Microscopic image analysis using mathematical morphology: Application to haematological cytology

Jesús Angulo[1] and Georges Flandrin[2]

[1] Centre de Morphologie Mathématique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau, France

[2] Laboratoire Central d'Hématologie, Hôpital Universitaire Necker, 149 rue de Sèvres, 75743 Paris, France

Keywords: microscopic image analysis, mathematical morphology, statistical pattern recognition, blood smear, automation in haematological cytology, lymphoproliferative disorders, motorised microscope, telepathology, telediagnosis

Abstract. The mathematical morphology is a non-linear image processing approach which is based on the application of lattice theory to spatial structures. This technique is proven to be a very powerful tool in microscopic image analysis. In this chapter, we present the architecture of a system of quantitative microscopy which yields an integrated platform for the morphological image analysis, pattern recognition, and visual content-based indexing of peripheral blood smear wide microscopic fields. This image analysis approach, based on tools of mathematical morphology, provides an objective description of blood cells and more specifically the system is designed to assist pathologists to diagnose lymphoproliferative disorders within a telepathology context.

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, to calculate the number and morphology of platelets, and in microbiology for detection of erythrocyte parasites (malaria) [1]. 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 to waste. 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, particularly to study the morphology of lymphocytes which is regarded as the principle basis for the identification and discrimination among the chronic lymphoproliferative disorders [2].

In a strategy of automation using a motorised microscope within a telepathology context and following some research works on haematological cytology image analysis for microscopic diagnosis [3][4], the aim of this chapter is to present a powerful application of mathematical morphology in order to propose an integrated image analysis platform for the segmentation, feature extraction and classification of lymphocytes from peripheral blood smear cells in microscopic large fields images. In a second phase, the extension of our platform to study the erythrocyte morphology and erythrocyte inclusions is currently developed.

The rest of the chapter is organised as follows. First, the technological advances which have motivated this work are presented in next section. Then a reminder of mathematical morphology, illustrated with several examples, is included. We continue with the presentation of our automated morphometric analysis of peripheral blood smears cells approach. Finally, conclusions are given.

Return of microscope in haematology

Morphological analysis is still essential for diagnosis in haematological disorders. From a methodological point of view, the desirable qualities of a blood smear are two. These are 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 [5][6]. Commercially available automated devices for performing blood film and staining lead to reproducible preparations.

The motorised microscopes allow to obtain the automated production of digital microscopic image fields. The microscopic images are acquired for storing, transmitting and processing. The classical solution is the concept of *image folder* (including few representative images) [7] which has the lack of slide representativeness, limiting some applications. The *virtual slide* or *teleslide* is currently proposed as an alternative solution: a teleslide is a high resolution image at diagnostic magnification of each slide, composition of hundredth of fields [8].

All these technological advances together with other steps forward in networking, visualisation and computer technology allow today to believe in the return of microscope in haematology diagnosis, complementing of course the immunophenotyping, citogenetic and molecular characterisations.

Mathematical morphology

First introduced as a shape-based tool for binary images, mathematical morphology has become a very powerful non-linear image analysis technique with operators capable of handling sophisticated image processing tasks in binary, grey-scale, color and multiresolution imaging modalities. Mathematical morphology is the application of lattice theory to spatial structures. A tutorial in the technique can be found in [9][10]. This technique is proven to be a very powerful tool in microscopic image analysis [11]. In this section we briefly define the basic morphological operators used in the chapter, Fig. 1, 2 and 3.

In the framework of digital grids, a *grey-tone image* can be represented by a function $f : D_f \rightarrow T$, where D_f is a subset of $\mathbb{Z}^2$ and $T=\{t_{min}, \ldots, t_{max}\}$ is an ordered set of grey-levels. Let B be a subset of $\mathbb{Z}^2$ and $s \in N^2$ 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: $\varepsilon_B(f(x)) = \inf_{-y \in B} \{ f(x-y) \}$

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 $\gamma^{rec}(g,f) = \delta_g^i(f)$, such that $\delta_g^i(f) = \delta_g^{i+1}(f)$. There are three composed operators derived from the geodesic reconstruction that are of great importance for the application in quantitative cytology: the *close-holes* operator (fills all holes in an image f that do not touch the image boundary f_∂, $\psi^{clohole}(f) = [\gamma^{rec}(f^c, f_\partial)]^c$

and therefore provides a parameter free approach to detect holes in an image); the *size-and-close-grain* operator (removes the small light structures or grains - smaller than λB_1 - using an opening by reconstruction

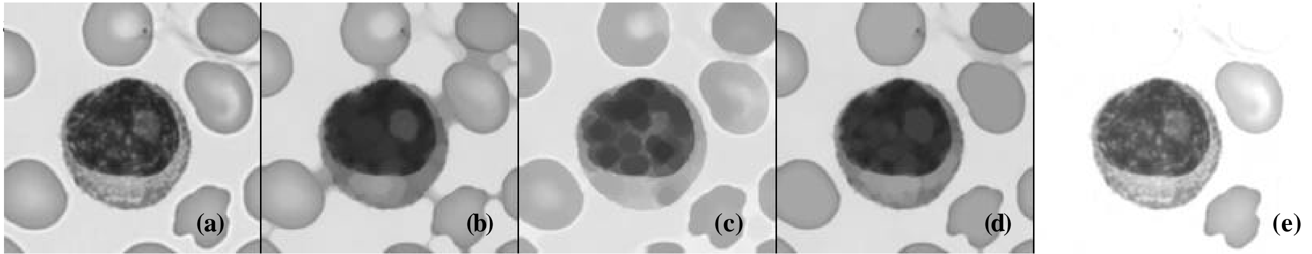

Figure 1. Lymphocyte and some erythrocytes: (a) initial image, (b) *opening* of size 15, (c) *closing* of size 15, (d) *close-holes* operator (on negative of initial image) and (e) *extract-edge-particles* operator (on negative of initial image).

smaller than μB_2 where B_2 is a circle- using an isotropic closing, $\psi^{clogra}_{\lambda B1, \mu B2}(f) = \varphi_{\mu B2}[\gamma^{rec}(f, \gamma_{B1}(f))]$); and the *extract-edge-particles* operator (removes the particles in f touching the image border $f_{¶}$, $\psi^{extractedge}(f) = f - [\gamma^{rec}(f, f_{¶})]$).

Families of openings or closings of increasing size are at the basis of the granulometric analysis. The granulometries or size distributions allow a good knowledge of the "objects"' or "structures" present in the images. Formally, a *granulometry* can be defined as a decreasing family of openings having a size parameter λ, $\Gamma=(\gamma_\lambda)_{\lambda\geq 0}$, and satisfying the absorption law: $\lambda\geq 0$, $\mu\geq 0$, $\gamma_\lambda\gamma_\mu=\gamma_\mu\gamma_\lambda=\gamma_{max(\lambda,\mu)}$. Moreover, granulometries by closings (or anti-granulometry) can also be defined as families of closings. Performing the granulometric analysis of an image f with Γ is equivalent to mapping each opening of size λ with a measure $m(\gamma_\lambda(f))$ of the opened image $\gamma_\lambda(f)$. This measure is typically the volume in the greyscale case. The granulometric curve, or *pattern spectrum*, of f with respect to γ, denoted $PS_\gamma(f,\lambda)$ is defined as the following normalised mapping: $PS_\gamma(f,\lambda) = (m(\gamma_\lambda(f)) - m(\gamma_{\lambda+1}(f))) / (m(f))$. The pattern spectrum $PS_\gamma(f,\lambda)$ maps each size λ to some measure of the bright image structures with this size. By duality, the concept of pattern spectrum extends to anti-granulometry by closings φ_λ, $PS_\varphi(f,\lambda) = (m(\varphi_\lambda(f)) - m(\varphi_{\lambda-1}(f))) / (m(f))$ and is used to characterise the size distribution of dark image structures. The normalised pattern spectrum is a probability density function, i.e. a size histogram of f. A large impulse in the pattern spectrum at a given scale indicates the presence of many image structures at that scale.

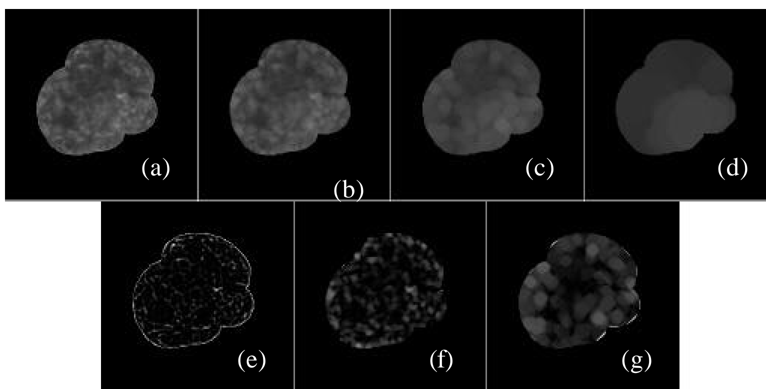

Figure 2. Lymphocyte chromatin texture analysis using an isotropic grey tone granulometry: (a) initial image f, (b-d) *openings* of increasing size $\gamma_\lambda(f)$: $\lambda=5$, $\lambda=10$ and $\lambda=30$, (e-g) residues $\gamma_\lambda(f) - \gamma_{\lambda+\mu}(f)$.

A plateau in a grey tone image is a connected component of uniform altitude. A *maximum* is a plateau without higher neighbors. Among the various features that can be extracted from an image, the maxima and the minima are the primary importance in the morphological segmentation and in the modern morphological filtering techniques, because they ``mark" the structures present in the image.

In pattern recognition the purpose of segmentation is to isolate the various objects from the background. In mathematical morphology, the *watershed transformation* is one of the most powerful tools for segmenting images [12]. The watershed line associates a catchment basin to each minimum of the function. The problem is that noisy and textured images have many minima, most of them being irrelevant for segmentation. Using the watershed on a gray tone image without any preparation leads to a strong over-segmentation. The best solution to this problem consists in initially determining *markers* for each region of interest, including the background of the image. It is then possible to construct a watershed line associated with these markers. Using the image of flooding, a hole is pierced in the relief at each marker, and floods the relief starting with these markers only. The dams which have to be erected are the limits of the catchment bassins associated with the markers (they are closed lines). Another important choice is the function used in the flooding process. Usually, the selected function is some form of image gradient. In image cytology, an important problem is the separation of overlapping particles and it is a well-resolved task in mathematical morphology. The solution involves using the *distance function* of a binary set: the maxima of the distance function mark the different particles.

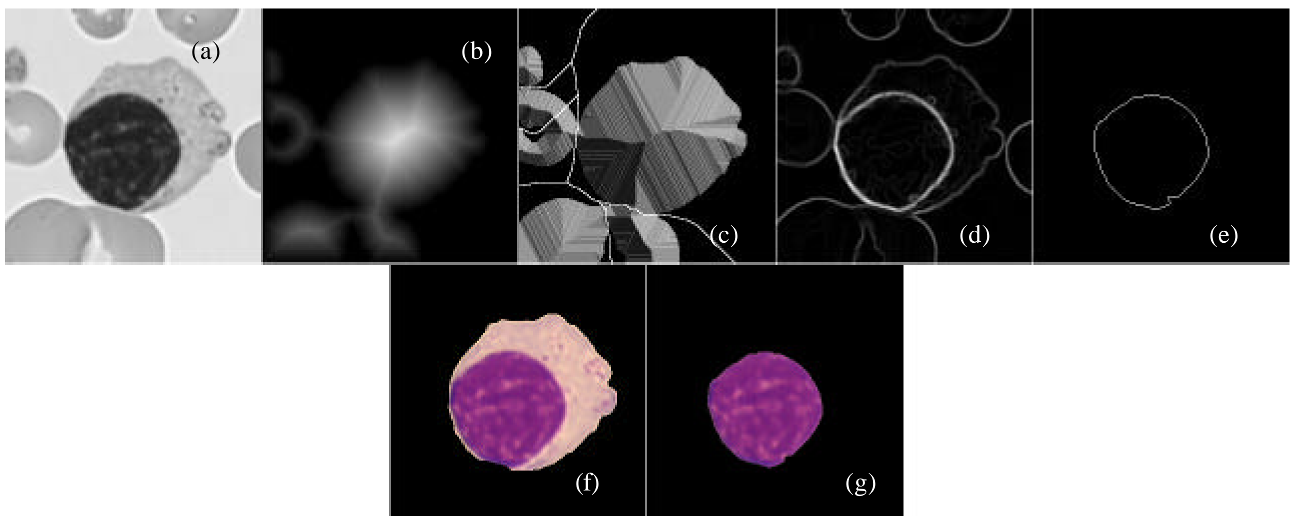

Figure 3. Lymphocyte segmentation: (a) initial grey-tone image, (b) distance function from a binary image after *close-holes* operator (cytoplasm with overlapped erythrocytes), (c) *watershed* line after a slight *filtering* to remove unimportant maxima: isolated cytoplasm, (d) morphological *gradient* of a simplified image, (e) watershed line of the gradient associated with markers (as inner marker, *ultimate erosion* of a thresholded nucleus and as outer marker, negative of previous detected cytoplasm): isolated nucleus, (f) and (g) segmented lymphocyte and segmented nucleus color images.

Automated morphometric analysis of peripheral blood smears cells

In this section, we present the architecture of a software system which yields an integrated platform for the morphological image analysis, pattern recognition, and visual content-based indexing of peripheral blood smear *teleslides*. This image analysis approach, based on tools of mathematical morphology, provides an objective description of blood cells and more specifically the system is designed to assist pathologists to diagnose lymphoproliferative disorders.

Approach overview. In Fig. 4 is depicted an overview diagram of the approach. A first pre-processing module implements the automated detection of working area of the blood smear in images scanned under low magnifying power (x25). When the best area is detected, the magnifying power is increased (x100) and the *teleslide* images are produced using a routine motorised microscope and a dedicated software. On a *teleslide*, the leukocyte are detected and the lymphocyte identified and extracted. Each lymphocyte is segmented into two regions: nucleus and cytoplasm. From the nucleus and cytoplasm color images and using mathematical morphology, a set of quantitative parameters are calculated. These measurements yield

a first classification of each morphological feature into a category. The set of morphological feature values constitutes the lymphocyte descriptor. The next step involves the use of the morphological descriptor for classifying the lymphocyte into a cellular typology, followed by a classification of the lymphocyte population. The morphological description provides an objective information which may be indexed into a database together with the *teleslide* and/or used for (tele)diagnosis procedures and research protocols.

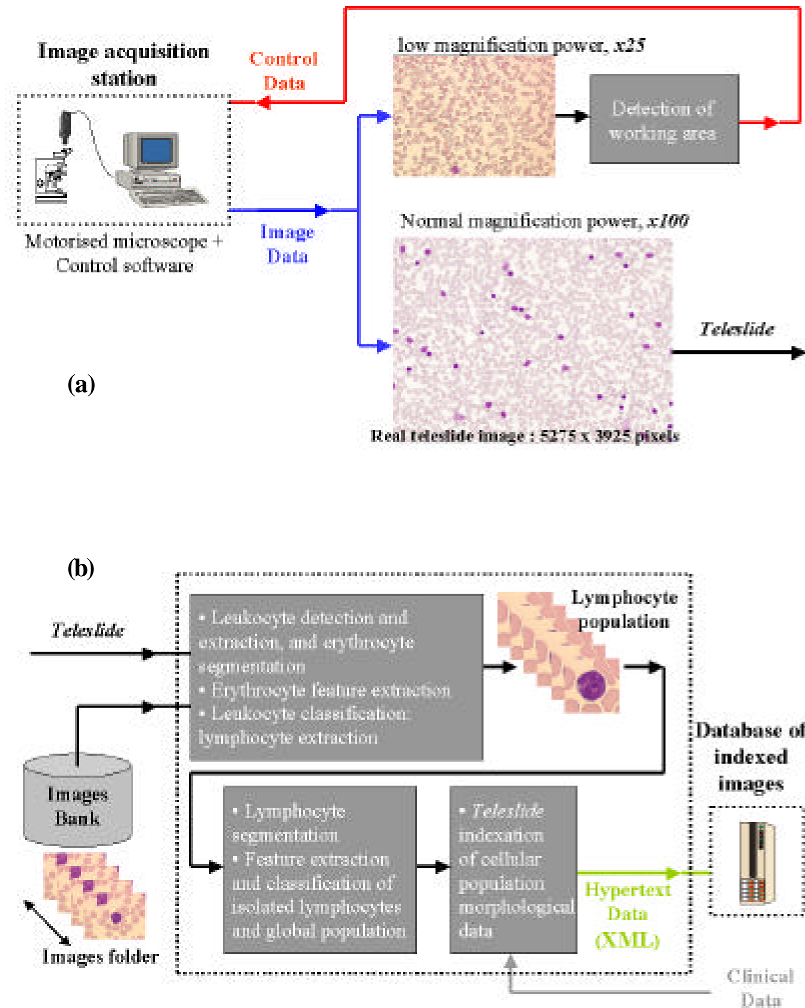

Figure 4. Automated morphometric analysis of peripheral blood smears cells: (a) production of *teleslides*, (b) integrated image analysis, pattern recognition and content-based indexing system.

Detection of working area. This module implements a technique to detect automatically the working area of peripheral blood smears stained with May-Grünwuald Geimsa. 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 *x25* or *x16*). 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, Fig. 5. Second, the number of connected components from the three kinds of particles is counted and the coefficient of spreading $\boldsymbol{r}_s = N_{centers} / N_{cells}$ and the coefficient of overlapping $\boldsymbol{r}_o = N_{cells_with_center} / N_{centers}$ are calculated. In summary, the working area is defined by the following boundaries: $\boldsymbol{r}_s^{work_area} > 0.5$ and $\boldsymbol{r}_o^{work_area} > 0.73$. Further details in [13].

(a)

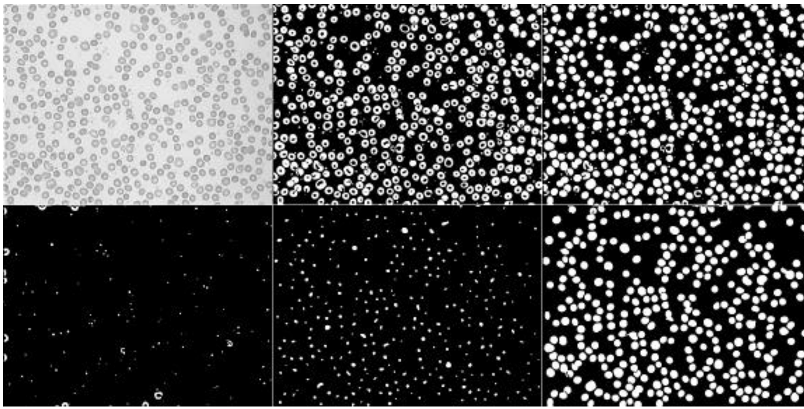

Figure 5. Erythrocyte image analysis algorithm (x25): (a) Filtered green component of color image, (b) binary image after thresholding, (c) binary mask of cells, (d) binary mask of platelets and artefacts, (e) binary mask of centers, (f) binary mask of cells with center.

Blood image segmentation. Motivated by the need of an optimal and robust segmentation technique for developing image-based haematological cytology applications, we have developed a hierarchical strategy for segmenting the visual entities of *teleslide* images based on some tools from mathematical morphology. The rationale for the chosen hierarchy is the semantic contents of the images and the subsequent application of the segmentation results. The segmentation scheme consists of three independent modules. The aims of the first module are the erythrocyte segmentation and leukocyte extraction by automatic thresholding and binary filtering. In the second level, the lymphocyte segmentation is achieved using the classical watershed segmentation (the result are the nuclear and cytoplasmic segmented images). And finally, the third level is about the nucleus chromatin segmentation by region growing in partition lattices [14].

Leukocyte classification. Leukocytes may be subdivided into five different categories: (1) monocytes, (2) neutrophils, (3) basophils, (4) eosinophils and (5) lymphocytes. When a peripheral blood smear, stained with May-Grünwuald Giemsa, is examined under a microscope at *x100* magnification, the five classes of leukocytes may be differentiated according to their morphological and spectral features. For this purpose, we introduced an image-based approach for automated leukocyte classifying using statistical techniques (template matching by histogram distances). Before a leukocyte can be identified, it must first be automatically located in a sub-image (leukocyte images obtained from previous segmentation module), in such a way that there is only one leukocyte in each image to study. The five leukocytes classes to detect may be differentiated according to the shape, texture and color information. In order to speed computation, the color and shape-texture features shall extract directly from the full leukocyte image (including some erythrocytes and the plasma background), i.e., no leukocyte segmentation is required. After the learning processing, where different feature selection and classifier definition alternatives were tested, a definitive approach was proposed: combined use of color (color histograms in the HSV and Lab spaces) and texture (morphological color granulometries) features. The performance of this approach, precision of 95%, allows the integration of the algorithm as an important module of our system which yields the identification of lymphocytes [15].

Lymphocyte feature extraction and classification. Taking the nuclear and cytoplasmic segmented color images as the starting point, a set of morphological quantitative parameters are calculated (around 30 parameters, including complex measurements as granulometric curves for chromatin texture analysis or combination of several parameters as the used ones for nuclear shape description), examples are shown in Fig. 6.

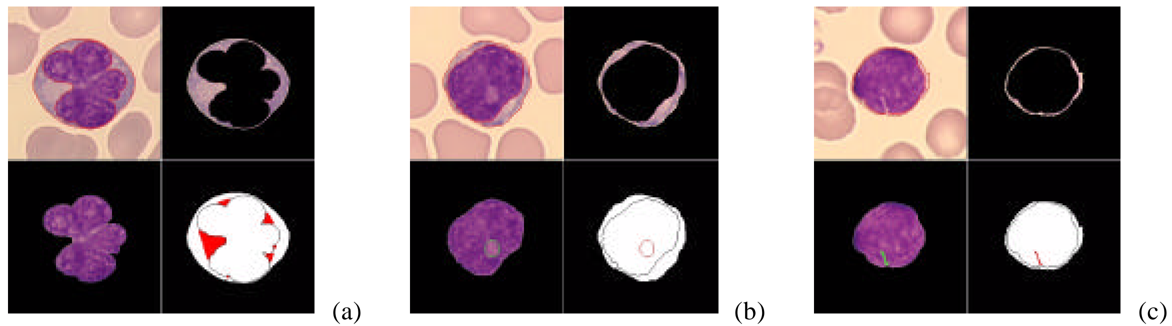

Figure 6. Examples of lymphocyte feature extraction: (a) polylobular nuclear shape, (b) nucleolus extraction, (c) cleaved nucleus detection.

These measurements yield a first classification of each morphological feature: (1) nuclear size, (2) cell size, (3) N/C ratio, (4) chromatin density, (5) nuclear shape, (6) nucleolus, (7) cytoplasmic basophilia, (8) cytoplasmic granulations, (9) cytoplasmic shape, and (10) nuclear excentration; into a category (for instance, the nuclear size may be very small, small, medium or large). The classification steps are carried out using classical statistical learning techniques and fuzzy logic. The set of morphological feature values constitutes the lymphocyte descriptor. The next step involves the use of the morphological descriptor for classifying the lymphocyte into a cellular typology (for instance, small lymphocyte, B-like lymphocyte, Hairy cell, etc.), followed by a classification of the lymphocyte population [16][17].

Information indexation. When the set of descriptors and classification information are computed, the resulting description takes the form on an XML document (normative format in MPEG-7, very efficient for editing, searching, filtering and processing information in databases), using an indexing module. This morphological description provides an objective information which may be indexed together with the *teleslide* and other complementary data: clinical features, immunophenotype, genetics, etc.

Discussion and conclusions

We have presented the satisfactory use of morphological operators for microscopic image analysis applied to haematological cytology. We have introduced a global image-based approach for the analysis and indexation of peripheral blood smear *teleslide* images which provides an objective description and classification of blood cells and which is designed to assist pathologist to diagnose haematological disorders. In a second phase, the extension of our platform to malaria study is currently developed; for instance, erythrocyte morphology (shape and color) and erythrocyte inclusions yields important knowledge for detection of malaria status and seriousness.

The algorithms have been based on methodological tests for a wide image database. New applications such as epidemiological databases, collaborative diagnosis, telehaematology, etc., can be developed adding networking abilities to the system, Fig. 7. Moreover, at present we are preparing an experience of teleconsensus, entailing the opinion of four experts, in order to validate the descriptors and classification results.

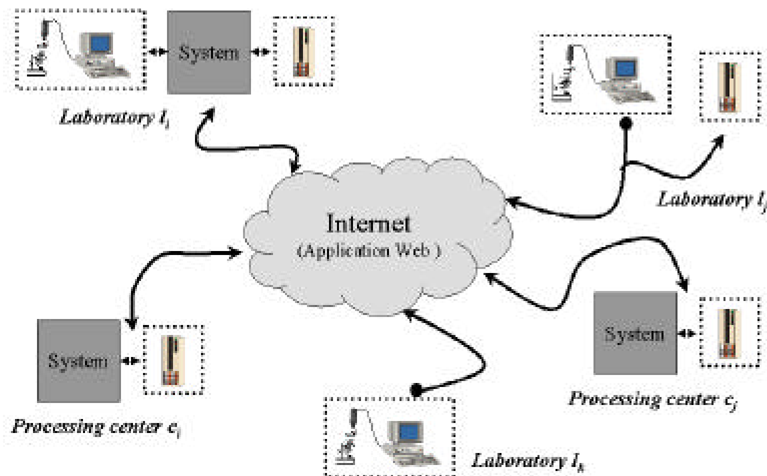

Figure 7. Adding networking abilities to the system: laboratories with or without processing capacity, local and centralised databases, distributed processing capacity, etc.

Acknowledgements

The authors would like to thank Prof. Jean Serra of *Centre de Morphologie Mathématique*, Fontainebleau, France, for scientific supervision of the work. Thanks are also addressed to Jacques Klossa of the company TRIBVN™, Paris, France for technical support.

References

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

[2] World Health Organization, *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*, Edited by E. S. Jaffe, N. L. Harris, H. Stein and J. W. Vardiman, IARC Press, Lyon, France, 2001 .

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

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

[5] 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), pp. 587-589.

[6] 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), pp. 1-5.

[7] J. Klossa, J. C. Cordier, G. Flandrin, C. Got and J. Hémet, A European de facto standard for image folders applied to telepathology and teaching, *Journal of Medical Informatics*, **48**(1998), pp. 207.

[8] J. Klossa, G. Flandrin and J. Hémet, Teleslide: Better slide representativeness for digital diagnostic microscopy applications, *6th Congress of European Group of Telepathology*, Crete, Greece, 14-21 September, 2002.

[9] J. Serra, *Image Analysis and Mathematical Morphology. Vol I and Vol II: Theoretical Advances,* Academic Press, London, 1982 and 1988.

[10] J. Serra, *Course on morphological operators*, http://cmm.ensmp.fr/~serra/cours/index.htm, Centre de Morphologie Mathématique, Ecole des Mines de Paris, 2002.

[11] F. Meyer, Automatic Screening of Cytological Specimens, *Computer Vision, Graphics, and Image Processing*, **35**(1986), pp. 356-369.

[12] F. Meyer and S. Beucher, Morphological segmentation, *Journal of Visual Communication and Image Representation*, **1**(1990), pp. 21-45.

[13] J. Angulo and G. Flandrin, Automated detection of working area of peripheral blood smears using mathematical morphology, To appear in *Analytical Cellular Pathology*, 2002.

[14] J. Angulo and J. Serra, Semantic-based hierarchical segmentation of color blood smear images using mathematical morphology, Submitted to *Medical Image Analysis*, August 2001.

[15] J. Angulo and J. Serra, Morphological color size distributions for image classification and retrieval, *Proceedings of ACIVS'2002 (Advanced Concepts for Intelligent Vision Systems)*, Ghent, Belgium, September 9-11, 2002.

[16] J. Angulo, J. Serra and G. Flandrin, Quantitative descriptors of the lymphocytes, *Analytical Cellular Pathology*, **22**(2001), pp. 69-70 (Abstracts of the 7th Congress of the European Society for Analytical Cellular Pathology).

[17] J. Angulo, J. Serra and G. Flandrin, Quantitative descriptors of the lymphocytes using mathematical morphology. First Part: Definition of descriptors and Second Part: Automated classification, *Internal Notes N-48/01/MM, N-/02/MM, CMM-Ecole des Mines de Paris*, France, 2001-2002.