

# A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei

Petter Ranefall<sup>a,\*</sup>, Lars Egevad<sup>b</sup>, Bo Nordin<sup>a</sup> and Ewert Bengtsson<sup>a</sup>

<sup>a</sup>*Centre for Image Analysis, Uppsala University, Sweden*

<sup>b</sup>*Department of Pathology, Uppsala University, Sweden*

Received 29 November 1996

Revised 1 July 1997

Accepted 7 October 1997

**Abstract.** A new method for segmenting images of immunohistochemically stained cell nuclei is presented. The aim is to distinguish between cell nuclei with a positive staining reaction and other cell nuclei, and to make it possible to quantify the reaction. First, a new supervised algorithm for creating a pixel classifier is applied to an image that is typical for the sample. The training phase of the classifier is very user friendly since only a few typical pixels for each class need to be selected. The classifier is robust in that it is non-parametric and has a built-in metric that adapts to the colour space. After the training the classifier can be applied to all images from the same staining session. Then, all pixels classified as belonging to nuclei of cells are grouped into individual nuclei through a watershed segmentation and connected component labelling algorithm. This algorithm also separates touching nuclei. Finally, the nuclei are classified according to their fraction of positive pixels.

**Keywords:** Colour image segmentation, immunohistochemical staining, pixelwise classification, relaxation, watershed segmentation, automated cell counting

## 1. Introduction

Quantification of immunohistochemistry is usually performed either subjectively according to a semiquantitative grading or by cell counting. The former has a poor reproducibility and the latter takes a lot of time. Both methods depend on the skills and mood of the pathologist. The purpose of this work is to develop an objective and stable computerised method for quantification. The proposed procedure initially produces by a pixelwise classification. It is shown that the area percentage of positive staining is highly correlated to the number of positively stained nuclei [4] and in many applications it is enough to compute the area percentage. When the staining is good and the conditions are controlled this could be computed by a method with fixed thresholds in the colour space [13]. But when the images are difficult, a more general classifier is needed, e.g., a Maximum Likelihood (ML)

---

\*Correspondence to: Petter Ranefall, Centre for Image Analysis, Lägerhyddvägen 17, S-752 37 Uppsala, Sweden. Tel.: +46 18 471 78 49; Fax: +46 18 55 34 47; E-mail: petter@cb.uu.se.

classifier [5]. In this paper we propose a classification method with a performance comparable to the ML classifier, but less time consuming.

In some applications it is necessary to be able to compute features for each individual nucleus, e.g., staining intensity, area, etc. To do this we suggest some postprocessing of the pixelwise classification, in order to separate the nuclei and to label them as either positively or negatively stained. In order to do this it is most important that the pixelwise classification is good.

## 2. Materials and methods

### 2.1. Hardware

The  $756 \times 572$  pixels colour images with  $3 \times 256$  grey levels were grabbed by a Sony DXC-151 colour video camera attached to a standard Olympus BH-10 optical microscope, using magnification  $40\times$ . This results in a pixel size of about  $0.4 \mu\text{m}$ . The Rayleigh resolution criterion [3] gives a resolution limit of  $0.24 \mu\text{m}$  for a wavelength of  $550 \text{ nm}$  and a numerical aperture of  $0.7$ . We are thus not fully resolving the images but our application is not concerned with details of the nuclear texture. And a larger field of view was considered more important than maximum resolution.

Example of such an image is shown in Fig. 1.

### 2.2. Overview of the method

The procedure for the work is based upon four steps on two different levels. First, all pixels are classified according to colour by a new supervised algorithm [9]. Then each pixel is changed to the most frequent class in its neighbourhood. At this stage we can only count the number of pixels for the different classes and compute the area proportions. The next level is the object level where we first group pixels together and separate touching cell nuclei. Then each cell nucleus is assigned to the class that was the most frequent one among its pixels.

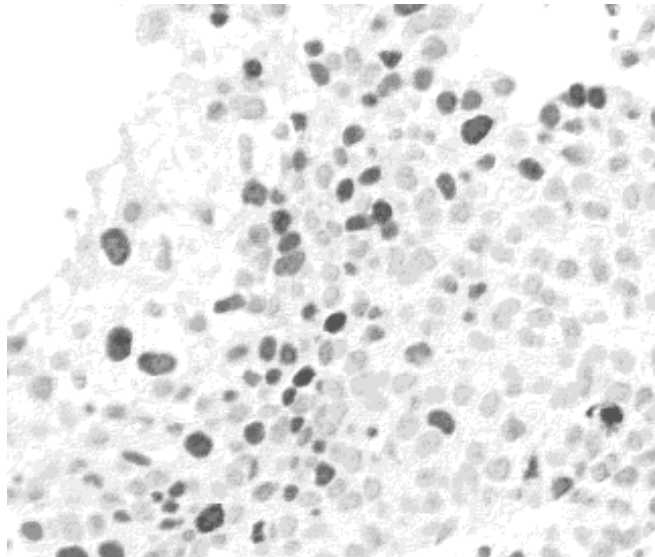


Fig. 1. Typical image of immunohistochemically stained cell nuclei.

### 2.3. Classification

The classification step is initialised by supervised training of a classifier [9], performed by the operator marking some reference pixels for the desired classes, using the mouse, on an image that is representative for the entire staining session. Here we do have some degree of subjectivity. Then a classifier is created by hierarchically splitting the colour space containing the colours from the training image. The colour space is split perpendicularly to one of the axes according to a criterion that maximises the difference between the spread within the sub space before the split and the sum of the spread within the new sub spaces after the split [7,9]. This criterion preserves clusters in the colour space. This procedure is continued until no subspace contains reference pixels belonging to different classes. The reference pixels are only guiding and the class limits are drawn based on the training image. This removes some of the subjectivity from the training phase.

The result of this training is a box classifier [1] which could be applied to any image. Therefore this training only needs to be performed once, as long as the images are from the same staining session, and the illumination is kept constant.

When the classifier is applied to an image each pixel is classified (see example in Fig. 2 and detail study in Fig. 3), but we will not have any knowledge about objects in the image.

### 2.4. Relaxation

Relaxation is then applied, i.e., each pixel is changed to the most common class in its neighbourhood. We have used a  $3 \times 3$  neighbourhood, but larger neighbourhoods are also possible. In this step we are thus bringing some contextual information into the classification. If the classification was reasonable the result of the relaxation will be that small holes in the objects are filled and single misclassified pixels are corrected, but still without any *a priori* knowledge about cell nuclei. See detail study in Fig. 4.

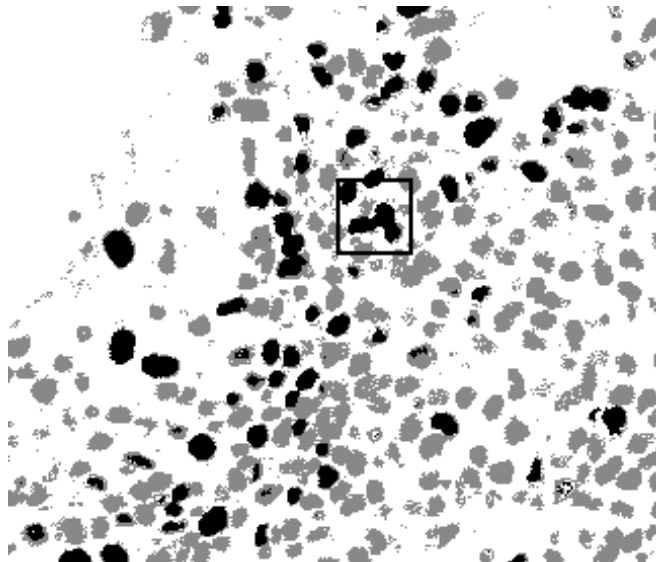


Fig. 2. Result when a classifier is applied to Fig. 1. Positive pixels are dark, negative pixels are light grey and background pixels are white.

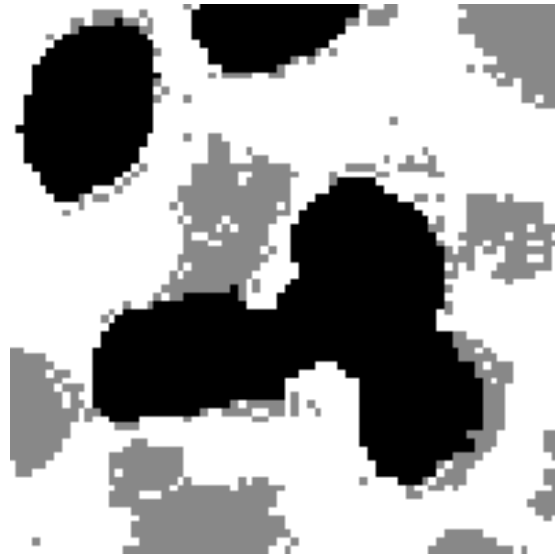


Fig. 3. Detail study of area marked by square in Fig. 2.



Fig. 4. Figure 3 after relaxation.

### *2.5. Watershed segmentation*

Now we move on to consider each connected set of pixels classified as objects as one or several touching cell nuclei. If they are touching we wish to separate them. As a preparation step before the separation a distance image is created by coding each pixel belonging to a cell nucleus with a label that represents the distance to the background. We have used the Chamfer 3–4 distance transform [2] (see detail in Fig. 5). Here the distances are represented as equidistant lines in an image that looks like heights on an orienteering map. But if we regard the distance image as holes



Fig. 5. Distance transformation of Fig. 4.

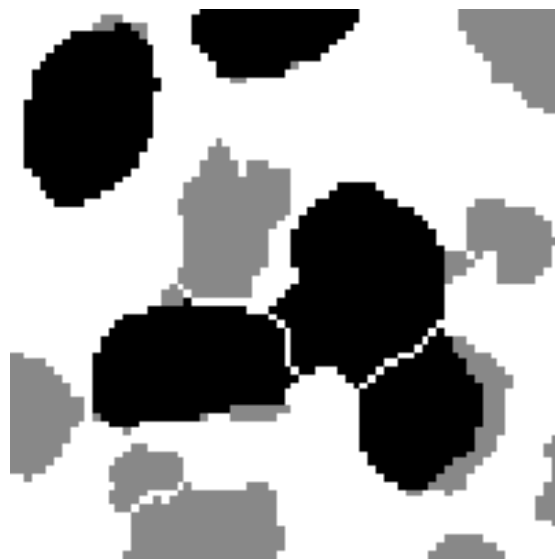


Fig. 6. Separated objects.

instead of heights, then we can imagine that we fill these holes with water until the water from two touching holes meet. There a dam is built which is the separation line between the objects. Here a threshold for the proportion between the radius at the “dam” and the maximum radius of the object, is used to decide whether to split the object or not. We have used the threshold of 80%, but this parameter should be trimmed for each different application. The result is an image with objects separated into individual cell nuclei (see detail in Fig. 6). This algorithm is called watershed segmentation [8,11].

### 2.6. Homogenisation

The next step is to homogenise the objects by giving all pixels in the object the colour that was its most common colour in the classification step. See detail in Fig. 7 and the final result in Fig. 8.

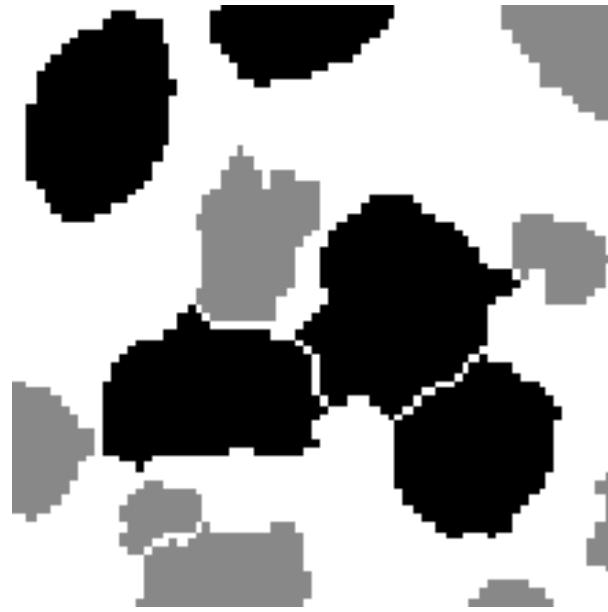


Fig. 7. After homogenisation, each cell nucleus is either completely positive (black) or negative (grey).

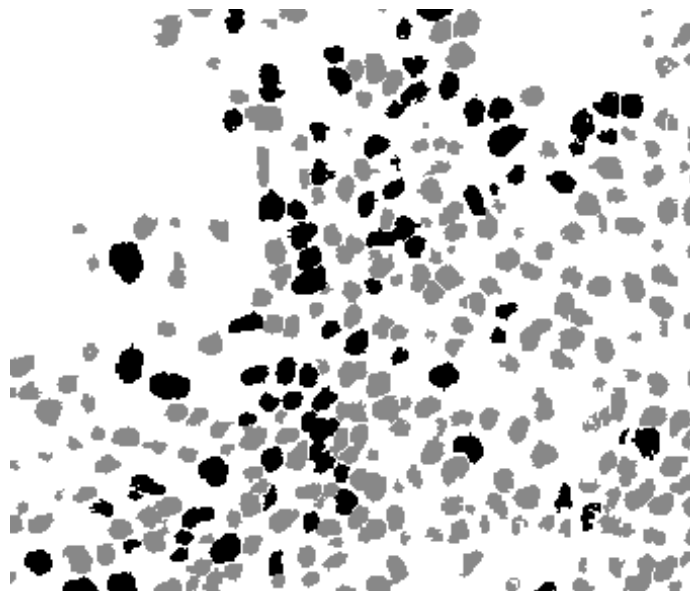


Fig. 8. Final result.

### 3. Results

#### 3.1. Evaluation of classification method

The proposed classification method has successfully been used extensively in a biomedical application to classify colour images of immunochemically stained human cell nuclei from bladder carcinoma. Sections, 5  $\mu\text{m}$  thick, of formalin fixed paraffin embedded bladder cancer were immunostained with the monoclonal antibody MIB-1. MIB-1 is specific for the proliferation related nuclear antigen Ki67. Since chromogen DAB was used, the aim was to distinguish between cell nuclei with a positive staining reaction (brownish), cell nuclei without such a reaction (bluish) and the rest of the tissue (see Fig. 1). Variations in staining reaction between different staining batches have caused any methods based on fixed thresholds in colour space to fail. We thus needed an adaptive method and this algorithm has served its purpose in a satisfactory way. The pathologists are of the opinion that the results agrees with their subjective visual estimations and the algorithm is currently being used in several large application studies. Some preliminary results have been presented at conferences [6,10,12].

The proposed segmentation algorithm works well for the images studied in our application in the sense that a pathologist will accept the images as reasonably correctly segmented; Figs 1 and 2 shows an example of this. We thus know that the results are “reasonable”. Still there may be considerable variation along the borders of the cell nuclei between different versions of a classifier.

In order to evaluate the performance of the algorithm we have tried the approach of interactively creating templates for true pixelwise classification but the results have been non conclusive. Subjective evaluations have also shown considerable variation. For this study we have chosen to compare our classifier with the well established Maximum Likelihood classifiers with respect to classification speed and average colour distance. We used the following test scheme: Bladder carcinoma specimen from 10 patients were stained in the same staining session. One image was selected from each case. On each of these images, training areas for three different classes were drawn: background, positive cell nuclei, negative cell nuclei. For each class there were three training areas, each with size approximately corresponding to one cell nucleus. These three areas were subjectively chosen to represent dark, medium and light objects in the class. From each image a classifier was derived by the proposed method and by the ML method. In the ML case four different assumptions about the covariance matrices were tested. This resulted in  $(1 + 4) \cdot 10 = 50$  different classifiers that were applied to the 10 images (see example in Fig. 2).

Since a reasonable criterion for a good colour based segmentation algorithm is that it creates compact regions in colour space for each class, one possible measure of the quality of a particular segmentation is the average distance from the cluster centres in colour space for all pixels in the image. We will call this measure “average colour distance”. (Note that this measure has been called “average colour error” by other authors when used in colour image segmentation, e.g., [14].) Table 1 shows the average colour distance for the different types of classifiers. Note that none of these differences are statistically significant.

Tables 2 and 3 shows the efficiency of the different classification methods for the training and the classification, respectively, when computed on a 233 MHz DEC Alpha 2000 workstation. In these tables  $x$ ,  $y$ ,  $p$  denotes the dimensions of the image and the number of layers, respectively, and  $n$  is the number of classes. The differences in training and classification times between the proposed method and the ML methods are significant, which is also the case for the differences between “ML general” and the rest.

Table 1  
Average colour distance in pixel units

Classifier	Average colour distance
ML (mindist)	22.06
ML (uncorrelated)	21.82
ML (equal covar)	21.40
ML (general)	26.13
Proposed method	24.60
Do subsampled $4 \times 4$	24.30

Table 2  
Time complexity for training the different classifiers (excluding the time for inter-actively selecting the training pixels)

Classifier	CPU-time (s)	Theoretically
ML (mindist)	0.15	$O(xyp)$
ML (uncorrelated)	0.15	$O(xyp)$
ML (equal covar)	0.15	$O(xyp)$
ML (general)	0.15	$O(xyp)$
Proposed method	2.88	$O(x^2y^2p)$ worst case (when $n = xy$ ) $O(xyp)$ best case
Do subsampled $4 \times 4$	2.23	$O(x^2y^2p)$ worst case (when $n = xy$ ) $O(xyp)$ best case

Table 3  
Time complexity for carrying out the classification of one image

Classifier	CPU-time (s)	Theoretically
ML (mindist)	1.45	$O(xynp)$
ML (uncorrelated)	1.45	$O(xynp)$
ML (equal covar)	1.45	$O(xynp)$
ML (general)	3.63	$O(xynp^2)$
Proposed method	0.23	$O(xyp)$
Do subsampled $4 \times 4$	0.20	$O(xyp)$

The main advantages with the proposed classification method is that the training has to be performed only once and that the classification that is performed for every new image is fast. In the example above it is sufficient to have three different images to gain time with the proposed method, and if the image is sub-sampled during the training, even more time is saved.

Another advantage is that the behaviour of the classifier is “deterministic”, because every pixel in the training set will be correctly classified (with exception for collisions of course).

### 3.2. Comparison with subjective cell counting

In 12 images from different patients but from the same staining session the nuclei were manually labelled as positive or negative. In each of the images, training areas were selected. Twelve classifiers of each type were created with the method described above and with the general ML method. Then the proposed method based on these classifiers were applied to all 12 images. This resulted in 12 estimations of the number of cells for each image and for each classification method, i.e., a total of 144 estimations for the proposed method and 144 estimations for the ML method. It should be noted that in order not to favour any of the classification methods, we did not allow ourselves to check the result of the classification and redo the training, which, by the way, is a suitable procedure



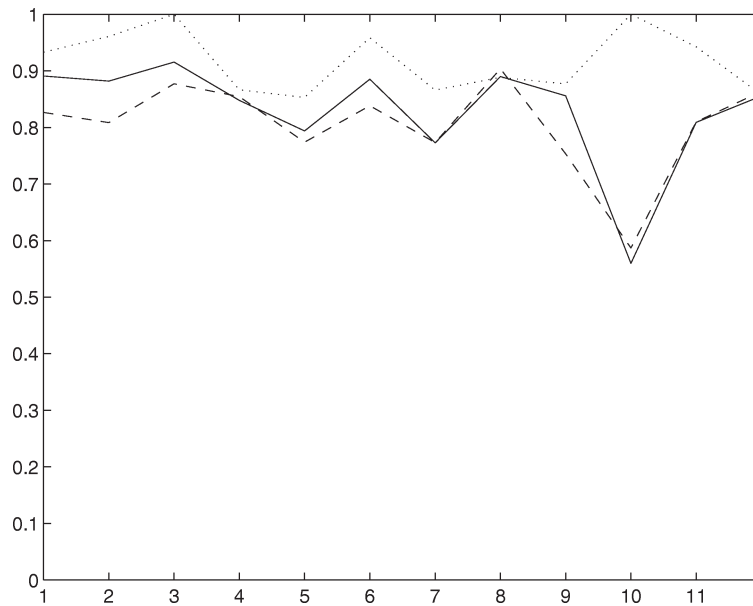


Fig. 9. Average proportion of manually marked nuclei that were classified as being of the same type. The line shows the results for the 12 images when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. The difference between the proposed classifier and the ML classifier is not statistically significant.

to get a good classifier. Therefore the performance of both methods are not “ideal” in this test, but this test gives us a possibility to compare the two methods. To show a more “ideal” behaviour of the proposed method we have also tried the following approach: we selected one of the 12 images as training image and trained on this image, checked the result of the pixelwise classification on this image, and allowed ourselves to improve the training if we were not satisfied. (In the daily routine work 0, 1 or at most 2 alterations are the normal procedure.) This classifier was applied to all 12 images, followed by the rest of the segmentation procedure.

The results from the automatic cell counting based on the classifiers were compared with the manual classification, which served as a “key”. For every classification we checked how many of the manually marked nuclei that were classified as being of the same type. We also checked the other way around, i.e., how many of the classified nuclei that were manually marked as being of the same type. Finally, we compared the proportions of positive nuclei.

Figure 9 shows, for each image, the average proportion of manually marked nuclei that were classified as being of the same type. The line shows the result when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. The difference between the proposed classifier and the ML classifier is not statistically significant.

Figure 10 shows, for each image, the average proportion of classified nuclei that were manually marked as being of the same type. The line shows the result when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. These curves show a slightly worse result than Fig. 9, which indicates that nuclei gets somewhat too fragmented by the watershed method. The difference between the proposed classifier and the ML classifier is not statistically significant.

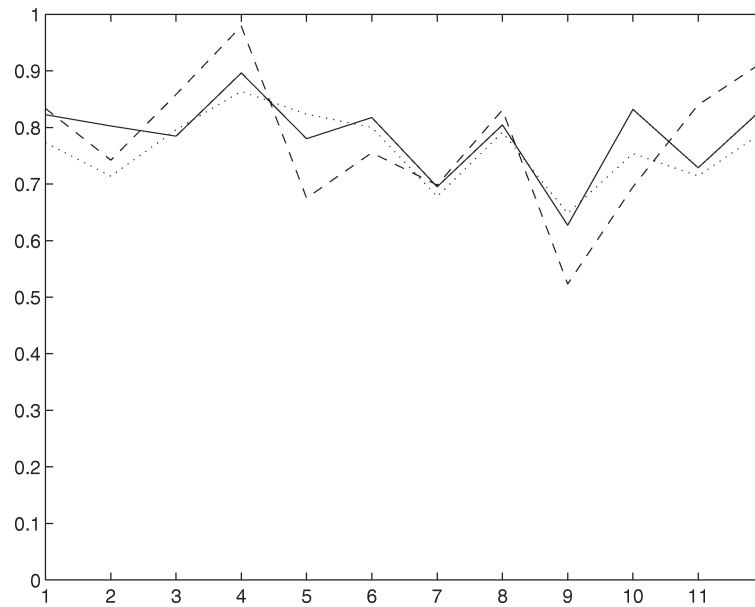


Fig. 10. Average proportion of classified nuclei that were manually marked nucleus of the same type. The line shows the result for the 12 images when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. These curves show a slightly worse result than Fig. 9 which indicates that nuclei gets somewhat too fragmented by the watershed method. The difference between the proposed classifier and the ML classifier is not statistically significant.

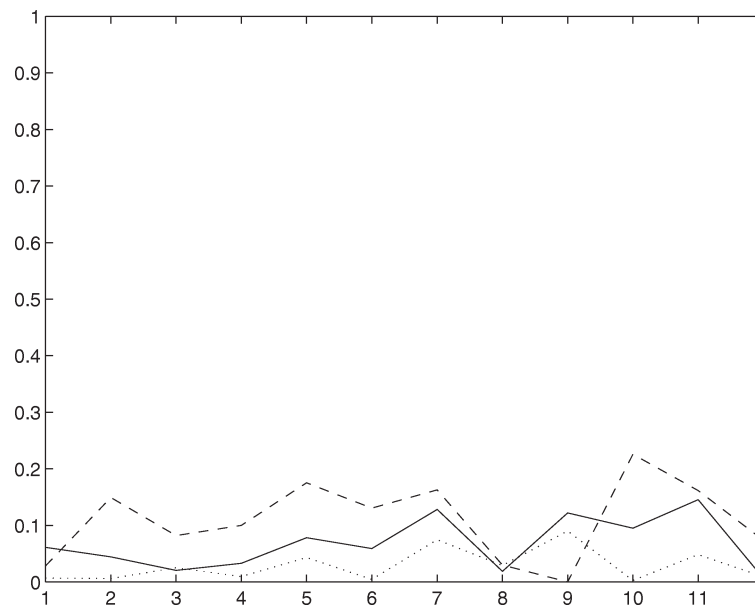


Fig. 11. Absolute value of the difference between the estimated proportions of cell nuclei and the proportions of cell nuclei for the manual "key". The line shows the result when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. Here the difference between the proposed classifier and the ML classifier is statistically significant in favour of the proposed classifier.

Figure 11 shows the absolute value of the difference between the estimated proportions and the proportions for the manual “key”. The line shows the result when the proposed classifier was used, and the dashed line shows the result when the ML classifier was used. The dotted line shows the result for the improved classifier. Here the difference between the proposed classifier and the ML classifier is statistically significant in favour of the proposed classifier.

The conclusion is that the automatic method, and especially with the proposed classifier instead of ML, gives good correlation to the manually marked “key”, but it tends to overestimate the number of cell nuclei, i.e., it splits too many objects. This problem is mainly due to the fact that some objects are fragmented. Suggested solutions: smooth the image with some kind of low-pass filter, use a larger relaxation filter, use a fill-holes algorithm, etc. If the latter is used there might be a problem with false holes, which occurs when three or more cell nuclei are touching and thereby forming a ring structure.

## 4. Discussion

### 4.1. Classification performance

The comparison between our proposed classification algorithm and the classical ML algorithms shows no significant difference in classification accuracy. The images in this material have colour regions that can be expected to produce rather compact, multi-normally distributed clusters, the ideal case for an ML algorithm. The fact that the restricted ML version showed better results than the more general ones is also an indication of this. With more difficult colour regions our algorithm should show a greater advantage relative ML since it does not rely on the multi-normality assumption.

Another advantage in this context is the fact that the behaviour of our classifier is “deterministic” in the sense that every pixel in the training set is correctly classified, with the exception of direct collisions. Concerning the efficiency issue, our results clearly shows the superior speed when it comes to executing the resulting box classifier using our implementation method. We here have a speed advantage of at least a factor of six compared with ML (15 when compared with the general version of ML). This is somewhat offset with a significantly slower training phase. Using sub-sampling in the training phase reduces the time disadvantage with no noticeable penalty in classification accuracy. Still the training takes almost 15 times as long as for ML. So the algorithm has its greatest advantage when several images are to be classified with the same classifier. In our study only three images need to be processed before our method produced a gain in time.

### 4.2. Separation of objects

The separation of objects makes it possible to count cell nuclei and comparisons with subjective cell counting shows good correlation between the results of the proposed method and the subjective method. It is often sufficient to do a pixelwise classification, but one important application for cell nuclei segmentation is computing the intensities for the positive nuclei. This kind of data will be studied more in the near future. Assessment of immunostaining is an important method of quantitative pathology. A major difficulty, however, is the poor reproducibility of subjective grading. The main purpose of the automated classification is to improve the reproducibility. An ongoing study is comparing the reproducibility of automated and subjective assessment of immunostaining.

## Acknowledgements

The authors are very grateful to our colleagues at the Department of Pathology for cooperation and for providing us with the cell images.

## References

- [1] E. Bengtsson, B. Nordin and F. Pedersen, MUSE – a new tool for image segmentation based on multivariate statistics, in: *Computer Methods and Programs in Biomedicine*, Vol. 42, Elsevier, Amsterdam, 1994, pp. 181–200.
- [2] G. Borgefors, Distance transformations in digital images, *Comput. Vision Graph. Image Process.* **34** (1986), 344–371.
- [3] K.R. Castleman, *Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1979, p. 268.
- [4] S. Caulet, C. Lesty, M. Raphael, D. Schoevaert, P. Brousset, J.L. Binet, J. Diebold and G. Delsol, Comparative quantitative study of Ki-67 antibody staining in 78 B and T cell malignant lymphoma (ML) using two image analyser systems, *Pathol. Res. Pract.* **188**(4/5) (1992), 490–496.
- [5] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973, pp. 24–30.
- [6] M. Hellström, S. Brändstedt, M. Häggman and C. Busch, Cytokeratin staining and image analysis of epithelium in relation to stroma in total prostatectomy specimens. A study of the effect of neoadjuvant hormone ablation in localized prostate cancer, *Urol. Res.* **23**(4) (1995), 267.
- [7] O. Milvang, An adaptive algorithm for color image quantization, in: *Proceedings of the 5th Scandinavian Conference on Image Analysis*, Vol. 1, Stockholm, Sweden, 1987, pp. 43–47.
- [8] C. Orbert, E. Bengtsson and B. Nordin, Watershed segmentation of binary images using distance transformations, in: *Proceedings of Image Processing: Nonlinear Image Processing IV*, SPIE 1902, San José, CA, 31 January–5 February 1993, pp. 159–170.
- [9] P. Ranefall, B. Nordin and E. Bengtsson, A new method for creating a pixelwise box classifier for color images, *Machine Graph. Vision* (to appear).
- [10] P. Ranefall, B. Nordin, E. Bengtsson and L. Jönsson, A new method for segmentation of colour images applied to immunohistochemically stained cells, in: *Abstracts of the Third Conference of the European Society for Analytical Cellular Pathology*, Vol. 6, Grenoble, France, 1994, p. 210.
- [11] L. Vincent and P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Machine Intell.* **13**(6) (1991), 583–598.
- [12] K. Wester, K. Engström, P. Ranefall and C. Busch, Chromogen-counterstaining and color based image analysis of immunohistochemistry (IHC), in: *Abstracts of the 22nd Congress of Medical Technology*, Oslo, Norway, 1996.
- [13] F. Willemse, M. Nap, S.C. Henzen-Logmans and H.F. Eggink, Quantification of area percentage of immunohistochemical staining by true color image analysis with application of fixed thresholds, *Anal. Quant. Cytol. Histol.* **16**(5) (1994), 357–364.
- [14] Y.-H. Yang and J. Liu, Multiresolution color image segmentation, *IEEE Trans. Pattern Anal. Machine Intell.* **16** (1994), 689–700.