

Identificação automática de cromossomos humanos

Guilherme C. Kurtz, Tiago B. Bonini¹, Giovani R. Librelotto (orientador), Luis A. Perles, Michele R. Sagrillo

UNIFRA – Centro Universitário Franciscano, Rua dos Andradas, 1614, Santa Maria, RS
{guitarro17,tiagobonini}@gmail.com,{giovani,perles,sagrillo}@unifra.br

Abstract

Karyotyping is the chromosome classification and identification. Such task may take some time. With a picture of chromosomes is possible to automatize such process. One of the first steps is an image segmentation where all chromosomes are extracted from the image. To do this, a sequence of blurring, edge detection, erode/dilate is applied to the image to make a mask. The last step creates islands inside some chromosome mask leading to data loss. An algorithm has been developed to delete those islands. A previous step for the classification has been implemented using the position of the centromere determined by detection of the narrow part of the chromosome. Further developments include arm proportion and banding pattern recognition systems.

Resumo

Cariotipagem é a classificação e identificação dos cromossomos. Tal tarefa leva algum tempo. Com uma imagem de cromossomos é possível automatizar tal processo. Um dos primeiros passos é a segmentação da imagem onde todos os cromossomos são extraídos da imagem. Para isso, uma seqüência de suavização, detecção de bordas, erosão e dilatação é aplicada a imagem para fazer uma máscara. O último passo cria ilhas em alguns cromossomos da máscara levando a perda de dados. Um algoritmo foi desenvolvido para deletar essas ilhas. Um passo anterior da classificação foi implementado usando a posição do centrômero determinada pela detecção da parte mais estreita do cromossomo. Desenvolvimentos futuros incluem a proporção dos braços e sistemas de reconhecimento de padrão de bandas.

1. Introdução

O desenvolvimento da tecnologia tem possibilitado a exploração criteriosa de diversos segmentos relacionados à área da ciência que estuda a cadeia

DNA e cromossomos, denominada citogenética. Com isso, têm sido oferecidos no mercado mundial alguns sistemas computadorizados dedicados à análise cromossômica. Estes sistemas apresentam funções tais como a geração da representação dos cromossomos em pares (pareamento) de acordo com padrões internacionais, conhecidos como Cariótipo.

Apesar da demanda crescente, é pequena a oferta de sistemas automáticos que auxiliem o trabalho dos geneticistas na coleta de dados e geração do Cariótipo. Portanto, o objetivo do presente trabalho é desenvolver um sistema computadorizado utilizando técnicas de visão por computador capaz de auxiliar o geneticista tanto na execução da análise da forma do cromossomo humano a fim de realizar o pareamento assistido dos cromossomos como também auxiliar na coleta e análise de dados para realização do estudo do cariótipo [2].

Para criar um sistema que identifique cromossomos automaticamente, optou-se pelo processamento de imagens. Essa área tem como objetivo a melhora do aspecto visual de imagens, além de fornecer métodos para a sua interpretação, possibilitando a submissão para posteriores etapas de processamento. Detalhes sobre este tema são descritos na seção 2.

Para efetuar um comparativo da abordagem utilizada com outras soluções existentes, a seção 3 apresenta trabalhos relacionados. A seção 4 apresenta a metodologia proposta para a identificação automática de cromossomos. Por fim, a conclusão do artigo é efetuada na seção 5.

2. Processamento de imagens

O processamento de uma imagem digital abrange tanto fundamentos teóricos quanto a parte de software e de hardware e é dividido em cinco etapas: a aquisição da imagem, o pré-processamento, a segmentação, o reconhecimento e a interpretação [3]. Dentre estas, as que serão abrangidas neste trabalho são as etapas de pré-processamento e segmentação, principalmente.

¹ Aluno bolsista de PROBIC – UNIFRA.

Após obter uma imagem digital, é necessário fazer o seu pré-processamento e em seguida sua segmentação, de modo a utilizar-se de técnicas para o realce de contraste e brilho e remoção de ruídos. Essas técnicas se dividem em duas categorias: métodos no domínio espacial e métodos no domínio de frequência [3]. O *domínio espacial* refere-se à manipulação direta dos *píxeis* da imagem. Técnicas no *domínio de frequência* são baseadas nas modificações através das transformadas de Fourier. O trabalho desenvolvido utilizou técnicas no domínio espacial.

Além disso, é necessário fazer a separação das regiões de interesse da imagem (segmentação). A saída do estágio de segmentação constitui-se de dados em forma *píxeis*, que corresponde tanto à fronteira de uma região como a todos os pontos dentro da mesma.

3. Trabalhos Relacionados

A ferramenta KEI [4] utiliza uma técnica de detecção de bordas para elaborar a segmentação da imagem. É possível que o operador altere o resultado da segmentação caso a imagem não traga um resultado final satisfatório. Após a segmentação e cariotipagem, o sistema apresenta imagens dos cromossomos classificados, permitindo ao usuário efetuar trocas entre eles caso o resultado não esteja totalmente correto.

Além disso, podemos citar o que foi desenvolvido por [7], no qual se utilizou a técnica de aprendizado com redes neurais para diminuir a taxa de erros na classificação de cromossomos. Neste trabalho, testaram-se diversas arquiteturas de redes neurais, visando comparar o resultado obtido com outras abordagens, como Algoritmos de Busca em Grafos, Redes de Markov e os Algoritmos de Transportes.

O Kario foi criado para facilitar o trabalho dos geneticistas no reconhecimento do cariótipo [8]. O Kario utiliza um banco de dados MS Access para o armazenamento dos dados textuais, sobre os quais ele permite que se façam consultas. Entretanto, tal como o KEI, o reconhecimento dos cariótipos não é 100% fiel.

3. Identificação Automática de Cromossomos

A análise e identificação de cromossomos humanos são chamadas de cariotipagem. A espécie humana possui 23 pares de cromossomos, sendo 22 pares autossomos e um par de cromossomos sexuais (XX ou XY). A automação deste processo torna-se útil visto que as imagens obtidas geralmente são de baixa qualidade, dificultando sua classificação.

O ponto de partida neste projeto é uma imagem digitalizada de um conjunto de cromossomos humanos, como demonstrado na figura 1. Esta imagem passará por etapas de pré-processamento e segmentação, podendo ter interferência direta de um operador, caso seja necessário. Após essas etapas, a imagem é submetida a um algoritmo de reconhecimento de padrões [5], que terá como critério de classificação o tamanho e o padrão de bandas dos cromossomos.



Figura 1. Imagem de cromossomos humanos

A partir da imagem original, cria-se uma nova imagem chamada de máscara (seção 4.1). Essa máscara será uma imagem binária obtida através de processos de suavização, limiarização, dilatação, erosão e remoção de ilhas. A partir da imagem pré-processada, a segmentação tratará de separar cada uma das regiões (cada cromossomo) e armazená-las separadamente.

Logo, o reconhecimento de padrões utiliza-se de ideogramas como modelos para a tentativa de reconhecimento dos cromossomos. Duas técnicas foram implementadas: o reconhecimento por tamanho e o reconhecimento por padrão de bandas. Por fim, exibe-se a seqüência dos cromossomos classificados.

3.1. Pré-processamento

Neste trabalho, a etapa de pré-processamento envolve a utilização de técnicas de realce de contraste, suavização, dilatação, erosão. Estas técnicas visam como resultado final uma imagem monocromática, destacando os cromossomos da imagem com a cor branca e as regiões que não interessam com a cor preta.

A suavização de imagens é aplicada com o fim de eliminar os ruídos, reduzindo detalhes que não são desejados. Para isto, é aplicado um filtro espacial de passa baixa, utilizando *vizinhos-de-8* [3]. O resultado da aplicação deste filtro será uma imagem com menos ruídos, o que facilita um realce de contraste mais uniforme sem extrapolar os limites desejados.

Neste caso, dado um intervalo em níveis de cinza, todos os *píxeis* que contenham um valor que pertença a este intervalo terão seu contraste ajustado para o branco total (nível 255) ou preto (nível 0). Entretanto, somente técnicas de suavização e realce de contraste

não são suficientes para se obter uma imagem capaz de destacar os cromossomos de uma imagem.

Para facilitar a visualização e a definição das regiões procuradas, aplica-se uma dilatação sobre a imagem, eliminando lacunas. Por sua vez, a erosão de uma imagem serve para a eliminação de detalhes que são irrelevantes em uma imagem.

Além das transformações de erosão e dilatação, desenvolveu-se uma nova transformação com base nesses dois, que será chamado de remoção de ilhas. Por exemplo, a figura 2 apresenta vários *pixels* da com nível de cinza igual a 0 (ou *pixels* pretos) na imagem. Se este *pixel* estiver em uma ilha de *pixels* pretos cercados de *pixels* de nível de cinza igual a 255 (ou *pixels* brancos), todos estes *pixels* deverão ter seu valor de nível de cinza alterado de 0 para 255, pois estes constituem de um buraco na máscara do cromossomo.

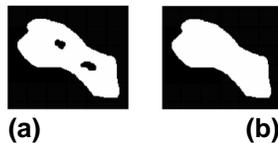


Figura 2. Exemplo da imagem antes e depois da aplicação da remoção de ilhas

Com estes filtros e transformações, a imagem original é mantida intacta e é criada a uma nova imagem contendo uma máscara sobre a original, como se percebe na figura 3.



Figura 3. Máscara

A partir de então, a etapa de segmentação irá trabalhar na separação de cada um desses cromossomos com a finalidade de iniciar o processo de identificação dos mesmos.

3.2. Segmentação

A partir de uma imagem pré-processada, o objetivo da segmentação é obter um conjunto de segmentos separados com informações em relação à imagem de origem. A segmentação utilizada foi por regiões, as quais são os cromossomos separados.

Após o pré-processamento, faz-se uma varredura pela máscara na busca por estes *pixel* com nível de cinza igual a 255 (branco total): a partir de um *pixel* encontrado com nível de cinza igual a 255, todos os *pixels* vizinhos-de-4 são selecionados e armazenados

até que se encontre um novo *pixel* com valor 0. A partir destes *pixels* armazenados, o processo se repete até que todos os *pixels* desta região de interesse tenham sido armazenados. Após isso, o processo se repete na tentativa de encontrar uma nova região de interesse.

Logo, todas as regiões que contenham pontos interligados de níveis de cinza com valor de 255 são retiradas da imagem original (tendo como referência a máscara) e armazenadas separadas em um vetor que contem os *pixels* de cada região e sua posição na respectiva imagem. A Figura 3 apresenta a máscara da imagem original (Figura 1) obtida após o seu pré-processamento e pronta para ser usada como referência na segmentação. Ao utilizar uma imagem binária como máscara, as regiões de interesse que devem ser segmentadas ficam limitadas aos *pixels* de nível de cinza igual a 255, e as regiões de não interesse com nível de cinza igual a 0.

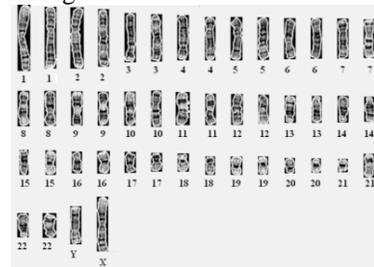


Figura 4. Resultado final após a identificação dos cromossomos da Figura 1

Por fim, tem-se armazenado todas as regiões de interesse da imagem separadas, possibilitando o reconhecimento de padrões. Porém, nessa fase surgem diversos problemas, tais como imagens com cromossomos tortos, inclinados, ou até mesmo sobrepostos. Os problemas de inclinação são ajustados após os cromossomos estarem segmentados, através da rotação dos mesmos. Com um pré-processamento bem elaborado, alguns destes problemas são resolvidos sem a necessidade de ajustes futuros, facilitando a identificação dos cromossomos.

3.3. Reconhecimento de padrões e identificação dos cromossomos

Dentre as características utilizadas na cariotipagem, o método aqui proposto utiliza inicialmente o tamanho do cromossomo e o seu padrão de bandas[1]. Para ambas as características avaliadas, foram utilizadas imagens padrões na forma de ideogramas. Os ideogramas contêm o padrão de bandas de todos os cromossomos e o seu tamanho proporcional em relação aos outros. Isso possibilitou a utilização de um único conjunto de objetos que serve para os dois casos de avaliação, facilitando uma futura mixagem dos

métodos. Na figura 5 percebe-se uma imagem contendo os modelos utilizados nesta abordagem.

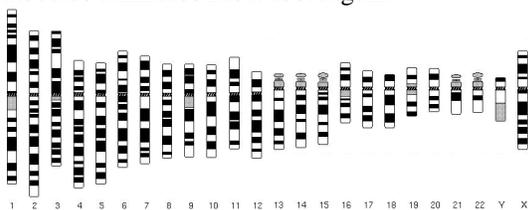


Figura 5. Ideograma

Na cariotipagem baseada no comprimento, é realizada uma medida longitudinal do cromossomo. Cada cromossomo na célula tem um comprimento único [1], o que possibilita a classificação dos mesmos utilizando métodos baseados no comprimento. O processo é comparar cada ideograma com todos os cromossomos de acordo com a proporção de seu tamanho com a proporção do tamanho dos demais cromossomos na célula. Então, é realizada uma busca visando a menor diferença possível entre os ideogramas e o cariótipo.

O padrão de bandas possibilita combinações diversas no processo de classificação da amostra [1]. Utilizou-se o padrão de bandas através de métodos no conjunto espacial. A busca se dá na seguinte forma: cada ideograma é comparado com todos os cromossomos. Nestes, é feita uma mudança de escala de acordo com o tamanho longitudinal do ideograma. Então, calcula-se o somatório da diferença dos níveis de cinza da imagem do ideograma (i) com a imagem o cromossomo (p), sendo o resultado final o índice de semelhança entre ambos:

$$\sum_{j=0}^h |p(w/2, j) - i(w/2, j)|$$

O cálculo é feito a partir da coluna do meio da imagem (sendo w a largura da imagem, $w/2$ são os *píxeis* do meio). A menor diferença dará a classificação do cromossomo como sendo o mesmo do ideograma que foi comparado.

5. Conclusão

Atualmente, existem diversas técnicas de processamento de imagens para o reconhecimento de padrões cromossômicos. Porém, nenhuma delas retorna um resultado considerado ótimo [1]. Se por um lado, a utilização de técnicas de aprendizado traz uma resposta melhor a este problema, por outro, a utilização das mesmas pode se tornar um processo demorado e com alto custo, como o uso de técnicas de aprendizado com redes neurais para a busca da classificação dos mesmos [7]. Entretanto, por vezes é necessária a utilização de técnicas menos exatas, mas com um custo inferior.

Desta forma, o presente artigo abordou a construção de uma ferramenta de identificação automática de cariótipos. A metodologia envolveu praticamente todas as etapas do processamento de imagens, de forma a permitir o reconhecimento de cromossomos humanos, inclusive, em imagens de baixa qualidade.

A ferramenta atual possui mecanismos para manipular a imagem antes das etapas de segmentação, tais como separação de regiões que não deveriam estar unidas, recorte, dilatação, erosão, entre outras.

Das técnicas utilizadas para o processo de identificação dos cromossomos, a classificação por tamanho (demonstrada na figura 5) teve um desempenho melhor. Existe a possibilidade de utilizar os pontos positivos de cada uma delas para criar um novo método de reconhecimento.

Para a intervenção do operador, desenvolveu-se uma interface que permite efetuar os ajustes na máscara da imagem, caso necessário. Esta intervenção auxilia nas etapas de segmentação e identificação. Entretanto, o objetivo é reduzir o trabalho do operador e tornar o processo o mais automático possível.

Os próximos passos envolvem o tratamento de cromossomos tortos e sobrepostos, bem como uma bateria de testes sobre o sistema envolvendo imagens das mais diversas qualidades, contendo uma variada disposição dos cromossomos nas mesmas.

6. Referências

- [1] Biyani, P., Wu, X., and Sinha, A.(2005) “Joint Classification and Pairing of Human Chromosomes”. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 2, No. 2, April-June.
- [2] Carothers, A., and Piper, J (1994). “Computer-aided classification of human chromosomes: a review”. Journal Statistics and Computing. Springer. Vol. 4, No. 3.
- [3] Gonzales, R. C., Woods, R. E., Junior, C. (2000) “Processamento de imagens digitais”. Blucher.
- [4] Qualiterm Eletrônica (2003). “Key: Editor de Cariótipos Profissional”. <http://www.qualiterm.com.br/folhetos/key.pdf>
- [5] Schalkoff, R. (1992). Patter Recognition – Statistical, Structural and Neural Approachs, John Willey & Sons.
- [7] Todesco, J. L. (1995). Reconhecimento de Padrões Usando Rede Neural Artificial com Função de Base Radial: Uma Aplicação na Classificação de Cromossomos Humanos. Tese de Doutorado, UFSC.
- [8] ZETA Imaging (2008) “Kario Ver. 8.8”. <http://imageanalysis.altervista.org>