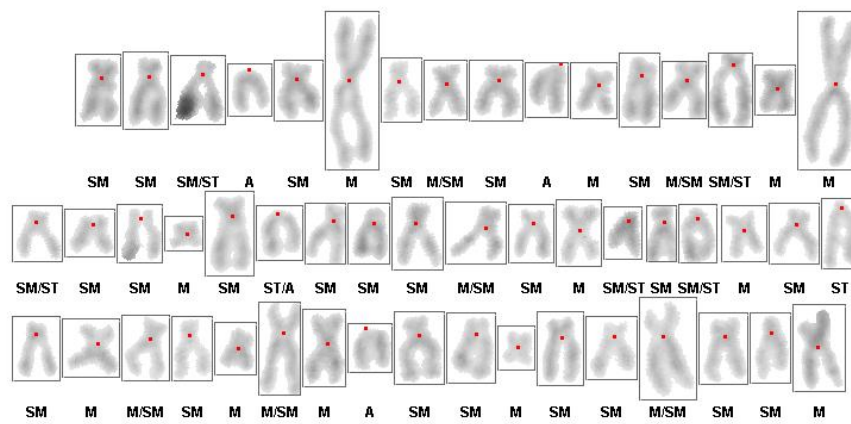
4.5 Considerações Finais do Capítulo

Este capítulo apresentou o processo de classificação e o processo de pareamento de cromossomos. Estes processos culminaram na montagem do cariótipo final de peixe. Agora que o cariótipo está montado, o citogeneticista pode estudar a espécie, conhecer suas características e fazer possíveis estudos evolutivos.

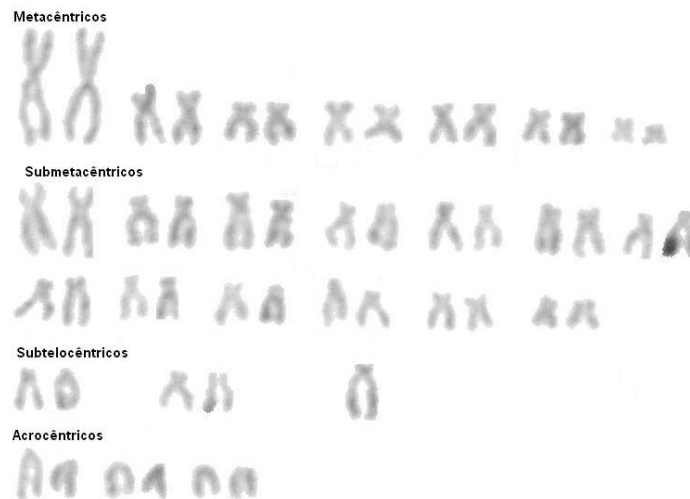
Além das funções mencionadas nos Capítulos 3 e 4, o sistema SiMCaPe, disponibiliza algumas facilidades gráficas que irão auxiliar o citogeneticista no processo de montagem do cariótipo. O próximo capítulo tem como objetivo mostrar estas funcionalidades em detalhes.



(a)



(b)



(c)

Figura 4.11: Cariótipo (a) imagem de cromossomos de peixe em metáfase (b) resultado do processo de classificação dos cromossomos (c) resultado do processo de pareamento

Capítulo 5

O Sistema para Montagem de Cariótipos

5.1 Introdução

O sistema SiMCaPe apresenta muitas tarefas automatizadas, tais como pré-processamento, segmentação, rotação dos cromossomos, localização do centrômero, classificação e pareamento. Mas devido à grande subjetividade do problema, o citogeneticista pode alterar os resultados fornecidos pelo sistema, para cada uma destas tarefas, usando as interfaces gráficas disponíveis no sistema, que possibilitam de maneira fácil esta interação.

O citogeneticista pode alterar os valores sugeridos para pré-processamento e segmentação da imagem, pode rotacionar interativamente um cromossomo, pode alterar a posição do centrômero e a classe do cromossomo e trocar os pares de cromossomos sugeridos pelo sistema. O sistema também disponibiliza uma interface onde é possível separar os cromossomos que se sobrepõem e marcar os cromossomos do tipo B, para que eles não participem do processo de pareamento.

Os dados relativos a cada cariótipo são salvos em uma base de dados modelada de forma a garantir consultas do tipo: seleção de um dado cariótipo já armazenado para visualização, alteração no pareamento dos cromossomos de um cariótipo já armazenado e recuperação de todos os cariótipos de uma dada população para escolha do melhor cariótipo.

Neste capítulo serão apresentadas as funcionalidades do sistema SiMCaPe, que auxiliam o citogeneticista na obtenção dos melhores resultados no processo de montagem do cariótipo de peixe. Na seção 6 serão apresentados a metodologia para a realização dos testes e uma avaliação dos resultados obtidos.

5.2 Recursos utilizados no desenvolvimento do SiMCaPe

Os requisitos que orientaram a escolha da linguagem de programação para o desenvolvimento do SiMCaPe foram: possuir recursos de processamento digital de imagens, possuir facilidades para construção de interface gráfica e permitir portabilidade. Diante destas necessidades, a linguagem Java foi escolhida por suportar múltiplas plataformas, por oferecer facilidades para o desenvolvimento de interface, pelo uso do `java.swing` e do `java.awt`, e por possuir uma biblioteca para trabalhar com imagens, o JAI (Java Advanced Image). Foi utilizada a versão Java 1.4.2 e a versão JAI 1.1.2.

Ferramentas como o MatLab, Visual C ou C++ Builder poderiam ter sido usadas, mas trariam mais dificuldades para o desenvolvimento. A primeira traz grandes facilidades para o trabalho com imagens, pois possui uma biblioteca de funções já implementadas, mas possui poucos recursos para a construção de interfaces. A segunda apresenta facilidades de interface gráfica mas não disponibiliza recursos para trabalhar diretamente com imagens. A terceira disponibiliza recursos de interface gráfica e funções para trabalhar com imagens, mas estas funcionalidades não são tão completas quanto as apresentadas na linguagem Java.

O Sistema Gerenciador de Banco de Dados escolhido para integrar o SiMCaPe é o PostgreSQL, versão 8.0, por atender aos requisitos exigidos pelo sistema, estar disponível na Internet, ser software livre e possuir fácil integração com o Java.

Nas próximas seções serão apresentadas as funcionalidades disponíveis no sistema SiMCaPe e as respectivas interfaces, que permitem ao usuário interagir com os valores automáticos propostos pelo sistema.

5.3 Pré-processamento e segmentação dos cromossomos

A etapa de pré-processamento já foi discutida no Capítulo 3. Como foi visto, o segundo passo do pré-processamento é a aplicação da função de pertinência difusa que requer um limite superior, como parâmetro. Este limite superior é calculado pelo sistema utilizando o histograma da imagem que foi realçada pela função quadrática. No entanto, em alguns casos, o valor calculado pelo sistema não produz um bom resultado, sendo necessário que o usuário altere este valor interativamente.

Para auxiliar o citogeneticista na escolha deste limite superior, foi proposto um *slider*, como o mostrado na Figura 5.1. À medida que o valor do limite superior é modificado, a função de pertinência difusa é alterada e aplicada sobre a imagem e o resultado é imediatamente mostrado na tela. Assim, o usuário encontra mais facilidade para encontrar o parâmetro para realizar esta etapa do pré-processamento. O mesmo recurso é aplicado na função de limiarização, executada após a função de pertinência difusa, que também necessita de um limiar, para ser executada.

Além disso, nas etapas de pré-processamento e segmentação é possível fazer um *zoom* na imagem, a fim de identificar detalhes, como por exemplo, a região de fronteira entre dois cromossomos, ou visualizar pequenos objetos que se parecem com um cromossomo, sendo possível usar uma borracha para apagar estes objetos. Também é possível salvar em um arquivo, a imagem resultante de cada processo que foi executado.

Outra funcionalidade gráfica disponível na etapa de pré-processamento é a possibilidade de desenhar um contorno não retangular na imagem original e aplicar funções de realce, brilho e contraste nesta imagem, com parâmetros informados pelo usuário. Esta função se tornou necessária visto que, algumas vezes, o fundo da imagem apresenta artefatos que impedem a aplicação de uma mesma função de pré-processamento com os mesmos parâmetros. Com esta funcionalidade, o usuário pode selecionar diferentes áreas da imagem onde os artefatos se mantêm constantes e aplicar funções de realce, brilho e contraste locais a cada uma delas, usando parâmetros configurados por ele.

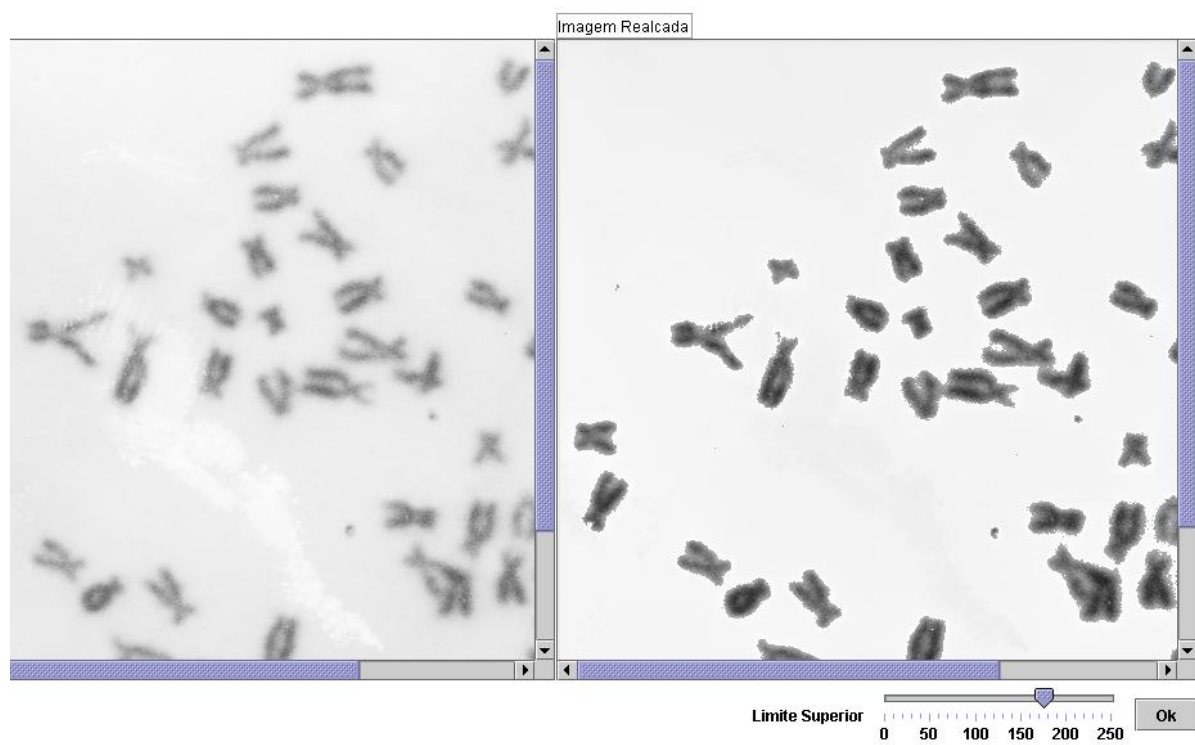


Figura 5.1: Escolha do limite superior para execução da função de pertinência



Figura 5.2: Cromossomos sobrepostos duplicados que serão separados interativamente

5.4 Separação dos cromossomos que se sobrepõem

Como foi apresentado na seção 3.4.1, o sistema SiMCaPe propõe um método automático de separação dos cromossomos que se tocam. No entanto o sistema não propõe nenhum método automático para a separação dos cromossomos sobrepostos. Para os casos de sobreposição de cromossomos e para os casos onde a rotina automática não conseguiu separar os cromossomos que se tocam, foi proposta uma interface capaz de separar estes cromossomos interativamente. Nesta interface, os cromossomos que estão juntos, são apresentados em uma janela suspensa e são apresentados de forma duplicada, como mostrado na Figura 5.2.

Para separar os cromossomos o usuário deverá apagar na primeira janela um dos cromossomo e apagar na segunda janela o outro cromossomo. Para apagar o cromossomo foi implementado uma borracha, que ao ser passada sobre a imagem, transforma os pixels em branco, ou seja, estes tornam-se fundo da imagem. A borracha é a apresentada em 3 tamanhos diferentes para que seja possível apagar pequenas e grandes regiões. O SiMCaPe também disponibiliza um recurso de *zoom* para auxiliar na remoção de regiões muito



Figura 5.3: Cromossomos sobrepostos que foram separados interativamente

pequenas do cromossomo. Além disso, caso o usuário tenha apagado algo errado, ele pode recomençar o processo, a qualquer momento. O resultado da separação dos cromossomos apresentados na Figura 5.2 é mostrado na Figura 5.3.

Além do recurso da borracha para apagar, também há o recurso da linha. Este recurso é melhor utilizado quando há cromossomos que se tocam. Assim, é possível passar uma linha, ao longo da região de fronteira entre os cromossomos, separando-os.

5.5 Rotação de cromossomos

Para a extração das características, inicialmente os cromossomos devem ser rotacionados para a posição vertical. Como foi mostrado na seção 3.5, o esqueleto do cromossomo é encontrado e com base nele é encontrado o ângulo de rotação. Como o esqueleto de alguns cromossomos são bastante ramificados, em alguns casos, o ângulo de rotação calculado pelo sistema não é adequado para rotacionar o cromossomo para a posição vertical. Nestes casos, o SiMCaPe disponibiliza uma interface gráfica para que o citogeneticista possa fazer



Figura 5.4: Rotação manual de um cromossomo



Figura 5.5: Rotação manual de múltiplos cromossomos

a rotação interativamente.

O cromossomo a ser interativamente rotacionado é mostrado numa janela suspensa na qual o usuário informa o ângulo de rotação, conforme mostra a Figura 5.4. Para rotacionar o cromossomo para o sentido horário o valor do ângulo informado é positivo, para rotacionar para o sentido anti-horário, o valor do ângulo é negativo. Se a rotação não ficar boa, o usuário poderá fazer uma nova rotação, à partir da posição em que o cromossomo se encontra, ou então cancelar a rotação que ele fez. Também é possível selecionar vários cromossomos, de uma só vez, e rotacioná-los com o mesmo ângulo, como mostra a Figura 5.5.

5.6 Pareamento de cromossomos

O sistema SiMCaPe disponibiliza uma rotina automática de pareamento, onde os cromossomos da mesma classe são ordenados em ordem decrescente de tamanho, e organizado em

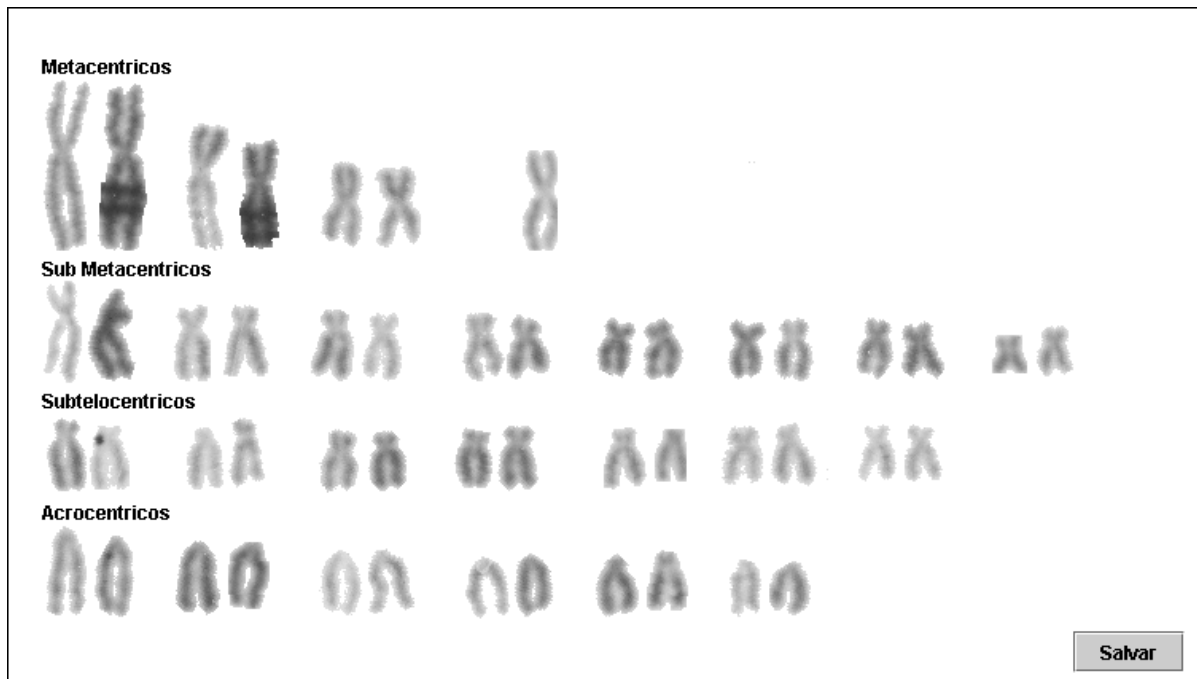


Figura 5.6: Resultado do pareamento

pares. Após executar o pareamento é exibida uma tela onde aparecem as quatro classes de cromossomos, com seus respectivos cromossomos pareados, conforme mostra a Figura 5.6.

O usuário poderá trocar os pares de cromossomos, selecionando dois cromossomos para a troca. A troca de pares pode ser feita entre cromossomos da mesma classe, ou entre cromossomos de classes diferentes. Também é possível alterar a classificação de um cromossomo. Além disso, é possível salvar o resultado em uma base de dados para posteriores consultas. Esta funcionalidade será apresentada na seção 5.8.

5.7 Escolha do melhor cariótipo

Para escolher o melhor cariótipo o citogeneticista deve primeiramente selecionar a espécie e a população a ser consultada. Em seguida, o sistema faz uma busca em todas os cariótipos armazenados na base de dados, que pertençam àquela espécie e população. Todos os cariótipos recuperados são então exibidos na tela, como mostra a Figura 5.7.

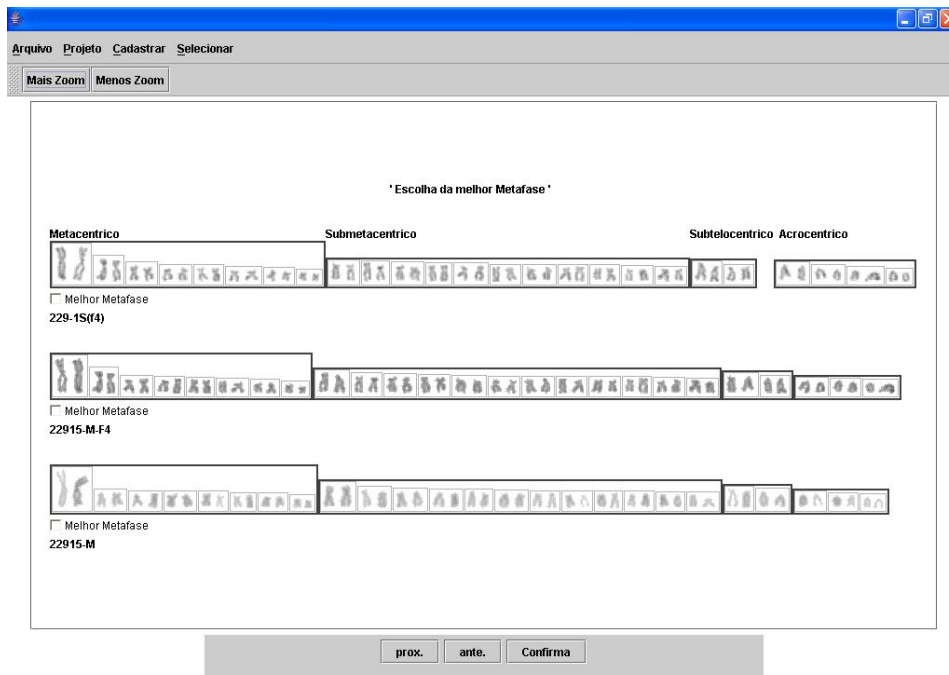


Figura 5.7: Tela de escolha do melhor cariótipo

Os cariótipos são dispostos com os cromossomos em tamanho reduzido para que o citogeneticista consiga visualizar vários cariótipos de uma só vez. Então o citogeneticista analisa em cada cariótipo características como: imagem com bom contraste, cromossomos com cromátides juntas, boa localização do centrômero e cromátides pouco tortas. Os cariótipos que não apresentam tais características são eliminados pelo citogeneticista. Por último o citogeneticista escolhe aquele que melhor representa a espécie.

Para auxiliar na escolha do melhor cariótipo, a interface permite eliminar cariótipos, executar um zoom para aumentar a visualização de um cromossomo e executar um zoom para diminuir o tamanho do cromossomos a fim de que eles caibam na tela. Depois de escolher o melhor cariótipo, este é armazenado na base de dados de cariótipos.

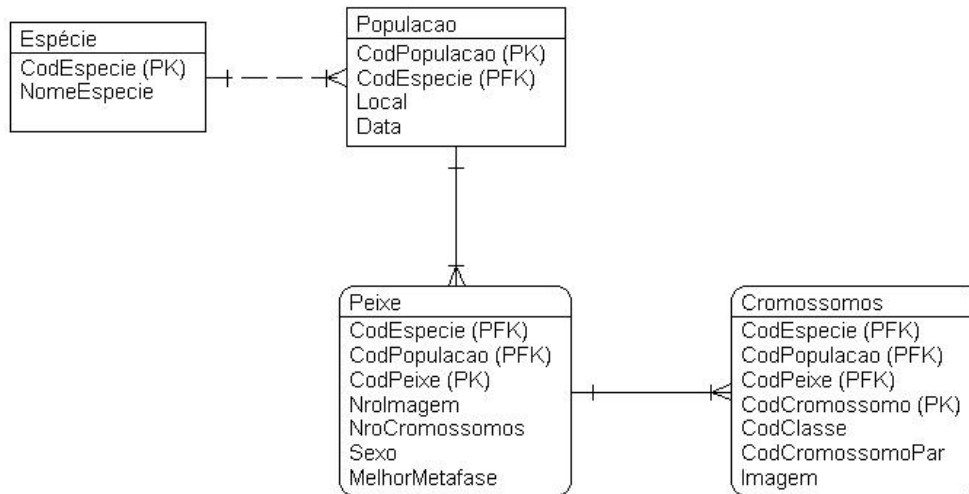


Figura 5.8: Modelo do Banco de Dados

5.8 Modelo da Base de Dados proposta para o SiMCaPe

A fim de armazenar os cariótipos montados pelos citogeneticistas, para futuras consultas e atualizações, foi utilizado uma base de dados, cujo modelo é mostrado na Figura 5.8. O modelo é composto por quatro tabelas: Espécie, População, Peixe e Cromossomos.

A tabela Espécie armazena todas as espécies de peixe que o laboratório de citogenética trabalha. Toda vez que o laboratório vai trabalhar com uma nova espécie, esta deve ser cadastrada nesta tabela. São exemplos de espécies de peixe: *Astyannax scabripinnis*, *Astyannax eigenmaniorium*, *Astyannax altiparanae* e *Hoplias malabaricus*.

A tabela População armazena todas as populações de uma espécie que estão sendo pesquisadas no laboratório. Por população entende-se um grupo de indivíduos da mesma espécie que foram retirados do mesmo lugar. São exemplos de população, os peixes da espécie *Astyannax scabripinnis* retirados do Córrego do Jataí, Córrego do Óleo e Rio das Pedras. A tarefa de cadastrar novas populações fica a cargo do citogeneticista.

A tabela Peixe armazena os dados de um cariótipo de peixe que foi montado pelo sistema. Esta tabela possui atributos como: a espécie e população a que o indivíduo pertence, o nome da imagem que foi executada, quantos cromossomos a imagem possui,

qual o sexo do peixe, se o cariótipo deste indivíduo foi escolhido como o melhor cariótipo e um código sequencial gerado pelo sistema, que identifica o indivíduo.

A tabela Cromossomos armazena os dados referentes aos cromossomos que compõem o cariótipo de um determinado indivíduo. Cada cromossomo pertence a um peixe de uma determinada espécie e população, portanto esta tabela possui os atributos CodEspecie, CodPopulacao e CodPeixe. Cada cromossomo é identificado por um número sequencial gerado pelo sistema e possui também atributos tais como: a classe a que o cromossomo pertence, o código do cromossomo que é seu par na montagem do cariótipo e a imagem em níveis de cinza do cromossomo.

As tabelas Peixe e Cromossomos são preenchidas automaticamente pelo sistema, no final da montagem de uma metáfase, quando o usuário seleciona o botão salvar e escolhe a espécie, a população e o sexo do indivíduo.

5.9 Considerações Finais do Capítulo

Este capítulo mostrou as principais funcionalidade gráficas disponíveis no sistema SiMCaPe que auxiliam o citogeneticista no processo de montagem do cariótipo de peixe. Estas funcionalidades permitem que o usuário altere interativamente os resultados propostos pelo sistema, nos casos onde ele julgar necessário. Além disso, este capítulo mostrou o modelo do Banco de Dados utilizado para armazenar os cariótipos montados pelo sistema.

O próximo capítulo tem como objetivo mostrar como foram realizados os testes e apresentar uma discussão dos resultados obtidos pelo SiMCaPe.

Capítulo 6

Avaliação e Discussão dos Resultados

Para testes foram utilizadas 40 imagens de cromossomos de peixe das espécies *Astyanax altiparanae*, *Astyanax eigenmanniorum*, *Astyanax scabripinnis*, *Hoplias malabaricus* e *Hoplias lacerdade*. As 40 imagens foram escolhidas aleatoriamente, não usando nenhum critério para seleção. Os resultados obtidos pelo sistema foram comparados com os resultados obtidos por dois grupos de citogeneticistas experientes.

6.1 Metodologia para realização dos experimentos

As imagens utilizadas para testes foram primeiramente pré-processadas, os cromossomos foram segmentados e rotacionados, utilizando o SiMCaPe. Cada imagem com os cromossomos rotacionados e numerados foram entregues para cada grupo de citogeneticista para que a partir desta imagem, eles pudessem marcar manualmente a posição do centrômero, medir o tamanho dos braços dos cromossomos, classificar os cromossomos e montar o cariótipo final. Os dois grupos de citogeneticistas trabalharam de forma independente para que os resultados não fossem influenciados.

Para marcar a posição do centrômero e medir o tamanho dos braços dos cromossomos os citogeneticistas utilizaram o *software* Micro Measure [34]. Neste *software* o usuário marca interativamente a posição do centrômero, as extremidades do cromossomo e alguns pontos que representem a curvatura do braço do cromossomo. A partir destes pontos o sistema calcula o tamanho do braço maior e menor do cromossomo.

6.2 Avaliação dos resultados

1. Avaliação dos resultados obtidos pelos processos de pré-processamento e segmentação:

Em 25 das 40 imagens usadas para testes havia ocorrência de cromossomos que se tocam. Nestas imagens havia cerca de 3,7 cromossomos se tocando por imagem. Depois de executar o processo de segmentação automática, disponível no SiMCaPe, o número de cromossomos se tocando por imagem mudou para 1,33. Isto resulta numa melhoria de 64,05%. Os cromossomos que continuaram se tocando após o processo de segmentação tiveram de ser interativamente separados pelo usuário, através da interface gráfica disponível no sistema, apresentada no capítulo anterior.

Para as etapas de pré-processamento e segmentação, que são tarefas difíceis do processo de cariotipagem, os resultados obtidos pelo sistema foram satisfatórios. Para o pré-processamento, a aplicação das funções de realce, de pertinência difusa e de limiarização não produziram bons resultados nos casos, em que a imagem apresentava simultaneamente regiões com muito ruído e outras regiões com pouco ruído. Este fato requer que as funções de pré-processamento utilizem parâmetros adaptativos, ainda não incorporados ao sistema. Para contornar este problema o SiMCaPe possui funções extras de brilho, contraste e intensidade para facilitar o pré-processamento das imagens que apresentam regiões heterogêneas. Neste caso o usuário seleciona uma parte da imagem, que mantém um ruído constante, e aplica estas funções extras utilizando parâmetros configurados pelo usuário.

O processo de segmentação, com a separação dos cromossomos que se tocam, apresentou bons resultados utilizando a teoria de conjuntos difusos. Nos casos em que a região de transição entre cromossomos era muito escura, o método proposto não produziu a separação automática dos cromossomos, necessitando da intervenção do citogeneticista. No entanto, ao contrário de muitos sistemas existentes, o sistema consegue separar agrupamentos que contenham mais que dois cromossomos que se tocam.

2. Avaliação dos resultados obtidos pelo processo de rotação:

Nas imagens usadas para testes, havia cerca de 49,2 cromossomos por imagem, sendo que 69,5% dos cromossomos por imagem, foram rotacionados corretamente pelo sistema. Os cromossomos, que foram incorretamente rotacionados pela rotina automática do SiMCaPe, foram rotacionados interativamente pelo usuário.

Os cromossomos foram rotacionados para a sua posição vertical de acordo com a orientação do eixo principal do seu esqueleto. Como algumas espécies de peixe apresentam cromossomos com as cromátides um pouco separadas, o esqueleto do cromossomo destas espécies apresenta muitas ramificações o que dificulta a rotação correta do cromossomo. Além deste caso, os cromossomos acrocêntricos, cujos esqueletos muitas vezes apresentam a forma de U ou a forma de V, também são rotacionados de forma incorreta. Para os casos em que o resultado da rotação foi incorreto, o sistema disponibiliza uma interface gráfica para que a rotação seja corrigida interativamente. Há também o caso dos cromossomos que após o processo automático de rotação ficam invertidos. Neste caso, estes cromossomos devem ser rotacionados interativamente em 180° graus. Estes cromossomos não foram contabilizados como erro do sistema, visto que os procedimentos subsequentes à rotação funcionam normalmente com o cromossomo invertido.

3. Avaliação dos resultados obtidos pelo processo de localização do centrômero:

Para o processo de localização do centrômero, os resultados obtidos pelos grupos de citogeneticistas foram comparados com os resultados obtidos pelo sistema SiMCaPe para esta tarefa. O sistema SiMCaPe e o primeiro grupo de citogeneticistas, no processo de localização do centrômero, concordaram em 71,95% dos cromossomos por imagem. O sistema SiMCaPe e o segundo grupo de citogeneticistas, no processo de localização do centrômero, concordaram em 78,83% dos cromossomos por imagem. Vale ressaltar que os dois grupos de citogeneticistas concordam quanto à localização do centrômero em 84,15% dos cromossomos por imagem.

Para a localização do centrômero foi usado o vetor de projeção horizontal do cromos-

somo, e a partir deste vetor foi localizado o ponto de mínimo da maior bacia encontrada. No entanto, alguns cromossomos não apresentam um ponto de mínimo no seu vetor de projeção horizontal, como é o caso dos cromossomos acrocêntricos. Quando o cromossomo não possui nenhum ponto de mínimo, o sistema caracterizou-o como sendo um cromossomo acrocêntrico. No entanto, alguns cromossomos subtelocêntricos também não apresentam ponto de mínimo, assim como alguns cromossomos acrocêntricos apresentam ponto de mínimo. Nestes casos, o sistema apresentou grande dificuldade na localização do centrômero. Para os cromossomos metacêntricos, submetacêntricos e os subtelocêntricos (que apresentam um ponto de mínimo) e os acrocêntricos (que não apresentam ponto de mínimo), o sistema obteve resultados bastante satisfatórios.

4. Avaliação dos resultados obtidos pelo processo de classificação:

Os resultados obtidos pelo sistema SiMCaPe, para o processo de classificação de cromossomos, foram comparados com os resultados obtidos pelos dois grupos de citogeneticistas. No processo de classificação, o sistema SiMCaPe e o primeiro grupo de citogeneticistas concordam em 78,52% dos cromossomos por imagem. O sistema SiMCaPe e o segundo grupo de citogeneticistas concordam em 81,16% dos cromossomos por imagem.

O processo de classificação alcançou bons resultados quando comparado com os resultados produzidos pelos dois grupos de citogeneticistas e levando-se em consideração que este é um processo bastante subjetivo. Alguns cromossomos foram classificados numa classe pelo primeiro grupo de citogeneticistas e em outra classe pelo segundo grupo de citogeneticistas. Comparando os resultados obtidos pelos dois grupos, notamos que o nível de concordância entre eles é de 70,74 % dos cromossomos por imagem.

Os erros no processo de classificação aconteceram, em geral, devido a erros na estimativa do tamanho dos braços do cromossomo. Este erro acontece principalmente quando o cromossomo possui braços muito tortos, gerando erros nas medidas índice centromérico (CI) e razão entre braços (AR) que afetam diretamente a classificação

do cromossomo.

O processo de classificação também apresenta dificuldades em classificar os cromossomos acrocêntricos e subtelocêntricos. Estes dois tipos de cromossomos apresentam o centrômero localizado bem na extremidade dos cromossomos, numa região bastante próxima, dificultando a sua caracterização. Quando a posição do centrômero dos cromossomos acrocêntricos está localizada bem na extremidade do cromossomo o sistema acerta em 100% dos casos.

5. Avaliação dos resultados para o processo de pareamento:

O pareamento é outra etapa do processo de cariotipagem que é bastante subjetiva. Cada citogeneticista pode montar pareamentos diferentes para uma mesma imagem de cromossomos e a cada vez que ele re-analisa o pareamento, pode realizar alterações, visto que esta é uma tarefa que exige bastante percepção visual. Sendo assim torna-se difícil uma avaliação objetiva dos resultados apresentados pelo SiMCaPe. No caso dos cromossomos metacêntricos, que em geral são poucos e apresentam alguns pares bem grandes e alguns pares bem pequenos, os resultados apresentados pelo SiMCaPe foram bons. Também no caso de cromossomos acrocêntricos que são em geral em pequeno número, os resultados foram bons. Já no caso de cromossomos submetacêntricos e subtelocêntricos, onde os pares apresentam tamanhos semelhantes, o sistema não conseguiu bons resultados.

Fazendo uma comparação dos pareamentos produzidos pelos dois grupos de citogeneticistas, é possível notar que não existe concordância entre eles. Os pares de cromossomos submetacêntricos e subtelocêntricos, encontrados pelos dois grupos, variam muito, o que justifica o baixo desempenho do SiMCaPe.

6.3 Considerações Finais do Capítulo

Este capítulo comparou os resultados obtidos pelo sistema SiMCaPe com os resultados obtidos por dois grupos de citogeneticistas, que trabalharam de forma independente.

A partir dos resultados obtidos é possível notar que o nível de concordância entre

o sistema e os dois grupos de citogeneticistas para o processo de classificação é maior que o nível de concordância entre os grupos, o que mostra que o sistema SiMCaPe está produzindo resultados próximos ao encontrado pelo especialista.

Uma grande vantagem do sistema SiMCaPe é o aumento de produtividade do citogeneticista. Sem utilizar o sistema SiMCaPe os citogeneticistas gastam em média 2h para montar um cariótipo, conseguindo ao fim do dia montar cerca de 4 cariótipos. Utilizando o sistema SiMCaPe o citogeneticista gasta cerca de 20 a 30 minutos para montar cada cariótipo, totalizando cerca de 16 cariótipos por dia. Portanto, temos um aumento de produtividade de 300%.

Outra vantagem proporcionada pelo uso do sistema SiMCaPe é a padronização do procedimento para montagem do cariótipo, o que facilitará a análise e comparação de resultados obtidos por diferentes grupos de pesquisa.

Capítulo 7

Conclusão

O estudo do cariótipo de diversos organismos é uma tarefa importante e rotineira nos laboratórios de citogenética. Muitos trabalhos foram propostos para automatizar a tarefa de montagem de cariótipos, mas a maioria deles são aplicados à montagem do cariótipo humano. Os sistemas de cariotipagem que podem ser usados para outros organismos além do humano, em geral, ou disponibilizam poucas tarefas automatizadas ou apresentam restrições quando usados para montar o cariótipo de peixe .

Visando auxiliar os citogeneticistas na montagem de cariótipos de peixe, foi proposto neste trabalho o sistema SiMCaPe. O sistema proposto disponibiliza tarefas automatizadas para as principais tarefas do processo de cariotipagem, que são: pré-processamento da imagem, segmentação dos cromossomos, rotação dos cromossomos, extração das características de cada cromossomo, classificação e pareamento. Além disso, o sistema disponibiliza recursos gráficos para que o citogeneticista possa interagir com os resultados propostos pelo sistema, caso julgue necessário.

Tendo em vista os objetivos do sistema e os resultados apresentados, pode-se concluir que o SiMCaPe atingiu os objetivos iniciais e está atendendo as expectativas dos citogeneticistas. Os resultados obtidos em cada uma das fases são bastante satisfatórios. O processo de montagem do cariótipo é bastante subjetivo, portanto já imaginava-se que interações deveriam ser feitas pelo citogeneticista a fim de alterar os resultados propostos pelo sistema. Portanto pode-se concluir que o sistema SiMCaPe poderá auxiliar muito o trabalho dos citogeneticistas que pesquisam o cariótipo de peixe.

Este trabalho proporcionou várias contribuições, sendo que as principais são:

- desenvolvimento de um sistema de inferência difusa para classificação dos cromossomos, sem utilização do padrão de banda
- desenvolvimento de um método para localização do centrômero do cromossomo
- realização do pareamento em espécies que não possuem um número fixo de cromossomos
- desenvolvimento de um método de segmentação que separa cromossomos que se tocam
- desenvolvimento de um sistema interativo para auxiliar na montagem de cariótipos

Existem algumas propostas de trabalhos futuros. A primeira proposta é testar o sistema para a montagem de cariótipos de outros seres vivos, entre eles os seres humanos. Os grupos de animais a serem testados são: mamíferos, aves, répteis e anfíbios. No caso dos humanos, o sistema poderá ser usado até a etapa de classificação. A etapa de pareamento, disponível no SiMCaPe, não produz bons resultados no caso de cromossomos humanos já que não leva em consideração a característica padrão de banda.

O processo de localização do centrômero para o caso dos cromossomos acrocêntricos é uma tarefa que necessita ser melhorada, visto que para esta classe de cromossomos, o algoritmo desenvolvido no sistema SiMCaPe não apresentou bons resultados. Neste caso, novas características devem ser agregadas à este processo, como por exemplo o contorno do cromossomo ou o seu esqueleto.

O processo de separação dos cromossomos sobrepostos apresentado pelo SiMCaPe, é um processo interativo, portanto necessita ser automatizado. Um método de separação dos cromossomos sobrepostos deve ser criado ou algum dos métodos já existentes, que separa cromossomos humanos sobrepostos, deve ser adaptado.

Como o processo de pareamento não obteve bons resultados, novas soluções para este problema devem ser desenvolvidas. Características como a forma do cromossomo, tamanho do braço menor e distância entre as cromátides devem ser também usadas no

processo de pareamento, visto que estas são características observadas pelo citogeneticista durante a execução do pareamento realizada interativamente.

A última proposta de trabalho futuro é o desenvolvimento de um método automático de escolha do melhor cariótipo, que leve em consideração a qualidade da imagem, a forma dos cromossomos e a qualidade dos pares de cromossomos.

Referências Bibliográficas

- [1] E. R. Faria, D. Guliato, and J. C. S. Santos. Segmentação de cromossomos de peixe usando conjuntos difusos. In *III Brazilian Workshop on Bioinformatics*, pages 121–124, 2004.
- [2] E. R. Faria, D. Guliato, and J. C. S. Santos. Segmentation and centromere locating methods applied to fish chromosomes images. *Lectures Notes in Bioinformatics, Springer-Verlag*, 3594:181–189, 2005.
- [3] M. L. Carvalho, C. Oliveira, and F. Foresti. Nuclear dna content of thirty species of neotropical fishes. *Genetics and Molecular Biology*, 21:47–54, 1998.
- [4] A. R. T. Mariano. Descrição citogenética de três espécies do gênero astyanax (pisces, characidae) da bacia do rio araguari - Uberlândia, MG. Dissertação, Universidade Federal de Uberlândia, 2001.
- [5] T. F. Denton, editor. *Fish Chromosome Methodology*. Charles Thomas Publisher, Springfield, 1973.
- [6] C. Kosswig. The role of fish in research on genetics and evolution. *Genetics and Fish Mutagenesis, Springer -Verlag*, pages 3–16, 1973.
- [7] R. F. Artoni, M. R. Vicari, and L. A. C. Bertollo. Citogenética de peixes neotropicais: métodos, resultados e perspectivas. *Biological and Health Sciences*, 6:43–60, 2000.
- [8] Conselho de Informações sobre Biotecnologia. Disponível em : <http://www.cib.org.br/glossario.php>. Acessado em 21 de Janeiro de 2006.
- [9] M. Guerra, editor. *Introdução à Citogenética Geral*. Ed. Guanabara, 1998.

- [10] S. Morelli and A. C. S Araujo. Caracterização cariotípica de duas populações do gênero *astyanax* da Bacia do Rio Paranaíba - MG. In *45º Congresso Nacional de Genética*, page 803, 1999.
- [11] C. H. Z. Pantaleão. *Contribuição à análise e classificação citogenética baseada no processamento digital de imagens e no enfoque lógico-combinatório*. Tese, Universidade Federal de Santa Catarina, 1998.
- [12] A. Levan, K. Fredga, and A. A. Sandberg. Nomenclature for centromeric position on chromosomes. *Hereditas*, 52:201–220, 64.
- [13] O. Moreira Filho and L. A. C. Bertollo. Análise cromossômica de *astyanax scabripinnis rivularis* (characiformes, characidae) da região de Três Marias MG. *Ciênc. Cult*, 35:855, 1986.
- [14] S. M. H. N. Mizoguchi and I. C. M. Santos. Cytogenetic and morphometric differences in populations of *astyanax scabripinnis* (pisces, characidae) from Maringá region, PR, Brazil. *Genetics and Molecular Biology*, 21:44–61, 1998.
- [15] Denver Conference. A proposed standard system of nomenclature of human mitotic chromosomes. *Lancet*, 1:1063–1065, 1960.
- [16] R. S. Ledley and F. H. Ruddle, editors. *O Pensamento Matemático nas Ciências do Comportamento*. Editora Renes, 1973.
- [17] L. J. Vliet, I. T. Young, T.K. ten Kate, B.H. Mayall, F.C.A. Groen, and R. Roos, editors. *Athena: A macintosh-based interactive karyotyping system*. Automation of Cytogenetics, Springer Verlag, 1989.
- [18] A. M. Jennings and J. Graham. A neural network approach to automatic chromosome classification. *Phys. Med. Biol*, 38:959–970, 1993.
- [19] E. Datcheva. Cais - an intelligent system for chromosome analysis and classification. In *Int. Workshop Artificial Intelligence and the Humanities*, pages 49–54, 1996.

- [20] E. Datcheva. Heuristic knowledge based chromosome classification. In *Proc. of the First National Conference INFORMATICS'94*, pages 83–89, 1994.
- [21] B. Lerner, H. Guterman, I. Dinstein, and Y. Romem. Medial axis transform-based features and a neural network for human chromosome classification. *Pattern Recognition*, 11:1673–1683, 1995.
- [22] Y. Xia. Skeletonization via the realization of the fire front’s propagation and extinction in digital binary shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11:1076–1086, 1989.
- [23] J. L. Todesco. *Reconhecimento de padrões usando rede neuronal artificial com uma função de base radial: uma aplicação na classificação de cromossomos humanos*. Tese, Universidade Federal de Santa Catarina, 1995.
- [24] J. M. Keller, P. Gader, O. Sjahputera, C. W. Caldwell, and H. T. Huang. A fuzzy logic rule-based system for chromosome recognition. In *Eighth IEEE Symposium on Computer-Based Medical Systems*, pages 125–132, 1995.
- [25] J. M. Keller and O. Sjahputera. Evolution of a fuzzy rule-based system for automatic chromosome recognition. In *IEEE International Fuzzy Systems Conference Proceedings*, pages 129–134, 1999.
- [26] G. Agam and I. Dinstein. Geometric separation of partially overlapping non-rigid objects applied to automatic chromosome classification. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-19:1212–1222, 1997.
- [27] B. Lerner. Toward a completely automatic neural-network based human chromosome analysis. *IEEE Transactions on Systems*, 28:544–552, 1998.
- [28] M. Popescu, P. Garder, J. Keller, C. Klein, J. Stanley, and C. Caldwell. Automatic karyotyping of metaphase cells with overlapping chromosomes. *Computers in Biology and Medicine*, 29:61–82, 1999.
- [29] J. Piper and E. Granum. On fully automatic feature measurement for banded chromosome classification. *Cytometry*, 10:242–255, 1989.

- [30] C. C. Graham and J. Graham. Trainable grey-level models for disentangling overlapping chromosomes. *Patter Recognition*, 32:1335–1349, 1999.
- [31] C. C. Graham and J. Graham. Disentangling chromosome overlaps by combining trainable shape models with classification evidence. *IEEE Transactions On Signal Processing*, 50:2080–2085, 2002.
- [32] J. M Cho. Chromosome classification using backpropagation neural networks. *IEEE Engineering in Medicine and Biology*, 19:28–33, 2000.
- [33] R. C. Gonzales and R. E. Woods, editors. *Digital Image Processing*. Addison-Wesley Publishing Company, 1992.
- [34] A. Reeves and J. Tear. Micromeasure. Disponível em: <http://www.colostate.edu/Depts/Biology/MicroMeasure/>, Acessado em 21 de Janeiro de 2006.
- [35] A. F. Bressane-Neto, F. Porto-Foresti, J. Bortolozzi, and F. Foresti. Desenvolvimento de um software para cariotipagem de cromossomos de peixe. In *Resumos do 51º Congresso Brasileiro de Genética*, page 189, 2005.
- [36] Applied Imaging. Cytovision karyotyping system. Disponível em: <http://www.cytovision.com>, Acessado em 21 de Janeiro de 2006.
- [37] D. Guliato, W. a. Carnielli, J. A. Zuffo, and J. E. Leo Desautels. Segmentation of breast tumors in mammograms. *Journal of Eletronic Imaging*, 12(3):369–378, 2003.
- [38] J. Facon. Home page de jacques facon - processamento e análise de imagens, pontifícia universidade católica do paraná. Disponível em: <http://www.ppgia.pucpr.br/~facon>, Acessado em 21 de Janeiro de 2006.
- [39] C. M. Holt, A. Stewart, and M. Clint e R. H. Perrot. An improved parallel thinning algorithm. *Communications of the ACM*, 30:156–160, 1987.
- [40] C. J. Hilditch. Linear skeletons from square cupboards. In *Proceedings of the 4th Annual Machine Intelligence Workshop*, 4:403–420, 1969.

- [41] Stentiford. Some new heuristics for thinning binary handprinted characteres for ocr. *IEEE Transactions on Systems, Man and Cybernetics*, 13:81–84, 1983.
- [42] T. Y. Zhang and C. Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 24 n3:236–239, 1984.
- [43] M. Moradi, S. K. Setarehdan, and S. R. Ghaffari. Automatic locating the centromere on human chromosome pictures. In *Proceedings 16th IEEE Symposium on Computer-Based Medical Systems (CBMS'03)*, pages 56–61, 2003.
- [44] J. B. T. M. Roerdink and A. Meijster. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41:187–228, 2000.
- [45] E.H. Mamdani and S.Assilian. An experiment in linguistic synthesis with logic controller. *Int. J. Man-Machine Studies*, 7:1–13, 1975.
- [46] A. Sriraman and R. V. Mayorga. A fuzzy inference system approach for greenhouse climate control. *Environment Informatics Archives*, 2:699–710, 2004.
- [47] V. Miranda and P. Calisto. A fuzzy inference system to voltage/var control in dms-distribution management system. In *14 th PCSS, session 3, paper 1*, pages 1–6, 2002.
- [48] H. Seker, O. M. Odetayo, D. Petrovic, and N. G. R. Naguib. A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. *IEEE Transactions on Information Technology in Biomedicine*, 7 n(2):114–122, 2003.
- [49] J. S. Wu, T. E. Lee, C. T. Tsai, T. H. Chang, and S. H. Tsai. A fuzzy rule-based system for crew management of distribution systems in large-scale multiple outages. In *International Conference on Power System Technology- POWERCON 2004*, pages 1084–1089, 2004.
- [50] T. Lin and P. Yu. Partition fuzzy median filter based on fuzzy rules image for image restoration. *Fuzzy Sets and Systems*, 147:75–97, 2004.

- [51] G. J. Klir and B. Yuan, editors. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall Ptr, 1995.