ONTOLOGY-BASED LYMPHOCYTE POPULATION DESCRIPTION USING MATHEMATICAL MORPHOLOGY ON COLOUR BLOOD IMAGES

J. ANGULO[1], J. KLOSSA[2] and G. FLANDRIN[3]

[1]Centre de Morphologie Mathématique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77 300 Fontainebleau, France
Fax: +33 1 64 69 47 07; Email: jesus.angulo@ensmp.fr
[2]TRIBVN Company, Paris, France
Email: jklossa@tribvn.com
[3]Unité de Télémédecine, Hôpital Necker-Enfants Malades, Paris, France
Email: gflandrin@wanadoo.fr

Abstract – Despite modern technologies (immunophenotyping, molecular probing, etc.) cytomorphologic examination of stained peripheral blood smears by microscopy remains the main way of diagnosis in a large variety of diseases (e.g. leukaemic disorders). Using tools from mathematical morphology for processing peripheral blood colour images, we have developed an image-based approach, to provide an objective and understandable description of lymphocyte populations according to a specifically designed ontology. This ontology-based framework needs a conceptualisation of the problem from a morphological viewpoint, the introduction of an adapted language, the generation of representative image databases, the development of image processing and data classification algorithms to automate the procedure and the validation of the system by human expertise. In this paper we present the main concepts, algorithms and some results to illustrate the high-performance of the approach. The aim of our work is to reconcile the automatisation with the medical expertise, so that they can reinforce each other.

Key words: lymphocyte, haematological cytology, cellular ontology, quantitative image analysis, mathematical morphology, segmentation, classification, colour, shape, texture.

INTRODUCTION

Despite modern technologies (immunophenotyping, molecular probing, etc) cytomorphologic examination of stained peripheral blood smears by microscopy remains the main way of diagnosis in a large variety of diseases (e.g. lymphoproliperative disorders like leukaemia, lymphoma, etc.) (13,33).

Automated flow cytometry-based systems for leukocyte recognition are currently available in the market and are used in clinical laboratory routines. These devices are limited to identifying normally circulating leukocytes and flagging abnormal circulating cells, without being able to classify the abnormal cells (16,8).

The desirable qualities of a blood film depend on two factors: the quality of the smearing (sufficient working area) and the quality and reproducibility of the staining procedure. Both have been previously studied (6,7). The currently available commercial automated devices for performing blood film and staining lead to reproducible preparations. In addition, other technological advances in image processing, robotised microscopy, computer technology and networking motivate us to work on a new generation of automated systems to assist diagnosis in haematological cytology.

Developing reproducible parameters / measurements to describe the blood cells has been the aim of many research projects during the last thirty years; systems have indeed been developed for an automatic classification of the lymphocytes/erythrocytes. There are typically two ways to tackle the problem and organize the image processing tasks.

The first one focuses on a single morphological characteristic, disregarding all others (24,22). The corresponding system will be able to discriminate or calibrate one particular population with respect to this parameter. A good example is the measurement of the DNA distribution of a particular cell population. Such methods are limited in their scope and unable to describe the cellular morphology in its totality. The second type of approaches address the question in a more global way: after the extraction of a large number of measurements and parameters describing all the interesting morphological characteristics, an advanced pattern recognition system provides an automatic classification of the cells in the different categories (20,10,17). The problem with these last techniques is the difficulty to interpret the results intuitively and hence the inability to validate the system by expert-observers.

The features often have no natural morphological meaning and cannot be interpreted by the medical doctors. The complete system behaves like a black box.

Using tools from mathematical morphology for processing peripheral blood colour images, we have developed an image-based approach to provide an objective and understandable description of lymphocyte populations, according to a specifically designed ontology, in order to provide a more objective and accurate estimation of the blood leukocyte morphology than that estimated by the standard microscopy analysis. The main aim is to reconcile automatisation with human medical expertise examination of the blood films.

At this stage of our knowledge the first application is to give a more reproducible observation of individual cells and to prepare their study at the level of cell population, in order to search in the future correlation with other biological parameters and/or with clinical and prognostic data.

This ontology-based framework needs a conceptualisation of the problem from a morphological viewpoint, the introduction of an adapted language, the generation of representative image databases, the development of image processing and data classification algorithms to automate the procedure and finally, the validation of the system by human expertise.

In our works, mathematical morphology results are compared with the experience of a panel of 5 expert morpho-haematologists who defend a logical decision tree and a descriptive language to base their description of cells and diagnosis. We present in this paper the main concepts, algorithms and some examples of results to illustrate the high-performance of the approach.

Mathematical morphology is a non-linear image processing approach which is based on the application of lattice theory to spatial structures (28). This technique is proven to be a very powerful tool in microscopic image analysis, as shown in this work.

The rest of the paper is organised as follows. In section 2 an overview to the approach is given, laying particular emphasis on the originality of our methodology. Then, section 3 presents the different image processing algorithms to study a lymphocyte population, illustrating the descriptions with several examples. The results are analysed and discussed in section 4. Finally, conclusions and perspectives are considered in section 5.

APPROACH OVERVIEW

Ontology-based methodology

Here we propose a new approach based on the experience of expert morpho-hemato-logists: their logical decision trees and even their descriptive language on which they base their description of cells and diagnosis will be converted to an automatic classification and decision.

This methodology is based on the definition of an ad-hoc lymphocyte description ontology, which contains the whole morphological information of the lymphocyte. The structured and hierarchical ontology is composed of all the significant morphological characteristics, considered in an individual way. For each characteristic (e.g. nuclear size), it is necessary to fix the qualitative values of the categories so that one could assign to this characteristic (e.g. very small, small, medium or large). The morphological characteristics and its values must be interpretable and "measurable" intuitively, so that the ontology is valid for automatic and manual morphological studies.

To accomplish this objective it was necessary to develop a multidisciplinary approach in several steps: 1st- to retain the characteristics studied and to define the possible categories; 2nd- to constitute databases with examples of lymphocyte populations for each characteristic and its categories (used for training the system, then to validate it); 3rd- to develop automated algorithms of quantification as well as classification, to go from the measurements/parameters to the values of the characteristics; 4th- to validate the system by manual review of results according to a panel of experts.

Lymphocyte ontology

Consequently, in order to characterise a lymphocyte we propose the following descriptor composed of ten relevant morpho-logical/colorimetric criteria, and each criterion is declined into several categories (each category is illustrated with an example of lymphocyte)

This ontology constitutes an evolution with respect to the classical previous semiological works in the field(9,8).

1 Nuclear Size	
very small	NSI 1-1
small	NSI 1-2
medium	NSI 1-3
large	NSI 1-4

1-1 1-2 1-3 1-4

2 Cell Size	
very small	CSI 2-1
small	CSI 2-2
medium	CSI 2-3
large	CSI 2-4

2-1 2-2 2-3 2-4

3 N/C Ratio	
very low	NC 3-1
low	NC 3-2
medium	NC 3-3
high	NC 3-4
very high	NC 3-5

3-1 3-2 3-3 3-4

3-5

4 Chromatin Density	
very low	CD 4-1
low	CD 4-2
dense	CD 4-3
very dense	CD 4-4

4-1 4-2 4-3 4-4

5 Nuclear Shape	
regular, round	NSH 5-1
regular, oval	NSH 5-2
weakly irregular, oval	NSH 5-3
weakly irregular, kidney shape	NSH 5-4
irregular, deep indentation (n)	NSH 5-5
irregular, narrow indentation	NSH 5-6
grooved chromatin pattern	NSH 5-7
binucleated	NSH 5-8

5-1 5-2
5-3 5-4
5-5 5-6
5-7 5-8

6 Nucleolus	
absent	NU 6-1
small, multiple	NU 6-2
single, prominent	NU 6-3

6-1 6-2 6-3

7 Cytoplasmic Basophilia	
very weak	CB 7-1
weak	CB 7-2
strong	CB 7-3
very strong	CB 7-4

7-1 7-2 7-3 7-4

8 Cytoplasmic Granulations	
absent	CG 8-1
present	CG 8-2

8-1 8-2

9 Cytoplasmic Shape	
regular	CSH 9-1
irregular, budding	CSH 9-2
irregular, villosity	CSH 9-3

9-1 9-2 9-3

10 Nuclear Excentration	
absent	NEX 10-1
moderate	NEX 10-2
strong	NEX 10-3

10-1 10-2 10-3

Consensus between the experts and consistence of data/results

This lymphocyte ontology and the associated databases are the result of a protocol between a panel of experts in the field, which guarantees the soundness and the consistency of the structure terminology and of the reference images. Typically, the protocol needs several iterations until stability of results to reach the consensus.

Fig. 1 shows the Web GUI developed for aiding in the manual reading of cells. The lymphocyte to be described is the image on the top part and the user must choose the corresponding category for each feature. The image on the bottom allows the visualization of the examples for the different categories. These kinds of tools are used during the constitution of the reference databases and also during the validation of the system. Each expert makes a different lecture, and then the coordinator proposes a consensual lecture according to the previous ones, which is finally validated by the other experts.
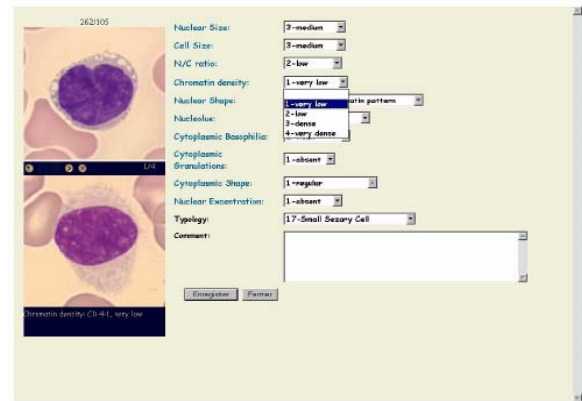

Fig. 1   Web GUI developed for manual reading of cells (see text).

System architecture

The architecture of the morphological image processing software is decomposed into several modules associated to the different steps; see the diagram depicted in Fig. 2. Two successive phases are involved in the system: blood image production and blood image analysis.

A motorised microscope allows the automated production of digital blood colour image fields. The microscopic images are acquired for storing, transmitting and processing.

The classical solution is the concept of image folder (including a few representative images) which has the lack of slide representativeness, limiting some applications. The virtual slide is currently proposed as an alternative solution: a high resolution image at diagnostic magnification of each slide, composition of hundredth of fields (21). We have also previously developed a module which implements 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 (typically immersion objective ×25 or ×16). See details in (1). When the best area is detected, the magnifying power is increased (typically ×100) and the folder of images or the virtual slide images are produced using a routine motorised microscope and a dedicated software.
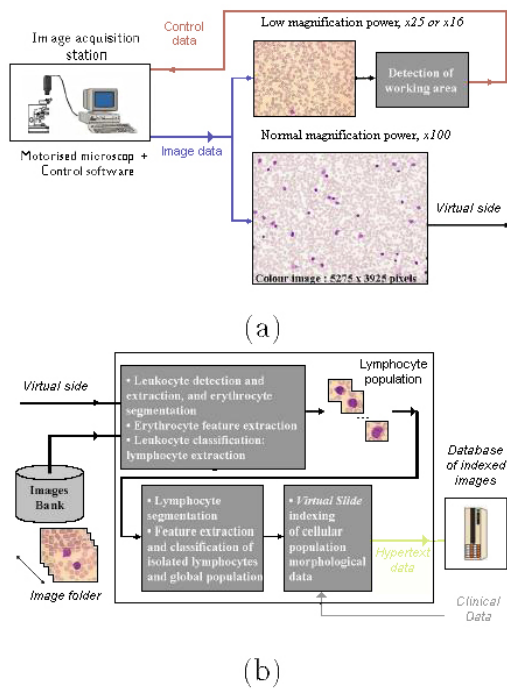
Fig. 2 Automated morphometric analysis of peripheral blood smear cells. (a) Colour image production, (b) integrated image analysis, pattern recognition and content-based indexing system.

The automated analysis of blood cell images is decomposed into several phases. The leukocytes are detected and the lymphocytes identified and extracted, to generate a lymphocyte population. Each lymphocyte is segmented into two regions: nucleus and cytoplasm. From the nucleus and cytoplasm colour images, a set of quantitative parameters are calculated. These measurements yield a classification of each morpho-logical feature into a category. The set of morphological feature values constitutes the lymphocyte descriptor, and by extension, the population descriptor. The morpho-logical description provides objective information which may be indexed into a database together with the initial images and/or used for (tele) diagnosis procedures and research protocols.

MORPHOLOGICAL IMAGE PROCESSING ALGORITHMS

Let f_{blood} be the original colour image ("small" image from a folder or "large" image from a virtual slide because the algorithms are valid for both kinds of images).

Colour blood segmentation

In a system using digital image processing methods for automation of quantitative cytology, the segmentation appears to be the most crucial part: errors in the segmentation process may propagate to the feature extraction and to the classification and finally result in an erroneous biomedical interpretation. The blood smear image segmentation has been studied using different approaches, based on different techniques: hierarchical thresholding (32) (adding geometric corrections (11)), colour clustering (12), region growing (19,18), watershed transformation (15), colour snakes (35).

The aim of the colour blood segmentation is to separate the cells entities. The segmentation module is divided into two algorithms. Firstly, a step for erythrocyte segmentation and leukocyte detection/extraction into subimages, followed by a precise lymphocyte segmentation of nucleus and cytoplasm.

Erythrocyte segmentation and leukocyte extraction – In this preliminary and fast step, the green component of the colour image has shown to be sufficient for segmenting the erythrocytes and leukocytes from the background. The algorithm involves mainly two parts: automatic thresholding and binary filtering; see fig. 3(a).
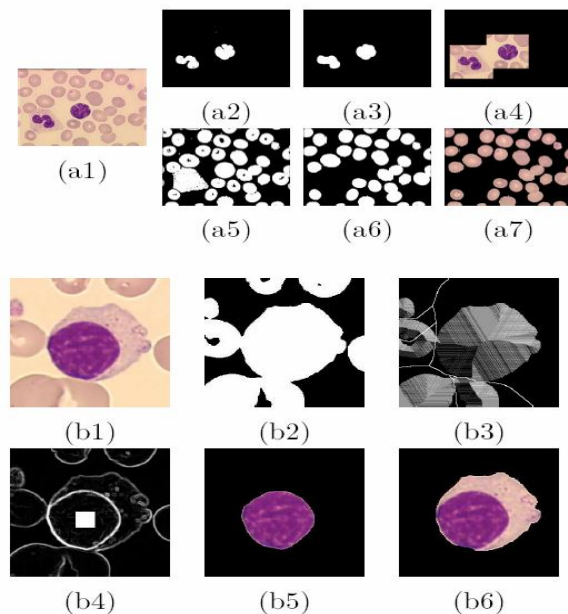

Fig. 3 (a) Algorithm for segmenting the red blood cells and extracting the leucocytes. (b) Procedure to segment a lymphocyte. See the text for full details.

5

In every blood image, four regions may be observed which are, namely, the background (the plasma), the erythrocytes and the nucleus/cytoplasm of the leukocytes. We have previously developed a morphological thresholding procedure, based on the analysis of the dynamics of histogram peaks, see (1), to robustly extract a binary image corresponding to the nuclear regions and a binary image with the erythrocytes and cytomplasic regions (fig. 3(a2) and (a5) respectively). Important mistakes are observed and therefore these binary images have to be filtered using classical morphological operators (28,1).

For the nucleus image, we sequentially execute the "Close-Holes" operator, the "Extract-Edge-Particles" operator (we are interested only in whole leukocytes not touching the border) and the "Size and Close-Grains" (resolving the cases where the nucleus is fragmented into several particles), fig. 3(a3). In order to obtain the position of each leukocyte (coordinates x and y), we use the morphological centroid (ultimate erosion) or the geometrical centre of mass (very fast computation from binary moments). These centroids define the regions of interest: a centered sub-image around each leukocyte can be automatically extracted from the whole blood image, reducing the size of images to be processed in subsequent steps, fig. 3 (a4). The leukocyte sub-images, f_{leuko_i}, are the input to the leukocyte classification module, which allows the identification of lymphocytes, a particular class of leukocytes.

For the erythrocytes-cytoplasms image, we perform a "Close-Holes" operator. Note that the erythrocytes overlapped on cytoplasmic particles have been separated and the cytoplasmic regions extracted, fig. 3(a6), for this purpose we use the geodesic reconstruction using the nuclear particles as markers (see details below). The image $f_{erythro}$ may be used to quantify the morphology of erythrocytes.

Lymphocyte segmentation – Each lymphocyte colour image, f_{lympho_i}, must be separated into two regions: the nucleus, $f^{nu}_{lympho\_i}$, and the cytoplasm $f^{cy}_{lympho\_i}$, see fig. 3(b). The lymphocyte segmentation has to be very precise as the masked colour nucleus and cytoplasm are used in the feature extraction step. We propose to combine the green component and the saturation component (from a luminance/saturation/hue representation).

The first step is a thresholding procedure on the green component, using the same algorithm than for the preliminary segmentation. After that, we have two binary images: the provisional cytoplasmic mask (fig. 3(b2)) and the provisional nuclear mask. For the cytoplasm, we start with a reconstruction using the nuclear mask as marker, in such a way that there will be only a cytoplasmic particle (the "Close-Holes" operator is also applied), fig 3(b2). Frequent contacts between the cytoplasm and the neighbouring erythrocytes are observed. The separation of this kind of overlapping particles is a problem well-resolved in mathematical morphology. The solution involves the use of the distance function of the binary set: the maxima of the distance function mark the different particles.

Then, the watershed transformation of this function (after a slight filtering to remove unimportant maxima) yields an ideal separation between these particles. Fig. 3(b3) depicts these separation lines and the shaded distance function. Note that the over-segmentation of erythrocytes is not a problem since a new reconstruction with the approximation to the nucleus extracts only the cytoplasmic profile.

In the case of the nucleus, the first step is the image simplification of the green and saturation components using a levelling (a symmetrical filter by reconstruction that simplifies the texture and preserves the contours). Then, the morphological gradient is obtained for each component by subtraction from a dilated and eroded function. The function to flood with the watershed transformation is the maximum of these two gradients, fig. 3(b4). The previously obtained cytoplasm is considered as the outer nucleus marker. For the inner markers, we construct the distance function of the provisional nuclear mask, then, the maxima of the distance function (ultimate erosion) are dilated. Using a marker for each particle of the nucleus, we guarantee the correct extraction (sometimes, the connection between the different particles of the nucleus is thin or, in fact, non-existent). Fig. 3(b4) shows the inner marker as a square. The construction of the watershed line of the gradient image associated with the inner and outer markers yields the final result. In fig. 3(b5) and (b6) the extracted nucleus and cell are shown.

Fig. 4(b) shows several lymphocytes segmented with this method (superposition of contours on the original images). In fact, as already pointed out, the problem of separation of overlapped erythrocytes on the cytoplasm and polynuclear lymphocytes is satisfactorily solved by the algorithm.
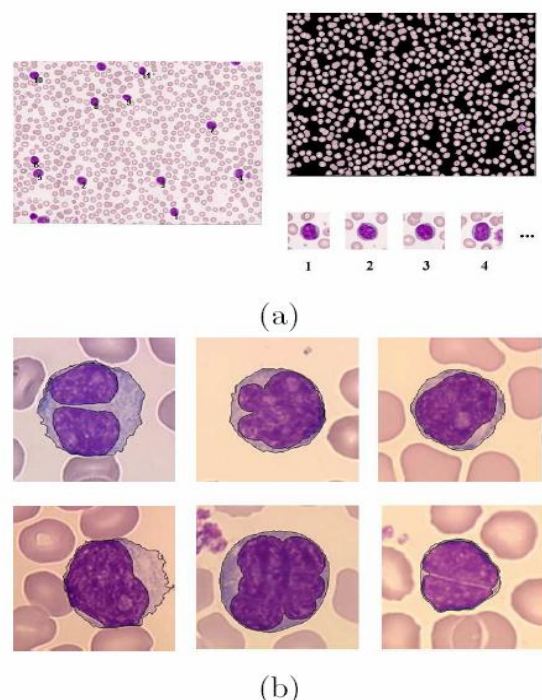

Fig. 4 Examples of colour blood segmentation: (a) segmentation of erythrocytes and extraction of leukocytes; (b) segmentation of lymphocytes (contours in black).

In fig. 4(a) an example of erythrocyte segmentation and leukocyte extraction is also given. Both algorithms have been used to segment hundreds of images and only a few of them have been considered as errors of segmentation.

Texture/colour leukocyte classification

Leukocytes may be subdivided into five different categories: 1- monocytes, 2- neutrophils, 3- basophils, 4- eosinophils and 5- lymphocytes. In a blood colour image (stained with May-Grünwuald Giemsa), the five classes of leukocytes may be differentiated according to their morphological and spectral features (29,5,31). In our approach, the texture/colour leukocyte classification is needed to detect the lymphocytes in order to constitute a representative lymphoid population.

This algorithm treats the full leukocyte image (including some erythrocytes and the plasma background), i.e., no leukocyte segmentation is required at this step, only the identified lymphocytes are then segmented.

The leukocyte classification is achieved by combining colour granulometries and colour histograms using statistical techniques (template matching by histogram distances); which identify the category of the extracted subimages $f_{leuko\ i}$. After the learning processing, where different feature selection and classifier definition alternatives were tested, a definitive approach has been proposed resulting in very high-performances. On the one hand, we use the colour histograms of the images in a luminance/saturation/hue representation (1D histogram of luminance and 2D histogram in polar coordinates of saturation/hue). In order to be independent and robust faced with the variations of the staining procedure, the histograms are normalised with respect to the colour of the plasma (background). On the other hand, an extension to colour images of the notion of granulometry curve or pattern spectrum (series of openings/closing of increasing size and a measure associated to each size) allows to define colour granulometric curves to complete the description of leukocytes. In fig. 5 an example of colour granulometric curves for a basophile is depicted, showing the pattern spectrum of luminance component and the pattern spectra of centered hue functions (weighted by the saturation), which describe the colour-texture information. As for the other algorithms, full details can be found in (2).
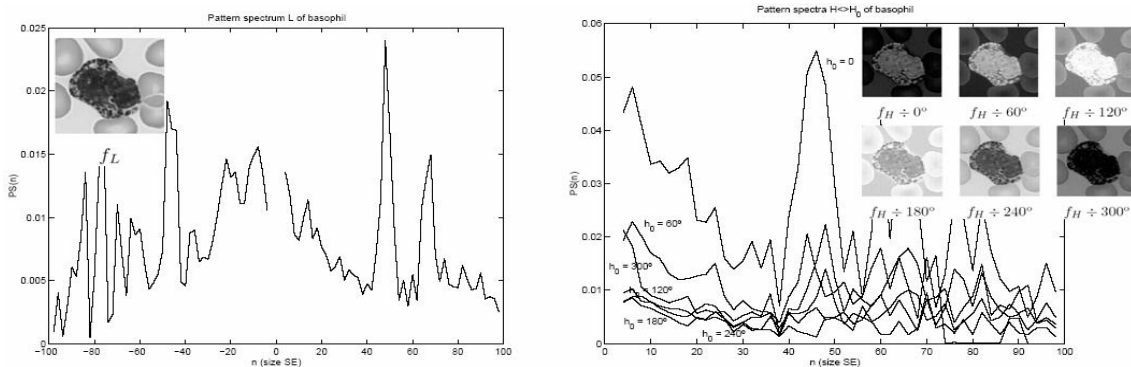

Fig. 5   Examples of colour granulometric curves for a basophil. On the left, pattern spectrum of luminance component and on the right part, pattern spectra of centered hue functions (weighted by the saturation).

Lymphocyte feature extraction and quantification

Taking the segmented nuclear and cytoplasmic colour images of each lymphocyte, $f^{nu}_{lympho\_i}$ and $f^{cy}_{lympho\_i}$ respectively, a set of morphological quantitative features (parameters, measurements, curves, etc.) is computed. We present in this section a brief summary of this quantification step.

(1) Nuclear size, (2) Cell size, (3) N/C ratio and (10) Nuclear excentration – The size of the nucleus and the cell (nucleus+cytoplasm) can be defined by means of the surface area (equal to the number of pixels) of their binary masks, S_{nu} and S_{cell}. Then, the classical nucleus to cytoplasm area ratio is defined as: $NCR = S_{nu} / S_{cell}$.

The nucleus location within the cell is a particular case of the general problem of the location of one structure within another. Many methods have been published describing some alternatives to deal with this problem (23). We propose a simple parameter to quantify the degree of eccentricity of the nucleus in the cell: the Euclidean distance of centroids, i.e., $d_{nu \rightarrow cell} = \sqrt{(x_{nu} - x_{cell})^2 + (y_{nu} - y_{cell})^2}$ where $(x,y)_{nu}$ and $(x,y)_{cell}$ are the morphological centroid of the binary mask of the nucleus and cell respectively.

(4) Chromatin density – Chromatin distribution reflects nucleus organisation of the DNA and contains important cellular diagnostic/prognostic

information. The use of the chromatin pattern in automated cell classification is complicated by several factors. In fact, the type of pattern obtained is dependent on the fixation and staining method used. In lymphocytes stained with May-Grünwuald Giemsa, the variations are less critical than in other DNA quantitative cytochemical staining procedures as the Feulgen reaction. In contrast to other morphological parameters, like shape and size, the textural parameters may not be easily assessed by visual observation. Many approaches have already been considered to automatic analyse the chromatin texture, see for instance (26,4,14,34).

We propose a morphological method for the analysis of nuclear texture based on the notion of granulometry (morphological size distributions), working on the luminance component. In fig. 6 (a) are shown some examples of nuclei for the different chromatin categories of our ontology. The density of the chromatin is given by the alternation from condensed (dark) structures and light (bright) structures. For this reason, we have combined four granulometric curves (using disks of increasing size as structuring elements): granulometry by openings (describing bright structures), granulometry by closings (dark structures) and pseudo-granulometries using erosions and dilations which complement the two other curves. An important advantage of

(a1) (a2) (b1)

(a3) (a4) (b2)

Fig. 6   Selection of nuclei with four different chromatin texture, from the least dense (a1) to the densest (a4). Granulometric curves of chromatin: (b1) pattern spectra by openings and closings and (b2) curves of erosions and dilations.

granulometries is their inherent normalisation. The four curves are normalised to the volume of the image, thus variations (under some conditions, i.e. *anamorphoses*) in the acquisition process do not have an influence on the granulometric results. In fig. 6 (b)

the corresponding results of chromatin description using the granulometric curves are depicted. We can observe their ability to differentiate the different classes of texture. In practice, Euclidean distances are used to compare the granulometries to classify with templates associated to reference populations.

(5) Nuclear shape – According to our lymphocyte ontology, the nuclear shape is the richest morphological feature in terms of category number (in fact, the case "grooved chromatin pattern" has not yet considered in our works due to its difficult morphological interpretation). Let us recall that the ideal lymphocyte nuclear shape is a circle.

Many quantitative studies on the nuclear morphology of lymphoid cells have been published, showing that simple parameters are not sufficient for reliable discrimination of the different cell typologies (24,3,22,27,30). In order to give a shape definition or classification of the nucleus, we have developed a strategy in different phases.

Firstly, we start by considering the extreme degree of irregularity: the presence of a polynuclear cell. This is achieved by thresholding the distance function of the binary mask shape, because of the parts of nucleus could be connected, see the example of a tri-nuclear lymphocyte in fig. 7(a).
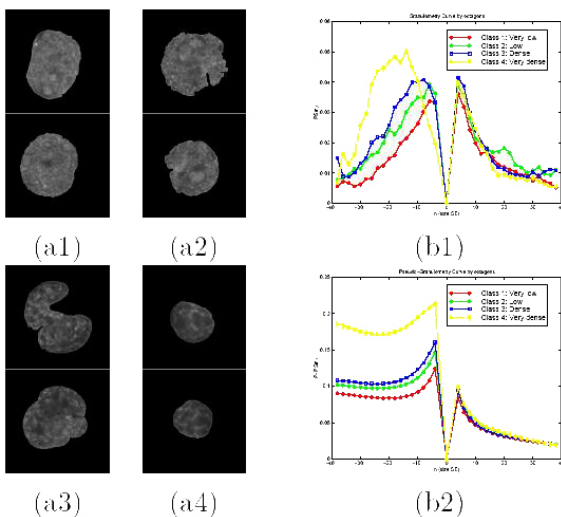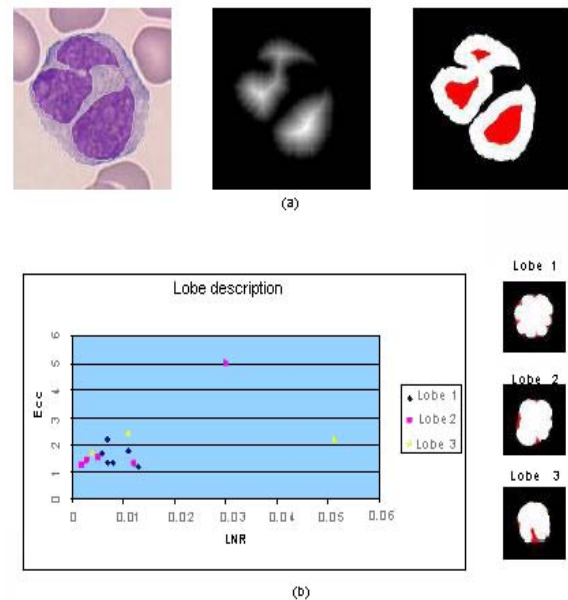

(a)

(b)

Fig. **7**   (a) Polynuclear lymphocytes and thresholding of distance function. (b) Example of polylobular nuclei description using parameters associated to the inter-lobe regions.

Then, using two simple scalar parameters, the form factor or circularity, $FF = P^2 / 4\pi S$, and the eccentricity

(obtained by using the binary moments to compute the ratio between the length of the major axis and the length of the minor axis of the ellipse of inertia), the regular shapes are separated from the irregular ones.

The irregular shapes are described by considering the nuclear lobes. In fact, computing the difference between the nuclear mask and its convex hull, the inter-lobe regions are extracted, fig. 7 (b), and these regions can be used to study the number, regularity and magnitude of nuclear shape irregularities. Starting from these significant concavity regions, several parameters of size and shape have been defined which allow a powerful description of irregular shapes and a sorting into several classes. Moreover, an additional morphological algorithm has been developed to identify, the cleaved cells (characterised by linear irregularities through the nucleus), see details in fig. 8.
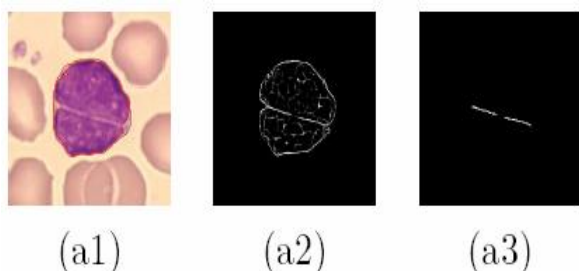

Fig. 8 Algorithm to extract linear irregular-rities, characterising the cleaved cells: (a1) original image, (a2) grey-level thinning from the green component of nucleus, (a3) result of morphological filtering (basically, the detection of lines is achieved by means a linear opening).

(6) Nucleolus – The nucleolus is a rounded body that is usually basophilic (bright), eccentrically placed in the nucleus and sharply demarcated from the surrounding nucleoplasm. We have developed a multi-step algorithm for detecting the presence of nucleoli, and in the positive case, their extraction and quantification, see fig. 9.

Our approach is mainly based on the hypothesis that the nucleolus is a bright structure surrounded by a very dark background, which presents a circular shape. Therefore, working on the green component, the first step is an adaptively stretching of the histogram in order to saturate the bright regions and to extract them. Then, several criteria of size and shape on the one hand, and of contrast between the region inside the region and outside allow to accept a candidate region as a nucleolus.
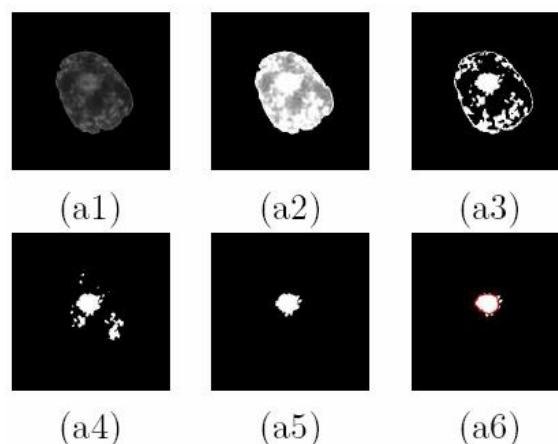

Fig. 9 Algorithm to extract the nucleolus: (a1) original green component of nucleus, (a2) contrast modification by adaptively stretching of histogram, (a3) thresholding, (a4) filtering of binary components according to size/shape, (a5) candidate used to evaluate the criterion of intensity difference, (a6) accepted nucleolus (rounded shape).

Nucleoli may be single or multiple and vary greatly in size from one lymphocyte type to another in different disorders. For our purposes we consider two situations: *single*, *large nucleolus* or *multiple nucleoli* (two or three small nucleoli, more difficult to detect and with uncertain significance). This is probably the most difficult morphological feature to study.

(7) Cytoplasmic basophilia, *(8) Cytoplasmic granulations* – Basophilia is the tendency to readily stain with basis dyes. In the case of images of smears stained with May-Grünwald-Giemsa, basophilia is the tendency to stain blue. Therefore, a study of the basophilia of cytoplasm must involve the study of cytoplasm colour.

Our approach is very simple and involves the description of the colour of the cytoplasmic region by using the L*a*b* colour representation (perceptually uniform). In fact, it is possible to discriminate the basophilia by means of the averaged values of these three colour components for the pixels of the cytoplasm. In fig. 10 an example illustrating the capacities of separation of cytoplasms according to their colour is given. A clustering on this space of three parameters allow us to define the frontiers of different classes.

The other interesting feature to be studied is the cytoplasm granulation, an important phenomenon since in lymphoid cells classification this feature is of fundamental importance. The presence of grains happens exclusively in cytoplasms with relatively weak basophilia. Consequently, the application of the detection algorithm will have to be associated to the study of basophilia level.
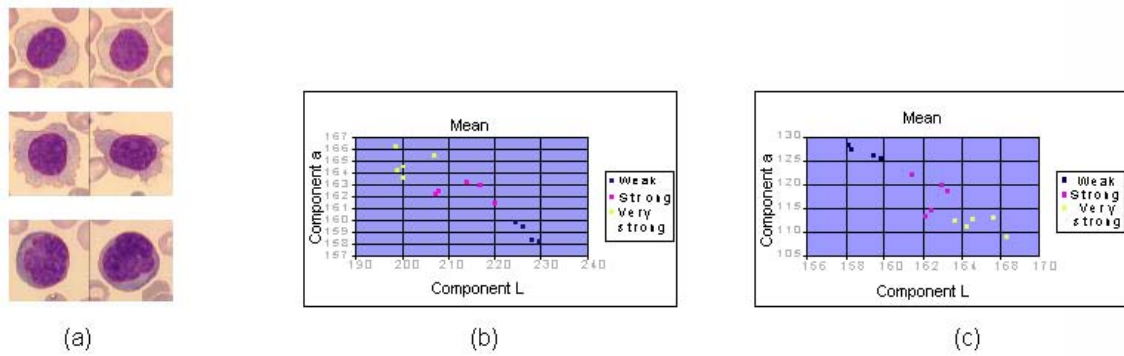
Fig. 10 (a) Examples of lymphocytes with different level of basophilia. (b1) and (b2) averaged values of colour components L*, a* and b* in the cytoplasmic region.

From an image processing viewpoint, the detection of the azurophilic granules in the cytoplasm is an easy problem: the grains are dark structures, adequately contrasted (especially in the green component). Using a top-hat transformation accompanied by a thesholding operation, these structures can be extracted (see fig. 11). Due to fact that the grains are almost circular dark particles, we use a black top-hat by a closing with a disk as structuring element.
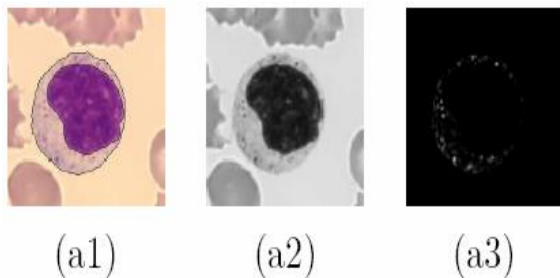

Fig. 11 Algorithm to extract the cytoplasmic granulations: (a1) original image, (a2) green component, (a3) result of black top-hat transformation (residue of a closing) limited to the cytoplasm.

(9) Cytoplasmic shape – The shape of the cell contour (or cytoplasm profile) is also useful to identify different atypical classes of lymphocytes. Fig. 12 (a) shows some examples of the three categories of cytoplasm profiles to characterise. Once again using a granulometry, by applying circular opening of increasing size on the binary masks of cytoplasms, it is very easy to identify if the shape is regular, or in the case of irregular shape, the degree of irregularity, see fig. 12 (b). In fact, two parameters are derived from the curves (integral of narrow irregularities and integral of wide irregularities) to synthesise the shape information.
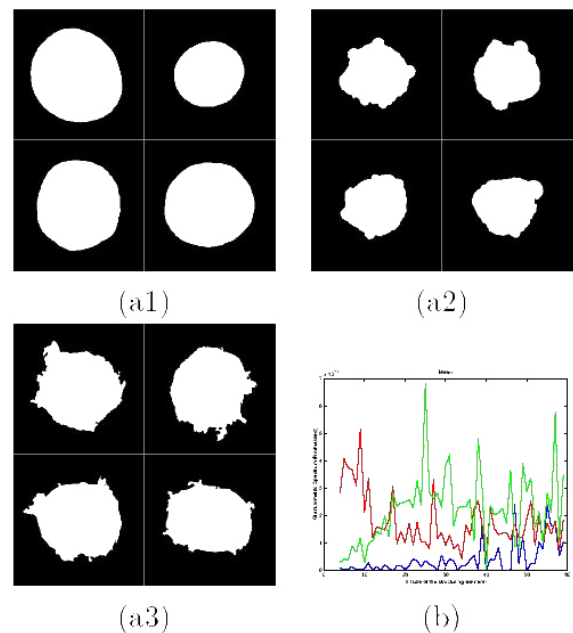

Fig. 12 Selection of nuclei with the three different cytoplasmic shape: (a1) regular, (a2) budding, (a3) villosity. (b) Granulometric curve of openings (blue for regular, green for budding and red for villosity).

Mapping lymphocytes onto the ontology

All these extracted measurements yield a classification of each morphological feature from our ontology into a pre-established category. For instance, according to the value of surface area, the nuclear size may be "very small", "small", "medium" or "large"; or according to the granulometric curve of cytoplasm binary mask, the cytoplasmic shape may be "regular", "budding irregular" or "villosity irregular". We consider this step as the procedure of mapping each lymphocyte of the population onto the ontology in order to generate the population morphological descriptor.

In practice, the morphological feature classification is a set of decision trees obtained after statistical methodological tests and training

procedures from our database. A classifier has been defined for each morphological feature and takes as input information the corresponding measurements and parameters, providing as output the corresponding category, and in some cases a level of confidence of classification. The complexity of classifiers is quite simple because the computed morphological parameters are already very adapted to the discrimination of corresponding categories; the interested reader can find all the details in (2).

In conclusion each lymphocyte of a population can be described by using well established terms of the ontology, corroborated by quantitative measurement. The resulting analysis of a population takes the form of an XML document, very efficient for editing, database indexing and searching, Web visualization, etc.

RESULTS AND DISCUSSION

A lymphocyte population can be summarised and compared to another population by means of the histograms

associated to the category values of each morphological feature (e.g. histogram of size categories or histogram of chromatin density). These representations are more useful in diagnosis purposes than the histograms of direct parameters: obviously the histogram of surface area is not easy to interpret by a hematologist or still worse, the multidimensional histograms of granulometric curves describing the texture of chromatin.

Fig. 13 shows a first example of comparison of two lymphocyte populations by means of the histograms. As we can observe from the images of analysed cells, the H400 (Chronic Lymphocytic Leukaemia) is morphologically very different from the H470 (T-Prolymphocytic Leukaemia). H400 is distinguished from H470 by distinctly smaller nuclear and cellular sizes, by a stronger density of chromatin, by more regular nuclear shapes, the quasi absence of nucleoli, and by a clearly weaker basophilia. The parameters of N/C ratio, nuclear excentration or the cytoplasmic features are indistinctive.
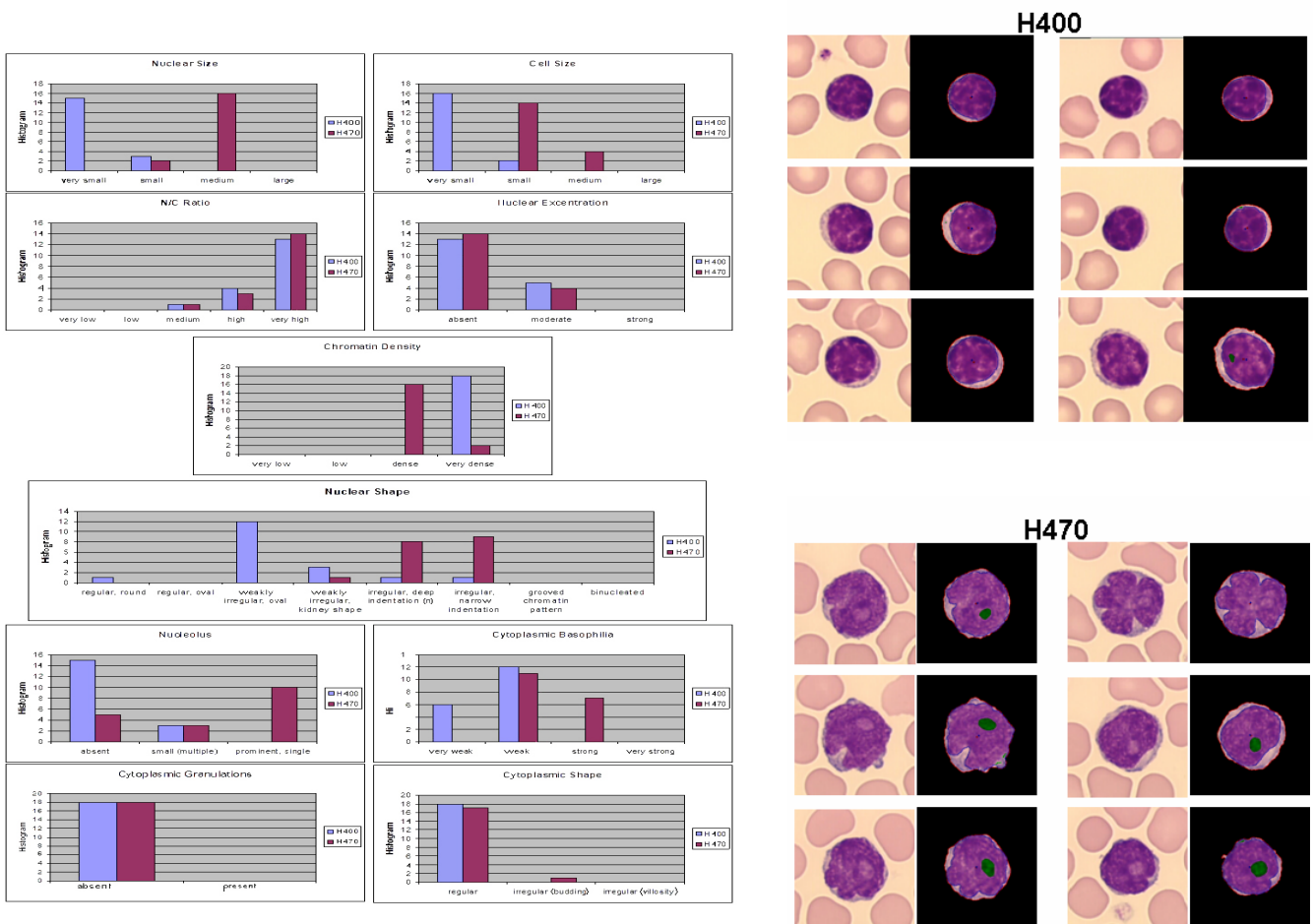

Fig. 13   Comparison of two lymphocyte populations (pathologies H400 and H470) using the histograms of the ten features. Six examples of processed cells for each population are shown on the top part.

But we can also consider the following examples of pathology, with close morphology, compared two to two, which demonstrate the discriminating character of these morphometric data; they are superimposed narrowly on the classically recognized morphological data from medical literature.

For instance, in fig. 14 the comparison of two populations of Hairy Cell Leukaemia (H410) and Splenic Lymphoma with Villous Lymphocytes (H412) are given, which are often morphologically mixed. From the four histograms shown, H410 is distinguished from H412 by higher nuclear sizes, a density of chromatin definitely weaker than in H412 (value "very dense" very rare in H410), a basophilia less marked (value "very weak" very frequent and seldom presents in H412), finally by a different distribution of nuclear excentration (more marked in H412).
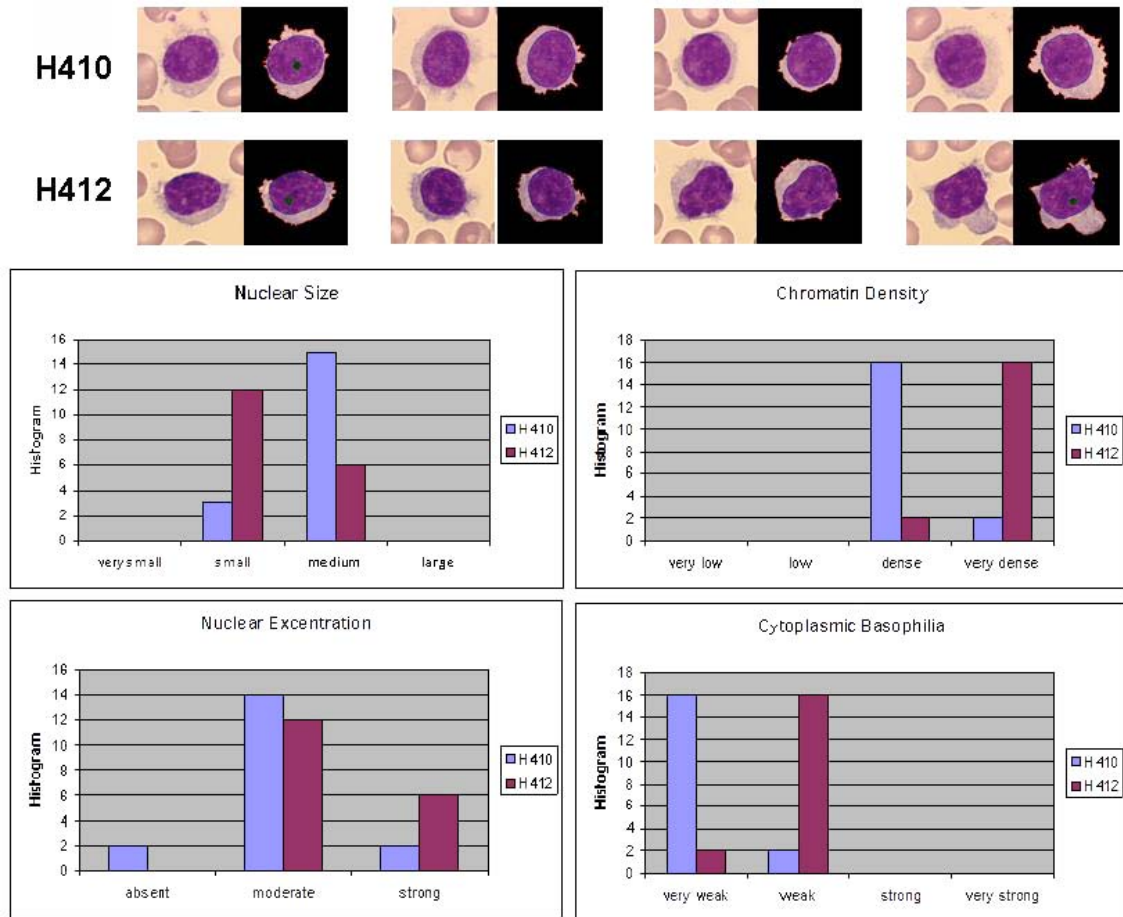

Fig. 14   Comparison of two lymphocyte populations: H410 Vs. H412.

Another example, depicted in fig. 15, compares a population of H413 (Follicular lymphoma) and Mantle Cell Lymphoma (H414). In this case, H413 is distinguished from H414 by a higher N/C ratio (less dispersed values), rarer nucleoli (form "prominent, single" always goes away), polymorphic nuclear shapes, but with a peak of value "irregular, narrow indentation" for H413. Other parameters which are not shown, like a less strong nuclear excentration, a less strong density of chromatin or a weaker cytoplasmic basophilia, are also very significant.

Particularly interesting is the comparison of two cases of Chronic Lymphocytic Leukaemia (H400) in fig. 16. The first H400 is taken after an evolution of 22 years without treatment. H400 is distinguished from the second H400-92 by distinctly smaller nuclear and cellular sizes, a higher N/C ratio (less dispersed values), a weaker nuclear excentration and by a stronger basophilia. The second H400-92 is transformed into a Plasmocytic leukaemia, H404-95, in three years of evolution. The profiles H400 and H400-92 (each one with the diagnosis of Chronic lymphocytic leukaemia) are in fact definitely different.
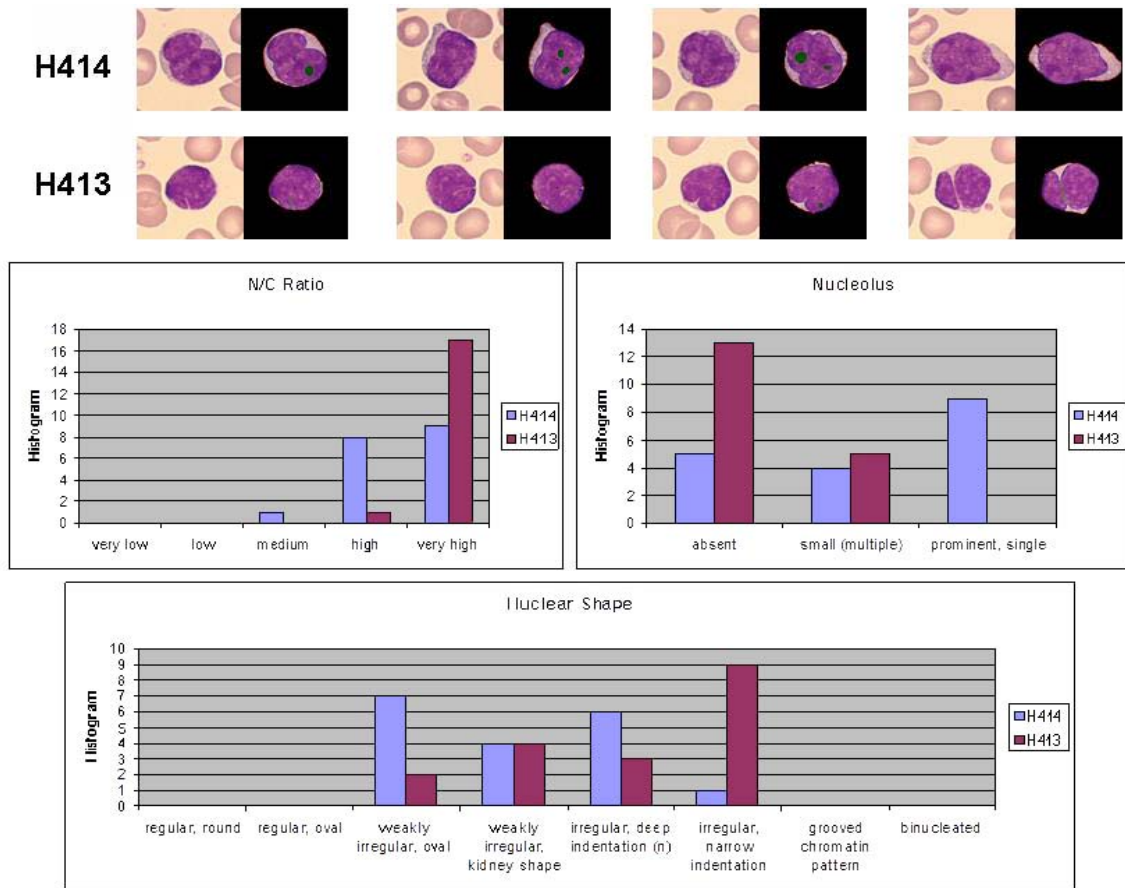
Fig. 15   Comparison of two lymphocyte populations: H414 Vs. H413.
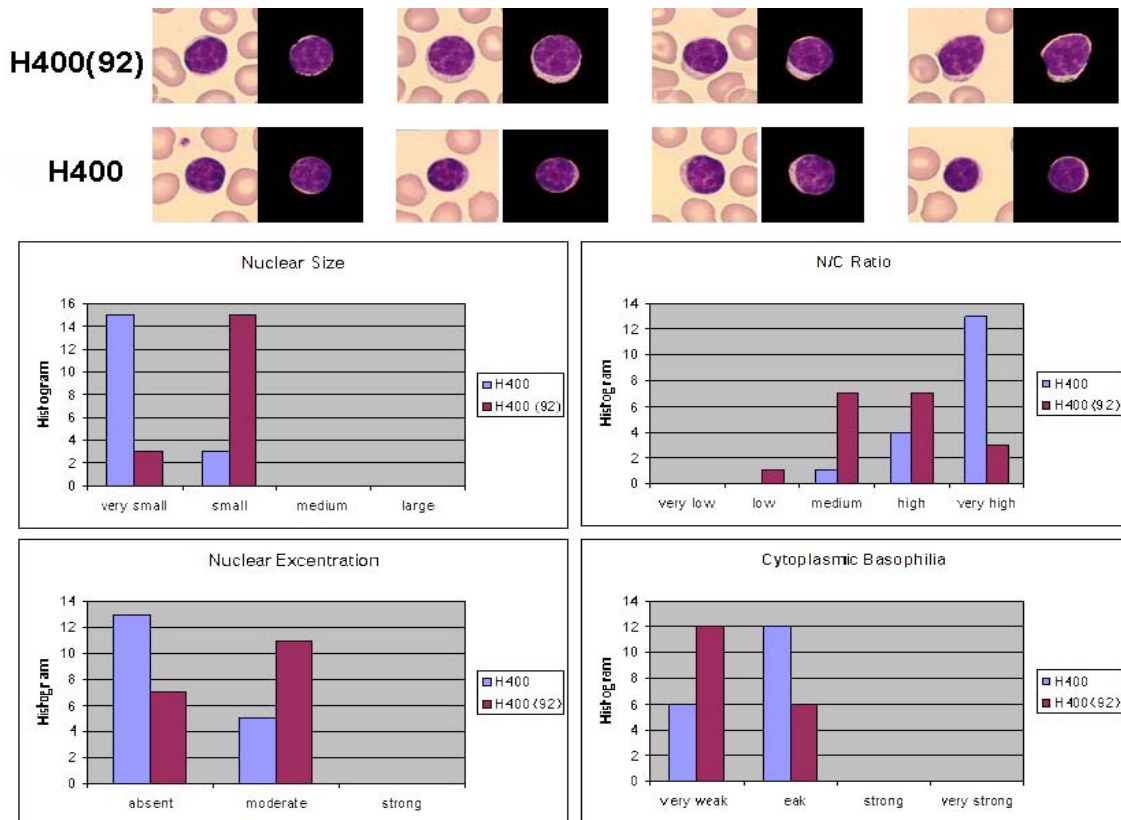

Fig. 16   Comparison of two lymphocyte populations: H400 Vs H400(92).

The H400-92 and its H494-95 transformation are also definitely morphologically different (fig. 17): H400 (92) is distinguished from H404-95 by nuclear and cellular sizes definitely smaller, a higher N/C ratio (and more dispersed values), a stronger density of chromatin and by a weaker cytoplasmic basophilia.
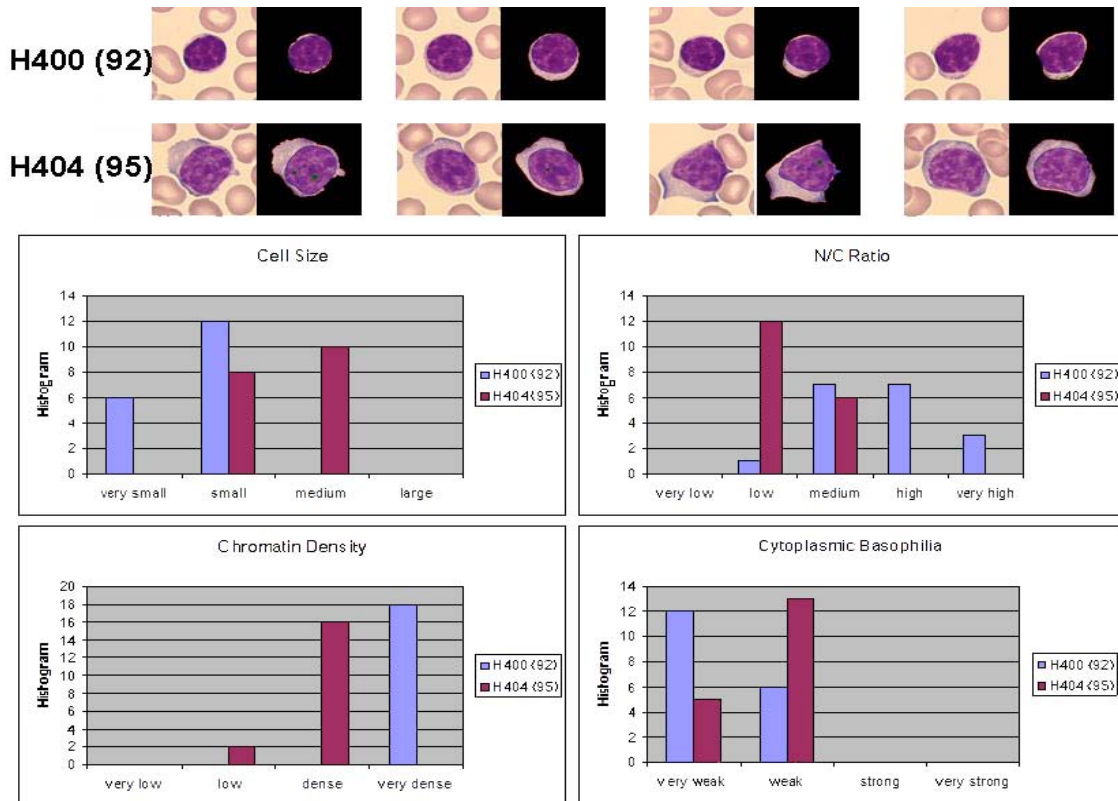

Fig. 17 Comparison of two lymphocyte populations: H400(92) Vs. H404(95).

CONCLUSIONS

Several advantages are associated with the present system. This ontology-based framework to describe automatically lymphocyte cells is coherent in its different steps with the manual procedures. The terminology, data and classification/decision trees are based on the medical expertise, which yields an interesting link between the classical medical knowledge and the automatisation of the laboratory routines. In practice, this methodology improves and makes easier the development and validation of automated image processing systems. In fact, other biomedical problems which are usually studied by subjective interpretation of images (e.g., retinal images in ophtalmology, etc.) can be approached by means of conceptually similar systems.

The automated image-based measurements and derived descriptors are objective and reproducible. The morphological data obtained can be used to automate the classical diagnosis. In addition, this type of quantitative approach provides the necessary tools to bring precise details within diagnoses already strongly elaborate and to advance in the consensus necessary to all the related disciplines in experimental medicine.

With the reservation of a larger evaluation of "observer versus measurement results" and an application on cell population by a more automatised procedure (selection of zone of interest on wide microscopic field), this approach may allows to study the evolution of a pathology according to the treatment and the system can also be applied in clinical protocols.

After this first validation of the approach, subsequent biomedical studies will be able to highlight its practical usefulness for the evolutionary study of lymphoid pathologies and for discrimination among various groups of similar diagnosis.

The results obtained by the system can be indexed and can be consulted from databases, with the possibility of searching for similar morphological content. Moreover, due to the fact that there are fewer haematologists than in the past who are specialists in cellular morphology, this kind of systems can also be used for teaching activities.

A similar approach has been developed for describing and quantifying the erythrocyte cell morphology (red blood cells) and the results have been also promising (25).

REFERENCES

1. Angulo, J., Flandrin, G. (2003) Automated detection of working area of peripheral blood smears using mathematical morphology. *Analytical Cellular Pathology*, **25(1)**:37–49.

2. Angulo J. *Morphologie mathématique et indexation d'images couleur. Application à la microscopie en biomédecine.* Ph.D. Thesis, Ecole des Mines, Paris, December 2003.

3. Ball, P.J., van der Valk, P, Kurver, P.H.J., Lindemann, J., Meijer, J.L.M. (1985) Large cell lymphoma. Differential diagnosis of centroblastic and B-immunoblastic subtypes by morphometry on cytologic preparations. *Cancer*, **55**:486–492.

4. Beil, M. (1992) Description of chromatin structures in cell nuclei. *Acta Stereologica*, **11**:129–134.

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

6. Benattar, L., Flandrin, G. (1999) Morphometry and Quality Control for a May-Grünwald Giemsa stained preparation. A 40 centers cooperative study. *Leukemia and Lymphoma*, **33**:587–589, 1999.

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

8. Benattar, L., Flandrin, G. (2001) Morphometric and Colorimetric Analysis of Peripheral Blood Smears Lymphocytes in B-Cell Disorders: Proposal for a Scoring System. *Leukemia and Lymphoma*, **42**:29–40, 2001.

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

10. Bergmann, M., Heyn, H., Müller-Hermelink, H.-K., Harms, H., Aus, H.M. (1990) Automated recognition of cell images in high grade malignant lymphoma and reactive follicular hyperplasia. *Analytical Cellular Pathology*, **2**:83–95.

11. Brenner, J.F., Lester, J.M., Selles, W.D. (1981) Scene segmentation in automated histopathology: Techniques evolved from cytology automation. *Pattern Recognition*, **13**:65–77.

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

13. D'Arena, G., Keating, M.J., Carotenuto, M. (2000) Chronic Lymphoproliferative Disorders: An Integrated Point of View for the Differential Diagnosis, *Leukemia and Lymphoma*, **36(3–4)**:225–237.

14. Dawson, A.E., Cibas, E.S., Bacus, J.W., Weinberg, D.S. (1993) Chromatin Texture Measurement by Markovian Analysis. Use of Nuclear Models to Define and Select Texture Features. *Analytical and Quantitative Cytology and Histology*, **15**:227–235.

15. Di Ruberto, C., Dempster, A., Khan, S., Jarra, B. (2000) Segmentation of Blood Image Using Morphological Operators. Dans *Proc. 15th International Conference on Pattern Recognition*, 401–404, Barcelona, Spain.

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

17. Foran, D. J., Comaniciu, D., Meer, P. et Goodell, L.A. Computer-Assisted Discrimination Among Malignant Lymphomas and Leukemia Using Immunophenotyping, Intelligent Image Repositories, and Telemicroscopy. *IEEE Trans. on Information Technology in Biomedicine*, **4(4)**:265–273, 2000.

18. Garbay, C. (1986) Image Structure Representation and Processing: A Discussion of Some Segmentation Methods in Cytology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**:140–146.

19. Haussmann, G., Liedtke, C.E. (1984) A Region Extraction Approach to Blood Smear Segmentation. *Compututer Vision, Graphics, and Image Processing*, **25**:133–150.

20. Juan, J., Sigaux, F., Flandrin, G. (1985) Automated Classification of Lymphoid Cells. *Analytical and Quantitative Cytology and Histology*, **7**:38–46.

21. Klossa J., Flandrin, G., Hemet, J. (2002) Teleslide: better slide representativeness for digital diagnostic microscopy applications. Dans *Proc. of the 6th Congress of the European Group of Telepathology*, Heracliton Crete, Greece, September.

22. Lesty, C., Raphael, M., Nonnenmacher, L., Leblond-Missenard, V., Delcour, A., Homond, A., Binet, J.L. (1986) An application of mathematical morphology to analysis of the size and shape of nuclei in tissue sections of non-Hodgkin's lymphoma. *Cytometry*, **7**:117–131.

23. Lesty, C. (1994) 2D modelling of nucleus location in lymphoid cells. *Analytical Cellular Pathology*, **6** : 65–82.

24. Litovitz, T.L., Lutzner, M.A. (1974) Quantitative measurements of blood lymphocytes from patients with chronic lymphocytic leukemia and Sezary syndrom. *J. Natl. Cancer Inst.*, **53**:75–77.

25. Luengo-Oroz, M.A., Angulo, J., Flandrin, G., Klossa, J. (2005) Mathematical morphology in polar-logarithmic coordinates. Application to erythrocyte shape analysis. In *Proc. of the 2th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2005)*, Estoril, Portugal, June 2005. Lecture Notes in Computer Science, Vol. LNCS **3523**, pp 199–205, Springer.

26. Meyer, F. (1980) Quantitative analysis of the chromatin of lymphocytes: An essay on comparative structuralism. *Blood Cells*, **6**:159–172.

27. Pelstring, R.J., Swerdlow, S.H. (1987) Improved nuclear contour indices for lymphoid morphometry. *Analytical and Quantitative Cytology and Histology*, **9**:469–468.

28. Serra, J. (1982, 1988) *Image Analysis and Mathematical Morphology. Vol I, and Image Analysis and Mathematical Morphology. Vol II: Theoretical Advances.* Academic Press, London.

29. Song, X., Abu-Mostafa, Y., Sill, J., Kasdam, H. (1997) Incorporating contextual information into white blood cell image recognition. In *Advances in Neural Information Processing Systems*, MIT Press.

30. Stevens, M.W., Fazzalari, N.L., Crisp, D.J. (1987) Quantitation of nuclear shape in non-Hodgkin's lymphoma. *Analytical and Quantitative Cytology and Histology*, **9**:459–468.

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

32. Wermser D., Haussmann, G., Liedtke, C.-E. (1984) Segmentation of Blood Smears by Hierarchical Thresholding. *Compututer Vision, Graphics, and Image Processing*, **25**:151–168.

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

34. Wolf, G., Beil, M., Guski, H. (1995) Chromatin Structure Analysis Based on a Hierarchic Texture Model. *Analytical and Quantitative Cytology and Histology*, **17**:25–34.

35. Yang L., Meer, P., Foran, D.J. (2005) Unsupervised Segmentation Based on Robust Estimation and Color Active Contour Models. *IEEE Trans. on Inf. Technol. in Biomedicine*, **9(3)**: 475–486.