A New Approach to Automatic Classification of the Curved Chromosomes

Mehrsan Javan-Roshtkhari
Control and Intelligent processing center of Excellence, ECE, University of Tehran, Tehran, Iran.
m.javan@ece.ut.ac.ir

S. Kamaledin Setarehdan
Control and Intelligent processing center of Excellence, ECE, University of Tehran, Tehran, Iran.
ksetareh@ut.ac.ir

Abstract

In this paper, an effective algorithm for chromosome image processing for straightening the curved chromosomes is presented. This is a very helpful procedure which extends the domain of success of most of the previously reported algorithms to highly curved chromosomes. The procedure is based on the calculation and analyzing the vertical and horizontal projection vectors of the binary image of the chromosome. The binary image is obtained by thresholding the input image after histogram modification. When applied to the real chromosome images the proposed algorithm could straighten all of the highly bent curved chromosomes within the image dataset. To assess the effectiveness of proposed algorithm, a neural network based chromosome classification system is developed. Wavelet transform domain features are extracted and used in an MLP structure for this purpose and a classification rate of 95.3% is obtained.

Keywords: *Chromosome classification, Feature extraction, Projection vectors, Wavelet transform, Artificial neural networks.*

1. Introduction

Karyotyping is a standard tool in medicine. In addition to well known genetic abnormalities like aneuploidy (improper number of chromosomes) or chromosomes with some missing parts, some of the fatal pathological conditions like leucemia are also correlated with chromosome defects [1].

Karyotyping consists of the identification, classification and presentation of the 23 pairs of the chromosomes in a single picture. This process, which is usually done manually by a human expert, is a difficult and time consuming task. In conventional karyotyping, giesma banded cells are photographed under a light microscope (an example picture is shown in figure 1a) during the metaphase stage (one of the four stages of the cell division namely: prophase, metaphase, anaphase and telophase). In metaphase, the chromatin is condensed inside the chromosomes making them to be easily observed with a light microscope. A band is defined as part of the chromosome which is clearly distinguishable from its adjacent segments by its darker or brighter appearance. The chromosomes are visualized as consisting of a continuous series of these bright and dark bands. Each of 23 chromosomes has a specific band pattern.

The result of the karyotyping process for figure 1a, which is done manually by a cytogenetist is shown in figure 1b. Two stages of this process are segmentation and classification of the chromosomes. Automatic classification of chromosomes has been a well studied problem in the last 3 decades [2, 3, 4]. Natural complexity of the problem is caused by various unpredictable appearances of the chromosomes due to non-rigid nature of them.

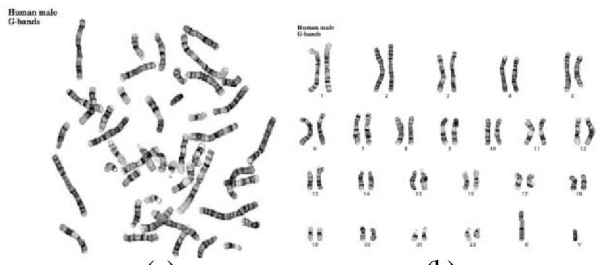
Features used in chromosome classification generally fall into two main categories of the geometrical features and the band pattern based features [5].

Many methods were developed based on the stated features, but most of them suffer a kind of weakness. One of the most common weaknesses is that they cannot work on the bended chromosomes. Some methods can classify bended chromosome if the bending is less than 90 degrees [6].

Few works were done for straightening chromosomes in the past [7]. There are some commercial software packages that offer very comprehensive chromosome analytical tools [2]. Some non-commercial image analysis software packages also provide the ability to straighten the images of curvilinear objects [8]. Two such packages are Image/J [9] and Image SXM [7]. The most recent software, Image SXM, is not fully automatic and needs the user to identify the two ends of the object. Moreover these methods need the central curve of the object (chromosome) as necessary information. The main drawback of these algorithms is the computation cost. For an image of size $n \times n$, the computation complexity for finding central curves is of order $O(n^2)$ up to $O(n^3)$ depending on the method used for calculation [10].

In comparison, the effective algorithm described in this article is computationally very effective. Moreover the results we achieved are among the best in the literature. Also this algorithm is fully automatic.

After straightening the chromosomes, any method can be used for classification. The previously reported approaches used the density profile signal (The density profile, or DP, is a one-dimensional vector obtained by intensity sampling of the chromosome along its longitudinal axis. [5]). In some methods a reduced version of the density profile (hence the feature vector) is developed and used based on such sub-optimal methods as the knockout method [11]. Some other studies have used those features extracted from the weighted density distribution [12], the Fourier transformation or the wavelet transformation [13] of the density profile signal. In our study, we used the wavelet transform to extract useful features from the density profile and a three layer perceptron used as the classifier. A correct classification rate of 95.3% is obtained by applying the ANN based classifier to the features obtained from straightened chromosomes by the proposed algorithm.



(a) (b)

Figure 1. a: G-banded chromosomes as seen under a microscope, b: Karyotype of a male.

The rest of the paper is organized as follows: Section 2 describes the chromosome image data set used in this research. In section 3 the block diagram of the algorithm together with the function of each block is explained. Section 4 describes feature extraction and classification method. Section 5 demonstrates the results of the application of the algorithm to real chromosome images. Finally section 6 concludes the paper.

2. The Dataset

The images used in this study are microscopic chromosome pictures. The images are acquired by conventional photography using a light microscope with a magnification factor of 100X Chromosomes and then scanned by a scanner with a resolution of 300 dpi [6].

3. Processing Method

Figure 2 demonstrates the block diagram of the proposed algorithm for straightening the bended chromosomes. In follows we will describe the function of each block in more details.

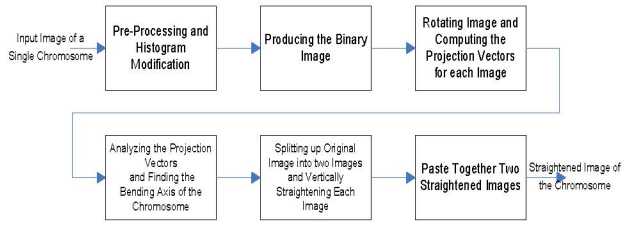


Figure 2. Block diagram of the proposed automatic algorithm.

3.1. Producing the Binary Image

Since we are looking for morphological features of a chromosome, the grayscale image contains additional information that is not needed at this stage. Therefore, producing a binary image will make the rest of the process easier.

The chromosome images are typically bimodal images [14]. It means that an object over a uni-color background forms these images. Histogram of such image usually includes two peaks, one of which corresponds to the background and the other to the object. The preferred threshold separating the object and background-related peaks in the histogram of such an image is usually the gray level representing the global minimum located between the two peaks [6]. By locating the two peaks representing the background and the object in the histogram, the gray level coincident with the global minimum between the two peaks is defined next. Using this value as a threshold, all the pixels with a gray level below this threshold are set to 1 (white) and the remaining pixels are set to 0 (black) producing the binary format of the input image.

3.2. Rotating Image and Computing the Projection Vectors for each Image

In order to calculate the horizontal projection vector, simply the pixel values of each row are summed up in the binary image. Considering that the binary image includes only 1s (white pixels) and 0s (dark pixels), therefore, each element of the horizontal projection vector is equal to the number of white pixels (1s) in the corresponding row.

Similarly, to calculate the vertical projection vector, the pixel values of each column are summed up in the binary image.

Theoretically, these two orthogonal projection vectors contain all the morphological information of the object and can be used for binary image reconstruction and/or for feature extraction. For example they can be used to calculate the extent of the chromosome (i.e. where the chromosome begins and ends on the image

plane), locating the centromere of the chromosome [6], or they can be used for locating bending axis of chromosome, which is explained next.

3.3. Analyzing the Projection Vectors and Determining the Bending Axis of the Chromosome

To find the bending point of the chromosome image, the information in the horizontal projection is just enough. It can be seen from the horizontal projection vector in figure 3 that if the object (chromosome) is vertically straight, it includes two peaks, and the global minimum between them corresponds to the centromere. But if the chromosome is bended 90 degrees or more, the horizontal projection vector can contain one or two peaks depending upon the bending degree, and position of the image. If the horizontal projection vector includes only one global peak, by rotating the chromosome image slightly in both direction (clockwise and counter-clockwise), in a controlled angle steps, another peak will appear gradually in the projection vector. Also, if the horizontal projection vector includes two peaks, by rotating the image of the chromosome the magnitude of each local maximum and local minimum will change. Figure 4 shows a sample chromosome and its projection vector for different degrees of rotations.

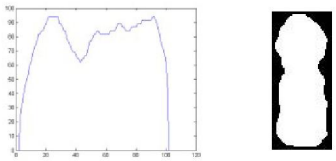


Figure 3. Horizontal projections of a binary image.

By inspecting the set of the rotated images, for any image that the magnitudes of the two peaks in the corresponding projection vector are approximately equal and also the magnitude of the global minimum is very little in comparison with the two peaks, the position of the global minimum corresponds to the bending location of the chromosome in the image.

As discussed above, two parameters are needed to find the proper image from the set of the rotated images, which are the magnitudes of the peaks and the global minimum. Thus these are the parameters that should be taken into consideration in order to identify the right image among the possible candidate images.

In our algorithm, for each rotated image a composite score was calculated as follows:

$$S = w_1 \times R_1 + w_2 \times R_2 \quad (1)$$

Where

$$R_1 = \frac{P_1}{P_2} \quad (2)$$

$$R_2 = \frac{P_3}{P_1 + P_2} \quad (3)$$

and P_1 is the magnitude of the greater peak, P_2 is the

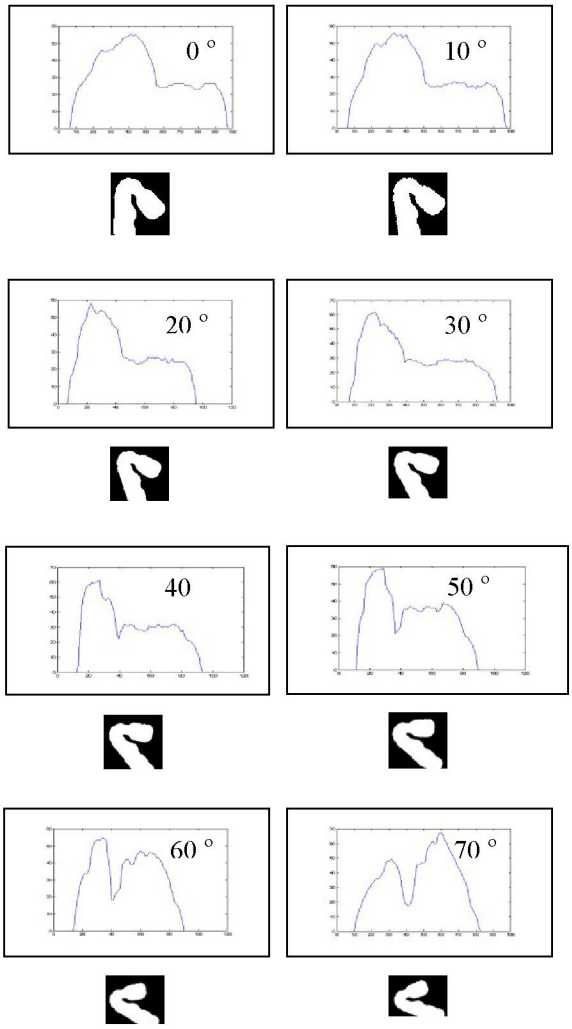


Figure 4. Rotated images and corresponded horizontal projections.

magnitude of the other peak and P_3 is the magnitude of the global minimum. Also w_1 and w_2 are the tuning parameters, where $w_1<1$ and $w_2<1$ and $w_1 + w_2 = 1$.

The suitable rotated image is the one for which the score (S) is minimum. The global minimum in the horizontal projection vector of such image corresponds to the bending point of the chromosome. The central point from the bending axis can be considered as the chromosome's bending point.

Sharp and small peaks can have a misleading effect on the process of finding local maxima and minimum. Therefore, it is necessary to eliminate the unnecessary details of the projection signal using a lowpass filter [15]. A lowpass FIR filter with cutoff frequency $\omega_c=0.15\pi$ and order 20, using Kaiser window was found appropriate for this purpose.

3.3.1. Determination of the tuning parameters

The tuning parameters, w_1, w_2, were selected in an

iterative process based on the closeness of the position of the automatically extracted bending points using a specific set of $\{w_1, w_2\}$ to the reference point that manually identified.

For this propose, 110 highly bent chromosome images were selected and the best set of weights obtained as: w_1=0.43, w_2=0.57.

3.4. Straightening Procedure

Finding the bending point of the chromosome, the image can now be easily straightened. The image is first separated into two sub images from the bending axis; each sub image contains one arm of the chromosome. In each sub image, the object (chromosome's arm) is approximately straight. For obtaining a vertically straightened chromosome, object in each sub image must be vertical. For this purpose, information in the horizontal projection vector is enough. If the object is vertical on the image, the magnitude of greatest peak in the vector is at minimum value.

By the above hypothesis, each sub image must be rotated in controlled angles (30 degrees) in both directions around the bending point, then for each rotated image the value of the greatest peak was calculated. When this value is at its minimum value, the object in the corresponding image is vertical.

According to the above explanations, the two resulting images each contain one of the arms of the chromosome that is vertically straight.

3.5. Applying the Method on Gray-scale Image

At this stage, the information obtained on the binary image (bending point and the rotation angles for each sub image) is now applied to the corresponding greyscale image. Figure 5 shows a sample chromosome with its bending axis identified.

Figure 5. Bending axis of a sample chromosome 17.

3.6. Connecting the Two Straightened Arms

Next, the two resulting grayscale images of the chromosome arms must be connected together to obtain a straightened chromosome picture.

Since the ends of the straightened arms are not flat, when the two pictures are connected together, there will be a gap between them. The values of pixels in the gap between images can be determined by interpolation, but it can change the pattern of dark bands on the chromosome. In order to retain dark bands pattern of the chromosome, the value of those pixels must equal to the mean value of the pixels that are located on the bending axis of the original image. So, the image obtained with proposed algorithm will have the same dark band pattern as the original image. Figure 6 shows a sample highly bent chromosome 17 and its straightened version.

Figure 6. A sample chromosome 17 before (a) and after (b) application of the straightening method.

4. Feature Extraction and Classification

For evaluating the quality of the proposed algorithm, the straightened chromosomes are used in a classification procedure. For this purpose we used the wavelet transform base feature extraction method of [13] and the neural network based classification procedure of [2].

4.1. Feature Extraction with Discrete Wavelet Transform (DWT) from the Density Profile

After straightening each chromosome, the central curve of the chromosome can be approximated by piece wise vertical lines. The density profile (DP) of a chromosome was defined in Section 1. In this work, each sample of the DP signal is defined as the average intensity of the pixels lying on a line perpendicular to the approximation of the central curve of the chromosome. For this purpose, at each pixel belonging to the approximation of the central curve, a perpendicular line was considered and the intensities of the object pixels belonging to this line were averaged to produce one sample of the DP signal (a number in the range of 0–255). Since the approximate central curve is a vertical line, each sample of the DP signal is the average of the pixels that are in the same row of the image.

The DP feature vector was normalized in both length and in value. Normalizing in length yields suitable feature representation (all classified vectors are in the same dimension) and invariance to scale change. The length of the normalized DP vector was set to be 64. The 64 values of the DP vector were normalized into the [0, 1] range.

In this work the pyramid algorithm developed by Mallat [16] was used for wavelet decomposition.

The eight coefficients of the fourth level of the Daubechies wavelet basis (db3) were found to be effective as features.

4.2. Classification Method

Utilization of Artificial Neural Networks for classification of chromosomes has been deeply studied. Lerner [17] has suggested that ANN's are the best chromosome classifiers, especially when the number of classes is limited. In the present study, the number of classes was limited to the three, classes of chromosomes 16, 17 and 18. So ANN was used as the classifiers.

A three layer feed-forward perceptron neural network trained by backpropagation learning rule were used for the classification. To evaluate the separability of the classes with different combinations of features, different architectures were utilized. The number of output nodes was fixed to the number of classes (3), and the number of input nodes was set equal to the dimension of the feature vector. The number of neurons in the hidden layer was the parameter used for the optimization of the structure of the network. A three layer perceptron with 7 neurons in its hidden layer was found suitable for this task.

5. Experimental Results

The proposed automatic bending point locating algorithm is now applied to 175 randomly selected highly bent chromosome images from the data set. For comparison purposes, this point is manually marked on the chromosomes for all images by an expert observer. The automatically defined bending point locations were then compared to those identified manually, and the results are summarized in Table1. In this comparison the Euclidean distance between the two results is considered as the absolute error of the automatic algorithm. Moreover, each error value is normalized with the length of the chromosome for a more meaningful comparison, which is also determined manually. Table 1 demonstrates the mean and the standard deviation of the absolute and normalized errors calculated over the entire data set.

Table 1. Results of the comparison of the automatically defined bending point locations to those identified manually.

Mean value of the absolute error	2.3 (pixel)
Standard deviation of the absolute error	1.6
Mean value of the normalized error	0.017
Standard deviation of the normalized error	0.012

Using the bending axis, the location of the centeromere can also be obtained. Centromere locating is important for feature extraction and classification of the chromosomes, which is a necessary step towards automatic Karyotyping.

As another comparison method, the density profile of the original image (by obtaining the central curve of the chromosome as explained in [5]) and the straightened image are calculated, and the mean square difference between them was calculated. The mean value of the difference was 0.037 with the standard deviation 0.00351 for the 175 selected chromosomes. Figure 7 shows the density profiles of chromosomes represented in figure 6.

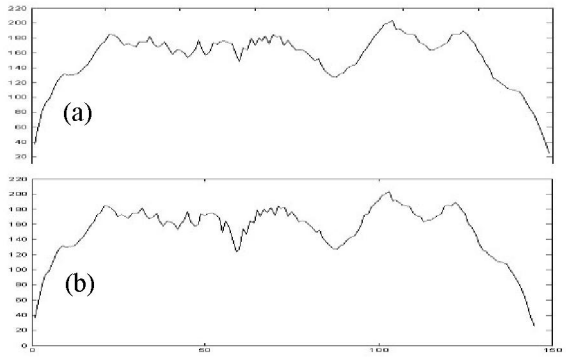


Figure 7. Comparison of the density profiles for the chromosome in Figure 6, before (a) and after (b) application of the straightening algorithm.

For classification, all chromosomes image in the dataset (303 images) were used. First, the straightening algorithm was applied to all images, then the features extracted from the density profile as stated in section 4.

The results of the classification using the three layer perceptron of section 4.2 is demonstrated in table 2.

Table 2. The results of classification

The average of correct classification results over 20 experiments on the training set (%)	100
The average of correct classification results over 20 experiments on the test set (%)	95.3
The standard deviation of the classification results over 20 experiments (%)	3.02
Worst classification result on the test set (%)	91.2
The best classification results on the test set (%)	99.1

The process of training and testing with early stopping strategy (with maximum possible epochs of 150) was conducted 20 times. For each round of experiments, 201 chromosomes were randomly selected to be included in the training set and the remaining 102 chromosomes constituted the testing set. To avoid over-fitting, a subset of the test set, including 30 feature vectors was used as the validation set.

6. Conclusion

Most previously reported automatic algorithms for classification of the human chromosomes are unable to

cope with the bended chromosomes. In this paper a simple, yet very effective algorithm for automatically straightening the microscopic image of the human chromosomes was presented. This algorithm can also find the centromere in the image, which is an important feature of the chromosome for classification of the chromosomes, and Karyotyping.

With the proposed algorithm, all previously reported methods for automatic Karyotyping can now be applied to the straightened chromosomes without any limitation.

The proposed algorithm is based on the calculation and analyzing the vertical and horizontal projection vectors of the binary image of the chromosome obtained by applying a threshold on the input image after histogram modification and analyzing. The algorithm was applied to 175 real highly bended chromosome images and the results were completely satisfactory from the view point of the expert.

Also, it showed that by using the proposed algorithm, the density profile signal can be computed easily and the common methods for feature extraction and classification can be used.

7. References

[1] F. Hong, and L. Mark, *Medical Cytogenetics*, Marcel Dekker, pp. 21-49, 2000.

[2] A. Carothers, and J. Piper, "Computer-Aided Classification of Human Chromosomes: A Review", *J. Statistics and Computing,* Vol. 4, No. 3, 1994, pp. 161-171.

[3] K. R. Castleman and J. H. Melnyk, "Automated System for Chromosome Analysis – Final Report", *JPL Document No. 5040-30,* Jet Propulsion Laboratory, Pasadena, California, 1976.

[4] J. Piper and E. Granum, "On Fully Automatic Feature Measurement for Banded Chromosome Classification," *J. Cytometry,* No. 10, 1989, pp. 242-255.

[5] M. Moradi and S. K. Setarehdan, "New features for automatic classification of human chromosomes: A feasibility study", *Pattern Recognition Letters,* Vol. 27, 2006, pp. 19–28.

[6] M. Moradi, S. K. Setarehdan and S. R. Ghaffari, "Automatic Locating the Centromere on Human Chromosome Pictures", *16th IEEE Symp. on Computer-Based Medical Systems,* 2003, pp. 1063-71.

[7] S. D. Barrett and C. R. de-Carvalho, "A software tool to straighten curved chromosome images", *Chromosome Research*, 2003, No. 11, pp. 83-88.

[8] E. Kocsis, B.L. Trus, C. J. Steer, M. E. Bisher, and A. C. Steven, "Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment", *Struc. Biol.,* No.107, 1991, pp. 6-14.

[9] W. S. Rasband, Image J. National Institutes of Health, Bethesda, Maryland, USA, 2002. Available at: http://rsb.info.nih.gov/ij.

[10] F. J. Jing and S. Sahni, "Serial and Parallel Algorithms for the Medial Axis Transformation", *IEEE Trans. On Pattern Anal. and Mach. Intel.,* Vol. 14, No 12, 1992, pp. 1218-1224.

[11] B. Lerner, H. Guterman, I. Dinstein, and Y. Romem, "Medial axis transform-based features and a neural network for human chromosome classification", *Pattern Recognition,* Vol. 28, No. 11, 1995, pp. 1673-1683.

[12] W. Qiang, and K.R. Castleman, "Automated Chromosome Classifier Using Wavelet-Based Descriptors", 13th *IEEE symp. on computer based medical systems*, 2000, pp. 189-194.

[13] N. Sweeney, R.L. Becker, and B. Sweeney, "A comparison of wavelet and Fourier descriptors for a neural network chromosome classifier", *19th IEEE EMBS Conference,* 1997, pp. 1359-1362.

[14] D. Vernon, "*Machine Vision*", Prentice-Hall, 1991.

[15] A.V. Oppenheim and R. W. Schafer, *"Discrete-Time Signal Processing"*, second edition, Prentice Hall, 1999.

[16] S. Mallat, " A theory for multiscale signal decomposition: the wavelet representation", *IEEE Trans. Pattern Anal. and Mach. Intell.,* Vol. 11, 1989, pp.674-693.

[17] B. Lerner, "Towards a Completely Automatic Neural-Network-Based Human Chromosome Analysis", *IEEE Trans. On Sys., Man. and Cybernetics,* Vol. 28, No. 4, 1998, pp. 544-552.