Automated detection of working area of peripheral blood smears using mathematical morphology

Jesús Angulo [a,*] and Georges Flandrin [b]

[a] *Centre de Morphologie Mathématique, Ecole des Mines de Paris, France*
[b] *Laboratoire Central d'Hématologie, Hôpital Necker-Enfants Malades de Paris, France*

Abstract. The paper presents a technique to automatically detect the working area of peripheral blood smears stained with May-Grünwuald Giemsa. The optimal area is defined as the well spread part of the smear. This zone starts when the erythrocytes stop overlapping (on the body film side) and finishes when the erythrocytes start losing their clear central zone (on the feather edge side). The approach yields a quick detection of this area in images scanned under low magnifying power (immersion objective ×25 or ×16). The algorithm consists of two stages. First, an image analysis procedure using mathematical morphology is applied for extracting the erythrocytes, the centers of erythrocytes and the erythrocytes with center. Second, the number of connected components from the three kinds of particles is counted and the coefficient of spreading ρ_s and the coefficient of overlapping ρ_o are calculated. The data from fourteen smears illustrate how the technique is used and its performance.

Colour figures can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

Keywords: Automation, quality control, telepathology, image processing, mathematical morphology, erythrocytes, blood smear, blood spreading

1. Introduction

Historically examination under microscope of good quality blood smears has been the best way to estimate the number of leukocytes or white blood cells; for leukocyte differentiation; to study the morphology of erythrocytes or red blood cells; to characterise the morphology of lymphocytes and; to calculate the number and morphology of platelets [9].

From the last 20 years automated systems for leukocyte recognition are currently available in the market and are used in clinical laboratory routines. These generally rely on flow cytometry techniques whereby a blood sample flows through a detector and is then discarded. In addition these devices reach the limited aims of identifying normally circulating leukocytes and of

flagging abnormal circulating cells. Thus examination of stained peripheral blood smears remains necessary for detecting and classifying abnormal cells [4], particularly to study the morphology of lymphocytes which is regarded as the principle basis for the identification and discrimination among the chronic lymphoproliferative disorders [5,10].

From a methodological point of view there are two desirable qualities of a blood smear. These are the sufficient working area, defined as the well *spread* part of the smear, and an adequate quality and reproducibility of the *staining* procedure. Both have been previously studied, see Benattar and Flandrin [3,4].

Following some research works on haematological cytology image analysis for microscopic diagnosis [11,12], the aim of this paper is to present a powerful application of mathematical morphology in order to automatically detect the working area of a peripheral blood smear using the fields scanned under low magnifying power (immersion objective ×25 or ×16,

*Corresponding author: Jesús Angulo, Centre de Morphologie Mathématique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau, France. Tel.: + 33 1 64 69 48 00; Fax: + 33 1 64 69 47 07; E-mail: angulo@cmm.ensmp.fr.
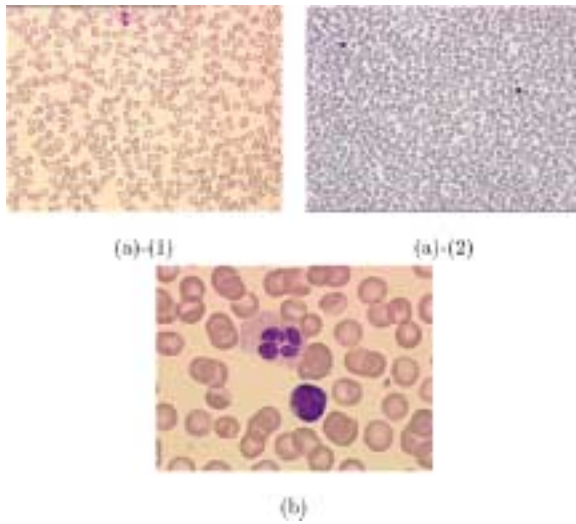
Fig. 1. Peripheral blood smear stains with May-Grünwald Giemsa: (a) Low magnifying power, (1) immersion objective ×25 and (2) immersion objective ×16. (b) High magnifying power, immersion objective ×100. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

see Fig. 1(a)). In a strategy of automation using a motorised microscope and within a telepathology context, once the optimal area is detected, the following step is to increase the magnifying power (immersion objective ×100, see Fig. 1(b)) in order to locate the interesting fields which contain some leukocyte and in this case to acquire the fields for storing, transmitting or processing the images.

1.1. State of the art

Classification of blood cells from peripheral blood smears is an established problem in image analysis and pattern recognition which started over four decades ago, Ingram et al. [17], Young [22]; and which continues still to arouse interest. See in Beksaç et al. [6], the use of a neural network for blood cell recognition, and in Theera-Umpom and Gader [21], the application of morphological granulometries for the same purpose. In the early 80 s, image analysis based systems for leukocyte differential counting were introduced and several of these systems were quite capable of differentiating between many types of immature and abnormal cells, see Haus et al. [16]. At present, several systems based on automated microscopy and image processing are under development for differential blood cell analysis, see Comaniciu et al. [8], and _CellaVision_ Company [7]. However we did not find any reference relating to the automated detection of the working zone of peripheral

blood smears using image analysis and the above referenced systems do not claim any characteristic about this issue.

2. Materials and methods

2.1. Background

In order to define the working area of a peripheral blood smear we have to examine the smear under low magnifying power in such a way that we have in the field a sufficiently large sample of the smear zone, i.e., enough number of cells for an understanding overview. Our methodology is valid for different magnification powers whenever this essential requirement is observed. For instance, in this study we worked with images obtained from oil-immersion objective ×25 and oil-immersion objective ×16. We believe that these two values are respectively the upper and lower limit of validity of the approach.

In the stained images, the erythrocytes (ennucleated cells) have a reddish color from the haemoglobin and usually appear round or oval with a pale-staining central region, caused by their biconcavity [9].

The working area or optimal area is defined as the well spread part of the blood smear. This zone starts on the body film side, when erythrocytes stop overlapping (piled on top of one another) and finishes on the feather edge side, when erythrocytes start to lose their clear central zone (they were not well spread). See in Fig. 2 the series of 20 images which illustrate this definition. The rationale behind our algorithm for automatic detection of the optimal area is the quantification of these two erythrocyte phenomena, the spread and the overlap.

2.2. Peripheral blood smear images

In this paper, we used the images from 14 peripheral blood smears stained with May-Grünwuald Giemsa using an uniform and stable technique [3]. For each one, the acquisition of several fields yields a succession of images which covers the complete smear area. For this study, the manual technique for selecting the field images on the film is the following: the vertical position corresponds approximately to the middle of the film and on the horizontal dimension, a carefully sampling allows to cover the total smear. These microscopic images were acquired at controlled light intensities (the RGB camera is calibrated to ensure correct color registration) using a ICG TRIBVN workstation.
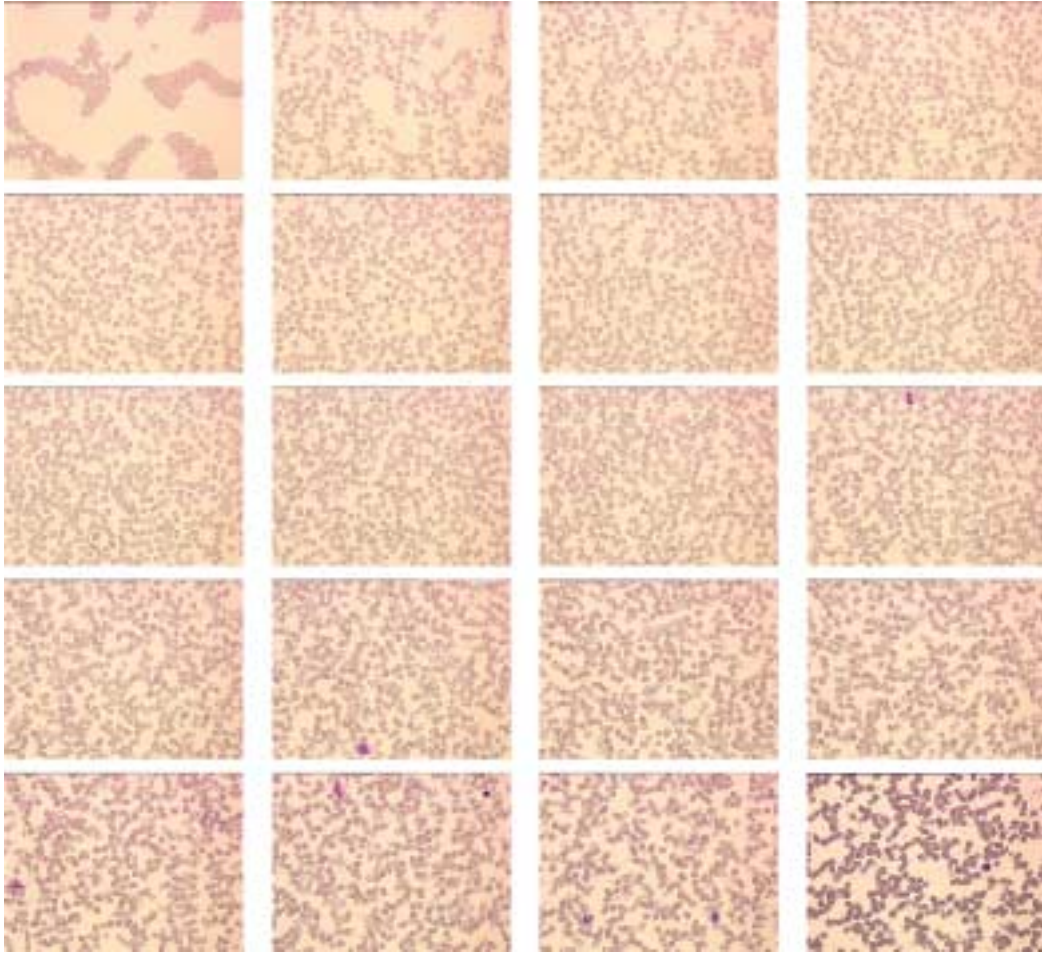
Fig. 2. Series of 20 blood images obtained from the same film and covering the whole smear (*Smear 0*, ×25). This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

The fourteen image folders are,

– *Smear 0* (×25): 20 images used to describe in detail the foundation of the algorithm and for the discussion which provides the morphological quantitative definition of working area.
– *Smear A*, *Smear B* and *Smear C* (×16): the results of the application of the approach on these smears (each one with 20 images) illustrate how the technique is used.
– *Smear 1, Smear 2, ..., Smear 10* (×16): the analysis of these smears (10 images by smear) shows by a statistical study the performance of the method.

2.3. Mathematical morphology

First introduced as a shape-based tool for binary images, mathematical morphology has become a very powerful nonlinear image analysis technique with operators capable of handling sophisticated image processing tasks in binary, grey-scale, color and multiresolution imaging modalities. A tutorial in the technique can be found in Serra [18,19]. In this section we briefly define the basic morphological operators used in this paper. The practical effects are shown in Fig. 3 by means of a grey-tone image series.

In the framework of digital grids, a *grey-tone image* can be represented by a function $f : D_f \to T$, where D_f is a subset of $\mathbf{Z}^2$ and $T = \{t_{\min}, \ldots, t_{\max}\}$ is an ordered set of grey-levels. Let B be a subset of $\mathbf{Z}^2$ and $s \in \mathbf{N}$ a scaling factor. sB is called *structuring element* (shape probe) B of size s. The basic morphological operators are

– Dilation: $\delta_B(f(x)) = \sup_{y \in B}\{f(x - y)\}$,
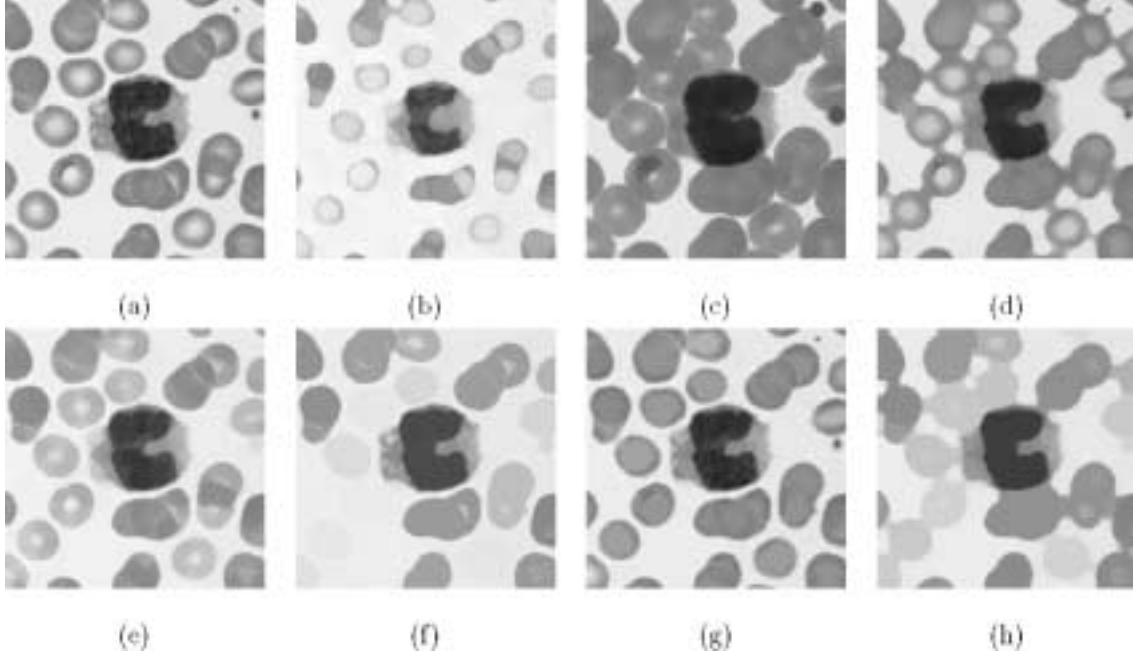– Erosion: $\varepsilon_B(f(x)) = \inf_{-y \in B}\{f(x - y)\}$.

Fig. 3. Effects of morphological operators on grey-tone structures: (a) Initial image, f, (b) dilation, $\delta_{sB}(f)$ (fading of dark structures and expansion of white structures), (c) erosion, $\varepsilon_{sB}(f)$ (dual effects), (d) opening, $\gamma_{sB}(f)$ (filter out light structures), (e) closing, $\varphi_{sB}(f)$ (filter out dark structures, in the four cases $s = 10$ and B is a circle), (f) closing by reconstruction, φ^{rec} (using as marker a closing of size $s = 20$ and a dual reconstruction; the dark structures with size smaller than 20 are removed and the contours are preserved), (g) dual *close-hole* operator, $\psi^{\text{clohole}}(f)$ (the dual holes–light structures inside dark bordered regions – that do not touch the image boundary are filled) and (h) dual *size-and-close-grain* operator, $\psi^{\text{clogra}}(f)$ (dark structures smaller than a size are removed and the dark structures are merged).

The two elementary operations of *erosion* and *dilation* can be composed together to yield a new set of operators having desirable feature extractor properties which are given by

– Opening: $\gamma_B(f) = \delta_B[\varepsilon_B(f)]$,
– Closing: $\varphi_B(f) = \varepsilon_B[\delta_B(f)]$.

The morphological *openings* (*closings*) filter out light (dark) structures from the images according to the predefined size and shape criterion of the structuring element.

A morphological tool that complements the opening and closing operators for feature extraction (extract the marked particles) is the morphological reconstruction, implemented using the *geodesic dilation* operator based on restricting the iterative dilation of a function marker f by B to a function mask g,

$$\delta_g^n(f) = \delta_g^1 \delta_g^{n-1}(f),$$

where $\delta_g^1(f) = \delta_B(f) \wedge g$. The *reconstruction* by dilation is defined by

$$\gamma^{\text{rec}}(g, f) = \delta_g^i(f)$$

such that $\delta_g^i(f) = \delta_g^{i+1}(f)$ (idempotence).

There are two composed operators derived from the geodesic reconstruction that are of great importance for the application in this paper are the *close-hole* operator and the *size-and-close-grain* operator.

– Close-hole: this operator fills all holes in an image f that do not touch the image boundary f_∂ (used as a marker) and therefore provides a parameter free approach to detect holes in an image:

$$\psi^{\text{clohole}}(f) = \left[\gamma^{\text{rec}}\left(f^c, f_\partial\right)\right]^c.$$

– Size-and-close-grain: this operator removes the small light structures or grains (smaller than λB_1) using an opening by reconstruction and closes the packets of grains when they are sufficiently near (gap smaller than μB_2 where B_2 is a circle) using an isotropic closing

$$\psi^{\text{clogra}}_{\lambda B_1, \mu B_2}(f) = \varphi_{\mu B_2}\left[\gamma^{\text{rec}}\left(f, \gamma_{\lambda B_1}(f)\right)\right].$$

2.4. Cell segmentation by thresholding

In blood images scanned under low magnification power, two regions may be observed: the plasma (background) and the cells (erythrocytes, leukocytes and platelets). The histogram summarises the grey-level contents of an image. In this case, the histogram of the filtered green channel $h_G[n]$ is typically bimodal. The cell segmentation by thresholding at the optimal value, u_T, between the two modal peaks, n_1 and n_2, yields a binary image which matches the leukocytes, the platelets and the erythrocytes without their central zone. Different methods, see the survey on image segmentation by Fu and Mui [13] or the overview and comparison of eleven common histogram-based thresholding methods in Glasbey [14], can be employed to determine u_T.

We propose a new approach which combines the classical selection of the thresholding value by minimising the sum of within class variances with a morphological technique for selecting the central mode values which speeds up the convergence to u_T. The algorithm, divided into three steps, is summarised as follows,

Smooth histogram In order to obtain the smooth histogram $\hat{h}_G[n]$, we filter the histogram by averaging every bin with its neighbours (anti-causal, zero-phase filter of size $N = 5$). In Fig. 4(b) the interest of this step is shown: after smoothing, the number of maxima is reduced and the computation of the next step is faster. On the other hand, this preprocessing step stabilises and enhances the histogram modes in order to avoid that in an extreme situation a secondary maxima could have a significant contrast.

Maxima of the histogram In order to find the two peak maxima, we use a morphological tool: the dynamics or contrast extinction value. Introduced by Grimaud [15], this measure of contrast maps each maximum with a value: the contrast of a maximum is the minimum descent necessary to move from the maximum to another higher maximum (the contrast of the highest maximum is defined as the difference between the maximum and minimum of the function). Figure 4(a) illustrates the contrast of histogram maxima. Then, we can select the two highest maxima, n_1 and n_2, which square with the central value of each mode. This definition of histogram modes as the histogram peaks of largest dynamic (the most relevant peaks) has been previously used by Soille [20].

Variance criterion Using a statistical test, we obtain the best histogram partition u_T with respect to a criterion of variance C,

$$C = \min_{u_T}\left\{ \sum_{n=t_{\min}}^{u_T} (n - n_1)^2\, \hat{h}_G[n] + \sum_{n=u_T}^{t_{\max}} (n - n_2)^2\, \hat{h}_G[n] \right\}.$$

2.5. Algorithm

Our algorithm can be divided into two parts. First, an image analysis procedure using mathematical morphology is applied for extracting the erythrocytes, the centers of erythrocytes and the erythrocytes with center. Second, the number of connected components from the three kinds of particles is counted and the coefficient of spreading and the coefficient of overlapping are calculated.

2.5.1. Image analysis

The aim of the image analysis step is to segment and classify the cells, mainly the erythrocytes. We are interested in the morphology of erythrocytes: presence or absence of clear center and overlapping of neighbouring cells. The distinction between erythrocytes with and without center and between separated and overlapped erythrocytes are required. The first task is easy, we extract on one hand the cells and on the other the centers (brighter than the body), afterwards by *reconstruction*, the cells with center are obtained. The second task is less direct, for identifying the overlapped cells we consider that after segmentation two overlapped cells are only a connected component nevertheless they have probably two centers. The presence of some leukocytes is not a serious error since their contribution from a statistical viewpoint is negligible. However, it is desirable to extract the platelets and other small artefacts which could be numerous.

Let f be the blood smear color image. After comparing several color spaces, we found that the contours of the erythrocytes appear most continuous and contrasted against the background as well as a better contrast between center and body in the green channel f_G of the RGB color space. After pre-filtering and binarisation, it is enough to apply morphological filtering to binary images. There are two reasons behind the use of binary operators: the first one is for particle counting we have to use a binary image and on the other hand using binary operators the algorithm is signifi-
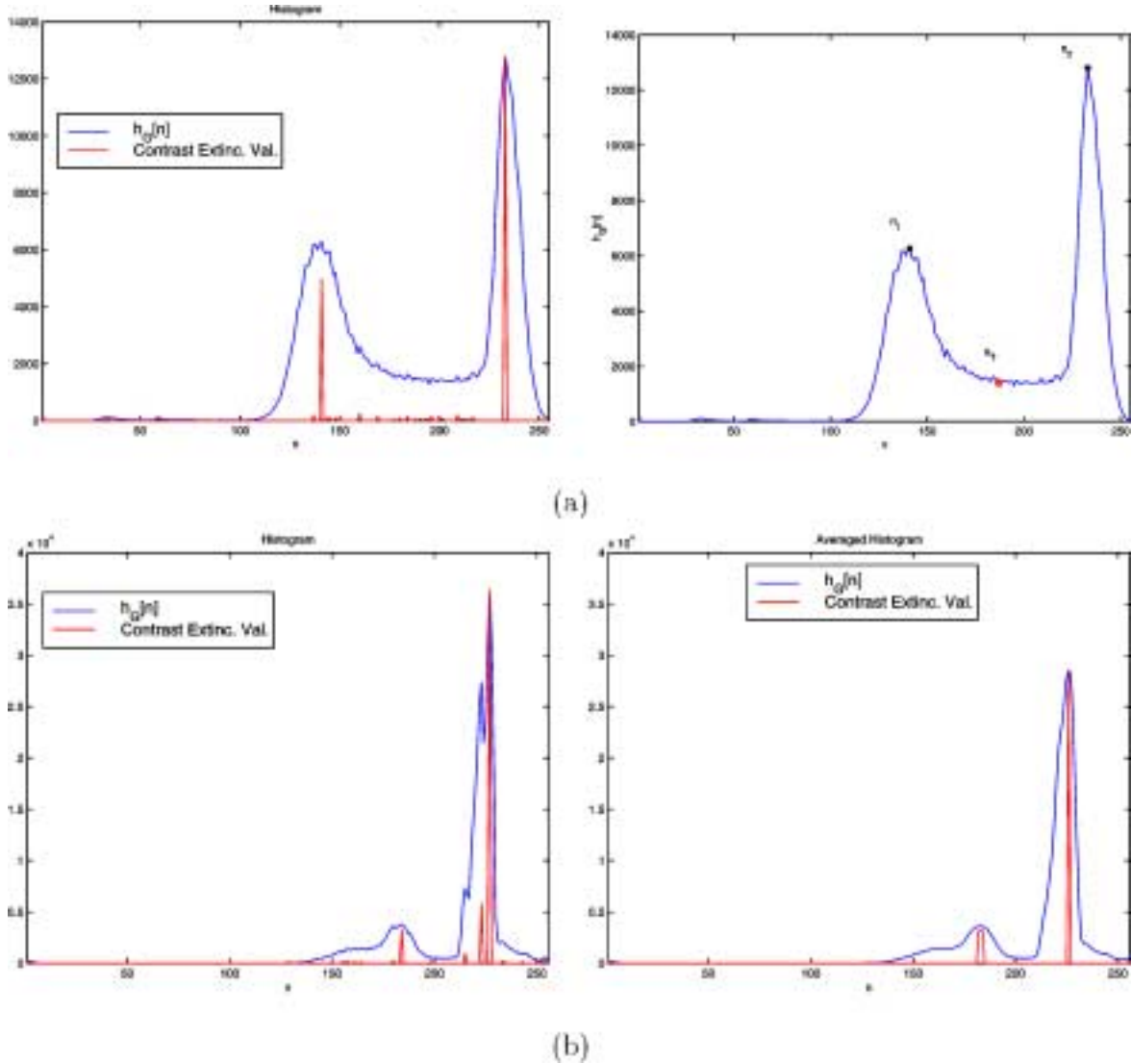
(a)

(b)

Fig. 4. Histograms and optimal thresholding: (a) On the left, histogram and contrast extinction values; on the right, maxima n_1 and n_2 (the two peaks of largest dynamic or the first two most relevant peaks), and optimal threshold value u_T. (b) On the left, histogram and contrast extinction values; on the right, smoothed histogram and new contrast extinction values. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

cantly faster (to envisage the integration of this algorithm in a microscopic system). The algorithm can be summarised as follows (the intermediate images which allow to follow the different steps are shown in Fig. 6),

Filtering In a first step we filter the image f_G using a median filter MF_n of size $n = 3$

$$\hat{f}_G = MF_n(f_G),$$

in order to remove the noise and small mistakes of digitalisation. In fact, this pre-filtering which

is the same for images under $\times 25$ and $\times 16$ allows that the possible Gaussian and flicker noises, introduced in the optical and image acquisition chain system, are controlled before the thresholding.

Binarisation The histogram of $\hat{f}_G$ is typically bimodal, i.e., two object categories: the cells (erythrocyte bodies, leukocytes and platelets) and the background (plasma and erythrocyte center). A simple thresholding process at u, i.e., the optimal value between the two modes, $T_{[t_{\min}, u]}$, seg-
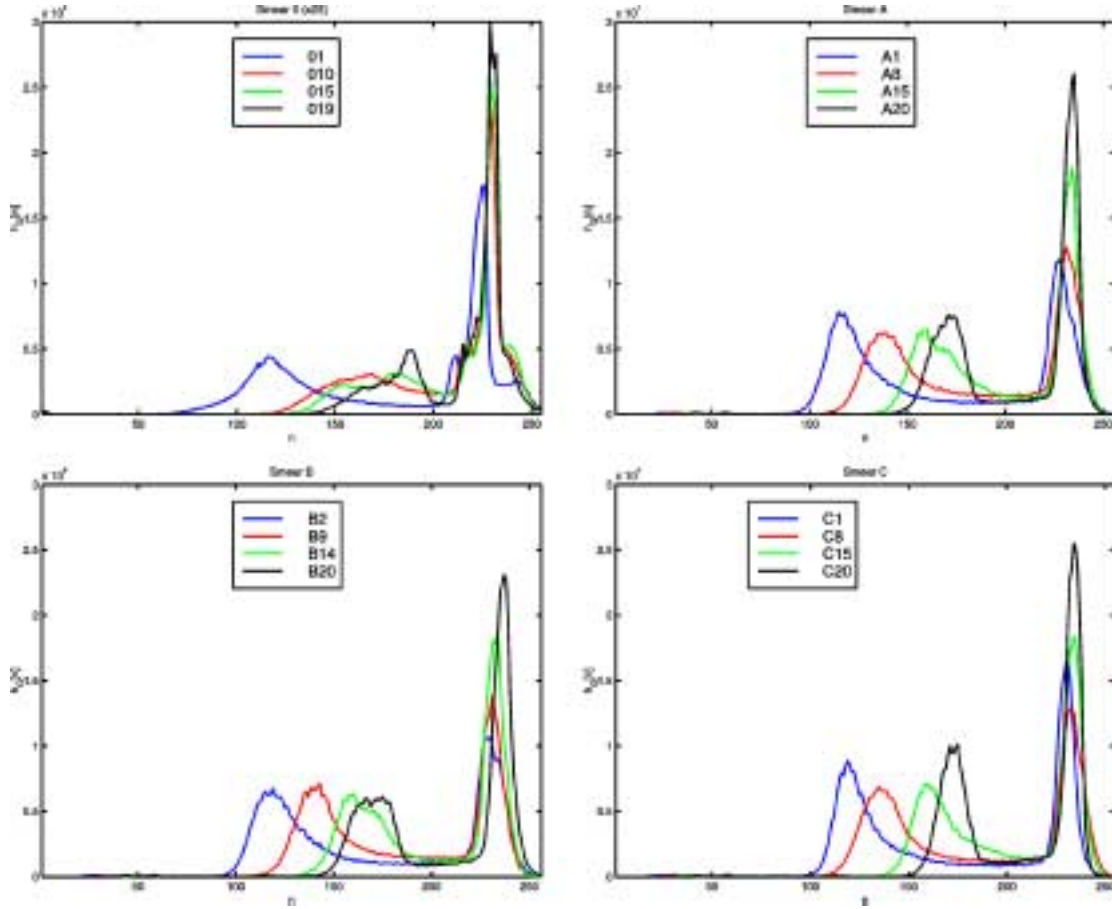
Fig. 5. Histograms of some images from the four smears test (for each one, the green component histogram of 4 random images is depicted): the histograms are typically bi-modal but the optimal value which separates the two categories depends on the image. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

ments the two categories obtaining a binary image

$$I = T_{[t_{\min},u]}(\hat{f}_G).$$

The choice of u is crucial since it is important to differentiate between center and body on the erythrocytes. If u is chosen too low, cell bodies are broken, if it is set too high, cell and background are not separated. In Fig. 5 some examples of histograms are shown.

The overlapping phenomenon determines the locus of the first mode and the variations in the staining procedure condition the relative position of histogram. In order to automatically control this lack of definition, the optimal u is chosen by means of the adaptive morphological thresholding, presented above.

Extraction of cells In I, the cells without their centers are included. In order to fill all the holes in the cells, we apply a close-hole operator

$$I_{\text{cl}} = \psi^{\text{clohole}}(I).$$

Extraction of platelets and artefacts As we know approximately their size, in order to eliminate the platelets we can apply a reconstruction using an opening as the marker

$$I_{\text{cl2}} = \gamma^{\text{rec}}(I_{\text{cl}}, \gamma_{s_1 B})$$

with s_1 such that $s_1 B$ is larger than the size of the platelets; the structuring element B is a circle. The result is a binary mask of cells, knowing that a certain amount of them could be overlapped. By difference, we can obtain
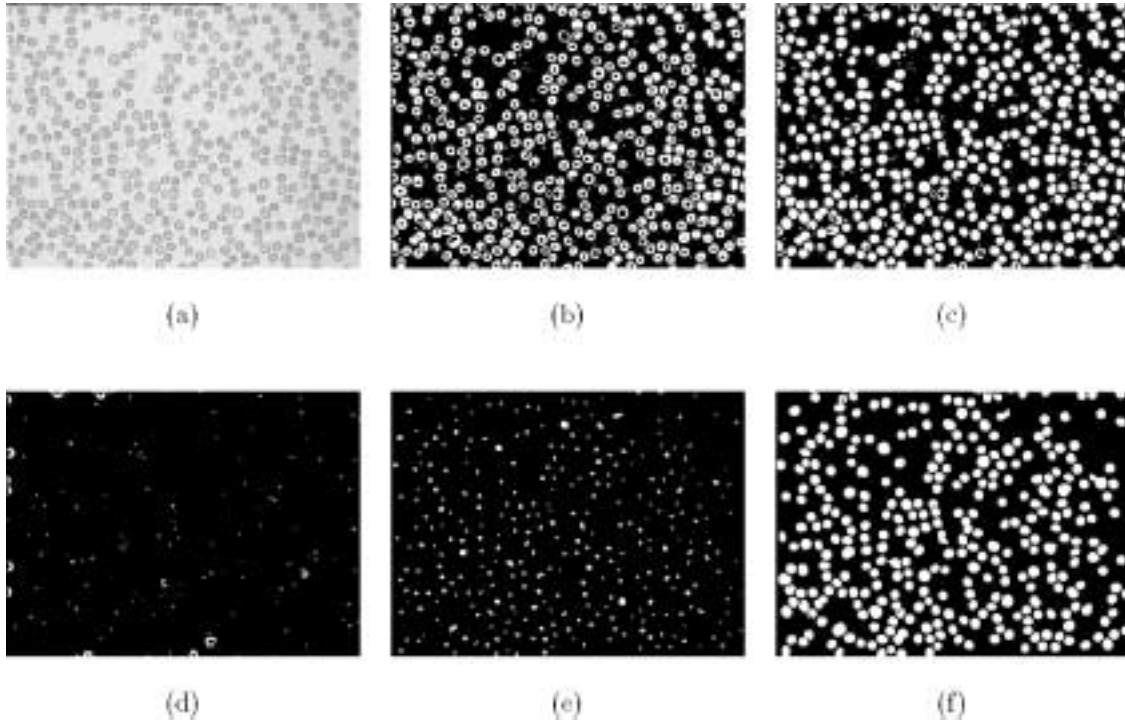
Fig. 6. Cell image analysis algorithm ($\times 25$): (a) Filtered green component of color image. (b) Binary image after thresholding. (c) Binary mask of cells. (d) Binary mask of platelets and artefacts. (e) Binary mask of centers. (f) Binary mask of cells with center.

$$I_{\text{pl}} = I_{\text{cl}} - I_{\text{cl2}},$$

that contains the platelets and some small artefacts, typically pieces of cell.

Extraction of centers The goal is to extract the centers of cells. To achieve this, we start by taking the difference between I and I_{cl}

$$I_{\text{ct}} = I - I_{\text{cl}}.$$

Then, we apply a size-and-close-grain operator in order to merge the grains which are part of the same center

$$I_{\text{ct2}} = \psi^{\text{clogra}}_{s_2 B, s_3 B}(I_{\text{ct}}),$$

where s_2 is the size of the opening by reconstruction which removes the small centers and s_3 the size of the closing which packs the grains sufficiently near. As result, we have the binary mask of centers.

Extraction of cells with center The final image is obtained by applying a reconstruction of the mask of cells I_{cl2} using the mask of centers I_{ct2} as the marker

$$I_{\text{clct}} = \gamma^{\text{rec}}(I_{\text{cl2}}, I_{\text{ct2}})$$

that provides a binary mask with the cells which have at least a clear center.

This algorithm has three parameters: the size of the platelets s_1, the size of the centers considered too small s_2 and the size of the gap between particles belonging to the same center s_3. Since they are parameters of size of biological structures, the use of different magnification powers of cells involves different sizes. For instance, under $\times 25$: $s_1 = 9$, $s_2 = 3$ and $s_3 = 5$ and under $\times 16$: $s_1 = 7$, $s_2 = 1$ and $s_3 = 3$. The robustness of an algorithm can be defined in respect to changes in the parameters or to image quality. We have tested the algorithm and the influence of parameters on an image database with 180 field images, introducing several alterations of quality (noise, resolution, etc.). In view of the results we can state that the behaviour of the algorithm with respect to limited changes is quite robust.

Figure 7 depicts the results of the algorithm for three fields belonging to a blood smear under $\times 25$ and another under $\times 16$, showing the differences of the possible three zones that we could find: too spread, acceptable and too thick.
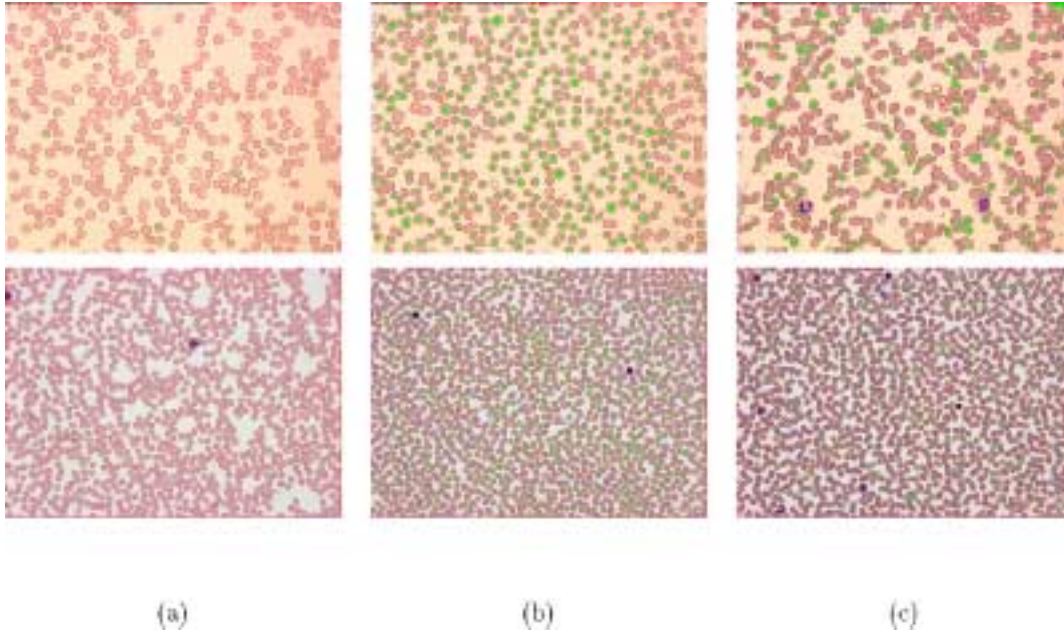
Fig. 7. Example of three fields of a blood smear (first row $\times 25$ and second row $\times 16$): (a) Spread area, (b) optimal area and (c) thick area. In green, clear centers and in red, boundaries of cells. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

2.5.2. Parameters for quantification

Taking these three binary images I_{cl2}, I_{ct2} and I_{clct} as a starting point, we can count for each one the number of connected components (separated particles) in such way that we have the number of cells N_{cells}, the number of centers N_{centers} and the number of cells with clear center $N_{\text{cells\_with\_center}}$. In order to calculate the connected components of a binary image, a fast algorithm based on the geodesic reconstruction is used. From these three measurements, the following parameters are defined.

The *coefficient of spreading* ρ_s quantifies the proportion of cells which have a clear center; i.e.,

$$\rho_s = \frac{N_{\text{centers}}}{N_{\text{cells}}}.$$

A low value of ρ_s involves that the cells are too spread on the slide or that they are too overlapped on top of one another and the central zones are also piled. Since N_{centers} can be greater than N_{cells}, the range of $\rho_s > 1$. This happens when there are some overlapped cells. However, in order to quantify more precisely this situation, we introduce a second parameter. The *coefficient of overlapping* ρ_o assesses the proportion of cells with clear central zone which are overlapped; i.e.,

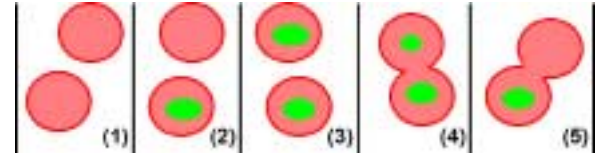$$\rho_o = \frac{N_{\text{cells\_with\_center}}}{N_{\text{centers}}}.$$

Fig. 8. Morphology of neighbouring erythrocytes. (1) Two separated cells without center, $\rho_s = 0$ and $\rho_o \sim 1$. (2) Two separated cells with a clear center, $\rho_s = 0.5$ and $\rho_o = 1$. (3) Two separated cells with two centers, $\rho_s = 1$ and $\rho_o = 1$. (4) Two connected (overlapped) cells with two centers, $\rho_s = 2$ and $\rho_o = 0.5$. (5) Two connected (overlapped) cells with a center, $\rho_s = 1$ and $\rho_o = 1$. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

In this case $\rho_o \leqslant 1$. A low value of ρ_o involves any amount of overlapped erythrocytes.

The synthetic diagram, in Fig. 8, contains the possible situations of two neighbouring erythrocytes and the respective values of ρ_s and ρ_o. The combined use of the two coefficients is very meaningful and yields a discrimination of the different zones of the smear.

3. Experiments and discussion

Colour figures can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

In order to define numerically the working area, we present a control experiment on *Smear 0* (see Fig. 2)

Table 1

Quantitative results of the study and subjective remarks by the haematological expert on *Smear 0*

Image	No. cells	No. centers	No. cells with center	ρ_s	ρ_o	Remarks
1	208	1	1	0.005	1.000	End of slide, too spread
2	322	12	11	0.037	0.917	Too spread, no clear center
3	346	16	16	0.046	1.000	"
4	373	63	63	0.169	1.000	"
5	386	211	211	0.547	1.000	Acceptable, clear center
6	382	306	303	0.801	0.990	"
7	374	301	297	0.805	0.987	"
8	396	360	351	0.910	0.975	"
9	408	387	370	0.948	0.956	"
10	416	410	380	0.985	0.927	"
11	402	393	354	0.977	0.900	"
12	382	404	345	1.057	0.854	"
13	376	383	328	1.018	0.856	Too thick, overlapped cells
14	360	363	289	1.008	0.796	"
15	352	360	271	1.023	0.753	"
16	332	336	245	1.012	0.729	"
17	271	342	205	1.262	0.599	"
18	304	316	212	1.040	0.671	"
19	223	268	155	1.202	0.578	"
20	116	115	56	0.991	0.487	Illegible, hard accumulation

for testing the ability of coefficients ρ_s and ρ_o against a human expert.

The images have been subjected to analysis by a haematological grader in order to remark the contents and to assign the corresponding zone of the smear. We consider three kinds of zone: too spread, acceptable (or working area) and too thick. Then the parameters have been computed for each image. The results have been compared with the subjective study. The values of parameters and human comments are summarised in Table 1.

From this table we were able to establish the values of the coefficients that allow the objective verification of the classification done by the human expert. For the coefficient of spreading, it is easy to observe that if we consider for instance $\rho_s > 0.50$ as criterion of acceptation, the fields too spread are rejected. The interpretation of this round value is also very intuitive: half the detected cells (50%) have at least a clear center. The coefficient of overlapping is less simple since its interpretation is also less intuitive. We took the limit value $\rho_o > 0.73$; i.e., at least 73% of cells with clear center are not overlapping. This choice implied that three images, doubtfully classified by human expert as belonging to the thick area, were acceptable for reading.
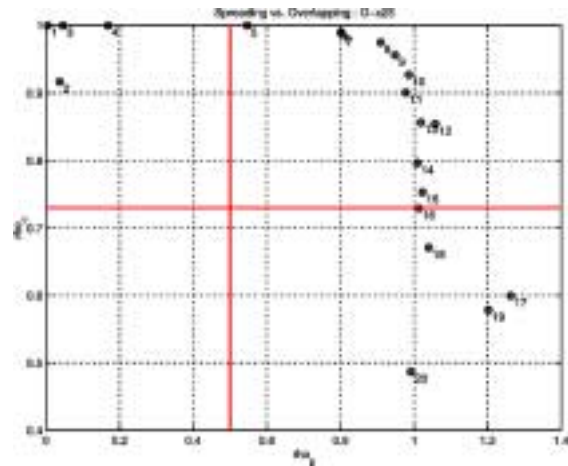

Fig. 9. Scatter plot of parameters ρ_s and ρ_o on *Smear 0*. The working zone belongs with the upper-right quadrant. This figure can be viewed on http://www.esacp.org/acp/2003/25-1/angulo.htm.

These undecided cases for a human operator are resolved according to objective criteria.

Figure 9 shows the scatter plot of parameters ρ_s and ρ_o. The established limits divide the space, demarcating the fields belonging to the working zone.

In summary, from this control case the working area is defined by the following boundaries

$\rho_s^{\text{work area}} > 0.50$ and $\rho_o^{\text{work area}} > 0.73$.

Moreover, we can state that

if $\rho_s \leqslant 0.50 \Rightarrow$ area too spread

and

if $\rho_o \leqslant 0.73 \Rightarrow$ area too thick.

We fixed these values as the classification stage of our approach.

Apart from the control example $\times 25$, we tested initially the algorithm on 3 smears $\times 16$, with 20 images from each one. By this scale changing we showed that the definition of working area by means of $\rho_s^{\text{work area}}$ and $\rho_o^{\text{work area}}$ is independent of magnification power and consequently independent of size of cell sample. Also independent of the field image size. This is natural because the definition of working area is given by two relative coefficients.

In order to compare the results obtained by the proposed algorithm with the performance of a human expertise, the image contents has been manually assessed into tree classes: too thick (TT), too spread (TS), acceptable (OK) and doubtful (DB). The results are shown in Table 2. We observe that the results are good enough. The manual results matched the automatic classification in the well-defined cases. A response was also obtained when the human has a doubt. The conflicts (image 16 of *Smear B* and image 16 of *Smear C*) happened just in the border between acceptable and spread area, where the decision to take is very difficult for the human grader.

Finally, in order to complete the results we carried out an additional analysis on other 10 smears $\times 16$, with 10 images from each one. The methodology used to validate the performance is the same, comparing the result of the proposed algorithm with the human expert evaluation. We propose to use the well-known notions of *true positive TP*, *true negative TN*, *false positive FP* and *false negative FN*. Besides the absolute number of classified images, we used the classical relative ratios: *sensitivity*, $Sens. = TP/(TP + FN)$ (measures the number of field images which truly belong to the working area and which correctly classified); *specificity*, $Spec. = TN/(TN+FP)$ (measures the number of field images which do not belong to the working zone which test negative) and the *positive predictive value* $PPV = TP/(TP + FP)$ (measures the proportion of fields which have been classified as working area are

Table 2

Comparison of the proposed automatic method with manual expertise on *Smear A*, *Smear B* and *Smear C*. Code for the manually classified zones: too thick (TT), too spread (TS), acceptable (OK) and doubtful (DB). The automated working zone belongs with the upper-right quadrant
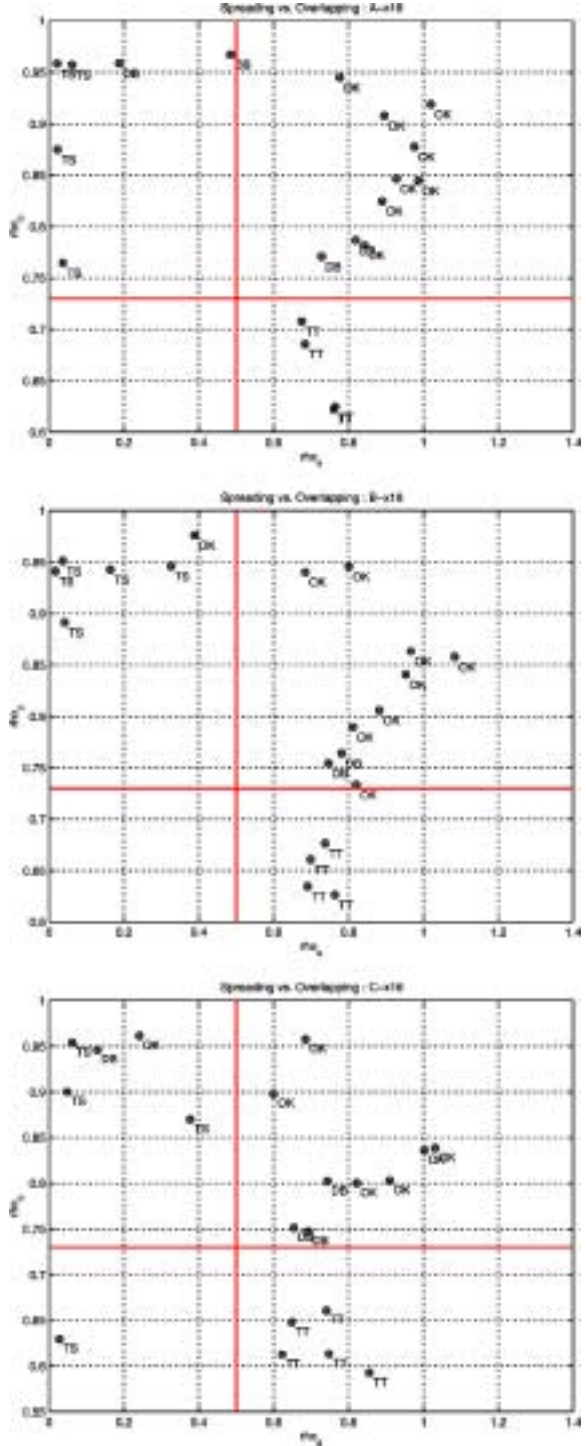
Table 3

Comparison of the results obtained by the proposed algorithm with the manually classified by a human expert

	TP	TN	FP	FN	Sens. (%)	Spec. (%)	PPV (%)
Smear A	10	10	0	0	100.0	100.0	100.0
Smear B	10	9	1	0	100.0	90.0	90.0
Smear C	9	10	1	0	100.0	90.0	90.0
Smears 1,2,…,10	43	48	7	2	95.5	87.2	86.0
Total	72	77	9	2	97.2	89.5	88.8

Table 4

Statistical values of the coefficients ρ_s and ρ_o on the 14 smears for the different smear zones: too thick (TT), doubtful in the border between TT and OK (DB 1), working area (OK), doubtful in the border between OK and TS (DB 2), and too spread (TS). Data are expressed as mean ± std. dev.

	ρ_s	ρ_o
Total	0.655 ± 0.357	0.825 ± 0.130
TT	0.567 ± 0.294	0.722 ± 0.150
DB 1	0.738 ± 0.054	0.768 ± 0.020
OK	0.845 ± 0.189	0.887 ± 0.072
DB 2	0.566 ± 0.430	0.905 ± 0.075
TS	0.409 ± 0.488	0.810 ± 0.154

really in the working area). In Table 3 are summarised the satisfactory values obtained for the 13 smears ×16.

In essence, the choice of decision boundaries for $\rho_s^{\text{work area}}$ and $\rho_o^{\text{work area}}$ is not only a question of automated classification. Despite that the fixed values of the coefficients are only based on a well-representative smear and a proved human experience these decision limits have yielded interesting results. We can not use the term "optimal" limits for the fixed values however we believe that for classical cell smear blood analysis the images have to be acquired from zones which coefficients of spreading and overlapping belong approximately to the presented values, see in Table 4 the statistical values of the coefficients ρ_s and ρ_o on the 14 smears for the different smear zones. Moreover, the physical sense of parameters allows to lift the boundaries depending on the requirements of the blood cell image analysis application.

Considering the success of the classification technique we could say that the information provided by these two coefficients and their computation speed can yield a simple tool to help an expert or a software system during the process of image acquisition.

4. Conclusions

Definition of working area of peripheral blood smear stained with May-Grünwuald Giemsa is a crucial step in order to automate the haematological cytology image analysis for microscopic diagnosis; detection of leukocytes using a motorised microscope; acquisition in digital images and quantitative analysis of morphological cell features based on mathematical models [1,2]. An efficient algorithm for the detection of the well spread part of the smear has been presented. Robustness and accuracy in comparison to human expertise have been evaluated on several smears under two magnification powers.

The results are encouraging and they open up new possibilities. Besides integrating the presented algorithm as a software module in a microscopic workstation, there are other indirect applications of the approach in practical haematology as a tool for quality control and standardisation of automatic devices and laboratory routines. Benattar and Flandrin [4] performed a comparison of an apparatus for automatically spreading peripheral blood films to the manual wedge-pull technique, showing that the average size of optimal area for counting is twice as large as compared to manual ones. Using our approach, it is possible to envisage comparing objectively the performance of different slide maker apparatus.

References

[1] J. Angulo, J. Serra and G. Flandrin, Quantitative descriptors of the lymphocytes, in: *7th Congress of the European Society for Analytical Cellular Pathology*, Caen, France, April 2001.

[2] J. Angulo, J. Serra and G. Flandrin, Haematological cytology image analysis and semantic indexing: Towards a global approach, in: *XIVth Congress of the International Society of Diagnostic Quantitative Pathology*, Oviedo, Spain, September 2001.

[3] L. Benattar and G. Flandrin, Morphometry and quality control for a May-Grünwald Giemsa stained preparation. A 40 centers cooperative study, *Leukemia and Lymphoma* **33** (1999), 587–589.

[4] L. Benattar and G. Flandrin, Comparison of the classical manual pushed wedge films, with an improved automated method for making blood smears, *Hematology and Cell Therapy* **41** (1999), 1–5.

[5] J.M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D.A.G. Galton, H.R. Gralnick and C. Sultan, Proposal for the classification of chronic (mature) B and T lymphoid leukaemias, *Journal of Clinical Pathology* **42** (1989), 567–584.

[6] M. Beksaç, M.S. Beksaç, V.B. Tipi, H.A. Duru, M.U. Karakas and A. Nur Çakar, An artificial intelligent diagnostic system on differential recognition of hematopoietic cells from microscope images, *Cytometry* **30** (1997), 145–150.

[7] CellaVision Company, www.cellavision.com.

[8] D. Comaniciu, P. Meer and D.J. Foran, Image-guided decision support system for pathology, *Machine Vision and Applications* **11** (1999), 213–224.

[9] J.V. Dacie and S.M. Lewis, *Practical Haematology,* Churchill Ltd., London, 1963.

[10] G. D'Arena, M.J. Keating and M. Carotenuto, Chronic lymphoproliferative disorders: An integrated point of view for the differential diagnosis, *Leukemia and Lymphoma* **36**(3–4) (2000), 225–237.

[11] G. Flandrin, Haematological cytology image bank and teletransmision for microscopic diagnosis, *Arch. Ana. Cytol. Path.* **43**(1995), 257–261.

[12] G. Flandrin, Image Bank, diagnostic codification and telediagnosis in hematology, *Leukemia and Lymphoma* **25** (1997), 97–109.

[13] K.S. Fu and J.K. Mui, A survey on image segmentation, *Pattern Recognition* **13** (1981), 3–16.

[14] C.A. Glasbey, An analysis of histogram-based thresholding algorithms, *CVGIP: Graph. Mod. Image Process.* **55** (1993), 532–537.

[15] M. Grimaud, New measure of contrast: the dynamics, *Image Algebra and Morphological Image Processing III,* **SPIE-1760** (1992), 291–305.

[16] H.M. Haus, H. Harms, V. Ter Meulen and U. Gunzer, Statistical evaluation of computer extracted blood cell features for screening populations to detect leukemias, in: *Pattern Recognition Theory and Applications*, P.A. Devjver and J. Kittler, eds, Spring-Verlag Berlin, Heidelberg, 1987.

[17] M. Ingram, P.E. Norgren and K. Preston, Automatic differentiation of white blood cells, in: *Image Processing in Biological Science*, Univ. of Calif. Press, 1968, pp. 97–117.

[18] J. Serra, *Image Analysis and Mathematical Morphology*, Vol. I, Academic Press, London, 1982.

[19] J. Serra, *Image Analysis and Mathematical Morphology, Vol. II: Theoretical Advances*, Academic Press, London, 1988.

[20] P. Soille, Morphological partitioning of multispectral images, *Journal of Electronic Imaging* **5** (1996), 252–265.

[21] N. Theera-Umpon and P.D. Gader, Counting white blood cells using morphological granulometries, *Journal of Electronic Imaging* **9** (2000), 170–177.

[22] I.T. Young, The classification of white blood cells, *IEEE Trans. Biomed. Eng.* **19** (1972), 291–298.