CELL IMAGE CLASSIFICATION USING HISTOGRAMS, HIGHER ORDER STATISTICS AND ADABOOST

Vinod Chandran, Jasmine Banks, Wageeh Boles, Brenden Chen, Inmaculada Tomeo-Reyes

Queensland University of Technology, Brisbane, Queensland, Australia.

ABSTRACT

A cell classification algorithm that uses first, second and third order statistics of pixel intensity distributions over pre-defined regions is implemented and evaluated. A cell image is segmented into 6 regions extending from a boundary layer to an inner circle. First, second and third order statistical features are extracted from histograms of pixel intensities in these regions. Third order statistical features used are one-dimensional bispectral invariants [1]. 108 features were considered as candidates for Adaboost [2] based fusion. The best 10 stage fused classifier was selected for each class and a decision tree constructed for the 6-class problem. The classifier is robust, accurate and fast by design.

***Index Terms*—** Cell, Classification, Histogram, Bispectrum, Higher order statistics, Adaboost

1. RATIONALE OF THE METHOD

An algorithm for automated cell classification is designed with the following considerations: (a) it must utilize texture information selectively from different parts, (b) it must be robust to rotation and noise, (c) it must capture third order statistical information from pixel intensity producing many features and (d) an optimal fused classifier should be generated for a given problem based on training data. The most original aspect of the methodology is the use of bispectral invariant features [1] from histograms to achieve the desired robustness and accuracy with support from region selection and feature selection using Adaboost [2].

2. IMAGE PRE-PROCESSING

It is assumed that individual cell images have been separated and a cell mask image is available. Based on the cell mask, each cell is divided into regions extending from a boundary layer to an inner circle. 6 regions are used in the system described here (listed in Table 1). They are not all disjoint. A histogram of pixel intensities is computed for each region.

3. FEATURE EXTRACTION

18 features are extracted from each region – brightness (first order), contrast (second order) and 16 one-dimensional bispectral invariants [1]. These features are designed to be robust to brightness and contrast changes. Histograms of the regions are invariant to rotation of the cell image by design and consequently, the features are as well. Bispectral invariant [1] features are angles and lie in the range between $[-\pi, \pi)$ and the other features are normalized to this range. 108 features were used to train classifiers for each of the HEp-2 cell classes using Adaboost [2]. The number of selected features by their order and the region of the cell they come from are given in table 1.

Class	Order			Region					
	1	2	3	Outer most	Next outer	Inner disc	Inner ring	Small disc	Inner most
HOM	4		6	4	3	1		1	1
CEN	4		6	5		1	1	1	2
NUC	4	1	5	3	3	2		2	
SPE	7	2	1	4	2	2	1	1	
NUM	5		5	4	4	1		1	
GOL	4		6	2	2	5		1	

Table 1 Number of selected features by order and region

It can be observed that third order statistical features are important for all classes and that the outermost regions close to the cell boundary contain significant discriminative information. The best feature (first stage) was third order for 3 of the classes and first order for the other three.

4. CLASSIFICATION

Adaboost [2] is used to generate 10 stage binary classifiers. They are combined in a decision tree. 60% of the ICIP HEp-2 cell classification competition training data were used for training. The best binary classifier was 91% accurate and the overall accuracy for 6 classes was 49.1% on all data. The order of complexity of the algorithm is roughly $O(N^2) + O(RFQ \log_2 Q)$ where N^2 is the number of pixels in the cell, R is the number of regions, F is the number of bispectral features and Q is the number of intensity levels. For training $R = 6, F = 16, Q = 256$ and for testing $RF = 10$.

5. REFERENCES

[1] V. Chandran and S. L. Elgar, "Pattern Recognition using Invariants Defined from Higher Order Spectra – One Dimensional Inputs," *IEEE Trans. Signal Processing*, vol. 41, pp. 205-212, Jan 1993.

[2] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm, in 13th Intl. Conf. Artificial Intelligence, 1996.