

Automatic quantification of immunohistochemically stained cell nuclei using unsupervised image analysis

Petter Ranefall^{a,*}, Kenneth Wester^b and Ewert Bengtsson^a

^aCentre for Image Analysis and ^bDepartment of Pathology, Uppsala University, Sweden

Received 16 January 1997

Revised 27 August 1997

Accepted 2 January 1998

Abstract. A method for quantification of images of immunohistochemically stained cell nuclei by computing area proportions is presented. The image is transformed by a principal component transform. The resulting first component image is used to segment the objects from the background using dynamic thresholding of the P^2/A -histogram, where P^2/A is a global roundness measure. Then the image is transformed into principal component hue, defined as the angle around the first principal component. This image is used to segment positive and negative objects. The method is fully automatic and the principal component approach makes it robust with respect to illumination and focus settings. An independent test set consisting of images grabbed with different focus and illumination for each field of view was used to test the method, and the proposed method showed less variation than the intraoperator variation using supervised Maximum Likelihood classification.

Keywords: Image analysis, principal components, automatic quantification, immunohistochemically stained cell nuclei

1. Introduction

Quantification of the proportions of specifically stained regions in images is of significant interest in a growing number of biomedical applications. These applications include histology and cytology where quantification of various stainings performed on histological tissue sections, smears, imprints, etc. is of utmost importance. Through the use of special stains, biological components of interest can be given a specific colour. Qualitatively, this can be evaluated visually as the presence of a specific colour. But to do a quantitative evaluation the number of stained cell nuclei and/or the proportion of specimen area that has been stained needs to be measured. Pure visual estimates of this provide very crude results with poor inter- and intraobserver reproducibility. For this purpose image analysis-based methods have been developed. A very extensive survey of general image segmentation techniques is found in [7], and the problem of quantification of cell images using colour information has been studied for some time by many researchers, e.g., [2,5,10]. In previous studies we developed a method

*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.Ranefall@cb.uu.se.

based on a supervised colour classifier from selected training areas [8,9]. But even with this method, inter- and intraoperator variability is a problem. An objective standardised method for quantifying the proportions of different stained regions is desirable. Such a method should be insensitive to changes in illumination and focus. The approach described in this paper is to normalise the data by making a principal component (PC) transformation of the colour space. The first principal component is used to extract the objects of interest and the principal component hue, defined as the angle around PC1, is used to segment the positive and negative objects.

2. Materials and methods

2.1. Immunohistochemistry

For this study we used images from immunohistochemically (IHC) stained bladder carcinoma and reactive tonsil tissue sections. The primary mouse monoclonal antibody, clone Mib1, recognises a proliferation associated protein localised in the nucleus. A biotinylated secondary, anti-mouse, antibody was applied followed by incubation in peroxidase-labelled streptavidin-biotin complex. We wanted to test different types of staining and therefore, developing was performed in both diaminobenzidine (DAB) and nickel-enhanced DAB (Ni-DAB) on consecutive tissue sections. Developing in DAB gives a dark brown reaction product whereas Ni-DAB results in a black, crisp staining of the nuclei. Harris haematoxylin was used as counterstain when DAB was used as chromogen, staining the IHC-negative nuclei distinctly dark blue. As counterstain to Ni-DAB a light blue staining in Ehrlich's haematoxylin was chosen in order to produce a better contrast to the black positive nuclei.

2.2. Image acquisition

The 756×572 pixel colour images with 3×256 grey levels were grabbed by a Sony DXC-151 colour video camera attached to a standard Olympus BH-10 optical microscope, using a magnification of $40\times$. This results in a pixel size of about $0.4 \mu\text{m}$. The Rayleigh resolution criterion [1] gives a resolution limit of $0.24 \mu\text{m}$ for a wavelength of 550 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 nuclei texture. A larger field of view was considered more important than maximum resolution.

For all images Köhler illumination was maintained and the aperture iris diaphragm ring was fixed to 0.5.

The method was trained on 75 images collected so that for five fields of view, 15 images differing only in illumination and focus were grabbed. Examples of such images are shown in Fig. 1.

2.3. Method

2.3.1. Overview of the method

1. Segment objects from background.
 - (a) Perform a principal component transform of the RGB data and use the first component image.
 - (b) Compute P^2/A -histogram from the PC1 image.
 - (c) Threshold the PC1 image at minimum in P^2/A -histogram.
2. Segment positive and negative objects.

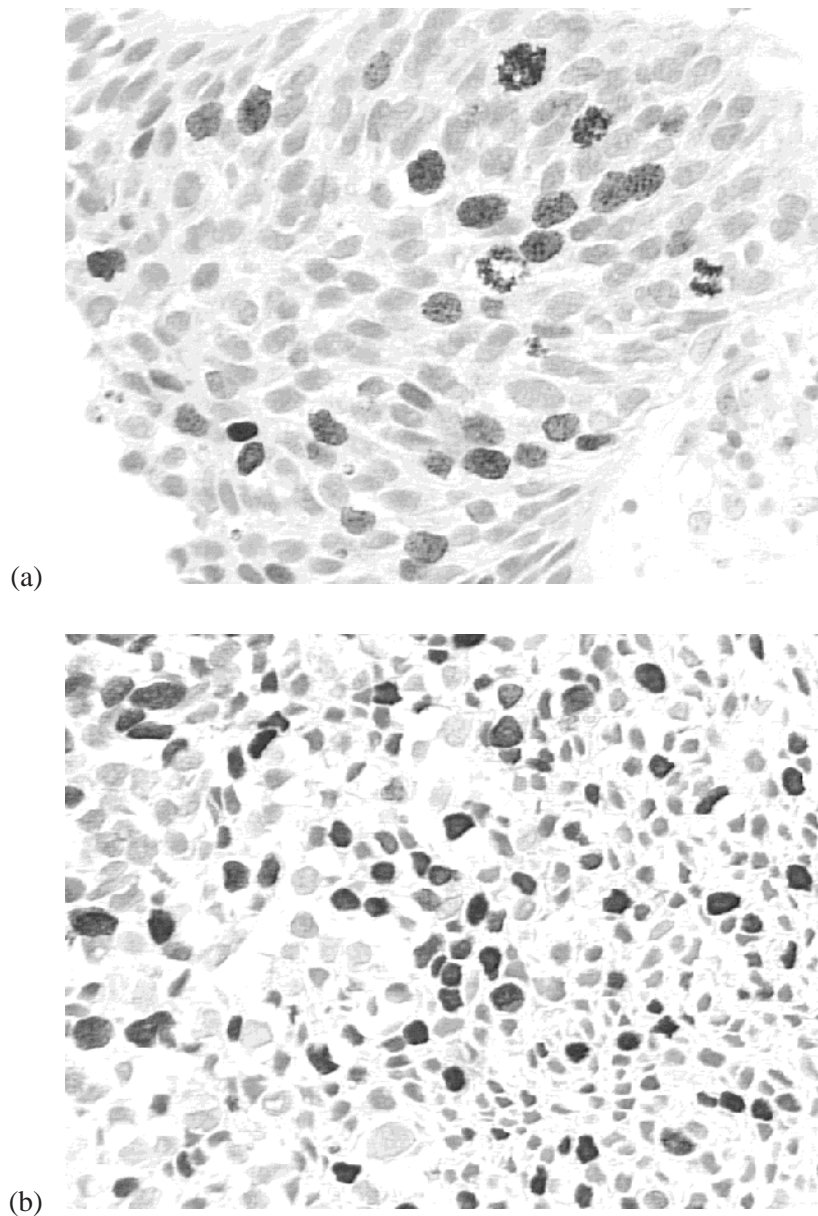


Fig. 1. Typical images of immunohistochemically stained cell nuclei.

- (a) Project the object pixels onto a plane orthogonal to PC1 and compute the direction to the centre of gravity for the object pixels on this plane. We call the angle between this direction and a reference direction principal component hue (PCH).
- (b) Compute P^2/A -histogram in clockwise direction of PCH.
- (c) Compute P^2/A -histogram in counter-clockwise direction of PCH.
- (d) Analyse the P^2/A -histograms to find the direction which gives the most significant peaks.
- (e) Perform a constrained threshold between the peaks.

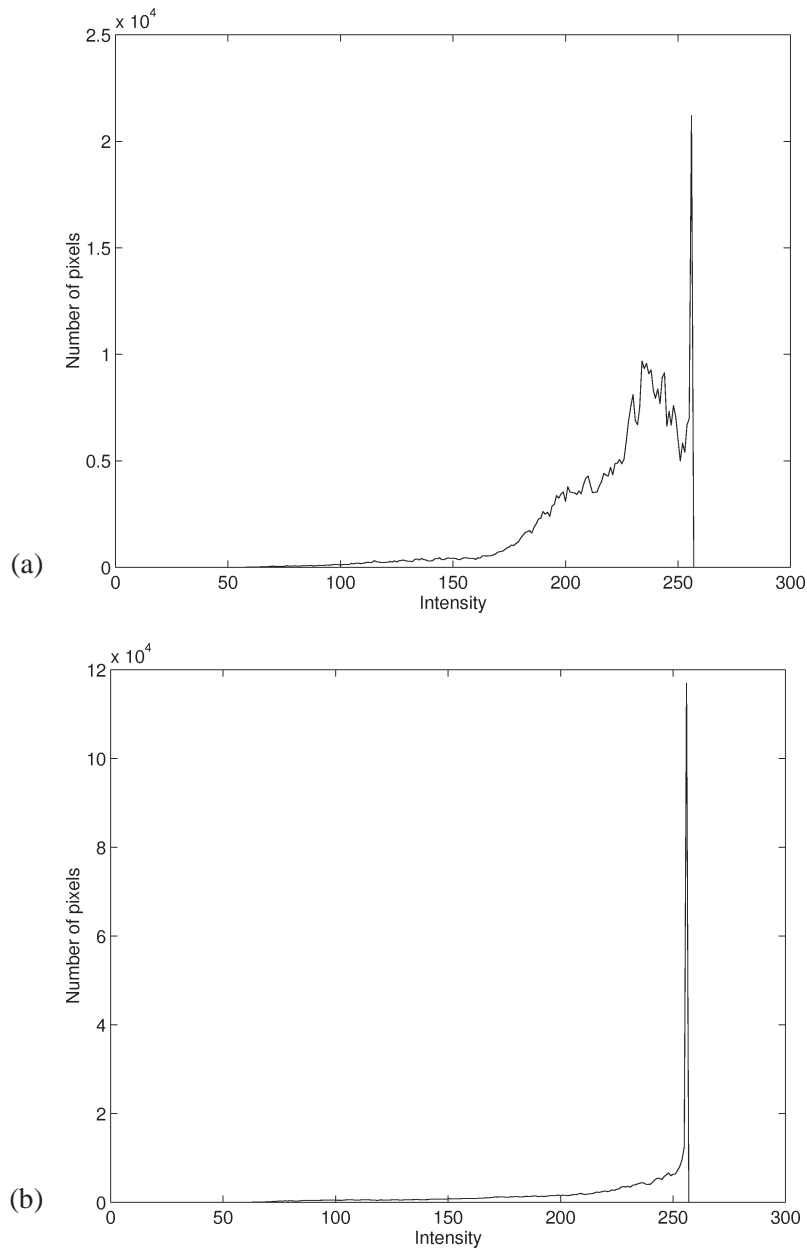


Fig. 2. Histogram of the first principal component for the images in Fig. 1. In this histogram and in all following histograms, the values on the x -axis are the intensity. Here the values on the y -axis are the corresponding number of pixels.

2.3.2. Segmenting objects from background

Perform a principal component transform of the RGB values [6]. The sign of the first principal component, PC1, is defined so that the sum of the components is positive. In this way high PC1 value corresponds to “light”. The data is scaled so that the entire interval (here 0–255) is used.

We now want to do a histogram-based global thresholding of the PC1 image, but it is difficult to automatically set the thresholds based on analysis of the ordinary grey level histograms (see examples

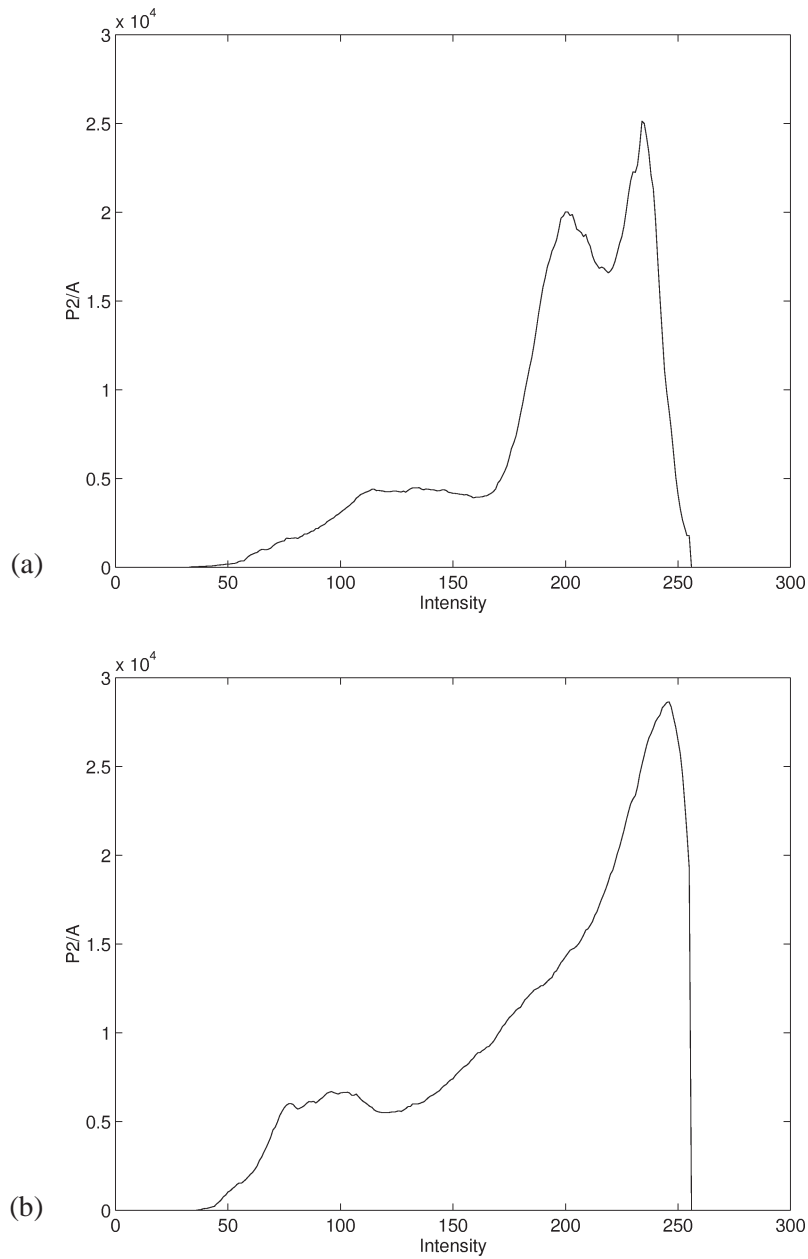


Fig. 3. P^2/A -histograms (P^2/A -values on the y -axis) of the images in Fig. 1. (a) has two significant peaks, but (b) has only one significant peak.

in Fig. 2). Therefore P^2/A -histograms are computed instead. P^2/A is the square of the perimeter divided by the area. A perfect circle gets the value and the more complex the object becomes, the higher the value, i.e., P^2/A is a scale invariant measure of “roundness”. By thresholding the image to a binary image and computing the P^2/A -value for this binary image, and doing so for all possible threshold values, a P^2/A -histogram is created (see examples in Fig. 3).

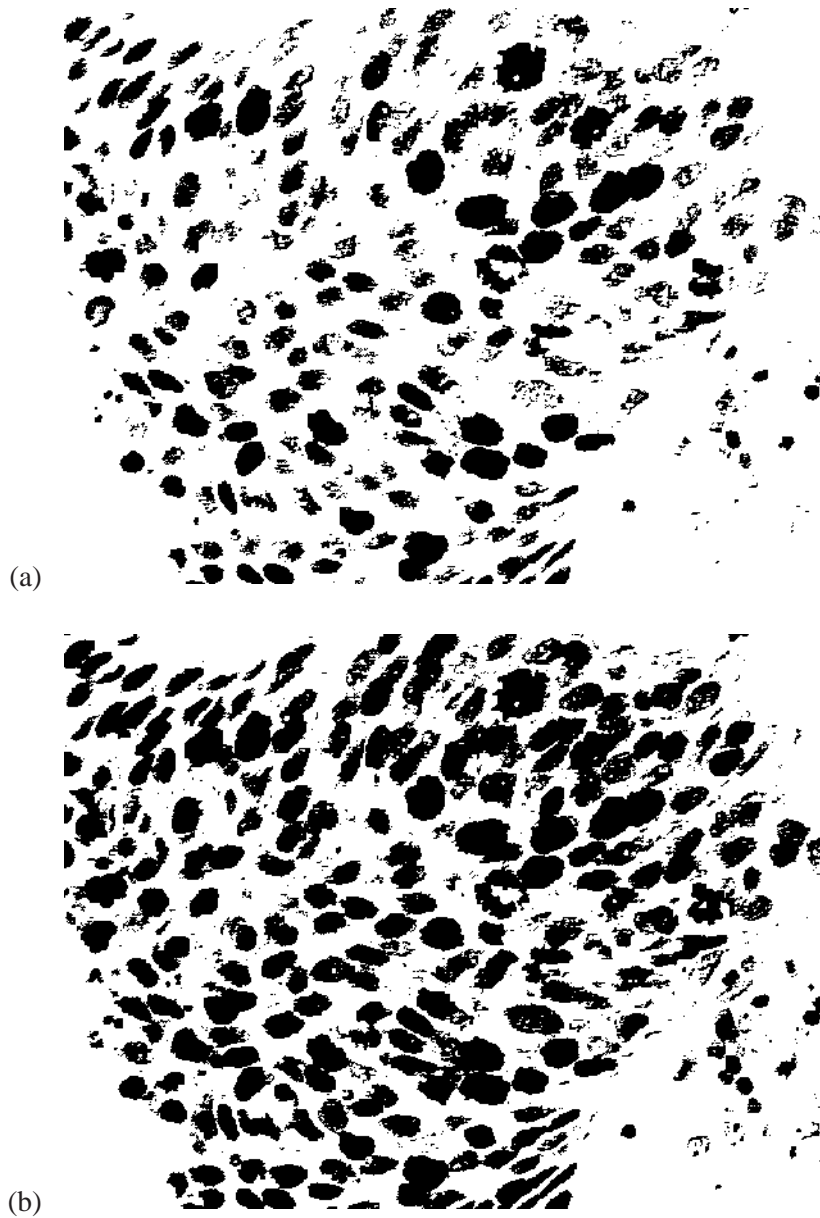


Fig. 4. Threshold for Fig. 1(a) at the left maximum, the minimum and the right maximum, respectively, of the P^2/A -histogram in Fig. 3(a).

The P^2/A -histogram is defined for each grey level i as the sum of, for each pixel with grey level $\leq i$, the number of neighbours with grey level $> i$.

This could be implemented efficiently with only one pass over the image (see Appendix). We have used 4-connectivity when computing neighbourhood, i.e., only pixels sharing an edge are regarded as neighbours.

This histogram has the following useful properties: the P^2/A -value increases when the number of objects increases and when the objects (or background) become more fragmented. The P^2/A will

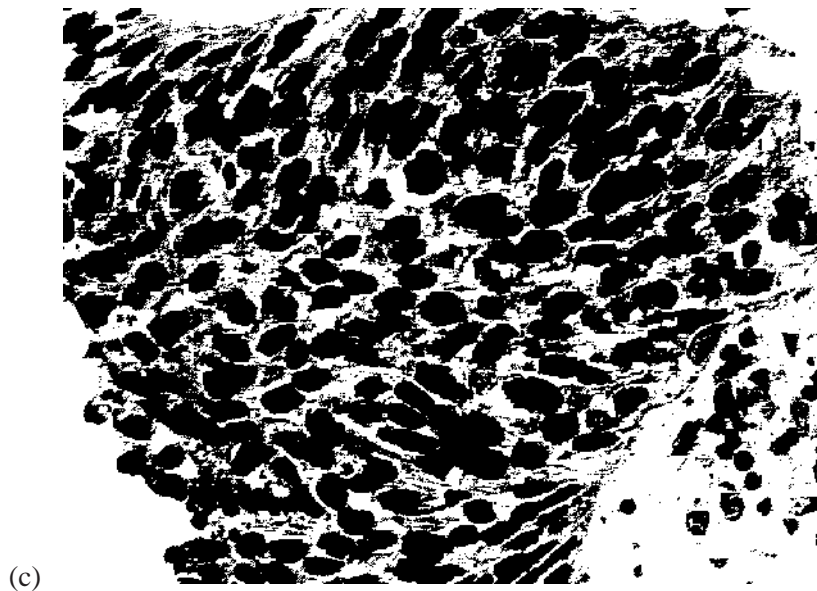


Fig. 4. (Continued).

be low when we have a threshold that gives a good segmentation into a small number of compact objects. For light microscopic images with dark objects on a light background it will be large for high thresholds that cause fragmented background, and also large for low thresholds that cause fragmented objects. The histogram should thus typically have two significant peaks with a minimum in between. The global threshold is set at the minimum. Effects of threshold settings at the minimum and maxima are shown in Fig. 4.

The significant peaks are defined primarily as the two maxima that come after the longest sequence of increasing values, and before the longest sequence of decreasing values, respectively (see example of peaks that fulfill this criterion in Fig. 3(a)). If no minimum exists between these maxima (i.e., they are identical), then the P^2/A -histogram is approximated with a mixture of two normal distributions (see [4] for details) and if there is a minimum between the mean values of these distributions, then the maxima are significant. It is also required that the two peaks are on different sides of the P^2/A -peak for the 11×11 mean filtered PC1 image. The motivation for this is that for a mean filtered image the P^2/A -histogram will typically have its peak rather close to where objects and background overlap (see example of peaks that fulfill this criterion in Fig. 5 and an example of nonsignificant peaks in Fig. 3(b)).

The procedure described above works well for most images, but if the range where the objects are becoming fragmented overlaps the range where the background is becoming fragmented, then the histogram will have only one peak (see example in Fig. 3(b)). If we threshold at the level of this peak, then we will get too many objects. There could be many ways to handle this problem. One heuristically derived approach is to use low pass filtered versions of the PC1 image, which will reduce the fragmentation. We thus first create a 5×5 filtered PC1 image, which is to be thresholded. We then inspect a second 11×11 filtered image. If the P^2/A -peak of this second image is to the left of the inflection point before the original P^2/A -peak, then this level is used as threshold, otherwise the threshold level will be at the inflection point before the peak of the grey level histogram of the 5×5 filtered image.

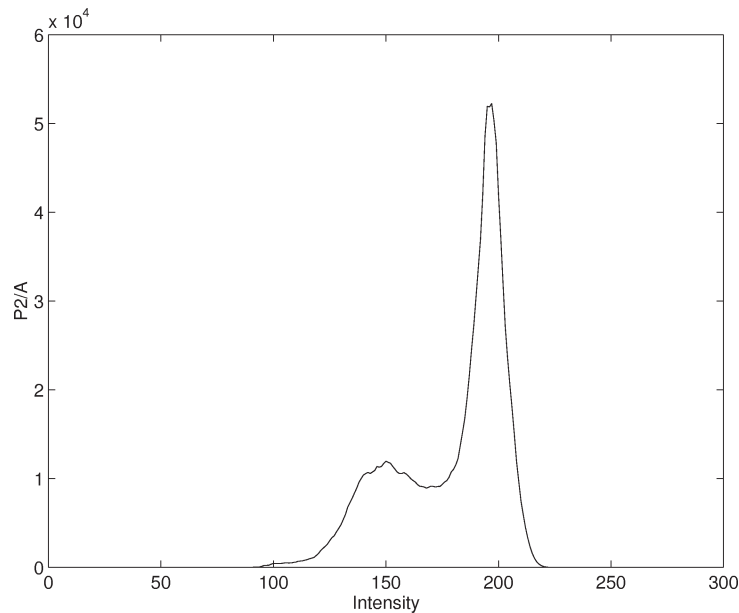


Fig. 5. Example of histogram with two significant peaks, despite the fact that the maximum after the longest sequence of increasing values is equivalent to the maximum before the longest sequence of decreasing values.

2.3.3. Segmenting positive and negative objects

The object pixels are projected onto a plane orthogonally to PC1. On this plane the centre of gravity for the object pixels is computed and then each pixel is associated with the direction to the centre of gravity, defined as the angle between this direction and a reference direction. We call this angle principal component hue (PCH). Since it is computationally desirable to have a linear, instead of a circular, scale, we go around the preliminary PCH histogram to find the most data-free section and set the reference direction (0 angle) in the middle of this section. To be able to relate the angles to colour we compute the angle for pure red, green and blue, respectively.

Next we compute the P^2/A -histogram clockwise and counter-clockwise for the PCH image. In practice this is done by computing the P^2/A as in Section 2.3.2 on the PCH image and the inverted PCH image, respectively, with the background set to 255 in both cases.

For the P^2/A -histogram with the most significant minimum, defined as the difference between the second highest maximum and the minimum between the two highest maxima, the inflection points to the right of the left maximum and to the left of the right maximum are used as input thresholds to constrained thresholding. Constrained thresholding is defined so that all pixels with grey values below threshold 1 are set to class 1, and pixels with grey values between threshold 1 and threshold 2, that have a neighbour with grey value below threshold 1, are also set to class 1. In the same manner pixels with grey values above threshold 2 are set to class 2, and pixels with grey values between threshold 1 and threshold 2, that have a neighbour with grey value above threshold 2, are also set to class 2. Pixels between threshold 1 and threshold 2 that only have neighbours with values between threshold 1 and threshold 2, or if they have neighbours both below threshold 1 and above threshold 2, are set to background. This could be implemented efficiently with one pass over the image (see Appendix).

The result of the proposed procedure applied to Fig. 1 is shown in Fig. 6.

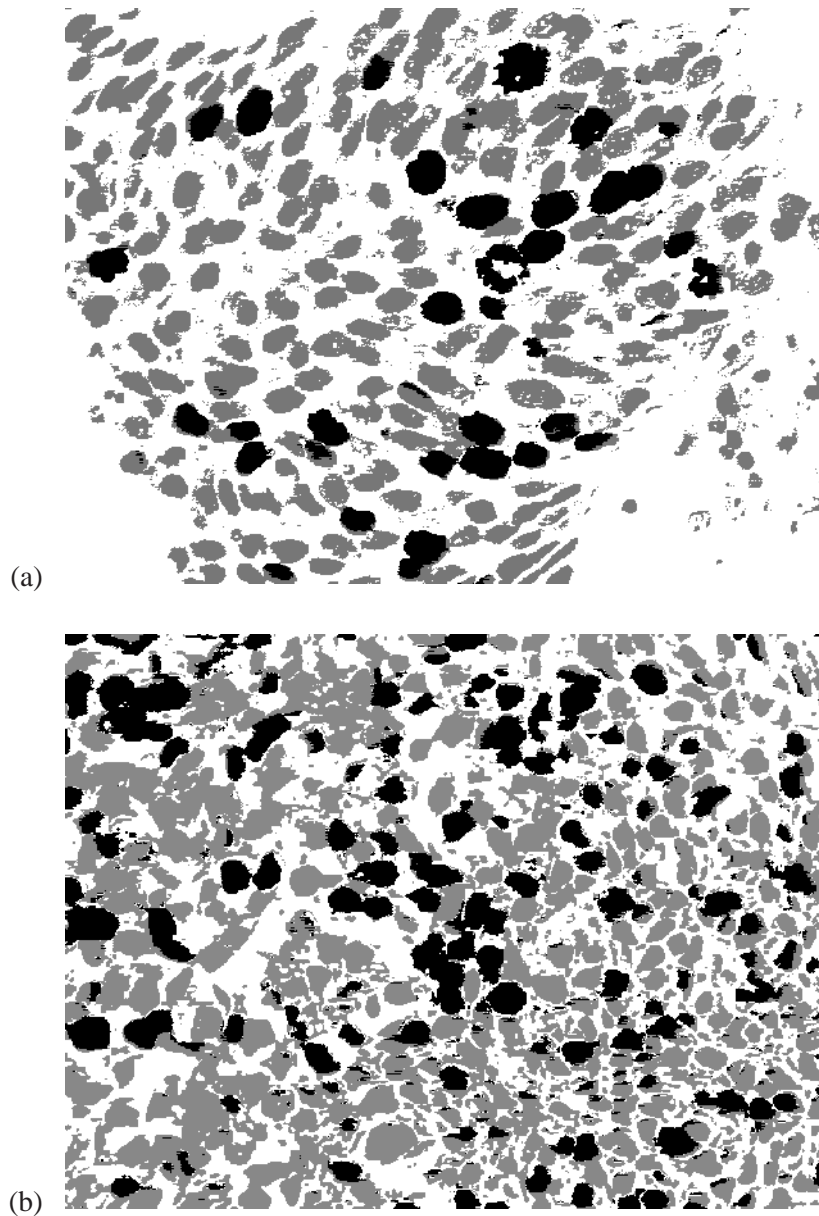


Fig. 6. Result when the described method is applied to the images in Fig. 1. Black is positive cell nuclei and grey is negative cell nuclei.

3. Results

Since different quantification methods can give reasonable results when viewed one at a time, it is difficult to tell what the true estimate should be. Another criterion for a good quantification method is that it is robust and gives low variation. We have tested the robustness and variation of the proposed method and of methods based on supervised Maximum Likelihood (ML) classification [3] by using images from the same field of view, captured with varying focus and illumination.

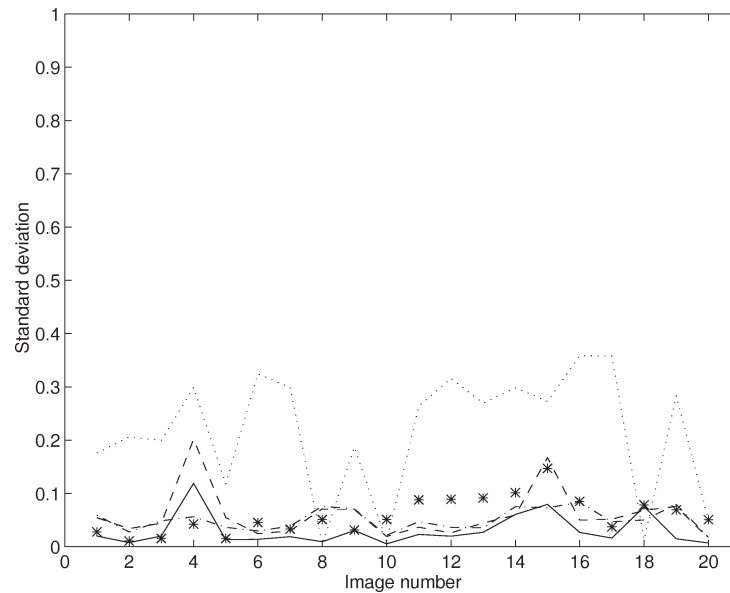


Fig. 7. Standard deviations for the four different ML methods and the proposed method on 20 tested fields of view, where 1–10 are stained using DAB and 11–20 are stained using Ni-DAB. Line: training image variation; dashed line: intraoperator variation; dash-dot line: intraoperator and training image variation; dotted line: classification image variation; asterisks: proposed method.

For evaluation we have used an independent test set consisting of 20 fields of view. Ten from IHC stainings using DAB as chromogen and the remaining 10 fields of view were the same histological areas as above but from Ni-DAB stained consecutive tissue sections. The fields of view were all selected from three specimens from three patients, two bladder carcinomas and one reactive tonsil. For each field of view 12 images were grabbed, differing in focus and illumination. Thus, in total we got 240 images. The difference in focus was approximately $\pm 1.5 \mu\text{m}$ and the difference in illumination was about ± 50 on a 0–255 scale.

Since we have 12 images for each field of view, the standard deviation of the estimated proportions is a measure of the variability of the results for a field of view. These standard deviations can be compared to the standard deviations obtained using a supervised ML classifier. We used a general ML classifier with no assumptions about the covariance matrices. In the following ways we used the ML classifier on the images from the same field of view (ordered after expected variation):

1. The same training region is applied to all images to create different classifiers and each image is classified with its corresponding classifier. This tests the variability due to the differences in the training images. This should give low variation.
2. Twelve training regions are applied to the same image to create 12 different classifiers and the classifiers are used to classify the same image. This tests the intraoperator variability.
3. Every image is trained and classified with its own training region. This tests the differences due to intraoperator variability and training image differences.
4. One training region is used to create a classifier on one image, and the resulting classifier is then used to classify all images. This tests the variability due to differences in the classified image. This should give high variation, since this method does not compensate for differences in intensity.

Table 1
Average standard deviations for the four ML methods and the proposed method

Method	Mean (SD (total))	Mean (SD (stain 1))	Mean (SD (stain 2))
Training image variation	0.0304	0.0258	0.0351
Intraoperator variation	0.0600	0.0604	0.0594
Intraoperator/training image variation	0.0505	0.0465	0.0545
Classification image variation	0.2155	0.1824	0.2485
Proposed method	0.0582	0.0325	0.0839

Note that only the differences to the classification image are significant.

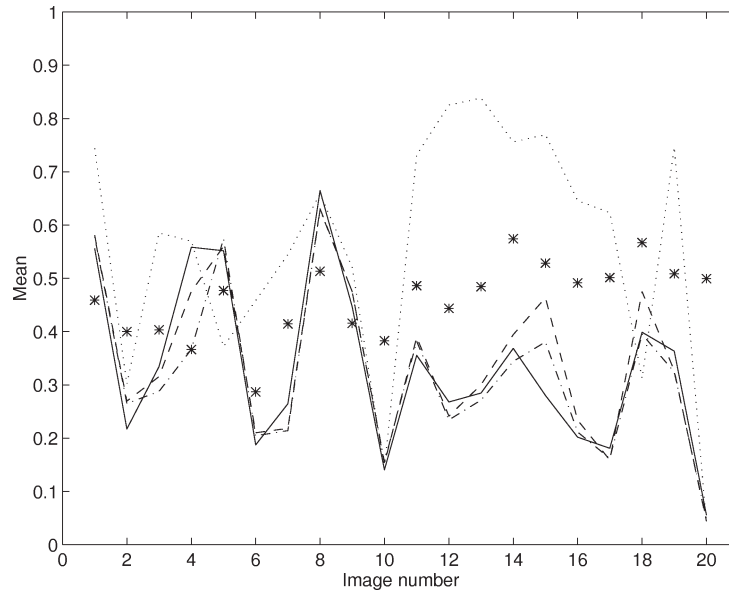


Fig. 8. Mean values for the four different ML methods and the proposed method on 20 tested fields of view, where 1–10 are stained using DAB and 11–20 are stained using Ni-DAB. Line: training image variation means; dashed line: intraoperator variation means; dash-dot line: intraoperator and training image variation means; dotted line: classification image variation means; asterisks: means for proposed method.

The standard deviations for these four methods and for the proposed method are shown in Fig. 7 and in Table 1. The results indicate that the proposed method is suitable for chromogens giving differences in hue, i.e., the DAB, but less suitable for chromogens giving differences mainly in grey scale, i.e., Ni-DAB. The variation for the proposed method on images from the same field of view is smaller than the intraoperator variation for a supervised ML classifier on the same image.

A comparison of the estimated proportions (see Fig. 8) gives some clear differences, even though most results seem reasonable when looking at the images. It could be argued that it is not meaningful to compare proportions computed with different methods, since what we want to achieve is not to copy the performance of another method, but a stable and automatic method. If it is desired to obtain estimations that are as similar as possible to those obtained by a different method, one possible way is to do a linear transformation between the resulting proportions to translate between the different methods.

It should be noted that when using supervised ML classification it is sometimes necessary to redraw the training areas in order to get satisfactory results, which involves more subjectivity as well as more user interaction time.

The proposed method uses about 5 s of CPU time per image on a 233 MHz DEC Alpha 2000, which is about the same as the classification time for the general ML classifier.

4. Discussion

Principal component hue as defined in this paper is a powerful tool for segmentation, but since it is related to the centre of gravity of only the object pixels, the PCH will always have values spread over a large range of the unit circle. This could make the method detect different classes even though there is only one class in the image.

Some pathologists suggest that the fields of view to be analysed must be subjectively selected by the observer. This is to ensure that images describing patterns and features of interest in the sample and images representative for the sample are included in the quantification. By using such an approach the images are guaranteed to contain both positive and negative staining. But for the proposed method to be part of a method that is unsupervised all the way from image acquisition, we need to automatically detect when the images contain only one class. One way to do this is to compute the average hue (note: not PCH!) for the classes, respectively, and test if the absolute difference between the two classes is below some threshold. Ten degrees seems to be a reasonable threshold value.

It is possible to postprocess the resulting thresholded segmentation to get rid of objects too small and to make objects more homogeneous. If the parameters (e.g., filter size, object size, etc.) for these operations are trimmed properly one can expect that the results would improve. In our study we have not used this kind of postprocessing.

Another problem is that if the image is too light, the shape of the data could be distorted towards white, and if the data shape is too dark then the colour range is not being used properly. This can be detected through a simple analysis of the grey level histogram.

Together with a standardised staining method, such as immunohistochemistry autostainer equipment, the algorithm described in this article could be used as a standard for quantification of immunohistochemically stained specimens. The remaining subjectivity is limited to the selection of images to analyse. Compared to interactive methods, the proposed method also has the advantage of working very quickly, a couple of seconds per image. The described method could also be applied to other types of images and extended to more classes, but so far we have concentrated on stained cell nuclei from human bladder carcinoma and reactive tonsil tissue.

Acknowledgements

We are grateful to Christer Busch and colleagues at the Department of Pathology for cooperation and for providing us with the images. We are also grateful to Prof. Lennart Bondesson at the Department of Mathematics for statistical discussions. This project was supported by the Swedish National Board for Technological Development under Grant no. 5523-93-4991.

Appendix

Pseudo-code for computing P^2/A -histogram:

Input: image f . Output: histogram h .

```

FOR i:=0 TO 255
  area[i]:=0;
  perimeter[i]:=0;
END FOR
FOR y:=0 TO YMAX
  FOR x:=0 TO XMAX
    this:=f(x,y);
    right:=f(x+1,y);
    down:=f(x,y+1);
    area[this]:=area[this]+1;
    IF this<right THEN
      perimeter[this]:=perimeter[this]+1;
      perimeter[right]:=perimeter[right]-1;
    ELSE
      IF this>right THEN
        perimeter[this]:=perimeter[this]-1;
        perimeter[right]:=perimeter[right]+1;
      END IF
    END IF
    IF this<down THEN
      perimeter[this]:=perimeter[this]+1;
      perimeter[down]:=perimeter[down]-1;
    ELSE
      IF this>down THEN
        perimeter[this]:=perimeter[this]-1;
        perimeter[down]:=perimeter[down]+1;
      END IF
    END IF
  END FOR
END FOR
CumulativeArea:=0;
CumulativePerimeter:=0;
FOR i:=0 TO 255
  CumulativeArea:=CumulativeArea+area[i];
  CumulativePerimeter:=CumulativePerimeter+perimeter[i];
  h[i]:=CumulativePerimeter/CumulativeArea;
END FOR

```

Note: The border outside the image could be set either to 0 or 255. We have used 255.

Pseudo-code for constrained threshold:

Input: image f , thresholds $t1$, $t2$. **Output:** image g .

```

FOR y:=0 TO YMAX DO
  FOR x:=0 TO XMAX DO
    g(x,y):=UNDEFINED;
  END FOR
END FOR
FOR y:=0 TO YMAX DO
  FOR x:=0 TO XMAX DO
    this:=f(x,y);
    right:=f(x+1,y);

```

```

down:=f(x,y+1);
IF this<=t1 THEN
  g(x,y):=CLASS1;
  IF t1<right<t2 THEN
    IF g(x+1,y)=UNDEFINED OR g(x+1,y)=CLASS1 THEN
      g(x+1,y):=CLASS1;
    ELSE
      g(x+1,y):=CONFLICT;
    END IF
  END IF
  IF t1<down<t2 THEN
    g(x,y+1):=CLASS1;
  END IF
ELSE
  IF t1<this<t2 THEN
    IF g(x,y)=UNDEFINED THEN
      IF (right<=t1 AND down<t2) OR (down<=t1 AND right<t2) THEN
        g(x,y):=CLASS1;
      ELSE
        IF (t1<down AND t2<=right) OR (t1<right AND t2<=down) THEN
          g(x,y):=CLASS2;
        ELSE
          IF (down<=t1 AND t2<=right) OR (right<=t1 AND t2<=down) THEN
            g(x,y):=CONFLICT;
          END IF
        END IF
      END IF
    ELSE
      IF g(x,y)=CLASS1 THEN
        IF t2<=right OR t2<=down THEN
          g(x,y):=CONFLICT;
        END IF
      ELSE
        IF g(x,y)=CLASS2 THEN
          IF right<=t1 OR down<=t1 THEN
            g(x,y):=CONFLICT;
          END IF
        END IF
      END IF
    END IF
  ELSE
    IF t2<=this THEN
      g(x,y):=CLASS2;
      IF t1<right<t2 THEN
        IF g(x+1,y)=UNDEFINED OR g(x+1,y)=CLASS2 THEN
          g(x+1,y):=CLASS2;
        ELSE
          g(x+1,y):=CONFLICT;
        END IF
      END IF
      IF t1<down<t2 THEN
        g(x,y+1):=CLASS2;
      END IF
    END IF
  END IF

```

```
END IF
END IF
END FOR
END FOR
```

Note: The edge of the image should be treated so that no references outside the image are made.

References

- [1] K.R. Castleman, *Digital Image Processing*, Prentice-Hall, 1979.
- [2] 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** (1992), 490–496.
- [3] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.
- [4] B. Everitt and D. Hand, *Finite Mixture Distributions*, Chapman and Hall, 1981.
- [5] C. Garbay, G. Brugal and C. Choquet, Application of colored image analysis to bone marrow cell recognition, *Anal. Quant. Cytol.* **3**(4) (1981).
- [6] I. Jolliffe, *Principal Component Analysis*, Springer, 1986.
- [7] N.R. Pal and S.K. Pal, A review on image segmentation techniques, *Pattern Recognition* **26**(9) (1993), 1277–1294.
- [8] P. Ranefall, L. Egevad, B. Nordin and E. Bengtsson, A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei, *Anal. Cell. Pathol.* **15** (1997), 145–156.
- [9] P. Ranefall, B. Nordin and E. Bengtsson, A new method for creating a pixelwise box classifier for color images, *Machine Graph. Vision* **6**(3) (1997), 305–323.
- [10] 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.