ISSN: 2456 – 3080
International Journal of Applied and Advanced Scientific Research
Impact Factor 5.255, Special Issue, July - 2017
6thInternational Conference on Innovations in Electronics & Communication Engineering
On 21st& 22nd July 2017 Organized By
Department of ECE, Guru Nanak Institutions, Hyderabad, Telangana

ENERGY MINIMIZATION FOR IDENTIFICATION OF BANDING PATTERN IN CHROMOSOMES USING OPTIMIZED GRAPH CUT ALGORITHM

K. B. Jayanthi*, Nirmala Madian & P. Kiruthika****

* Professor, Department of ECE, K. S. Rangasamy College of Technology, Tiruchengode, Tamilnadu
** Assistant Professor, Department of ECE, K. S. Rangasamy College of Technology, Tiruchengode, Tamilnadu

Cite This Article: K. B. Jayanthi, Nirmala Madian & P. Kiruthika, "Energy Minimization for Identification of Banding Pattern in Chromosomes Using Optimized Graph Cut Algorithm", International Journal of Applied and Advanced Scientific Research, Special Issue, July, Page Number 22-26, 2017

Abstract:

Intensity inhomogeneity is a significant cause in reducing the accuracy of image segmentation. This paper proposes an algorithm for identification of bands in chromosomes using graph cut segmentation that uses global and local image statistics. The global energy is an estimate of the intensity distribution of the image and background and local energy provide the information related with neighboring pixels that eliminates the impact of intensity inhomogeneities. Efficient energy minimization helps in better pixel labeling and this is done by optimized Graph cut process. The shape prior of the band at each location of the image is considered with shape probability energy functions. The experimental results demonstrate that the approach is robust and efficient in detecting the band information in chromosomes to a larger extent.

1. Introduction:

Automated classification of chromosomes has been a topic of interest since 1960. Due to the difficulties in staining techniques and non-rigid nature of chromosomes, it is not possible to give decisive results. Still, research is going on and many algorithms are being proposed to automate the entire process or part of it. Few papers are published for the identification of banding patterns whereas many are working with segmentation of chromosomes and identification of centromere position. Chromosome classification based on local band descriptors is proposed earlier [1]. Here they have used local band description and Weighted Density Distribution (WDD). The former is sensitive to band segmentation whereas the latter is global and is not applied for segmentation of band structures. Few papers have used global features like Fourier, Gaussian and WDD whereas few papers have used structural features for segmentation of banding pattern. WDD is proven to be the best but it fails in the identification of local features like centromere position. The basis function of WDD proves to be better in finding the band pattern descriptors [2]. The basic functions derived from wavelet packet transform are also applied [3]. But the results are only comparable and not much improvement is demonstrated over WDD. WDD which exhibits better results extracts numerical features from band pattern profiles. This needs the exact medial axis which is crucial. Another method uses dominant points of the contour and cubic splines for extracting the features [4]. This claims to reduce the error rate in classification by 0.6%. Pattern vector is also used in the identification of chromatids' banding pattern [5]. This paper proposes only the theory in the process. Classification is done with banding pattern analysis without locating the centromere position using fast search and match algorithm in the chromosome database which provides better results [6].

Many papers are published for extracting the banding features [7]. Local band patterns search and match [8] is less effective in identifying the banding patterns. Banded regions are identified for classification only in abnormal chromosomes [9] which are fewer in number. Chromosomes belonging to class E (16, 17, 18) alone are considered for classification [10] and only straight individual chromosomes are taken for analysis of banding patterns [11]. A simple, effective classification algorithm for the sample of chromosomes obtained from the laboratory is not available. As mentioned few considered only normal chromosomes, whereas few others work only with certain classes of chromosomes. The sample obtained usually has all classes of chromosomes and may have abnormal chromosomes along with normal chromosomes and also the chromosomes may be in different orientations. It is therefore necessary to find algorithms which are successful in all aspects. This paper proposes optimized graph cut for identification of banding pattern which is useful in the classification of chromosomes along with other features like centromere position, length and width of chromosomes. Energy minimization is being recently proposed by many researchers for image segmentation. Graph cut gives good results when applied for image segmentation. Dynamic graph cut based Otsu's method [12] is used to segment the masses in mammogram images. The combinatorial graph cuts help in obtaining global optima and provides good efficiency, numerical robustness and applications related to N-D problems [13]. A graph-cut approach is used for the extraction of the biliary structures in MRCP images [14]. Graph based approach proves to be efficient computational technique for segmentation [15].

Graph cut algorithm is used for segmentation of magnetic resonance images [16], intensity inhomogeneity correction [17], retinex based image inhomogeneity correction for vessel enhancement [18] and also helps in eliminating the energy minimization problems [19]. Segmentation helps in partitioning the image into multiple regions. Image segmentation is used for assigning label to each pixel. This provides better visual characterization of pixels holding the same label. Segmentation using minimal cut in a graph approach guarantees better solution. In human beings, each cell has 23 pairs of chromosomes. These pairs are arranged in an organized profile known as karyotype. Karyotyping is the process of arranging the chromosomes in order for further analysis. The images obtained from the laboratories after staining usually have clusters of chromosomes as shown in Figure 1. It is necessary to karyotype the chromosomes as shown in Figure 2.

ISSN: 2456 – 3080
International Journal of Applied and Advanced Scientific Research
Impact Factor 5.255, Special Issue, July - 2017
6thInternational Conference on Innovations in Electronics & Communication Engineering
On 21st& 22nd July 2017 Organized By
Department of ECE, Guru Nanak Institutions, Hyderabad, Telangana

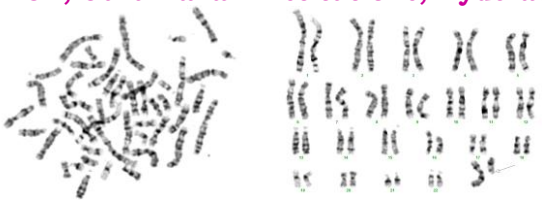

Figure 1: Input Chromosome images

Chromosomes are characterized by dark and bright bands. This banding pattern is unique for each pair of chromosomes. G band chromosomes are taken for analysis as they exhibit better banding pattern. Banding pattern indicates the series of dark and bright regions in the chromosomes. This proposed work uses graph based segmentation for identifying the darker bands in the chromosomes and hence use it for classification and karyotyping.

2. Methodology:

The block diagram for the proposed method is shown in Figure 2. The individual G band chromosome images are taken and intensity corrections are applied using K-means clustering. After this, energy minimization is done using graph cut. The energy minimization across bands identifies the dark bands in each chromosome which is further used for classification.

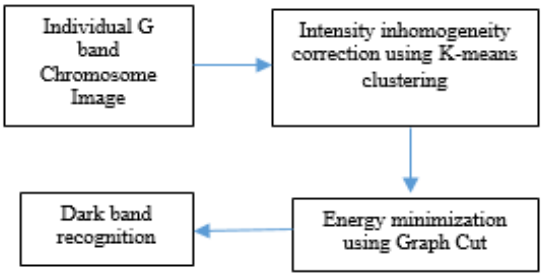

Figure 2: Overall Work Flow

A. Intensity Inhomogeneity Correction Using K-Means Clustering:

K-means [20] is a clustering algorithm that groups similar pixels. The similarity of data set elements in an n- dimensional space is expressed by Euclidean distance. The procedure for K-means is as follows:

- ✓ Choose K different points M_i (i=1, 2,….K) in the data space that serve as a prototype for a cluster C_i.
- ✓ Each data element with the closest prototype is assigned to the cluster.
- ✓ The mean of all elements of C_i for each $M_i$is updated.
- ✓ Steps 2 and 3 are repeated until there is no change in the values of M_i and C_i.

B. Graph Cut Theory: Energy minimization using graph cut helps to get finer details of the image which are usually lost in other methods. This is possible due to correlation of local features. Even though this is a complex process the labeling done for the pixels helps in better segmentation of local features, especially the bands in the chromosomes. Graph based approach is used to extract the region of interest and its minimal cut helps in removing the edge pixels related to other regions. A graph is an abstract representation of a set of objects. The pairwise relationship between objects is given by a graph. Let G be a graph represented as *G= (V,E)* where V is a set of vertices and E is the edges connecting the vertices. Graph cut [15] helps in finding the optimal solution for binary problems and also used in energy minimization problems. The labeling 'f' that minimizes energy is given as

$$E(f) = \sum_{\{p,q\}\in N} u_{\{p,q\}} \cdot T\left(f_p \neq f_q\right) + \sum_{p\in P} D_p(f_p)$$

The first term is the smooth term which ensures the labelling is smooth where *T* equals 1 if the input condition is true and D_p ensures how well f_p matches pixel *p*. The first term corrects the neighboring labels which are too different. The second term is data term and ensures the current label is coherent with the observed data. Energy based methods are used for analysing global properties in images. Many methods involve standard moves where only one pixel can change its label at a time. When the labels change results are significantly affected. Boykov [ ] proposed two different algorithms for finding the approximate solution. They are α-β swap and α- Expansion. Both the algorithms allow a large number of pixels to change their labels simultaneously. Both the algorithms are applied for chromosomes. There is no significant difference in the output between the two algorithms. α-expansion does not allow any pixel which is initially labelled α to change their labels whereas α-β swap allows this change from α-β and vice versa. The algorithms are given below.

α-β Swap for Band Identification:

The algorithm Steps are

- ✓ Initially, labeling is done from the clusters identified from K-means clustering algorithm
- ✓ For each pair of labels, another labeling with minimum energy is found; this is only one α-β swap away from the initial one
- ✓ Here only pixels labeled as α or β are allowed to be swapped whereas all other pixels will retain their initial labels
- ✓ If the energy of the newly labeled pair is less than the earlier one, it is replaced; otherwise the earlier label is retained
- ✓ This process is repeated over all pairs of labels

ISSN: 2456 – 3080
International Journal of Applied and Advanced Scientific Research
Impact Factor 5.255, Special Issue, July - 2017
6thInternational Conference on Innovations in Electronics & Communication Engineering
On 21st& 22nd July 2017 Organized By
Department of ECE, Guru Nanak Institutions, Hyderabad, Telangana

The graphical representation of the α-β swap algorithm is shown in Figure 3.Thestructure of $G_{\alpha\beta}$for any image is given as follows:

- ✓ Select two labels α and β. Consider p in the sets P_α and P_β be the image pixels .i.e. $V_\alpha = \{\alpha, \beta\} \cup P_\alpha \cup P_\beta$
- ✓ Each pixel $p \in P_{\alpha\beta}$is connected to vertices α and β by t-linkst_p^αandt_p^β.
- ✓ Each pair of pixels p, which are neighbours$(\{p,q\} \in N)$are connected by n-links$e_{\{p,q\}}$.

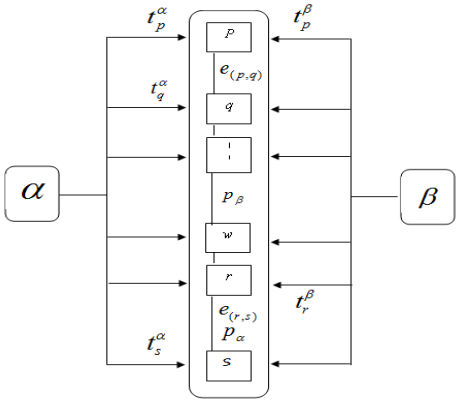

Figure 3: Structure of graph for swap move

- ✓ Edge weights are shown in table I

Table 1: Weights of Edges

Edge	Weight	for
t_p^α	$D_p(\alpha) + \sum_{q \in N_p, q \notin P_{\alpha\beta}} V(\alpha, f_q)$	$p \in P_{\alpha\beta}$
t_p^β	$D_p(\beta) + \sum_{q \in N_p, q \notin P_{\alpha\beta}} V(\alpha, f_q)$	$p \in P_{\alpha\beta}$
$e_{\{p,q\}}$	$V(\alpha, \beta)$	$(\{p,q\} \in N), p, q \in P_{\alpha\beta}$

i.e., t-links with data term and n-links by coherence term are weighted.

For every pixel p, only one of the links can be broken during labeling. IF both the links are broken the pixel will be isolated. The cut is always made such that it does not break the links between the pixels connected to the same terminal; otherwise the cut is made, but again based on the energy function.

3. Results and Discussions:

Individual G band chromosome images as shown in fig.1 are the inputs. G band images are considered as they have unique characteristic recognition for chromosomes. These inputs are subjected to K means clustering. Image clustering helps in grouping the similar pixel intensity values together in the same image. K means has a problem with two neighbouring pixels that have high probability to belong to the same cluster. This issue is solved by considering the distance of each pixel to one of the cluster centroids. This information improves the performance of K means and also helps in solving the problem related with intensity inhomogeneity. The graph shown in the information obtained after K-means clustering is considered as the data for the graph cut algorithm. The advantages of using graph cut method include i) neighbouring pixels belong to the same cluster and ii) pixel is not alone recognized automatically to the closest cluster, but the exact distance of each pixel is considered. The ideogram image is shown in Figure 4. The observation on calculation of number of bands are compared and tabulated and it is shown in Table II.

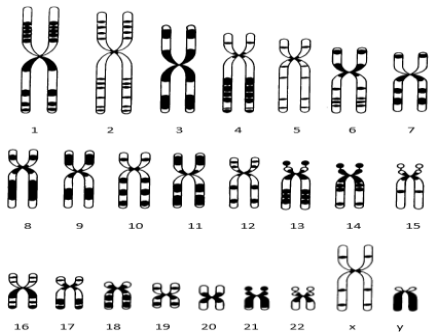

Figure 4 Ideogram of the banding pattern [21]

Table 2: Band Determination in Chromosomes

Chromosome No.(Ideogram image)	No. of dark bands	Chromosome No.(Real-time image)	No. of dark bands
1	9	1	9

ISSN: 2456 – 3080
International Journal of Applied and Advanced Scientific Research
Impact Factor 5.255, Special Issue, July - 2017
6thInternational Conference on Innovations in Electronics & Communication Engineering
On 21st& 22nd July 2017 Organized By
Department of ECE, Guru Nanak Institutions, Hyderabad, Telangana

2	10	2	10
3	4	3	6
4	8	4	7
5	6	5	6
6	6	6	6
7	4	7	6
8	3	8	7
9	3	9	4
10	5	10	5
11	4	11	4
12	5	12	4
13	5	13	5
14	4	14	4
15	2	15	3
16	5	16	3
17	4	17	2
18	3	18	3
19	3	19	1
20	1	20	3
21	1	21	1
22	1	22	2
X	4	x	4
Y	1	y	1

The k means algorithm used here for clustering provides the cluster values ci. These cluster values are used to find the data function. The smoothness term in the energy function is initially assumed to be unitary matrix with normalized values. The data function tries to minimize the energy flow by finding the cut with maximum flow for reducing the disagreement between the observed and assumed values. Iterations are more in minimising the energy function when disagreement is higher. Exact minimization is difficult and a near optimum solution is tried with α-β swap and α-expansion algorithms. But the results indicate that there is not much difference in outputs when the algorithms are implemented. The former one tries to swap positions during iterations whereas the later one does not swap labels but allows to move to a single label 'α'. This is because the clustering chosen in the initial stage is successful and not much of relabeling is required. This provides the initial labelling needed and the number of iterations are also less. Figure 4 (a-f) shows the outputs obtained. In each set the first image is the input chromosome, the second image is the dark band enhanced chromosome and the third image is the graph which exhibits the number of dark bands available in that chromosome.

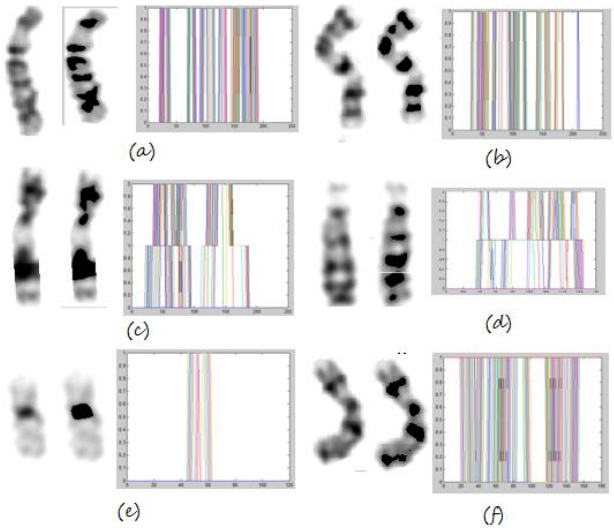

Figure 4 (a-f): Graph cut output image

4. Conclusion and Future Work:

Minimization of energy is a difficult task because the chromosomes have more number of dark bands. Though G-band chromosomes are taken for visibility, difficulties are faced in energy minimization, partly due to energy of dark and light bands. The other problem that is faced is that, same number of bands exist in more than one chromosome pair which makes it difficult for classification. For example chromosomes 3, 7, 11, 14 and 17 have 4 dark bands each but the size and position of bands vary for each chromosomes. The first problem is solved by applying special enhancement for the darker regions. The second one cannot be solved and the process fails to classify the chromosomes with same number of dark bands. The matching of number of bands is approximately around 50%. There are ten pairs of chromosomes with this problem. Area of the dark bands is included as an

additional parameter which is successful in classifying 8 more pairs of chromosomes. Difficulties still remain in the remaining two pairs of chromosomes. If the size and position of the bands are included as additional features it is possible to get betters outputs. Future work is to combine any other energy optimization algorithm with graph cut to solve the issue.

5. References:

1. Frans CA Groen, Ton K ten Kate, Arnold W M Smeulders & Ian T Young, "Human Chromosome Classification Based on Local Band Descriptors", Pattern Recognition Letters, 1989, vol. 9, pp.211-222.
2. Jim Piper & Erik Granum, "On Fully Automatic Feature Measurement for Banded Chromosome Classification", Wiley-Liss, Inc., Cytometry, 1989, vol. 10, pp.242-255.
3. Q. Wu and K. R. Castleman, "Automated chromosome classification using wavelet-based band pattern descriptors," in Proc. 13th Symp. On Computer-Based Medical Systems, Houston, Texas, 2000, pp. 189–194.
4. G. Ritter and G. Schreib, "Profile and feature extraction from chromosomes", in Proc. 15th Int. Conf. on Pattern Recognition, Barcelona, Spain, 2000, vol.2, pp. 2287.
5. Rahmadi Trimananda, 'Chromosome Centromere and Chromatid's Banding Identification Using Pattern Vector', Proceedings of 2nd International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2010, pp.132-134.
6. Jau-hong Kao, Jen-hui Chuang &Tsaipei Wang, "Chromosome classification based on the band profile similarity along approximate medial axis", Pattern Recognition, 2008, vol. 14, no. 1, pp. 77 – 89.
7. Shadab Khan, Joao Sanches, and Rodrigo Ventura, "Robust Band Profile Extraction Using Constrained Nonparametric Machine-Learning Technique", IEEE Transactions on Biomedical Engineering, 2010, Vol. 57, No. 10,pp 2587-2591.
8. Toru Abe, Chieko Hamada, Tetsuo Kinoshita, "Chromosome Image Recognition with Local Band Patterns", MVA2009 IAPR Conference on Machine Vision Applications, 2009, pp. 308-311.
9. Benoît Legrand, Che Sau Chang, Sim-Heng Ong, Soek-Ying Neo, and Nallasivam Palanisamy, "Automated Identification of Chromosome Segments Involved in Translocations by Combining Spectral Karyotyping and Banding Analysis", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 2008, Vol. 38, No. 6, pp. 1374-1384.
10. Mehdi Moradi,S. Kamaledin Setarehdan, "New features for automatic classification of human chromosomes: A feasibility study", Pattern Recognition Letters, 2006,Vol. 27, No.1, pp. 19–28.
11. W. Sethakulvichai , S. Manitpornsut , M. Wiboonrat , W. Lilakiatsakun , A. Assawamakin , S. Tongsima, "Estimation of Band Level Resolutions of Human Chromosome Images", International Joint Conference on Computer Science and Software Engineering (JCSSE), 2012.
12. S. Pitchumani Angayarkanni, Nadira Banu Kamal and Ranjit Jeba Thangaiya, "Dynamic graph cut based segmentation of mammogram", Springer plus, 2015, pp. 1-9.
13. Yuri Boykov and Gareth Funka-lea, "Graph Cuts and Efficient N-D Image Segmentation", International Journal of Computer Vision, 2006, Vol. 70 No. 2, pp. 109–131.
14. Rajasvaran Logeswaran, Dongho Kim, Jungwhan Kim, Keechul Jung and Bundo Song, "Graph-Cut Energy Minimization for Object Extraction in MRCP Medical Images", Journal of Medical Systems, 2012, Vol. 36,No.1, pp.311–320.
15. Meghshyam G. Prasad, "Graph Cuts for Image Segmentation", pp.1-20.
16. Qing Luo, Wenjian Qin , JiaGu, 'Kernel Graph cuts segmentation for MR images with Intensity Inhomogeneity Correction', Applied Mechanics and Materials, 2013, Vols. 333-335, pp. 938-943.
17. Hui Wang and Hong Zhang, 'Adaptive shape prior in graph cut segmentation', Proceedings of 17th IEEE International Conference on Image Processing, 2010.
18. Yitian Zhao, Yonghuai Liu, Xiangqian Wu, Simon P. Harding, and Yalin Zheng, "Retinal Vessel Segmentation: An Efficient Graph Cut Approach with Retinex and Local Phase", 2015, PLoS One, Vol.10, No.4.
19. Richard Szeliski, Ramin Zabih and Daniel Scharstein, "A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, Vol. 30, No. 6, pp. 1068 – 1080.
20. J. Mac Queen, "Some methods for classification and analysis of multivariate observations" in Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1967, pp. 281–297
21. Maximo E. Drets, Margery W. Shaw, "Specific banding patterns of human chromosomes (heterochromatin/ Giemsa stain/ chromosome bands)",1971, Proc. Nat. Acad. Sci.USA, Vol.68, No.9, pp.2073-2077.