33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Fully automated detection of the counting area in blood smears for computer aided hematology

Stephan Rupp and Timo Schlarb and Erik Haßlmeyer and Thorsten Zerfaß
Image Processing and Medical Engineering Department
Fraunhofer-Institute for Integrated Circuits IIS
91058 Erlangen, Germany
timo.schlarb@iis.fraunhofer.de

***Abstract*— For medical diagnosis, blood is an indispensable indicator for a wide variety of diseases, i.e. hemic, parasitic and sexually transmitted diseases. A robust detection and exact segmentation of white blood cells (leukocytes) in stained blood smears of the peripheral blood provides the base for a fully automated, image based preparation of the so called differential blood cell count in the context of medical laboratory diagnostics. Especially for the localization of the blood cells and in particular for the segmentation of the cells it is necessary to detect the working area of the blood smear. In this contribution we present an approach for locating the so called counting area on stained blood smears that is the region where cells are predominantly separated and do not interfere with each other. For this multiple images of a blood smear are taken and analyzed in order to select the image corresponding to this area. The analysis involves the computation of an unimodal function from image content that serves as indicator for the corresponding image. This requires a prior segmentation of the cells that is carried out by a binarization in the HSV color space. Finally, the indicator function is derived from the number of cells and the cells' surface area. Its unimodality guarantees to find a maximum value that corresponds to the counting areas image index. By this, a fast lookup of the counting area is performed enabling a fully automated analysis of blood smears for medical diagnosis. For an evaluation the algorithm's performance on a number of blood smears was compared with the ground truth information that has been defined by an adept hematologist.**

I. INTRODUCTION

For medical diagnosis, blood is an indispensable indicator for a wide variety of diseases, i.e. hemic, parasitic and sexually transmitted diseases. For routine examinations the blood is typically analyzed automatically by flow cytometry or the use of Fluorescence Activated Cell Sorting (FACS). Unfortunately, these automatic methods do not deliver satisfying results in 30% to 40% of the analyses. Hence, in those cases, the subsequent visual inspection of the blood is essential. In addition, there are many use cases where such a visual inspection of a blood smear is required, especially if a pathological suspicion exists. A blood smear is a method for blood preparation for a microscopical examination. It is used for cytological and hematological assessments of blood cells (white blood cell count, WBC) and for proving the existence of bacteria or parasites within the blood such as nematoda or plasmodia.

In order to obtain a blood smear, a very small amount of blood is placed on one end of a microscope slide and dispersed along its full length. The goal is to get a region

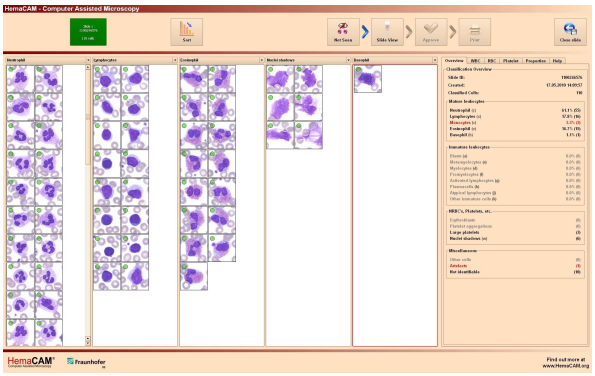


Fig. 1. The HemaCAM system aids the lab assistant analyzing blood smears for hematology-based diagnoses.

where the cells are spaced far enough apart to be counted and differentiated. This region is frequently referred to as *counting area* or *working area*. The preparation of a slide can be done automatically or manually. Unfortunately, slide maker devices are very expensive and result in high costs. Thus, even small or medium sized labs have to resort to a manual preparation. However, it requires specialized skills and experience by the lab assistant to achieve a good smear with a counting area that is large enough and suitable for the examination. Even though an appropriate slide preparation is not guaranteed and often has to be redone.

Once the blood smear is completely air-dried, the slide is immersed in methanol for fixating the smear. Finally, the slide is stained in order to distinguish the cells from each other.

At Fraunhofer IIS the HemaCAM system has been developed for an automated investigation of blood smears in order to support the physician in his WBC-based diagnosis. The system comprises of a Zeiss Axio Imager with a automatic height adjustable table (z-axis), a Zeiss Achroplan objective with a 10-fold magnification, a 1-fold tube adapter, an AVT Pike F-100C CCD camera and LED illumination. In addition, the system features a Märzhäuser stage in order to move the slide under the objective in two directions (x- and y-axis). The HemaCAM software comes out on top with the user interface, hardware control and image processing algorithms.

A robust detection and exact segmentation of white blood cells (leukocytes) in stained blood smears of the peripheral

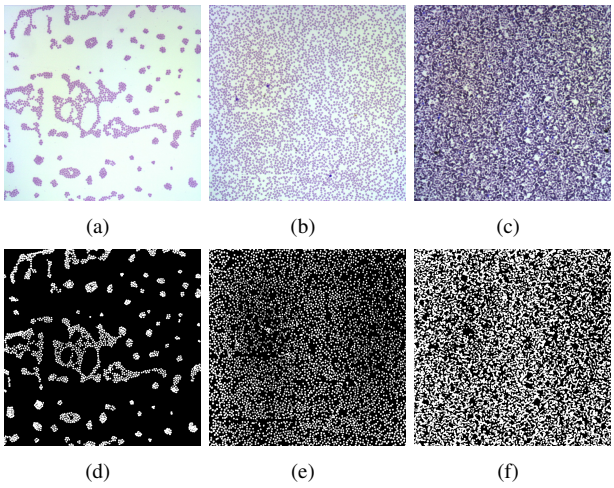


Fig. 2. Images of different regions within a blood smear (top line) along with their binarizations (bottom line) showing different concentrations of cells. The low cell densitiy (a,d) exhibits only small amount of cells and extended regions with background. The counting area (b,e) is characterized by predominantly separated and non interfering cells. In contrast, a large number of cells with a significant tendency to cluster represents a high cell density regions (c,f)

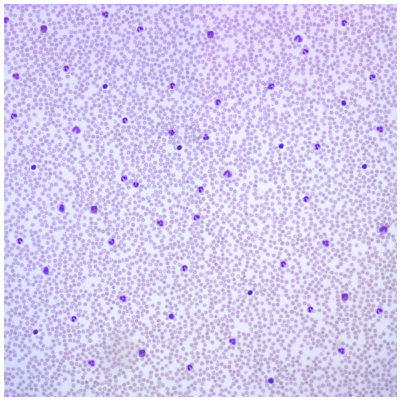


Fig. 3. Sample image section with a high number of leucocytes that requires a prior removal of leucocytes in order to prevent from biasing the computation of the indicator function.

blood provides the base for a fully automated, image based preparation of the WBC in the context of medical laboratory diagnostics (so-called Computer-Assisted-Microscopy, CAM). The variety of the different white blood cells, each with it's characteristic color and texture features, raises the difficulty of classification within the task of a full automation. A basic prerequisite for the detection of blood cells is the reliable localization of the counting area within the blood smear, because there the cells are predominantly separated and hence do not interfere with each other. See Fig. 2 for example images of a light, valid and dense region of blood cells of a blood smear.

The paper is structured as follows. In Sec. II, related work is discussed and the key idea of our contribution is shown. Sec. III details the algorithm. Finally, we present the experimental setups and results in Sec. IV and close the paper with a discussion and conclusions in Sec. V.

II. RELATED WORK AND CONTRIBUTION

Cell segmentation and classification have been studied intensively in the past years [1], [2]. Despite the fact that the location of the counting area plays a decisive role in image-based blood analysis systems, only a few papers could be found in literature that cover the robust detection of the counting area in the images of blood smears [3], [4].

This work focuses on the problem of robustly finding the counting area in a blood smear. For this, multiple images of such a smear are taken and analyzed in order to select the image corresponding to this area. The analysis involves the computation of an unimodal function directly from the image content. This requires a prior segmentation of the cells that is carried out by a binarization in the HSV color space. Finally, the indicator function is derived from the number of cells and the cells' surface area. Its unimodality guarantees to find a maximum value that corresponds to the counting area's image index, enabling a fast determination of the counting area on blood smears.

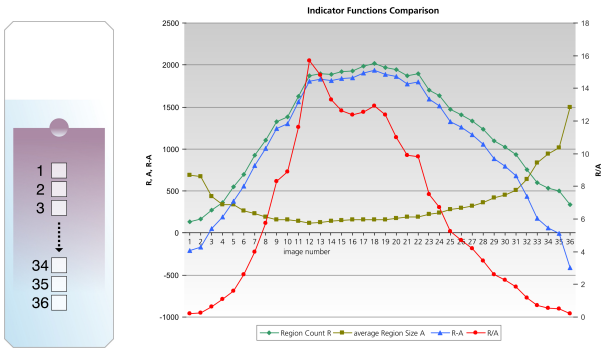
III. METHODS

In the following, we consider a set of M images. For the sake of simplicity, we assume that these images have been obtained by a systematic, full scan of a slide. However, in order to speed-up the scanning process, arbitrary heuristics can be used in order to constrain the search space as applied for the evaluation described in sec. Sec. IV.

For example, the scanning process in the HemaCAM system performs a search along the horizontal center line of a slide, starting at the thick end and heading for the smear's light end as shown in Fig. 4(a). From practical experience there is evidence that this heuristic does neither reduce the robustness nor the accuracy of finding the right area. Likewise, in everyday use HemaCAM reduces the amount of images by moving the x positions of the scan's start and end towards the center of the slide. By this, the localization process can be accelerated significantly. However, for evaluation purposes as addressed in Sec. IV, the full horizontal range is scanned yielding a constant number of 36 images for each slide.

Typically, the slide is hosted by a stage, that presents its contents (the blood smear) to the optics of the microscope. Hence, a full scan is obtained by subsequently moving the stage, acquiring an image with an objective with a magnification factor of 10 and storing the stage's position until the full slide has been captured. Due to the fact, that the stage defines a global coordinate system with x-, y- and z-axis, each image of the blood smear cooresponds to the position where it has been captured and vice versa. This property is essential for the reliable detection of the counting area in a blood smear. In general, our contribution is independent of a special staining as long as the staining causes a sufficent high contrast between the background and the cells. For demonstration purposes, we'll consider blood smears that have been stained according to Pappenheim using May-Grünwald-Giemsa (MGG) solutions.

In such a scan, let us denote the i^{th} image by $\mathbf{I}_i$ whereas its corresponding position in slide coordinates is given by



(a) (b)

Fig. 4. (a) The images are acquired along the y-axis starting at the thick end and heading to the light end of a smear. Under real conditions, the search space is further restricted by moving x positions of the scan's start and end towards the slide's center, i.e. by starting at index 3 and stopping at index 34. By this, the number of images is reduced resulting in a significant acceleration of the localization process. (b) Two different indicator functions (blue, red) along with the trend of the values. They are derived from the number of regions (green) and the average region size (khaki).

$\mathbf{p}_i = (x_i, y_i, z_i)$. The superscript of an image denotes the color space it refers to. Thus, the image $\mathbf{I}_i^{\text{HSV}}$ represents the image $\mathbf{I}_i$ in the HSV color space and $\mathbf{I}_i^{(S)}$ refers to its saturation channel, whereas $I_i^{(S)}(x, y)$ represents the value of an individual pixel at (x, y) in the referenced image plane. For the sake of clarity, we'll drop the color space superscript, if an RGB image is considered.

For each image $\mathbf{I}_i$ of the sequence, the following steps are performed:

- Leucocytes are segmented and labeled as background. For MGG stained slides, this step can be carried out efficiently by setting all pixels to the background level that suffice the condition:
$$\mathbf{I}_i^{(B)}(x, y) < \alpha\, \mathbf{I}_i^{(G)}(x, y),$$
with α being a weighting factor between $\frac{2}{3}$ and $\frac{1}{2}$.
- Transformation of $\mathbf{I}_i$ into cylindrical HSV representation: $\mathbf{I}_i \rightarrow \mathbf{I}_i^{\text{HSV}}$
- Separation of cells from the background. For this, a binarization operator $\mathcal{B}\{\cdot\}$ (i.e. Otsu [5]) is applied to the saturation channel $\mathbf{I}^{(S)}$ yielding a threshold value τ_i:
$$\tau_i = \mathcal{B}\{\mathbf{I}_i^{(S)}\}.$$
- Extraction of the number N_i of connected regions composed by an accumulation with a minimum of A_{min} pixels to prevent the detection of small regions caused by dirt or noise on the slide or in the optical path. A region R_j is defined by all the adjacent pixels $\mathbf{p} = (x, y)$ and $\tilde{\mathbf{p}} = (\tilde{x}, \tilde{y})$ that belong to the foreground:
$$N_i = \left| \bigcup \{R_j : |R_j| \geq A_{min}\} \right| ,$$
$$R_j = \Big\{ \mathbf{p}, \tilde{\mathbf{p}} : I_i^{(S)}(\mathbf{p}) \geq \tau_i \wedge I_i^{(S)}(\tilde{\mathbf{p}}) \geq \tau_i \wedge ||\mathbf{p} - \tilde{\mathbf{p}}|| \leq 1 \Big\}.$$
- Extraction of the areas A_j of the regions R_j that is the number of pixels a certain region R_j consists of:
$$A_j = |\{R_j : |R_j| \geq A_{min}\}| \ , j = 1, \ldots, N_i.$$
- Calculation of the average area $\bar{A}_i$ from the areas A_j:
$$\bar{A}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} A_j.$$
- Calculation of the indicator function $\vartheta(i)$ that yields a characteristic value i^* that corresponds to the position p_{i^*} that is for sure a part of the blood smear's counting area and thus a suitable starting point for the subsequent determination of the WBC.

The segmentation and removal of leucocytes is an essential step. For example the clinical picture of chronic lymphatic leukemia is characterized by a high number of leucocytes that dominate the image content, see Fig. 3. Under such conditions the binarization operator may yield a suboptimal threshold value which impairs the process of separating the cells from the background. Consequently, the leucocytes must be removed beforehand in order to guarantee that counting area is determined correctly under such special conditions, too.

Indicator Functions $\vartheta(i)$

The calculation of an indicator function is motivated by the fact that images containing the counting area of a blood smear are characterized by a high number of blood cells and a small average cells' surface area. In the following two different indicator functions based on $\bar{A}_i$ and N_i are considered for a robust determination of the index i^* that corresponds to a position p_{i^*} in the counting area.

Fig. 4(b) depicts four graphs of an example slide. The trends of the values $\bar{A}_i$ and N_i as well as two derived indicator functions $\vartheta(i) = N_i - \bar{A}_i$ and $\vartheta(i) = \frac{N_i}{\bar{A}_i}$ are shown. The indicator function's calculations are kept quite simple in order to allow for short computation times and thus fast determinations of the counting area's position $\mathbf{p}^*$ which is given by the global maximum i^* of the indicator values $\vartheta(i)$.

The first indicator function is given by the difference of the the number N_i of connected regions and the average region size, $\vartheta(i) = N_i - \bar{A}_i$. Despite its simplicity it yields good results. However, it does not consider the average cluster size adequately. Consequently, the resulting global maximum that identifies the counting area is slightly biased as it identifies the number 18 instead of the reference value of 12.

In contrast, the ratio of both values $\vartheta(i) = N_i/\bar{A}_i$ increases the impact of the cluster size. For this second indicator function, the maximum function value i^* is reached at image 12 coinciding perfectly with the reference value.

Consequentially, it turns out that the second function seems to have more expressive power and thus can be considered as the more reliable merit function where the function's global maximum of $i^* = \vartheta(i)$ yields the index of the image taken in the counting area.

image set	hematologist	algorithm
1	9-12	9
2	17-22	21
3	14-17	15
4	16-19	18
5	26, 27	27
6	13-15	13
7	17-19	19
8	17-19	19
9	16-20	20
10	17-20	19
11	16-20	20
12	15-19	18
13	8-10	10
14	16-20	19
15	24-27	25
16	14-17	14
17	12, 13	11
18	10-15	10
19	17-25	21
20	10-15	13
21	11-14	13
22	20	20

TABLE I

COMPARISON OF IMAGES SHOWING THE COUNTING AREA

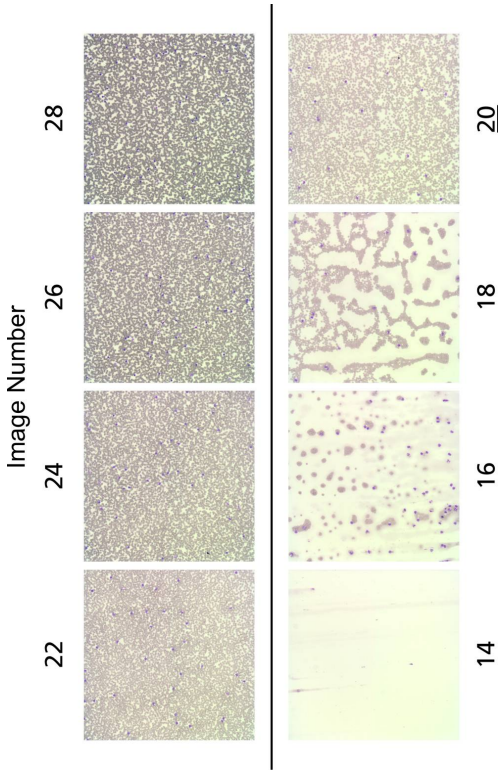


Fig. 5. Example images of slide 22. The counting area of this manually smeared slide was found on image 20.

IV. EXPERIMENTS AND RESULTS

To evaluate the accuracy of the presented approach for locating the counting area on blood smears multiple image sets have been taken and analyzed. The images are aquired using the above mentioned HemaCAM system. Aside from one slide all slides are prepared with a slide making device and stained with a stainer. Each of the image sets consists of 50 adjoint pictures taken along the center line of the longitudinal direction of the blood smear representing the major part of the linear expansion of one blood smear beginning from the end of the smear where the cell density is low. These image sets have been assessed by an adept hematologist and analyzed by the algorithm to find the images that show the counting region. The results are listed in table I for comparison.

For most image sets the expert found more than one image showing the counting area of the blood smear whereas the algorithm delivers only one image number. For each image set the algorithm did match with one of the image numbers the expert has choosen even in set 22 where a manually smeared and stained slide has been analyzed and only one image is showing the counting area.

V. CONCLUSIONS AND FUTURE WORKS

A robust detection and segmentation of blood cells in stained blood smears provides the base for an automated, image-based preparation of the differential blood cell count in the context of medical laboratory diagnostics. In this contribution, we addressed the problem of finding the counting area on a blood smear in order to support subsequent segmentation and analysing tasks. Our approach is based on the computation of an unimodal function that serves as an indicator for the counting area. The unimodal design guarantees to find a extremal value that corresponds to the counting area's image. The derived function is easy to compute and derives its values directly from the image content. In the experiments the algorithm's performance was evaluated on a number of blood smears and it was compared with the ground truth information that has been defined by experienced hematologists. The experimental results have shown that the indicator function reliably and robustly detects the counting area.

Currently, the selection of the maximum peak position is computed from the raw trend. In fact, analysing a filtered version of the trend will most likely increase the robustness of both the detection of the maximum and the overall performance of the algorithm. Furthermore the trend of the function can be used to specify the major search direction for the localization of the cells after finding the counting area.

REFERENCES

[1] Thorsten Zerfaß, Thomas Rehn, and Thomas Wittenberg. Boundary precise segmentation of nucleus and plasma of leukocytes. In Joseph M. Reinhardt and Josien P. W. Pluim, editors, *Medical Imaging 2008: Image Processing*, volume 6914, page 69143Q. SPIE, 2008.

[2] Tobias Bergen, Dirk Steckhan, Thomas Wittenberg, and Thorsten Zerfaß. Segmentation of leukocytes and erythrocytes in blood smear images. *Conf Proc IEEE Eng Med Biol Soc*, 2008:3075–3078, 2008.

[3] Jesús Angulo and Georges Flandrin. Automated detection of working area of peripheral blood smears using mathematical morphology. *Anal Cell Pathol*, 25(1):37–49, 2003.

[4] W. Xiong, J. H. Lim, S. H. Ong, N. N. Tung, J. Liu, D. Racoceanu, K. Tan, A. Chong, and K. Foong. Automatic working area classification in peripheral blood smears without cell central zone extraction. *Conf Proc IEEE Eng Med Biol Soc*, 2008:4074–4077, 2008.

[5] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:62–66, 1979.