Feature Normalization via Expectation Maximization and Unsupervised Nonparametric Classification For M-FISH Chromosome Images

Hyohoon Choi*, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, and Kenneth R. Castleman, *Member, IEEE*

Abstract—Multicolor fluorescence *in situ* hybridization (M-FISH) techniques provide color karyotyping that allows simultaneous analysis of numerical and structural abnormalities of whole human chromosomes. Chromosomes are stained combinatorially in M-FISH. By analyzing the intensity combinations of each pixel, all chromosome pixels in an image are classified. Often, the intensity distributions between different images are found to be considerably different and the difference becomes the source of misclassifications of the pixels. Improved pixel classification accuracy is the most important task to ensure the success of the M-FISH technique. In this paper, we introduce a new feature normalization method for M-FISH images that reduces the difference in the feature distributions among different images using the expectation maximization (EM) algorithm. We also introduce a new unsupervised, nonparametric classification method for M-FISH images. The performance of the classifier is as accurate as the maximum-likelihood classifier, whose accuracy also significantly improved after the EM normalization. We would expect that any classifier will likely produce an improved classification accuracy following the EM normalization. Since the developed classification method does not require training data, it is highly convenient when ground truth does not exist. A significant improvement was achieved on the pixel classification accuracy after the new feature normalization. Indeed, the overall pixel classification accuracy improved by 20% after EM normalization.

Index Terms—Chromosome, classification, expectation maximization (EM), maximum-likelihood, multicolor fluorescence *in situ* hybridization (M-FISH), normalization, unsupervised.

I. INTRODUCTION

THE FLUORESCENCE *in situ* hybridization (FISH) microscopic imaging modality has been widely used for the analysis of genes and chromosomes. Multiple fluorophores are often used combinatorially to visualize several biological target types simultaneously. Using combinatorial labeling methods, $2^n - 1$ specimens can be discriminated using n fluorophores. When three fluorophores are used, seven specimens can be an-

alyzed by the binary combinations (presence or absence) of the fluorophores. n gray scale images of specimens, stained with n fluorophores, can be obtained using a monochrome camera and a set of optical bandpass filters that are specifically designed for the excitation and emission wavelengths of the fluorophores [1].

In particular, multicolor (multiplex) fluorescence *in situ* hybridization, so called M-FISH, uses minimum five fluorophores to uniquely identify all 24 chromosome types of the human genome. A sixth fluorophore, DAPI (4'-6-di-amidino-2-phenylindole, a blue fluorescent dye), is used to counterstain the chromosomes [1], [2]. Thus, each pixel of an M-FISH image is typically composed of six values that correspond to the intensities of six fluorophores. Fig. 1 shows an example of M-FISH images. By analyzing the combinations of the six spectral intensities, all of the chromosome pixels in an image are identified, and a pseudocolor is assigned based on the class the pixel belongs to [3], [4]. After the pixel classification, chromosomes are displayed according to a standard format.

The M-FISH technique has been used for the characterization of translocations, to search for cryptic rearrangements, to study mutagenesis, tumors, and radiobiology [5]. In cancerous cells, translocations, or exchanges of chromosomal material between chromosomes, are extremely common.

Currently available M-FISH systems still exhibit misclassifications of multiple pixel regions due to a number of factors, including nonhomogeneity 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. To reliably detect subtle and cryptic chromosomal aberrations, a highly accurate pixel classification method has to be developed. Along with a reliable pixel classification method, automation of karyotyping process is another important goal. The automation requires segmentation of chromosomes, which not only involves object/background separation but also involves separating touching and overlapping chromosomes. While automating the segmentation of partially occluded chromosomes is an extremely challenging problem, a pixel classification method that satisfies both high accuracy and minimum human intervention has not been realized.

In order to achieve a high accuracy in pattern recognition, selection or extraction of good features is the most important stage of processing. Different classifiers may produce different accuracies, but the accuracy is fundamentally bounded by the sample distribution in the feature space. Thus, feature normalization is also a crucial part of classification after feature se-

*H. Choi is with Sealed Air Corporation, 2033 Gateway Place, San Jose, CA 95110 USA (e-mail: hyohoon@alumni.utexas.net).

A. C. Bovik is with the Laboratory of Image and Video Engineering, The University of Texas-Austin, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

K. R. Castleman was with the Advanced Digital Imaging Research, League City, TX 77573 USA (e-mail: ken@castleman.org).
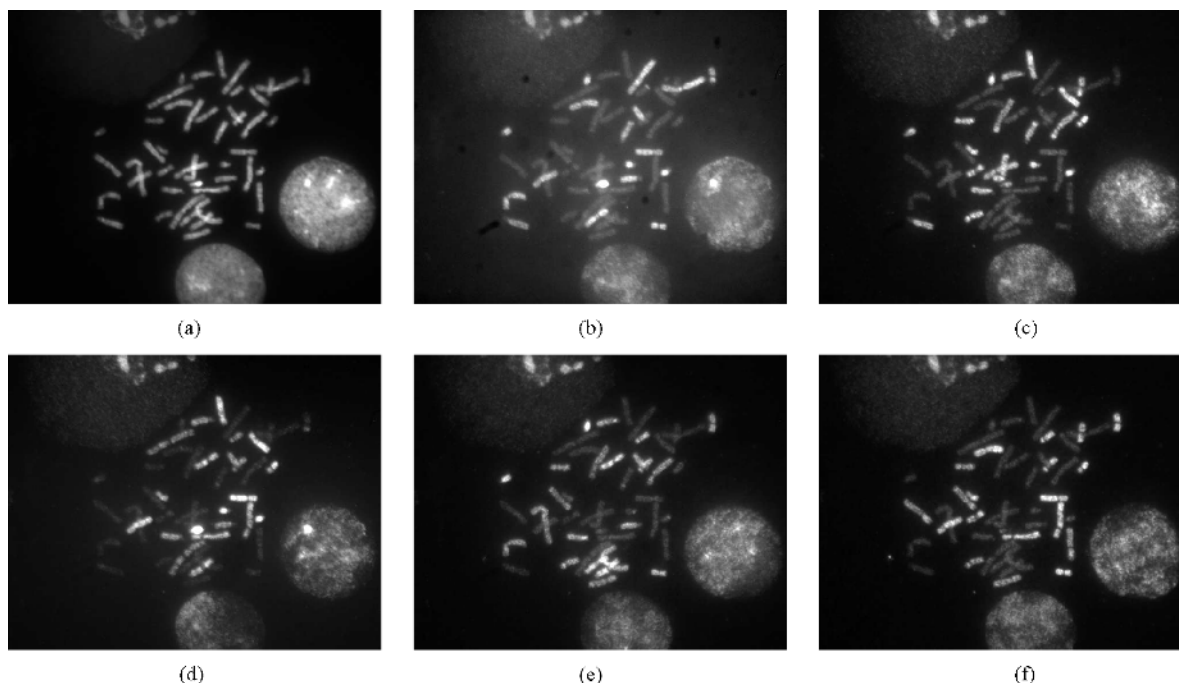
Fig. 1. An M-FISH image. Chromosomes are combinatorially labeled using five fluorophores and counterstained using DAPI. Each gray scale image corresponds to the sum of intensities of the emission wavelengths (a narrow range of wavelengths) of each fluorophore. (a) DAPI. (b) Aqua. (c) Green. (d) Gold. (e) Red. (f) Far Red.

lection. In particular, when features are obtained independently, the normalization must be performed in order to reduce the intravariance of the feature distribution among different images. In M-FISH, each channel is captured independently, and each channel has a different integration time due to different signal strengths of fluorophores. As the relative intensity values across the six channels are used as features, intensity variations should be normalized prior to pixel classification.

In this paper, we present a new normalization method for M-FISH images using the expectation maximization algorithm. The developed normalization method significantly increases the pixel classification accuracy for any classifier. We also present a new classification method for M-FISH images that does not require training of a classifier (unsupervised) nor does it require class parameter estimation (nonparametric).

II. NORMALIZATION OF M-FISH IMAGES

A. Motivation

In M-FISH, six fluorophores are combinatorially used to discriminate 24 chromosome types. For those who are not familiar with M-FISH, the color map of Vysis probe sets are shown in Table I. According to the color map, chromosome 1, for example, is stained with DAPI and spectrum Gold dyes. Ideally chromosome 1 should be observed only in the DAPI and Gold channels and should not be visible in other channels. However, due to the overlap of excitation and emission spectra and the broad sensitivity of image sensors, the obtained images contain a certain amount of crosstalk between the color channels. This phenomenon is called color spread [6]. Thus, all chromosomes are visible on all channels with different intensity levels (see Fig. 1). Furthermore, each fluorophore has a different sensitivity to the excitation wavelength. Thus, some fluorophores

TABLE I
CHROMOSOME LABELING CHART OF VYSIS M-FISH PROBE

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
6	x		x			
7	x					x
8	x				x	x
9	x			x	x	
10	x	x		x		x
11	x	x			x	
12	x		x	x		
13	x	x	x			
14	x		x	x	x	
15	x	x		x	x	
16	x		x			x
17	x		x		x	x
18	x			x	x	x
19	x		x	x		x
20	x	x			x	x
21	x	x	x	x		
22	x	x	x		x	
X	x	x				x
Y	x	x		x		

require a short integration time while others require a long exposure time. Especially Aqua and Gold require long exposure times in order to visualize the hybridized chromosomes. An example of integration times is [DAPI, Aqua, Green, Gold, Red, Far Red]$= [0.14, 6, 0.76, 6, 2.96, 1.4]$ seconds. When a pixel belongs to chromosome 1, the obtained intensity values are expected to have a pattern of [high, low, low, high, low, low] for Vysis probes. Unfortunately, this pattern can be easily broken when each channel is independently acquired. A long exposure time amplifies the leaked intensity, and in some cases it

TABLE II
PIXEL VALUES OF CHROMOSOME 1. EVEN THOUGH CHROMOSOME 1
IS STAINED WITH DAPI AND GOLD, THERE IS NO OBVIOUS PATTERN
IN FEATURE VALUES BECAUSE OF CHANNEL CROSSTALK AND
INDEPENDENT INTEGRATION TIME PER CHANNEL

Images	Location (x,y)	Spectrum					
		DAPI	Aqua	Green	Gold	Red	Far Red
v1301xy	243,172	79	75	56	79	52	51
v1312xy	352,194	54	75	54	101	48	50
v1401xy	251,314	176	75	60	44	50	27

can be higher than the chromosome intensities on other channels at the same pixel location. The different dc offset levels of each channel of the imaging device [e.g., three channel (color charge-coupled device (CCD)] and nonflat background elevation also bias the signal intensity upward. Furthermore, chromosomes appearing in one spectral channel exhibit different intensity levels: some are darker or brighter than others, partially because of the nonflat background, but more substantially because of the different fluorophore sensitivities for different chromosomes. Examples of real pixel values of chromosome 1 across multiple images are shown in Table II. As shown in the table, there is no obvious pattern in the feature values for the preceding reasons.

When the variation of the feature distribution across images is significant, which means the feature distribution of an unknown image is unpredictable, classification methods that rely on the estimation of class parameters will yield low accuracy.

As long as k classes are grouped separately in the feature space even if the feature distribution differs from image to image, pixels can be accurately classified without estimating class parameters using unsupervised-nonparametric clustering methods such as k-means clustering or fuzzy k-means clustering. However, when the number of classes is not fixed (e.g., chromosome images: the number of chromosome classes differs by gender or diseases), finding the right number of classes after the clustering by cluster validation adds complexity and may cause inaccuracy. Pixels can be clustered into a maximum number of classes (24 clusters for M-FISH data) using these methods, and clusters that are closer than a threshold should be merged. The threshold will be again data dependent, which will be different for different images.

Therefore, regardless of the choice of classifiers the variations of the feature distribution should be minimized in order to obtain overall high accuracy in pixel classification.

B. Background Correction and Color Compensation

This section briefly describes how each preprocessing step improves the image quality and the quality of pattern (feature vector). Castleman first introduced a signal model for FISH images, and showed that the true signal can be recovered based on the model [7]. Microscope images of biological specimens often contain nonflat background surfaces. The removal of the nonflat background surface is called background correction, and is commonly performed as a preprocessing step. The removal of the channel crosstalk is called color compensation.

The observed signal at a pixel $\mathbf{y}$ is modeled as

$$\mathbf{y} = \mathbf{E}\{\mathbf{Cx} + \mathbf{b}\} + \mathbf{n} \tag{1}$$

where, $\mathbf{x}$ is the 6×1 vector of the true signal, $\mathbf{C}$ is the 6×6 color spread matrix, $\mathbf{b}$ includes the dc-offset of the CCD and various factors that cause background intensity elevation, $\mathbf{n}$ is the noise of the imaging device such as white noise and shot noise, and $\mathbf{E}$ is the 6×6 diagonal matrix of exposure times. This model assumes that the gray levels are linear with brightness of the fluorophores.

The noise term $\mathbf{n}$ can be minimized by median filtering and lowpass filtering with a 3×3 kernel for both operations.

The nonflat background surface can be approximated by a 2-D cubic surface. The surface that has the minimum mean square error relative to the background pixels is the estimated 2-D cubic surface [6]. By subtracting the surface from the image (six surfaces are estimated for six channels), $\mathbf{b}$ is removed.

After the background correction, the signal model becomes

$$\mathbf{y} = \mathbf{ECx}. \tag{2}$$

Once the color spread matrix $\mathbf{C}$ is found from a labeled image or images, it can be applied to other images to correct color spreadings (the details of estimating $\mathbf{C}$ from the labeled images are out of the scope for this paper). An image I is a set of pixels of $\mathbf{y}_j$, where $j = (1, \ldots, N_p)$ and N_p is the number of pixels in an image. An image without the channel crosstalk is computed by $\mathbf{x}_j = \mathbf{C}^{-1}\mathbf{E}^{-1}\mathbf{y}_j$ for all j. To account for the fluorophore sensitivities, exposure times $\mathbf{E}$ can be multiplied by $\mathbf{x}_j$.

Color compensation is an effective method of improving the quality of M-FISH images by removing the channel crosstalk. Fig. 2 shows an example of before and after the color compensation. It is not easy to distinguish which chromosomes are truly hybridized and which are due to crosstalk in Fig. 2(b). In particular, chromosomes marked with number 1 and 2 are due to crosstalk and they are effectively removed in Fig. 2(f).

Table III shows pixel values after the background correction and color compensation. As the values show, the intensity corresponding to the channel crosstalk has been removed effectively. The background correction helps reveal the pattern and color compensation further enhances the pattern, as shown in Fig. 2 and in Table III. Accordingly, pixel classification accuracy also increased significantly after the background correction (results are shown in Section IV). However, our experiments on a small number of images showed that color compensating images after the background correction did not improve the overall classification accuracy, which contradicts our expectation. This suggests that revealing the pattern helps classification but enhancing the pattern is not enough. The pattern must satisfy a certain criteria, which is explained in the following section. The color compensation improves the image quality significantly but may not be a necessary preprocessing step for pixel classification.

C. Expectation Maximization Normalization

Even after background correction and color compensation, intensity variations within a chromosome and among chromosomes in a channel and between channels, caused by uneven hybridization in a chromosome and unequal fluorophore sensitivities depending on chromosomes, remain as a source of classification error. Given a channel, chromosomes that are supposed to be bright in that channel are expected to have a similar intensity
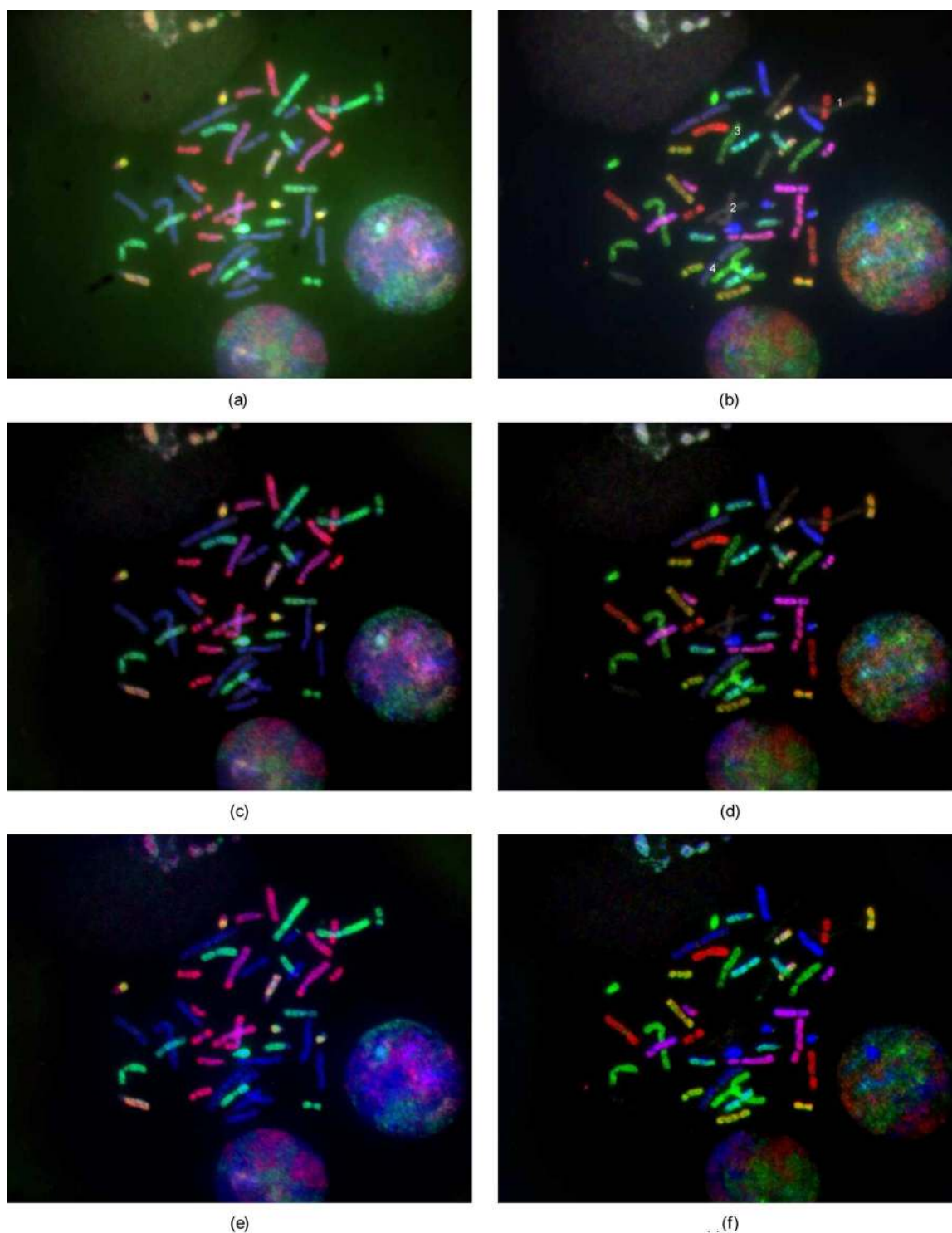
Fig. 2. Color compensation result on image V1301XY. (a) Green, Aqua, and DAPI channels of V1301XY are combined as a color image. (b) Far red, Red, and Gold channels are combined and shown as a color image. (c) and (d) Result of background correction. (e) and (f) Color compensation result (simple scaling has been applied to the color compensated image). As shown in (e) and (f) the quality of the image has been improved significantly by removing channel crosstalk. Pixel values numbered on (b) are shown in Table III. (a) Before color compensation. (b) Before color compensation. (c) Background correction of (a). (d) Background correction of (b). (e) Color compensation of (a). (f) Color compensation of (b).

level among them, but often chromosome intensities considerably differ as some are much brighter than others. Those bright chromosomes in one channel are not consistently brighter than other chromosome in other channels where they are supposed to appear. Often a chromosome with a certain intensity level on one channel appears on another channel with a significantly

lower or higher intensity. This inconsistency causes classification errors since the pattern, more specifically the texture, of a feature vector $\mathbf{y}$ becomes inconsistent. Given an individual feature value, e.g., gray scale of 60, it is uncertain whether it comes from a hybridized chromosome or from noise. Only when a feature vector is formed does the relative intensity difference

TABLE III
PIXEL VALUES NUMBERED ON FIG. 2(B). NP, BC, AND CC
MEANS NO PROCESSING, BACKGROUND CORRECTION, AND
COLOR COMPENSATION, RESPECTIVELY

Pixel	Processing	Aqua	Green	Gold	Red	Far Red	Class
				Spectrum			
1	NP	191	57	38	44	55	3
	BC	133	26	11	17	34	
	CC	196	1	0	0	11	
	Pattern	1	0	0	0	0	
2	NP	137	219	75	84	85	6
	BC	61	173	32	43	50	
	CC	18	208	1	27	16	
	Pattern	0	1	0	0	0	
3	NP	95	154	60	128	53	4
	BC	30	118	27	97	28	
	CC	0	151	0	163	0	
	Pattern	0	1	0	1	0	
4	NP	104	77	120	80	71	1
	BC	29	29	76	37	34	
	CC	0	0	128	0	1	
	Pattern	0	0	1	0	0	

among feature values deliver meaningful information about the pixel membership. The relative intensity difference among feature values is called texture, which is independent of the mean value of the vector. Two feature vectors $\mathbf{y}_1 = [1, 0, 1, 0, 1]$ and $\mathbf{y}_2 = [100, 60, 100, 60, 100]$ have the same texture, while a third vector $\mathbf{y}_3 = [68, 5, 240, 10, 210]$ has a similar intensity pattern of High and Low as $\mathbf{y}_1$ and $\mathbf{y}_2$ but has a different texture, if we define the texture as $(\mathbf{y}_j - \mu_{\mathbf{y}_j})/\sigma_{\mathbf{y}_j}$.

Suppose that $\mathbf{y}_3 \in \omega_{10}$ (chromosome 10), and the pattern of $\mathbf{y}_3$ is consistent throughout all $\mathbf{y} \in \omega_{10}$, then a supervised classification method should work well without further normalizing the data. Even though background correction significantly reduced the variations in all feature vectors that belong to a chromosome ω_i for all i, there are pixels misclassified due to the aforementioned variations. Therefore, hybridized chromosomes must have a certain intensity level across all spectral channels, and at the same time, noise including intensity due to spectral crosstalk should have a certain intensity level that is lower than the intensity of hybridized chromosomes across all spectral channels. The normalization process should minimize the difference of the sample distributions (determined by the joint density functions) for all images. This can be achieved by normalizing the variables (the features).

An M-FISH image M is composed of six gray scale images $\{I_1, I_2, I_3, I_4, I_5, I_6\}$, each corresponding to a spectral channel. Each gray scale image I_κ contains gray scale values y that belong to background $I_b(\kappa)$ and chromosomes $I_c(\kappa)$, i.e., $I_\kappa = \{y | y \in I_b(\kappa) \cup y \in I_c(\kappa)\}$, and $I_c(\kappa) = \{y | y \in w_1 \cup y \in w_2\}$, where $w_1 = $ intensity due to none fluorophore and $w_2 = $ intensity due to a fluorophore. The distribution of y in $I_c(\kappa)$ is assumed to be a mixture of two Gaussians: $p(y|w_1) \sim N(\mu_1, \sigma_1)$ and $p(y|w_2) \sim N(\mu_2, \sigma_2)$, and $\mu_1 < \mu_2$. Then $I_c(\kappa)$ is a set of unlabeled samples drawn independently from the mixture density

$$p(y) = p(y \mid w_1)P(w_1) + p(y \mid w_2)P(w_2). \quad (3)$$

Since the models are identical for all channels, the channel index κ is not specified for y. A parameter vector $\boldsymbol{\theta}$ contains
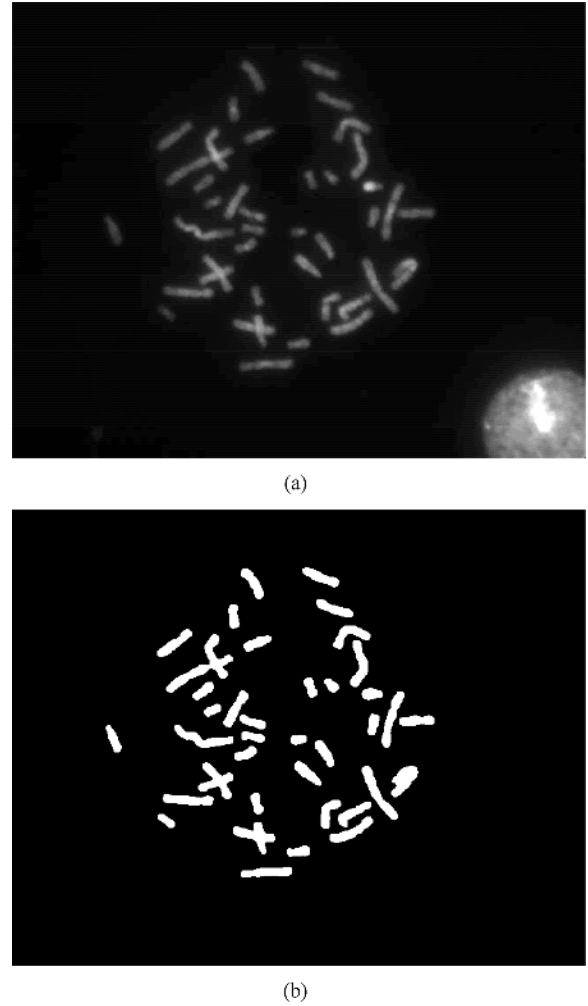

(a)

(b)

Fig. 3. Segmentation result. Chromosomes are automatically segmented from background by utilizing six spectral information, global and local intensity, and edge information. Cells are also removed based on the size and circularity. (a) V130740XY DAPI Channel. (b) Segmentation result.

$(\mu_1, \mu_2, \sigma_1, \sigma_2, P(w_1), P(w_2))$. $P(w_1)$ and $P(w_2)$ are prior probabilities and also called mixing parameters.

The separation between $I_c(\kappa)$ and $I_b(\kappa)$ is obtained by a new automatic segmentation method [8], which combines global and local intensity, spectral information, and edge information to segment chromosomes from the background. Cells are also removed based on size and circularity (see Fig. 3).

After the segmentation, only pixels inside chromosome area are classified. Among the six features, the DAPI channel provides information regarding whether a pixel belongs to chromosomes or to background. Since chromosome-background classification (segmentation) is already accomplished, DAPI information becomes redundant when classifying only chromosome pixels. Thus, the remaining five features are normalized and used for classification.

Fig. 4 shows an example of the mixture density distributions of $I_c(\kappa)$ of an M-FISH image, V1401XX. The black bars in Fig. 4 represent the range of gray scale values for a pixel that has a [high low high low high] pattern. As one can notice, a significant portion of high values in $I_c(3)$ overlaps with low values in $I_c(2)$ and $I_c(4)$, resulting in a totally unexpected pattern. This
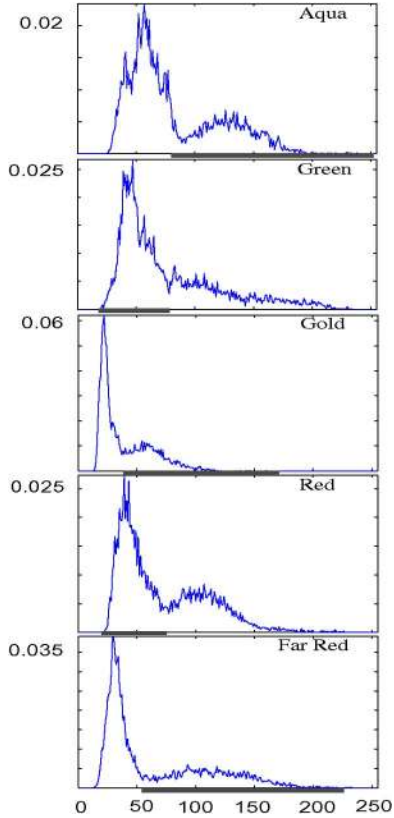
Fig. 4. The mixture density distribution of $I_c(\kappa)$ of V1401XX.

unexpected pattern will result in a low classification accuracy simply because the distributions of this image and the training data (or expected patterns) are different. We want to emphasize that this low classification accuracy comes from the difference in the patterns between the training and the testing data, and the accuracy is less dependent on the fundamental error rate (Bayes error) of the testing data. In other words, the joint distribution of five features of the testing data may have extremely small overlaps (low errors) among classes, but has its own distribution that is different from the training data, which will result in a low classification accuracy.

Fig. 5 further illustrates the aspects about this point. Suppose that each feature has a bimodal distribution, and there are two features describing four classes. The straight lines in the figure are the decision boundaries for the four classes. As one can see, the fundamental error rate of each data set is determined by the distribution of its marginal density functions. As the overlaps between the two modes in each feature increase, the error rate of the data increases. Both of the data in Fig. 5 seem to have small error rates. However, the testing data's classification accuracy will be low because the distribution is considerably different from the distribution of the training data set. The two distributions should be made as similar as possible by the normalization process in order to minimize the classification error.

Given a bimodal marginal density function (3) and its parameters, the normalization process should cause $y \in w_1$ and $y \in w_2$ to fall within certain ranges and the decision boundary between w_1 and w_2 to lie at a certain point. The parameters

$\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, P(w_1), P(w_2))$ are unknown, and the samples are unlabeled. $\boldsymbol{\theta}$ can be found by the maximum-likelihood estimation procedure.

When all parameters are unknown and if no constraints are placed on the covariance matrix (for multidimensional data), the maximum-likelihood principle yields useless singular solutions. However, meaningful solutions can still be obtained if we restrict our attention to the largest of the finite local maxima of the likelihood function, assuming that the likelihood function is well behaved at such maxima [9]. Then the parameter vectors for each class $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, P(w_i))$, assuming the feature vector $\mathbf{z}$ is multidimensional, can be estimated iteratively using the following equations (expressed in general terms):

$$\hat{P}(w_i) = \frac{1}{N} \sum_{j=1}^{N} \hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}}) \tag{4}$$

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{j=1}^{N} \hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}}) \mathbf{z}_j}{\sum_{j=1}^{N} \hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}})} \tag{5}$$

$$\hat{\boldsymbol{\Sigma}}_i = \frac{\sum_{j=1}^{N} \hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}})(\mathbf{z}_j - \hat{\boldsymbol{\mu}}_i)(\mathbf{z}_j - \hat{\boldsymbol{\mu}}_i)^T}{\sum_{j=1}^{N} \hat{P}(w_i | \mathbf{z}_j, \hat{\boldsymbol{\theta}})} \tag{6}$$

where N = number of unlabeled samples drawn independently from the mixture density of c classes, $i = 1, \ldots, c$, and

$$\hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}})$$
$$= \frac{p(\mathbf{z}_j \,|\, w_i, \hat{\boldsymbol{\theta}}_i) \hat{P}(w_i)}{\sum_{l=1}^{c} p(\mathbf{z}_j \,|\, w_l, \hat{\boldsymbol{\theta}}_l) \hat{P}(w_l)}$$
$$= \frac{|\hat{\boldsymbol{\Sigma}}_i|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z}_j - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{z}_j - \hat{\boldsymbol{\mu}}_i)\right] \hat{P}(w_i)}{\sum_{l=1}^{c} |\hat{\boldsymbol{\Sigma}}_l|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z}_j - \hat{\boldsymbol{\mu}}_l)^T \hat{\boldsymbol{\Sigma}}_l^{-1}(\mathbf{z}_j - \hat{\mu}_l)\right] \hat{P}(w_l)}. \tag{7}$$

Note that the number of classes c is two for our problem. Thus, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ per feature, and samples are grayscale values and $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are scalars per class from the bimodal distribution.

Among the various techniques that can be used to obtain a solution, one approach is to use an initial estimate to evaluate (7) for $\hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}})$, then use (4)–(6) to update the estimates [9]. This iterative method is also called expectation-maximization (EM). Since the solution depends on the initial estimates and to obtain fast convergence, a k-means clustering method is used to estimate the initial parameters. k-means clustering is a simple but popular method of finding the c mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_c$. Given the c initial mean vectors $\boldsymbol{\mu}_m$, the samples are classified to the nearest $\boldsymbol{\mu}_m$. Then by approximating $\hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}})$ in (5) as

$$\hat{P}(w_i \,|\, \mathbf{z}_j, \hat{\boldsymbol{\theta}}) \cong \begin{cases} 1, & \text{if } i = m \\ 0, & \text{otherwise} \end{cases}$$

anew estimates of the c mean vectors are obtained. The iteration repeats until the means converge. Usually c randomly chosen samples are used as the initial c means. In our case, the minimum and the maximum gray scale values in each channel are used as the initial mean values. Once $\hat{\mu}_1$ and $\hat{\mu}_2$ are found via the k-means clustering, the values $\hat{\sigma}_i^2$ are estimated from the samples classified to w_1 and w_2. These means and variances along
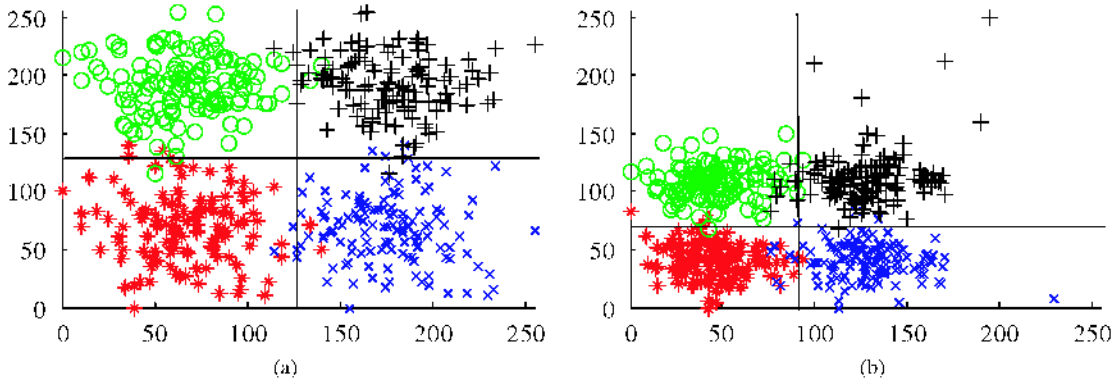
Fig. 5.   Distributions of training data and testing data. x and y axes are the feature values (thus, a feature vector forms a point in the figure). Each data set has its own fundamental error rate by its own distribution, but the classification accuracy for the testing data will be low because the distributions are different between the two data sets. The distributions should be normalized in order to obtain a high classification accuracy. (a) Training data. (b) Testing data.
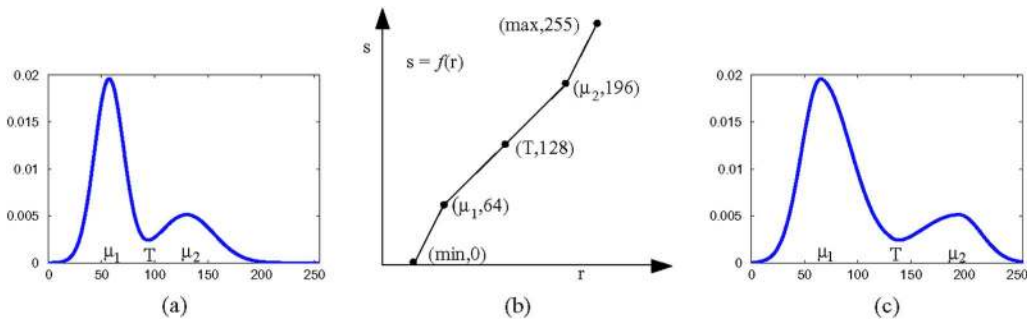

Fig. 6.   Marginal density function in (a) is normalized as in (c) by the piece-wise linear gray level mapping function in (b). The horizontal axes represent gray scale range. (a) Before normalization. (b) Normalization. (c) After normalization.

with equal priors are used as initial estimates for (4)–(6). Once the parameters are estimated by the EM method, the decision boundary between w_1 and w_2 is found by

$$T = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \qquad (8)$$

where

$$A = \hat{\sigma}_2^2 - \hat{\sigma}_1^2$$
$$B = 2\hat{\sigma}_1^2\hat{\mu}_2 - 2\hat{\sigma}_2^2\hat{\mu}_1$$
$$C = \hat{\sigma}^2\hat{\mu}_1^2 - \hat{\sigma}_1^2\hat{\mu}_2^2 - 2\hat{\sigma}_2^2 ln\left(\frac{\hat{\sigma}_2\hat{P}(w_1)}{\hat{\sigma}_1\hat{P}(w_2)}\right).$$

Given the parameter vectors and the decision boundary, the sample distribution is normalized by piece-wise linear transformations as shown in Fig. 6. The input intensity r is mapped to the output intensity s by

$$f(r) = \begin{cases} \frac{64}{\mu_1 - \min(r)}(r - \min(r)), & \text{if } \min(r) \leq r < \mu_1 \\ \frac{64}{T - \mu_1}(r - \mu_1) + 64, & \text{if } \mu_1 \leq r < T \\ \frac{64}{\mu_2 - T}(r - T) + 128, & \text{if } T \leq r < \mu_2 \\ \frac{63}{\max(r) - \mu_2}(r - \mu_2) + 192, & \text{if } \mu_2 \leq r < \max(r) \end{cases}$$
$$(9)$$

where $\min(r)$ is the minimum intensity level and $\max(r)$ is the maximum intensity level in r.

In the following sections, we will show how this new normalization helps classification accuracies.

III. PIXEL CLASSIFICATION METHODS

In this section, we briefly describe previous classification methods and introduce our new unsupervised and nonparametric classification method.

A. Previous Classification Methods

The first M-FISH system described in the literature was introduced by Speicher *et al.* [1] in 1996. Their classification method was based on the binary combinations of fluorophore intensities at each pixel. Binary values were obtained after thresholding each channel. This method is simple and fast (considering only the pixel classification time, excluding the time involved in manual corrections of the segmentation map), and does not require generation of a training data set. Their approach demonstrated the usefulness of the M-FISH technique.

In 1998, Eils *et al.* [10] introduced a method called the adaptive region-oriented approach. An image was initially divided into forty Voronoi polygons, and the polygons were subdivided iteratively until all polygons satisfied a homogeneity criterion. Neighboring polygons were merged if they were closer than a threshold distance in feature domain (6-D space). The accuracy of their method was not reported.

Recently, a supervised 6-feature, 25-class maximum likelihood classification method was introduced in [11]–[13]. Twenty five classes included 24 chromosome types and the background. Class distributions were assumed to be normal, and the class parameters were extracted from the training set,

TABLE IV
TRAINING IMAGES

Vysis	ASI	PSI
V1301XY	A0101XY	P0801XY
V1302XY	A0102XY	P0802XY
V1303XY	A0103XY	P0803XY
V1305XY	A0104XY	P0804XY
V1306XY	A0105XY	P0805XY
V1801XY	A0201XY	P0808XY
V1802XY	A0202XY	P1102XY
V1803XY	A0205XY	P1103XY
V1805XY		P1104XY

a subset of ADIR's M-FISH database.[1] By classifying every pixel in the image including both background and chromosome pixels, chromosomes were successfully segmented from the background. The pixel classification accuracy of this method was about 90% on a small number of images (the list of images and rates are shown in Table VI).

Schwartzkopf *et al.* [4] developed a joint pixel classification and segmentation method which can handle overlapping and touching chromosomes, using a maximum likelihood framework. After chromosome pixels were classified using a 6-feature, 24-class maximum likelihood classification method, touching and overlapping chromosomes were separated into single chromosomes by maximizing the likelihood of pixel membership and chromosome. While separating overlapping and touching chromosomes, misclassified pixels were corrected resulting in an increased classification accuracy from the initial pixel classification. The initial pixel classification accuracy significantly varied depending on the images, ranging from 20% to 90%. The mean pixel classification accuracy was 68% with a standard deviation of 17.5% [14]. The pixel classification error rate decreased by nearly 50% after using the joint segmentation and classification method [4]. Their method focuses on postprocessing rather than preprocessing the data to increase the initial pixel classification accuracy.

Choi *et al.* [13] have emphasized the importance of feature normalization, and performed background correction and color compensation in order to reduce the background elevation and channel crosstalk. Wang and Castleman [15] also performed background correction as a normalization step, and reported that after testing on five images, the pixel classification accuracy increased on average from 83% to 91% (the list of images and rates are shown in Table IV).

In all these previous methods, a proper feature normalization was not performed.

B. Unsupervised Nonparametric Classification Methods

Supervised classification methods, such as the Bayes classifier (parametric) and k-nearest neighbor clustering (nonparametric), require training data. If the number of classes and the forms for the class-conditional probability density functions are known, the class parameters can be estimated from the training data, and a parametric classification method can be used. If the number of classes is known but the forms for the class-conditional probability functions are unknown, then a nonparametric method such as k-nearest neighbor clustering can be used. In

[1]http://www.adires.com/05/Project/MFISH_DB/MFISH_DB.shtml

general, collecting and labeling a large set of samples can be extremely costly. Fortunately, we have available a large M-FISH image database. Thus the use of a supervised method is an adequate approach. However, in an early stage of investigation regarding the structure of the data based on some features, an unsupervised method is desired since the samples are unlabeled. Then, unsupervised methods can be used to generate the training data set and further to extract useful features.

Popular unsupervised methods are k-means clustering and fuzzy k-means clustering, which group the samples into k clusters whether or not k classes actually exist in the data. These methods can be readily used for normal XX (23 classes) or XY (24 classes) samples where the number of classes is known and fixed. When only the maximum number of classes is known, the use of one of these methods requires cluster validation to assess the right number of classes by finding the right threshold, which may or may not be feasible depending on the data.

In order to overcome the limitations of these unsupervised methods, we introduce a simple but effective unsupervised and nonparametric classification method for M-FISH images. The concept starts from the fact that a set of samples bound to a particular probe set has an expected intensity pattern for each class. Those fundamental patterns can be used as templates or ideal prototypes for classes. If our normalization process was effective, then the distance between a normalized sample to its correct class mean (template) should become as close as possible. If that is true then the minimum-distance classifier [9] can be used to classify pixels without actually training the classifier, and the classification accuracy will express the effectiveness of the normalization.

The derivation of the minimum-distance classifier is as follows. In Bayesian decision theory, the minimum-error-rate classification can be achieved by using a set of discriminant functions $g_i(\mathbf{y}), i = 1, \ldots, k$, and the classifier assigns $\mathbf{y}$ to class ω_i if

$$g_i(\mathbf{y}) > g_j(\mathbf{y}) \quad \text{for all } j \neq i \tag{10}$$

where

$$g_i(\mathbf{y}) = \ln p(\mathbf{y} \mid \omega_i) + \ln P(\omega_i).$$

If the density functions are multivariate normal in d dimensions, then the class-conditional probability density functions are expressed as

$$p(\mathbf{y} \mid \omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{y} - \boldsymbol{\mu}_i)\right]$$

and then the discriminant functions can be written as

$$g_i(\mathbf{y}) = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{y} - \boldsymbol{\mu}_i)$$
$$-\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i).$$

If we assume that the features are statistically independent and have the same variance σ^2, then the discriminant functions become

$$g_i(\mathbf{y}) = -\frac{\|\mathbf{y} - \boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

after ignoring the additive constants. After further simplification, we obtain the linear discriminant functions whose decision boundary is the hyperplane perpendicular to the line linking the class means. If the prior probabilities are assumed to be the same for all c classes, then the classifier assigns a feature vector $\mathbf{y}$ to the class that yields the minimum Euclidean distance $\|\mathbf{y} - \boldsymbol{\mu}_i\|$. This classifier is essentially the same as the nearest neighbor classifier. In general, multiple samples are used to represent each class in the nearest neighbor method, and after computing the distances to all the samples from an unknown sample $\mathbf{y}$, the sample $\mathbf{y}$ is assigned to the most frequent class among the k-nearest neighbors. In the template matching case, only one sample (ideally the class mean) per class is used to represent the class.

In our case, the template patterns are determined by the color table (e.g., Table I). Let a template sample from class ω_1 be denoted as $\mathbf{x}_1 = [0, 0, x, 0, 0]$, from ω_2 as $\mathbf{x}_2 = [0, 0, 0, x, 0]$, and so on, where x can be any positive real number. Then the template patterns are defined as

$$\boldsymbol{\mu}_i^t = \frac{\mathbf{x}_i - \mu_{\mathbf{x}_i}}{\sigma_{\mathbf{x}_i}}.$$

After EM normalization, it is important that the samples $\mathbf{y}$ should be further normalized before pixel classification by $\mathbf{y}' = (\mathbf{y} - \mu_{\mathbf{y}})/\sigma_{\mathbf{y}}$. Thus an unknown sample $\mathbf{y}$ is assigned to ω_i if

$$\|\mathbf{y}' - \boldsymbol{\mu}_i^t\| < \|\mathbf{y}' - \boldsymbol{\mu}_j^t\| \text{ for all } j \neq i. \tag{11}$$

IV. RESULTS

A. M-FISH Database

The developed normalization and classification methods were tested on the ADIR M-FISH database. The database contains M-FISH images of 203 metaphase spreads from 33 slides. Applied Spectral Imaging, PSI (former ADIR), and Vysis are the three probe sets that were used for the specimens. Three sets of file formats are available: PSI format (requires PSI's software to read), PNG format, and JPEG format. Each image is accompanied by ground truth except 17 images that are marked as extreme (EX). The set of PNG format images were used in our experiment, and a total of 185 images were tested (85 images for Vysis, 71 images for ASI, and 29 images for PSI). There are 86 Vysis probe images but V1301XY and V1304XY are the same images (only V1301XY was used). In the ground truth image, background pixels are assigned value 0, pixels in overlapped region are assigned value 255, and chromosome pixels are assigned a value from 1 to 24. In the case of a translocation, the whole chromosome is labeled as the class which makes up most of the chromosome. The dimension of the images is $647 \times 517 \times 6$ for all images except for two images, V261054 and V270659, whose dimension is $768 \times 568 \times 6$.

B. EM Normalization

Each data set has its own unique error rate (Bayes error) based on the feature distribution. While the fundamental error rate of each data set is one problem that causes classification
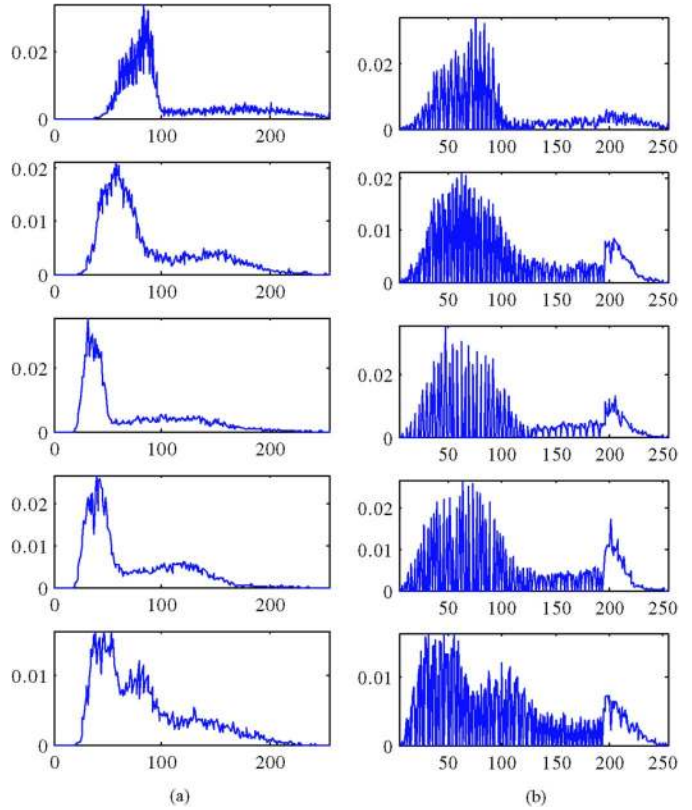

Fig. 7. Feature distribution (normalized histogram) of V1290562 before and after the EM normalization. Horizontal axis represents gray scale and vertical axis represents the normalized frequency of a gray scale. EM normalized images are shown in Fig. 8, and the classification result is shown in Fig. 10. (a) Before normalization. (b) After normalization.

error, the significant error comes from having different distributions for different data sets. The EM normalization process is focused on reducing the distribution differences among the different data sets. Thus, the classification accuracy improves significantly after normalization (rates are shown in the following section).

The mixture density parameters for each feature were found by (4)–(6), and then the decision boundary between the modes was found by (8). Given the parameters and the decision boundary T, the features were normalized by (9).

In particular, Fig. 7 shows the intensity distributions of the features of V290562 before and after EM normalization. As the figure shows, the uncertainty of a gray scale at a channel being high (hybridized) or low (not hybridized) is removed in the normalized data.

Fig. 8 shows the gray scale images before and after the EM normalization. As the figure shows, chromosomes that are hybridized have higher intensity levels than the intensities due to nonhybridized chromosomes. This normalization ensures that the patterns become consistent throughout all images.

C. Comparison of Classification Methods

The pixel classifications were performed with three different conditions: no preprocessing, background correction, and EM normalization. Both unsupervised-nonparametric (the minimum-distance classifier) and supervised-parametric (the
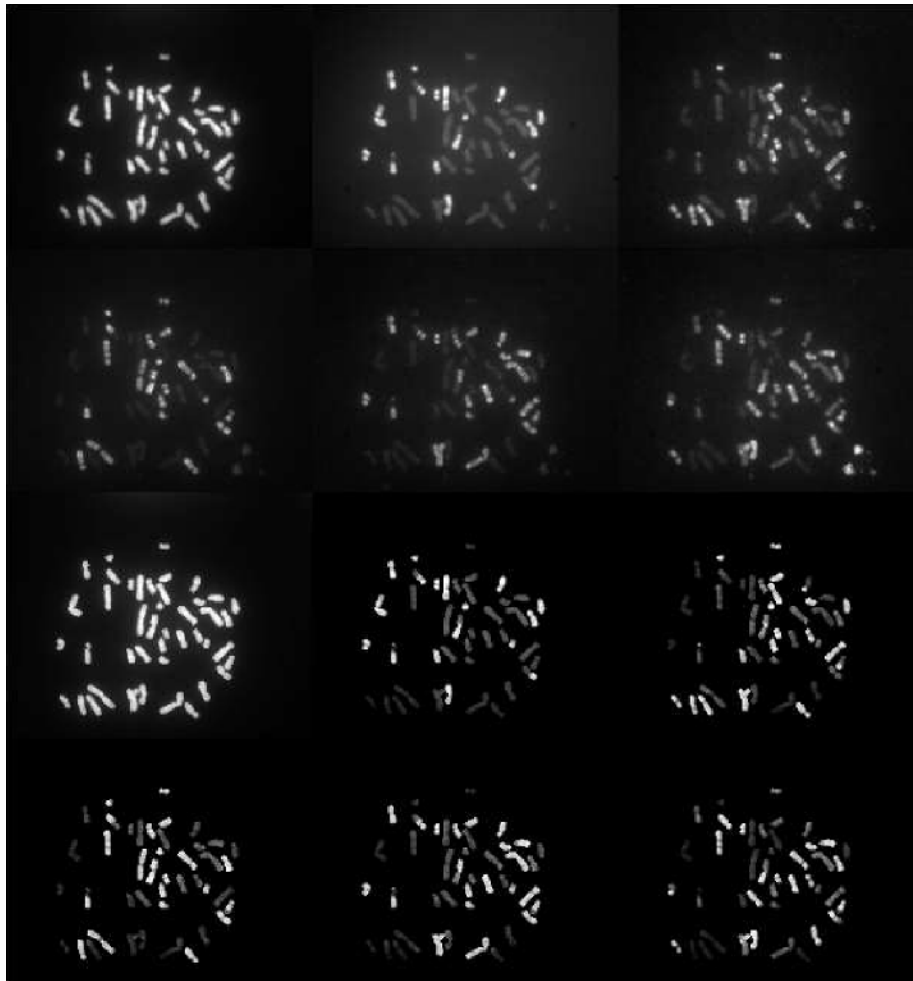
Fig. 8.   EM normalization result of V1290562 before (upper two rows) and after (lower two rows) the EM normalization.

TABLE V
OVERALL CLASSIFICATION ACCURACY. MD = MINIMUM-DISTANCE, ML = MAXIMUM-LIKELIHOOD, NP = NO PREPROCESSING, BC = BACKGROUND CORRECTION, AND EM = EXPECTATION MAXIMIZATION NORMALIZATION

	MD classifier			ML classifier		
Methods	NP	BC	EM	NP	BC	EM
Accuracy [%]	47.12	60.11	68.70	47.86	62.46	72.72

TABLE VI
CLASSIFICATION ACCURACIES [%] OF THE COMMONLY CITED IMAGES. IMAGES WITH EMPTY VALUES IN THE ML METHOD ARE USED AS TRAINING

Images	MD classifier			ML classifier		
	NP	BC	EM	NP	BC	EM
V1301XY	63.77	88.51	90.81			
V1302XY	83.35	92.35	92.99			
V1303XY	81.15	90.13	92.77			
V1305XY	93.01	92.36	94.72			
V1306XY	87.64	89.81	94.66			
V1308XY	77.95	86.63	96.28	84.45	93.89	96.49
V1309XY	56.16	83.58	84.50	70.34	84.55	86.29
V1310XY	80.57	88.00	84.19	86.03	88.50	86.90
V1311XY	94.01	93.48	94.50	90.79	90.99	94.54
V1312XY	87.31	93.44	94.83	95.09	95.25	95.20
V1313XY	89.92	91.69	94.53	93.76	94.82	94.88
Average	81.35	90.00	92.25	86.74	91.33	92.38

maximum-likelihood classifier) methods were used as classification methods.

Since the maximum-likelihood classifier requires training, a set of images of normal male specimens were selected as training samples for each probe set, as shown in Table IV. A total of 26 out of 185 images were used for training: 9 out of 85 images for Vysis images, 8 out of 71 images for ASI images, and 9 out of 29 images for PSI images. All 185 images were tested using both classification methods. Equation (10) was used for the maximum-likelihood classifier assuming the distributions were normal and (10) was also used for the minimum-distance classifier to classify pixels.

As Table V shows, the overall classification accuracy without any normalization was about 50%, which increased significantly after background correction to about 60%, and further improved

with EM normalization to about 70% for both classification methods. EM normalization increased the classification accuracy from 50% to 70%, which is a 40% increase in accuracy.

Table VI shows the classification accuracies for the commonly cited images in previous papers on M-FISH pixel classification. Note that the results shown in this paper are the initial pixel classification accuracies without any postprocessing to correct obvious misclassifications using such as
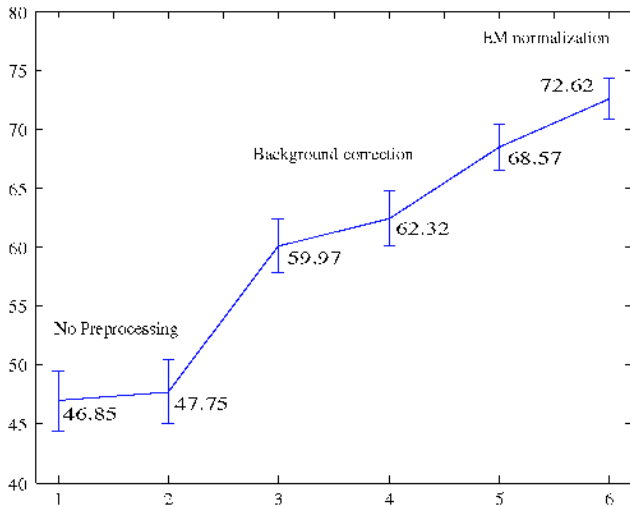
Fig. 9. Statistical significance of each classification method and bootstrapping of each method. Left to right: NP_MD, NP_ML, BC_MD, BC_ML, EM_MD, and EM_ML. Error bars are drawn at the 95th percentile.

majority filtering, and also note that the rates are regarding the chromosome pixels only. Since chromosomes occupy less than 10% of the image, even if the entire pixels in the image are classified, the rates for background and chromosomes should be separately reported. Our results are by far the most accurate compared to the other classification methods such as fuzzy logic classifier (unsupervised nonparametric method) [8], fuzzy k-means clustering (supervised nonparametric method) [15], k-nearest neighbors method (supervised nonparametric method) [16], and the maximum-likelihood classifier (supervised parametric method) [13]. All these classifiers would likely produce an improved classification accuracy after EM normlization. When a supervised classification method is used, the classification accuracy will be high when the sample distributions of both training and testing data are the same (an improved result is shown in this paper using a ML classifier). When a clustering method is used, given the maximum number of classes, for a data set whose number of classes is not known, a right number of classes should be estimated after clustering using the statistical parameters of known data. However, when the sample distributions vary significantly depending on data sets, the estimated statistical parameters will not be accurate for all data sets. Furthermore, clustering results also depend on the initial guess values (starting points). When they are not carefully chosen, the solution can fall into local minima or maxima, i.e., incorrect grouping of the data. These problems will be alleviated when data is normalized.

In order to evaluate the statistical significance of the effect of the EM normalization, bootstrap estimation was used. Given 185 data points for classification accuracies per method, 185 samples were selected at random (from a uniform distribution) iteratively 1000 times and at each iteration the mean was calculated. The distribution of the means for each method is shown in Fig. 9. The error bars represent the 95th percentile of the means. As the graph shows, the accuracies after the EM normalization are statistically significant. The difference between

the two classifiers are not significant except after the EM normalization. However, it should be remembered that 26 images were used as training and their classification results were also included in the ML accuracy. Therefore, it is reasonable to assume that the difference is slightly smaller than 4%, and whether it is statistically significant or not, the difference is marginal between the two classifiers.

Fig. 10 shows an example of a color coded classification result (its spectral images are shown in Fig. 8). The classification accuracies using the MD classifier without preprocessing, with background correction, and with EM normalization were 55.34%, 75.64%, and 84.03%, respectively. This particular image has six translocations but they are unmarked in the ground truth in the database. After carefully examining all six spectral images and manually constructing the new ground truth, the recalculated accuracy was 91.52%. There are 104 images that contain abnormalities, and among them 63 images contain translocations. If the ground truth in the database was marked with translocations, the true overall classification accuracy may have been slightly higher.

There are many images that give low classification accuracies even after EM normalization. The common factor among those images is that the image quality is poor. All 185 images were individually self trained and tested to evaluate the quality of the feature distribution. The mean accuracy was 89.95% with 51.30% as the minimum and 99.00% as the maximum (see Fig. 11). Images that gave lower than an 85% correct classification rate were identified (also visually confirmed) as bad images. In addition, three images that had higher than a 90% rate was added to the bad because they had wrong probe labeling. They were labeled as Vysis when in fact they were hybridized using the PSI probe. A total of 40 images were identified as bad, and the list is shown in Table VII.

In M-FISH, all six spectral channels are expected to be perfectly aligned to each other. While that is true for most cases, only four images were identified as misaligned. The misalignment can arise in various ways including mechanical shift during image capture and the use of poor quality lenses which will cause spherical and chromatic aberrations. The misalignment in V291562 was a vertical shift of 10 pixels only in channels 4 and 6. The misalignment in P080628, P080729, and P0804XY was found in the DAPI channels only.

It is interesting that the image quality varied depending on the probes. Images with the Vysis probe were captured with good quality in general. Many ASI probe images display poor quality. In many cases, at least one or more spectral channels occupied only the low intensity range. The cause can be either that the hybridization process was done poorly or the exposure times were not set correctly. Many PSI probe images showed low signal-to-noise ratio, as the signals do not display the high contrast. However, this comparison may not generalize the quality of the probes, since the image quality will depend on the quality of the specimen preparation and the settings in the microscope. Since images in the database were collected from five different labs, we do not know whether the image quality difference comes from human error or from the probe difference.
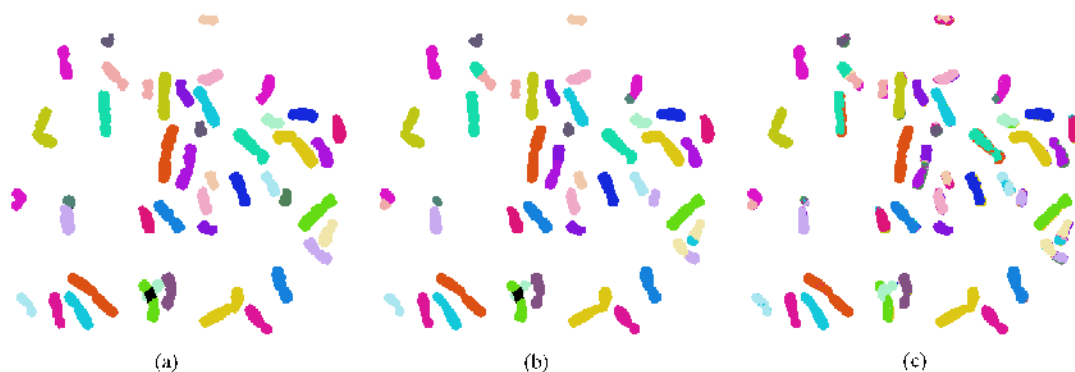
Fig. 10. Classification result of V290562 (spectral channels are shown in Fig. 8). (a) Ground truth. (b) Corrected ground truth. (c) EM_MD.

TABLE VII
LIST OF BAD QUALITY IMAGES. PQ = POOR QUALITY DUE TO EITHER ILL-HYBRIDIZATION OR WRONG EXPOSURE TIMES, CT = Channel Crosstalk, MA = Misalignment, WP = Wrong PROBE

| File name	Condition	File name	Condition	File name	Condition	File name	Condition
V250253	PQ	A020818	PQ	A0507XY	PQ	P070218	PQ
V260754	CT	A0202XY	PQ	A0604XY	PQ	P080628	PQ, MA(1)
V260856	CT	A0205XY	PQ	A0609XY	PQ	P080729	PQ, MA(1)
V290162	CT	A0206XY	PQ	A0614XY	PQ	P080930	PQ
V290362	CT	A0207XY	PQ	A0621XY	PQ	P0802XY	PQ
V290962	CT	A0209XY	PQ	A200344	PQ	P0804XY	PQ, MA(1)
V291562	MA(4,6)	A0402XY	PQ	A200444	PQ	P0808XY	PQ
A0102XY	PQ	A0403XY	PQ	A200544	PQ	V1701XY	WP
A020402	PQ	A0503XY	PQ	A200644	PQ	V1702XY	WP
A020315	PQ	A0506XY	PQ	P070109	PQ	V1703XY	WP
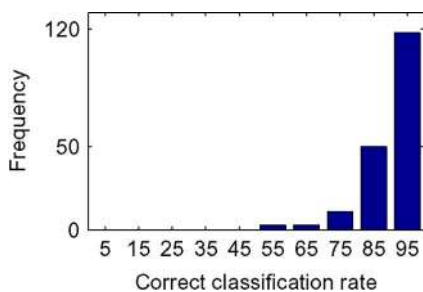
Fig. 11. Correct classification rate of individually self trained and tested images. Ten bins are used from 0 to 10, 10 to 20, ..., 90 to 100.

Excluding the bad images, the classification accuracy of remaining 145 images are shown in Fig. 12. The mean accuracies were 51.66%, 52.42%, 65.40%, 67.58%, 74.49%, and 77.80% for NP_MD, NP_ML, BC_MD, BC_ML, EM_MD, and EM_ML respectively.

V. CONCLUSION

In this paper, we have shown the importance of feature normalization in M-FISH images in order to obtain improved pixel classification accuracies, and we introduced a new normalization method using the expectation maximization algorithm. Previously the variation in the feature distributions among the different M-FISH images was not emphasized as a source of misclassification. Even if it was recognized, there was no good method of reducing the variation. Assuming the distribution of each feature in the chromosome region is a mixture of two normal density functions, the maximum-likelihood parameters
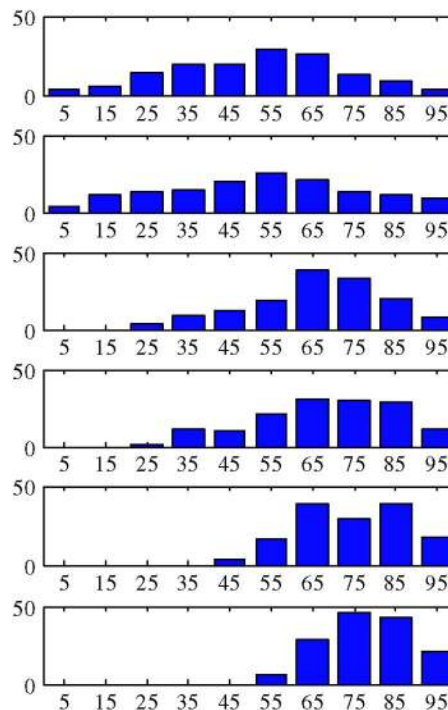

Fig. 12. Histogram of classification accuracies. x axis represents the classification accuracy [%], y axis represents the frequency. Ten bins are used from 0 to 10, 10 to 20, ..., 90 to 100. Top to bottom: NP_MD, NP_ML, BC_MD, BC_ML, EM_MD, and EM_ML, respectively.

were estimated for the mixture density and each feature was normalized based on the parameters. The overall pixel classification

accuracy improved by 40% after EM normalization from 50% (with no proprocessing) to 70% (with EM normalization). The improvement was statistically significant with no preprocessing and with background correction.

We have also introduced a new unsupervised, nonparametric classification method for M-FISH images. The performance was as accurate as the maximum-likelihood classifier, whose accuracy also significantly improved after EM normalization. In fact any classifier will likely produce an improved classification accuracy after EM normalization. Since the new classification method does not require training data, it is an effective and convenient tool when ground truth does not exist. And it can be readily used not only for M-FISH image of future probe sets but for other FISH images in general.

To build an automated karyotyping system, not only pixel classification but also the segmentation of overlapping and touching chromosomes should be done automatically. Utilizing the cluster shape, pixel membership, and prior knowledge about the chromosome size, overlapping and touching chromosomes can be separated. Based on the chromosome segmentation result, misclassifications can be corrected. However, the segmentation accuracy depends on the initial pixel classification accuracy. Therefore, our developed normalization and classification methods will help improve segmentation accuracy without the need for generating extensive ground truth.

ACKNOWLEDGMENT

REFERENCES

[1] M. R. Speicher, S. G. Ballard, and D. C. Ward, "Karyotyping human chromosomes by combinatorial multi-fluor FISH," *Nature Genet.*, 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] 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.

[4] W. Schwartzkopf, A. Bovik, and B. Evans, "Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images," *IEEE Trans. Med. Imag.*, vol. 24, no. 12, pp. 1593–1610, Dec. 2005.

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

[6] K. R. Castleman, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[7] K. R. Castleman, "Color compensation for digitized FISH images," *Bioimaging*, vol. 1, no. 3, pp. 159–165, Sep. 1993.

[8] H. Choi, K. R. Castleman, and A. C. Bovik, "Segmentation and fuzzy-logic classification of M-FISH chromosome images," in *Int. Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 69–72.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley Interscience, 2000.

[10] R. Eils, S. Uhrig, K. Saracoglu, K. Sätzler, A. Bolzer, I. Petersen, J. M. Chassery, M. Ganser, and M. R. Speicher, "An optimized, fully automated system for fast and accurate identification of chromosomal rearrangements by multiplex-fish (M-FISH)," *Cytogenetics Cell Genetics*, vol. 82, pp. 160–171, 1998.

[11] F. A. Merchant, K. N. Good, H. Choi, and K. R. Castleman, "Automated detection of chromosomal rearrangements in multicolor fluorescence in-situ hybridization images," in *Conf. Proc. 2nd Joint EMBS-BMES Conf. 2002 24th Annu. Int. Conf. Eng. Medicine Biol. Soc. Annu. Fall Meeting Biomed. Eng. Soc.*, Houston, TX, Oct. 2002, vol. 2, pp. 1074–1075.

[12] M. P. Sampat, K. R. Castleman, and A. C. Bovik, "Pixel-by-pixel classification of M-FISH images," in *Conf. Proc. 2nd Joint EMBS-BMES Conf. 2002 24th Annu. Int. Conf. Eng. Medicine Biol. Soc. Annu. Fall Meeting Biomed. Eng. Soc.*, Houston, TX, Oct. 2002, vol. 2, pp. 999–1000.

[13] H. Choi, K. R. Castleman, and A. C. Bovik, "Joint segmentation and classification of M-FISH chromosome images," in *Conf. Proc. 26th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, San Francisco, CA, 2004, vol. 3, pp. 1636–1639.

[14] W. Schwartzkopf, "Maximum likelihood techniques for joint segmentation-classification of multi-spectral chromosome images," Ph.D. dissertation, Dept. Electrical Computer Eng., Univ. Texas, Austin, Dec. 2002.

[15] Y. Wang and K. R. Castleman, "Normalization of multicolor fluorescence in situ hybridization (M-FISH) images for improving color karyotyping," *Cytometry Part*, vol. 64A, no. 2, pp. 101–109, 2005.

[16] M. Sampat, A. Bovik, J. Aggarwal, and K. Castleman, "Supervised parametric and non-parametric classification of chromosome images," *Pattern Recognit.*, vol. 38, no. 8, pp. 1209–1223, Aug. 2005.