MDL BASED INTERPRETATION FOR OVERLAPPING CELL NUCLEI IN HISTOLOGICAL IMAGES

Jenni Hukkanen[1]*, Edmond Sabo*[2]*, and Ioan Tabus*[1]

[1] Department of Signal Processing, Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, FINLAND, Jenni.Hukkanen@tut.fi
[2] Department of Pathology, Rappaport Faculty of Medicine, Technion,
Haifa, Israel, e_sabo@rambam.health.gov.il

ABSTRACT

In this extended abstract we present our current work concerning the development of minimum description length (MDL) principle based solutions for medical image segmentation problems. Our application is the interpretation of overlapping nuclei in histological images in terms of individual nuclei within a parametric shape family. MDL is a well suited approach for this application as MDL provides an efficient tool for comparing various competing geometrical structures composed of different numbers of elliptical shapes, each shape representing one nucleus. Our MDL criterion, developed for solving the problem of overlapping objects, involves different cost terms than the existing MDL criteria for image segmentation and in addition our resulting description is fully implementable, since it does not use asymptotic expressions for the involved codelengths. In experiments we have compared the results of automatic segmentation and human subject segmentations including the segmentation given by an expert pathologist.

1. INTRODUCTION

Histological images are 2D images taken from thin slices of tissue samples. They provide important information to expert pathologists for medical diagnosis and evaluation of the grade of the disease. The staining used in histological images is typically hematoxylin and eosin (H&E) staining, which highlights nuclei of the cells in the image. Unfortunately, the thickness of the tissue slice is in practice higher than a single nuclei layer, which results into clumps of overlapping nuclei in the two-dimensional image. Hence, the nuclei segmentation algorithm should not only segment well separated individual nuclei, but also should separate overlapping and occluding nuclei into individual ones.

Traditional segmentation algorithms are intended only for producing binary segmentation results, where the problem of overlapping nuclei is not resolved. There exist algorithms for splitting clumps of nuclei from binary segmentation. Unfortunately, binary segmentation results for H&E stained histological images can be noisy and unreliable.

The shape of the nuclei is in many cases almost elliptical and hence in our approach we interpret clumps of nuclei as unions of elliptical shapes such that each ellipse represents one nucleus. Our recently proposed SNEF algorithm [1] is an ellipse fitting based algorithm for cell nuclei segmentation. It has a number of options, which may lead to different ellipse proposals for the interpretation. In [2, 3] we proposed a minimum description length (MDL) principle [4] based criterion for comparison between different interpretations involving different numbers and arrangements of ellipses.

2. METHODS

MDL provides a tool for comparison between different statistical models representing geometrical structures, since by using the MDL criterion one can choose that description, which best explains the data, for a given data set and class of models, providing a natural trade-off between the complexity of the model and the fitting of the data.

The MDL for image segmentation was first introduced in Leclerc [5]. The main idea in [5] and our approach [2, 3] is the lossless description of the image to be segmented using a total codelength $L(Y, \Omega, \beta)$, involving the following terms:

$$L(Y, \Omega, \beta) = L(\Omega) + L(\beta|\Omega) + L(Y|\Omega, \beta),$$

where $L(\Omega)$ is the codelength for describing the contour, which splits the image into foreground and backgound, $L(\beta|\Omega)$ is the cost of describing the coding parameters in each of the regions, and $L(Y|\Omega, \beta)$ is the codelength for encoding the image given the contour and using the coding distributions. Since our problem is not the segmentation, but interpreting a region of possibly overlapping nuclei by ellipses, the resulting costs for the contour have a different form than in [5]. We use Golomb-Rice codes for encoding the residuals, which is known to be efficient in the field of lossless image coding and provides a fully implementable coding algorithm.

In order to minimize locally the MDL criterion we have introduced an iterative algorithm for updating the parameters of ellipses [2].

3. RESULTS AND CONCLUSIONS

We run the SNEF algorithm for each image four times, using different thresholds for binarization of gradients and intensity values, resulting in four alternative segmentations and then we compute for each segmentation the value of MDL criterion. For each image the best SNEF segmentation is taken as the one with the lowest value of MDL criterion. We compared the obtained best SNEF segmentation results to the segmentations provided by human subjects. The subjects were allowed to give multiple interpretations for each image and also to specify their belief towards the interpretation. The ground truth was computed as a weighted average of the subjects interpretations.

We found in [2] that our local iterative algorithm, which optimizes the parameters of ellipses in order to minimize the MDL, decreases the variability in the MDL value of the provided human interpretations. We also noticed that after the iterative algorithm the deviations of the MDL values obtained by the best SNEF with respect to the MDL of the ground truth are in general lower than two times the standard deviation of MDL for the human segmentations.

One can use before the SNEF algorithm different preprocessing stages, which can effectively protect against the artifacts in the original images and in [3] we studied the effects of preprocessing using smoothing by Gaussian filtering and rescaling at various downsizing scales. We noticed a good correlation between the highest MDL and the highest (supervised) similarity index, which measures the overlapping areas between the ground truth and the provided segmentation.

As a conclusion, the proposed MDL based criterion for comparison between different interpretations of cell nuclei in histological images offers a good selection tool, matching closely supervised criteria, which will require segmentations provided by human subjects.

4. REFERENCES

[1] J. Hukkanen, A. Hategan, E. Sabo, and I. Tabus, "Segmentation of cell nuclei from histological images by ellipse fitting," in *Proc. of the European Signal Processing Conference*, Aalborg, Denmark, August 2010.

[2] J. Hukkanen, E. Sabo, and I. Tabus, "Representing clumps of cell nuclei as unions of elliptic shapes by using the MDL principle," in *Proc. of the European Signal Processing Conference*, Barcelona, Spain, August 2011.

[3] J. Hukkanen, E. Sabo, and I. Tabus, "MDL based structure selection of union of ellipse models at multiple scale descriptions," in *Proc. of the Second Symposium on Interdisciplinary Approaches in Fractal Analysis*, Bucharest, Romania, May 2011.

[4] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[5] Y.G. Leclerc, "Constructing simple stable descriptions for image partitioning," *International Journal of Computer Vision*, vol. 3, pp. 73–102, 1989.