

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS COMPUTACIONAIS
PARA O AGRONEGÓCIO

MATHEUS DE LIMA GOEDERT

**SISTEMA PARA MONTAGEM DE CARIÓTIPO DE PEIXE BASEADO
EM PROCESSAMENTO DIGITAL DE IMAGENS**

DISSERTAÇÃO

MEDIANEIRA-PR

2018

MATHEUS DE LIMA GOEDERT

**SISTEMA PARA MONTAGEM DE CARIÓTIPO DE PEIXE BASEADO
EM PROCESSAMENTO DIGITAL DE IMAGENS**

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio – PPGTCA – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Medianeira, como requisito parcial para obtenção do título de Mestre em Tecnologias Computacionais para o Agronegócio / Área de Concentração: Tecnologias Computacionais Aplicadas à Produção Agrícola e Agroindústria.

Orientador: Prof. Dr. Pedro Luiz de Paula Filho

Coorientador: Prof. Dr. Daniel Rodrigues Blanco

MEDIANEIRA-PR

2018

Dados Internacionais de Catalogação na Publicação

G594s

Goedert, Matheus de Lima

Sistema para montagem de cariótipo de peixe baseado em
processamento digital de imagens / Matheus de Lima Goedert – 2018.
55 f. : il. ; 30 cm.

Orientador: Pedro Luiz de Paula Filho

Coorientador: Daniel Rodrigues Blanco

Dissertação (Mestrado) – Universidade Tecnológica Federal do
Paraná. Programa de Pós-Graduação em Tecnologias Computacionais
para o Agronegócio. Medianeira, 2018.

Inclui bibliografias.

1 Cromossomos. 2. Citogenética. 3. Tecnologias Computacionais -
Dissertações. I. Paula Filho, Pedro Luiz de, orient. II. Blanco, Daniel
Rodrigues, coorient. III. Universidade Tecnológica Federal do Paraná.
Programa de Pós-Graduação em Tecnologias Computacionais para o
Agronegócio. IV. Título.

CDD: 004

Biblioteca Câmpus Medianeira
Marci Lucia Nicodem Fischborn 9/1219



TERRMO DE APROVAÇÃO

SISTEMA PARA MONTAGEM DE CARIÓTIPO DE PEIXES BASEADO EM PROCESSAMENTO DIGITAL DE IMAGENS

Por

MATHEUS DE LIMA GOEDERT

Essa dissertação foi apresentada às catorze horas e trinta minutos, do dia cinco de abril de dois mil e dezoito, como requisito parcial para a obtenção do título de Mestre em Tecnologias Computacionais para o Agronegócio, Linha de Pesquisa Tecnologias Computacionais Aplicadas à Agroindústria, no Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio - PPGTCA, da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Dr. Pedro Luiz de Paula Filho (Orientador – PPGTCA)

Prof. Dr. Arnaldo Candido Junior (Membro Interno – PPGTCA)

Prof. Dr. Roberto Laridondo Lui (Membro Externo – UNIOESTE)

A via original com as assinaturas encontra-se na secretaria do programa.

AGRADECIMENTOS

A Deus, por conceder a oportunidade de concluir mais essa etapa da minha formação acadêmica, provendo força e sabedoria para enfrentar os desafios impostos.

A minha esposa, por todo amor, companheirismo, incentivo e inspiração, mas principalmente, por estar ao meu lado em mais esta etapa da minha vida.

A minha família, em especial aos meus pais, por investirem em minha educação e pelo suporte e incentivo em todos os momentos de minha vida, sem esta base nada seria possível.

Aos meus orientadores, que dividiram essa missão de maneira igualitária, pela compreensão, disponibilidade e humildade ao repassar todo o conhecimento necessário. Foi um privilégio poder trabalhar ao lado de pesquisadores excepcionais.

A minha colega Marcela Marques, pela paciência e presteza em ajudar para a melhoria e enriquecimento deste trabalho.

Aos amigos e colegas, pela força e pela vibração em relação a esta jornada.

Aos professores e colegas de curso, pois juntos trilhamos uma etapa importante de nossas vidas.

A Universidade Tecnológica Federal do Paraná, *Campus* Medianeira, e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela concessão da bolsa de mestrado.

A todos que, com boa intenção, colaboraram com o desenvolvimento desta pesquisa.

"Inteligência é a habilidade de
se adaptar à mudança."

Stephen Hawking

RESUMO

GOEDERT, Matheus L. **Sistema para montagem de cariótipo de peixe baseado em processamento digital de imagens**. 2018. Dissertação (Mestre em Tecnologias Computacionais para o Agronegócio) - Universidade Tecnológica Federal do Paraná.

Partindo do princípio de que as explicações evolutivas, geralmente, são explicações genéticas, obter informações referentes ao genoma das espécies mostra-se fundamental no processo de validação de algumas teorias evolutivas. Na realização deste tipo de pesquisa, os citogeneticistas lançam mão de algumas ferramentas, como a montagem de cariótipos e sua representação, o que permite ao pesquisador identificar diferenças e semelhanças morfológicas nos cromossomos, de forma relativamente simples. Contudo, esta tarefa pode ser considerada trabalhosa e exaustiva. Diante disso, diversos sistemas computacionais envolvendo Processamento Digital de Imagens (PDI) vêm sendo propostos com o intuito de auxiliar nesta atividade. É notável que o PDI apresenta-se como uma importante ferramenta para a automação de processos, seja pela agilidade, precisão ou ainda pelo conforto. Com isso, este trabalho tem por objetivo desenvolver um sistema baseado nos conceitos de visão computacional para classificação automática de cromossomos de peixes a partir de imagens individuais de metáfases. Para isso foram utilizadas técnicas de PDI implementadas em C++, por meio da biblioteca OpenCV. O software proposto tomou por base imagens obtidas através de microscopia de campo claro, de metáfases mitóticas de células individualizadas do peixe *Hoplias malabaricus* (traíra), levantadas do banco de dados do Laboratório de Biodiversidade Molecular e Cromossômica do *Campus* Santa Helena da UTFPR. A primeira etapa do processamento das imagens foi a segmentação, dividida em eliminação do fundo, primeira segmentação e segunda segmentação. Os fundamentos desta etapa foram a limiarização automática da imagem, eliminando seu fundo, e a separação de cromossomos unidos, por meio de um algoritmo de clusterização *k-means*. Devido às diversidades de cores e tonalidades dos seus respectivos fundos, esta etapa pode ser considerada a mais desafiadora, e conseqüentemente a mais extensa e onerosa computacionalmente. Por fim, as imagens individualizadas dos cromossomos foram classificadas de acordo com a proposta de Levan (1964). Para isso desenvolveu-se um algoritmo para varredura do cromossomo em busca da posição do centrômero e conseguinte medição dos braços. Os resultados da segmentação foram satisfatórios, apresentando uma eficiência de 91,25%. Já a classificação apresentou uma eficiência menor, em torno de 63,64%, conferindo ao sistema uma eficiência global de aproximadamente 60,00%.

Palavras-chave: Visão Computacional; Cariotipagem; Cromossomo.

ABSTRACT

GOEDERT, Matheus L. **System for assembling fish karyotype based on digital image processing**. 2018. Dissertação (Mestre em Tecnologias Computacionais para o Agronegócio) - Universidade Tecnológica Federal do Paraná.

By the principle that evolutionary explanations are usually genetic explanations, obtaining information about the species genome is fundamental in the validation process of some evolutionary theories. In this kind of research, the cytogeneticists use some tools, such as the assembly of karyotypes and their representation, which allows the researcher to identify morphological differences and similarities in the chromosomes, in a relatively simple way. However, this task can be kind of laborious and exhaustive. Therefore, many computational systems involving Digital Image Processing (DIP) have been recommended with the purpose of assisting in this activity. It is notable that DIP presents itself as an important tool for the processes automation, either by its agility, precision or comfort. Thereby, this work aims to develop a system based on the concepts of computer vision for automatic classification of fish chromosomes from individual metaphases images. In order to do this, it was used DIP techniques implemented in C ++ through the OpenCV library. The software used light field microscopy images of mitotic metaphases from individual cells of the *Hoplias malabaricus* (traíra) fish, collected from the Molecular and Chromosome Biodiversity Laboratory database of UTFPR Santa Helena campus. Segmentation was the first stage of image processing, and it has been divided into background elimination, first segmentation and second segmentation. This step's fundamentals were the automatic image thresholding, eliminating its background, and the separation of joined chromosomes, using a k-means clustering algorithm. Due to the colors and shades diversity of its respective backgrounds, this step can be considered the most challenging and consequently the most extensive and computationally costly. Ultimately, the chromosomes individualized images were classified according to Levan (1964) proposal. An algorithm was developed in order to scan the chromosome in the searching of centromere position and posteriorly to measure the arms. The segmentation results were satisfactory, presenting 91.25% efficiency. However, the classification showed a lower efficiency, around 63.64%, giving the system an overall efficiency of approximately 60.00%.

Palavras-chave: Computer vision; Karyotyping; Chromosome.

LISTA DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1 - Ampliações sucessivas do organismo humano até o DNA | 12 |
| Figura 2 - Morfologia do cromossomo metafásico..... | 15 |
| Figura 3 - Cariótipo (a) e ideograma (b) do peixe <i>Hoplias malabaricus</i> | 16 |
| Figura 4 - Interação entre as diversas áreas relacionadas a sistemas de imagens digitais | 18 |
| Figura 5 - Histogramas de intensidade que podem ser divididos (a) por um limiar único e (b) por limiares duplos | 21 |
| Figura 6 – Fluxograma das etapas do sistema proposto..... | 27 |
| Figura 7 - Metáfases mitóticas de peixe obtidas por microscopia de campo claro utilizando diferentes filtros de luz | 28 |
| Figura 8 – Histogramas das imagens de referência | 30 |
| Figura 9 – Histogramas aproximados das imagens de referência 1 e 2..... | 30 |
| Figura 10 – Sequência de operações do processo de eliminação do fundo: (a) original; (b) blur (3x3); (c) cinza; (d) limiarização; (e) 3 x dilatação (3x3); (f) 3 x erosão (3x3); (g) subtração | 32 |
| Figura 11 – Gráfico do valor do limiar em relação a intensidade do fundo..... | 34 |
| Figura 12 – ROI segmentada (a) com ruído e (b) limpa | 34 |
| Figura 13 – ROIs com contornos preenchidos e retângulos englobantes | 35 |
| Figura 14 – ROIs recortadas da imagem original em um fundo branco | 36 |
| Figura 15 – Grupo de cromossomos unidos (a) com os pixels agrupados em três cores (<i>k-means</i>), (b) cromossomos da imagem original individualizados | 37 |
| Figura 16 – Cromossomo individualizado com binarização invertida | 38 |
| Figura 17 – Linhas correspondentes ao centrômero e comprimento dos braços | 38 |
| Figura 18 – Imagens após a eliminação do fundo com (a) limiar alto e (b) limiar baixo | 39 |
| Figura 19 – Imagens originais e sem o fundo..... | 40 |
| Figura 20 – Grupo de cromossomos unidos (a) recortados da imagem original e (b) com os pixels agrupados em três cores (<i>k-means</i>)..... | 41 |
| Figura 21 –Cromossomos com os pixels agrupados por cores (<i>k-means</i>), (a) c com a borda em destaque, (b) unidos com apenas duas cores e (c) comprometido | 42 |
| Figura 22 –Cromossomos segmentados do grupo da Figura 10..... | 44 |
| Figura 23 – Imagem original desfocada e cromossomos unidos..... | 44 |

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO | 9 |
| 2 OBJETIVOS..... | 11 |
| 2.1 OBJETIVO GERAL | 11 |
| 2.2 OBJETIVOS ESPECÍFICOS | 11 |
| 3 REVISÃO DE LITERATURA | 12 |
| 3.1 CITOGENÉTICA | 12 |
| 3.1.1 Citogenética de Peixes..... | 13 |
| 3.1.2 Cromossomo | 14 |
| 3.1.3 Cariótipo | 15 |
| 3.2 PROCESSAMENTO DIGITAL DE IMAGENS | 17 |
| 3.2.1 Imagem Digital | 19 |
| 3.2.2 Filtros Espaciais de Suavização | 20 |
| 3.2.3 Limiarização | 20 |
| 3.2.4 Morfologia Matemática | 21 |
| 3.2.5 Método de Agrupamento <i>k-means</i> | 22 |
| 3.3 ESTADO DA ARTE | 23 |
| 4 METODOLOGIA | 27 |
| 4.1 AQUISIÇÃO DA IMAGEM | 28 |
| 4.2 SEGMENTAÇÃO | 29 |
| 4.2.1 Eliminação do Fundo..... | 29 |
| 4.2.2 Primeira Segmentação | 34 |
| 4.2.3 Segunda Segmentação | 36 |
| 4.3 CLASSIFICAÇÃO..... | 37 |
| 5 RESULTADOS E DISCUSSÃO | 39 |
| 5.1 SEGMENTAÇÃO | 39 |
| 5.2 CLASSIFICAÇÃO..... | 44 |
| 6 CONCLUSÃO | 46 |
| 7 SUGESTÕES PARA TRABALHOS FUTUROS | 47 |
| REFERÊNCIAS..... | 48 |

1 INTRODUÇÃO

A busca pela compreensão dos fenômenos naturais, como a origem e evolução da vida, levou ao surgimento das diversas áreas do conhecimento, como por exemplo, a Genética e a Citologia. Com a expansão dos conhecimentos, estas duas áreas da Biologia passaram a sobrepor seus conhecimentos em uma área comum, denominada Citogenética. De maneira geral e simplista, pode-se dizer que esta área da ciência envolve todo e qualquer estudo relativo aos cromossomos.

Partindo do princípio de que as explicações evolutivas, geralmente, são explicações genéticas, obter informações referentes ao genoma das espécies mostra-se fundamental no processo de validação de algumas teorias. Com isso, nos últimos anos, diversos estudos citogenéticos vêm sendo desenvolvidos nos mais variados grupos de seres vivos, tendo destaque para as pesquisas realizadas em peixes neotropicais.

Segundo Nelson (2006), os peixes oferecem várias oportunidades de pesquisa devido, principalmente, a sua diversidade morfológica e de habitats que exploram. Essa grande diversidade justifica o interesse de pesquisadores, fazendo com que diversos estudos de caracterização cromossômica sejam desenvolvidos neste grupo de vertebrados, a fim de subsidiar um melhor entendimento acerca da evolução das espécies fomentando o manejo adequado e a conservação da biodiversidade.

Na realização deste tipo de pesquisa os citogeneticistas lançam mão de algumas ferramentas, como a montagem de cariótipos e sua representação gráfica através de cariogramas. Trata-se de uma imagem na qual os cromossomos são classificados e ordenados segundo critérios morfológicos propostos por Levan *et al.* (1964). Esta ferramenta permite ao pesquisador identificar diferenças e semelhanças morfológicas nos cromossomos, de forma relativamente simples.

No entanto, a montagem do cariótipo propriamente dita não é o foco do cientista, mas sim a sua interpretação. Adicionalmente, esta tarefa pode ser considerada trabalhosa e exaustiva. Frente a isso, diversos sistemas computacionais vêm sendo propostos com o intuito de auxiliar neste processo, principalmente envolvendo cromossomos humanos e a detecção de síndromes relacionadas com alterações cromossômicas numéricas e estruturais.

Apesar de existirem diferentes *softwares* para produção de cariótipos, a maioria deles requer certa interação com o usuário. Neste contexto, a utilização dos conceitos de visão computacional mostra-se uma opção válida para automatizar o processo de montagem do cariótipo. A visão computacional é uma das áreas do Processamento Digital de Imagens (PDI) em pleno desenvolvimento nos últimos anos, tendo como principal desafio desenvolver sistemas autônomos capazes de reagir a estímulos visuais auxiliando, assim, na resolução de problemas altamente complexos. Consiste na extração de informações das imagens para identificação e classificação dos objetos de interesse, relacionando ferramentas de análise de imagem com técnicas de inteligência artificial (CONCI *et al.*, 2008; PEDRINI; SCHARTZ, 2008).

É notável que o PDI apresenta-se como uma importante ferramenta para a automação de processos, seja pela agilidade, precisão ou ainda pelo conforto. Além disso, algumas técnicas de PDI como a visão computacional e reconhecimento de padrões têm alcançado excelentes resultados, superando até mesmo os sistemas naturais. O que justifica a grande quantidade de técnicas desenvolvidas, estatísticas e sintáticas, para a classificação de padrões (TING, 2009; PLOTZE; BRUNO, 2007; CONCI *et al.* 2008).

Considerando o contexto supracitado, este trabalho teve por objetivo desenvolver um sistema baseado nos conceitos de visão computacional para classificação de cromossomos de peixes a partir de imagens individuais de metáfases. Para isso foram utilizadas técnicas de PDI implementadas em C++, por meio da biblioteca OpenCV.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Criar um sistema computacional, por meio de técnicas de processamento digital de imagens, para automatizar a classificação de cromossomos de peixes, a partir de imagens de metáfases obtidas por microscopia de campo claro.

2.2 OBJETIVOS ESPECÍFICOS

- a) Obter imagens de metáfases mitóticas, capturadas por microscopia de campo claro, do peixe *Hoplias malabaricus*, originando um banco de dados da UTFPR, *campus* Santa Helena;
- b) Manipular as imagens e extrair dados por meio de técnicas de PDI, para segmentar os cromossomos e rotacioná-los verticalmente;
- c) Medir o tamanho dos braços dos cromossomos e classificá-los quanto à sua morfologia e posição do centrômero;
- d) Comparar a eficiência do sistema com classificações obtidas manualmente.

3 REVISÃO DE LITERATURA

3.1 CITOGENÉTICA

Embora os citologistas já soubessem que o núcleo é uma estrutura comum a todas as células, que se divide formando outras novas e que deste processo surgem estruturas alongadas, mais tarde chamadas de cromossomos. A ciência da Genética propriamente dita, somente teve início com os trabalhos de Gregor Mendel em 1865, quando o cientista austríaco obteve os resultados de seus experimentos com o cruzamento entre diferentes linhagens de ervilhas (GUERRA, 1988; GRIFFITHS *et al.*, 2008).

Deste marco até a década de 1920, o grande volume de evidências levou a conclusão de que o DNA é o material genético. Contudo, somente no início do século XX foi possível afirmar com clareza que são os cromossomos do núcleo celular os portadores das informações que especificam o desenvolvimento dos organismos. A Figura 1 ilustra ampliações sucessivas do organismo humano, sua constituição celular e estruturas adjacentes (GRIFFITHS *et al.*, 2008).

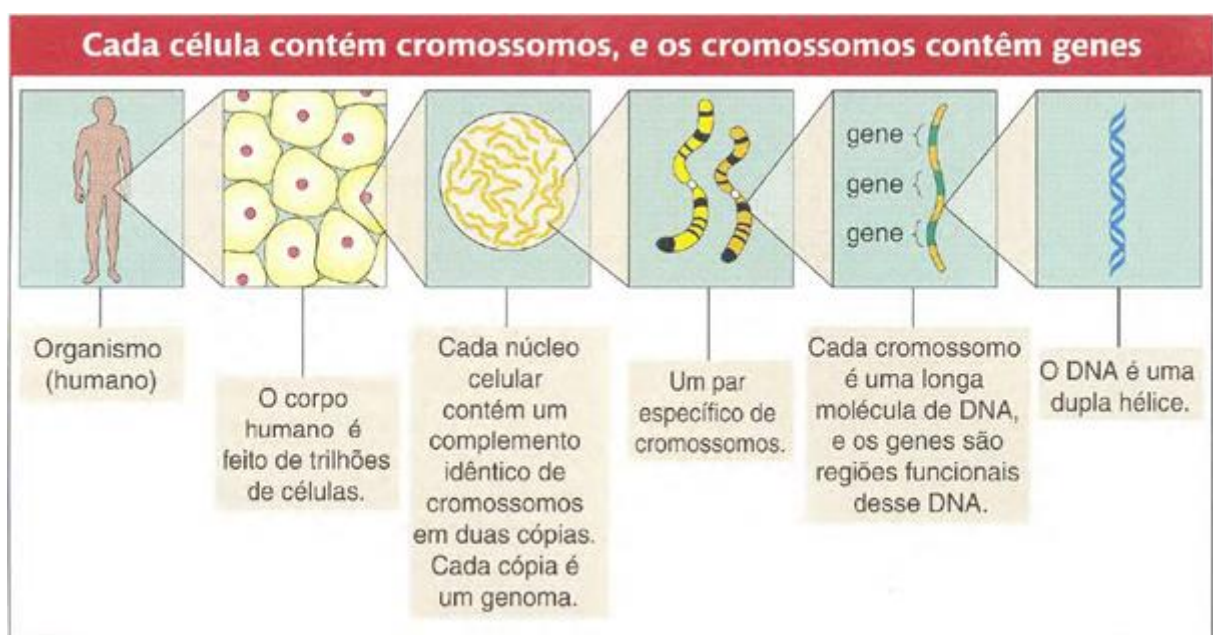


Figura 1 - Ampliações sucessivas do organismo humano até o DNA
Fonte: Griffiths et al. (2008)

Com a proposta da teoria da herança cromossômica, a Citologia e a Genética sobrepuseram seus conhecimentos numa área comum, posteriormente denominada Citogenética. Esta área da ciência expandiu-se ao longo dos anos envolvendo diversos outros campos da biologia como, por exemplo, a medicina clínica e o melhoramento animal e vegetal. “Atualmente a Citogenética refere-se a todo e qualquer estudo relativo ao cromossomo, tanto no que diz respeito a sua morfologia, organização, função e replicação quanto a sua variação e evolução” (GUERRA, 1988).

3.1.1 Citogenética de Peixes

Os peixes apresentam uma enorme diversidade morfológica e ocupam uma grande variedade de habitats, dificultando relativamente à compreensão da sua história evolutiva. No entanto, este cenário proporciona diferentes possibilidades de análises que se apresentam como uma oportunidade promissora ao avanço de estudos no âmbito citogenético (NELSON, 2006).

A citogenética teve início na década de 70 e hoje reúne um dos maiores grupos de pesquisadores voltados a um campo de interesse comum, com diversos laboratórios de pesquisa distribuídos entre diferentes estados brasileiros, desde o Amazonas até o Rio Grande do Sul, alguns já mais antigos e bem consolidados, ao lado de outros mais recentes e em fase de consolidação.

Estudos genéticos têm se mostrado decisivo para o conhecimento e entendimento da diversidade e da evolução ictiofaunística mundial. Os peixes representam o maior grupo de vertebrados existentes no planeta, com 34.725 espécies válidas (ESCHMEYER; FONG, 2018). Na região Neotropical, que compreende a América Central, a parte sul do México, a península da Baixa Califórnia, o sul da Flórida, todas as ilhas do Caribe e a América do Sul, encontramos a ictiofauna de água doce mais diversificada e rica do planeta, com mais de 5.700 espécies conhecidas e muitas outras sendo descritas anualmente (ALBERT; REIS, 2011).

Dentro deste universo de espécies destacam-se os representantes da superordem Ostariophysi, representando mais de 70% da ictiofauna de água doce Neotropical (REIS *et al.*, 2003).

A ictiofauna brasileira é distribuída de forma bastante heterogênea, dados de

Reis *et al.*, (2003), indicam a ocorrência de cerca de 2.300 espécies dulcícolas, enquanto que marinhas em torno de 1.298 espécies (MENEZES *et al.*, 2003). Apenas cinco ordens detêm mais de 95% das espécies de peixes de água doce conhecidas, sendo Characiformes a mais numerosa, seguida dos Siluriformes, Perciformes, Cyprinodontiformes e Gymnotiformes (ROSA; LIMA, 2008).

A obtenção de informações referentes ao genoma tais como polimorfismos estruturais e/ou numéricos, rearranjos cromossômicos e sistema de cromossomos sexuais heteromórficos, favorece a compreensão dos processos evolutivos que originaram a diversidade desta ictiofauna (BLANCO, 2012).

Nos últimos anos, diversos estudos relacionados à citogenética de peixes vêm sendo desenvolvidos no Brasil e demais países da região neotropical, dos quais pode-se citar os realizados por: Sánchez-Romero *et al.* (2015), Blanco *et al.* (2014), Lui *et al.* (2013), Alves *et al.* (2012), Carvalho *et al.* (2011), dentre outros. Graças a estudos desta natureza, diversas informações relevantes vêm sendo levantadas, sejam inéditas, conforme o trabalho de Traldi *et al.* (2012), ou comparativas, como Jacobina *et al.* (2009).

3.1.2 Cromossomo

Uma sequência completa de informações genéticas (DNA) de um organismo é denominada genoma, nos eucariontes é encontrado no núcleo celular dividido em unidades chamadas cromossomos, que contém uma molécula de DNA (Figura 1). Dentro de uma mesma espécie de indivíduos, o conjunto cromossômico geralmente apresenta o mesmo aspecto e quantidade de cromossomos (GRIFFITHS *et al.*, 2008). Esses cromossomos se condensam durante o período de divisão celular a fim de viabilizar uma divisão igualitária do material genético entre as células filhas recém formadas.

De acordo com Guerra (1988), na maioria das vezes que é feita a menção a cromossomos, esta refere-se a cromossomos metafásicos mitóticos de eucariotas. Isto ocorre, pois na fase da metáfase da divisão celular é que encontra-se o maior grau de condensação cromossômica. O cromossomo metafásico é formado por duas subunidades paralelas (Figura 2) denominadas cromátides, que por sua vez são

constituídas por um único fio de DNA, associado a um arcabouço proteico denominado *scaffold*. As cromátides são unidas por uma região menos condensada denominada de constrição primária ou centrômero, que divide a cromátide em dois segmentos chamados braços cromossômicos. A extremidade terminal do braço cromossômico é denominada telômero. Em um ou poucos cromossomos do conjunto cromossômico de cada espécie podem ocorrer uma constrição secundária. Pode-se dizer que, do ponto de vista funcional, as regiões que merecem maior atenção são o centrômero, a constrição secundária e o telômero.

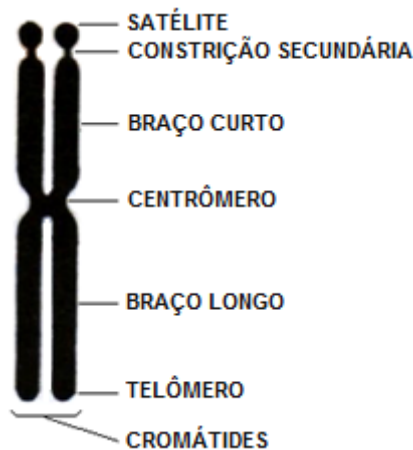


Figura 2 - Morfologia do cromossomo metafásico
Fonte: Guerra (1988)

3.1.3 Cariótipo

As explicações evolutivas são geralmente explicações genéticas, e de acordo com Griffiths *et al.* (2008), “a análise genética, usando organismos com mutações, é um método poderoso para investigar vias bioquímicas, fisiológicas e desenvolvimentais”.

Neste contexto, em estudos citogenéticos, os cientistas utilizam-se de algumas metodologias e ferramentas, buscando identificar diferenças e semelhanças genéticas. Uma das técnicas, considerada relativamente simples, porém de fundamental importância, é a representação do cariótipo, haja vista que fornece informações essenciais, principalmente em estudos de ordem evolutivo e

citotaxonômico (KLACZKO; VIEIRA, 2003; GUERRA *et al.*, 1988).

Um cariótipo é a caracterização do conjunto cromossômico de uma célula organizado por número, forma, tamanho e posição do centrômero. A representação do cariótipo conhecida como cariograma é obtida através de uma fotografia ou desenho detalhado de uma metáfase, de onde os cromossomos são recortados e os homólogos emparelhados e enumerados sistematicamente em ordem decrescente de tamanho, conforme Figura 3a (ALBERTS *et al.* 2010; KLACZKO; VIEIRA, 2003).

Outra forma de apresentar as informações cariotípicas de um organismo é através de um ideograma (Figura 3b), que consiste em uma representação esquemática dos cromossomos, levando em consideração seu tamanho e posição do centrômero (GUERRA *et al.*, 1988).

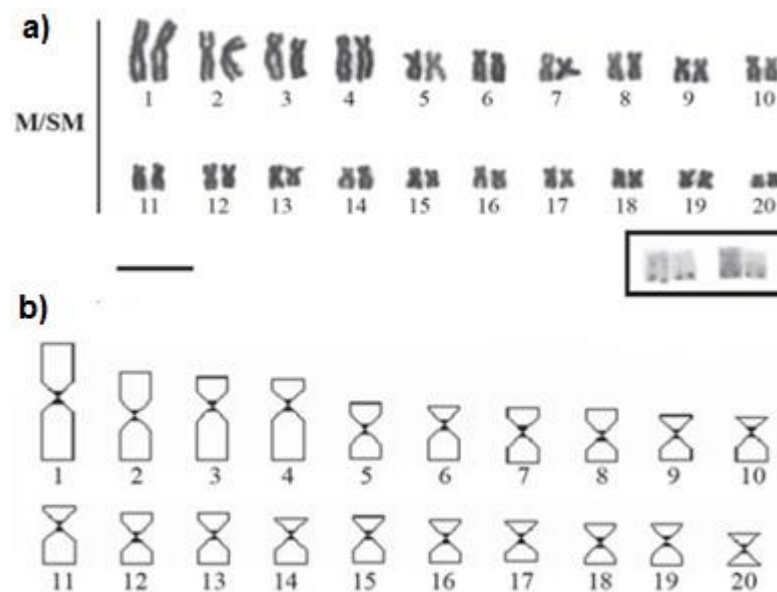


Figura 3 - Cariótipo (a) e ideograma (b) do peixe *Hoplias malabaricus*
Fonte: Jacobina (2009)

Para obter-se a fotografia, utiliza-se de material apropriado rico em divisões celulares, mitóticas ou meióticas, que pode ser tratado em colchicina para aumentar a probabilidade de se observar células em divisão, na metáfase, fase na qual os cromossomos encontram-se com maior grau de condensação. Feito isso o material é corado, esmagado e analisado em microscópio (KLACZKO; VIEIRA, 2003).

No cariótipo, os cromossomos devem ser agrupados em classes

morfológicas que são delimitadas pela razão entre braços (RB), que consiste na divisão do maior braço cromossômico maior (q) pelo braço menor (p), resultando nas seguintes morfologias: metacêntricos (RB de 1,0 a 1,70), submetacêntricos (RB de 1,71 a 3,0), subtelocêntricos (RB de 3,01 a 7,0) e acrocêntricos (RB > 7,0) (LEVAN, 1964). Esta divisão dos cromossomos em diferentes morfologias auxilia os citogeneticistas a identificar diferenças e similaridades importantes na caracterização cromossômica.

3.2 PROCESSAMENTO DIGITAL DE IMAGENS

A visão é considerada o mais avançado sentido humano, todavia os seres humanos são limitados à banda visual do espectro eletromagnético. Diferentemente, o Processamento Digital de Imagens (PDI), dispõe de equipamentos capazes de cobrir praticamente todo o espectro eletromagnético. Esta tecnologia permite trabalhar com imagens geradas desde câmeras digitais e scanners, até sensores de ultrassom e infravermelho, além de satélites, microscópios, dentre outros. Tantas possibilidades conferem a esta tecnologia um amplo campo de aplicações (GONZALEZ; WOODS, 2011; CONCI *et al.* 2008).

Conforme relata Gonzalez e Woods (2011), distinguir o ponto em que o PDI termina e outras áreas relacionadas começam é um assunto em desacordo entre os diferentes autores do tema. Algumas vezes, uma distinção é traçada definindo o PDI como uma disciplina na qual tanto a entrada quanto a saída de um processo são imagens. No entanto essa fronteira é restritiva e, de certa forma, artificial.

Os autores sugerem três níveis sequenciais de processos computacionais nessa linha, classificando-os como nível baixo, médio e alto. O primeiro consiste na realização de operações primitivas, como por exemplo, o pré-processamento de imagens para redução de ruídos, o realce de contraste e o aguçamento de imagens. O processamento nível médio refere-se às tarefas como a segmentação de regiões e objetos, bem como à sua redução a uma forma adequada para o processamento digital e classificação (reconhecimento). O último processo, o de nível alto, busca “dar sentido” ao reconhecimento dos objetos, e por fim realizar funções cognitivas normalmente associadas à visão humana (ramo da inteligência artificial).

Os limites do PDI estabelecidos por Gonzalez e Woods (2011), consideram os processamentos de nível baixo e nível médio, assumindo dois tipos de métodos: os que tanto a entrada quanto a saída do processo são imagens; e os que têm como entrada uma imagem, mas sua saída são atributos obtidos destas. Outra abordagem interessante é a proposta por Conci *et al.* (2008), para os autores os sistemas de imagens digitais possuem um escopo bem definido, abrangendo diversas áreas que relacionam dados e imagens, tais como o processamento de imagens, a análise de imagens, a síntese de imagens, a visão computacional e o reconhecimento de padrões. A diferença entre estas áreas está na relação entre o uso de técnicas de inteligência artificial, dados e imagens, conforme apresentado na Figura 4.

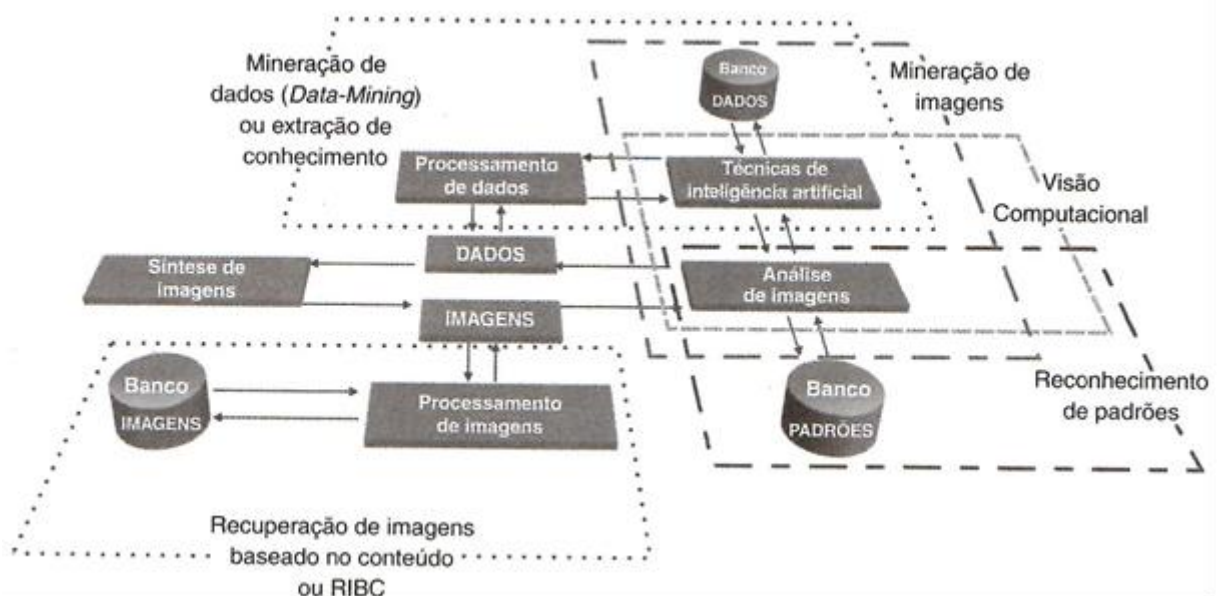


Figura 4 - Interação entre as diversas áreas relacionadas a sistemas de imagens digitais
Fonte: Conci *et al.* (2008)

Uma das áreas em pleno desenvolvimento nos últimos anos é a visão computacional, que tem como principal desafio desenvolver sistemas autônomos capazes de reagir a estímulos visuais, auxiliando na resolução de problemas altamente complexos. Consiste na extração de informações das imagens para identificação e na classificação dos objetos de interesse, relacionando ferramentas de análise de imagem com técnicas de inteligência artificial (Figura 4) (CONCI *et al.* 2008; PEDRINI; SCHATZ, 2008).

Para Pedrini e Schwartz (2008), o grande avanço tecnológico, aliado ao

desenvolvimento de novos algoritmos, tem promovido o aumento da utilização de técnicas de PDI, como a visão computacional, nas mais diversas áreas: na medicina, em exames como ecocardiogramas (SKORTON *et al.*, 1985); na biologia, buscando facilitar a classificação de cromossomos (RAMAKUMAR *et al.*, 2015); no controle de qualidade, identificando peças defeituosas (YANG *et al.*, 2015); na geração de energia elétrica, auxiliando na estimativa da incidência de radiação solar em painéis fotovoltaicos (ALONSO-MONTESINOS; BEATLLES, 2015), na manutenção de equipamentos, por meio de análise de imagens térmicas (TAHERI-GARAVAND *et al.*, 2015), entre outras.

3.2.1 Imagem Digital

Uma imagem digital é descrita espacialmente e em amplitude por uma função $f(x, y)$ e pode ser representada por uma matriz $N \times M$, onde as linhas e colunas identificam um ponto na imagem (x, y) , com valor correspondente ao nível de cinza naquele ponto. Na Equação 1, a função $f(x, y)$ representa a interação entre a iluminância $i(x, y)$, que está associada a quantidade de luz incidente sobre o objeto no ponto (x, y) , e a refletância $r(x, y)$ que indica a quantidade de luz refletida no ponto (FILHO; NETO, 1999).

$$f(x, y) = i(x, y) \times r(x, y) \quad (1)$$

Cada ponto na grade bidimensional que representa a imagem digital é denominado pixel. O pixel é caracterizado pelo valor da tonalidade de cor e pela sua localização no arranjo de pixels de uma imagem bidimensional, que é determinada utilizando-se a notação matricial, em que o índice m representa a posição da linha na qual o pixel se encontra e o termo n a posição da coluna. Em uma imagem digital representada no sistema RGB, um pixel pode ser visto como um vetor cujas componentes representam as intensidades de vermelho, verde e azul da imagem no ponto (x, y) . Assim, a imagem colorida pode ser vista como a composição de três imagens monocromáticas (QUEIROZ; GOMES, 2001).

Em termos de processamento digital de imagens, um dos sistemas mais utilizados para representação de cores é o RGB. Esse modelo é baseado em um sistema coordenadas cartesianas, que pode ser visto com um cubo, no qual os valores RGB primários estão em três vértices, as cores secundárias estão em outros três vértices, o preto está na origem e o branco no vértice mais distante da origem. A escala de cinza estende-se do preto ao branco através de um segmento de reta que une esses dois pontos (FILHO; NETO, 1999).

3.2.2 Filtros Espaciais de Suavização

No processamento digital de imagens, o uso de máscaras espaciais é denominado filtragem espacial. A suavização de imagens no domínio espacial baseia-se no uso máscaras de convolução adequadas para o borramento da imagem e para a redução de ruídos nela presentes. Essas atividades são realizadas na etapa de pré-processamento para eliminar detalhes que não são de interesse nas etapas subsequentes do processamento, como remoção de pequenos objetos e conexão de pequenas descontinuidades em linhas ou curvas (FILHO; NETO, 1999; GONZALEZ; WOODS, 2010).

No filtro espacial linear de suavização, a resposta é a média dos pixels contidos na vizinhança da máscara de filtragem. Esses filtros são chamados de filtros de média ou também filtros passa-baixa. Baseiam-se na ideia que ao substituir o valor de cada pixel de uma imagem pela média dos níveis de intensidade dos pixels situados na vizinhança, definida pela máscara, o processo resulta em uma imagem com perda de nitidez, ou seja, com redução das transições abruptas nas intensidades (GONZALEZ; WOODS, 2010).

3.2.3 Limiarização

A limiarização, ou binarização baseia-se no histograma de intensidade da imagem para determinar um limiar T . O histograma de uma imagem explica a

distribuição estatística dos seus níveis de cinza. Trata-se de uma representação gráfica que fornece o número de pixels associado a cada nível de intensidade presente na imagem (FILHO; NETO, 1999).

Um exemplo trivial é o apresentado por Gonzalez e Woods (2011), em que considera uma imagem $f(x,y)$ composta por objetos claros em um fundo escuro, formando dois grupos distintos de intensidade que podem ser divididos por um limiar T , conforme a Figura 5. Logo, pode-se dizer que qualquer pixel (x,y) da imagem que apresente intensidade $f(x,y) > T$ pode ser classificado como objeto, e contrariamente a isso, quando $f(x,y) < T$ é classificado como fundo. Sendo assim a região de interesse (ROI - *Region of Interest*), $g(x,y)$ é dada pela Equação 2.

$$g(x,y) = \begin{cases} 1 & \text{se } f(x,y) > T \\ 0 & \text{se } f(x,y) \leq T \end{cases} \quad (2)$$

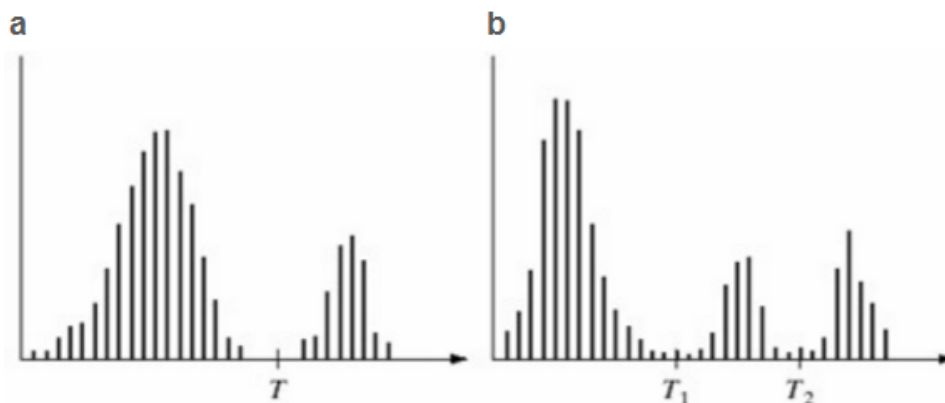


Figura 5 - Histogramas de intensidade que podem ser divididos (a) por um limiar único e (b) por limiares duplos

Fonte: Gonzales e Woods (2011)

3.2.4 Morfologia Matemática

A morfologia matemática é, primordialmente, a análise das estruturas geométricas presentes em uma imagem a partir da vizinhança de cada *pixel*. Esta vizinhança é delimitada por meio de um elemento estruturante, formado por um conjunto de elementos de fundo, com forma e dimensões completamente definidas. O processamento morfológico permite representar e descrever regiões como fronteiras e esqueletos, além de ser utilizado em técnicas de pré-processamento como

filtragem de ruídos, afinamento e poda. Para isso duas operações básicas da morfologia matemática binária são fundamentais, sendo elas: a erosão, que diminui ou afina os elementos da imagem; e a dilatação, que aumenta ou engrossa os objetos da imagem (FACON, 1996; GONZALEZ; WOODS, 2011).

A operação em que se realiza uma erosão seguida de uma dilatação, é chamada de *abertura*, sendo utilizada para suavizar contornos, romper istmos e eliminar saliências delgadas. Já a operação de *fechamento*, consiste em realizar-se uma dilatação seguida de uma erosão, esta técnica também é utilizada para suavizar contornos, mas ao contrário da *abertura*, preenche pequenas lacunas, fundindo descontinuidades estreitas e eliminando pequenos buracos (FACON, 1996; FILHO; NETO, 1999).

3.2.5 Método de Agrupamento *k-means*

O algoritmo *k-means* é uma técnica utilizada em mineração de dados, para particionar um conjunto de dados em k grupos (*clusters*). Pode ser utilizado para agrupar as intensidades dos pixels de uma imagem, segmentando a imagem em k regiões, atribuindo para cada pixel o valor de intensidade do *cluster* a que este foi considerado (MACQUEEN, 1967; JAIN, 2010).

O algoritmo inicia determinando-se k elementos denominados sementes, podendo ser arbitrados aleatoriamente. Com isso, é calculada a distância de cada elemento do conjunto de dados em relação às sementes, considerando os elementos como pertencentes ao *cluster* aqueles que possuem as menores distâncias. Feito isso, o centroide de cada grupo é determinado pela média dos valores dos pontos, sendo este adotado como referência para a próxima iteração, em que as distâncias são novamente calculadas em relação a este ponto. Sendo assim, a cada iteração um novo centroide é calculado para cada grupo e este processo se repete até que os centroides se mantenham fixos (TAN *et al.* 2018; JAIN, 2010).

3.3 ESTADO DA ARTE

O desenvolvimento de ferramentas que buscam automatizar o processo de montagem de cariótipos, a partir de uma imagem de uma célula em metáfase, há anos vem sendo alvo de pesquisas por parte de citogeneticistas em conjunto com desenvolvedores de programas. O avanço da tecnologia e, conseqüentemente, das técnicas de PDI, tem colaborado com este cenário e diversos sistemas vêm sendo propostos neste sentido, tornando cada vez menor a interação entre software e usuário.

Uma das primeiras ferramentas para análise automática de cromossomos foi o FIDAC, um equipamento desenvolvido por Ledley (1964) na década de 60, que integrava um computador (IBM 7094) a um circuito eletrônico de digitalização de imagens em tempo real. Obtida a imagem digital, um sistema próprio, o FIDACSYS, realizava diversas operações para segmentar, contar, medir, parear e classificar os cromossomos. Para isso, primeiramente o sistema convertia a imagem para escala de cinza, eliminava o fundo, a partir de um valor limiar de cinza, e gerava linhas de contorno por meio de vários segmentos de reta, o que possibilitava extrair características como número, tamanho e área dos cromossomos para a realização de cálculos para classificação e pareamento.

Este equipamento, embora não realizasse a montagem do cariótipo, fornecia informações essenciais que facilitavam a tarefa dos geneticistas. Além disso, foi um marco importante para alavancar novas pesquisas acerca do tema. Nos trabalhos citados a seguir é possível perceber que, de maneira geral, as etapas do processamento são basicamente as mesmas, com algumas variações de metodologias e ferramentas aplicadas, que foram surgindo ao longo do tempo.

Nickolls *et al.* (1981) propuseram um sistema para análise de cromossomos humanos e testaram diferentes técnicas no pré-processamento das imagens com o objetivo de avaliar o impacto na localização automática do centrômero. Foram utilizadas variações de quatro metodologias: a convolução, através de filtro laplaciano; a Transformada de Fourier; transformações não-lineares; e a manipulação do contraste da imagem. Destas, o filtro de Laplace foi a que apresentou melhores resultados, haja vista que as demais técnicas de filtragem fragmentavam os cromossomos da imagem, dificultando as etapas seguintes.

Um dos primeiros sistemas comerciais para auxílio na montagem do cariótipo foi o software Athena proposto por Vliet *et al.* (1989), que funcionava como um módulo do sistema Acuity para análise de células. Seu funcionamento consistia basicamente na segmentação da imagem, separação dos cromossomos que se tocam ou estão sobrepostos e classificação baseada no comprimento, posição do centrômero e padrão de banda. Contudo, este processo não acontecia de forma automática, eram necessárias diversas parametrizações e interferências realizadas pelo usuário, principalmente no processo de separação dos cromossomos, o que dificultava sua operação por aqueles que não estavam habituados com o ambiente computacional.

Em 1989, Piper e Granum (1989) propuseram e compararam quatro metodologias com diferentes procedimentos para classificação totalmente automática de cromossomos humanos. Para isso utilizaram técnicas de reconhecimento de padrões e classificadores estatísticos baseados em três bancos de dados distintos. Os autores concluíram que cada metodologia apresenta resultados diferentes para cada base de dados, além do mais, a utilização das diferentes bases de dados dentro da mesma metodologia elevou em 10,00% a taxa de erro dos sistemas.

Nesta mesma linha, também utilizando imagens de cromossomos humanos, Wu *et al.* (1990) comparou os resultados obtidos por uma rede neural artificial (RNA) do tipo *multi-layer perceptron*, implementada e treinada pelo algoritmo *back-propagation*, com um classificador estatístico baseado no Teorema de Bayes. Alegando que o uso de RNAs confere melhor desempenho ao sistema quando comparadas aos modelos estatísticos tradicionais considerados limitados.

Wu *et al.* (1990), comprovou tal afirmação em seu estudo, onde se obteve uma taxa de erro de 24,20% para a rede neural e 31,40% para o modelo estatístico. Além disso, os classificadores estatísticos não possuem a capacidade de se adaptarem ou “aprenderem” ao longo da sua utilização. Contudo, mais tarde, Wu (2000) e Wu *et al.* (2005) avaliaram novamente o modelo baseado no Teorema de Bayes variando a técnica de segmentação e obtiveram um resultado bastante satisfatório com um desempenho de 96,80% e aproximadamente 95,00%, respectivamente.

Utilizando o mesmo tipo de RNA (*multi-layer perceptron*) e treinamento (*back-propagation*), Cho e Hong (1994) e Cho (2000), examinaram 460 imagens de cromossomos humanos, e obtiveram bons resultados com uma taxa de erro de 6,52%.

Em outra oportunidade Cho *et al.* (2004), utilizando um algoritmo de pré-seleção, conseguiram melhorar os resultados da RNA atingindo uma taxa de erro de 5,90%.

Lerner *et al.* (1994), utilizando outros dois algoritmos de treinamento, o *knock-out* e o *Principal Component Analysis* (PCA), chegaram à mesma conclusão de que o uso de RNAs é promissor para a classificação de cromossomos humanos. Visando reduzir o tempo de processamento através da redução do tamanho da RNA, Delshadpour (2003), também obteve resultados satisfatórios em seu sistema, que apresentou uma taxa de reconhecimento de 88,30%.

Existem ainda outros trabalhos acerca do tema, envolvendo RNAs, dentre eles cita-se os de Moradi e Setarehdan (2006), com uma taxa de reconhecimento de 98,60%; Javan-Roshtkhari e Setarehdan (2007) com uma taxa de 95,30%; Wang *et al.* (2008) com uma taxa superior a 90,00%; Oskouei e Shanbehzadeh (2010) com 95,50% de taxa de reconhecimento; Yan e Bai (2013) com uma taxa de 96,8%; Gagula-Palalic e Can (2014) com uma taxa de classificação de 98,27%.

Outra metodologia utilizada para classificação de cromossomos é o classificador *Support Vector Machines* (SVM), utilizado por Kou e Zhang (2002) em sua pesquisa, com resultados superiores a 90,00% de taxa de acerto. Também utilizaram o SVM, Karvelis *et al.* (2009a) e Markou *et al.* (2012), atingindo bons resultados na classificação dos cromossomos.

Com o surgimento da metodologia M-FISH na década de 90, que resulta em imagens de cromossomos em fluorescência, surgiram também novos estudos propondo a classificação dos cromossomos com este tipo de imagem, haja vista que os sistemas até então desenvolvidos não funcionariam neste tipo de imagem. Dentre as pesquisas relacionadas podem-se citar as realizadas por: Schwartzkopf *et al.* (2005), em que buscou-se desenvolver um sistema capaz de segmentar e classificar cromossomos sobrepostos e agrupados por meio da classificação pixel-a-pixel da imagem, obtendo um desempenho de 90,00%; e Karvelis *et al.* (2006), que comparou a metodologia pixel-a-pixel com a sua proposta, que realiza a segmentação dos cromossomos com a ferramenta *watershed* seguida da classificação por meio do Teorema de Bayes. Neste experimento foram utilizadas 15 imagens, e os resultados obtidos foram favoráveis à sua pesquisa, que apresentou um desempenho médio de 89,00% versus 77,00% para a metodologia pixel-a-pixel.

Em outras publicações o autor realizou mudanças em seu sistema, testando e/ou comparando diferentes classificadores e sets de treinamento, Karvelis

et al. (2008) testou seu classificador Bayesiano com 183 imagens e obteve um desempenho de 82,50%. Na sequência, Karvelis *et al.* (2009b), atingiu um desempenho de 82,61%. Por fim, Karvelis *et al.* (2009a) comparou os resultados do classificador Bayesiano com o classificador SVM, obtendo uma performance de 70,04% e 80,20% respectivamente.

Existem ainda sistemas que não utilizam os conceitos de inteligência artificial ou modelos estatísticos para classificação dos cromossomos, a exemplo disso, Guimarães *et al.* (2003) propôs em seu estudo analisar e comparar características morfológicas do corpo dos cromossomos, como linhas de contorno e distâncias euclidianas.

Diante do exposto, percebe-se que, para cromossomos humanos, já foram realizados vários trabalhos abordando e comparando diferentes métodos com o foco na classificação. Diferentemente a isso, estudos mais recentes, têm demonstrado uma tendência de aprimoramento das técnicas de segmentação (CAO *et al.*, 2012), pareamento (FENG *et al.*, 2012; JANANI *et al.*, 2012) e seleção da metáfase a ser utilizada para análise (UTTAMATANIN *et al.*, 2013; RAMAKUMAR *et al.*, 2015).

Com o levantamento bibliográfico apresentado neste capítulo, foi possível perceber a escassez de estudos direcionados à classificação e cariotipagem automática de cromossomos de espécies que não a humana. Contudo, no contexto do estudo proposto, Bressane *et al.* (2005) desenvolveu em sua pesquisa um *software* para auxílio na obtenção de cariótipos de peixes.

Faria (2006), também propôs um programa automatizado para obtenção de cariótipo de peixes, utilizando inferência difusa na classificação dos cromossomos. Ainda assim, este sistema requer a intervenção do usuário na parametrização da imagem, e em eventuais correções que se façam necessárias, como por exemplo, erros na rotação e na segmentação de cromossomos sobrepostos. Os resultados da pesquisa de Faria (2006) foram comparados com cariótipos obtidos por dois grupos de citogeneticistas, sendo que para o primeiro grupo a classificação dos cromossomos concordou em 78,52% e para o segundo grupo 81,16%. Vale destacar que quando comparados, os cariótipos dos dois grupos concordaram em 70,74%, demonstrando o quanto a obtenção de cariótipos é subjetiva e sujeita a erros, o que ratifica o desenvolvimento de sistemas automatizados.

4 METODOLOGIA

Esta pesquisa consistiu no desenvolvimento de um sistema computacional, para classificação de cromossomos de peixes, conforme Levan (1964), em linguagem C++, utilizando uma biblioteca aberta (*open source*) multiplataforma denominada OpenCV. Nesta biblioteca estão contidas diversas ferramentas e metodologias de PDI que serão utilizadas para manipular as imagens digitais das metáfases. Portanto, a fase inicial da pesquisa consistiu em um levantamento documental de imagens de metáfases de *Hoplias malabaricus* (traíra), já obtidas por citogeneticistas.

O sistema proposto é constituído basicamente por três etapas: aquisição, segmentação e classificação. A segmentação pode ser considerada a etapa mais desafiadora, frente à diversidade das imagens de entrada, principalmente no que diz respeito às diferenças de cores e tonalidades dos seus respectivos fundos. Por isso, foi necessário subdividir esta etapa em três partes, apresentados juntamente com os demais, no fluxograma da Figura 6. Por fim, os cromossomos de seis imagens foram classificados manualmente, com o intuito de comparar com os resultados obtidos.

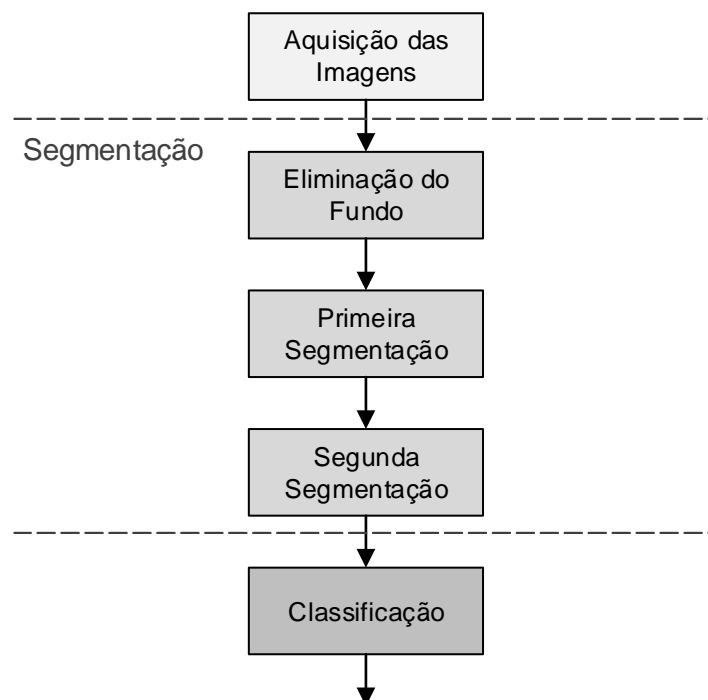


Figura 6 – Fluxograma das etapas do sistema proposto
Fonte: Autoria própria

4.1 AQUISIÇÃO DA IMAGEM

A primeira etapa de um sistema baseado em visão computacional é a aquisição da imagem, podendo ocorrer por meio de inúmeros tipos de sensores, ou ainda a partir de uma imagem já digitalizada. Esta pesquisa propôs realizar o processamento de imagens que deverão ser inseridas no *software*, logo, deverão ser previamente obtidas e digitalizadas pelo citogeneticista.

O sistema proposto tomou por base 100 imagens obtidas por meio de microscopia de campo claro de metáfases mitóticas de células individualizadas de peixes da espécie *Hoplias malabaricus* (Figura 7). Estas imagens foram fornecidas pelo banco de dados do Laboratório de Biodiversidade Molecular e Cromossômica do *campus* Santa Helena da UTFPR, onde foram capturadas por meio de uma câmera (Color Q – 5M pixels) acoplada a um microscópio trinocular (Olympus BX51).

Portanto, o sistema proposto atende aos padrões das imagens fornecidas, que estão no formato JPEG (*.jpg), com uma resolução de 5M *pixels* e dimensões de 1392 x 1040 *pixels*. A diversidade de cores dos fundos das imagens é atribuída ao uso de filtros de luz do microscópio, utilizados para modificar o comprimento de onda, a fim de melhorar o poder de resolução.

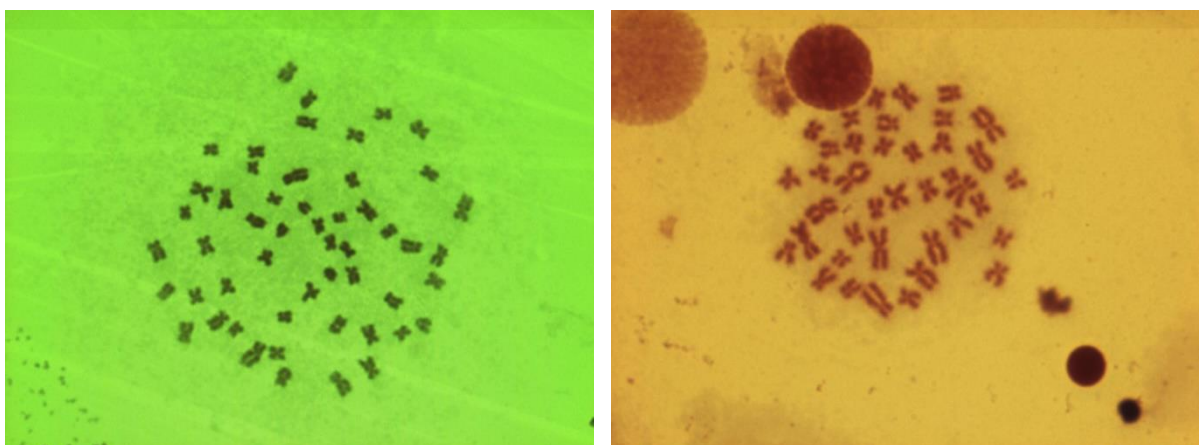


Figura 7 - Metáfases mitóticas de peixe obtidas por microscopia de campo claro utilizando diferentes filtros de luz
Fonte: Autoria própria

4.2 SEGMENTAÇÃO

Esta etapa pode ser considerada a mais morosa, frente às diferentes características das imagens de entrada, principalmente devido às diversidades de cores e tonalidades dos seus respectivos fundos, o que dificultou o processo de binarização da imagem. O objetivo dos três algoritmos descritos a seguir é retornar imagens individualizadas dos cromossomos, que serão as entradas da etapa de classificação.

4.2.1 Eliminação do Fundo

A partir desta etapa, a pesquisa assumiu um caráter experimental, haja vista a necessidade de verificação do comportamento da imagem quando submetida aos procedimentos subsequentes, e caso necessário, modificar parâmetros e/ou testar outras ferramentas em etapas anteriores. Devido à diversidade de cores e tonalidades dos fundos das imagens, e com o intuito de prover consistência ao sistema, desde o início os procedimentos foram executados em uma amostra de 11 imagens, das 100 obtidas. Estas imagens foram organizadas e escolhidas com base na intensidade do seu fundo.

Pode-se observar na Figura 7 que as imagens de entrada possuem dois grupos de intensidades bem definidos: o fundo e os objetos (cromossomos e ruídos), o que levou a hipótese de se utilizar os conceitos de limiarização para eliminação do fundo das imagens. Sendo assim, o primeiro passo foi realizar uma análise do histograma das intensidades de cinza das imagens a fim de confirmar o agrupamento dos valores em dois grupos dominantes e assim determinar o valor do limiar. Contudo, foi possível observar a presença de apenas um grande agrupamento, situado nos valores mais elevados de intensidade (Figura 8), ou seja, a parte mais clara da imagem, constituída pelos pixels que formam o fundo.

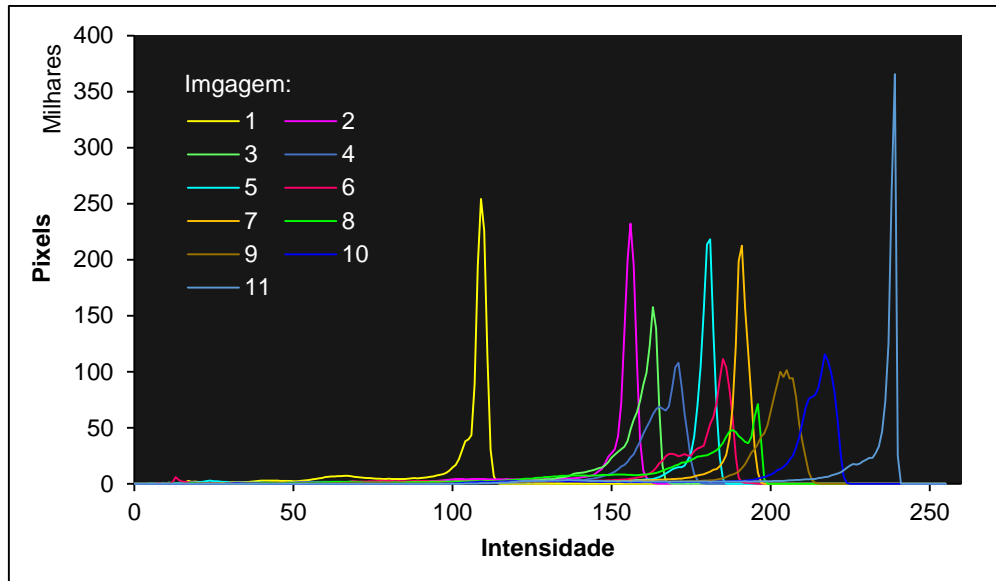


Figura 8 – Histogramas das imagens de referência
Fonte: Autoria própria

Como não foi possível identificar nenhum vale de forma clara e com comportamento similar nos histogramas das diferentes imagens analisadas, foi realizada uma análise mais aproximada do gráfico, deixando de fora os valores de intensidades mais altas, conforme a Figura 9. Nesta segunda análise foi possível perceber um comportamento comum nos histogramas: um pequeno agrupamento de pixels mais escuros, levando a acreditar que esta região representa os cromossomos.

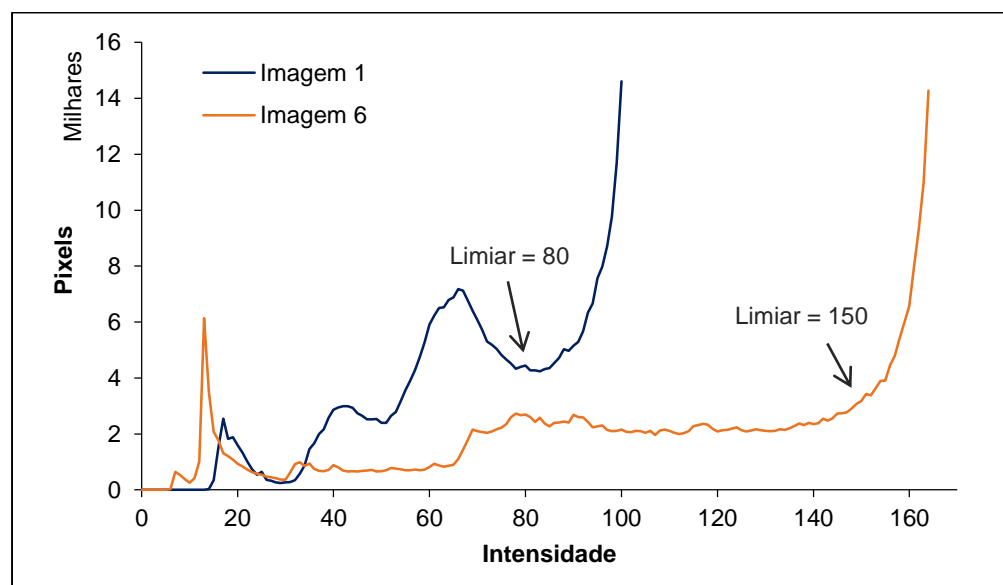


Figura 9 – Histogramas aproximados das imagens de referência 1 e 2
Fonte: Autoria própria

Sendo assim optou-se por testar a binarização utilizando como limiar valores entre essas duas regiões. Os resultados foram melhores, utilizando-se valores em que a concentração de *pixels* nas intensidades mais altas começa a aumentar, ou logo antes deste aumento repentino (Figura 9). Este valor foi determinado com o auxílio de um *trackbar* para variar o limiar e acompanhar o efeito dessa mudança na eliminação do fundo. Esta operação permitiu eliminar praticamente todo o fundo das 11 imagens analisadas. Porém restaram sombras em torno de algumas ROIs, elementos emendados e fragmentos do fundo.

O problema da sombra, oriundo principalmente da falta de foco das imagens, foi amenizado utilizando uma operação de borramento na imagem original por meio de um filtro passa baixa com uma máscara de dimensão 9x9, executado uma vez. Por reduzir os ruídos da imagem, esta operação tornou os fundos das imagens mais homogêneos, o que melhorou o resultado da binarização, por reduzir a quantidade e o tamanho dos fragmentos de fundo. O efeito colateral deste processo é a união de alguns cromossomos muito próximos e o apagamento dos mais claros.

Alguns dos pequenos elementos presentes na imagem foram eliminados após a binarização, por meio de sucessivas dilatações e erosões (abertura), com três repetições cada, utilizando uma máscara 3x3. Esta ação, colaborou ainda para a separação de alguns elementos emendados. Todas as operações citadas até esta etapa podem ser melhor visualizadas nas imagens da Figura 10.

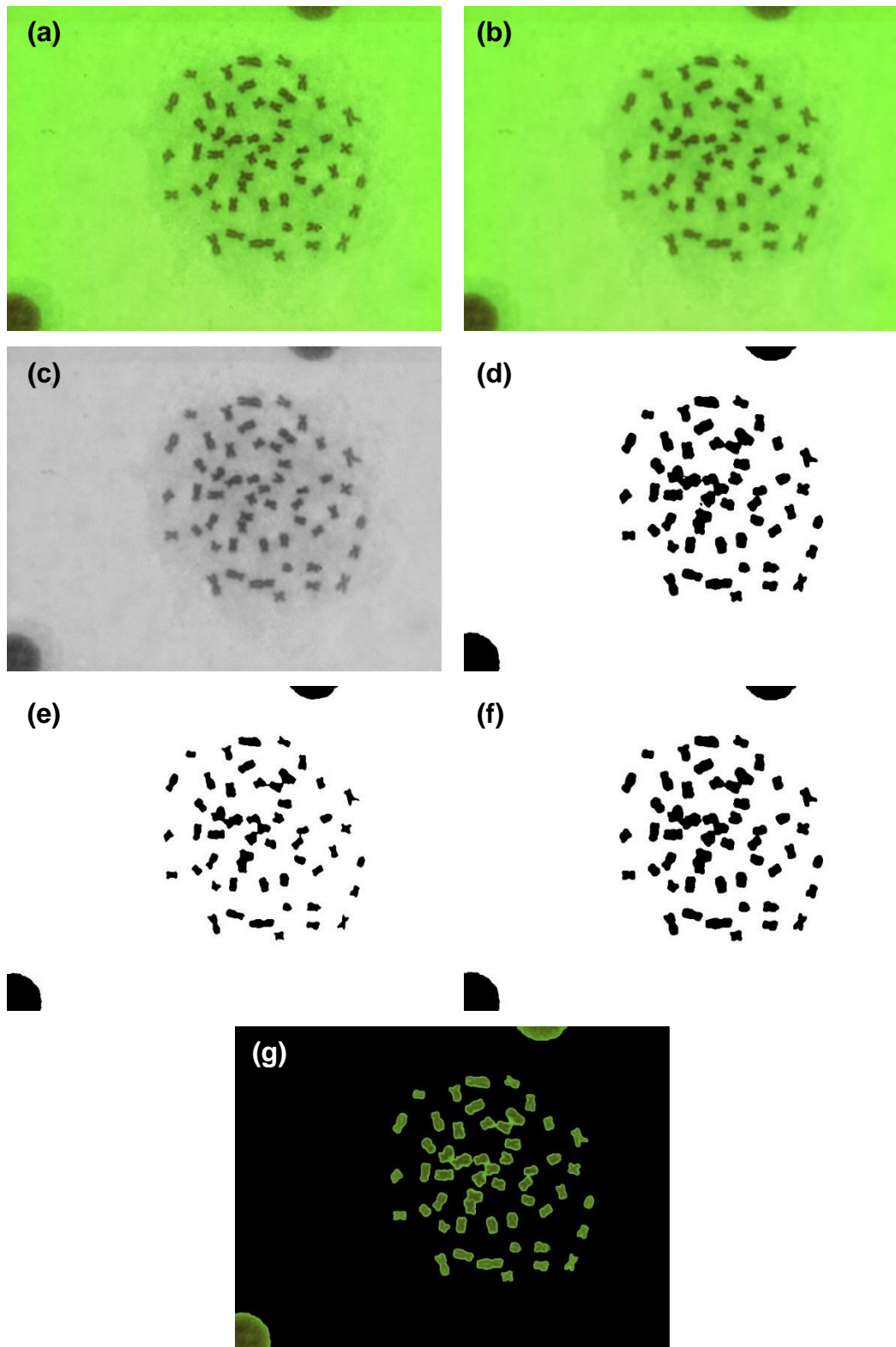


Figura 10 – Sequência de operações do processo de eliminação do fundo: (a) original; (b) blur (3x3); (c) cinza; (d) limiarização; (e) 3 x erosão (3x3); (f) 3 x dilatação (3x3); (g) subtração
 Fonte: Autoria própria

Os resultados nesta etapa foram satisfatórios, porém com valores de limiar diferente para cada uma das 11 imagens (Tabela 1), encontrados por tentativa e erro. Para executar a binarização da imagem de forma automática em todas as imagens, buscou-se criar uma relação entre a intensidade do fundo (*ifundo*) e o valor do limiar ideal, para que este pudesse assumir um valor variável em função da intensidade do fundo.

Tabela 1 – Imagens referência organizadas com base na intensidade do fundo

| Posição | Máximo (fundo) | Limiar Manual | Limiar Calculado |
|---------|----------------|---------------|------------------|
| 1 | 109 | 80 | 76,49 |
| 10 | 156 | 120 | 118,27 |
| 20 | 163 | 120 | 124,49 |
| 30 | 171 | 125 | 131,60 |
| 40 | 181 | 130 | 140,49 |
| 50 | 185 | 150 | 144,05 |
| 60 | 191 | 150 | 149,38 |
| 70 | 196 | 150 | 153,83 |
| 80 | 205 | 175 | 161,83 |
| 90 | 217 | 180 | 172,50 |
| 100 | 239 | 185 | 192,06 |

Fonte: Autoria própria

Os dados da Tabela 1 foram utilizados para realizar uma regressão linear, a fim de se obter uma equação para determinar o valor do limiar em função do pico máximo dos histogramas das imagens (Equação 3).

$$T = 0,889 \times ifundo - 20,415 \quad (3)$$

Com a Equação 3 foi possível obter uma linha de tendência, disposta no gráfico da Figura 11, juntamente com os dados das 11 imagens de referência. A análise visual do gráfico, sugere que os resultados, representados pela linha de tendência, foram satisfatórios, fato reforçado pelo valor de R^2 que demonstrou que 94,71% da variável dependente (limiar) pode ser expressa pela equação da reta.

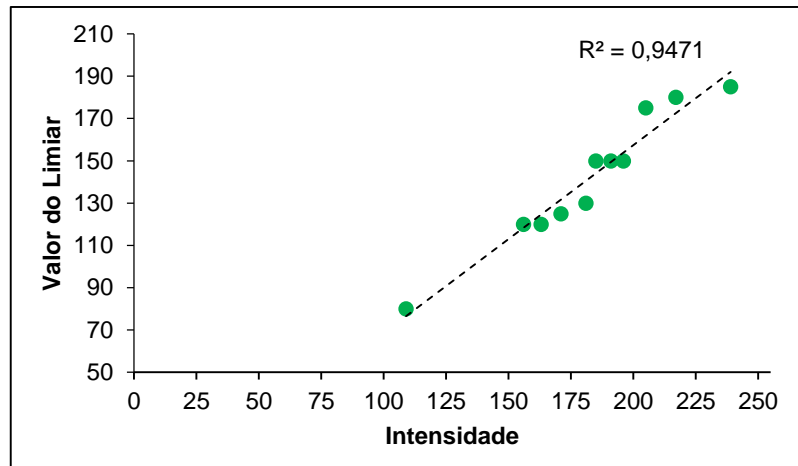


Figura 11 – Gráfico do valor do limiar em relação a intensidade do fundo

Fonte: Autoria própria

4.2.2 Primeira Segmentação

O objetivo desta etapa foi segmentar as imagens originais, gerando outras imagens menores com as ROIs isoladas na etapa anterior. Para isso foi utilizada uma ferramenta de detecção de borda, denominada *findContours*, baseada no método proposto por Suzuki e Abe (1985), que permitiu determinar um retângulo englobante em cada ROI, possibilitando gerar novas imagens “recortadas” das imagens originais. Contudo, esta etapa apresentava uma falha, conforme observa-se na Figura 12, pedaços de outras ROIs eram segmentadas juntamente.

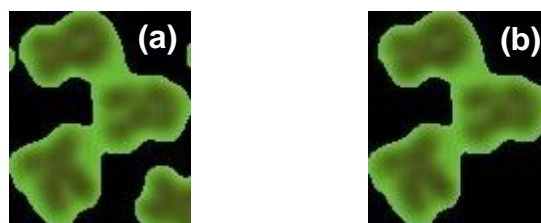


Figura 12 – ROI segmentada (a) com ruído e (b) limpa

Fonte: Autoria própria

A metodologia utilizada para resolver este problema, foi realizar o desenho e preenchimento dos contornos com cores aleatórias (Figura 13), o que possibilitou isolar cada ROI com base na sua respectiva cor, e obter o resultado da Figura 12 (b).

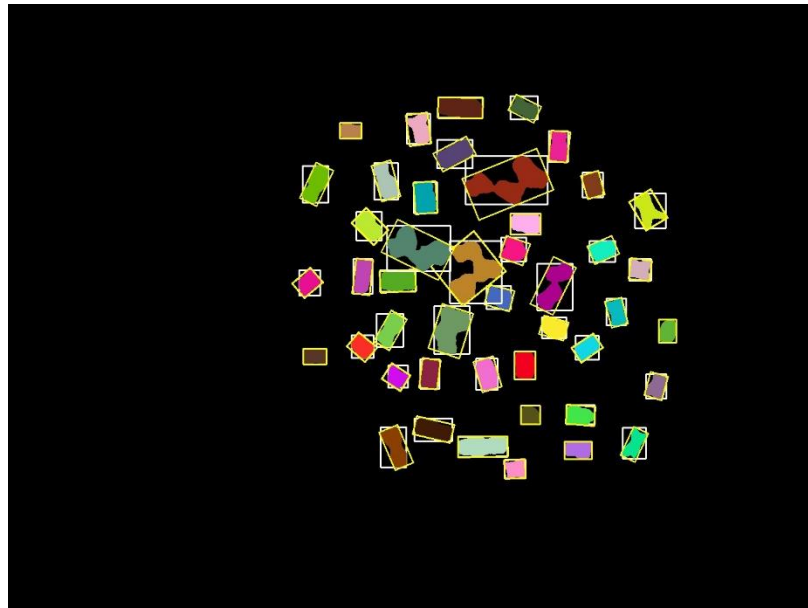


Figura 13 – ROIs com contornos preenchidos e retângulos englobantes
Fonte: Autoria própria

As saídas desta etapa foram segmentos das imagens originais com uma borda do fundo original entorno dos elementos, inseridos em um novo fundo branco (Figura 14). Optou-se pelo fundo branco para facilitar a binarização da imagem utilizada na etapa de medição dos braços dos cromossomos, pois desta forma os histogramas apresentaram três agrupamentos dominantes, sendo que as intensidades mais escuras se tornaram exclusividade das ROIs.

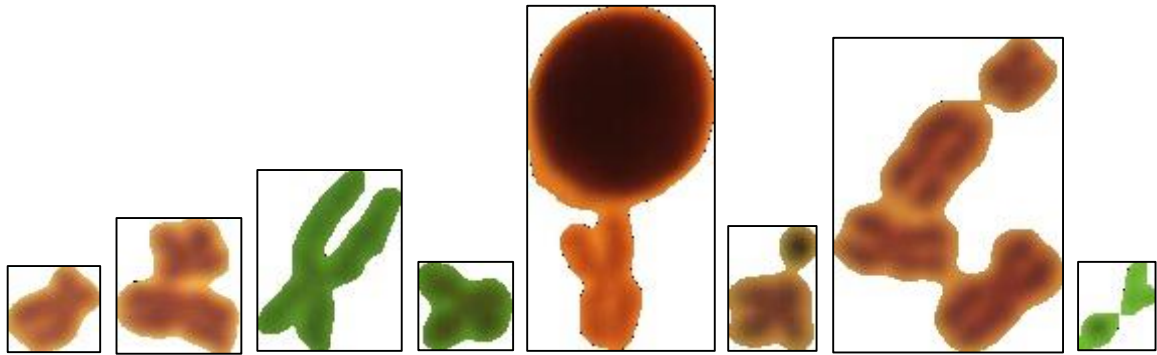


Figura 14 – ROIs recortadas da imagem original em um fundo branco
Fonte: Autoria própria

Mesmo após estas etapas, continuaram existindo alguns cromossomos unidos e muitos elementos que não são cromossomos, tais como: círculos, borrões, fragmentos e partes mais escuras do fundo. Portanto, neste ponto foi realizada uma limpeza, onde foram eliminados todos os fragmentos com dimensões inferiores a 30 *pixels*, por ser o tamanho da menor dimensão da menor imagem de cromossomo das 11 analisadas. Esta operação teve ainda a intenção de tornar o sistema mais leve, evitando processamentos desnecessários. Objetos maiores foram eliminados em etapas futuras, haja vista que neste ponto ainda haviam muitos cromossomos unidos.

4.2.3 Segunda Segmentação

Para separar os cromossomos unidos utilizou-se a ferramenta *k-means* nas imagens resultantes da primeira segmentação. O objetivo deste processo foi alterar a cor dos pixels para uma das três classes ($k=3$): o cromossomo, pixels mais escuros; a borda, formada por restos do *background* original, e que manteve vários cromossomos ligados; e o novo fundo branco (Figura 15 a). Esta operação possibilitou eliminar as duas intensidades mais claras, restando apenas os cromossomos isolados, que foram segmentados com o mesmo algoritmo utilizado na primeira segmentação, em que as ROIs eram preenchidas com cores aleatórias e depois selecionadas. Com o intuito de manter o mesmo padrão nas imagens fragmentadas, este processo foi realizado em todas as ROIs, mesmo àquelas com um único cromossomo já isolado.

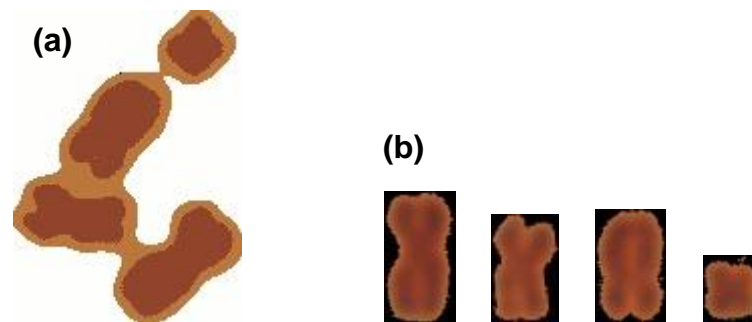


Figura 15 – Grupo de cromossomos unidos (a) com os pixels agrupados em três cores (*k-means*), (b) cromossomos da imagem original individualizados
 Fonte: Autoria própria

Na segunda segmentação, os fragmentos foram rotacionados para a posição vertical (Figura 15 b), e para isso foi utilizado o ângulo do menor retângulo englobante possível, representados em amarelo na Figura 13. Nesta etapa foi realizada ainda uma segunda limpeza, desta vez removendo imagens com ao menos uma dimensão menor que 20 e maior que 131 *pixels*, tamanho da menor e maior dimensão dos cromossomos das 11 imagens de referência. Por fim, foram salvas cópias das imagens binarizadas invertidas, com os cromossomos brancos em um fundo preto, utilizadas na classificação.

4.3 CLASSIFICAÇÃO

Para classificar os cromossomos é necessário calcular a razão entre o tamanho dos seus braços, que somente podem ser medidos identificando-se a posição do centrômero. Este ponto é facilmente visualizado nas imagens, por formar uma estricção no cromossomo, ou seja, uma região mais estreita na ROI (Figura 16). Para localizar esta região, realizou-se uma varredura ao longo das linhas da imagem verificando a quantidade de *pixels* com intensidade 255 (brancos), sendo considerado como centrômero a coluna com menor quantidade de *pixels* brancos (Figura 16).



Figura 16 – Cromossomo individualizado com binarização invertida
Fonte: Autoria própria

Considerando que o formato do cromossomo se assemelha a retas transversais que se cruzam formando um “X”, e valendo-se do teorema de Tales, em que a intersecção entre retas paralelas e transversais formam segmentos proporcionais, foi possível determinar o comprimento dos braços contando as linhas do centrômero até as extremidades da imagem, obtendo-se valores de comprimento em *pixels* (Figura 17). Sendo assim é possível afirmar que o braço maior (diagonal) está para o braço menor (diagonal), assim como, o braço maior (vertical) está para o braço menor (vertical).

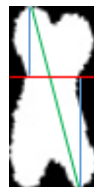


Figura 17 – Linhas correspondentes ao centrômero e comprimento dos braços
Fonte: Autoria própria

As medidas encontradas não representam as medidas reais dos braços, contudo, podem ser consideradas por se tratar de uma relação, o que possibilitou calcular a razão entre os braços. Consequente a isso, os cromossomos foram classificados em metacêntricos (RB de 1,00 a 1,70), submetacêntricos (RB de 1,71 a 3,00), subtelocêntricos (RB de 3,10 a 7,00) e acrocêntricos (RB > 7,00) (LEVAN, 1964).

5 RESULTADOS E DISCUSSÃO

Em termos de aplicabilidade, a eficiência do sistema deve levar em consideração a perda dos cromossomos, uma vez que este fato impossibilita a utilização da imagem da metáfase, para obtenção do cariótipo. Contudo, esta abordagem condena todos os outros cromossomos da imagem, que estavam conformes, tornando a eficiência geral do sistema mais baixa. Sendo assim, do ponto de vista computacional, é conveniente realizar outra abordagem, desconsiderando, para o cálculo da eficiência, apenas os cromossomos realmente afetados: que se perderam, ou que continuaram unidos.

5.1 SEGMENTAÇÃO

No que diz respeito à primeira etapa do processamento das imagens, a eliminação do fundo, pode-se dizer que o método proposto, baseado na variação do limiar de acordo com a intensidade do fundo (Equação 3), foi satisfatório. Haja vista que todas as 100 imagens analisadas tiveram o fundo totalmente ou quase totalmente eliminado (Figura 18).

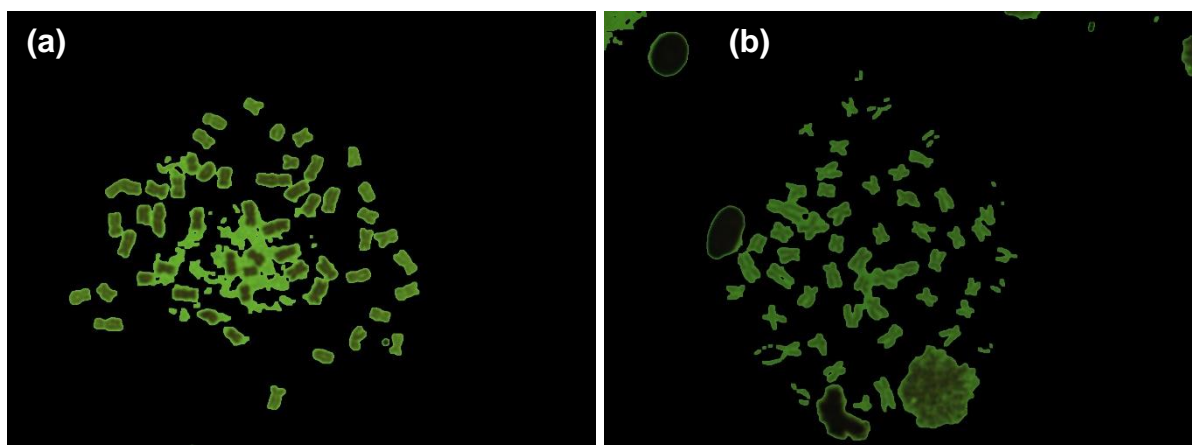


Figura 18 – Imagens após a eliminação do fundo com (a) limiar alto e (b) limiar baixo
Fonte: Autoria própria

Contudo, cinco imagens perderam cromossomos, total ou parcialmente, durante este processo. Isso ocorre devido ao fato de alguns cromossomos possuírem pixels com valores de intensidade iguais, ou muito próximos, aos valores do fundo. Vale destacar que para obter-se um cariótipo, todos os cromossomos da imagem da metáfase devem ser preservados. Mesmo assim, as imagens que perderam cromossomos continuaram sendo processadas, a fim de avaliar a eficiência nas demais etapas do sistema.

Quanto mais distante foi o valor do limiar calculado em relação ao valor do limiar ótimo, positivamente, maior a quantidade de cromossomos unidos, pois mais pixels com cores claras (como o fundo) foram selecionados (Figura 18 a). Quanto mais distante foi o valor de maneira negativa, mais cromossomos foram perdidos no processo, pois somente os pixels com intensidades mais baixas (escuras) foram selecionados, conforme Figura 18 (b).

Mesmo as imagens que apresentaram grandes desvios nos valores do limiar, o sistema apresentou bons resultados na eliminação do fundo, como pode ser observado nas imagens da Figura 19, em que os cromossomos foram preservados e a maior parte do fundo eliminada.

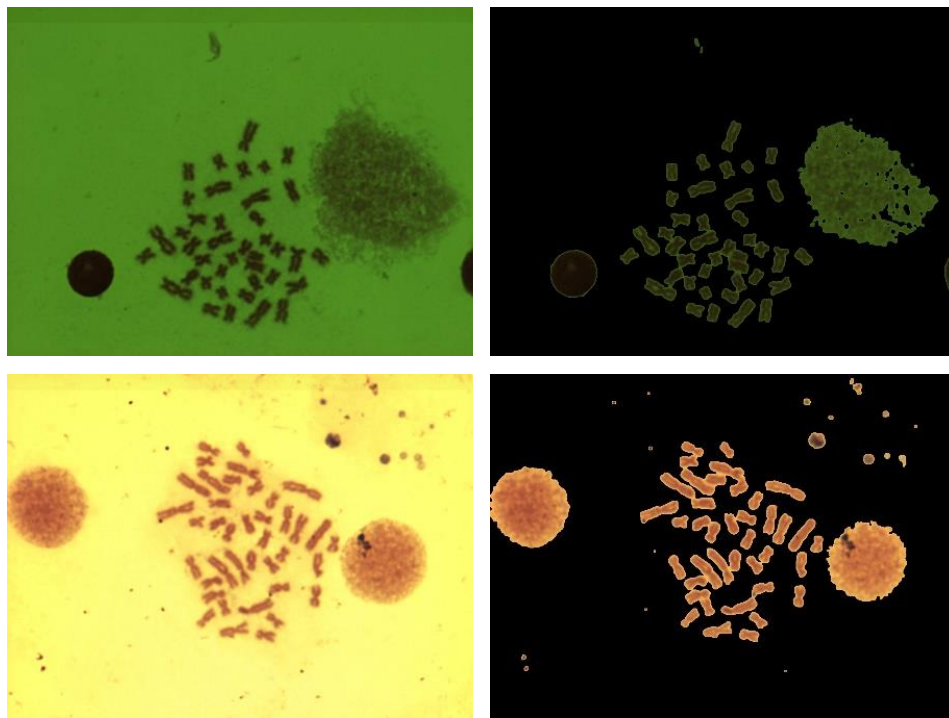


Figura 19 – Imagens originais e sem o fundo
Fonte: Autoria própria

Na sequência as imagens originais foram submetidas ao primeiro processo de segmentação das ROIs, que atenderam perfeitamente ao objetivo. Em 100% das imagens, tudo aquilo que não era fundo foi salvo em uma nova imagem chamada de fragmento, além disso, todos os cromossomos que entraram nesta etapa foram mantidos. Contudo, praticamente todas as imagens, 95 das 100 analisadas, apresentavam ao menos 1 grupo de cromossomos unidos entre si, ou unidos com ruídos (tudo o que não for cromossomo). Os grupos variaram de um único cromossomo unido a um elemento estranho, à grupos envolvendo 90% dos cromossomos da imagem (Figura 20).

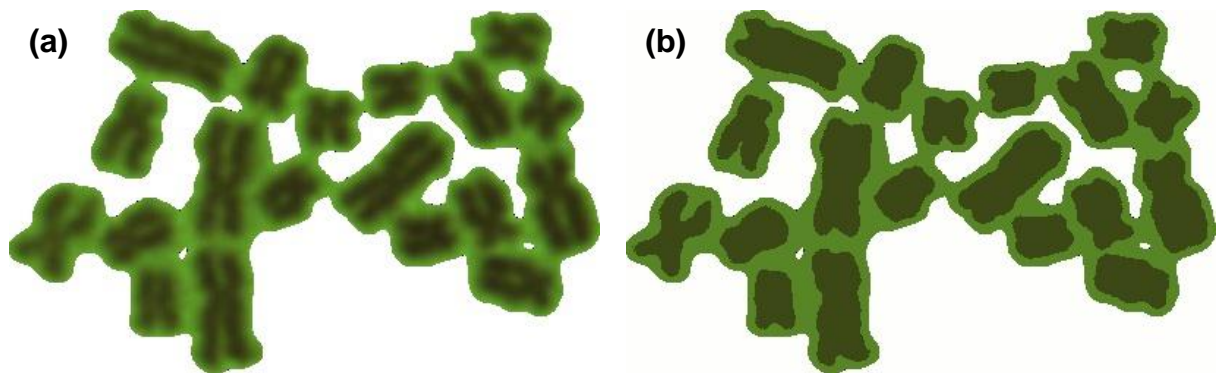


Figura 20 – Grupo de cromossomos unidos (a) recortados da imagem original e (b) com os pixels agrupados em três cores (*k-means*)
Fonte: Autoria própria

Como 95% das imagens apresentavam cromossomos unidos a algo, realizou-se então a segunda segmentação, desta vez, baseada na eliminação das regiões que ligavam os cromossomos. Para isso, primeiramente, as imagens segmentadas passaram por um processo de agrupamento das intensidades em 3 grupos, normalmente: cromossomos (intensidade baixa), região de união (intensidade intermediária) e fundo branco (intensidade alta).

A ferramenta *k-means* funcionou em todos os fragmentos, não somente para segmentar os cromossomos unidos, mas como também para melhorar a limpeza em torno dos cromossomos já individualizados (Figura 21 a). Porém em alguns, os pixels se agruparam em apenas duas intensidades (Figura 21 b), ou ainda classificando partes do cromossomo como região de união (Figura 21 c), o que impediu a separação de alguns cromossomos e ocasionou a eliminação de outros,

respectivamente.

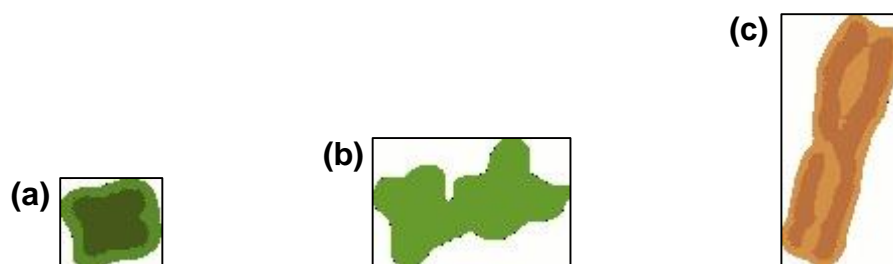


Figura 21 – Cromossomos com os pixels agrupados por cores (*k-means*), (a) c com a borda em destaque, (b) unidos com apenas duas cores e (c) comprometido

Fonte: Autoria própria

Nesta etapa, das 95 imagens que apresentavam cromossomos unidos, em 56 eles foram totalmente separados, e com as outras 5 imagens em que os cromossomos já haviam sido separados na primeira etapa, totalizou 61 imagens com cromossomos totalmente separados. Destas 61 imagens, 39 (64%) não perderam nenhum cromossomo, sendo que destas, uma já havia sido comprometida na eliminação do fundo, restando então 38 imagens possíveis de se obter um cariótipo completo. Algumas perdas de cromossomos nesta etapa, muito pequenos ou muito grandes, devem-se ao processo de limpeza.

Sabendo que a quantidade de cromossomos deve ser a mesma nos cariótipos de diferentes metáfases, procurou-se descobrir a quantidade de cromossomos nesta fase do processo, por ser a última segmentação. No entanto, 32 das 38 imagens apresentavam elementos segmentados que não eram cromossomos, impedindo a quantificação exata. Das seis imagens possíveis, três delas apresentaram 40 cromossomos, o primeiro indicativo de que poderia ser esta a quantidade a ser considerada.

Sendo assim, realizou-se uma contagem manual dos cromossomos das 38 imagens possíveis, e a quantidade de cromossomos com maior incidência foi de 16 imagens com 39 cromossomos cada. Por se tratar de um número ímpar, impossível de ocorrer pelo fato dos cromossomos existirem aos pares, passou-se então a considerar a segunda maior incidência, que foi de 11 imagens com 40 cromossomos cada, reforçando a suspeita do primeiro indicativo. O mesmo se repete quando analisado a contagem dos cromossomos nas imagens originais, a maior ocorrência também é para a quantidade de 39, e em seguida 40, conforme a Tabela 2. As

imagens possuem diferentes quantidade de cromossomos, pois muitos ficam escondidos atrás de outros elementos da imagem, ou até mesmo vão parar na imagem de outra metáfase.

Tabela 2 – Imagens por quantidade de cromossomos

| Quantidade de cromossomos | Ocorrência nas imagens | | |
|---------------------------|------------------------|-------------|----------|
| | Originais | Segmentadas | Limpas |
| 15 | 1 | 0 | 0 |
| 24 | 1 | 0 | 0 |
| 26 | 1 | 0 | 0 |
| 31 | 1 | 0 | 0 |
| 32 | 2 | 0 | 0 |
| 33 | 2 | 0 | 0 |
| 34 | 3 | 0 | 0 |
| 35 | 1 | 0 | 0 |
| 36 | 7 | 0 | 0 |
| 37 | 8 | 2 | 0 |
| 38 | 10 | 0 | 0 |
| 39 | 27 | 16 | 2 |
| 40 | 18 | 11 | 3 |
| 44 | 1 | 0 | 0 |
| 49 | 2 | 1 | 1 |
| 50 | 4 | 2 | 0 |
| 51 | 5 | 1 | 0 |
| 52 | 6 | 5 | 0 |
| TOTAL | 100 | 38 | 6 |

Fonte: Autoria própria

No total foram contabilizados 4182 cromossomos nas 100 imagens, sendo preservados 3973 (95,00%) ao longo das etapas de segmentação, 35 foram perdidos na eliminação do fundo e 174 na segunda segmentação, totalizando 209 cromossomos (5,00%), em 61 imagens diferentes. Nas 59 imagens que mantiveram os cromossomos, 21 apresentaram cromossomos que não se separaram, restando as 38 imagens possíveis de se obter um cariótipo completo.

Entraram no sistema 439 agrupamentos com 1447 cromossomos, destes, 366 (83,37%) agrupamentos foram desfeitos ao término de todas etapas de segmentação, separando 1290 cromossomos, ou seja, 89,15% dos cromossomos foram segmentados adequadamente. A última operação desta etapa foi a limpeza dos

fragmentos, que eliminou 45,31% dos ruídos. As imagens da Figura 22, são resultadas do segundo processo de segmentação, realizado no grupo de cromossomos da Figura 10.



Figura 22 –Cromossomos segmentados do grupo da Figura 10
Fonte: Autoria própria

A ineficiência desta segunda segmentação, pode ser atribuída ao fato de que alguns cromossomos estarem encostados, ou sobrepostos, como é o caso daqueles apresentados na Figura 23. Isto é agravado ao passo que o nível de desfoque da imagem é maior, pois torna os limites das ROIs mais borradas, sobrepondo uma à outra.

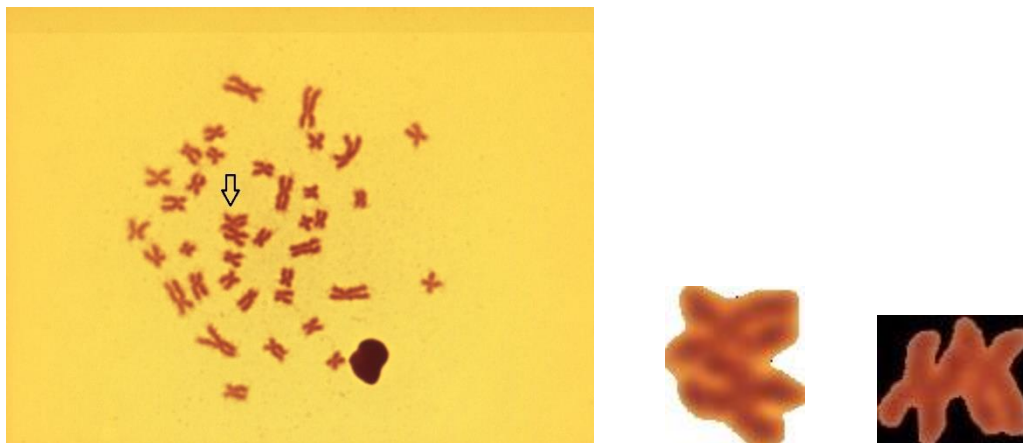


Figura 23 – Imagem original desfocada e cromossomos unidos
Fonte: Autoria própria

5.2 CLASSIFICAÇÃO

Nesta etapa, foram avaliados os resultados obtidos em relação às imagens

que entraram no algoritmo classificador, e não em relação às imagens originais das metáfases. Portanto, mesmo que algum cromossomo da imagem tenha sido perdido ao longo das segmentações, nesta etapa os que restaram foram processados. Apesar de não ser mais possível obter um cariótipo destas metáfases desfalcadas, esta análise permitiu uma avaliação mais fidedigna da eficiência do processo de classificação em si.

A primeira operação desta etapa, que foi a varredura pela imagem binarizada do cromossomo individualizado, em busca da linha com menos pixels brancos, funcionou adequadamente para todas as imagens analisadas. Consequentemente, a medição das distâncias do ponto encontrado até as extremidades da imagem, foi realizada com sucesso.

Para avaliar a eficiência desta etapa, foram analisados 253 cromossomos individualizados, sendo que 161 foram classificados de maneira correta, conforme a matriz de confusão do Quadro 1. Sendo assim, a taxa de acerto do classificador foi de 63,64%. Pode-se observar na matriz que todos os cromossomos subtelocêntricos foram classificados corretamente, em relação a classificação manual, contudo as metáfases dos indivíduos analisados deveriam apresentar somente cromossomos metacêntricos e submetacêntricos. Isso reforça o quanto a classificação manual pode ser subjetiva.

| | Meta | Submeta. | Subtelo. | Acro. | Ruídos |
|------------------------|-------------|-----------------|-----------------|--------------|---------------|
| Metacêntrico | 112 | 32 | 5 | 1 | 2 |
| Submetacêntrico | 20 | 45 | 4 | 0 | 1 |
| Subtelocêntrico | 0 | 0 | 4 | 0 | 0 |
| Acrocêntrico | 0 | 0 | 0 | 0 | 0 |
| Ruídos | 14 | 12 | 3 | 1 | 0 |

Quadro 1 – Matriz de confusão dos cromossomos classificados

Fonte: Autoria própria

Considerando então as classificações possíveis, a maior taxa de acerto recai sobre os cromossomos metacêntricos e em seguida os submetacêntricos, com 73,68% e 64,28% de acerto, respectivamente. Por fim, nenhum ruído foi classificado como tal.

6 CONCLUSÃO

A etapa de aquisição e organização das imagens foi o primeiro passo para a criação de um banco de imagens de metáfases de peixes e seus respectivos cariótipos. Além da importância para a biologia, do ponto de vista computacional, este banco de imagens e cariótipos pode ser útil para futuras utilizações e aperfeiçoamentos do sistema proposto, como por exemplo, no treinamento de sistemas de inteligência artificial.

No que diz respeito ao desenvolvimento do sistema, as técnicas e ferramentas disponibilizadas pela biblioteca OpenCV foram suficientes para realização do processamento das imagens, além de facilitar a programação por trazer diversos algoritmos de visão computacional já implementados.

A segmentação das imagens foi a etapa mais extensa e onerosa computacionalmente, tendo em vista a quantidade de processos envolvidos e imagens geradas. Esta etapa foi fator limitante para o sucesso da classificação dos cromossomos, uma vez que para realizar a medição dos braços e encontrar o centrômero, o cromossomo precisa estar individualizado em uma imagem fragmentada da imagem original.

Apesar de muitas imagens terem sido descartadas por perderem cromossomos, considerando que entraram na etapa de segmentação 4182 cromossomos e saíram 3816 devidamente individualizados e preservados, o algoritmo mostrou-se eficiente com uma taxa de segmentação de 91,25%. O processo de eliminação do fundo, baseado na regressão linear, foi o grande trunfo para a automação do sistema, que apresentou 95,00% de eficiência.

O algoritmo de classificação obteve uma eficiência considerada moderada (63,64%), haja vista que o sistema foi alimentado com muitos elementos que não eram cromossomos e ele não foi capaz de identificar estas sujeiras. Tal fato, fez com que a eficiência global do *software* caísse para aproximadamente 58,07% (91,25% de 63,64%), porém, de maneira geral, é possível afirmar que os resultados foram satisfatórios, tendo em vista a subjetividade das medições manuais.

7 SUGESTÕES PARA TRABALHOS FUTUROS

A fim de complementar esta pesquisa, sugere-se para trabalhos futuros:

- Obtenção de mais cromossomos classificados manualmente para nova comparação;
- Desenvolver e implementar algoritmos para localizar os pares homólogos e montagem da imagem final do cariótipo;
- Testar o sistema em outras espécies de peixes;
- Testar a etapa de segmentação em imagens com diferentes padrões, como formato, tamanho, resolução e intensidade;
- Implementar um algoritmo de inteligência artificial para reconhecimento dos cromossomos e ruídos;
- Melhorar os resultados das limpezas, eliminando elementos, com base em outros atributos, além das dimensões, como por exemplo, perímetro, área e demais informações morfológicas da imagem;
- Introduzir interação humana para selecionar cromossomos não selecionados automaticamente pelo sistema;
- Analisar contornos para melhorar a taxa de acerto da classificação de cromossomos curvados.

REFERÊNCIAS

ALBERT, J. S.; REIS, R. E. **Historical biogeography of neotropical freshwater fishes**. California: University of California Press, 2011.

ALONSO-MONTESINOS, J.; BEATLLES, F. J. The use of a sky camera for solar radiation estimation based on digital image processing. **Energy**. p. 10, jun. 2015.

ALVES, ANDERSON. L. Karyotypic diversity and evolutionary trends in the Neotropical catfish genus *Hypostomus* Lacépède, 1803 (Teleostei, Siluriformes, Loricariidae). **Comparative Cytogenetics**. v. 6, n. 4, p. 443-452, dez. 2012.

BLANCO, Daniel R. Origin of the X1X1X2X2/X1X2Y sex chromosome system of *Harttia punctata* (Siluriformes, Loricariidae) inferred from chromosome painting and FISH with ribosomal DNA markers. **Genetica**. v. 142, p. 119-126, 2014.

BLANCO, Daniel R. **Estudos citogenéticos clássicos e moleculares em espécies do gênero *Harttia* (Siluriformes, Loricariidae), com enfoque no papel dos DNAs repetitivos na evolução cariotípica do grupo**. São Carlos, 2012. 167 p. Tese de doutorado – Doutorado em Ciências da Universidade Federal de São Carlos – UFSCar, 2012.

BLUM, H. A Transformation for Extracting New Descriptors of Shape. **MIT Press**. Cambridge, p. 362-380, 1967

BONETTO, A.A. (1986). The Paraná River Systems.11. In: Davies B.R. E Walker, K.F. (eds). **The Ecology of River Systems**. The Netherlands: Dr. W. Junk Publishers, 541-555.

BRESSANE-NETO, A. F.; PORTO-FORESTI, F.; BORTOLOZZI, J.; FORESTI, F. Desenvolvimento de um Software Para Cariotipagem de Cromossomos de Peixes. In CONGRESSO BRASILEIRO DE GENÉTICA, 51. 2005, Águas de Lindóia. **Anais do 51º Congresso Brasileiro de Genética**. Águas de Lindóia: SBG, 2005 p. 189.

CAO, Hongbao; DENG, Hong-Wen; WANG, Yu-Ping. Segmentation of M-FISH Images for Improved Classification of Chromosomes With an Adaptive Fuzzy C-means Clustering Algorithm. **IEEE Transactions on Fuzzy Systems**, v. 20, n. 1, p. 1-8, fev. 2012.

CARDOSO, Andréa. **Fundamentos da Pesquisa Operacional**. Alfenas: UNIFAL-MG, 2011.

CARVALHO, Rafael A. Chromosomal analyses in *Megalonema platanum* (Siluriformes: Pimelodidae), an endangered species from South American rivers. **Neotropical Ichthyology**. v. 9, n. 1, p. 177-182, 2011.

CHO, Jongman; HONG, Seunghong. Classification of Giemsa-Stained Human Chromosomes Using Multi-Layer Neural Network. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 16. 1994, Baltimore. **Proceedings of the 16th Annual International Conference of the IEEE**. Baltimore: IEEE, 1994, p. 1115-1116.

CHO, Jong Man. Chromosome Classification Using Backpropagation Neural Networks. **IEEE Engineering in Medicine and Biology**, v. 19, n. 1, p. 28-33, jan/fev. 2000.

CONCI, Aura; AZEVEDO, Eduardo; LETA, Fabiana R. **Computação gráfica**. v. 2. Rio de Janeiro: Elsevier, 2008.

ESCHMEYER, W. N.; FONG, J. D. Species by family/subfamily. In: ESCHMEYER, W. N. **Catalog of fishes**. California: California Academy of Sciences. Online version. Disponível em: <<http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>>. Acesso em 22 mar. 2018.

FACON, J. **Morfologia matemática: teoria e exemplos**. Editora Universitária Champagnat da Pontifícia Universidade Católica do Paraná, 1996.

FARIA, Elaine R. **SiMCaPe – Sistema para a montagem de cariótipo de peixe baseado em conjuntos difusos**. 2006. 86 f. Dissertação (Mestrado em Ciência da Computação) – Faculdade de Computação, Universidade Federal de Uberlândia. Uberlândia, 2006.

FENG, Xinwei; CONG, Peisheng; ZHU, Zhongliang; DU, Xinyuan. Automated Pairing of Human Chromosomes Applying Gradient Profile and Similarity Matching Algorithm. **Chemometrics and Intelligent Laboratory Systems**, v. 111, n. 1, p. 46-52, fev. 2012.

FILHO, Ogê M.; NETO, Hugo V. **Processamento Digital de Imagens**, Rio de Janeiro: Brasport, 1999.

GAGULA-PALALIC, Sadina; CAN, Mehmet. Human Chromosome Classification Using Competitive Neural Network Teams (CNNT) and Nearest Neighbor. In: INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS IEEE-EMBS. 2014, Valencia. **Proceedins of International Conference on Biomedical and Health Informatics IEEE-EMBS**. Valencia: IEEE, 2014 p. 626-629.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4ª ed. São Paulo: Atlas, 2002.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento digital de imagens**. 3. Ed. São Paulo: Pearson, 2011.

GRIFFITHS, Anthony J. F. **Introdução à Genética**. 9.ed. Rio de Janeiro: Guanabara Koogan, 1988.

GUERRA, Marcelo S. **Introdução à Citogenética Geral**. Rio de Janeiro: Guanabara Koogan, 1988.

GUIMARÃES, L. V.; SCHUCK JR, A.; ELBERN, A. Chromosome Classification for Karyotype Composing Applying Shape Representation on Wavelet Packet Transform. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 25. 2003, Cancun. **Proceedins of the 25th Annual International Conference of the IEEE**. Cancun: IEEE, 2003 p. 941-943.

JACOBINA, Uedson P. Biogeography and comparative cytogenetics between two populations of *Hoplias malabaricus* (Bloch, 1794) (Ostariophysi: Erythrinidae) from coastal basins in the State of Bahia, Brazil. **Neotropical Ichthyology**. v. 7, n. 4, p. 617-622, 2009.

JAIN, A. K., Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, p. 651-666, 2010.

JANANI, S.NIRMALA, M.; NANDAKUMAR, R. Feature Extraction and Pairing of G-Band Chromosome Images Using K-Nearest Neighbour Classifier. **International Journal of Computer Science and Technology**, v. 3, n. 2, p. 137-140, abr/jun. 2012.

JAVAN-ROSHTKHARI, Mehrsan; SETAREHDAN, S. Kamaledin. A New Approach to Automatic Classification of the Curved Chromosomes. In: INTERNATIONAL SYMPOSIUM ON IMAGE AND SIGNAL PROCESSING AND ANALYSIS, 5. 2007,

Istanbul. **Proceedins of 5th International Symposium on Image and Signal Processing and Analysis**. Istanbul: IEEE, 2007 p. 19-24.

KARVELIS, Petros S.; FOTIADIS, Dimitrios I.; GEORGIU, Ioannis A.; SYRROU, Marika. A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 28. 2006, Nova York. **Proceedings of 28th Annual International Conference of the IEEE**. Nova York: IEEE, 2006 p. 3009-3012.

KARVELIS, Petros S.; TZALLAS, Alexandros T.; FOTIADIS, Dimitrios I.; GEORGIU, Ioannis. A Multichannel Watershed-Based Segmentation Method for Multispectral Chromosome Classification. **IEEE Transactions on Medical Imaging**, v. 27, n. 5, p. 697-708, mai. 2008.

KARVELIS, P. S.; FOTIADIS, D. I.; GEORGIU, I.; SAKALOGLOU, P. Enhancement of the classification of multichannel chromosome images using support vector machines. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE-EMBS, 31. 2009, Minneapolis. **Proceedings of the 31th Annual International Conference of the IEEE-EMBS**. Minneapolis: IEEE, 2009a p. 3601-3604.

KARVELIS, Petros S.; FOTIADIS, Dimitrios I.; TSALIKAKIS, Dimitrios G.; GEORGIU, Ioannis A. Enhancement of Multichannel Chromosome Classification Using a Region-Based Classifier and Vector Median Filtering. **IEEE Transactions on Information Technology in Biomedicine**, v. 13, n. 4, p. 561-570, jul. 2009b.

KLACZKO, Louis B.; VIEIRA, Roberto D. **Avaliação do estado do conhecimento da diversidade biológica do Brasil: Genética**. Campinas: MMA/UNICAMP. 2003.

KOU, Zhenzhen; JI, Liang; ZHANG, Xuegong. Karyotyping of Comparative Genomic Hybridization Human Metaphases by Using Support Vector Machines. **Cytometry**, v. 47, p. 17-23, 2002.

LEDLEY, Robert S. Hight-Speed Automatic Analysis of Biomedical Pictures. **Science**, v. 146, n. 3641, p. 216–223, out. 1964.

LERNER, B.; LEVINSTEIN, M.; ROSENBERG, B.; GUTERMAN, H.; DINSTEIN, I.; ROMEM, Y. Feature Selection and Chromosome Classification Using a Multilayer Perceptron Neural Network. In: WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE, 6. 1994, Orlando. **Proceedings of the 6th World Congress on Computational Intelligence**. Orlando: IEEE, 1994 p. 3540-3545.

LEVAN, Albert. Nomenclature for centromeric position on chromosomes. **Hereditas**. v. 52, p. 201-220, nov. 1964.

LUI, Roberto L. Comparative chromosomal analysis and evolutionary considerations concerning two species of genus *Tatia* (Siluriformes, Auchenipteridae). **Comparative Cytogenetics**. v. 7, n.2, p. 63-71, abr. 2013.

MACQUEEN, J. 1967. Some Methods for classification and Analysis of Multivariate Observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 1. **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**. California: University of California Press, p. 281-297

MARKOU, Christoforos; MARAMIS, Christos; DELOPOULOS, Anastasios. Automatic Chromosome Classification using Support Vector Machines. In: **Handbook of Pattern Recognition: Methods and Applications**, K. Hosny, and J. De La Calleja (ed.), iConcept Press, 2012.

MENEZES, N.A.; BUCKUP, P.A.; FIGUEIREDO, J.L.; MOURA, R.L. **Catálogo das espécies de peixes marinhos do Brasil**. São Paulo: Museu de Zoologia, Universidade de São Paulo, 2003. 160 p.

MORADI, Mehdi; SETAREHDAN, S. Kamaledin. New Features for Automatic Classification of Human Chromosomes: A Feasibility Study. **Pattern Recognition Letters**, v. 27, p. 19-28, 2006.

NELSON, Joseph S. **Fishes of the world**. 4. ed. New Jersey: John Wiley & Sons, 2006.

NICKOLLS, P.; PIPER, J.; RUTOVITZ, D.; CHISHOLM, A. Pre-Processing of Images In An Automated Chromosome Analysis System. **Pattern Recognition**, v. 14, n 1-6, p. 219-229, 1981.

OSKOEI, Baharak C; SHANBEHZADEH, Jamshid. Chromosome Classification Based on Wavelet Neural Network. In: INTERNATIONAL CONFERENCE ON DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS. 2010, Sydney. **Proceedings of the International Conference on Digital Image Computing: Techniques and Applications**. Sydney: IEEE, 2010 p. 605-610.

PEDRINI, Hélio; SCHARTZ, William R. **Análise de imagens digitais: princípios, algoritmos e aplicações**. São Paulo: Thomson Learning, 2008.

PIPER, Jim; GRANUM, Erik. On Fully Automatic Feature Measurement for Banded Chromosome Classification. **Cytometry**, v. 10, p. 242-255, out. 1988.

PLOTZE, R. O.; BRUNO, O. M. Análise de formas e reconhecimento de padrões por meio da assinatura fractal multi-escala. **Infocomp**, São Carlos, vol 6, pp 91-100, 2007.

PRODANOV, C. C; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2ª ed. Novo Hamburgo: Feevale, 2013.

QUEIROZ, José E. R.; GOMES, Herman M. Introdução ao processamento digital de imagens. **Rita**. v. 8, n. 1. 2001.

RAMAKUMAR, Adarsh. High-throughput sample processing and sample management; the functional evolution of classical cytogenetic assay towards automation. **Mutation Research**. p. 10, jul. 2015.

REIS, R. E.; KULLANDER, S. O.; FERRARIS, C. **Check List of freshwater Fishes of South and Central America**. Porto Alegre: Edipucrs, 2003.

ROSA, S. R.; LIMA, F. C. Os Peixes Brasileiros Ameaçados de Extinção. **Livro vermelho da fauna brasileira ameaçada de extinção**. 2008. v. 2, p. 9-18.
Disponível em: <<http://www.icmbio.gov.br/portal/images/stories/biodiversidade/fauna-brasileira/livrovermelho/volumell/Peixes.pdf>>. Acesso em 24 mar. 2018. 278 p.

SÁNCHEZ-ROMERO, Omar. First description of the karyotype and localization of major and minor ribosomal genes in *Rhoadsia altipinna* Fowler, 1911 (Characiformes, Characidae) from Ecuador. **Comparative Cytogenetics**.v. 9, n. 2, p. 271-280, jun. 2015.

SANTOS, G.M.; FERREIRA, E.J.G. (1999). **Peixes da bacia Amazônica**. R.H. Lowe-McConnell, Estudos ecológicos de comunidades de peixes tropicais. Edusp, São Paulo, 345-373.

SCHWARTZKOPF, Wade C.; BOVIK, Alan C.; EVANS, Brian L. Maximum-Likelihood Techniques for Joint Segmentation-Classification of Multispectral Chromosome Images. **IEEE Transactions on Medical Imaging**, v. 24, n. 12, p. 1593-1610, dez.

2005.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 3ª ed. Florianópolis: UFSC, 2001.

SUZUKI, S.; ABE, Keiichi. Topological Structural Analysis of Digitized Binary Images by Border Following. **Computer Vision, Graphics, and Image Processing**. v. 30, n. 1, p. 32-46. abr. 1985.

SKORTON, D. J. Digital signal and image processing in echocardiography. **Progress in Cardiology**. v. 110, n. 6, p. 1266-1283.

TAHERI-GARAVAND, Amin. An intelligent approach for cooling radiator fault diagnosis based on infrared thermal image processing technique. **Applied Thermal Engineering**. v. 87, p. 434-443, jun. 2015.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 2 ed. Londres: Pearson, 2018.

TING, W. S. Sistemas de informações gráficas, Campinas - SP, 2009, unpublished.

TRALDI, Josiane B. First karyotype description of *Hypostomus iheringii* (Regan, 1908): a case of heterochromatic polymorphism. **Comparative Cytogenetics**. v. 6, n. 2, p.115-125, mar. 2012.

UTTAMATANIN, Ravi; YUVAPOOSITANON, P.; INTARAPANICH, A.; KAEWKAMNERD, S.; TONGSIMA, S. Chromosome Classification for Metaphase Selection. In: INTERNATIONAL SYMPOSIUM ON COMMUNICATIONS AND INFORMATION TECHNOLOGIES (ISCIT), 13. 2013, Surat Thani. **Proceedings of the 13th International Symposium on Communications and Information Technologies (ISCIT)**. Surat Thani: IEEE, 2013 p. 464-468.

VARI, R.P.; MALABARBA, L.R. (1998). Neotropical ichthyology: An Overview. **Phylogeny in the classification of neotropical fishes** (L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M.S. Lucena, C.A.S. Lucena, Eds.). Edipucrs, Porto Alegre, 1-11.

VLIET, Lucas J. Van; YOUNG, I. T.; KATE, T. K.; MAYALL, B. H.; GROEN, F. C. A.; ROOS, R. Athena: A Macintosh-Based Interactive Karyotyping System. **Automation of Cytogenetics**, p. 47-66, 1989.

WANG, Xingwei; LI, Shibo; LIU, Hong; WOOD, Marc; CHEN, Wei R.; ZHENG, Bing. Automated Identification of Analysable Metaphase Chromosomes Depcted on Microscopic Digital Images. **Journal of Biomedical Informatics**, v. 41, n. 2, p. 264-271, abr. 2008.

WU, Q.; SUETENS, P.; OOSTERLINCK, A. Chromosome Classification Using a Multi-Layer Perceptron Neural Net. **Annual International Conference of the IEEE Engineering in Medicine and Biology Society**, v. 12, n. 3, 1990.

WU, Qiang; CASTLEMAN, Kenneth R. Automated Chromosome Classification Using Wavelet-Based Band Pattern Descriptors. In: SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS, 13. 2000, Houston. **Proceedings of the 13th Symposium on Computer-Based Medical Systems**. Houston: IEEE, 2000 p. 189-194.

WU, Qiang; LIU, Z.; CHEN, T.; CASTLEMAN, K. R. Subspace-Based Prototyping and Classification of Chromossome Images. **IEEE Transactions on Image Processing**, v. 14, n. 9, p. 1277-1287, set. 2005.

YAN, Wenzhong; BAI, Lei. Algorithms for Chromosome Classification. **Engineering**, v. 5, p. 400-403, 2013.

YANG, Yi. Online quality optimization of the injection molding process via digital image processing and model-free optimization. **Journal of Materials Processing Technology**. v. 226, p. 85-98, jul. 2015

ZHANG, T. Y.; SUEN, C. Y. A fast parallel algorithm for thinning digital patterns. **Communications of the ACM**, v.27, n.3, p.236-239. 1984.