A Water Immersion Algorithm for Cytological Image Segmentation

Pascal Bamford         Brian Lovell
Cooperative Research Center for Sensor Signal and Information Processing, Department of Electrical and Computer Engineering, University of Queensland, Australia.

Abstract - A non-parametric multi-resolution method is presented for cell segmentation and some of the advantages it has over other techniques, when applied to this application, are outlined. The proposed algorithm is an extension on the method of Wilson and Spann [6] and uses a technique based on the concept of water immersion which ensures closed boundaries. Experiments with cervical cells have shown that the algorithm is robust, fast and accurate.

1 Introduction

The ratio of the area of the cytoplasm to the area of the nucleus (CN) is one of the best discriminatory features for the classification of cells in the screening for cancer of the cervix. The larger the ratio, the more likely the cell is to be dysplastic (pre-cancerous) or even cancerous. This ratio naturally depends upon an accurate segmentation of the cytoplasm from the background and of the nucleus from the cytoplasm. Unfortunately, cytological images present difficult segmentation problems such as overlapping cells, clumped groups (sometimes several cells in depth), out of focus cell boundaries and edges that fade into the background [1].

Traditionally, cell segmentation is performed by the use of thresholding or edge detection [3]. However, thresholding tends to cause voids in the resulting image that must be filled by some means as a post-processing step and edge detectors tend to produce incomplete boundaries and amplify image noise. Morphological methods of segmentation include the technique of flooding to find the watersheds of a gradient image [7]. Inner and outer markers are required as starting points for flooding and watersheds are found as the zones of confluence from these markers. However, the generation of the inner markers is not straightforward in this application. It is possible to have areas inside the cell that have a steeper gradient than the actual cytoplasm boundary and unsuitable inner markers will cause these areas of large gradient to be found, rather than the required cell boundary. With the use of further morphological techniques the inner markers can be improved, producing reasonable results. However, this approach was found to be rather computationally expensive. The algorithm described in this paper not only produces a rapid and robust method of finding the initial markers, but also achieves an accuracy that enables a very simple boundary re-estimation stage.

A great deal of interest lies in the analysis of nuclear texture, especially with respect to malignancy associated changes (MACs) [5]. The markers produced by this algorithm provide an excellent mask from which nuclei may be simply extracted for analysis.

This research is being performed in conjunction with the Royal Womens Hospital (Brisbane) who have made available 5000 slides prepared using the ThinPrep® technique to allow extensive testing.

2 Method

The method developed by Wilson and Spann [6][8] consists of three stages:

- Quadtree Smoothing.
- Lowest level classification.
- Boundary re-estimation.

2.1 Quadtree Smoothing

The quadtree approach is a multi-resolution segmentation process. Each block of four pixels of the original image is assigned to a parent node. This parent node then assumes a value equal to the average of its children, effectively smoothing the image. These parent nodes are recursively subjected
to the same procedure, each new set of nodes forming a smaller image. This procedure is equivalent to viewing the image at a series of lower magnifications (figures 1 and 2). The quadtree has the effect of smoothing the image so that small background artefacts, such as blood cells in this case, become less significant relative to the objects of interest. As the resolution is changed, spatial resolution is traded for class resolution. The aim is to choose the magnification where the class resolution for the object of interest is the highest. Usually, this can be determined from the object size a priori. The quadtree structure is shown in figure 3.

2.2 Lowest Level Classification

Once the quadtree containing the various representations of the image has been realised an initial classification needs to be made at the lowest resolution. The grayscale images may be visualised
as topographical maps where the cells become hills above a relatively smooth plain. One could imagine that the image surface is lowered into ‘water’ so that ‘air’ pockets become trapped in the hills of the map. These air pockets become cell markers. Occasionally, it is possible that small areas in the background may also develop pockets of air due to a residual unevenness in the surface. Frequently these can be discarded by simple post-processing operations, such as discriminating on the basis of area. The method inherently ensures closed contours and therefore greatly facilitates the boundary re-estimation stage. Below is the pseudo-code for the algorithm.

Start: Every \( p_{i,j} = \text{UNBOUND} \)

FOR waterlevel = 0 to 255
FOR all \( p_{i,j} \) over image
  IF \( p_{i,j} = \) waterlevel and \( p_{i,j} = \text{UNBOUND} \)
    THEN \( p_{i,j} = \text{WATER} \)
ENDFOR
IF (\( p_{i,j} \) on edge of image) SET \( p_{i,j} = \text{FLAG} \)
ITERATE OVER IMAGE
  IF \( p_{i,j} = \text{UNBOUND} \) and neighbour = \text{FLAG} 
    THEN \( p_{i,j} = \text{FLAG} \)
UNTIL NO CHANGE
FOR all \( p_{i,j} \) over image
  IF \( p_{i,j} = \text{UNBOUND} \) THEN \( p_{i,j} = \text{BOUND} \)
  IF \( p_{i,j} = \text{FLAG} \) THEN \( p_{i,j} = \text{UNBOUND} \)
ENDFOR
ENDFOR

Upon completion of the algorithm, all pixels labelled BOUND are classified as object and pixels labelled UNBOUND are classified as background. Figure 4 shows the boundaries of the markers (objects) found during the water immersion process, superimposed upon the original image. It can be seen that at this level a good mask has been generated for ‘areas of interest’. In an automated screener, this mask would be used by the microscope controller to ‘zoom in,’ enabling the efficient capture of high resolution images from which feature extraction would be performed.

2.3 Boundary Re-estimation

At this stage, each pixel at the lowest magnification has been assigned to a particular class (object or background) and has been given that classification based on the water immersion algorithm. Any pixel that has a nearest neighbour (8-connectivity) in a different region is then classified as a boundary pixel. The classification is then propagated back down the tree. This means that only pixels whose parents were boundary pixels need to be reclassified at the lower levels. This is done by performing a re-immersion upon these pixels in order to re-define the border at the next level of the quadtree. Figure 5 shows the final segmented image after boundary re-estimation. Notice that despite the two areas in the top right of the image being treated as one by the initial marker, they become separated during the re-estimation. This shows how the immersion algorithm overcomes one of the problems of the use of quadtrees. Because the quadtree reduces the size of the image by averaging pixels together, regions of close proximity will often become merged. If these regions are preserved during the remaining processes, an under-segmentation will result. However, because each level is re-immersed here, merged regions become split and correct segmentation results.

3 Results

Tests upon many captured images have shown that a fast and accurate segmentation of cells is achieved by this method. Although the iterative technique of water immersion is fairly slow, this does not present a serious problem as only a small number of pixels are operated upon at the top level of the quadtree and during the boundary re-estimation. The use of the immersion method to re-estimate the border has several advantages over other schemes tested. Faster methods, such as Lloyd-Max quantization [2], can result in noisy edges from the choice of imperfect thresholds. Additional processing is required in this case in order to estimate where the true boundary lies through these noisy areas. Water immersion, however, has the property of finding a continuous edge and any ‘islands’ formed outside of the main object body can simply be discarded.
Figure 4: Original image with the boundaries of the lowest level markers superimposed.

Figure 5: Final segmentation after boundary re-estimation.

4 Conclusions

A fast and robust cell segmentation method is presented. The algorithm is particularly suited to implementation in a cytological screener, as areas of interest on the slide are quickly identified at low-resolution. This represents a very large saving in the processing time of each slide as only these areas need to be analysed further. This segmentation scheme currently forms the basis of an experimental cytological software package. Coupled with a frame grabber, camera and microscope, it provides an environment in which a user may ana-
lyse samples by scanning slides and selecting cells of interest. The user would then be presented with the statistics and even a preliminary classification of each cell. Figure 6 shows the program user interface.

5 Acknowledgement

The authors would like to acknowledge the contribution of Mark Jeacocke for early development of this algorithm[4].

References


