

Feature Representation and Signal Classification in Fluorescence *In-Situ* Hybridization Image Analysis

Boaz Lerner, William F. Clocksin, Seema Dhanjal, Maj A. Hultén, and Christopher M. Bishop, *Member, IEEE*

Abstract—Fast and accurate analysis of fluorescence *in-situ* hybridization (FISH) images for signal counting will depend mainly upon two components: a classifier to discriminate between artifacts and valid signals of several fluorophores (colors), and well discriminating features to represent the signals. Our previous work has focused on the first component. To investigate the second component, we evaluate candidate feature sets by illustrating the probability density functions (pdfs) and scatter plots for the features. The analysis provides first insight into dependencies between features, indicates the relative importance of members of a feature set, and helps in identifying sources of potential classification errors. Class separability yielded by different feature subsets is evaluated using the accuracy of several neural network (NN)-based classification strategies, some of them hierarchical, as well as using a feature selection technique making use of a scatter criterion. The complete analysis recommends several intensity and hue features for representing FISH signals. Represented by these features, around 90% of valid signals and artifacts of two fluorophores are correctly classified using the NN. Although applied to cytogenetics, the paper presents a comprehensive, unifying methodology of qualitative and quantitative evaluation of pattern feature representation essential for accurate image classification. This methodology is applicable to many other real-world pattern recognition problems.

Index Terms—Color image segmentation, feature representation, fluorescence *in-situ* hybridization, image analysis, neural networks, signal classification.

I. INTRODUCTION

FLUORESCENCE *in-situ* HYBRIDIZATION (FISH) allows the detection of specific DNA sequences in intact cell chromosomes. It enables selective staining of various sequences in interphase nuclei and therefore the detection, analysis, and quantification of specific numerical and structural chromosomal abnormalities within these nuclei. FISH is a widespread and diversely applied technology. The fields of biology in which FISH is employed include prenatal diagnosis, gene mapping, DNA replication and recombination, clinical diagnosis, and monitoring of disease and radiation dosimetry [1].

Manuscript received June 20, 1999; revised March 22, 2000 and September 9, 2001. This work was supported in part by EPSRC, U.K., under Contract GR/L51072: *Automatic Analysis of FISH Images* and in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel. This paper was recommended by Associate Editor V. Murino.

B. Lerner is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: boaz@ee.bgu.ac.il).

W. F. Clocksin is with the Computer Laboratory, University of Cambridge, New Museums Site, Cambridge CB2 3QG, U.K.

S. Dhanjal is with Department of the Biological Sciences, Warwick University, Coventry CV4 7AL, U.K.

M. A. Hultén is with Department of the Biological Sciences, Warwick University, Coventry CV4 7AL, U.K.

C. M. Bishop is with Microsoft Research, Cambridge CB2 3NH, U.K.

Publisher Item Identifier S 1083-4427(01)10805-2.

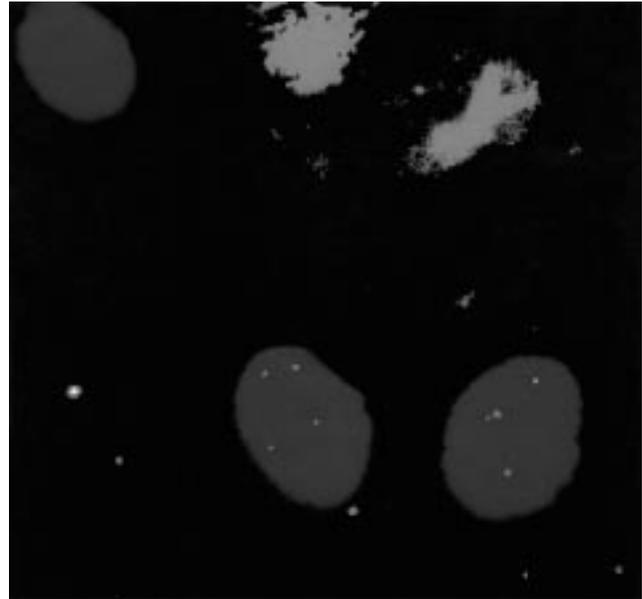


Fig. 1. Example (in black and white) of a FISH image used for dot counting.

Dot counting, the enumeration of signals (also called dots or spots) within the nuclei, is considered one of the most important applications of FISH, as the dots in the image represent the inspected chromosomes (see Fig. 1). Unfortunately, dot counting has to be performed in a dark laboratory room viewing the biological samples uncomfortably through a microscope in a time-consuming procedure. In addition, the expertise required for the completion of dot counting is gained through extended and expensive training. Digital microscopy in FISH allows the application of image analysis techniques for the automation of dot counting. The conventional approach [2] to automatic dot counting relies on an auto-focusing mechanism to select the “clearest” image for the analysis. However, basing dot counting on auto-focusing has some shortcomings [3], such as high cost, a long analysis period, and counting errors. Recently, it has been proposed [3] to base FISH dot counting on a neural network (NN) classifier discriminating between in-focus and out-of-focus images taken at different focal planes of the same field-of-view (FOV), as an alternative to the use of auto-focusing mechanism. Images at different focal planes are analyzed, and signals of each image are classified by the NN as valid data or artifacts, which are the result of out-of-focusing. Following the discrimination of valid signals and artifacts in each image, the image that contains no artifacts is selected as the in-focus image to represent that FOV, whereas the other

out-of-focus images are rejected. When the required number of (in-focus) images (or nuclei) is collected, proportion estimation of the number of cells having specific numbers of signals can be performed [4] as in auto-focusing-based dot counting methods. The suggested method overcomes most of the shortcomings of auto-focusing and when combined with multispectral analysis (which will be discussed in Section III) it shortens the length of image acquisition. Moreover, the classifier can be integrated into any existing dot counting system as it only replaces the auto-focusing mechanism. Also, the NN classifier can be replaced by any other classifier.

However, as the suggested system is required to classify both valid signals and artifacts of several fluorophores, its ability to discriminate between focused and unfocused signals should be more accurate than that of the discriminating element of a system employing an auto-focusing mechanism, as the latter encounters only (in-focus) valid signals. Therefore, the proposed system depends upon two components: a highly-accurate classifier to distinguish between valid and artifact signal data of several fluorophores, and well-discriminating signal representation. A hierarchical NN trained by a scaled conjugate gradient algorithm was found [3] to be an accurate classifier of FISH signals into real (valid) signals and artifacts of two fluorophores. In this paper, we aim to find well-discriminating feature representations of FISH signals of both in-focus and out-of-focus images that ensure an efficient and even more accurate signal (and thereby image) classification. Together, the two components provide a reliable methodology of fully automatic signal classification for dot counting in FISH images. Moreover, this methodology incorporates qualitative and quantitative evaluation techniques of pattern feature representation, making it both complete and generic for application to problems in other domains as well.

Automatic FISH image analysis in this paper is compiled of several ingredients. First, the analysis is multispectral utilizing color in both image processing, segmentation, feature description, and classification. Second, emphasis is placed on feature representation using vast and independent families of features, e.g., size, shape, intensity, color, and features projected on the data principal axes. Third, feature selection employing both qualitative and quantitative evaluation mechanisms is extensively applied to assure compact and accurate feature representation. Finally, classification is based on hierarchical NNs partitioning the feature space sequentially ensuring speedy, efficient, and precise discrimination.

Section II of this paper describes the procedure used to acquire FISH images, while Section III discusses the advantages of multispectral FISH image analysis. Section IV depicts the extraction of representative features of FISH signals, while feature selection is addressed in Section V. Section VI presents three NN-based classification strategies of signals into valid and artifact signals of two colors. A complete evaluation of feature representations of FISH signals by visual analysis of scatter plots and probability density functions (pdfs), as well as by two class separability criteria: a scatter criterion and the probability of misclassification, is given in Section VII. Section VIII concludes the paper.

II. BIOLOGICAL MATERIALS AND METHODS

A. Slide Preparation

The interphase nuclei preparations from amniotic fluid were made using the method by Klinger *et al.* [5] with minor modifications: 1–2 ml of amniotic fluid was centrifuged and the cell pellet washed in PBS warmed to 37 °C. The cells were resuspended in 75 mM Potassium Chloride (KCl) and put directly onto slides coated with APES (Sigma) and incubated at 37 °C for 15 min. Evaporation of PBS was compensated with filtered distilled water. Excess fluid was carefully removed and replaced with 100 ml of 3% Carnoy's fixative, 70% 75 mM KCl at room temperature for 5 min. The excess fluid was carefully removed and five drops of fresh fixative were dropped onto the cell area. Slides were briefly dried on a 60 °C hotplate, and then either used immediately for hybridization or dehydrated through an alcohol series and stored at –20 °C until required.

B. Hybridization

Target areas were marked on the slides using a diamond-tipped scribe. Target DNA was denatured by immersing in 70% formamide: 30% 2 × SSC at 73 °C for 5 min. Then, 10 μL of probe mix containing spectrum orange LSI 21 and spectrum green LSI 13 (Vysis UK) was applied to the target area and a coverslip was placed over the probe solution. Coverslips were sealed using rubber cement and slides placed in a prewarmed humidified container in a 37 °C incubator for 16 h. Coverslips were removed and slides washed in 0.4 × SSC/0.3% NP-40 solution at 73 °C for 2 min. Slides were then placed in 2 × SSC/0.1% NP-40 solution at room temperature for 1 min. When completely dried, 10 μL of DAPI II counterstain (Vysis UK) was applied to the target area and sealed under a coverslip.

C. Fluorescence Microscopy and Screening Procedure

Slides were screened under a Zeiss axioplan epifluorescence microscope using Zeiss × 100 objective. Signals were viewed using Vysis DAPI/Green/Orange triple bandpass filter set and images acquired using a CCD camera (Photometrics CH250/A) and SmartCapture software (Vysis UK). Red and green signals, corresponding to chromosomes 21 and 13, respectively, were seen on blue DAPI stained nuclei. Slides were scanned by starting in the upper left corner of the coverslip and moving from top to bottom. A total of 400 in-focus and out-of-focus images were collected from five slides, stored in tagged image file format (TIFF) format, and used in the feature evaluation experiments (see Section VII).

III. MULTISPECTRAL FISH IMAGE ANALYSIS

A. Motivation

Color systems such as color television and color photography usually follow the human visual system and describe and synthesize color images using the three primary colors—red, green, and blue (RGB). Similarly, a tricolor digital image can be considered as a two-dimensional (2-D) image having three intensity levels (red, green, and blue) at each pixel. By analyzing each of the three color channels of the RGB image separately and in

various combinations, preprocessing and segmentation of multispectral images can be facilitated.

In FISH preparation, multiple probes, labeled by different fluorophores, are frequently employed in combination. In the current study, for instance, chromosomes 13 and 21 are detected as green and red signals, respectively, whereas the nuclei are indicated by blue. Although the position in the image and the characteristics of the fluorophores are of importance, previous analysis [2] converts color information into a grey-intensity scale, and FISH image analysis is then based on brightness contrast. However, difficulties encountered during the analysis of intensity-based FISH images can be avoided if color information is maintained and used [3], [4]. Nuclei can be analyzed using the blue channel of the RGB image, whereas red and green signals using the red and green channels, respectively. Multispectral FISH image analysis is beneficial not only to facilitate preprocessing and segmentation, but also to yield color-based features that may contribute to efficient signal classification. Finally, the advantages of using multispectral analysis compared with the conventional intensity-based analysis are expected to increase with the complexity of the specific FISH application and the number of fluorophores employed.

B. Color Specification

Using the RGB color format, which is the most basic quantitative description of a color image, we represent color by the scaled red, green, and blue intensities of each image pixel. In the hue, saturation, intensity (HSI) format, which is more suitable for approximating human color judgments, the color of a pixel is represented by its hue and saturation, whereas the intensity indicates the pixel overall brightness regardless of its color. The RGB format is useful to facilitate preprocessing and nucleus and signal segmentations (see Section III-C). However, as valid signals have intensity components in only one channel (red signals in the red channel and green signals in the green channel), channel intensities of different color signals are very similar to each other. Therefore, only the hue parameter (H) of the HSI color format can detect differences in color, and hence be employed for measuring multispectral signal features.

To convert RGB to HSI format, we use [6]

$$H = \arctan 2(3^{1/2}(G - B), (2R - G - B)) \quad (1)$$

$$S = 1 - 3(\min(r, g, b)) \quad (2)$$

and

$$I = (R + G + B)/3 \quad (3)$$

where $r = R/(R + G + B)$, $g = G/(R + G + B)$, and $b = B/(R + G + B)$, and R, G, and B are the intensities in the three channels, respectively.

C. Color Image Segmentation

Special multistage (usually TopHat-based) procedures that rely on heuristically-derived thresholds and parameters are conventionally employed to segment FISH nuclei and signals [2]. Color image segmentation, however, avoids the use of these procedures. It is performed separately on each of the three different channels of the RGB image using global thresholds. In

this paper, threshold values of 0.5 and 0.8 of the maximum channel intensity are found suitable for the segmentation of signals and nuclei, respectively. Finding “good” global thresholds in the RGB image is straightforward compared with thresholding an intensity image since the channels contain no background and only blue (red, green) objects are found in the blue (red, green) channel. For these reasons, moderate changes in the threshold values barely affect the overall classification accuracy or the results of feature evaluation. Moreover, computationally speaking, the red and green channels of a color image are represented by sparse matrices (to see this, note that the area of a typical signal of, for example, ten pixels, is less than 0.01% of a typical image area, for example, 400×400 pixels²). Therefore, special algorithms for sparse matrices can be exploited to enable faster performance of multispectral analysis compared with the intensity-based analysis that requires the full matrices.

Thereafter, the blue, red, and green thresholded objects are used as candidates for nuclei and red and green signals, respectively. Noise reduction, boundary smoothing of the nuclei by morphological operations, and spatio-spectral correlation between nuclei and signals are then implemented to complete the segmentation.

IV. SIGNAL MEASUREMENT

A set of features is measured for each of the segmented signals to be classified. The set includes area (a size measure) and eccentricity (a shape measure), which have been previously suggested [2]. In addition, we measure a number of spectral features. We compute at each of the relevant RGB channels three intensity-based measurements: the total and average channel intensities and the channel intensity standard deviation. We also compute four HSI hue-based measurements: maximum hue, average hue, hue standard deviation, and Delta Hue. Delta Hue is the difference between the maximum and average hue normalized by the average hue. This last feature has been added to the set because it was observed that the difference between values of the average and maximum hue of real signals is usually near zero, whereas for some kinds of artifacts (e.g., overlap of signals of two different fluorophores) this difference is substantially large.

In the next subsection, we will elaborate the feature set in order to enhance signal representation.

A. Spectral Feature Extraction

A compact feature representation can be obtained by linear feature extraction of signal channel intensities using principal component analysis (PCA). Applying PCA [7, pp. 400–403], we can represent signal intensity using a lower-dimensional linear combination of projections of that intensity. Intensity is projected onto principal axes that maximize the data variance. Let $\mathbf{X} = f(\mathbf{Y})$ be a linear mapping of a random (intensity) vector \mathbf{Y} , $\mathbf{Y} \in R^d$, $\mathbf{X} \in R^m$, and $m < d$. The approximation

$$\hat{\mathbf{Y}} = \sum_{i=1}^m \mathbf{x}_i \phi_i \quad (4)$$

with the minimum mean-square error $\varepsilon = E\{(\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}})\}$ is obtained when ϕ_i ($\forall i = 1, d$) satisfy

$$\Sigma_Y \phi_i = \lambda_i \phi_i. \quad (5)$$

The m most effective principal axes ϕ_i are those eigenvectors associated with the m largest eigenvalues λ_i of the covariance matrix of the mixture density Σ_Y . $\mathbf{x}_i = \phi_i^T \mathbf{Y}$ are the projected values of the intensity vector \mathbf{Y} on ϕ_i .

PCA is applied here to signal channel intensities. Red and green signals have intensity components also in the blue channel of the RGB image, as the signal is part of a nucleus. To find projections which are capable of discriminating between red and green signals, we apply the PCA to 2-D vectors representing the intensities in only the red and green channels of the signal pixels. We would expect that the principal axes of red and green real signals will coincide with the R- and G-axes, since the signals have intensities in only one channel (either R or G). However, overlap between signals of different fluorophores, or between a signal and background fluorescence due to other fluorophores, leads to artifacts for which the principal axes are expected to be between the R and G axes. Using the eigenvectors as features can hence improve the ability to distinguish between real signals and artifacts of different colors, and we therefore include the two coordinates of ϕ_1 of the signal red and green intensity components in our feature set (see Section IV).

PCA-based features are very effective, yet their computation may become demanding. However, projection of a typical, natural image along the eigenvector which corresponds to the largest eigenvalue captures most of the variance (information) contained in the image. For these images, this eigenvector has a value of around $(1/3, 1/3, 1/3)^T$, which remains similar across different natural images [6]. Therefore, projecting a typical color image onto its first eigenvector is equivalent to computing $I_1 = (R + G + B)/3$, which is the average grey intensity of the image. This intensity is an effective color feature, and the load involved in its computation is negligible compared with that required to perform PCA. Therefore, I_1 is added to the signal feature set.

In summary, the nine features of Section IV as well as the additional three features presented here are all evaluated for FISH signal representation. The 12 features are listed and numbered in Table I to facilitate their identification in the rest of the paper.

V. FEATURE SELECTION

Classification based on a large feature set may be complex, costly to compute, and because of the ‘‘curse of dimensionality’’ even inaccurate. Moreover, even for a feature set of a moderate size, like the one employed here, some of the features can be found to contribute very little to the classification accuracy and others to be correlated to each other. The exclusion of redundant features from the set simplifies and shortens training of a classifier, and frequently also improves its accuracy [8]. Hence, the purpose of feature selection is to select a (small) subset of the feature set yielding accurate classification in minimal computational cost. In practical problems and for a not very large feature set, we can search among all the possible feature subsets and evaluate each one of them using a criterion of class separa-

TABLE I
SET OF FEATURES STUDIED IN THE WORK. NUMBERS ARE USED IN THE REST OF THE PAPER TO IDENTIFY THE FEATURES. TEXTURE INDICATES STANDARD DEVIATION OF CHANNEL INTENSITY (5) OR HUE (8). EIG. 1 AND 2 ARE ABBREVIATIONS FOR THE TWO COORDINATES OF THE EIGENVECTOR CORRESPONDING TO THE LARGEST EIGENVALUE OF THE RED AND GREEN INTENSITY COMPONENTS OF THE SIGNAL

Number	Feature	Number	Feature
1	Area	7	Average Hue
2	Eccentricity	8	Hue Texture
3	Total Channel Intensity	9	Delta Hue
4	Average Channel Intensity	10	Eig. 1
5	Texture	11	Eig. 2
6	Maximum Hue	12	Average Grey Intensity (I_1)

bility. The subset that achieves the highest value of the criterion is then selected to represent the patterns to the classifier.

The criterion of separability that is considered here, called J_1 , is based on the within-class scatter matrix [7, pp. 446–447]

$$S_w = \sum_{i=1}^L P_i E\{(X - M_i)(X - M_i)^T | \omega_i\} = \sum_{i=1}^L P_i \Sigma_i \quad (6)$$

and the between-class scatter matrix

$$S_b = \sum_{i=1}^L P_i (M_i - M_0)(M_i - M_0)^T \quad (7)$$

where

$$M_0 = E\{X\} = \sum_{i=1}^L P_i M_i \quad (8)$$

is the mean pattern of the mixture distribution. $X | \omega_i$ are patterns of class ω_i ($i = 1, L$) with mean M_i , covariance matrix Σ_i , and *a priori* probability P_i . The criterion

$$J_1 = \text{tr}(S_w^{-1} S_b) \quad (9)$$

where $\text{tr}(A)$ is the trace of matrix A, is expected to be larger when the between-class scatter matrix is larger and/or the within-class scatter matrix is smaller, thus indicating strong class separability.

VI. SIGNAL CLASSIFICATION

Recently [3], we have demonstrated the feasibility of automatic signal classification in in-focus and out-of-focus FISH images. The signal classification methodology employed in the current work is based on this demonstration and briefly summarized as follows. Signals (representing the red and green fluorophores) are classified into four classes: ‘‘real red,’’ ‘‘artifact red,’’ ‘‘real green,’’ and ‘‘artifact green.’’ Within the ‘‘artifact’’ classes we expect to find unfocused and overlap signals, and signals which are the result of background fluorescence. These signals will have patterns with different values of features than those of real signals, and hence will be classified as artifacts. Labels for the patterns, as belonging to one of the four classes, are

needed to train and evaluate the classifier, and they are obtained by an expert cytogeneticist using a custom-built graphical user interface for labeling FISH images [9].

Following the normalization of the features to zero mean and unit variance, patterns of signals extracted from all the images are divided randomly into training and test sets and classification into one of the four classes is implemented using cross-validation. A validation set which is drawn from the training set assures that the classifier is not over-trained, and the “optimal” configuration is selected. This guarantees rapid training and improved generalization capability of the classifier.

Three NN-based classification strategies are examined here. In the first, called the “monolithic strategy,” patterns are classified into the four classes using a single NN. The two other strategies are hierarchical and based on the assumption that the classification problem can be considered as a 2×2 class problem rather than a four-class classification problem. In the second strategy, termed the “independent,” patterns are classified into “red” and “green” classes using the “color network” and independently by a second network, the “real network,” into reals and artifacts. Classification of a pattern into the four classes is achieved by a common decision of both networks. In the third strategy, called “combined,” patterns are first classified into “red” and “green” classes using the “color network” and then based on the results of this network they are classified by two other networks, the “real-red network” and the “real-green network,” into reals and artifacts of the two colors. The building block of each of the classification strategies examined here is a two-layer perceptron NN trained by the scaled conjugate gradient algorithm [10, pp. 282–285]. Classification is based on the approximation of the two-layer perceptron outputs to the *a posteriori* probabilities for the classes. The classification accuracy is used here as the second criterion of class separability, in addition to J_1 (see Section V), employed for the evaluation of feature representation.

VII. EXPERIMENTS AND RESULTS

Several experiments were performed to investigate feature representations of FISH signals as well as to compare feature evaluation techniques and class separability criteria applied to the representations. Before the experiments, we had created a database of 400 in-focus and out-of-focus FISH images captured from five slides. Following nuclei segmentation, the system identified 944 objects within these images as nuclei, of which 613 also contained signals (the remaining 331 objects were unfocused nuclei that therefore contain no signals). Following signal segmentation, 3,144 objects within the above nuclei were identified as potential signals and features were measured for them. Based on labels provided by expert inspection (see Section VI), 1,145 of the signals were considered as “reals” (among them 551 were red) and 1,999 as “artifacts” (among them 1,224 were red).

A. Visual Analysis Using pdfs

Features have first been evaluated visually using conditional pdfs. Fig. 2 shows four examples of histogram estimates of

one-dimensional (1-D) conditional pdfs for three features of Table I—area, average channel intensity, and average hue. In the first three examples, red signals—reals and artifacts are compared, whereas in the last example real signals—red and green are evaluated. The first example [see Fig. 2(a)] demonstrates that the area parameter of real signals is much more confined than that of artifacts, but that classification of reals and artifacts based on the area feature could be inaccurate due to large overlap of the two distributions. Fig. 2(b) indicates less overlap between the distributions of reals and artifacts for the average channel intensity, where the values of the artifact intensities are usually lower than those of the reals. Average hue is found in Fig. 2(c) and (d) to be a well-discriminative feature for distinguishing between red and green signals [see Fig. 2(d)] and even between reals and artifacts [see Fig. 2(c)]. Similar graphs have also been derived for other combinations of classes and features. The large extent of overlap between distributions for different classes demonstrates some of the expected difficulties in classifying FISH signals into reals and artifacts of two colors.

B. Class Separability (Single Feature)

To extend the visual evaluation of single features for signal classification, we have performed additional experiments using two class separability criteria. In the first experiment, feature selection (see Section V) was applied to the original set of 12 features. Criterion J_1 (9) was computed for each and every feature to give an indication of the amount of class separability the feature provides. In the second experiment, the probability of misclassification estimated by the “monolithic strategy” (see Section VI) was evaluated for signals represented using each of the features. For each feature representation, the optimal configuration of the NN classifier was determined on a validation set, and training continued for 200 epochs. Table II shows the value of J_1 , rank according to J_1 of the feature among all the features (“lowest” is “best”), the optimal configuration of the classifier and the classification accuracy on the training and test sets. Results are given for each of the features of Table I.

We can draw two main conclusions from Table II. First, there is a general agreement between criterion J_1 and the classification accuracy. The three features with the highest values of J_1 [average hue (7), maximum hue (6) and Eig. 2 (11)] are those responsible for the highest classification accuracy. In addition, two of the features with the lowest values of J_1 [area (1) and hue texture (8)] also yield very low classification accuracy. For the rest of the features, the agreement between the two criteria is weaker, probably since the difference between two classification accuracies or two values of J_1 is frequently marginal. Moreover, this agreement additionally supports results coming from the study of the pdfs (see Fig. 2). The results also demonstrate the benefit of applying a simple class separability criterion such as J_1 to the data as a screening procedure before performing classification.

The second conclusion drawn from Table II is that in order to achieve sufficiently accurate classification we would need multifeature signal representations. Classification based on most of the single-feature representations failed since the representations could not lead to sufficient discrimination of signals of

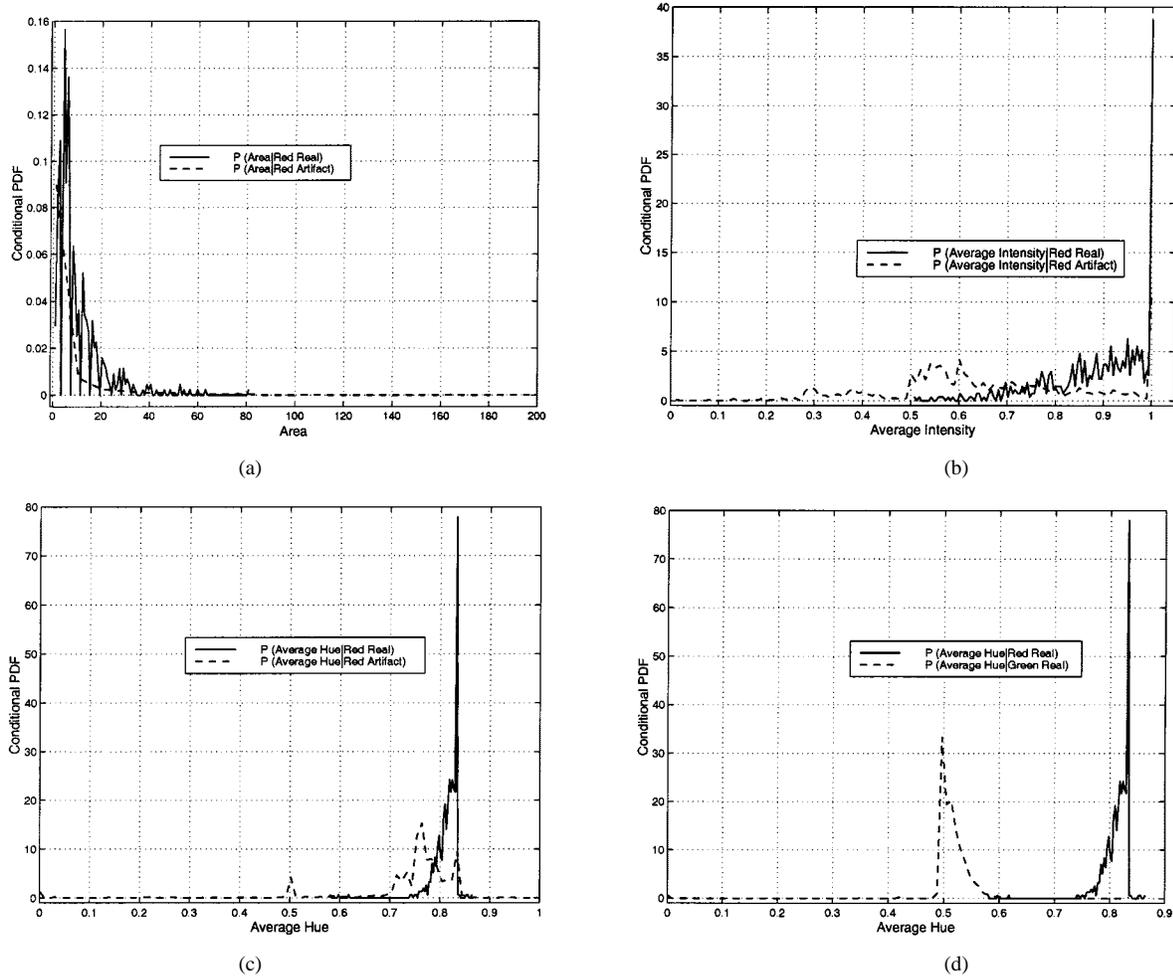


Fig. 2. Histogram estimates of the 1-D conditional pdfs for red signals: (a), (b), and (c): reals versus artifacts and for real signals: (d) red versus green. Density functions are plotted for the features: (a) area, (b) average channel intensity, and (c) and (d) average hue.

TABLE II

EVALUATION OF SINGLE FEATURES FOR SIGNAL CLASSIFICATION BY TWO CLASS SEPARABILITY CRITERIA. THE TABLE SHOWS THE VALUE OF J_1 (9) AND RANK ACCORDING TO J_1 , AS WELL AS THE OPTIMAL CONFIGURATION OF THE "MONOLITHIC" NN CLASSIFIER AND THE CORRESPONDING ACCURACIES ON THE TRAINING AND TEST SETS IN CLASSIFYING THE SIGNALS INTO THE FOUR CLASSES WHEN THE SIGNALS ARE REPRESENTED BY EACH OF THE FEATURES OF TABLE I

Feature Number	J_1	Rank	NN Configuration	Training (%)	Test (%)
1	0.0048	10	1:1:4	45.9	45.9
2	0.1058	7	1:2:4	46.5	46.6
3	0.0048	11	1:2:4	48.2	48.2
4	0.3237	5	1:3:4	46.4	46.3
5	0.1801	6	1:2:4	42.8	42.5
6	1.1398	2	1:7:4	65.8	65.4
7	1.1649	1	1:4:4	70.6	70.6
8	0.0013	12	1:9:4	44.9	44.5
9	0.0061	9	1:10:4	47.8	47.5
10	0.3534	4	1:2:4	41.3	40.8
11	0.6690	3	1:6:4	53.8	53.7
12	0.0704	8	1:3:4	49.2	49.0

two colors (except maybe for the maximum and average hue features).

C. Scatter Plots

Visual analysis of scatter plots of feature pairs may help in evaluating the impact of adding more features to the classification process and in detecting correlations between features. Fig. 3 shows scatter plots of four pairs of features. To facilitate the visual analysis, only 200 patterns of each of the four classes are randomly selected and presented. In Fig. 3(a), the average hue reveals good color discrimination, whereas the area is not able to resolve overlap between reals and artifacts. The linear dependency of the average channel intensity upon the hue feature in Fig. 3(b), and the "tendency" of points on the graph toward points of the other color require some explanations. Equation (1) shows that the hue of a signal depends nonlinearly upon the channel intensity according to an arctangent function. However, those signal intensity values which we find in the color channels of FISH images, fall onto the linear section of the arctangent function. Therefore, in our case, hue changes linearly with the average intensity. In addition, since artifacts are mostly the result of unfocused signals, their intensities are weaker than those of real signals. As the intensity of such an artifact signal,

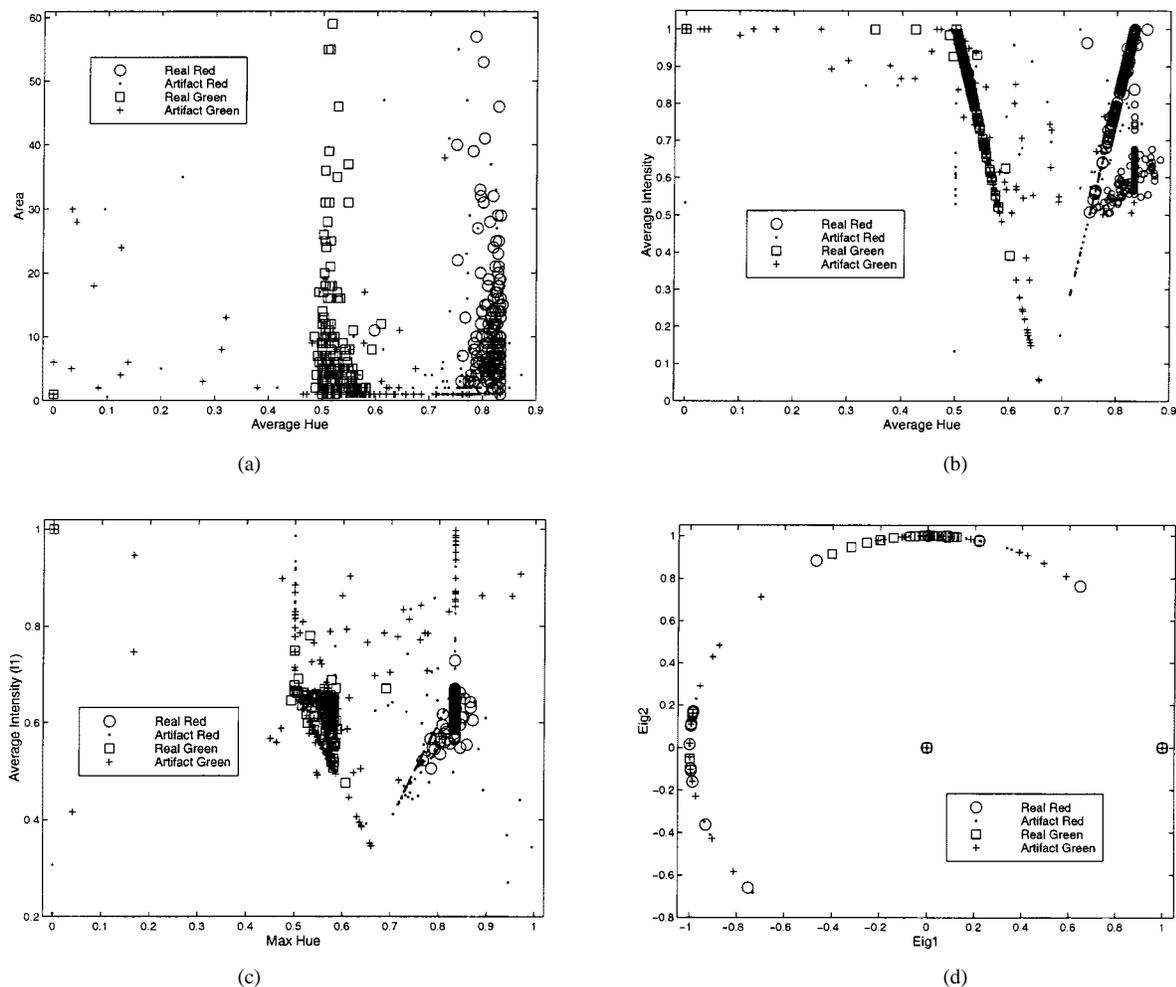


Fig. 3. Scatter plots of four pairs of features: (a) area versus average hue, (b) average channel intensity versus average hue, (c) average grey intensity (I_1) versus maximum hue, and (d) the two coordinates of the eigenvector corresponding to the largest eigenvalue (Eig. 1 and 2).

e.g., red, decreases, its main component, e.g., R, also decreases while the other two components remain the same ($G = 0$ and $B = 1$ in this example). Based on (1), the outcome is that the signal “changes” its color toward the other color as its intensity decreases. “Shift” of color is observed for both red and green signals, and having very weak intensities, signals of two colors may have almost similar hues [see Fig. 3(b)].

This also explains the dependency in Fig. 3(c) of the average grey intensity (I_1) upon the maximum hue. In this case, however, lines have different slopes as the average grey intensity (I_1) depends on one-third of the specific channel intensity (see Section IV-A). The two major clusters in Fig. 3(c) are mostly due to real signals. These signals have almost fixed values of I_1 since all of their three color components are fixed (e.g., for red real signals $R \simeq 1$, $G \simeq 0$, and $B = 1$). Artifacts that are caused by overlap of signals of different colors (or signals and fluorescence background of the other color) create some anomaly in the graph, which is not seen in Fig. 3(b). These artifacts have an additional intensity component of the occluded signal that increases the average grey intensity (I_1) but not the average channel intensity. However, as the intensity of the top signal is much stronger than that of the occluded signal, hue is determined almost entirely by the top signal. These artifacts are

responsible for the two almost “Max Hue” = constant-lines in Fig. 3(c). In addition, within these two lines we can find points (related to artifacts) that “swap” classes. This interesting phenomenon can be again explained by (1). When one intensity component, e.g., G, of overlap signals is larger than the second component, e.g., R, and the two intensities are large, there is an agreement between visual analysis and analysis based on (1) about the signal hue, e.g., green. Therefore, both the system and the expert cytogeneticist will agree on the signal hue. However, when the two intensity components are small (for $G > R$) or close to 0.5 (for $R > G$) the expert will still judge the hue by the top (stronger) signal, but as (1) predicts, hue will “shift” toward the color of the occluded (weak) signal, and the system will eventually decide on that latter color.

Finally, in Fig. 3(d), the two coordinates of the eigenvector corresponding to the largest eigenvalue are plotted against each other. Real signals have only one color intensity component, either R or G, and therefore, are projected onto either (1,0) [or (-1,0)] or onto (0,1) [or (0,-1)]. Color-mixed artifacts are projected inbetween. As PCA (see Section IV-A) cannot be applied to single-pixel signals, these signals (most of them artifacts) are projected artificially on (0,0).

TABLE III

ACCURACY OF THE THREE NN-BASED STRATEGIES: “MONOLITHIC,” “INDEPENDENT,” AND “COMBINED” IN CLASSIFYING FISH SIGNALS REPRESENTED BY DIFFERENT COMBINATIONS OF FEATURES. FEATURES ARE DEFINED BY THEIR NUMBERS ACCORDING TO TABLE I. REPORTED ARE THE NUMBER OF HIDDEN UNITS IN THE NN CLASSIFIER (hid.) AND THE PER CENT PROBABILITIES OF ACCURACY ON THE TRAINING (Tr.) AND TEST (Tst.) SETS FOR EACH OF THE STRATEGIES. THE TWO VALUES UNDER hid. FOR THE “INDEPENDENT” AND “COMBINED” CLASSIFIERS ARE THE NUMBERS OF HIDDEN UNITS OF THE “COLOR” AND “REAL” NETWORKS, RESPECTIVELY, WHICH MAKE UP THESE TWO STRATEGIES

Feature Combination	‘monolithic’			‘independent’			‘combined’		
	hid.	Tr.	Tst.	hid.	Tr.	Tst.	hid.	Tr.	Tst.
4, 7, 12	7	79.0	78.2	3, 11	78.5	77.7	3, 11	81.9	81.4
4, 5, 6	16	79.0	77.3	12, 7	79.1	77.3	12, 7	82.4	81.3
6, 9, 12	21	80.4	79.3	8, 8	79.9	79.0	8, 8	83.8	83.4
1, 4, 12	9	56.8	54.9	1, 4	55.7	54.6	1, 4	89.5	89.0
1, 4, 7	15	84.3	83.0	1, 14	82.1	81.5	1, 14	88.3	87.5
1, 4, 7, 9	20	85.3	83.4	13, 13	85.8	83.9	13, 13	88.9	88.1
1, 4, 7, 10, 11	19	86.9	84.1	9, 6	86.2	83.8	9, 6	89.4	87.9
1, 4, 7, 12	20	88.4	86.3	12, 7	88.2	86.3	12, 7	89.9	89.2
1, 4, 6, 7, 10–12	15	89.0	87.4	7, 12	89.3	87.0	7, 12	90.5	89.1

D. Class Separability (Multifeature)

The previous experiments have showed that, in order to achieve accurate classification we may need multifeature signal representations. Therefore, we employed the two class separability criteria, scatter criterion J_1 , and the probability of misclassification of the NN, in evaluating such representations.

Input and output dimensions of each of the NN-based classification strategies (see Section VI) were set by the feature space dimension and the number of classes, respectively. The number of hidden units was determined such that the network had the highest generalization capability. This was achieved by evaluating networks of different numbers of hidden units on an independent validation set drawn from the training set. The network which had the lowest error measured on the validation set was selected for training. Training of each of the networks, in each of the experiments reported here, was continued for 200 epochs and used three random network initializations. The results were averaged over these initializations (a committee) as part of a cross-validation (CV-5) experiment. This procedure was repeated for each signal feature representation and each classification strategy (see Section VI).

Table III shows the results of this evaluation for several manually-selected feature subsets, which are based on the most discriminative single features found in the previous experiments. Unseen signals represented by different combinations of features are classified as reals or artifacts of two colors with accuracies of up to 89.2% depending on the classification strategy. Feature subsets consisting of hue features [maximum hue (6) or average hue (7)] and intensity features [average channel intensity (4) or average grey intensity (I_1)(12)] are found to provide the best representations of the signals. Closer examination of the classification accuracies of the “color” and “real” networks

TABLE IV

EVALUATION USING CRITERION J_1 OF COMBINATIONS OF THREE FEATURES FOR FISH SIGNAL CLASSIFICATION. THE TABLE INCLUDES RESULTS FOR THE TEN COMBINATIONS WITH THE HIGHEST VALUES OF J_1 AND FOR OTHER COMBINATIONS OF THREE FEATURES FROM TABLE III. FOR EACH COMBINATION THE RANK ACCORDING TO J_1 AMONG THE 120 SUBSETS IS ALSO GIVEN. TO BE CONSISTENT WITH PREVIOUS FEATURE NUMBERS, WE KEEP NUMBER 12 FOR THE AVERAGE GREY INTENSITY (I_1), ALTHOUGH ONLY TEN FEATURES ARE INVOLVED IN THE SELECTION (SEE TEXT)

Feature Combination	J_1	Rank
4, 7, 12	1.7543	1
4, 5, 6	1.6789	2
4, 5, 7	1.6580	3
4, 6, 7	1.6046	4
4, 7, 8	1.6022	5
4, 6, 12	1.5781	6
2, 4, 7	1.5699	7
4, 7, 9	1.5692	8
2, 4, 6	1.5511	9
4, 6, 8	1.540	10
6, 9, 12	1.2218	56
1, 4, 12	0.4342	74
1, 4, 7	1.4958	14

(Section VI), which are responsible for the results of the “independent” and “combined” strategies, reveals that hue features (6, 7) are crucial for separating multicolor signals, where intensity features (4, 12) are essential for separating real signals from artifacts. Table III also shows that selecting specific subsets of features from the entire set is useful for FISH signal classification, as the accuracy of classification based on these sets is comparable with that based on the entire set. For example, classification accuracies on the test set based on the feature subset (1, 4, 7, 12) are 86.3%, 86.3%, and 89.2% using the “monolithic,” “independent,” and “combined” strategies, respectively, compared with accuracies of 85.1%, 87.0%, and 89.1% when the entire set is employed. Similar evidence for the benefit of applying feature selection to pattern classification in other real-world problems was given in [8], where texture features were used for pixel classification of synthetic aperture radar images.

We also applied feature selection using criterion J_1 (the second criterion of class separability) to evaluate feature subsets chosen from the entire set. Before the application, however, we had removed the two coordinates of the eigenvector corresponding to the largest eigenvalue in [Fig. 1 (10) and Fig. 2 (11) in Table I] from the data. This was done since the coordinates of single-pixel signals (most of them artifacts)

TABLE V
PERCENTAGE OF TIMES IN WHICH EACH OF THE TEN FEATURES OF TABLE I (EXCLUDING FEATURES 10 AND 11) APPEARS IN THE “BEST” (ACCORDING TO J_1) 30 COMBINATIONS OF THREE FEATURES

Feature Number	1	2	3	4	5	6	7	8	9	12
(%)	2.2	6.7	3.3	16.7	12.2	17.8	20.0	6.7	5.5	8.9

TABLE VI
“BEST” (ACCORDING TO J_1) FEATURE SUBSET FOR EACH FEATURE SUBSET SIZE

Feature Sub-Set Size	Feature Sub-Set	Feature Sub-Set Size	Feature Sub-Set
1	7	7	2, 4, 5, 7–9, 12
2	4, 7	8	2, 4–9, 12
3	4, 7, 12	9	1, 2, 4–9, 12
4	4, 5, 7, 12	10	1–9, 12
5	4, 5, 7, 8, 12	12	1–12
6	4, 5, 7–9, 12		

were set artificially to be (0,0), as the PCA could not be applied to single-element vectors [as explained in Section VII-C for Fig. 3(d)]. This removal prevented the expected bias of values of criterion J_1 toward these two features.

Since only ten features were included in the reduced feature set, we could allow exhaustive search for the “best” (according to criterion J_1) subset of, for example, three features. This search was done quickly since it involved the evaluation of only 120 subsets. The ten combinations of three features with the highest values of class separability criterion J_1 (and thereby lowest ranks) are presented in Table IV. The table also shows values of J_1 and ranks of the combinations of three features of Table III. Table V shows the percentage of times each of the ten single features appears in the 30 “best” (according to J_1) combinations of three features. Both Tables IV and V demonstrate the superiority, regarding criterion J_1 , of the features: average channel intensity (4), maximum hue (6) and average hue (7). In each of the ten “best” combinations, the average channel intensity and either or both the maximum or average hue were selected.

Finally, we employed the classification accuracy of the NN-based “monolithic strategy” as a criterion to determine the “optimal” size of the subset of “best” features. The “optimal” size was determined by the highest classification accuracy and the “best” features for each subset size were selected using criterion J_1 . Table VI presents the “best” features for each subset size. Fig. 4 shows the classification accuracy of the “monolithic strategy” obtained using the feature subsets given in Table VI. For small subset sizes, the classification accuracy increases almost linearly with size, but employing larger subsets only improves the accuracy moderately until the “curse-of-dimensionality” deteriorates the results. The

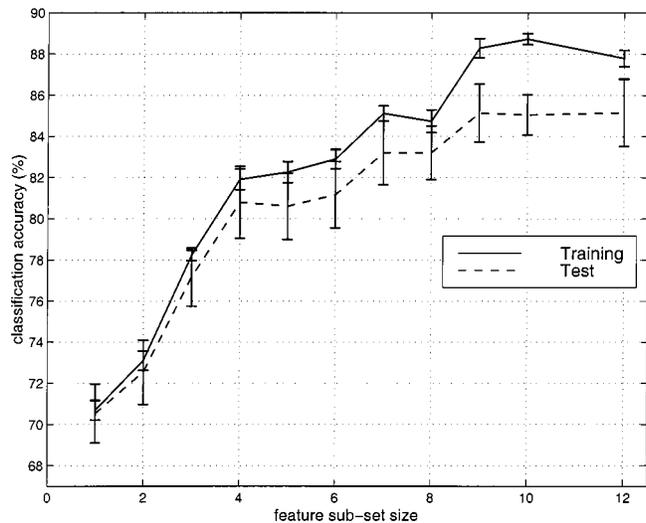


Fig. 4. Classification accuracy (mean and standard deviation) of the NN-based “monolithic strategy” for increasing sizes of feature subsets. Each subset includes the “best” (according to J_1) features (Table VI). The last subset includes the entire feature set.

effect of the curse-of-dimensionality is even more evident after comparing the highest accuracy (85.1%) of the “monolithic strategy” for the entire set (see Fig. 4) with those accuracies (higher than 86%) obtained using several manually selected subsets of four and seven features (see Table III). The success of the latter sets also hints to the inferiority of criterion J_1 compared with the classification accuracy in selecting optimal feature subsets for classification.

E. Classifier Comparison

In the last experiment, we compared the accuracy of the NN classifier with that of three other state-of-the-art techniques: Bayesian neural network (BNN), support vector machine (SVM), and naive Bayesian classifier (NBC). For computational reasons, we simplified the problem by dividing the classification task into two tasks: 1) classification of signals into color (red or green) and 2) classification of signals as “real” or “artifact.” This simplification determined indirectly the “independent strategy” as the technique of choice to represent the NN classifier in the comparison. The configuration and parameters of each of the classifiers were determined very carefully to enable peak performance of each of the four techniques [11]. The entire feature set was employed without applying feature selection. The comparison shown in Table VII reveals that the BNN is the most accurate (although not always significantly) classification technique for both tasks, and the NN and SVM are comparable and second best. The inferiority of the NBC compared with the other techniques is attributed to the relatively large amount of dependency among features of the set (e.g., average and maximum hue, total and average channel intensities). This dependency violates the independence assumption of the NBC [12], and thereby decreases its accuracy. This result emphasizes again the vital role of preliminary feature selection (or feature extraction performed as part of the classification process) in removing correlated features in order to facilitate pattern classification.

TABLE VII

ACCURACY OF FOUR TECHNIQUES ON THE FISH TEST SET WHEN SIGNALS ARE CLASSIFIED INTO COLOR AND AS "REAL" OR "ARTIFACT" (REAL/ARTIFACT) AND REPRESENTED BY THE ENTIRE FEATURE SET

Model	Colour (%)	Real/Artifact (%)
Neural Network (NN)	98.1	86.4
Bayesian Neural Network (BNN)	98.8	88.2
Support Vector Machine (SVM)	98.4	87.2
Naive Bayesian Classifier (NBC)	94.0	83.0

VIII. DISCUSSION

This paper has explored suitable feature representations for FISH signal classification. For this purpose, a family of features, consisting of measurements of size, shape, intensity, texture, and color, was examined. In addition, the application of linear feature extraction to signal intensities provided features which were capable of improving the accuracy of the classification, mainly due to the identification of artifacts resulting from signal overlap of two types of fluorophores. These last features, along with the average grey intensity, representing the projection of the image on its major principal axis, were measured as well.

A set of 12 features has been evaluated by different criteria. Histogram estimates of conditional pdfs and scatter plots provided preliminary visual insight into the relative merit of features to the classification procedure, dependencies between features, and potential causes of misclassification. Feature selection enabled the choice of feature subsets of any type and number, which maximized scatter criterion J_1 measuring class separability. However, the ultimate and most reliable criterion for evaluating features for class separability in problems with nonparametric class conditional pdfs (see Fig. 2) is the probability of mis-classification. When evaluated single features for FISH signal representation, the two class separability criteria mostly agreed. However, for more complex feature representations, some mismatches in selecting optimal subsets by the two criteria were found. These mismatches can be attributed to several factors. First, the scatter criterion we have used, J_1 , is based on the Euclidean metric, which is useful for discrimination purposes only when the class patterns have equal covariance matrices. Second, the J_1 criterion is sensitive to the relative locations of the classes in the feature space. For example, the probability of mis-classification of well-separated classes is not dependent on the distance between the classes (or their centroids) and is always zero, whereas J_1 is changed with this distance. These two factors deteriorated the ability of J_1 to act as a reliable class separability criterion in classification tasks. Third, the hidden layer of the NN classifier performed an additional feature extraction stage, which expanded the classifier discrimination power beyond that of a scatter criterion. We have found that the "optimal" subsets recommended by J_1 are not necessarily "optimal" for classification, i.e., they do not lead to the most accurate classification. Nevertheless, when the

calculation or estimation of the probability of mis-classification are difficult to be performed, a simple, easy to implement feature selection technique is of great value, especially as a preliminary analysis tool.

The extensive analysis has demonstrated the superiority of hue and intensity-based features for FISH signal classification. When features of the two families were combined together, even a single hue feature could separate entirely signals of two fluorophores, leaving the task of discriminating real signals from artifacts to intensity features. Consequently, feature sets consisting of both hue and intensity features enabled an NN-based hierarchical strategy to classify nearly 90% of FISH signals as reals or artifacts of two fluorophores. Utilizing the maximum information contained in the data to accomplish high classification accuracy with low computational requirements, the NN classifier provided performance comparable with those of other state-of-the-art classification techniques.

Finally, as almost only the measured features are specific to the classification problem, the methodology presented here for a complete, qualitative, and quantitative evaluation of feature representations can also be applied to other real-world pattern classification problems.

ACKNOWLEDGMENT

The authors thank the anonymous referees for their valuable comments.

REFERENCES

- [1] N. P. Carter, "Fluorescence *in situ* hybridization—State of the art," *Bioimaging*, vol. 4, pp. 41–51, 1996.
- [2] H. Netten, I. T. Young, L. J. van Vliet, H. J. Tanke, H. Vrolijk, and W. C. R. Sloos, "FISH and chips: Automation of fluorescent dot counting in interphase cell nuclei," *Cytometry*, vol. 28, pp. 1–10, 1997.
- [3] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hultén, and C. M. Bishop, "Automatic signal classification in fluorescence *in-situ* hybridization images," *Cytometry*, vol. 43, pp. 87–93, 2001.
- [4] K. R. Castleman and B. S. White, "Dot count proportion estimation in FISH specimens," *Bioimaging*, vol. 3, pp. 88–93, 1995.
- [5] K. Klinger, G. Landes, D. Shook, R. Harvey, L. Lopez, P. Locke, T. Lerner, R. Osathanondh, B. Leverone, T. Houseal, K. Pavelka, and W. Dackowski, "Rapid detection of chromosome aneuploidies in uncultured amniocytes by using fluorescence *in situ* hybridization (FISH)," *Amer. J. Human Genetics*, vol. 51, pp. 55–65, 1992.
- [6] Y. Ohta, *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. New York: Pitman, 1985.
- [7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [8] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 153–158, Feb. 1997.
- [9] B. Lerner, S. Dhanjal, and M. A. Hultén, "GELFISH—Graphical environment for labeling FISH images," *J. Microsc.*, vol. 203, pp. 258–268, Sept. 2001.
- [10] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- [11] B. Lerner and N. D. Lawrence, "A comparison of state-of-the-art classification techniques with application to cytogenetics," *Neural Comput.*, vol. 10, pp. 39–47, 2001.
- [12] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, P. Besnard and S. Hanks, Eds. San Mateo, CA: Morgan Kaufmann, 1995, pp. 338–345.



Boaz Lerner received the B.A. degree from the Hebrew University, Jerusalem, Israel, in 1982, the M.Sc. degree from Tel-Aviv University, Tel-Aviv, Israel, in 1988, and the Ph.D. degree from Ben-Gurion University, Beer-Sheva, Israel, in 1996.

He performed research at the Neural Computing Research Group at Aston University, Birmingham, U.K., and the Computer Laboratory of the University of Cambridge, Cambridge, U.K. In September 2000, he joined the Department of Electrical and Computer Engineering at Ben-Gurion University, as a Lecturer.

His research focuses on machine learning approaches to data analysis with applications to “real-world” (especially medical and biological) problems. His current interests include neural networks, statistical pattern recognition, and Bayesian networks for image classification, segmentation, and interpretation, the Bayesian methodology to inference and learning, and automation of image-based biomedical applications.



William F. Clocksin received the B.A. degree from Evergreen College, Olympia, WA, in 1976 and the Ph.D. degree from Cambridge University, Cambridge, U.K., in 1994.

After pursuing research in robot vision at Edinburgh University, U.K., from 1976 to 1980, and at Oxford University, U.K., from 1980 to 1983, he joined the Computer Laboratory at Cambridge University in 1983, where he is currently a Lecturer. His research interests include computer vision and image processing, intelligent factory automation,

and a social constructionist framework for artificial intelligence.

Seema Dhanja is currently pursuing the Ph.D. degree under the supervision of Prof. M. Hulten in the Department of Biological Sciences, University of Warwick, Coventry, U.K.



Maj A. Hultén received the M.D. degree in 1966 and the Ph.D. degree in 1974, both from the University of Stockholm, Stockholm, Sweden, and the FR-CPATH degree from the Royal College of Pathologists, London, U.K., in 1992

She is a Professor of medical genetics, Biological Sciences Department, Warwick University, Coventry, U.K. During 1975 to 1977, she was the Clinical Director and Medical Consultant, West Midlands Regional Genetic Services, Heartlands, Hospital, Birmingham, U.K., where she developed a comprehensive

genetics service for a population of around 2.5 million. Her main research interests concerns mechanisms of origin, improved diagnosis, and counseling of genetic disease with special reference to chromosome disorders.



Christopher M. Bishop (M'97) received the B.A. degree in physics from Oxford University, Oxford, U.K., in 1980, and the Ph.D. degree in theoretical physics from the University of Edinburgh, Edinburgh, U.K., in 1983. His thesis concerned quantum field theory.

He joined the Culham Laboratory where he worked on the theory of magnetically confined plasmas as part of the European controlled fusion program. He subsequently developed an interest in pattern recognition, and became Head of the Applied Neurocomputing Center at AEA Technology. In 1993, he was elected to a Chair in the Department of Computer Science and Applied Mathematics at Aston University, Birmingham, U.K. He was Principal Organizer of the six-month international research program on Neural Networks and Machine Learning at the Isaac Newton Institute for Mathematical Sciences, Cambridge, U.K., which ran from July to December 1997. After completion of the Newton Institute Program, he joined the Microsoft Research Laboratory, Cambridge, U.K., where he is currently Assistant Director. His research centers on probabilistic inference and graphical models. His current interests include latent variables, hierarchical generative models for data visualization, and variational methods for inference and learning in probabilistic networks. He has also been elected to a Chair of Computer Science at the University of Edinburgh.

Dr. Bishop is a Fellow of Darwin College, Cambridge.