SEGMENTATION AND FUZZY-LOGIC CLASSIFICATION OF M-FISH CHROMOSOME IMAGES

Hyohoon Choi∗, Kenneth R. Castleman†, Alan C. Bovik‡

∗ Department of Biomedical Engineering, University of Texas, Austin, Texas
† Advanced Digital Imaging Research, LLC., 2450 South Shore Blvd., Suite 305, League City, Texas
‡ Department of Electrical and Computer Engineering, University of Texas, Austin, Texas
Email: ∗ hyohoon_choi@mail.utexas.edu, † castleman@adires.com, ‡ bovik@ece.utexas.edu

ABSTRACT

Multicolor fluorescence *in-situ* hybridization (M-FISH) technique provides color karyotyping that allows simultaneous analysis of numerical and structural abnormalities of whole human chromosomes. Currently available M-FISH systems exhibit misclassifications of multiple pixel regions that are often larger than the actual chromosomal rearrangement. This paper presents a novel unsupervised classification method based on fuzzy logic classification and a prior adjusted reclassification method. Utilizing the chromosome boundaries, the initial classification results improved significantly after the prior adjusted reclassification while keeping the translocations intact. This paper also presents a new segmentation method that combines both spectral and edge information. Ten M-FISH images from a publicly available database were used to test our methods. The segmentation accuracy was more than 98% on average.

Index Terms— **M-FISH, Fuzzy-logic, Segmentation, Chromosome, Classification, Unsupervised.**

1. INTRODUCTION

Multicolor fluorescence *in-situ* hybridization is a combinatorial labeling technique developed for the analysis of human chromosomes [1], [2]. The technique has been used for the characterization of translocations, to search for cryptic rearrangements, to study mutagenesis, tumors, and radiobiology [3]. In cancerous cells, translocations, exchanges of chromosomal material between chromosomes, are extremely common. To be able to distinguish 24 human chromosomes (22 somatic chromosomes and X and Y sex chromosomes), a minimum of 5 fluorophores are used. Each chromosome is stained with a unique combination of fluorophores so that every chromosome is uniquely identified. An extra fluorophore, DAPI (4'6-diamidino-2-phenyl indole dihydrocloride), is counter stained to all chromosomes. Currently, there are two types of M-FISH imaging systems: a system developed by Speicher *et al.* [4], so called 'multiplex fluorescence in situ hybridization' (M-FISH), which uses a set of optical bandpass filters, and a system developed by

Schröck *et al.* [2], so called 'spectral karyotyping' (SKY), which uses an interferometer. Both systems provide color karyotyping (visualization of chromosomes in a specific format) by assigning a pseudocolor to each pixel based on the spectral information.

Currently available M-FISH systems still exhibit misclassifications of multiple pixel regions due to a number of factors including non-homogeneity of staining, variations of intensity levels within and between image sets, and emission spectra overlaps between fluorophores. The size of the misclassified regions are often larger than the actual chromosomal rearrangement. Thus a careful examination of the classification results by a well trained human operator must be performed. Furthermore, segmentation of chromosomes that are overlapping and touching each other are performed manually in most cases. To reliably automate the process, accurate segmentation and classification must be achieved.

In this paper, we present a new segmentation method between chromosomes and background and a novel unsupervised classification method based on a fuzzy logic classifier specifically designed for M-FISH images. After the segmentation and pixel classification, pixels are reclassified by adjusting *a priori*. Detailed explanations are described in Section 2. In Section 3, segmentation and classification results are shown along with example images. We conclude the paper in Section 4.

2. METHODS

2.1. Foreground-background segmentation

M-FISH images have six channels. Each channel contains the intensity of a corresponding fluorophore. Since each chromosome is uniquely stained, an intensity combination across 6 channels is unique for each chromosome. Previously, we have designed a 6-channel 25-class maximum likelihood classifier [5], [6]. 25 classes include 24 chromosomes plus background. By classifying every pixel in the image using this maximum-likelihood classifier, both segmentation and classification of chromosomes were achieved simultaneously. The overall accuracy of the segmentation was relatively high (about 90 %) using this method. When a portion

of the chromosome pixels are classified as background or vice versa, the lost region cannot be recovered without prior knowledge of the chromsome boundaries. Thus, prior to the pixel classification, an accurate segmentation method is desired.

In order to compute a reliable boundaries between objects and background, we combined multiple methods that utilize not only spectral information but also edge information. Laplacian of Gaussian (LoG) edge detection on the DAPI channel provides nice closed boundaries of chromosomes that correspond well to human perception. However, it also picks up unwanted artifacts from the background. In general, chromosome intensities are brighter than the neighboring background, although the background surface is not globally uniform. When object intensity is brighter than the neighboring pixels, adaptive thresholding is an effective segmentation method. This method effectively separates chromosomes from background. Due to its simplicity and effectiveness, adaptive thresholding is widely used for chromosome image segmentation. However, when a number of pixels in the foreground are darker than neighboring foreground pixels, adaptive thresholding creates holes inside the chromosome. To utilize the spectral information, 6-feature 2-class K-means clustering method is used. This clustering method is preferable to the maximum-likelihood method because it does not require training. It groups six dimensional data into two classes while iteratively regrouping the data points until the class means converge. Its classification results are similar to those of the maximum-likelihood classifier since they both utilize the same information. Adaptive thresholding, LoG edge detection, K-means clustering, and a global thresholding methods are combined together to achieve a final segmentation result. A composite threshold image is obtained after voting among those 4 methods. For example, a pixel becomes foreground when a majority (3 out of 4) are foreground.

2.2. Fuzzy-logic classifier

Supervised, parametric classification methods require training of classifiers. Training can be accomplished only when samples are labeled. Labeling samples can be extremely time consuming. A fuzzy-logic classifier is an unsupervised classification method that does not need to assume the underlying distribution, nor does it estimate the distribution. Furthermore, the computational complexity is far less (at least 10 times) than that of the maximum-likelihood classifier while the classification accuracy is comparable. It only requires information regarding spectral combinations for each class (e.g. Table I).

The discriminant functions of the fuzzy logic classifier are formulated as follows:

$$g_i(\mathbf{x}) = \prod_{j=1}^{6} f(x_j) P(\omega_i) \qquad (1)$$

Chromosome	Spectrum					
	DAPI	Aqua	Green	Gold	Red	Far Red
1	x			x		
2	x				x	
3	x	x				
4	x		x		x	
5	x			x		x
⋮						

Table I. Color table. Chromosome labeling chart for Vysis M-FISH probe. Chromosome 1 is stained with DAPI and Gold.

where i is the class index ($i = 1 \sim 24$), and j is the spectrum index ($j = 1 \sim 6$), $P(\omega_i)$ is the *a priori* probability for class i, and $\mathbf{x}$ is a sample vector.

$$f(x_j) = \begin{cases} x_j & \text{if } T(i,j) = 1 \\ 1 - x_j & \text{if } T(i,j) = 0 \end{cases} \qquad (2)$$

where T is the color table (e.g. Table I). For example, the discriminant function for class 1 will be (assuming equal priors)

$$g_1(\mathbf{x}) = x_1 \times (1 - x_2) \times (1 - x_3) \times x_4 \times (1 - x_5) \times (1 - x_6) \qquad (3)$$

A pixel $\mathbf{x}$ belongs to class ω_i if $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for all $j \neq i$. Only pixels inside foreground are classified using this classifier.

2.3. Prior adjusted reclassification

After the classification, overlapping and touching chromosomes can be separated by analyzing the classification result [7]. Currently we are developing algorithms for the automatic separation of overlapping and touching chromosomes. In commercial systems, chromosomes are segmented using only the DAPI channel prior to classification, and most of overlapping and touching chromosomes are separated manually. Thus in these systems the boundary information is given before the classification. The boundary information is extremely useful when correcting misclassifications. Misclassifications usually occur where chromosomes touch or overlap and near the boundaries of chromosomes. Here we introduce a method that eliminates misclassifications effectively while preserving the translocations intact when the boundary information is available. For the current status of our research, the boundary information is obtained from the ground truth.

In M-FISH data we have observed that when a pixel $\mathbf{x}_1$ belongs to ω_1 but misclassified as ω_2, the posterior probability difference is small, $P(\omega_2|\mathbf{x}_1) > P(\omega_1|\mathbf{x}_1)$ and $P(\omega_2|\mathbf{x}_1) - P(\omega_1|\mathbf{x}_1) = \epsilon$. The posterior probability is derived from Bayes rule as follows:

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i) P(\omega_i)}{p(\mathbf{x})}$$

In this case, $\mathbf{x}_1$ could be easily reclassified as ω_1 by a small increase in the prior for ω_1. However, when $\mathbf{x}_1$ truly belongs to ω_2, the posterior probability difference is large: $P(\omega_2|\mathbf{x}_1) >> P(\omega_1|\mathbf{x}_1)$, and a small increase in the prior for ω_1 would not change the classification result. Therefore, the developed method effectively corrects misclassifications while preserving the translocations. Of course, for this method to work, the right class to increase the prior must be determined for a given boundary.

A set of pixels that belongs to a boundary B_i is defined as S_i. S_i may contain pixels that belong to multiple classes due to misclassifications and a translocation. Given B_i, there exists the most likely class ω_m among $\{\omega_1, \ldots, \omega_{24}\}$ that S_i belongs to. Given B_i, m is found by the following formula:

$$m = \arg\max_m \left\{ P_s(\omega_m|\mathbf{s}) \sum_{i=1}^{24} P_p(\omega_m|\mathbf{x}_i) P_N(\omega_m) \right\} \quad (4)$$

where $P_s(\omega_m|\mathbf{s})$ is the posterior probability given $\mathbf{s}$, $\mathbf{s}$ is the normalized size of B_i, $P_p(\omega_m|\mathbf{x}_i)$ is the posterior probability given a vector that belongs to class ω_i, and $P_N(\omega_m)$ is the normalized number of pixels that belong to ω_m. Three factors are considered in determining the most likely class: the chromosome size, the sum of *a posteriori* probabilities for each class, and the class population. These three factors are effectively incorporated in order to compensate for errors. Once ω_m is found, all pixels in B_i are reclassified with a higher prior for ω_m.

3. RESULTS

We tested our algorithms on 10 images from Advanced Digital Imaging Research's M-FISH image database, which contains 200 hand-segmented M-FISH images. The database is available at `http://www.adires.com/05/Project/MFISH_DB/MFISH_DB.shtml`.

Foreground and background segmentation was performed using the voting method as explained in Section 2.1. Prior to the segmentation, a non-uniform background was corrected by fitting a cubic surface to the estimated background pixels and subtracting it from each channel [8]. The background pixels for each channel were estimated by a global thresholding method, an iterative clustering method, in which the threshold was found while iteratively grouping pixels into two classes until the class means converge. The decision boundary between the two classes was the threshold. To safely exclude chromosome pixels from the surface estimation, dilation was performed on the thresholded image. Pixels below threshold were used for the surface estimation. After the background correction, cells are identified based on the circularity and size measures. Once the background was corrected and cells are removed from the image, adaptive thresholding, LoG edge detection, and 6-feature 2-class K-means clustering were performed. A composite threshold image was created after voting. An example is shown

Images	Fuzzy	ML
v1303xy	89.5	90.7
v1304xy	86.0	87.2
v1305xy	94.6	.
v1306xy	91.1	.
v1308xy	91.0	.
v1310xy	83.4	87.4
v1311xy	92.4	90.1
v1312xy	89.0	.
v1313xy	88.9	93.3
Ave	88.0	89.8

Table II. Classification rates

in Fig. 1. Fig. 1 (b) was generated by thresholding the DAPI channel and by manually correcting mistakes. During the manual correction, some chromosomes were mistakenly drawn larger than their proper sizes such as the chromosome indicated by an arrow in Fig. 1 (b). Fig. 1 (c) agrees well with human perception. The segmentation accuracy was also quantitatively measured by comparing with the ground truth. Among 10 images, the lowest and the highest correct rates were 97.5% and 98.7%, and the average was 98.2%.

After the segmentation, pixels inside the chromosome region were classified using the Fuzzy-logic classifier. The initial classification rates are shown in Table II. Its classification accuracy is comparable to that of the maximum-likelihood classifier as shown in Table II. For the maximum-likelihood classification, 5 images were used as training. Since the fuzzy-logic classification does not require training, all 10 images were classified.

Reclassification was performed by adjusting the prior for each chromosome. The most likely class for a given chromosome was computed using eq. (4). It always found the right class when the chromosome segmentation was done correctly. Incorrect segmentations of chromosomes were found in the ground truth. Those chromosome were usually segmented larger than their original sizes. A wrong class was computed when the chromosome size was significantly larger than normal. In such cases, the size probability was so high for the large chromosome that the size pdf won over the other two terms in eq. (4). In images without incorrect segmentation, the classification accuracy increased (as shown in Fig. 1) as the prior increased. As the prior approaches unity for a class, all the pixels inside the chromosome will become one class. The misclassified pixels were corrected first, then the translocation was removed as the prior increased. Therefore, there is a limit on how much the prior can increase to achieve improved accuracy while preserving translocations.

4. CONCLUSION

We introduced a novel unsupervised classification method for M-FISH chromosome images and a prior adjusted reclassification method that corrects misclassifications effectively while keeping translocations intact. The accuracy of the
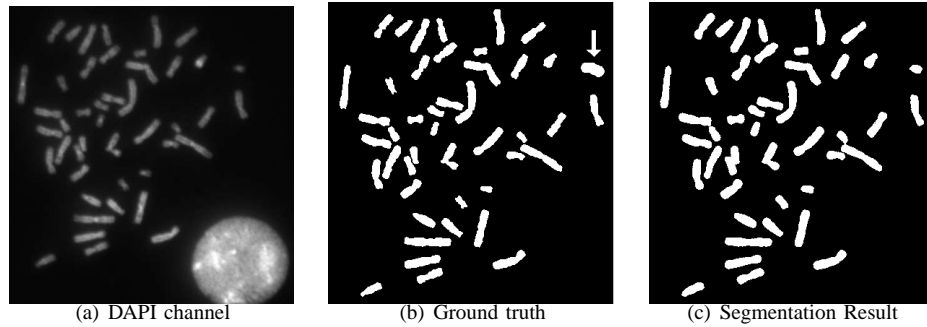
(a) DAPI channel   (b) Ground truth   (c) Segmentation Result

Fig. 1. Segmentation result

fuzzy-logic classifier is comparable to that of maximum-likelihood classifier. We also introduced a novel segmentation method that utilizes both spectral and edge information. High segmentation accuracy was obtained using our segmentation method.

5. REFERENCES

[1] M. R. Speicher, S. G. Ballard, and D. C. Ward, "Karyotyping human chromosomes by combinatorial multi-fluor fish," *Nature Genetics*, vol. 12, pp. 368–375, 1996.

[2] E. Schrock, S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, M. A. Ferguson-Smith, Y. Ning, D. H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, and T. Ried, "Multicolor spectral karyotyping of human chromosomes," *Science*, vol. 273, pp. 494–497, 1996.

[3] T. Liehr and U. Claussen, "Multicolor-fish approaches for the charaterization of human chromosomes in clinical genetics and tumor cytogenetics," *Current Genomics*, vol. 3, pp. 213–235, 2002.

[4] M. R. Speicher, S. G. Ballard, and D. C. Ward, "Computer image analysis of combinatorial multi-fluor fish," *Bioimaging*, vol. 4, pp. 52–64, 1996.

[5] H. Choi, K. R. Castleman, and A. C. Bovik, "Joint segmentation and classification of m-fish chromosome images," *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, 2004.

[6] M. P. Sampat, K. R. Castleman, and A. C. Bovik, "Pixel-by-pixel classification of m-fish images," *Proceedings of the Second Joint Conference of the IEEE EMBS and BMES Conference*, vol. 2, pp. 999 – 1000, Oct 2002.

[7] W. Schwartzkopf, A. Bovik, and B. Evans, "Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1593–1610, 2005.

[8] K. R. Castleman, *Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1996.

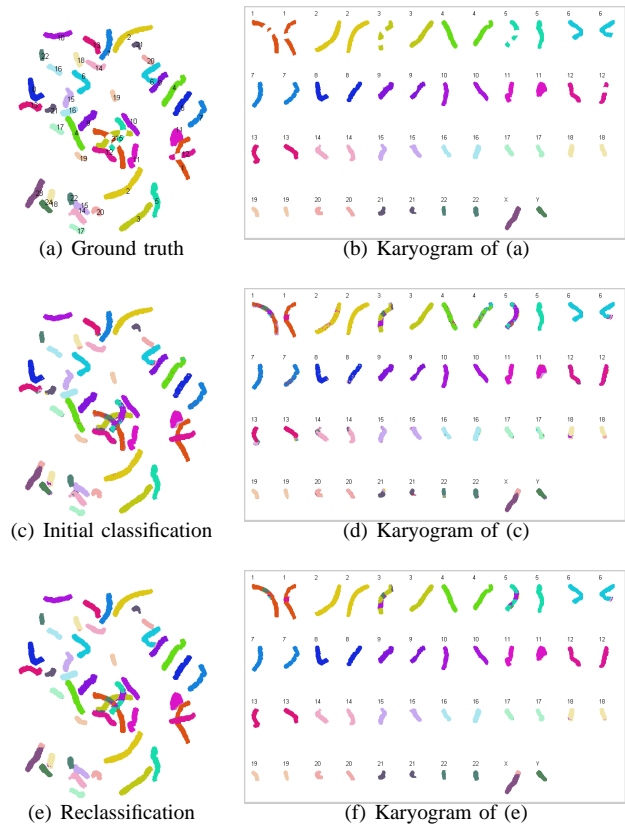(a) Ground truth   (b) Karyogram of (a)

(c) Initial classification   (d) Karyogram of (c)

(e) Reclassification   (f) Karyogram of (e)

Fig. 2. Fuzzy logic classification and prior adjusted reclassification