Technical Report

UCAM-CL-TR-466
ISSN 1476-2986

Number 466

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

Automatic signal classification in fluorescence in-situ hybridization images

Boaz Lerner, William Clocksin, Seema Dhanjal, Maj Hultén, Christipher Bishop

May 1999

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Abstract

Previous systems for dot counting in fluorescence in-situ hybridization (FISH) images have relied on an automatic focusing method for obtaining a clearly defined image. Because signals are distributed in three dimensions within the nucleus and artifacts such as debris and background fluorescence can attract the focusing method, valid signals can be left unfocused or unseen. This leads to dot counting errors, which increase with the number of probes. The approach described here dispenses with automatic focusing, and instead relies on a larger statistical sample of the specimen at a fixed focal plane. Images across the specimen can be obtained in significantly less time if a fixed focal plane is used. A trainable classifier based on a neural network (NN) is used to discriminate between valid and artifact signals represented by a set of features. This improves upon previous classification schemes that are based on non-adaptable decision boundaries and are trained using only examples of valid signals. Trained by examples of valid and artifact signals, three NN classifiers, two of them hierarchical, each achieve between 83% and 87% classification accuracy on unseen data. When data is pre-discriminated in this way, errors in dot counting can be significantly reduced.

Keywords: Auto-focusing; Colour image analysis; Dot counting; Fluorescence in situ hybridization (FISH); Neural networks; Signal classification;

1 Introduction

In recent years, FISH has emerged as one of the most significant new developments in the analysis of human chromosomes. FISH offers numerous advantages compared with conventional cytogenetic techniques since it allows numerical chromosome abnormalities to be detected during normal cell interphase. One of the most important applications of FISH is dot counting, i.e., the enumeration of signals (also called dots or spots) within the nuclei. Dot counting is used for studying numerical chromosomal aberrations in e.g., haematopoietic neoplasia, various solid tumours, prenatal diagnosis and for demonstrating disease-related chromosomal translocations [9].

However, a major limitation of the FISH technique for dot counting is the need to examine large numbers of cells. This is required for an accurate estimation of the distribution of chromosomes over cell population, especially in applications involving a relatively low frequency of abnormal cells. As visual evaluation of large numbers of cells and enumeration of hybridization signals is very tedious, laborious and time-consuming, FISH analysis for dot counting can be expedited by using an automatic procedure [2, 3, 7, 8, 9].

To perform dot counting, an automatic system has to exploit three-dimensional (3D) information of cells contained in the specimen. The system needs an automatic focus control that enables obtaining the 'sharpest' image along the Z-axis, similar to that obtained by manual adjustment of the mi-

croscope stage. Moreover, this mechanism has to be activated for each and every field of view (FOV). Employing an auto-focusing mechanism, however, suffers from a number of problems. First, dedicated hardware and software for the operation of the mechanism are needed [7, 8]. Second, automatic acquisition is dependent upon finding the 'sharpest' image. It can fail however, if the mechanism focuses on a source of noise such as debris or background fluorescence, or if the FOV is empty [7, 8]. Therefore, subsequent manual inspection for discarding such images is sometimes inevitable. Third, even if the 'sharpest' image is indeed found, it can only represent a section of a 3D-image, where signals in other sections which are above or below that section are left unfocused. Fourth, automatic focusing has been found to be time-consuming. Ten seconds are required to complete auto-focusing of one FOV [8], which is around 24 times longer than the time of moving from one FOV to another. And finally, research shows [8] that auto-focusing contributes about 3% of the total 11% error rate of the analysis.

We suggest that FISH dot counting be based on images that are sampled at a fixed focal plane as an alternative to the use of auto-focusing mechanism. This method is motivated by the assumption that nuclei are approximately uniformly distributed in the sample, so that translations at a fixed focal plane will provide a statistically equivalent sample as projections through different focal planes. Images can be captured by any scanning method of the slide, and the microscope stage can stop for collecting images arbitrarily, even at

random. Randomly-captured images in a fixed focal plane 'intersect' nuclei on the slide at random sections, which are equivalent to those encountered by the auto-focusing mechanism. This method enables most of the shortcomings of auto-focusing to be overcome, since it shortens the length of image acquisition and requires no special instrumentation. However, the system needs to acquire sufficient analysable images and to exploit most of the information contained within these images in order to enable dot counting. It may have to deal with more unfocused nuclei and signals, and so its ability to distinguish between focused and unfocused signals should be better than that of a system employing an auto-focusing mechanism. Therefore, the proposed system depends upon two components: a classifier to discriminate between valid and artifact signal data, and well-discriminating features to represent the signals.

Our previous work [5] has investigated the second component of feature representations for FISH signals. In the present work, we study the use of a classifier to discriminate between valid and artifact signals. Focused signals that have characteristics of valid signals are more likely to be classified by the system as 'reals'. Unfocused signals and signals created by background fluorescence or due to overlap between signals of different fluorophores are more likely to be classified as 'artifacts'. A two-layer perceptron neural network (NN) trained using large numbers of examples of these classes is employed for the classification.

Section 2 of the paper describes biological preparation and image acquisition, while Section 3 depicts the applied image analysis stages, namely colour analysis, nuclei and signal segmentations, and signal feature measurement. Section 4 presents a classifier of signals into 'reals' and 'artifacts' of two colours (fluorophores), while the results of applying the classifier to the problem are given in Section 5. Finally, the benefits of an accurate signal classification and its application to dot counting are discussed in Section 6.

2 Biological material and image acquisition

2.1 Slide preparation

The interphase nuclei preparations from amniotic fluid were made using the method by Klinger *et al.* [4] with minor modifications. 1-2ml of amniotic fluid was centrifuged and the cell pellet washed in PBS warmed to 37°C. The cells were resuspended in 75mM Potassium Chloride (KCl) and put directly on to slides coated with APES (Sigma) and incubated at 37°C for 15 minutes. Evaporation of PBS was compensated with filtered distilled water. Excess fluid was carefully removed and replaced with 100ml of 3% Carnoys fixative, 70% 75mM KCl at room temperature for 5 minutes. The excess fluid was carefully removed and 5 drops of fresh fixative were dropped on to 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.

2.2 Hybridization

Target areas were marked on the slides using a diamond tipped scribe. Target DNA was denatured by immersing in 70% formamide:30% 2xSSC at 73°C for 5 minutes. 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 placed over the probe solution. Coverslips were sealed using rubber cement and slides placed in a pre-warmed humidified container in a 37°C incubator for 16 hours. Coverslips were removed and slides washed in 0.4xSSc/0.3%NP-40 solution at 73°C for 2 minutes. Slides were then placed in 2xSSC/0.1% NP-40 solution at room temperature for 1 minute. When completely dried 10 μL of DAPI II counterstain (Vysis UK) was applied to the target area and sealed under a coverslip.

2.3 Instrumentation and screening procedure

Slides were screened under a Zeiss axioplan epifluorescence microscope using x100 objective. Signals were viewed using appropriate filters and images acquired using a CCD camera and SmartCapture software (Vysis UK). Slides were scanned by starting in the upper left corner of the coverslip and moving

from top to bottom. Images were captured by stopping at random intervals. Red and green signals were seen on blue DAPI stained nuclei, corresponding to chromosomes 21 and 13 respectively. The focus and colour ratios were adjusted for the first captured image from each slide, and then kept at those values for all the following images from that particular slide. A total of 400 images were collected from five slides and stored in TIFF format.

3 Image analysis

3.1 Colour analysis

Multiple probes, labelled by different fluorophores, are often used in conjunction in FISH preparation. For example, in the present study, chromosomes 13 and 21 are indicated by green and red signals, respectively, while the nuclei are coloured in blue (Figure 1a). The position in the image and the characteristics of each of these fluorophores have significant meaning to the researcher or clinician. Nevertheless, in most of the previous research of automatic FISH image analysis (see e.g. [7, 8]), and regardless of whether a monochromatic or a colour camera is being used, colour information is converted into gray-level scale. FISH image analysis is then based on brightness contrast and not on colour information, which is lost in the process. However, much of the difficulties which are encountered during the analysis of

intensity-based FISH images can be avoided if colour information is maintained and used. This is especially true for nuclei and signal segmentations. Many user-defined thresholds and heuristics are needed to segment signals from nuclei and nuclei from background, when intensity-based analysis is employed. Colour image analysis does not only facilitate pre-processing and segmentation [3], but it also yields hue-based features, which are found very efficient for FISH signal representation and classification [5].

In this work, colour is kept and specified by the RGB (red, green, blue) format, where each image pixel is represented by the normalized red, green and blue brightness values. Nuclei are analysed in the blue channel of the RGB image, whereas red (chromosome 21) and green (chromosome 13) signals are analysed separately in the red and green channels, respectively. The HSI (hue, saturation, intensity) colour format has been also used when measuring hue-based features to represent signals, as in previous research [5].

3.2 Colour image segmentation

Special multi-stage (usually TopHat-based) procedures that rely on heuristically-derived thresholds and parameters are conventionally employed to segment nuclei and signals [7, 8]. Colour 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. Finding 'optimal'

thresholds is almost trivial compared with thresholding an intensity image since only blue (red, green) objects are found in the blue (red, green) channel.

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. A signal whose area is larger than 5% of the area of the corresponding nucleus is rejected as 'background fluorescence'. Finally, since our interest in this work is to study FISH signal classification, we allow the system to accept signals of nuclei of irregular shape or which are part of a cluster. Such nuclei, as well as unfocused nuclei, can be rejected for the purpose of dot counting based on their size, shape and intensity [8].

3.3 Signal feature measurement

A few features are measured for each of the candidates for signals segmented from the RGB image. Features that are measured include area (a size measure), eccentricity (a shape measure), total and average intensities (intensity measures) and intensity standard deviation (texture measure). All but the last feature have been suggested previously [7] to represent signals, albeit measured using the intensity image. We also measure the maximum and average hue (colour measures) as they are more appropriate for signal dis-

crimination than RGB-based features [5]. Hue features can be measured only if colour information is kept, and the RGB image is then converted into HSI format. A few features are found to be very representative for signal classification when evaluated using scatter plots, probability density functions, a class separability criterion and the probability of misclassification [5]. Among these features are the area, average intensity (RGB) and average hue (HSI) of the signals. Therefore, these three features are measured here from the segmented signals and compose the signal patterns to be classified.

4 Signal classification

The main purpose of this work is to investigate the feasibility of automatic signal classification in randomly-captured FISH images. Although the application of the research is mainly in dot counting, we are not interested here in estimating the proportions of cells having specific numbers of signals, but rather in the ability to accurately distinguish between valid signals ('reals') and artifacts. This ability, if proven, will form the basis of a dot counter.

In the common procedure for automatic dot counting, signals whose relative intensity and either total intensity [7] or area [8] are in specific intervals are classified as 'reals', while other signals are rejected. The interval is defined by the minimum and maximum values of the features as measured on a training set composed of valid signals. Such a strategy is not appropriate for

the methodology we are suggesting here for a number of reasons. First, even the 'best' one (or two) discriminative features would fail to provide sufficient classification accuracy when signals have to be classified as 'reals' or 'artifacts' of one of a few fluorophores [5]. Dealing with a complex multi-class classification problem usually requires the use of multi-variate patterns. Second, as the training set includes only valid signals, the classifier is limited in its ability to model artifacts. Therefore, it may 'miss' those decision boundaries between the classes which yield the minimum probability of misclassification. Third, as the decision boundaries are determined by the minimum and maximum feature values, they are only a rough approximation of the real decision boundaries determined by feature values of the entire training data set. Moreover, these boundaries are sensitive to outliers. In the presence of outliers, the probability density functions of valid signals and artifacts may overlap to a greater extent and the probability of misclassification of the classifier may then be increased.

Therefore, the classification procedure proposed here is as follows. Three-dimensional-patterns of signals (and in another study [5] higher-dimensional-patterns), which are based on the signal area, average intensity (RGB) and average hue (HSI) (Section 3.3) are examined. The patterns are classified into four classes– 'real red', 'artifact red', 'real green' and 'artifact green'. Within the 'artifact' classes we expect to find mostly 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 valid signals, and hence will be classified as artifacts (Figure 1b). 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 environment for labelling FISH images [6].

Before performing each classification experiment, outliers (around 3% or less of the data) are automatically removed from the data and the features are then normalized 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 [1, pp. 374–375]. In the variant of cross-validation technique which is used in this work, the data is partitioned into five equal parts, where 4/5 of the data are used for training and the remaining 1/5 are kept for the test. The experiment is repeated five times where in each time another 4/5 (1/5) of the data are employed for the training (testing). Classification accuracy is then averaged over the five experiments (CV-5). The classifier is a two-layer perceptron NN [1, Ch. 4] trained by the scaled conjugate gradient algorithm [1, pp. 282–285]. Classification is based on the approximation of the multi-layer perceptron outputs to the *a posteriori* probabilities for the classes. A validation set which is drawn from the training set assures that the classifier is not over-trained. It also allows the selection of a minimal network configuration based on only a few hidden units. Both factors ensure

rapid training and improved generalisation.

Three classification strategies are examined here. In the first, called the 'monolithic strategy', patterns are classified into the four classes using a single NN. In the second, termed the 'independent strategy', patterns are classified into 'red' and 'green' classes using the 'colour network' and independently by a second network, the 'real network', into reals and artifacts. Classification of a pattern into one of 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 'colour 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 colours.

5 Results

Before beginning the experiments, we established a database of 400 FISH images, which were randomly-captured from five slides. Following nuclei segmentation, the system identified 944 objects within these images as nuclei, of which 613 also contained 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 (Section 4), 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).

First, experiments to find suitable configurations for classifiers of each of the strategies were performed. Input and output dimensions for the networks were set by the feature space dimension and the number of classes, respectively. The number of hidden units is determined such that the network has the highest generalisation capability. This was achieved by evaluating networks of different numbers of hidden units on an independent validation set [1, p. 372] drawn from the training set. The network which had the lowest error measured on the validation set was selected for training. Figure 2 shows the results of experiments with the 'monolithic' and the 'combined' strategies for determining the number of hidden units for each network, and therefore their configurations. Table 1 (first row) gives the configurations selected for the networks of each of the classification strategies, where the number of hidden units is selected by the highest classification accuracy on the validation set. Finally, training of each of the networks was continued for 100 epochs (presentations of the entire training set), and the results were averaged for each network over three random initialisations.

The classification accuracy for the 'monolithic' strategy, using its optimal configuration, was 84.0% and 82.9% for the training and test sets, respectively (Table 1 first column). We have examined the sensitivity of the classification accuracy of this strategy against the sample size by repeating the experiment for training sets of different sizes. The size of the training

set is increased from 10% to 90% of the data, where the same unseen 10% of the data are used for the test. The results in Fig. 3 demonstrate that the classification accuracy on the test set follows, as expected, the increase of the training sample size until its maximum level. However, the classification accuracy on the training set has a minimum. The explanation is that for a very small sample size, training is very simple and classification of a few training patterns can be very accurate. It is, however, more difficult to maintain this accuracy as the sample size increases and more variants of the training patterns are added. The classification accuracy, hence, decreases until it reaches a minimum for a 'critical mass' of learned patterns. After this point, as sample size continues to grow, the additional patterns are not so different from those of the 'critical mass'. Thus, learning of the patterns of the (extended) 'critical mass' is intensified, while at the same time the fraction of misclassified patterns becomes lower. The result of both trends is towards the improvement of the classification accuracy on the training set as is shown in Fig. 3.

Experiments with the other two strategies, the 'independent' and the 'combined', reveal that these strategies can improve the classification accuracy of the 'monolithic' strategy by 0.4% (to 83.3%) and 4.2% (to 87.1%), respectively, when tested on unseen data (Table 1). The table also demonstrates that classification of signals into their colours is more successful than that of signals into 'reals' and 'artifacts'. Finally, the 'combined' strategy,

when tested using an extended feature set, has achieved classification accuracy of 89.2% [5].

6 Discussion

Usually, the application of an auto-focusing mechanism to the acquisition of FISH images enables the analysis of nuclei and signals from focused images. However, the distributions of signals within a nucleus and nuclei within a specimen are uniform. Because of this, and the fact that the auto-focusing mechanism can focus on debris and background fluorescence, signals are often left unfocused. Consequently, missing images are analysed, and dot counting suffers from errors. Moreover, this flaw is enhanced significantly as the number of probes, and therefore signals, increases. In addition, auto-focusing requires special hardware and software and a large fraction of the analysis period. Around 50% to 75% of the total time needed for analysing a specific FOV is devoted to auto-focusing [8], where eventually all but one of the captured and analysed images are discarded. In summary, auto-focusing is a critical and time-consuming step of FISH image analysis that upon failure will undermine the whole analysis and will lead to unreliable results [8].

An alternative methodology is proposed in this work, which is not limited to well-focused images. The methodology is applied to randomly-captured images, and hence makes the use of an expensive auto-focusing mechanism

redundant. Unlike a methodology that is based on an auto-focusing mechanism, all the randomly-captured images are utilized. These images provide enough examples of focused and unfocused (and other artifact) signals which are necessary for training a classifier to accurately discriminate between valid signals and artifacts. When the system is later tested on unseen images of cells, a nucleus can be rejected automatically if it is unsuitable for the analysis (due to, e.g., overlap or irregular shape), or if signal classification demonstrates that the nucleus contains artifacts.

Our work has aimed to study the accuracy of automatic signal classification in FISH images. Signals are classified regardless of the quality of the image or the corresponding nucleus. Inference by the classifier is probability-driven, where decision is based on the maximum *a posteriori* probability. The probability framework enables the exploitation of the maximum discrimination information about the decision boundaries between the classes to be classified. Nevertheless, the proposed acquisition methodology provides the flexibility to use also other inference frameworks such as decision rules or trees.

Processing of colour in FISH images makes the segmentation and classification of nuclei and signals easier and less sensitive to artifacts and noise. Furthermore, it eliminates the employment of user-defined thresholds and prevents the excessive application of heuristics, both of which are common in intensity-based analysis. Consequently, colour-based FISH image analy-

sis reduces the dependence upon environmental conditions and preparation techniques, and thus provides a more generic technique. It also suggests an improvement to current methods [7, 8] since it is suitable for use with more than two fluorophores.

Classification experiments in this work have revealed that NN-based hierarchical classification strategies, besides shortening training sessions compared with a monolithic strategy, also decrease the classification complexity, and thus improve the classification accuracy. Signals are classified as valid signals and artifacts with an accuracy of 87.1% (or an accuracy of 89.2% for other feature sets [5]). This accuracy can be considered a promising result when classifying multi-colour signals of not necessarily single nuclei in not necessarily focused images. This is especially significant when compared to the state-of-the-art accuracy of 89.3%, which has been achieved on single-colour signals of single nuclei in well-focused images [8].

Combined with an accompanying probabilistic framework (e.g., [2, 3]), this accuracy can improve the estimation of the proportions of cells having different numbers of signals (0,1,2,...), and thereby lead to precise dot counting. As previous research has shown [8], an average of 11% of the cells are counted incorrectly. About 6% of these errors are due to the detection algorithm [7], whereas the remaining 94% are caused by wrong classification of the detected signals. We believe that by employing the classification methodology described here, these errors can be significantly reduced. Cur-

rent research is aimed at proving this.

Acknowledgements: This work is supported by EPSRC contract GR/L51072: *Automatic Analysis of FISH Images.*

References

[1] C. M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

[2] A. D. Carothers. Counting, measuring, and mapping in FISH-labelled cells: Sample size considerations and implications for automation. *Cytometry*, 16:298–304, 1994.

[3] K. R. Castleman and B. S. White. Dot count proportion estimation in FISH specimens. *Bioimaging*, 3:88–93, 1995.

[4] 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 hybridisation (FISH). *Am. J. Hum. Genet*, 51:55–65, 1992.

[5] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hultén, and C. M. Bishop. Feature representation for the automatic analysis of fluorescence in-situ

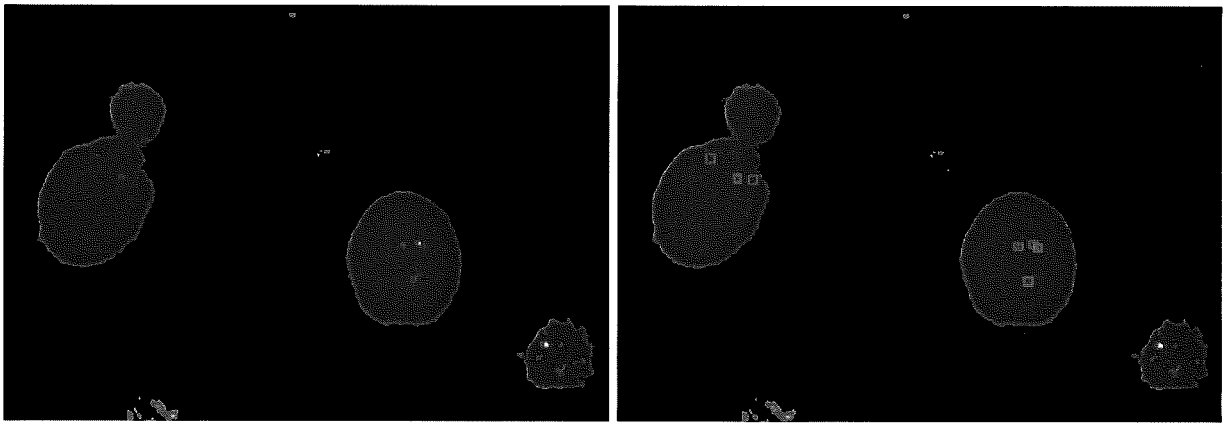
hybridization images. Technical Report 464, Computer Laboratory, University of Cambridge, May 1999.

[6] B. Lerner, S. Dhanjal, and M. A. Hultén. GELFISH- graphical environment for labelling FISH images. Technical Report 465, Computer Laboratory, University of Cambridge, May 1999.

[7] H. Netten, L. J. van Vliet, H. Vrolijk, W. C. R. Sloos, H. J. Tanke, and I. T. Young. Fluorescent dot counting in interphase cell nuclei. *Bioimaging*, 4:93–106, 1996.

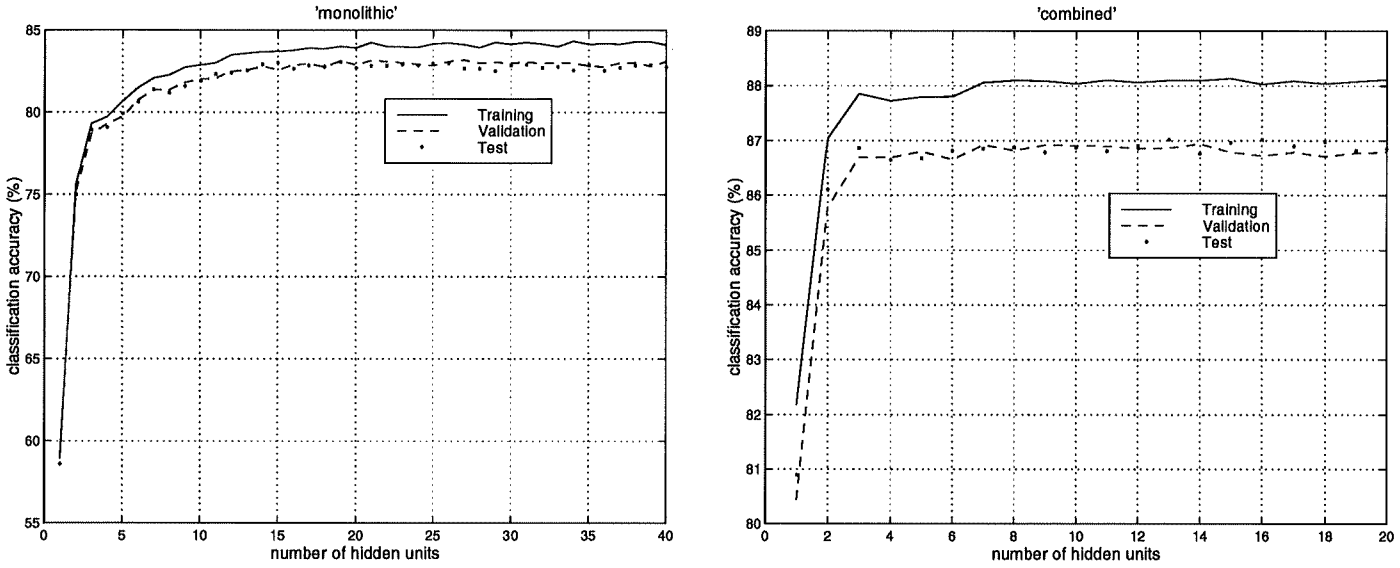
[8] H. Netten, I. T. Young, L. J. van Vliet, H. J. Tanke, H. Vroljik, and W. C. R. Sloos. FISH and chips: Automation of fluorescent dot counting in interphase cell nuclei. *Cytometry*, 28:1–10, 1997.

[9] H. J. Tanke, R. J. Florijn, J. Wiegant, A. K. Raap, and J. Vrolijk. CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization. *Histochemical Journal*, 27:4–14, 1995.



(a) (b)

Figure 1: (a) An example of an image used for FISH signal classification. (b) Valid signals ('reals') in Fig. 1a, as labelled by a cytogeneticist using a graphical interface [6], are marked by squares in corresponding colours. All other (unfocused and 'background fluorescence') signals are considered, and therefore labelled, 'artifacts'. These labels are used to train and evaluate the signal classifier.



(a) (b)

Figure 2: Classification accuracy of the (a) 'monolithic' and (b) 'combined' strategies for increasing numbers of hidden units (notice the different scales along the two y-axes). (The slight deviations in classification accuracy of the 'combined' classifier for 13 hidden units compared with Table 1 is attributed to the different experiments using random classifier initialisations and randomly selected data sets).

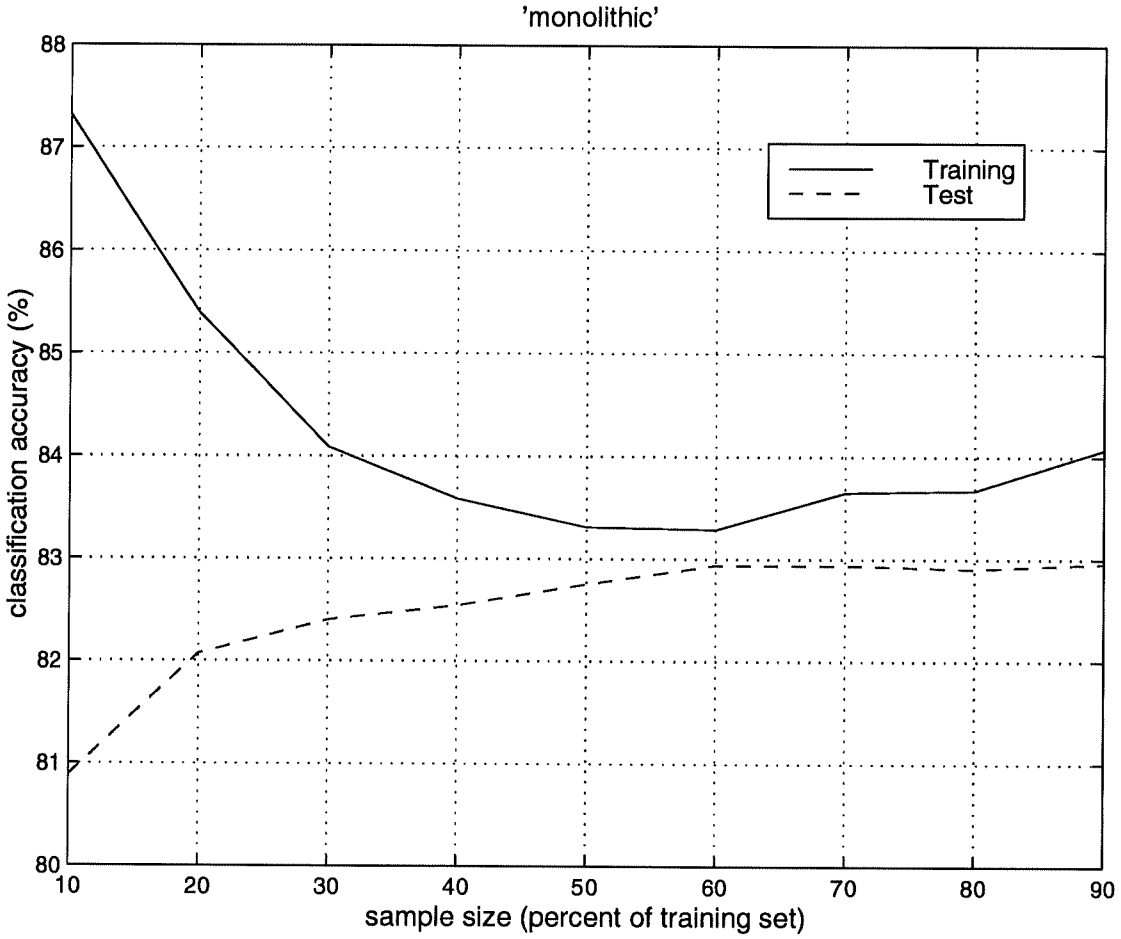


Figure 3: Classification accuracy for the 'monolithic' strategy for increasing sample sizes.

Table 1: Configurations as well as classification accuracies on the training and test sets of the three classification strategies– 'monolithic', 'independent' and 'combined'. Configurations are specified by the numbers of units in each layer of the network (input:hidden:output). Results for both the 'real' and the 'colour' networks are needed to obtain the overall classification accuracies of the 'independent' and 'combined' strategies (Section 4).

	'monolithic'	'real'	'colour'	'independent'	'combined'
Configuration	3:27:4	3:13:1	3:13:1	3:13:1	3:13:1
Training (%)	84.0	87.5	96.4	84.1	87.9
Test (%)	82.9	87.3	95.7	83.3	87.1