Multiple Sequential Thresholds Technique in Automated White Blood Cells Classification

BAO Han-fei (盒苍飞)
Department of Histo-embryology, Tongji Medical University, Wuhan
E.S. Gelsema, H.C. den Harink, A.W.M. Smeulders
Department of Medical informatics, Free University, Amsterdam, The Netherlands

Summary: This article describes a novel approach to the problem of automated white blood cell classification. Whereas in most earlier attempts, the segmentation of the cells has been recognized as the most difficult and most critical step in the sequence of operations, resulting in the classification, the method described here eliminates the necessity of the detection of the contour of the nucleus and of the cytoplasm, and is therefore less sensitive to such disturbing factors as the presence of granules, or other cells touching the cell of interest, etc.

The multiple sequential threshold method to be described here in two slightly different variants yields a correct classification rate of 94.7% for a 4 class problem (90 cells in the test set), and 91.8% for an 8 class problem (279 cells in the test set). Both experiments include immature cell types.

Key words: white blood cells, segmentation, automated cytology, pattern recognition

Research on the automation of the white blood cell classification and, more recently, on the recognition of other (immature) white blood cells has led in the early and mid seventies to the construction by various manufacturers of white blood cell differential machines for clinical routine use. Gelsema E.S and Landeweerd G.H have reviewed the early research efforts in this area[12].

In most of the image processing procedures used in the research work and in the automatic machines, the segmentation of the image, i.e. the separation of the cell from the background and of the nucleus from the cytoplasm has been identified as the most difficult part of the process. The final outcome, i.e. the determination of the cell type, depends critically on the correct localization of the cell con-
tours and of the border of the nucleus. Many ingenious techniques for this segmentation process have been invented and applied\(^{22}\), but still the failure rate of this process has severely limited the correct classification rate.

In principle, of course, in cases of improper segmentation, the cell could be disregarded. This, however, assumes that the segmentation process is able to evaluate its own performance and, moreover, that the failure rate is independent of the cell type. Both these assumptions are not true in practice.

For these reasons it was decided to experiment with a recognition procedure designed to be independent of the localization of the boundaries corresponding to the biological constituents of the cell, i.e., the cytoplasm and the nucleus. In the following, the method will be outlined and preliminary results will be given in terms of correct recognition rates in a case of four cell types and in another case of eight cell types. In these two experiments, slightly different variants of the multiple sequential threshold technique were used.

MATERIAL AND METHODS

1. The cytological material on which these experiments are based consists of 21 smears of peripheral blood taken from 17 leukemia patients. The slides were prepared at the Department of Haematology of the Academic Hospital of Groningen, the Netherlands. After spinning of the slides to obtain a monolayer of cells, also minimizing the occurrences of touching cells, the slides were stained with a May-Grünwald-Giemsa stain, as is also used in routine practice.

2. The equipment on which the image processing part of the experiment was done consists of a Leitz Orthoplan microscope, using a 100× oil immersion objective. The images were registered by an Eyecom Image Scanner, model 700 SV (Spatial Data Systems, Inc.). The video signal was digitized to a 512×512 image array, with a densitometric resolution of 256 levels. For these experimental conditions, the nominal spatial resolution amounts to 0.1 μm per pixel. Image processing was performed on an IBAS II system (Kontron Bildanalyse GmbH), which consists of 8 image memories, each of 512×512×8 bits, an array processor for fast image manipulations, driven by a Z-80 computer, running under the CPM operating system. A special procedure was designed for this purpose, consisting of parts for cell isolation, segmentation, feature extraction and output.

These feature files were then analyzed, using the system for Interactive Statistical Pattern Recognition and Analysis, ISPAHAN\(^{33}\). The final performance figures in the form of confusion matrices were calculated using feature evaluation procedures and classifiers available in ISPAHAN.

3. The segmentation method and feature definition

The usual way of segmenting a cell image concentrates on finding contours in the digitized image, corresponding as well as possible to the border of the cell and the border of the nucleus in the cell. Methods may range from simple global thresholding in the grey level histogram through adaptive (local) thresholding, to more sophisticated methods using heuristic search techniques\(^{12}\).

The methods to be described here are two variants of segmentation through multiple sequential thresholding.

1) The first segmentation method

In the first variant it is assumed that one can find a fixed threshold \(t(0)\), and that segmentation at \(t(0)\) corresponds roughly to the isolation of the nucleus in some types of cells. One may then define a sequence of