

Homologue Classification of Human Chromosome Images Using an Iterative Centromere Segmentation Algorithm

Parvin Mousavi¹, Rabab K. Ward¹, Mohammad Sameti¹, Peter M. Lansdorp² and Sidney S. Fels¹

Abstract - Segmentation of centromeres is a major step towards classification of homologous chromosomes, which in turn, is essential to advanced studies of cancer genetics. This paper describes an iterative fuzzy algorithm, which successfully segments the centromeres of human chromosome images. The algorithm is based on assigning a fuzzy membership value to each pixel and iteratively updating an error function. Chromosome 22 is then used to verify the centromere segmentation method.

Keywords - Homologue classification, segmentation, fuzzy sets.

I. INTRODUCTION

At present, advances in cancer research are largely dependent on developments in image processing techniques. This is partly due to the vast usage of Fluorescence In-Situ Hybridization (FISH) microscopy of human chromosomes for cancer research.

The nucleus of cells in the human body is made up of 23 pairs of chromosomes. In each pair, one chromosome is inherited from the mother known as the maternal homologue and the other chromosome is inherited from the father known as the paternal homologue. Scientists believe cancer is related to specific chromosome abnormalities. To study the characteristics of cancerous cells, it is essential to have the ability to analyze separate homologues [1].

Despite the vast amount of research conducted on analyzing chromosome images, it has only been a few years since the problem of homologue classification (into maternal and paternal classes) has been addressed [1,2,3,4]. Microscopy images of chromosomes available in the past, did not contain enough quantitative information. However, recently, novel Peptide Nucleic Acid (PNA) probes were developed at Terry Fox laboratory of BC Cancer Research Center (BCCRC) which provide chromosome images with high quantitative information [5].

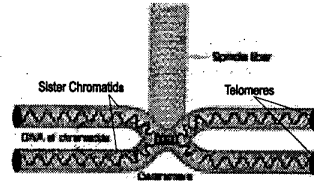


Figure 1 – Structure of a human chromosome

In this paper, a new algorithm is proposed for centromere (central part of chromosome) segmentation (Figure 1) [6]. The algorithm is based on fuzzy set theory and gradient descent method. Centromeres of human chromosomes are segmented and distinguished from the background using this algorithm.

There is no biological method to verify the segmentation results; however, for the specific chromosome 22, paternal and maternal homologues are differentiable due to morphological variations. We use this information to validate our segmentation algorithm. Homologues of chromosome 22 are classified based on the results from the segmentation algorithm as well as their morphological differences. The results of these classifications are compared to confirm our segmentation method.

The layout of the paper is as follows. In Section II, we give an overview of the database preparation and image acquisition. Section III discusses the centromere segmentation algorithm. In Section IV we present the results of this algorithm as well as homologue classification for chromosome 22. The conclusion of the paper is presented in Section V.

II. IMAGE ACQUISITION AND DATABASE PREPARATION

Prior to image acquisition, a slide of metaphase chromosomes is treated with certain types of PNA probes. Probes are fluorescent nucleic acid segments that bind to specific sites (sub-structures) on chromosomes and make these sites observable. Twelve images of the prepared slide

¹ Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, B.C Canada

² Terry Fox Laboratory, BC Cancer Research Center, Vancouver, B.C Canada



Figure 2 - Snapshot Snapshots of a metaphase image using DAPI (left) and FITC-centromere (right) probes

are then taken by a fluorescence microscope (Widefield Zeiss Axioplan) and preprocessed prior to being used. As probes are fluorescent, only the strings the probes bind to, are visualized by the fluorescence microscope [1]. The probes used for preparing our database are DAPI and FITC-Centromere probes. DAPI has a blue appearance and highlights the whole chromosome. FITC, on the other hand, has a green appearance and makes only centromeres of the chromosomes visible (see Figure 2).

The image acquisition system is divided into three subsystems as follows, 1) A fluorescence microscope which produces the images of the prepared slide, 2) A CCD camera to capture and digitize the image and 3) A computer system which stores the images, sends instructions to the microscope stage and is used in the preprocessing of the stored images. Preprocessing is another important stage in preparing our images. Acquired images of a single slide may differ from one another due to a number of factors such as the preparing conditions of the slide, the location of the slide in the field of view of the microscope, the illumination stability and photobleaching. Preprocessing eliminates most of these effects and enables us to have consistent output images. The algorithms used at this stage include background subtraction, wavelength compensation and photobleaching correction methods [2].

The database is created by cutting (isolating) every homologous chromosome pair (from the DAPI image) as well as its corresponding centromeres (from the FITC image) and forming separate images for each. We developed an automated algorithm for this purpose. The algorithm uses a table of co-ordinates of all the chromosomes to find and cut each chromosome and its corresponding centromere as a rectangular region. These rectangular regions are saved as images that make up our database. The objective is to segment all the centromere images from their background and extract features from the segmented centromeres for homologous classification purposes.

III. CENTROMRE SEGMENTATION

Although centromeres in the FITC images have higher intensities than their background, segmentation is not a

straight forward task. The boundaries of centromeres in microscopy images are not well defined and have extremely gradual transitions, therefore simple thresholding methods are not successful in segmenting them from their background. Our proposed segmentation method is based on fuzzy sets approach [6]. As the centromeres have unclear boundaries, the fuzzy set approach makes the process of assigning pixels to centromere/background regions more accurate. On the other hand, sometimes, there exists a small spectral overlap of DAPI images in the FITC-centromere image. In order to classify a pixel as part of the centromere, we need to examine the intensity of the surrounding pixels as well as the gray level value of that particular pixel. The centromere segmentation is performed in three steps described in the following sections.

A. INITIALIZATION

The acquired microscopy images of centromeres are indexed images with a gray-level range of 0-255. Each image is initially converted to an intensity image by determining the minimum and maximum gray levels of the image and normalizing all the pixel values in the image to a number between 0.0 and 1.0.

As mentioned above, for determining the membership of a pixel to the centromere, the surrounding pixels of that particular pixel need to be examined. Therefore a 3x3 neighborhood mask with different weight coefficients is defined as:

$$W(i, j) = \exp\left(\frac{\|(i, j)\|}{\alpha}\right) \quad (1)$$

where $\|(i, j)\|$ is the Euclidean distance between position (i, j) and the center of the mask [6]. The constant α controls the shape of the exponential and is set to 2.

B. FUZZY MEMBERSHIP VALUES

A fuzzy membership function is applied to the centromere image assigning a membership value to each pixel of the image. The fuzzy membership value of each pixel, determines the degree of its membership to the centromere or the background. The threshold T always corresponds to the membership value of 0.5, i.e. the location of T relative to the function curve is fixed. Usually the threshold, T , is chosen in the valley between the two peaks of the image histogram, however, for our images this is not applicable. The histograms of centromere images do not have valleys. This is due to the low number of centromere pixels in the image as well as the gray level distribution of the centromere images. In our application, we choose T at a point right after the peak of the image histogram averaged over all centromere images. In this case $T=0.3$.

C. ERROR CALCULATION AND UPDATING

After assigning a membership value to each pixel, an iterative process determines the membership of each pixel to a region. The error function is defined as:

$$E = \sum_{i,j} O(i,j)(1 - O(i,j)) \quad (2)$$

where $O(i,j)$ is fuzzy membership value of the image pixel (i,j) . The error, E , is a non-negative number and is minimized only if the fuzzy membership value, $O(i,j)$, of each image pixel is either 1 or 0. Inspired by the Gradient method and Back-Propagation algorithm, the updating rule is written as:

$$O^*(i,j) = O(i,j) + \eta \left(-\frac{\partial E(i,j)}{\partial O(i,j)} \right) \cdot O(i,j)(1 - O(i,j)) \quad (3)$$

where $E(i,j)$ is the error caused by pixel (i,j) , and η is the learning rate. In order to include the role of neighboring pixels in defining a pixel as centromere, and after substituting Eq. 2, the latter becomes:

$$O^*(i,j) = O(i,j) + 2\eta \left[\sum_{k,l \in N^0} w(k,l)(O(k,l) - 0.5) \right] \cdot O(i,j)(1 - O(i,j)) \quad (4)$$

where $W(i,j)$ are the weight coefficients of the mask defining a neighborhood N^0 around $O(i,j)$.

Note that for pixels with a fuzziness value close to 0 (or 1), it is highly unlikely to become members of region "1" ("0"), and therefore, the updating process can be accelerated. To do so, Eq.4 is changed to:

$$O^*(i,j) = O(i,j) + \nabla O(i,j)$$

where

$$\nabla O(i,j) = \begin{cases} 0 & ; O(i,j) = 0 \text{ or } 1 \\ 2\eta \left[\sum_{k,l \in N^0} W(k,l)(O(k,l) - 0.5) \right] & ; \text{else} \end{cases} \quad (5)$$

Equation 5 makes a quick decision for pixels with fuzzy values close to 0 or 1 (because $(O(k,l) - 0.5)$ has a large value), but for a fuzzier pixel $((O(k,l) - 0.5)$ closer to zero), there is a smaller change in its intensity, causing a delay in segmentation of the pixel.

IV. RESULTS AND DISCUSSION

As described in the previous section, a segmentation algorithm was developed based on fuzzy sets theory and error back-propagation. The algorithm was implemented

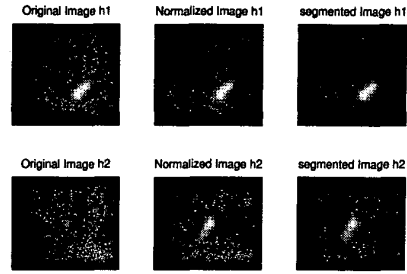


Figure 3- Centromere segmentation for chromosome 22 in a Metaphase

and used to segment the centromeres from their background for the prepared database. Segmentation was performed successfully on all of the centromere images and the 3x3 mask seems a reasonable size for testing the neighborhood of the pixel to be classified. An example of the segmentation is shown in Figure 4 for two homologues of chromosome 22 in a particular metaphase. The original, normalized and segmented centromeres for homologue one and homologue two of chromosome 22 are shown in the top and bottom rows, respectively.

As mentioned before, there is no biological method to validate our segmentation. To verify our results, we used chromosomes 22 as it is possible to classify their paternal and maternal homologues using differences in their DAPI images. Segmentation results (using our method above) of the FITC images are also used to classify chromosome 22 by measuring an intensity feature over the segmented area. The classifications of the two methods should agree.

A. HOMOLOGUE CLASSIFICATION OF CHROMOSOME 22

Once the centromere regions in FITC images are segmented for all chromosomes 22, the total integrated fluorescence intensities are calculated over each region. The IFI values are computed for the two homologues of

chromosome 22, in twelve metaphase images of a patient (the prepared database) as shown in Table 1. Homologues of chromosome 22 are classified into two groups using the differences in their IFI values.

On the other hand, the two homologues of chromosome 22 show apparent differences in their upper arms (known as p-arms) on DAPI images (see Figure 5). One homologue has short p-arms (top left) while the other homologue has no p-arms (bottom left). Homologues of chromosome 22 are also classified using this morphological difference. In Figure 4 an example of chromosome 22 homologues with and without p-arms as well as the segmented centromeres and their total IFI values are shown. As seen in this figure, the homologue with a p-arm has a higher IFI value than the homologue without p-arm.

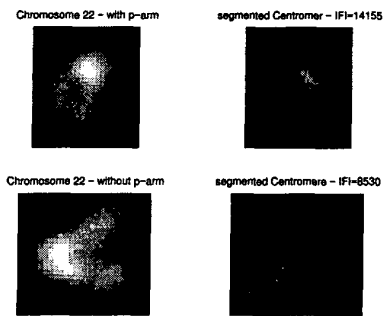


Figure 4- DAPI and segmented centromere images of chromosome

Looking at the total IFI values of all the segmented centromeres of chromosomes 22 in Table 1, it is reconfirmed that chromosomes 22 with a p-arm have higher centromere intensities than chromosomes 22 without a p-arm. Classification can be performed successfully using both the difference in the p-arm and the difference in centromere IFI. The accuracy of the classification is 100% and the results of the two methods comply perfectly.

Chrom.22	Image1	Image2	Image3	Image4
w p-arm	11881	10712	7157	6421
w/o p-arm	7356	7525	4687	5166
Chrom.22	Image5	Image6	Image7	Image8
w p-arm	7348	5619	7656	8513
w/o p-arm	5403	4181	5134	4913
Chrom.22	Image9	Image10	Image11	Image12
w p-arm	8433	7991	9818	7337
w/o p-arm	5029	4753	5206	4827

Table 1 - IFI values of segmented FITC-centromeres of homologous chromosomes 22 from the database

V. CONCLUSION

In this paper an iterative fuzzy segmentation algorithm was implemented and applied to FITC-centromere images of chromosomes. Using this algorithm, the centromeres were successfully segmented from their backgrounds. To verify the results, the total integrated fluorescence intensity of the segmented regions were calculated for chromosome 22 and the homologues of this chromosome were classified into two groups of maternal and paternal homologues. Since chromosomes 22 show differences between homologues in the DAPI images, we used this criterion to classify chromosome 22 into parental homologue classes as well. The classification results using these two criteria were compared and the two methods complied completely. This proves that our segmentation algorithm was successful in segmenting the centromeres from their backgrounds. The

results of this paper are a major step towards multi-feature analysis of chromosome images for homologue classification purposes.

Cancer is known as a somatic genetic disease characterised by specific mutations and genetic instability. Studies have shown various predispositions to cancer are inherited as abnormal genes. Statistics also suggest occurrence of some familial cancers (e.g. breast cancer). Following the approaches described, the results of this paper will allow classification of homologous chromosomes and will help track the responsible genes for cancer thus better understand the genetics of cancer.

BIBLIOGRAPHY

- [1] P. Mousavi, R. K. Ward, P. M. Lansdorp, "Feature analysis and classification of chromosome 16 homologues using fluorescence microscopy image," *IEEE Canadian Journal of Electrical and Computer Engineering*, Vol.23, No. 4, pp. 95-98, 1999.
- [2] S. S. S. Poon, *Telomere length measurements using fluorescence microscopy*, Ph.D. thesis, Dept. of Elec. and Comp. Eng., University of British Columbia, 1997.
- [3] P.M. Lansdorp, N.P. Werwoerd, F.M. Van de Rijke, V.D. Little, R.W. Dirks, A.K. Raap, H.J. Tanke, "Heterogeneity in telomere length of human chromosomes," *Human Molecular Genetics* Vol.5, No.5, pp. 685-691, 1996.
- [4] U.M. Martens, J.M. Zijlmans, S.S.S. Poon, V. Drogoulsz, J. Yui, E. Chavez, R.K. Ward, P.M. Lansdorp, "Short telomere on the P-arm of human chromosome 17," *Nature Genetics*, Vol.18, pp. 76-80, 1998.
- [5] P. M. Lansdorp, V. Dragowska, N. Rufer, T. Brummendorf, S. S. S. Poon, P. Mousavi, T. Duncan, U. Martens, "Applications of Peptide Nucleic Acid Probes in Cytometry," *Proceedings of the ISAC XLIX International Congress*, 1998.
- [6] M. Sameti, R. K. Ward, "A fuzzy segmentation algorithm for mammogram partitioning," *Digital Mammography 96*, eds. K. Doi, M. L. Giger, R. M. Nishikawa and R. A. Schmidt, Elsevier Science B.V., Amsterdam, Newyork, 1996.