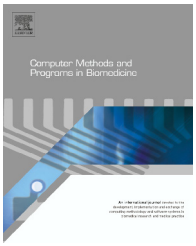
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 111 (2013) 128–138

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Debris removal in Pap-smear images

Patrik Malm[a,*], *Byju N. Balakrishnan*[b], *Vilayil K. Sujathan*[c], *Rajesh Kumar*[b], *Ewert Bengtsson*[a]

[a] *Centre for Image Analysis, Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Box 337, 751 05 Uppsala, Sweden*

[b] *Centre for Development of Advanced Computing, Thiruvananthapuram, India*

[c] *Regional Cancer Centre, Thiruvananthapuram, India*

ARTICLE INFO

Article history:
Received 31 August 2012
Received in revised form
25 February 2013
Accepted 26 February 2013

Keywords:
Debris removal
Pap-smear
Cervical cancer screening
LBC

ABSTRACT

Since its introduction in the 1940s the Pap-smear test has helped reduce the incidence of cervical cancer dramatically in countries where regular screening is standard. The automation of this procedure is an open problem that has been ongoing for over fifty years without reaching satisfactory results. Existing systems are discouragingly expensive and yet they are only able to make a correct distinction between normal and abnormal samples in a fraction of cases. Therefore, they are limited to acting as support for the cytotechnicians as they perform their manual screening.

The main reason for the current limitations is that the automated systems struggle to overcome the complexity of the cell structures. Samples are covered in artefacts such as blood cells, overlapping and folded cells, and bacteria, that hamper the segmentation processes and generate large number of suspicious objects. The classifiers designed to differentiate between normal cells and pre-cancerous cells produce unpredictable results when classifying artefacts.

In this paper, we propose a sequential classification scheme focused on removing unwanted objects, debris, from an initial segmentation result, intended to be run before the actual normal/abnormal classifier. The method has been evaluated using three separate datasets obtained from cervical samples prepared using both the standard Pap-smear approach as well as the more recent liquid based cytology sample preparation technique. We show success in removing more than 99% of the debris without loosing more than around one percent of the epithelial cells detected by the segmentation process.

© 2013 Elsevier Ireland Ltd. Open access under CC BY-NC-ND license.

1. Introduction

According to the World Health Organization (WHO), cervical cancer is the second most common type of cancer among women, annually killing close to 300,000 world wide. 86% of these deaths occur in developing countries [1]. The main reason behind this discrepancy is the absence in developing countries of organized screening programmes using the Papanicolaou test (Pap test) developed by Dr. Georges Papanicolaou in the 1940s [2].

A *Pap-smear* is prepared by acquiring cellular material from the uterine cervix using a spatula or a brush. The collected material is then smeared on a microscope slide, fixated using a spray fixative and then stained using the *Pap-stain* [3].

* *Corresponding author.*
E-mail address: patrik.malm@cb.uu.se (P. Malm).
URL: http://www.cb.uu.se/~patrik (P. Malm).

0169-2607© 2013 Elsevier Ireland Ltd. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.cmpb.2013.02.008

When a smear is analysed under a microscope, trained cytologists can not only find evidence of invasive cancer but also detect certain cancer precursors, allowing for early and effective treatment. If detected early, invasive cancer is curable and the 5-year survival rate is as high as 92% [4].

Although the Pap-smear has shown its worth through decades of use, it is hampered by a number of difficulties, e.g., variable smear thickness, uneven cell distribution, obscuring elements such as blood and inflammatory cells, and variable fixation and staining results. To overcome some of these problems a number of so-called liquid-based cytology (LBC) preparation methods have been developed. Using LBC the sample is immersed in a solution which is then subjected to a number of steps that work to homogenize the sample, remove unwanted components (e.g., red blood cells) and finally deposit a suitable mono-layer sample on a glass slide [5]. The result is a sample that should contain a better representation of the biological material presented in a mono-layer fashion which according to several studies, e.g. Jhala and Eltoum [6], lead to better screening results.

Most screening programmes are based on visual screening performed by cytotechnicians in specialized laboratories. The screening work is tedious and, often due to fatigue, error prone. Because of the hazards of fatigue some recommendations say that a cytotechnician should not work with screening more than 7 h a day and analyse no more than 70 samples per day [7].

To overcome some of the human limitations and hopefully to reduce screening costs several attempts to automate the screening process have been made since the 1950s with varying degree of success. Today there are systems that are able to perform a scan and subsequent analysis of a sample but they all have in common that they require specific sample preparation and are complicated and expensive to run [8].

When analysing a Pap-smear the cytotechnician will look for a number of tell-tale signs that a sample contains evidence of malignancy [3]. Many of these signs are related to the appearance of the cell nuclei (i.e. shape, color, nucleus/cytoplasm ratio, size, chromatin distribution). Because of the importance of nuclear characteristics the main focus when developing automated smear analysis has been on the processes of locating/delineating [9–13] as well as extracting features [13–15] from nuclei. Segmenting nuclei in Pap-smears is then a key process, but it is made very difficult by the same complications that make the smears hard for humans to analyse, i.e., variable smear thickness, obscuring elements, et cetera. The LBC preparation methods will reduce these problems but not remove them altogether.

The early attempts at screening systems used various kinds of automated greyscale thresholding [16,17] but more recent projects have used more complicated approaches. Bergmeir et al. [9] uses Canny edge detection followed by the randomized Hough transform [18]. Bamford and Lovell [10] use a dual active contour algorithm. Lin et al. [11] uses a two group object enhancement technique. Malm and Brun [12] uses Canny edge detection followed by anisotropic curve closing. Gençtav et al. [13] use a form of multi-scale watersheds to generate hierarchical partitioning of nuclei and cytoplasms.

All segmentation algorithms in this context are intrinsically created to avoid picking up unwanted objects, henceforth referred to as debris. Still there will in most cases be many debris objects among the segmented "nuclei". When such debris objects are subject to feature extraction and classification designed to detect signs of malignancy the outcome is more or less random leading to great difficulties in designing a system with sufficiently low false positive and false negative rates. In this paper we propose an initial classification stage with the sole purpose of detecting and removing the debris objects. To the best of our knowledge no previous paper has had that focus.

2. Methods

The objective of the work presented in this paper was to develop a robust method for filtering out debris from an initial segmentation result. The method has been tailored to tackle many of the difficulties present in Pap-smear images (Fig. 1). The approach centres around a sequential elimination scheme (Fig. 2) where objects from an initial segmentation are removed if deemed unlikely to be one of the relevant types of cell nuclei. The benefits of a sequential approach are two-fold. First, it allows for a lower-dimensional decision to be made at each stage, thus reducing the effects of the curse of dimensionality, and second, it makes it possible to place more computationally heavy object descriptors at the end of the pipeline where fewer objects remain.

Where applicable, a standard Bayesian quadratic classifier [19] has been used. Furthermore, since each step of the method only tackles a limited number of features, the complexity of the classifier is not as critical.

In the initial step of the proposed method, objects are thresholded based on their area (see Section 2.1). Following the basic thresholding, objects are analysed based on their shape. The second step of the algorithm evaluates objects using region-based and contour-based shape representations (see Section 2.2). The third step constitutes a custom algorithm that measures the elliptical deviation (see Section 2.3). Remaining objects at this stage are evaluated based on their texture (see Section 2.4) and finally their average greyvalue (see Section 2.5).

2.1. *Area*

Area is perhaps the most basic feature available and also the first one used within the field of automated cytology to separate cells from debris [20]. This is of course not an inherently specific feature but for segmentation algorithms where size is not taken into account implicitly, e.g. [12], it is a necessary one.

Finding a lower size threshold is generally not an issue in automated cervical cytology applications. Because it is such a well studied field much prior knowledge regarding cell characteristics, such as average size distribution, is available [21]. However, one of the key changes a cancerous cell undergoes is the substantial increase of nuclear size [3] (Fig. 3). Therefore, determining an upper size threshold that does not systematically exclude diagnostic cells is much harder.

The method described in this paper only uses a lower size threshold to avoid the exclusion of diagnostic cells. The

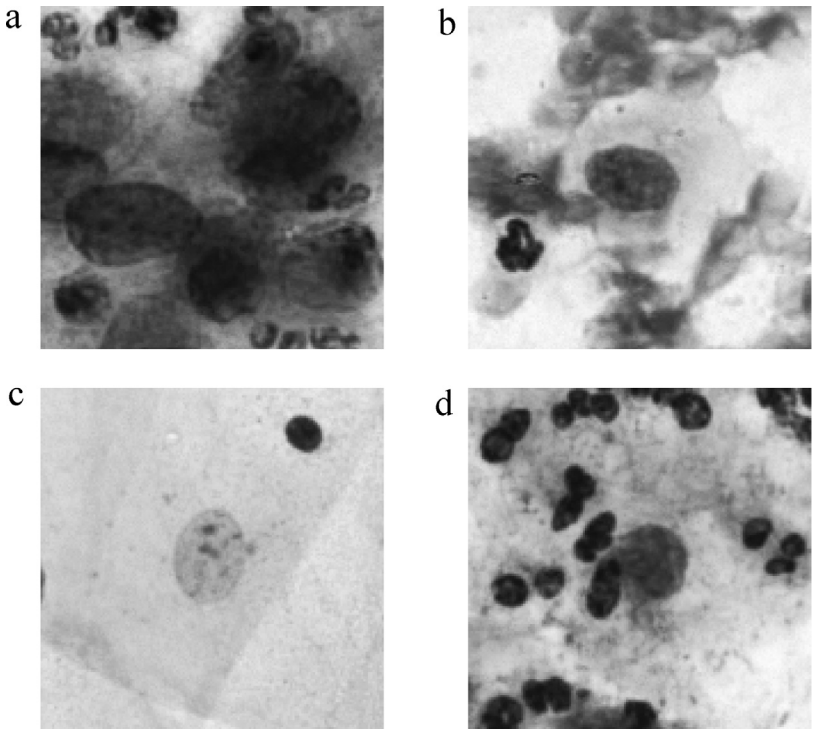


Fig. 1 – Figure depicting a few of the difficulties found in cervical smears: (a) overlapping cells, (b) red blood cells, (c) uneven staining, and (d) white blood cells.

threshold is found by looking at the distributions of the cell and debris populations used for training. We have chosen to place the threshold at the crossing point of the two distributions, thus accepting a certain loss of smaller cell nuclei for the gain of the increased number of debris objects removed (Fig. 4).

2.2. *Basic shape*

A key feature in the differentiation between cells and debris is the shape of the object. There exists a large number of methods for shape description [22]. One can divide the approaches into region-based, i.e., taking the entire object into account, and contour-based, i.e., only analysing the boundary pixels. In general, region-based methods are less sensitive to noise but more computationally heavy to calculate, whereas contour-based methods are relatively cheap to calculate but more noise sensitive since only boundary pixels are taken into account. We have chosen to use one region-based method, $perimeter^2/area$ (P2A), and one contour-based method, Fourier shape descriptors (FSD).

The P2A descriptor was chosen on the merit that it describes the similarity of an object to a circle [19]. This makes it well suited as a cell nucleus descriptor since nuclei are generally circular in their appearance.

Fourier shape descriptors are able to capture global features as well as the finer details of an object's shape. They are also relatively insensitive to noise and easy to normalize [23].

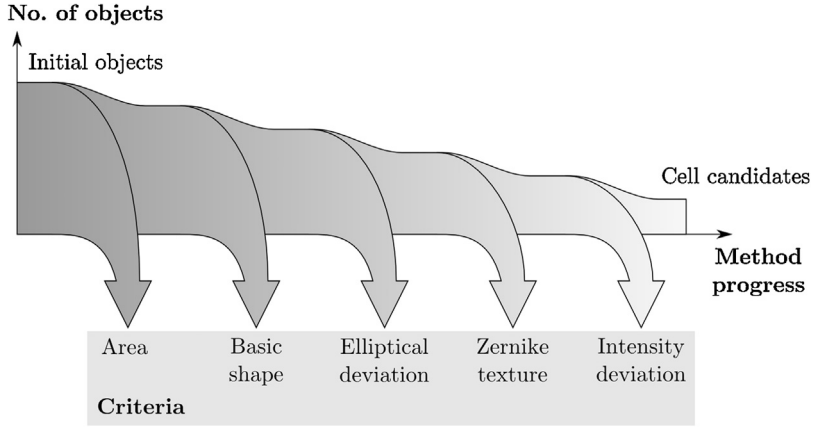


Fig. 2 – Overview of the approach for classifying cell candidates into cells and debris.

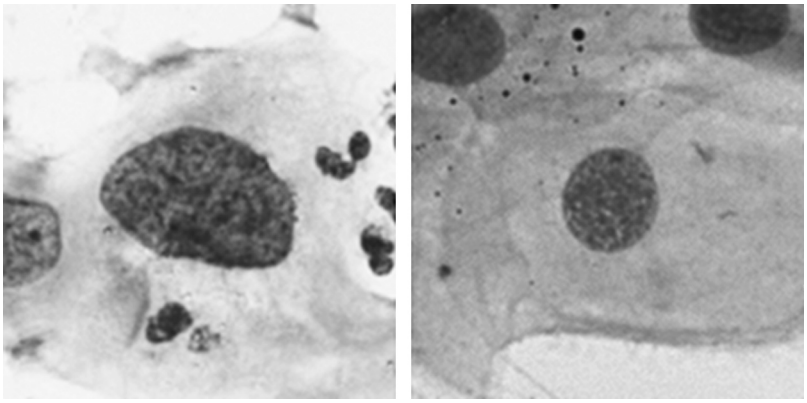


Fig. 3 – Comparison between cancerous nucleus (left) and normal nucleus (right).

When calculating FSDs, one generally starts with a shape signature, i.e., a 1D function representation of the boundary of an object. If we extract the object boundary coordinates ($x(t)$, $y(t)$, $t = 0, 1, \ldots, L-1$) for an object with an L-point boundary (Fig. 5), we can describe that boundary as a complex coordinate function

$$z(t) = x(t) + iy(t). \tag{1}$$

We can also choose to eliminate the effect of translation bias by using a shifted coordinate function

$$z(t) = [x(t) - x_c] + i[y(t) - y_c] \tag{2}$$

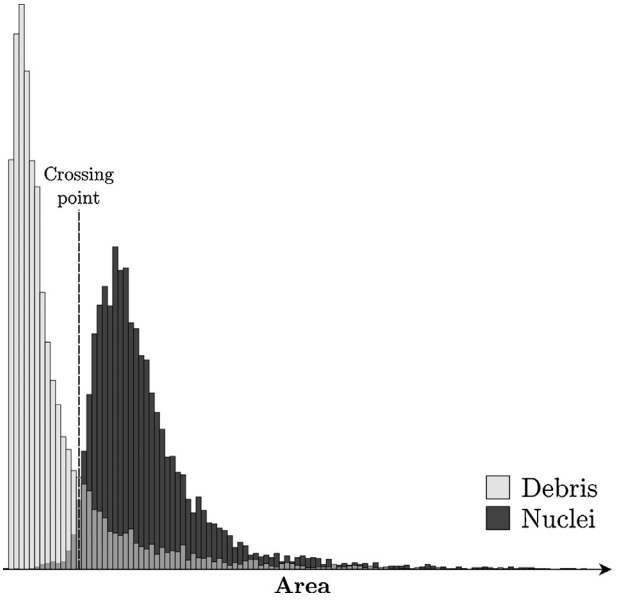


Fig. 4 – Illustration showing the size distribution of the debris and nucleus populations generated using the dataset described in Section 3.4. The crossing point between the two populations has been indicated.

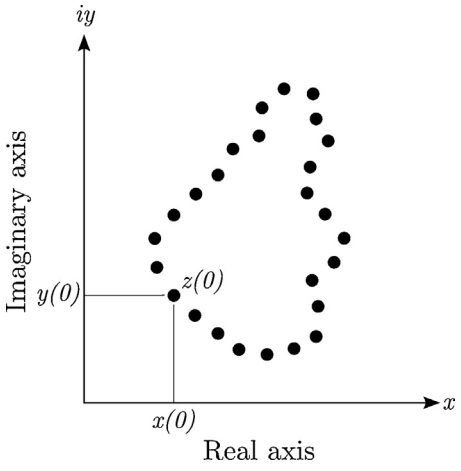


Fig. 5 – A digital boundary represented as a complex sequence of points. The arbitrary first point (x(0), y(0)) is shown.

where (x_c, y_c) are the centroid coordinates of the object calculated as

$$x_c = \frac{\sum_x \sum_y x b(x,y)}{\sum_x \sum_y b(x,y)}, \quad y_c = \frac{\sum_x \sum_y y b(x,y)}{\sum_x \sum_y b(x,y)}. \tag{3}$$

Here $b(x, y)$ constitutes the characteristic function of a binary object in an image $I(x, y)$ and is formally written as

$$b(x,y) = \begin{cases} 1, & (x,y) \text{ is an object pixel} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

We can now calculate the discrete Fourier transform (DFT) of $z(t)$ as

$$u(n) = \frac{1}{L}\sum_{t=0}^{L-1} z(t) e^{-i2\pi nt/L} \tag{5}$$

where $n = 0, 1, \ldots, L-1$. The complex coefficients $u(n)$ are called the FSD of the boundary. Fourier shape descriptors are especially good when it comes to shape invariants. By using the shifted coordinate function (Eq. (2)), we already have

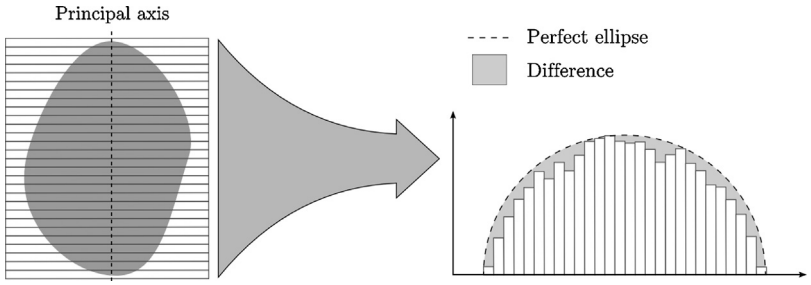


Fig. 6 – Density profile of the segmented object taken along the principal axis compared to the perfect ellipse.

translational invariance, $u(0)=0$. We can also achieve rotational invariance by ignoring the phase information, taking only the magnitude values of the FSD. Finally, we can achieve scale normalization by dividing the magnitude values of all descriptors by the value of the second descriptor, $u(1)$. An invariant Fourier descriptor vector, $\bar{d}$, can thus be written as

$$\bar{d} = (d_1, \ldots, d_{L-1}) = \left(\frac{|u(1)|}{|u(1)|}, \frac{|u(2)|}{|u(1)|}, \ldots, \frac{|u(L-1)|}{|u(1)|} \right). \tag{6}$$

For this paper three features have been calculated from $\bar{d}$. The first feature describes the eccentricity of the object, *ecc*, and is calculated as

$$ecc = \frac{d_1 - d_{L-1}}{d_1 + d_{L-1}}. \tag{7}$$

The second and third features constitute the low-frequency, $freq_{low}$, and the high-frequency, $freq_{high}$, shape energy. These are calculated by adding the values of certain intervals in the spectrum together [14]. The intervals are defined by a lower and a upper limit, $[A, B]$, $B > A$, giving us the feature scores

$$\begin{aligned} freq_{low} &= \sum_{i=1}^{A} d_i + \sum_{i=L-(A+1)}^{L-1} d_i \\ freq_{high} &= \sum_{i=A+1}^{B} d_i + \sum_{i=L-(B+A+1)}^{L-(A+2)} d_i. \end{aligned} \tag{8}$$

2.3. *Elliptical deviation*

In general, for Pap-smears, cell nuclei appear as elliptical objects. Because of this, shape descriptors based on different comparisons between segmented shapes and ellipses are fairly common [14]. Bengtsson et al. [24] try to separate overlapping cells by analysing the integrated profile of objects along the principal axis. For the method described in this paper a similar approach has been developed.

Given a binary object we initially need to locate its principal axis. To do this we first translate the object so that the centre of mass becomes the origin of the coordinate system. We can then calculate the second moments as

$$\begin{aligned} a &= \sum_x \sum_y (x - x_c)^2 b(x, y) \\ b &= 2 \sum_x \sum_y (x - x_c)(y - y_c) b(x, y) \\ c &= \sum_x \sum_y (y - y_c)^2 b(x', y'). \end{aligned} \tag{9}$$

If we put a, b and c together in matrix form we get the covariance matrix, C,

$$C = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}. \tag{10}$$

Finally, we can find the principal axis orientation of the object by calculating the eigenvectors $V = [v_1, v_2]$ of C.

With the principal axis of the object located we now sum the pixels of the object perpendicular to the axis (Fig. 6). This gives us a density function, Z, that represents the shape of the object which then can be compared to the density profile of an ellipse.

Given the basic equation of an ellipse

$$b^2x^2 + a^2y^2 = a^2b^2 \tag{11}$$

where a, b, $|a| \geq |b|$ constitutes the limits (Fig. 7). If we define c to be a point on the x axis we can rewrite Eq. (11) as

$$b^2c^2 + a^2y^2 = a^2b^2 \Leftrightarrow y = \frac{b}{a}\sqrt{a^2 - c^2}. \tag{12}$$

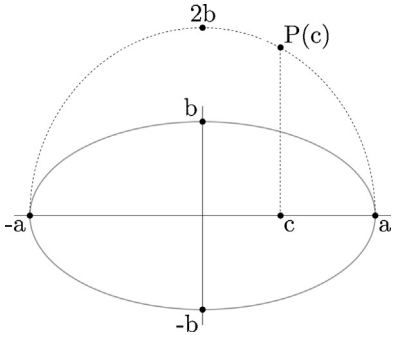


Fig. 7 – Calculating the density profile of a perfect ellipse.

The corresponding point on the density function can then be found to be

$$P(c) = 2\frac{b}{a}\sqrt{a^2 - c^2}. \quad (13)$$

We now have the ability to compare the object's density profile, Z, to the perfect ellipse density profile, $P(c)$, using the root mean square error (RMSE). For this comparison we approximate the variables a and b for $P(c)$ to be N_Z and $max(Z)/2$ respectively. Here N_Z is defined as the height of the object along its principle axis in pixels.

2.4. Texture analysis using regional Zernike moments

In addition to the size and shape of the nucleus, the texture is a very characteristic feature that can differentiate between nuclei and debris. There are many texture analysis methods described in the cell image analysis literature, see for instance [14]. We have chosen to use Zernike moments.

Zernike moments (ZMs) are a well-known group of mathematical tools that were introduced around 30 years ago. ZMs are complex moments orthogonal on a disk whose magnitude can be used as a rotation invariant image descriptor [25]. They are used for a variety of pattern recognition applications, and are known to be robust with regards to noise and to have a good reconstruction power [22]. Recently Sintorn and Kylberg [26] developed *regional Zernike moments* (RZMs) that combine the descriptive strength of the Zernike moments with the general and well established local filter operator. This allows for ZMs to be applied to the field of texture analysis. Sintorn and Kylberg show that RZMs work well as texture measures, outperforming Gabor filters, local binary patterns and Haralick features based on co-occurrence matrices on both discriminative power and noise sensitivity.

The ZM of order n with repetition l of function $f(r, \theta)$, in polar coordinates, inside a disk centred in a square image $I(x, y)$ of size $M \times M$, is written as:

$$\begin{aligned} A_{nl} &= \frac{n+1}{\pi}\sum_x\sum_y V^*_{nl}(r,\theta)I(x,y), \\ x,y &\in \sqrt{\left(x-\frac{M-1}{2}\right)^2 + \left(y-\frac{M-1}{2}\right)^2} < M/2, \end{aligned} \quad (14)$$

where the integer $n \geq 0$, $|l| \leq n$, and $n - |l|$ is even. $V^*_{nl}(r, \theta)$ denotes the complex conjugate of the Zernike polynomial $V_{nl}(r, \theta)$, that consists of an angular part $e^{il\theta}$ multiplied by a radial part R_{nl}, defined as

$$R_{nl} = \sum_{s=0}^{(n-|l|)/2} (-1)^s \frac{(n-s)!}{s!((n+|l|)/2)-s)!((n-|l|/2)-s)!} r^{n-2s} \quad . \quad (15)$$

To produce a rotation invariant texture measure, magnitudes from A_{nl} centred at each pixel in the texture image or region of interest (excluding the border where the patch $M \times M$ cannot fit within the region) are averaged. This can be seen as filtering the image or region with V_{nl} and averaging the result. The intensity is also normalized within each disk prior to calculating A_{nl} for each position.

2.5. Mean intensity deviation

Staining intensity varies considerably between different Pap-smears. The stain is not stoichiometric and, furthermore, stain intensity is highly batch dependent [5]. This means that samples are going to display both intra- and inter-sample variability with regard to nucleus stain intensity. This makes it impossible to set fixed thresholds for allowed intensity ranges of cell nuclei. Still, within one smear the distribution of average nuclear stain intensity is much more narrow than the stain intensity variation among debris objects. This fact can be used to help remove debris.

To cope with the significant stain intensity variations we use an adaptive approach. Rather than using pre-set gray level limits we determine the thresholds based on the average nuclear intensity distributions obtained from the remaining population of objects following the prior exclusion steps. By having this test at the end of the debris removal process, the objects that remain should consist of a majority of cell nuclei and thus define the distribution well. We can thus remove objects with an average gray value that lies outside an offset threshold of the average population intensity. This average can also be recalculated as more data is collected from the same specimen making the descriptor stronger as the sample analysis goes on.

3. Materials

The methods described in the previous section were tuned and tested on a range of Pap-stained material obtained as described in this section.

3.1. Image acquisition

Specimens were imaged using an Olympus BX51 bright-field microscope equipped with a 40×, 0.95 NA lens and a Hamamatsu ORCA-05G 1.4 Mpx monochrome camera, giving a pixel size of 0.25 μm. The microscope light path was filtered using a 570 nm bandpass filter, a wavelength previously shown to maximize the contrast of nuclei in Pap-smears [27].

To avoid getting nuclei that are not in focus, the microscope was fitted with a E-662 Piezo server controller (Physik Instrumente GmbH & Co. KG, Karlsruhe, Germany). This allowed for z-axis step control with a 0.1 μm resolution during image acquisition. Each field of view was photographed at 41 focus levels with a 0.4 μm step length. The focus stack was then combined into a single image using an extended depth-of-focus algorithm [28].

3.2. Datasets

The images were acquired from cervical cell specimen obtained from the Regional Cancer Center (RCC) in Thiruvananthapuram, Kerala, India. The specimen collection itself contained 63 standard Pap-smears and 5 LBC samples prepared using the ThinPrep system [29]. All specimens were diagnosed by expert cytotechnicians according to the Bethesda system [30]. Furthermore, for a subset of the specimens, an expert cytologist helped to collect a cell nucleus

database where ∼12,000 individual cells were diagnosed and annotated.

In total, 890 fields of view were collected from the available specimens. Based on the available material, three datasets were composed for this paper. The first dataset contained a total of 12,000 objects. Out of these, 6000 were diagnosed cell nuclei, of which 1472 were classified as either being cancer precursors or cancerous cells. The remaining 6000 were debris objects that were collected from segmentation results of randomly selected fields of view. The addition of malignant cells into the training data is important to avoid a systematic removal of cells with malignant appearance in the debris removal pipeline.

The second dataset consisted of 54 fields of view randomly selected from 37 of the classical Pap-smear specimen. Each field of view was initially segmented using the approach described in Section 3.3. The debris removal algorithm described in this paper was then used as a post-segmentation step. The result of the segmentation and debris removal algorithm was manually screened by placing a marker in each object that was incorrectly classified. The third dataset was generated the same way as the second one, with the sole exception that the 54 fields of view instead were collected from three of the LBC samples.

3.3. *Initial segmentation*

The method used for initial segmentation was developed by Kumar et al. [31]. The algorithm initially uses a Laplacian of Gaussian (LoG) filter to highlight edges in an image. The edgemap is thresholded to extract significant edges, and finally a set of binary operations, i.e., dilations, erosions and object filling, are used to obtain a collection of binary object masks.

This is a fairly simple approach to the segmentation task, often locating the cells in the image but also producing a good number of debris objects. This is actually an advantage for the study in this paper as the wide variety of debris objects serve as a good testing ground for the debris removal algorithm.

3.4. *Parameter settings*

All parameters except the intensity offset (Section 2.5) were optimised using two datasets created by randomly dividing the first dataset described in Section 3.2 into a training set, containing 4000 objects (2000 debris and 2000 nuclei of which ∼500 were non-normal), and a test set, containing the remaining 8000 objects.

For the area thresholding, the value was obtained manually using the method described in Section 2.1. The resulting threshold was set to 450 px^2, which amounts to a minimum nucleus diameter of 6 μm.

The values of [*A*, *B*] for the Fourier shape analysis (Section 2.2) were determined using a grid search approach with a quadratic Bayesian classifier as the method of evaluation. The values found were $A = 12$ and $B = 14$.

The order n and the kernel size M used for the regional Zernike moment (Section 2.4) calculations were optimized using a grid search approach with a quadratic Bayesian classifier as the method of evaluation. The values were set to $n = 5$ and $M = 6$.

The span of the intensity offset (Section 2.5) was set using the second dataset described in Section 3.2, containing whole image fields taken from standard Pap-smears. A total of 10 images randomly selected were used and the width of the intensity variation iteratively increased. The number of errors were minimized and the final offset was set to be ±50 graylevels.

3.5. *Computational complexity*

The algorithms described in this paper were implemented in MATLAB (2011b, The MathWorks, Natick, MA.) using the toolbox DIPimage [32] for image processing. Trials were run on a PC with a 2.66 GHz Intel Xeon® processor and 4 GB RAM. The average processing time for a single field of view was 30 s. No attempts towards speed optimisation have been made.

4. Results

The performance of the proposed debris removal algorithm was evaluated on three separate datasets described in Section 3.2. The main goal of each evaluation was to obtain a sensitivity and specificity score for separating an initial segmentation result (Section 3.3) into cell nuclei and debris. The parameter settings were accounted for in Section 3.4 and the experiments for the evaluation are presented in Section 4.1.

4.1. *Debris removal evaluation*

The described algorithms were evaluated on three levels using each of the datasets described in Section 3.2. An initial test was performed on the first dataset containing the 12,000 objects that were also used for the parameter optimization (Section 3.4). The same dataset was then used as the training set for the two following evaluations that were performed on random fields of view from LBC and smear samples, respectively. A collection of nuclei obtained from the LBC and smear dataset together with the classification result can be seen in Figure 8.

4.1.1. *Evaluation on single object database*

The first evaluation of the proposed method was performed using the object collection containing 6000 cell nuclei and 6000 debris objects. Initially, 4000 objects, equally divided between cells and debris, were randomly selected to be used as training data. The remaining 8000 objects were then fed into the pipeline illustrated in Fig. 2. The final step of the method, the mean gray value deviation, was omitted in this experiment since the objects are collected from a wide variety of samples and thus do not share the necessary intensity characteristics.

The results of the evaluation can be seen in Table 1. Rather than looking at the classification accuracy we noted the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values. These could be used to calculate the $Sensitivity = TP/(TP + FN)$ and the $Specificity = TN/(TN + FP)$ of the method which gave a more nuanced description of the method's performance.

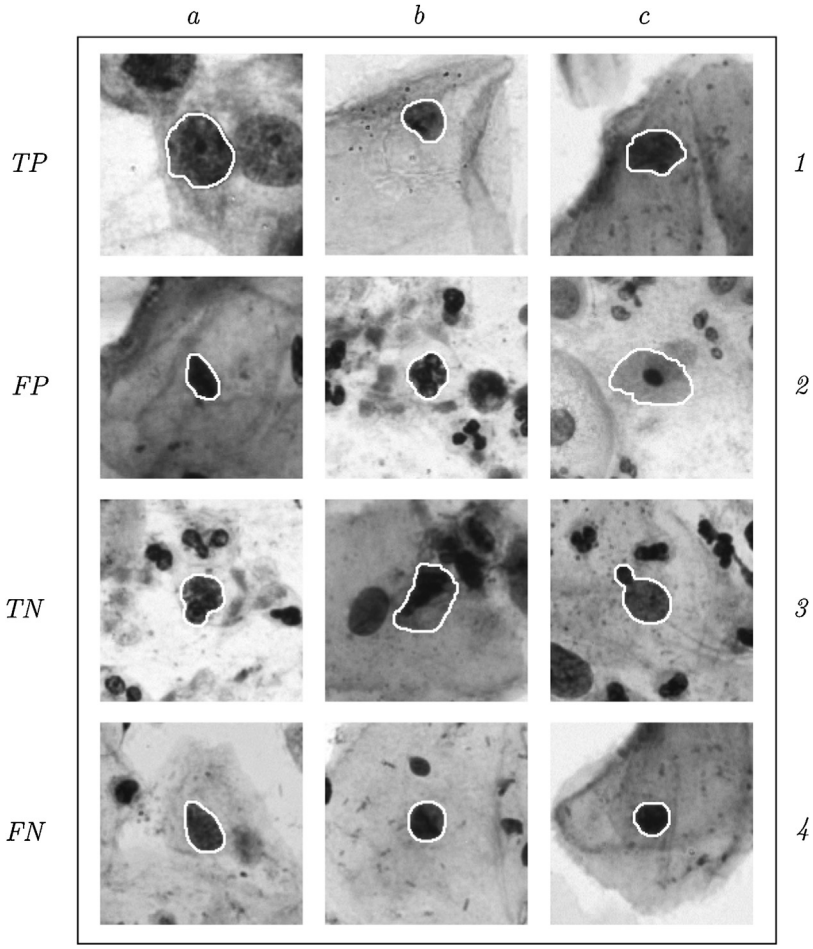


Fig. 8 – Classification results for a number off nuclei from the LBC and smear datasets. Each row (true positive (TP), false positive (FP), true negative (TN) and false negative (FN)) contain three different examples of classification outcomes (*a–c*). The white boundary around the objects shows the segmentation mask.

Using the this dataset we had the possibility to see the number of malignant cell nuclei that were classified as debris. This is, as has already been mentioned, important since a method that systematically filters out malignant cells will not work in an actual screening situation. For this evaluation, the number of malignant cells correctly classified as "cells" was 817 out of 963 (84.8%).

4.1.2. *Evaluation on LBC specimen*

For the second evaluation, the described method was applied to entire fields of view obtained from LBC specimen. In total, 8624 objects were segmented in the 54 fields of view. Out of these, the 793 cell nuclei present in the images have been manually annotated. The results of the debris removal method can be seen in Table 2. Out of the 54 fields used in the evaluation, false positives were found in 10.

4.1.3. *Evaluation on smear specimen*

For the final evaluation, the described method was applied to entire fields of view obtained from classical smear specimen. As for the LBC evaluation, a total of 54 images were analysed. However, the number of objects in the smear images

Table 1 – Classification results for the single object dataset.

		True class	
		Nucleus	*Debris*
Test outcome	*Nucleus*	**TP** 3743	**FP** 0
	Debris	**FN** 257	**TN** 4,000
		Sensitivity 0.9358	*Specificity* 1.000

Table 2 – Classification results for LBC specimen field of views.

		True class	
		Nucleus	*Debris*
Test outcome	*Nucleus*	**TP** 777	**FP** 13
	Debris	**FN** 16	**TN** 7818
		Sensitivity 0.9798	*Specificity* 0.9983

Table 3 – Classification results for smear specimen field of views.

		True class	
		Nucleus	*Debris*
Test outcome	*Nucleus*	**TP** 1114	**FP** 69
	Debris	**FN** 14	**TN** 11,138
		Sensitivity 0.9876	*Specificity* 0.9938

amounted to 12,335, out of which 1128 were determined to be nuclei with diagnostic value. The classification results can be found in Table 3. Out of the 54 image fields used in the evaluation, false positives were found in 26.

4.2. *Feature importance*

To evaluate the importance of the different steps of the algorithm, the number of discarded objects at each stage, for the 108 images included in the second and third evaluations were collected for analysis. The results can be found in Table 4.

5. Discussion

One can deduce by the many varied approaches to the segmentation task that, as is also true for many other biological applications, cervical cell specimens are very hard to partition reliably. Focusing on the process of nucleus segmentation, difficulties such as blood cells, inflammatory cells, distorted cells, overlapping objects, etc. need to be handled. However, since these problems can present themselves in an infinite number of variations, it is a task that is very hard. Therefore, most segmentation algorithms will produce a significant proportion of nucleus candidates that are not properly segmented nuclei, but various other structures. When these objects are used in a classifier designed to find indications of malignancy, they may just as likely be classified as malignant or as normal. As much effort as is spent on creating a robust segmentation algorithm should therefore be spent on removal of these unwanted objects.

A further motivation for the importance of this can be found in the statistics related to what the algorithms are actually looking for. A standard Pap-smear may typically contain 100,000–200,000 cells of the relevant cell-types. Out of these, a cytotechnician, or an automated screening system, is expected to detect a pre-cancerous condition even if only as few as 10–20 diagnostic (pre-malignant or malignant) cells are present; ideally a single malignant cell should be enough. A classifier with a one percent false positive error rate will find 1000–2000 diagnostic cells on a healthy sample, making the system useless. One approach to deal with this problem is to make the classifier highly asymmetrical between false-positive and false-negative, i.e., allowing it to miss-classify a large fraction of the actually malignant cells as normal. This may seem to defeat the purpose of the system which is to detect (pre-)malignancy. But the highly unbalanced numbers work that way. If the system has a false negative rate as high as 80% it will still detect 2–4 of the diagnostic malignant cells if we have 10–20 available. This is acceptable as long as the false positive rate is virtually zero, less than 0.001%, assuming we have a simplistic classification strategy of calling a specimen positive if we find at least one diagnostic (pre-)malignant cell.

Creating such a classifier is a daunting task but in principle possible if it is working with perfectly imaged cells and carefully extracted features. We can also have more sophisticated approaches, e.g., grading cells rather than classifying them in a binary way. A very different strategy is to look for very subtle malignancy associated changes among also the seemingly normal cells [33–35]. In that case the number statistics becomes more manageable but we will have very high demands on the quality of the features we extract, still making it mandatory not to have debris among the objects that are analysed as cells. Thus, independent of what classification approach we try to adapt, it rapidly breaks down if we have to deal with debris of various kinds. Debris may equally likely look like a malignant or a normal cell. If only a few percent of the objects that are brought to the classifier are debris, the classifier will likely fail.

In this paper we have presented a method for locating cell nuclei of diagnostic value in a segmentation result containing both nuclei and debris objects. The method operates in a sequential fashion allowing more detailed analysis to be performed at a stage where simpler to discard objects already have been removed, saving computational power and reducing the dimensionality of the classification problem. Each object passed on as a cell nucleus has been analysed based on size, shape, texture and average intensity. Table 4 records the effect of the different steps. Since the performance of each step of the algorithm is dependant on the previous one, the values do not correspond to the individual effectiveness of the descriptors. However, the results show that each step does contribute. It is also clear that the initial, low cost steps are able to remove the bulk of the objects leaving only the more difficult debris objects for the more complex, and computationally expensive, descriptors.

The method has been evaluated on three datasets ranging from isolated objects to standard fields of view from both LBC and smear specimen. The results from the evaluations show that the proposed method performs well over the entire range

Table 4 – Number and percentage of objects being removed at each stage collected from 108 fields of view (56 LBC and 56 smear).

Area	Basic shape	Elliptical deviation	Zernike textures	Intensity deviation	Remaining FP
13,130 (68.62%)	5,249 (27.43%)	505 (2.64%)	194 (1.01%)	45 (0.30%)	82 (0.42%)

of increasingly difficult sampling situations. As is seen in Figure 8, the algorithm is able to correctly classify difficult objects. However, certain objects in the figure, such as the cluster of leukocytes (*FP-b*) and the parabasal cell (*FP-c*), indicate that there still exists room for improvement.

Our goal with this paper was to describe a method that was entirely focused on the problem of debris removal. Future work includes evaluating the method in a clinical setting, exposing the approach to a continuous flow of samples. Initially, however, the algorithm needs to be sped up by a substantial amount, something that will be achieved by implementing the method in the C++language. Furthermore, a continuous update of the training data will be required to adapt to new variations likely to appear during extended trials.

Conflicts of interest

None declared.

Acknowledgements

The Swedish part of this work is funded by the Swedish Research Council (2008-2738) and VINNOVA (2008-01712).The part of the work which is done in India is funded by the Department of Information Technology of the Indian Government.

REFERENCES

[1] WHO/ICO;1; Information Centre on HPV and Cervical Cancer (HPV Information Centre), Human papillomavirus and related cancers in world. summary report 2010 (June 2012). www.who.int/hpvcentre

[2] G.N. Papanicolaou, H.F. Traut, M. Stanton, A. Friedberg, Diagnosis of Uterine Cancer by the Vaginal Smear, Oxford University Press, New York, 1943.

[3] World Health Organization, Comprehensive Cervical Cancer Control: A Guide to Essential Practice, WHO Press, 2006.

[4] D. Saslow, D. Solomon, H.W. Lawson, M. Killackey, S.L. Kulasingam, J. Cain, F.A.R. Garcia, A.T. Moriarty, A.G. Waxman, D.C. Wilbur, N. Wentzensen, L.S. Downs, M. Spitzer, A.-B. Moscicki, E.L. Franco, M.H. Stoler, M. Schiffman, P.E. Castle, E.R. Myers, American cancer society, American society for colposcopy and cervical pathology, and American society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer, American Journal of Clinical Pathology 137 (4) (2012) 516–542.

[5] H. Grohs, O. Husain (Eds.), Automated Cervical Cancer Screening, IGAKU-SHOIN Medical Publishers, Inc., 1994.

[6] D. Jhala, I. Eltoum, Barriers to adoption of recent technology in cervical screening, Cytojournal 4 (2007) 16–22.

[7] T.M. Elsheikh, R.M. Austin, D.F. Chhieng, F.S. Miller, A.T. Moriarty, A.A. Renshaw, American society of cytopathology workload recommendations for automated pap test screening: developed by the productivity and quality assurance in the era of automated screening task force, Diagnostic Cytopathology 1 (2012), n/a-n/a.

[8] E. Bengtsson,;1; Computerized cell image processing in healthcare, in: Choi, HK (Ed.), Healthcom 2005: 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry, Proceedings, Korea Multimedia Soc; IEEE; Inje Univ; Busan Convent Bur, IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2005, pp. 11-17.

[9] C. Bergmeir, M.G. Silvente, J.M. Benítez, Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework, Computer Methods and Programs in Biomedicine 107 (2012) 497–512.

[10] P. Bamford, B. Lovell, Unsupervised cell nucleus segmentation with active contours, Signal Processing 71 (2) (1998) 203–213.

[11] C.-H. Lin, Y.-K. Chan, C.-C. Chen, Detection and segmentation of cervical cell cytoplast and nucleus, International Journal of Imaging Systems and Technology 19 (3) (2009) 260–270.

[12] P. Malm, A. Brun, Closing curves with Riemannian dilation: application to segmentation in automated cervical cancer screening, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, J.-X. Wang, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M.L. Encarnaç ao, C.T. Silva, D. Coming (Eds.), Advances in Visual Computing, vol. 5875 of Lecture Notes in Computer Science, Springer, 2009, pp. 337–346.

[13] A. Gençtav, S. Aksoy, S. Önder, Unsupervised segmentation and classification of cervical cell images, Pattern Recognition 12 (2012) 4151–4168.

[14] K. Rodenacker, E. Bengtsson, A feature set for cytometry on digitized microscopic images, Analytical Cellular Pathology 25 (1) (2003) 1–36.

[15] B. Nielsen, F. Albregtsen, H.E. Danielsen, Statistical nuclear texture analysis in cancer research: a review of methods and applications, Critical Reviews™ in Oncogenesis 14 (2–3) (2008) 89–164.

[16] E. Bengtsson, O. Eriksson, J. Holmquist, B. Nordin, B. Stenkvist, High resolution segmentation of cervical cells., Journal of Histochemistry & Cytochemistry 27 (1) (1979) 621–628.

[17] N. Tanaka, H. Ikeda, T. Ueno, S. Watanabe, Y. Imasato, Fundamental study of automatic cyto-screening for uterine cancer. ii. Segmentation of cells and computer simulation., Acta Cytologica 21 (1) (1977) 79–84.

[18] L. Xu, E. Oja, P. Kultanen, A new curve detection method: randomized Hough transform (RHT), Pattern Recognition Letters 11 (5) (1990) 331–338.

[19] R. Gonzalez, E. Woods, Digital Image Processing, 3rd ed., Pearson Education, 2008.

[20] R.C. Mellors, R. Silver, A micro-fluorometric scanner for the differential detection of cells; application of exfoliative cytology., Science 114 (2962) (1951) 356–360.

[21] A. Mehnert,;1; Image analysis for the study of chromatin distribution in cell nuclei with application to cervical cancer screening, Ph.D. thesis, School of Information Technology and Electrical Engineering, The University of Queensland, 2004.

[22] D. Zhang, G. Lu, Review of shape representation and description techniques, Pattern Recognition 37 (1) (2004) 1–19.

[23] D. Zhang, G. Lu, A comparative study on shape retrieval using Fourier descriptors with different shape signatures, Journal of Visual Communication and Image Representation 1 (14(1)) (2003) 41–60.

[24] E. Bengtsson, O. Eriksson, J. Holmquist, T. Jarkrans, B. Nordin, B. Stenkvist, Segmentation of cervical cells: detection of overlapping cell nuclei, Computer Graphics and Image Processing 16 (4) (1981) 382–394.

[25] J. Flusser, T. Suk, B. Zitová, Moments and Moment Invariants in Pattern Recognition, John Wiley & Sons Ltd., 2009.

[26] I.-M. Sintorn, G. Kylberg, Regional Zernike moments for texture recognition., in: 21th International Conference on Pattern Recognition (ICPR), 2012, 2012, pp. 1625–1638.

[27] J. Holmquist, Y. Imasoto, E. Bengtsson, B. Olsen, B. Stenkvist, A microspectrophotometric study of papanicolaou-stained

cervical cells as an aid in computerized image processing., Journal of Histochemistry & Cytochemistry 24 (12) (1976) 1218–1224.
[28] B. Forster, D. Van de Ville, J. Berent, D. Sage, M. Unser, Extended depth-of-focus for multi-channel microscopy images: a complex wavelet approach., in: 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano, vols. 1 and 2, IEEE, Arlington, VA, 2004, pp. 660–663.
[29] Hologic inc.,;1; URL http://www.hologic.com (July 2012).
[30] R.J. Kurman, D. Solomon, The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses, Springer-Verlag, New York, 1994.
[31] R. Kumar, V. Kumar, K. Sharath, S. Sudhamony, R. Ravindrakumar, Detection and removal of artifacts in cervical cytology images using support vector machine, in: International Symposium on IT in Medicine and Education (ITME), 2011, vol. 1, 2011, pp. 717–721.
[32] C. L. Luengo Hendriks, L. J. van Vliet, B. Rieger, M. van Ginkel,;1; DIPimage: a scientific image processing toolbox for MATLAB, Computer Program, http://www.diplib.org/ (1999-).
[33] M. Guillaud, A. Doudkine, D. Garner, C. MacAulay, B. Palcic, Malignancy associated changes in cervical smears: systematic changes in cytometric features with the grade of dysplasia., Analytical Cellular Pathology 9 (3) (1995) 191–204.
[34] R. Kemp, C. MacAulay, D. Garner, B. Palcic, Detection of malignancy associated changes in cervical cell nuclei using feed-forward neural networks, Analytical Cellular Pathology 14 (1) (1997) 31–40.
[35] J. Hallinan,;1; Detection of malignancy associated changes in cervical cells using statistical and evolutionary computation techniques, Ph.D. thesis, The University of Queensland, 1999.