Automated segmentation of blood cells in Giemsa stained digitized thin blood films

Margarita Walliander1*, Riku Turkki1,2, Nina Linder1, Mikael Lundin1, Juho Konsti1, Ville Ojansivu1,2, Taru Meri3, Ville Holmberg3, Johan Lundin1

From 11th European Congress on Telepathology and 5th International Congress on Virtual Microscopy Venice, Italy. 6-9 June 2012

Background
Assessment of erythrocytes and leucocytes in thin blood films can be used as an inexpensive diagnostic aid in a series of disease states, e.g. infections, anemia and hematological malignancies. Manual counting of cells is still considered the gold standard for example to establish the level of parasitemia in malaria. However, manual cell counting is time consuming and subject to variability [1]. We here propose an image analysis method that is a combination of adaptive histogram thresholds and morphologic characteristics for the segmentation of red blood cells (RBCs) and white blood cells (WBCs) in digitized thin blood films. The method is implemented on a virtual microscopy platform, the Webmicroscope [2].

Methods
Ten Giemsa stained thin blood films were digitized with a microscopy slide scanner (Axio Imager Z2, Carl Zeiss MicroImaging, Jena controlled by Metafer software, MetaSystems, Altlussheim) using a 63x objective with a numerical aperture (NA) of 1.4 (Plan-Apochmat, Carl Zeiss, Jena) and oil immersion. Image acquisition was performed with a monochrome CCD camera with a 1360×1024 pixel sensor and a pixel size of 6.45 µm (CoolCube 1, MetaSystems, Altlussheim), a 1.0 camera adapter and illumination with an RGB illuminator (MetaLED Z, red 619 nanometer, green 515 nanometer, blue 465 nanometer, MetaSystems, Altlussheim). The pixel size in the digital images was approximately 0.10µm and the original TIFF images were converted into a wavelet format (Enhanced Compression Wavelet, ERDAS/Intergraph, Norcross, GA) and transferred to a virtual microscopy image server (http://fimm.webmicroscope.net/Research/Momic/tp2012) [3]. Approximately five-hundred (473–505) fields of view from each blood film sample were captured and stored in the database. Five of the samples were infected with Plasmodium falciparum and five were non-infected control samples.

The described method (Fig. 1) involves 1) separation of background and foreground, 2) recognition of objects that compose the foreground and 3) cell counting (i.e. RBCs and WBCs).

Image preprocessing
As a preprocessing step for each thin blood film sample, the green channel was selected from the original RGB image [4] and smoothed by applying a median filter 3X3 to reduce the 'salt and pepper' noise [5]. The green channel is extracted using a color deconvolution between the original image and a vector [0,1,0].

Adaptive histogram thresholds
Let each pixel of the preprocessed image have intensity levels in \([0, 1, 2, \ldots, L-1]\) with \(L = 256\). The number of pixels with intensity level \(i\) is denoted by \(n_i\), \(\forall i = 0, 1, 2, \ldots, L-1\), where \(N = \sum_{i=0}^{L-1} n_i\) is the total amount of pixels.

We defined the histogram distribution as \(p(i) = n_i/N\), \(p(i) \geq 0\), where \(\sum_{i=0}^{L-1} p(i) = 1\).

For any monolayer stained blood film, the histogram is bimodal. A typical histogram shape for a monolayer thin blood film is shown in Fig. 2.

There are two local maxima located at \(m_1\) and \(m_2\), where \(m_2 < m_1\) and \(P = p(m_2) < p(m_1) = Q\).

A threshold to differentiate the background from the foreground \(B\), is defined by finding the maximum distance...
Figure 1 Flowchart showing the cell segmentation process. Overview of the cell segmentation and counting process for red and white blood cells. From the original image the green channel is selected and smoothed. Dynamic thresholding allows the separation of the image in foreground and background. The foreground is split into Round Cells, WBC Candidates, Found Cells and the remaining part being deformed red blood cells and clumps of red blood cells. These subimages are used to count the objects of interest i.e. red blood cells and white blood cells.
between the histogram and the line \( l_1 \) described between \((m_1, P)\) and \((m_2, Q)\) as shown in Fig. 2.

\[
d(p(B), l_1) = \max \{ d(p(x), l_1) \mid x \in [m_2, \ldots, m_1] \}
\]

with

\[
l_1 = \left\{ (x, y) \mid y = ax + b, \forall x \in [m_2, \ldots, m_1], a = \frac{P - Q}{m_2 - m_1}, b = Q - m_2 \cdot \frac{P - Q}{m_2 - m_1} \right\}
\]

In a similar manner, a threshold to find heavily stained objects \( H \), was defined by finding the maximum distance between the histogram distribution and the line \( l_2 \) described between \((0, R)\) and \((m_1, P)\), with \( R = p(0) \);

\[
d(p(H), l_2) = \max \{ d(p(x), l_2) \mid x \in [0, \ldots, m_2] \}
\]

and

\[
l_2 = \left\{ (x, y) \mid y = ax + b, \forall x \in [0, \ldots, m_2], a = \frac{Q}{m_2}, b = 0 \right\}
\]

The image background is given by Eq. (1), while the heavily stained objects are described by Eq. (2).

\[
\text{Background} = \{ f(x, y) \geq B \mid f(x, y) \in D \}, \quad (1)
\]

\[
\text{Heavily Stained Objects} = \{ f(x, y) < H \mid f(x, y) \in D \}, \quad (2)
\]

where \( = \) gray level Image.

**RoundCells separation**

Using the image histogram and based on its bimodal shape, two important thresholds were extracted (Fig. 2). The first threshold (B) defines a binary separation between the image background and foreground. From the image foreground, all objects with roundness bigger than 0.6 were selected and the area of each of the objects was measured. The mean diameter to be 7.52µm and standard deviation of 0.06µm for the whole set of objects was calculated. Finally, only the subset of objects with an area equal to \( m \pm \sigma \) (3848±/−688 pixels) was chosen and defined as RoundCells. From this set of round objects of similar size, the average diameter was calculated and used to define a representative red blood cell, designated as \( \text{AvgRBC} \) (diameter ~7µm) and to establish limit diameters for WBCs (~7-21 µm) and platelets (~2-3µm).

The second threshold (H), defines the heavily stained objects in the foreground (i.e. WBC, platelets, artifacts and debris). The heavily stained objects larger than \( \text{AvgRBC} \) are the FoundWBCs.

**Detection of circular shapes by Hough transform**

Hough transform is calculated on gray level images that contain only the regions of interest while the remaining is set to zero. The region of interest is composed by foreground without RoundCells, FoundWBCs and debris. The maximization of Hough transform for a radius
interval \( r \left( \frac{3r}{2} \right) \) is performed, where \( r = \text{radius} \) (AvgRBC). The result is a set of accumulations of hits (votes). The accumulations are concentrated around the centers of the circular shapes. Hough transform detected cells are filtered by selecting the pixel with the maximum vote and deleting all the pixels with less than 20% of the votes. Thus the selection of nearly circular shapes is ensured. Finally, a morphological opening is performed to discard accumulations with less than 50 pixels. The remaining objects are centers of FoundCells.

**FoundCells detection**

After subtracting the RoundCells and the heavily stained objects from the original image, to compensate for the holes left from the subtraction of the platelets, debris and parasites, a morphologic filling was performed. By using Hough transform, circular shapes were detected in the grayscale image and designated as FoundCells, resolving the center positions of the nearly circular objects. A second representative red blood cell AvgRBC2 was defined from the area of FoundCells.

**ApproxCells detection**

After subtracting the RoundCells, the FoundWBCs and the FoundCells, the remaining image contains fragments of RBCs and deformed RBCs which Hough transform was not able to define as circular shapes. The total area covered by these objects, named ApproxCells was divided by the area of AvgRBC2 which is estimation for the number of cells that still remain without being counted #ApproxCells.

Finally, the total number of RBCs in the image is calculated by summing up the partial results;

\[
\#\text{TotalRBC} = \#(\text{RoundCells}) + \#(\text{FoundCells}) + \#(\text{ApproxCells})
\] (3)

**Results and discussion**

RBCs were manually annotated in 30 fields of views per thin blood film and WBCs were annotated in the entire data set (Table 1). The results from the manual counting and automated counting are shown (Table 1). Using the annotated fields of view, automated quantification of RBCs and WBCs was compared against the manual annotations and RBCs showed an overall error rate of 0.06%, WBCs counting showed an overall error rate of 0.21%. A test for the automated counting of RBCs and WBCs was performed on whole slides of thin blood films and approximately half a million red blood cells and 477 white blood cells were counted (Table 2.). Previously published studies have addressed the separation and counting blood cells, but fixed thresholds for colors, sizes and intensity values restrict the use to particular data sets [6-10]. Here, we make use of adaptive thresholds for size and intensity values, which converges to a solution.

**Conclusions**

The segmentation of RBCs and WBCs is an easy task for a human observer. Humans have the ability of distinguishing large number of colors, shades and hues, also estimating shapes and size similarities while referring to prior knowledge, making global and local comparisons simultaneously. However, performing large

### Table 1 Results comparing manual and automated cell counting

| Sample | Annotated Cells | Automated Cells | Error % | Annotated Cells | Automated Cells | Error % |
|--------|----------------|-----------------|---------|----------------|-----------------|---------|
| I1     | 3145           | 3160            | 0.476948| 20             | 20              | 0       |
| I2     | 4048           | 4058            | 0.247036| 34             | 34              | 0       |
| I3     | 2796           | 2782            | 0.500715| 22             | 22              | 0       |
| I4     | 2972           | 2958            | 0.471063| 30             | 32              | 6.66    |
| I5     | 3047           | 3042            | 0.164096| 77             | 75              | 2.59    |
| C1     | 3396           | 3389            | 0.206125| 75             | 75              | 0       |
| C2     | 3491           | 3482            | 0.257806| 55             | 56              | 1.81    |
| C3     | 3093           | 3087            | 0.193986| 42             | 42              | 0       |
| C4     | 3197           | 3206            | 0.281514| 49             | 50              | 2.04    |
| C5     | 3513           | 3514            | 0.028466| 71             | 72              | 1.38    |
| TOTAL  | 32698          | 32678           | 0.061166| 476            | 477             | 0.21    |
| %      | 100            | 99.9388         | 100     | 99.79          |                 |         |

Results comparing manual and automated counting for red blood cells and white blood cells (WBCs). Red blood cells were annotated in a region equivalent to 30 fields of view per film, while the annotations for WBCs were performed on 500 fields of view per film. The samples I1-I5 are Plasmodium falciparum infected cases, C1-C5 are non-infected controls.
scale quantification is a time consuming and tedious task.

We present an unsupervised tool for separating the foreground from the background in Giemsa stained thin blood films and an automated cell counter for RBCs and WBCs. The segmentation of blood cells in thin blood films can be used as a pre-processing step to specify the regions of interest for a secondary algorithm, e.g. the detection of malaria parasites in RBCs, morphological analysis of RBCs and WBCs and follow-up during treatment of hematological malignancies or measurement of response to chemotherapy.

List of abbreviations
RBC: Red blood cell; WBC: White blood cell; #: Number of cells in

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
The authors wish to thank Elisabet Tyyni for sample preparation and analysis. The study was kindly supported by the national Biomedinfra and Biocenter Finland projects.

Authors’ details
1Institute for Molecular Medicine Finland (FiMM), Finland. 2Center for Machine Vision Research, Department of Computer Science and Engineering, University of Oulu, Finland. 3Haartman Institute, Finland.

Published: 30 September 2013

References
1. O’Meara WP, Baccus M, Wongsrichanalai C, Muth S, Maguire JD, Jordan RG, Prescott WR, McIntosh FE: Reader technique as a source of variability in determining malaria parasite density by microscopy. Malaria Journal 2006, 5:118.
2. Linder E, Lundin M, Thors C, Lebdad M, Winecka-Krusnell J, Helin H, Leiva B, Isola J, Lundin J: Web-based virtual microscopy for parasitology: a novel tool for education and quality assurance. PLoS Negl Trop Dis 2008, 2(10):e315, Epub 2008 Oct 22.
3. The WebMicroscope virtual microscopy environment. [http://www.webmicroscope.net/].
4. Wermser D, Haussman G, E LC: Segmentation of blood smears by hierarchical thresholding. Computer Vision, Graphics and Image Processing 1984, 25:151-168.
5. Gonzalez RC, Woods RE: Digital Image Processing. New Jersey: Prentice Hall, 2002.
6. Kumar R, Joseph DK, Sreenivas TV: Teager Energy based blood Cell Segmentation. 14th International Conference on Digital Signal Processing 2002 2:619-622.
7. Tek FB, Dempster Ag, Kale I: Computer vision for microscopy diagnosis of malaria. Malaria Journal 2009, 8(53).
8. Donini LB, Miretti R, Lente NJ: White Blood cell segmentation using morphological operators and scale-space analysis. SBGRAP, 2007. XX Brazilian Symposium on Computer Graphics and Image Processing 294-304.
9. Ramesh N, Dangott B, Salama ME, Tasdizen T: Isolation and two-step classification of normal white blood cells in peripheral blood smears. Journal of Pathology Informatics 2012, 149-170.
10. Purwar Y, Shah S, Clarke G, Almugairi A, Muehlenbachs A: Automated and unsupervised detection of malaria parasites in microscopic images. Malaria Journal 2011, 10:364.

doi:10.1186/1746-1596-8-S1-S37
Cite this article as: Walliander et al.: Automated segmentation of blood cells in Giemsa stained digitized thin blood films. Diagnostic Pathology 2013 8(Suppl 1):S37.

Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit