Pixel-by-Pixel Classification of MFISH Images

Mehul P. Sampat[1], K.R. Castleman[2], A.C. Bovik[1]

[1]The University of Texas at Austin,TX 78712-1084, USA

[2]Advanced Digital Imaging Research,TX,USA

Abstract— **Multiplex Fluorescence In-Situ Hybridization (M-FISH) is a recently developed chromosome imaging method in which each chromosome is labelled with 5 fluors (dyes) and is also counterstained with DAPI. This paper proposes an automatic pixel by pixel classification algorithm for M-FISH images using a Bayes Classifier. The M-FISH pixel classification was approached as a 25 class 6 feature pattern recognition problem. The classifier was trained and tested on non-overlapping data sets and an overall classification accuracy of 95% was obtained.**

Keywords— **M-FISH, Bayes Classifier.**

I. Introduction

Images of chromosomes contain vital information about the health of a human being. Multiplex fluorescence in situ hybridization is a recently developed technology for chromosome analysis. In this technique chromosomes are dyed with multiple dyes so that each chromosome class appears to be a distinct color. In comparison to conventional chromosome analysis, superior detection of subtle and complex chromosomal rearrangements can be achieved with M-FISH analysis. M-FISH images are captured with a fluorescent microscope. To view each of the fluorescent dyes multiple optical filters are used. Each of the dyes are visible in different optical wavelengths referred to as spectral channels. Thus a set of M-FISH images can be viewed as a multi-spectral set. A typical M-FISH set consists of 6 images. Hence each pixel can be viewed as a 6 feature vector in a 6 dimensional space where each element represents the response of the dye. Traditionally, segmentation and classification are carried out as two separate sequential steps. We have combined these two steps into a single step, to classify the pixels into 24 chromosome types (22 autosomes + 2 sex chromosomes). In Sections II and III the Bayes Classifier and the methodology are discussed respectively. In Section IV the results are presented.

II. Bayes Classifier

Suppose we wish to classify objects into N different classes based on certain measured features of the object. If we measure d features of each object then each object can be described by a feature vector which has d components. Each object can also be viewed as a point in a d dimensional space called the *feature space*. All possible d dimensional feature vectors reside in this space. Let x denote a feature vector. Let $P(w_i)$ denote the probability that an object belongs to class w_i where i varies from 1 to N, for a N class classification problem. This is called the *a priori* probability. Note that $\sum_i P(w_i) = 1$. Let $p(x|w_i)$

denote the class-conditional probability distribution function. It represents the probability distribution function for a feature vector x given that x belongs to class w_i. The joint probability distribution function of an object belonging to class w_i and having feature vector x is $p(w_i, x)$. Let $P(w_i|x)$ be the *posteriori* probability which is the probability that the object belongs to class w_i given the feature vector x. Given $P(w_i)$ and $p(x|w_i)$, the *a posteriori* probability for an object represented by the feature vector x is given by the Bayes theorem [2]:

$$P(w_i|x) = \frac{p(x|w_i)P(w_i)}{p(x)} \quad (1)$$

where $p(x) = \sum_{i=1}^{N} p(x|w_i)P(w_i)$. $P(w_i|x)$ represents the probability that an object described by the feature vector x, belongs to the class w_i. The theorem is applicable for all probability density functions however depending on the nature of the data, the Gaussian density function is often used to model the distribution of feature values of a particular class. The general multivariate Gaussian density function in d dimensions is given by:

$$p(x) = \frac{1}{(2\pi)^{d/2} |\sum|^{1/2}} exp\left[-\frac{1}{2}(x-\mu)^t \sum^{-1}(x-\mu)\right] \quad (2)$$

where x is a d component feature vector, μ is the d component mean vector, $\sum$ is the $d \times d$ covariance matrix, $|\sum|$ and $\sum^{-1}$ are it's determinant and inverse respectively. Also $(x-\mu)^t$ denotes the transpose of $(x-\mu)$. During the training phase, the mean vector μ, and the $d \times d$ covariance matrix $\sum$ are calculated for each class from the training data. To classify an object described by the feature vector x, we calculate $P(w_i|x)$ for each class i. The class to which the object belongs is given by the Bayes Decision Rule which is,

$$decide\ w_i\ if\ P(w_i|x) > P(w_j|x) \forall\ j \neq i \quad (3)$$

If certain classification mistakes are more expensive than others then a $N \times N$ cost matrix C can be used to account for this. The elements of the cost matrix, $C_{i,j}$ represent the cost of assigning an object to class i when it actually belongs to class j. For example if all errors are equally costly, then

$$C_{i,j} = 1\ for\ j \neq i\ and\ C_{i,j} = 0\ for\ j = i \quad (4)$$

Then the risk of assigning an object to class i is given by $R_i = \sum_{j=1}^{N} C_{i,j}P(w_j|x)$.

III. Methodology

Our objective is to classify human chromosomes. The 46 human chromosomes consist of 22 pairs of similar, homologous and 2 sex-determinative chromosomes. These represent 24 classes. The background was included as a class. Since an M-FISH set consists of 6 images, each pixel can be represented by six features, which are the gray-scale values

(a) Actual Classmap

(b) Computed Classmap

Fig. 1

RESULTS

in the six color channels (five labels and DAPI counter-stain). Therefore, a 6-feature 25 class Bayes Classifier was used to do a pixel by pixel classification. The images for training and testing were selected from a public database of hand segmented M-FISH images. This database is made available online by Advanced Digital Imaging Research at:*http://www.adires.com/projects/mfish_db.shtml.* For each set of M-FISH images the database also contains a labelled classmap image in which each pixel is labelled according to the class to which it belongs to. Such a class-map is show in Fig 1(a). In this figure a separate color is used for each class. The classifier was trained on a set of 6 M-FISH images. The mean vector and the 6×6 covariance matrix for each class was calculated from the training data. The classifier was tested on different sets of MFISH images where each set consisted of 6 images. To remove the background fluorescence from each image, a two-dimensional cubic function was fitted to a set of background points from the image and then subtracted from each image[3].
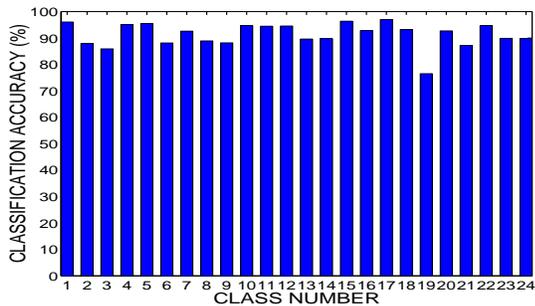
Fig. 2

AVERAGE CLASSIFICATION ACCURACY

IV. RESULTS

The classification was done for 5 sets of M-FISH images and a classmap was generated by the classifier for each set. The classifier accuracy was calculated by comparing the calculated class-map to the actual class-map provided in the database. One such pair of the actual classmap and the calculated class-map is shown in Fig.1(a) and Fig.1(b) respectively. The average overall classification accuracy obtained was 95.694%. The overall classification of non-background pixels was 91.428%. The overall classification accuracy of the background pixels was 96.004%. Fig. 2 shows the average classification accuracy for each of the chromosome classes.

V. CONCLUSION

An automatic pixel-by-pixel classification method for M-FISH images was implemented and tested. Promising results were obtained with the preliminary tests. The classification accuracy can be improved by adjusting the values of the cost matrix as classifying a pixel belonging to a particular class incorrectly is more expensive than classifying a background pixel incorrectly. The confusion matrix can be studied to see which errors were most frequent and the costs for making those errors can be increased. Also, the dimensionality of the data can be reduced by using Principal Component Analysis [2] and the transformed data could now be the input to the classifier.

REFERENCES

[1] M. R. Dpeicher, S.G.Ballard, and D.C.Ward, "Karyotyping human chromosomes by combinatorial multi-fluor fish," *Nature Genetics*, vol. 12, pp. 368–375, 1996.
[2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* San Diego: Harcourt Brace Jovanovich, Second ed., November 2000.
[3] K. R. Castleman, *Digital Image Processing.* Prentice-Hall, 1996.
[4] M. M. Le Beau, "One fish, Two fish, Red fish, Blue fish," *Nature Genetics*, vol. 12, pp. 341–344, 1996.
[5] K. Castleman, T. P. Riopka, and Q. Wu, "Fish Image Analysis," *IEEE Engineering in Medicine and Biology*, vol. 15, pp. 67–75, 1996.
[6] W. Schwartzkopf, B. L. Evans, and A. C. Bovik, "Minimum entropy segmentation applied to multi-spectral chromosome images," *Proc. IEEE Int. Conf. on Image Processing*, vol. II, pp. 865–868, Oct 2001.