Classification of lymphoproliferative disorders by spectral imaging of the nucleus

H. Greenspan¹, C. Rothmann², T. Cycowitz², Y. Nissan¹, A.M. Cohen³ and Z. Malik²
¹Faculty of Engineering, Tel-Aviv University, Tel-Aviv, ²Faculty of Life Sciences, Bar-Illan University, Ramat-Gan and
³Hematology Unit, Rabin Medical Center, Petach-Tiqwa, Israel

Summary. Spectral nuclear morphometry was used for the classification of lymphocytes in lymphoproliferative disorders. May-Grunwald-Giemsa-stained blood specimens were taken from thirty patients with infectious mononucleosis, non-Hodgkin lymphoma or chronic lymphocytic leukemia, and from ten healthy individuals. Blood specimens were analyzed by spectral imaging. Seventeen distinct spectra were collected into a spectral library and a distinct pseudo color was assigned to each one of them. The library was used to scan all the cells in the database and to create a spectrally classified image of each cell. The spectral map, per cell, reveals distinct spectral-response regions in each cellular compartment, via the distinct region colors.

Computational analysis of the spectral maps allows for the objective quantification of a set of parameters, or features, representing the cell. The features used in this work include the area and perimeter of the nucleus, circularity, edginess and the spectral pattern. The analysis pursued showed that each class of cells is associated with a set of unique parameters. We conclude that spectral analysis combined with feature analysis provides significant information in the analysis of lymphoproliferative disorders and may serve as an additional tool for the histopathological evaluation of disease.

Key words: Lymphocytes, Chronic lymphocytic leukemia, Non-Hodgkin’s lymphoma, Infectious mononucleosis, Spectral imaging

Introduction

Precise diagnosis and cell-classification play a critical role in the choice of treatment regimens of lymphoproliferative disorders. Routine diagnosis is based on the examination of May-Grunwald-Giemsa (MGG) stained specimens using a bright-field microscope. However, morphology alone is not sufficient for precise classification of lymphoproliferative disorders and usually an additional method is used such as immunohistochemical staining of membranal-antigens.

Our study deals with three lymphocyte-associated diseases: infectious mononucleosis, non-Hodgkin’s lymphoma and chronic lymphocytic leukemia. Infectious mononucleosis (IM) defines any blood lymphocytosis induced in response to an infectious disease. The principal feature of the diagnosis is the presence of a predominant blood lymphocytosis exceeding 50 percent of the blood white count, with an "atypical" (reactive) morphology in at least 10 percent of the blood lymphocytes. The latter are T lymphocytes undergoing blastogenesis as part of a cellular immune response to infected host cells. Diagnosis is established using immunofluorescence techniques to measure specific antiviral antibodies.

Non-Hodgkin’s lymphomas (NHLs) are a heterogeneous group of lymphoproliferative malignancies (Armitage, 1993). Needle biopsies are sometimes insufficient to specify the histologic type of lymphoma and lymph node biopsies are recommended whenever possible (Pinkus, 1996). The diagnosis is based on histological evaluation and determination of surface markers.

Chronic lymphocytic leukemia (CLL) is diagnosed mainly by an over-accumulation of small, mature-appearing B-lymphocytes in the peripheral blood (Kipps, 1995; Cheson et al., 1996; Avila-Carino et al., 1997). A sustained cell count of more than 5x10³/µl mature-appearing lymphocytes may suggest the transformation from a normal state to leukemia. The majority of normal lymphocytes and the dominant cellular population of CLL both consist of small cells with dense, clumped nuclear chromatin, which makes the distinction between these lymphocytic populations difficult by conventional light microscopy (Silber et al., 1990). Some cases differ in morphological feature from the typical mature, small cell B-CLL: (i) a mixture of small lymphocytes and prolymphocytes (>10% and <55%) designated as CLL/PL, or (ii) CLL mixed with large lymphocytes (Bennett et al., 1989). The French-American-British group (FAB) has proposed criteria based on
Spectral imaging of lymphocytes

Cytological specimens

Cytochemical and immunological methods in order to establish a clear diagnosis (Bennett et al., 1989; Bain, 1993). Immunophenotypic analysis reveals large amounts of surface immunoglobulin (slg) in normal B cells which is only weakly expressed, or undetectable, on B-CLL cells. The CLL cell expresses the pan-B antigens CD19 and CD20 and the activation antigens CD5 and CD23, however, it does not express the terminal B-cell differentiation antigens exhibited by plasma cells (Freedman and Nadler, 1993; Harris et al., 1994). B-CLL cells express either kappa or lambda-light chains and the monoclonality is essential to establish the diagnosis. Receptors for mouse red blood cell rosettes (MRBC-R) are detectable on both B-CLL cells and normal B cells, however, the two populations express different patterns of complement receptor.

The first stage of any lymphocyte-associated diagnosis (including all of the cases above), is based on the examination of a cytochemically-stained blood specimen using a bright-field microscope. Stained cytological specimens absorb, reflect, or emit photons in ways characteristic to their interaction with specific dyes (Rothmann et al., 1998). The metabolic properties of each cell are therefore reflected by its spectral image. The nuclear chromatin structure is modulated by histone-histone, histone-DNA interactions, histone modifications and by the presence of non-histone proteins (Woodcock and Horowitz, 1995). The accessibility of a specific DNA sequence is a consequence of the compaction-state of the chromatin in which it is located. Thus, the accessibility of highly condensed chromatin is likely to be extremely restricted (Woodcock and Horowitz, 1995). Variability in chromatin organization in differentiating cells and in abnormal cells is regarded as a common tool in cytology and histology. Spectroscopic analysis acquires the spectral information and enables the mapping of spectrally similar domains in the analyzed specimens. The collection of a large amount of data is required for the construction of an image database in order to achieve profound conclusions regarding the metabolic nature of the cells.

Spectral imaging has already been used in a number of cytological applications such as the differentiation and apoptosis of red blood cells (Rothmann et al., 1997), spectral morphometry of lymphocytes in chronic lymphocytic leukemia (Malik et al., 1998) and characterization of nuclear aberrations in differentiating red blood cells (Rothmann et al., 1997). In the present study we combine spectral imaging and microscopy, for the representation of white blood cells, towards the detection and classification of lymphoproliferative disorders.

Materials and methods

Cytological specimens

Peripheral blood was obtained from 30 patients with typical B-CLL, Hodgkin’s lymphoma and infectious mononucleosis, which were diagnosed according to established criteria (Bennett et al., 1989; Cheson et al., 1996) and from 10 healthy individuals (Hematology Unit, Rabin Medical Center, Golda-Hasharon Campus). The patients did not receive chemotherapy for at least 3 months before samples were obtained. Staining was performed according to the standard MGG method (specimens were stained in May-Grunwald for 7 minutes and washed in running tap water for 1 minute. The specimens were placed in Giemsa solution for 12 minutes). Ten cells were scanned from each specimen.

Spectral imaging

Cytological specimens stained by MGG were analyzed using the spectral imaging system, SpectraCube SD-200 (Applied Spectral Imaging, Migdal HaEmek, Israel) (Garini et al., 1996) attached to a microscope (Olympus AX, 100x/1.3 NA oil immersion condenser). The SpectraCube system consists of an interferometer situated in the parallel beam between an objective lens (infinity corrected) and a lens equivalent to an eyepiece, whose purpose is to form an image on a CCD camera. The light beam passing through the specimen is split in the interferometer in opposite directions, and is united again at the exit with an optical path difference (OPD) which is a function of the angle between the incoming beam and the interferometer itself. The OPD arises because for non-zero angles the two beams undergo different optical paths in the beamsplitter. The inherent mechanical stability of this interferometer allows a successful application of Fourier analysis to the visible spectral region. The measurement is done by recording successive CCD frames in synchronization with the steps of the motor used to rotate the collimated beam, so that the instantaneous OPD is known for every pixel in every recorded frame and can be used in the FFT calculation (Malik et al., 1996). During a measurement (20 seconds), each pixel of the CCD (512x512) is recording the interfered light from the interferometer, which is then Fourier transformed to give the spectrum (A detailed description of the optical system is described in Malik et al., 1996). Each pixel thus represents one of several tens of thousands of microspectrometers, acting simultaneously and independently. Spectral imaging acquires a so-called transmitted-spectra ‘cube’, with the two spatial dimensions of a flat sample (x and y), and the third, spectrum dimension, representing light intensity at any wavelength. The calculated pixel size in a spectral image is 0.04 µm². The spectral resolution (full width at half maximum) is 12 nm at 600 nm and the spectral range (more than 5% response) is 400-1000 nm (Garini et al., 1996).

Spectral similarity mapping

A given sample may be comprised of a number of
Spatially separated components, each characterized by a known and unique spectrum. In spectral mapping the task is to detect and spatially map the components per sample. The mapping algorithm used in this work consisted of the following four steps: first, approximately 10 × 10^7 spectra from the 400 cells were analyzed (10 cells from each of the 40 individuals, 512 × 512 spectra measured for each cell). The spectra were clustered into 17 visually distinct spectral groups. A representative spectrum from each group was stored in a "spectral library"; second, for every pixel of the input spectral cube, a comparison was made between its measured spectrum and all the reference spectra of the library; third, each pixel in the input image (corresponding to the spectral cube pixel), was identified with the most similar spectrum in the library; fourth, each pixel was displayed pseudo-colored according to the selected library spectrum.

Normalization steps were needed in order to uniformly treat individual samples and to preserve a consistent comparison with the prelabeled spectral library. Variation in spectra may result from differences in intensity. For example, the spectra of two identical sites in the same erythrocyte may appear different as a result of different concentrations of hemoglobin. In order to eliminate intensity variations, scanning was performed using normalized spectra (0-1 range). Standardization is needed to overcome differences in spectra that result from the fact that the specimens are taken from different patients and are stained on distinct days and by a different batch of stain. In order to eliminate the variance in dye concentration between different stained specimens, the spectra of all cubes were standardized according to the spectrum of a standard stained erythrocyte.

In the spectral mapping algorithm, the spectrum of every pixel of the input spectral cube is compared with the spectral-library spectra (second step). Let \( n \) be the number of spectra in the library. Define \( n \) functions, \( f^{n}_{x,y} \), for every pixel of spatial coordinates \( x \) and \( y \), as follows:

\[
f^{n}_{x,y} = \left( \int_{\lambda_1}^{\lambda_2} [I_{x,y}(\lambda) - I_{x,y}(\lambda)]^{2} \, d\lambda \right)^{1/2}
\] (1)

where the integral over \( \lambda \) stands for an integral over a predetermined spectral range \( \lambda_1 - \lambda_2 \). \( I_{x,y}(\lambda) \) is the spectrum of the pixel in question, and \( I_{x,y}(\lambda) \) are the \( n \) spectrum of the library. The minimal value for Equation (1) indicates the selected library spectrum per pixel (along with the corresponding pseudo-color for that pixel). The resulting pseudo-colored image reveals areas with similar chromophore compositions.

Quantitative measures assessing region characteristics and region morphology can now be extracted, in order to provide objective measures for cell characterization.

Computational analysis

Automatic computational analysis starts with preprocessing of the input image, to extract a region-of-interest from the pseudo-colored image. In a follow-up feature extraction stage, computational descriptors of the nuclear content can be computed. These include region descriptors, such as color and texture, and boundary descriptors, including nuclear area and contour.

Preprocessing

The spectral similarity mapping procedure provides an image that contains a continuous set of colors around the discrete set selected for the spectral library. The preprocessing phase in this work entails an initial stage of discretization of the image color content. The image is automatically segmented into the discrete set of pseudo-colors (from the spectral library) by affiliating each image pixel to the library index closest to the pixel color value. In the color-discretization process, a transition is made from the RGB color space to the HSV color space, in which color distances can be measured more accurately (Jain, 1989). The overall procedure was made computationally efficient by generating a look-up table, as follows: 1). The RGB space was quantized by multiples of 8 in each dimension, to produce a 32 × 32 table. 2). Each RGB table entry was converted into the respective HSV value, and related to an index, representing the pseudo-color chosen for the corresponding spectra. 3). The pseudo-color index can now be extracted immediately from the RGB table entry, calculated as follows:

\[
\text{index} = \text{fix}\left(\frac{R}{8}\right) + \text{fix}\left(\frac{G}{8}\right) \times 32 + \text{fix}\left(\frac{B}{8}\right) \times 32 \quad (2)
\]

As a second phase of the preprocessing, the input image is segmented into two main regions: a nuclear region (the region-of-interest) and a background region. The masking out of the nuclear region from within a given image was accomplished by defining the first two spectral indices (representing "black" and "gray") as background. Pixels affiliated with these spectra are labeled as background pixels while all other pixels are labeled nuclear pixels. A binary mask was generated, with black representing the background pixels and white representing the nuclear region. Such a mask is helpful for extracting nuclear shape and boundary characteristics.

Feature extraction

Following the segmentation process, color and texture descriptors were automatically extracted to represent the nuclear region characteristics.

Color statistics

The color statistics of the nuclear region were
summarized in a histogram, with each histogram entry defined as:

\[ p(x) = \frac{\text{total number of pixels with color level } x}{\text{total number of pixels in the region}} \quad (3) \]

The histogram was computed from pixels within the masked nuclear region. Each histogram contains as many bins as extracted via the spectral analysis and present in the extracted spectral library.

**Texture statistics**

Textural features provide computational measures for the description of the spatial distribution of the color regions within the nucleus. Each color (representing a certain nuclear composite) is analyzed separately, to determine if present as a single uniform region within the nucleus or if spread out within the nuclear region.

A new measure, the *Edge density measure*, is introduced, to define the spatial distribution of color. The edge density measure is computed as the ratio of the number of edges per color, to the total number of edges computed across all colors for the given nucleus. An edge map of the image is generated (Jain, 1989). For each color desired, a 2D binary mask is formed, with "1" for all pixels of the selected color, "0" for all other pixels in the image. The color mask is convolved with the edge map to extract the set of edges within the selected color region. The number of pixels that are marked as edge pixels are counted, per color k. Formally, the edge density for color k is computed as follows:

\[ ED_k = \frac{\text{Number of edges of color } k}{\text{Nucleus area}} \]

\[ \text{TotalED} = ED_1 + ED_2 + \ldots + ED_{17} \]

\[ e_k = \frac{\text{Number of edges of color } k}{\sum_i \text{Number of edges of color } i} = \frac{ED_k}{\text{TotalED}} \quad (5) \]

where, \( e_k \) is the normalized edge density measure per color. The respective ratios for each color are tabulated in an Edge-density histogram.

**Shape descriptors**

Several boundary descriptors may be used to characterize a nucleus. The area of a region is defined as the number of pixels contained within its boundary. The perimeter of a region is the length of the boundary. Although area and perimeter are sometimes used as descriptors, they apply primarily to situations in which the size of the objects of interest is invariant. A more robust measure, combining the two, is the compactness or circularity measure, defined as:

\[ \frac{\text{(perimeter)}^2}{\text{area}} \]

Compactness is a dimensionless quantity (insensitive to scale and orientation). It is minimal for a disk-shaped region.

**Results**

Spectral imaging was performed on MGG-stained blood specimens from thirty distinct patients with infectious mononucleosis, non-Hodgkin lymphoma or chronic lymphocytic leukemia, and from ten healthy individuals. A sample cell from each class is shown in Fig. 1. The spectra of all cells in the database were clustered into seventeen visually distinct spectral groups as represented by the spectral-library, shown in Fig. 2. A distinct pseudo color was arbitrarily assigned to each one of the spectra.

Spectral mapping was conducted on all the cells in the database, using the above-defined spectral-library.

---

**Fig. 1.** Spectral imaging of MGG-stained blood specimens from ten healthy individuals (A), and from thirty distinct patients with: chronic lymphocytic leukemia (B), infectious mononucleosis (C), and non-Hodgkin lymphoma (D). Each class is represented by a single cell. Bar: 10 µm.

**Fig. 2.** Spectral library. Seventeen distinct spectra sampled from the analyzed cells. Each spectrum is designated by a unique color.
Several examples of pseudo-colored spectral mapped images, from each cell class, are shown in Fig. 3. Each colored region corresponds to one spectrum from the spectral library. The collection of colored regions per nucleus, along with their spatial layout within the nucleus, represent distinct spectral regions in each nuclear compartment and their spatial order.

A computational analysis was performed on the pseudo-colored spectral maps. The preprocessing stage included the binarization of the image into nucleus and background regions and the extraction of the cell boundary from the binary image. Following the segmentation process, numerical features were extracted from each cell region. The pixels within the distinct nuclear regions were used to generate histograms that represent the color content and the edge-density content of the region. Examples of edge-density histograms are shown in Fig. 4. A statistical analysis of the nuclear color content is shown in Fig. 5. From the presented statistical results across the 4 cell categories, we note that each class of cells is associated with a unique spectral pattern, in both the color statistics as well as the edge-density statistics. Looking at the edge-density histograms (Fig. 4) we note that most of the edge energy in normal lymphocytes is present at the spectra designated as 6 and 7. In CLL cells the dominant edge-density spectra were 12 and 15. The lymphocytes from

![Fig. 3](image1.png)  
**Fig. 3.** Examples of spectrally classified cells from each of the 4 categories of interest in this work: (A) Ly: lymphocytes from healthy individuals, (B) CLL: chronic lymphocytic leukemia, (C) IM: infectious mononucleosis, and (D) NHL: non-Hodgkin’s lymphoma. Images are shown pseudo-colored with the 17 colors comprising the spectral library. The same magnification was applied as in Fig. 1.

![Fig. 4](image2.png)  
**Fig. 4.** Statistical analysis of edginess. Shown are edge-density histograms per cell category: (A) lymphocytes from healthy individuals, (B) CLL cells, (C) Infectious mononucleosis and (D) non-Hodgkin’s lymphoma. Bins (x-axis) allocated per spectrum (1-17).
infectious mononucleosis patients displayed a pattern in which the spectra designated as 12 and 13 had most of the edge energy; while in non-Hodgkin’s lymphoma the spectra designated as 4 and 16 were most common. An inspection of the color histograms (Fig. 5) reveals that the dominant spectra in normal lymphocytes were the spectra designated as 3, 4, 6 and 10, while in CLL cells the dominant spectra were 4, 12 and 15. The lymphocytes from infectious mononucleosis patients displayed a pattern in which the spectra designated as 4, 12, 13 and 16 were dominant; in non-Hodgkin’s lymphoma the spectra designated as 4, 5 and 16 were most common.

Area and boundary measures were computed over sets of 10 cells per category as shown in Table 1. The area of normal lymphocytes and CLL appeared to be approximately one third of the area of non-Hodgkin’s lymphoma and infectious mononucleosis cells. In addition, the non-Hodgkin’s lymphoma cells displayed high perimeter and circularity values.

Discussion

Our preliminary results show that spectral data combined with feature analysis provide significant information regarding the analyzed specimen. In the analysis of lymphoproliferative disorders, our system revealed specific spectral and morphological patterns associated with each particular disease. Normal lymphocytes and CLL appear to both have a relatively small nucleus with low circularity value (disk shaped). However, the color and edge density histograms reveal a distinct pattern for the two types of cells. Although, both infectious mononucleosis and non-Hodgkin’s lymphoma display large nuclei, the circularity value of the non-Hodgkin’s lymphoma cells is much higher (most non-disk like). Overall, the five nuclear descriptors: color, edge, area, perimeter and circularity, computed for spectrally analyzed cells, provide a unique set of characteristic values for each type of cells.

The goal of the advanced computational analysis is twofold. First, the numerical descriptors of the cell content are automatically extracted and enable objective evaluation of the cell characteristics. Second, the set of computational descriptors provides for objective measures of comparison between cells and cell count.
categories. The extracted descriptor set may be composed into feature vectors and distance metrics may be defined in feature space, to allow for an overall comparison measure between cells. The system presented is a step forward towards example-based content retrieval (Belongie et al., 1998; Wong et al., 1997), a growing field in which computational tools are developed enabling the professional user to provide a cell example as input to the system and to request the retrieval of similar content cells from archives available to the system. The development of such systems will provide an additional tool for histopathological evaluation of diseases.

Acknowledgements. We gratefully thank Ms Judith Hanania for her help in editing the manuscript. Hayit Greenspan’s work was supported by the Eshkol Grant, of the Israeli Ministry of Science. The work was supported by the Ela Kodesz Institute for Medical Engineering and Physical Sciences, Tel-Aviv University.

References


Accepted March 8, 2002