Karyotyping Techniques of Chromosomes: A Survey

Revathy M Nair, Remya R S, Sabeena K

Department of Computer Science and Engineering
College of Engineering, Karunagappally
Kollam, Kerala, India

***Abstract*— Chromosome analysis is used by clinicians for the early detection, prognosis and treatment evaluation of genetic disorders like leukaemia, Down syndrome etc. The detection is done with the help of karyotypes. Karyotyping is a challenging and difficult task. So an automated system of karyotyping is needed for doing it in less time and with more accuracy. In literature various methods have been proposed for efficient and automated Karyotyping. But an accurate and more efficient computer assisted system needs further research and studies. In this paper we are making a survey of some of the methods for karyotyping in the literature proposed so far and critically reviewing them by listing the strengths and weakness of each of them.**

***Keywords* — Chromosome, Metaphase, Karyotyping, Medial Axis, Feature Extraction, Classification**

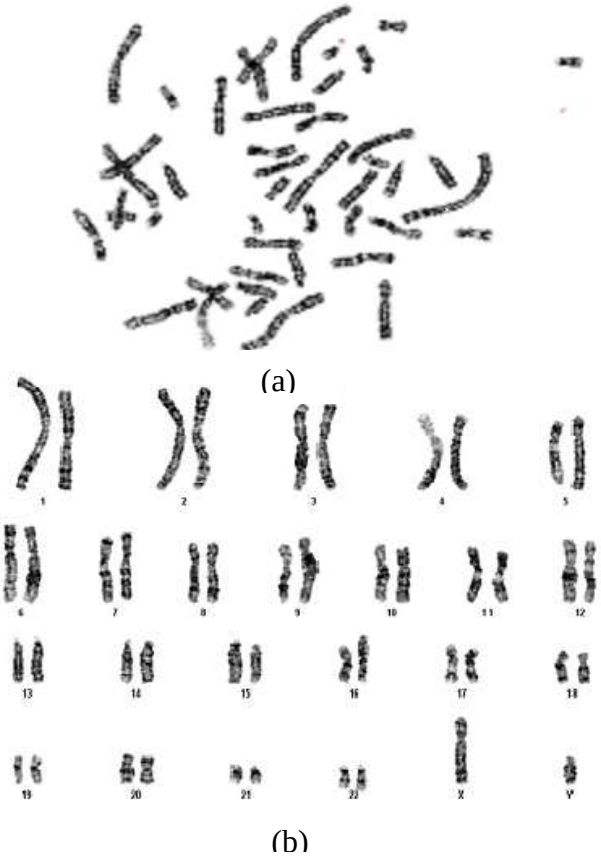
I. INTRODUCTION

Chromosomes are the hereditary material present inside the human cell for carrying genetic information through generations. During cell division the chromatids in the cell are condensed to form the chromosomes. For human beings 46 chromosomes are present that are arranged in 23 pairs. In these 22 pairs are called autosomes and one pair is the sex chromosome. For detecting the genetic diseases the analysis of chromosome is needed. This is done with the help of karyotypes. Chromosome analysis is done on karyotypes also for prognosis of disease and for treatment.

Karyotyping is the process of generating a karyotype. The karyotype of an organism is usually displayed in photomicrographs wherein chromosomes are arranged in homologous pairs, and in descending order of size and relative position of the centromere. Fig. 1a shows the metaphase image of chromosomes in which the chromosomes are spread and randomly distributed. Fig. 1b shows the karyotype image in which the chromosomes are paired and arranged in descending order of their lengths. From this karyotype various diseases are finding out by identifying the numerical and structural abnormalities.

The process of karyotyping is completed through many steps like enhancement, segmentation, feature extraction, classification etc as shown in Fig 2. The metaphase image taken for karyotyping, is noisy or usually with low contrast. Pre-processing methods are applied to the metaphase image to avoid noise and enhance the contrast of the image. Then for the separation of individual chromosomes from the spreaded image, a segmentation method is used. From each segmented chromosomes various features are extracted and based on these features the pairing and classification are done.

Then an analysis is performed for the detection of abnormalities in chromosome. The number of chromosomes are either less or greater due to some diseases. This can be identified by using the numerical abnormality detection methods. The translocation, deletion etc. of chromosomes can be identified by using structural abnormality detection methods.



(a)

(b)

Fig.1 (a) Metaphase image (b) Karyotype of the chromosomes

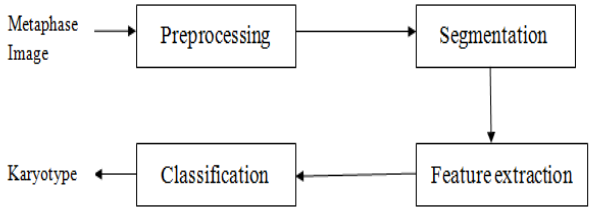


Fig.2 Steps of karyotyping

The overview of this paper is as follows. Section I is about the pre-processing methods. Section II describes various

segmentation methods and section III describes feature extraction methods. Section IV focuses on different classification methods for efficient karyotyping.

I. PREPROCESSING

Since the image is taken from the blood samples using different techniques some noises is added to the image. In preprocessing, steps are taken for removing the noise from the metaphase image and to improve the contrast. Sometimes this noises leads to misclassification of chromosomes. Some smoothing filters [1] such as Gaussian low pass filter and median filters are generally used for reducing the noise in the image.

Sometimes due to the lighting or microscope properties the metaphase image posses low contrast. In this case the classification becomes difficult. To improve the contrast some image processing techniques can be used. Sobel filter, Roberts filters [2], Laplacian pyramids [3] are generally used to improve the contrast. Technically it is a discrete differentiation operator computing an approximation of the gradient of the image intensity function. At each point in the image the result of the Sobel filter is either the corresponding gradient vector or the norm of the vector. Roberts filter is also a differential operator, which calculate gradient through discrete differentiation which is achieved by computing the sum of the squares of the differences between diagonally adjacent pixels. These are applied because they also provide important features like edge, band etc. They also sharpen the image.

Histogram equalization [4] is another method to improve the quality of the image. It improves the global contrast of the image. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

Recently a wavelet based enhancement algorithm [5] is used for chromosomes which give a good contrast improvement. These all image enhancement techniques improve the quality of the metaphase image and hence improve the classification efficiency.

III.SEGMENTATION

Segmentation is the separation of individual chromosome from the metaphase spread of chromosomes. The main challenge during the segmentation step is overlapping of chromosomes. The chromosomes in the metaphase spread may be overlapping. The separation is a difficult task. Different segmentation methods are present.

A. *Thresholding*:

Otsu thresholding [6], Kapur thresholding [7], fuzzy c means clustering on intensities [8], multistage adaptive thresholding [9], region based level sets [10] and genetic algorithm based methods [11]. The idea beyond the Otsu's method is that given two classes of pixels in an image, it is desirable to find the separation that minimizes the combined intra class variance. In Kapur's method it is assumed that an image is the outcome of two probability distributions, one for the object and other for the background. In order to obtain the threshold level maximizes the total entropy. The genetic algorithm consists of initialization, selection, single point cross over, mutation and decoding. In fuzzy c mean clustering method segment the chromosome from the background to tackle the non sharp margins of the chromosomes, and the fact that the distributions of features of the pixels belonging to chromosomes and background are often not well separated. It is mainly used in M-FISH images. In these algorithms the region based level set is the best method for chromosome segmentation and it gives highest segmentation efficiency.

B. Watershed algorithm

Watershed algorithm [12] is also used for the segmentation purpose, which divides the region in the image into different parts. This gives a matrix where each pixel has the value corresponding to the class in which the pixel included.

C. Region based segmentation

The individual chromosomes can then be identified using the region based approach, which is used to detect the connected components in a region. The algorithm starts with a seed region and searches for the neighbourhood and merges them to the same region if they have the value same value, otherwise it is labelled as a new region. The neighbourhood may be either 4-connected or 8- connected. For the overlapping segmentation the minimum entropy segmentation [13] and cut point detection algorithms [14] are mainly used.

IV. FEATURE EXTRACTION

Features are extracted from the individual chromosomes to classify them based on the features. Feature extraction is an important step because the classification efficiency is based on the features. Several features are present for the chromosome classification. After segmentation features are extracted for classification of chromosome. Different types of image features including, texture, numerical, morphological features, density profile and frequency domain features sampled from the results of wavelet or Fourier transformation have been investigated and compared to optimally represent chromosomes. Texture features includes edges, contrast, correlation, entropy etc. Numerical features include chromosome numbers and morphological features include length, area, convex hull perimeter etc.

A. Relative length of chromosome:

Length is an important feature for classification because it gradually decreases from class 1 to 22. Length is obtained by using the medial axis. Medial axis is the centerline of the chromosome that passes through the centromere. The length of the medial axis is considered as the length of the chromosome [16]. Relative length is the ratio of the particular chromosome and the sum of length of all the chromosomes in a set [15]. This relative length is a better feature than the length only.

B .Relative area:

Area is the number of pixels present in the chromosome. Relative area [15] is the ratio of the area of the chromosome to the sum of areas of all the chromosomes in the set.

C. Centromere index:

Centromere is the portion of the chromosome from where the chromatids are started to separate during cell division. The centromere is identified by intensity profile through medial axis [17]. The ratio of length of the p arm to the length of q arm is called centromere index. This is the important feature for karyotyping.

D .Band profile:

Band profile [16] [18] is the graph of the average intensity along the perpendicular line through the points in medial axis to the distance of that point from one end of the chromosome. This is the most important feature of chromosome classification because for the banded chromosome classification this gives the banding patterns. Since the banding patterns are unique for each chromosome, the classification can be easily performed. From this band profile several features are extracted like position of each band, average intensity of each band, width of bands etc. [19]. In [12], the standard deviation, mean etc are used for classification.

E. Wdd features:

These features are also computed by using the band profile of chromosomes [16]. These are obtained by multiplying the band profile with some predefined density distributions.

F. Mutual Information:

Mutual information [18] is an important feature for comparing two chromosomes. Histogram based mutual information is taken between each pair of chromosome and using the values pairing can be done. The mutual information can be computed by using the histogram of each images and the joint histogram of two images

V. CLASSIFICATION

Classification is the identification of each pair of chromosomes based on the extracted features. Classification can be done by using different methods. The main method used in chromosome classification is the artificial neural networks [15]. The extracted features are given to the neural network and train using the known samples. The number of input neurons is the number of features and the number of output class is the number of classes. The numbers of hidden layer neurons are adjusted for obtaining maximum classification efficiency. This ANN is the best known chromosome classifier. In this technique, the available data is divided into three subsets: training, validation and testing sets. The training subset is used for updating the ANN parameters; the testing subset is used for final assessment, and the classification. Error on the validation set is monitored during the training process to avoid over fitting. The validation error will normally decrease during the initial phase of training similar to the training set error. However, when the ANN begins to over fit the training data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights that produced the minimum error on the validation set are retrieved.

A hierarchical multi-layer neural network with an error back-propagation training algorithm has been adopted for the automatic classification of Giemsa-stained human chromosomes. The first step classifies chromosomes data into 7 major groups based on their morphological features such as relative length, relative area, centromere index, and density profiles. The optimal number of processing elements (PEs) in the hidden layer of this neural network was determined so that it showed the best classification result. The second step classifies each 7 major groups into 24 subgroups using each group classifier. The classification error decreased by using two steps of classification and the classification error was 5.9%. Seven neural networks were used to classify chromosomes of each subgroup into individual chromosome. Parameters for the neural network for each subgroup were different according to the subgroups and selected as the same method as the first step classifier.

Other classification methods such as k-Nearest neighbour classifier [12], knowledge based classifier [20], maximum likelihood classifier [16], Bayesian classifier [21] etc are used for chromosome classification. KNN is the method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance based learning, or lazy learning where the function is only approximated locally and all computations is deferred until classification. The k nearest neighbour algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours. If k=1 then the object is simply assigned to the class of its nearest neighbour. The features are given to the classifier and it classifies the chromosomes. The shortcoming of this classifier is that is sensitive to the local structure of the data. In [18] the classification problem is converted into an integer programming problem and then it can be solved by using GLPK kit. In [22] support vectors are used for chromosome classification. This method gives high efficiency and is less complex. But this is an expensive method. So this is not widely used for chromosome classification. The critical review is shown in Table 1.

Table 1:Comparison of different methods of karyotyping

Paper	Methods	Advantage	Disadvantage
On fully automatic feature measurement for banded chromosome classification Jim Piper, Eric Granum Cytometry 10:242-255 (1989)	Segmentation: Thresholding, done by human interaction, find chromosome axis, profile extraction, centromere detection Feature extraction: Area, convex hull perimeter, length, wdd features, centromere index Classification: Maximum likelihood classifier	Maximum number of features are selected	Centromere cannot be detected for all type of images
New features for automatic classification of human chromosomes: A feasibility study Mehdi moradi, S. Kamaledin Setarehdan Elsevier, Pattern recognition letters 27(2006) 19-28	Segmentation: Done by human expert, medial axis extraction, profile extraction Feature extraction: Extraction of bands using windows, features based on bands Classification: ANN	Unique features of chromosomes	Bands are difficult to identify
Linear Discriminate Analysis of the Wavelet Domain Features for Automatic Classification of Human Chromosomes. M. Javan Roshtkhari and S. Kamaledin Setarehdan ICSP 2008 proceedings	Segmentation: done by human expert Feature extraction: Relative length, centromere position, density profile. Wavelet is applied to DP and then LDA is applied Classification: ANN	Use key features for classification	Centromere cannot be extracted for all type of images
A Novel Metric for Bone Marrow Cells Chromosome Pairing. Artem Khmelinskii, Rodrigo Ventura , Joao Sanches IEEE transactions on biomedical engineering, vol. 57, no. 6, June 2010	Segmentation: done by human expert, medial axis extraction Feature extraction: area, length, band profile, mutual information based on histogram, weighted distance of features is taken. D is obtained Classification: integer programming problem	Weighted feature value is taken	Complex
Feature Extraction and Pairing of G-Band Chromosome Images using K-Nearest neighbour Classifier S. Janani, R. Nandakumar, M. Nirmala IJCST Vol. 3, Issue 2, April - June 2012	Segmentation: watershed transform Feature extraction: Band profile. Standard deviation, covariance are extracted by applying 2D wavelet transform Classification: K nearest neighbour classifier	Features are specific, simple	Less efficient than ANN, slow classification, memory intensive
Automatic Chromosome Classification using Support Vector Machines. Christoforos Markou, Christos Maramis, Anastasios Delopoulos ISBN: 978-1-477554-821. iConcept Press	Segmentation: human expert Feature extraction: Band profile Classification :SVM	High efficiency, less complex	Expensive

VI. CONCLUSION

The research for a computer assisted system of chromosome classification has been started for many years. A computer assisted chromosome classifier is required because the diagnosis by a human takes more time and expert knowledge. A computer assisted system diagnosis it in very less time and it uses the knowledge that we give to the system at the development stage. The early developed computer assisted systems classify small sets of chromosomes very efficiently. But since the chromosomes are very different for different individuals, for large datasets the classifiers give very less results. So an efficient system for chromosome classifier is needed.

The banding of chromosome is the best feature for classification process because for each chromosome they are unique. Using the related features of band patterns like the width, average intensity, position etc gives better efficiency. Reasons for better efficiency of artificial neural network are humanoid quality and more number of output classes unlike SVM.

REFERENCES

[1] Weeks A r, Myler H R and Wenaas H G " Computer generated noise images for the evaluation of image processing algorithms" Opt.Eng 32982-92, 1993

[2] Jain A K "Digital Image Processing"

[3] Anderson C H, Greenspan H, Akber S "Image Enhancement by nonlinear extrapolation in frequency space" IEEE Trans. Image Process.9 1035-48

[4] Gonzalez "Digital Image Processing"

[5] Wang Y, Wu Q, Castleman K R and Xiong Z "Chromosome image enhancement using multiscale differential operators" IEEE Trans. Med.Imag.22 685-93, 2003

[6] N Ostu, "A threshold selection method from grey level histogram" IEEE transactions on system man and cybernetics 9(1979) 62-66

[7] J Kapur, P Sahoo, A Wong "A new method for grey level picture thresholding using the entropy of the histogram" Computer graphics and image processing 29(3) (1985) 273-285

[8] B.J.C, "Pattern recognition with fuzzy objective function algorithms" Plenum press, New York 1981

[9] F Yan, H Zhang, R Kube "A Multistage adaptive thresholding method" , Pattern recognition letters(2005) 1183-1191.

[10] G Li, C Y Kao, J G Gore, Z Ding "Minimization of region scalable fitting energy for image segmentation" IEEE Transactions on Image processing 17(10) (2008) 1940-1949.

[11] J Zhang, Y Chen, X Huang "Edge detection of images based on improved Sobel operator and genetic algorithms", International conference on Image analysis and signal processing ISAP09,China, 2009 pp. 31-35

[12]S Janani, R Nandakumar, M Nirmala "Feature extraction and pairing of G banded Chromosome image using K nearest neighbour classifier" IJCST vol 3, Issue 2 April-June 2012

[13] Schwartzkopf W, Evans B L and Bovik A C "Minimum entropy segmentation applied to multi spectral chromosome images" 2001 Int. Conf. on Image Processing Vol 2,pp 865-8

[14] Mousmi V Munot, Madhuri Joshi, Nikhil Sharma, Gulsheen Ahuja "Automated detection of cut points for disentangling overlapping chromosomes" 2013 IEEE point of care healthcare technologies, Bangalore, India 16-18 January 2013

[15] J Cho, S Y Ryu, S H Woo, "A study for the hierarchical artificial neural network model for Giemsa stained human chromosome classification" Proceedings of the 26th annual international conference of the IEEE EMBS Sept 1-5,2004

[16] Jim piper, Erik Granum, "On fully automatic feature measurement for banded chromosome classification" Cytometry 10:242-255(1989)

[17] Nirmala Madian, K B Jayanthi "Analysis of human chromosome classification using centromere position" Elsevier meas 47(2014) 287-295

[18] Artem khmelinskii, Rodrigo Venture, Joao Sanches, "A novel metric for bone marrow cells chromosome pairing" IEEE Transactions on biomedical engineering, vol 57 no.6,June 2010

[19] Mehdi Moradi, S Kamaledin Setarehdan, "New features for automatic classification of human chromosome: A feasibility study" Pattern recognition letters 27(2006) 19-28

[20] Ozy sjahputera, James M Keller, "Evolution of a fuzzy rule based system for automatic chromosome recognition", 1999 IEEE International fuzzy systems conference proceedings Aug 22-25,1999

[21] Rodrigo Ventura, Artem Khmelinskii, J Miiguel Sanches "Classifier assisted metric for chromosome pairing" IEEE conf on Aug 31-sep 4, 2010

[22] Christoforos Markou, Christos Maramis, Anastasios Delopoulos "Automatic chromosome classification using support vector machines".