

# Image-guided decision support system for pathology

Dorin Comaniciu<sup>1</sup>, Peter Meer<sup>2</sup>, David J. Foran<sup>3</sup>

<sup>1</sup> Imaging Department, Siemens Corporate Research, Princeton, NJ 08540, USA

<sup>2</sup> ECE Department, Rutgers University, Piscataway, NJ 08854, USA; e-mail: meer@caip.rutgers.edu

<sup>3</sup> Pathology Department, UMDNJ-RWJ, Piscataway, NJ 08854, USA

Received: 20 November 1998 / Accepted: 16 August 1999

**Abstract.** We present a content-based image retrieval system that supports decision making in clinical pathology. The image-guided decision support system locates, retrieves, and displays cases which exhibit morphological profiles consistent to the case in question. It uses an image database containing 261 digitized specimens which belong to three classes of lymphoproliferative disorders and a class of healthy leukocytes. The reliability of the central module, the fast color segmenter, makes possible unsupervised on-line analysis of the query image and extraction of the features of interest: shape, area, and texture of the nucleus. The nuclear shape is characterized through similarity invariant Fourier descriptors, while the texture analysis is based on a multiresolution simultaneous autoregressive model. The system performance was assessed through ten-fold cross-validated classification and compared with that of a human expert. To facilitate a natural man-machine interface, speech recognition and voice feedback are integrated. Client-server communication is multithreaded, Internet-based, and provides access to supporting clinical records and video databases.

**Key words:** Content-based retrieval – Decision support system – Color segmentation – Information fusion – User interfaces

## 1 Introduction

One of the most common procedures when patients are admitted into a hospital is the drawing of blood. The resulting blood samples are routinely scanned by an automated complete blood count (CBC) device. If the CBC flags a specimen as suspicious because of an abnormal blood cell count for example, a blood smear specimen is forwarded to a medical technologist for review. Based upon a subjective, morphologic inspection of the constituent cells, a determination is made whether the cells are benign.

While the nuclear morphologic attributes can theoretically lead to a diagnosis, in practice, the clinical interpreta-

tion is often ambiguous. The subtle visual differences exhibited by some malignant lymphomas and chronic lymphocytic leukemia give rise to a significant number of false negatives (malignant cells classified as normal). If suspicious cells are detected, subsequent morphological evaluation of specimens by even experienced pathologists is often inconclusive. In these cases, differential diagnosis can only be made after supporting tests such as immunophenotyping by flow cytometry.

The classification of mantle cell lymphoma (MCL), a recently described disorder [2, 4], is of particular interest among the indolent lymphomas. MCL is often misdiagnosed as chronic lymphocytic leukemia (CLL) or follicular center cell lymphoma (FCC) [5]. In addition, the survival of patients with MCL is much shorter than that of patients with other low-grade lymphomas, and standard therapy for CLL and FCC is ineffective with MCL. Therefore, timely and accurate diagnosis of MCL has significant therapeutic and prognostic implications. Extensive medical research is underway to understand the significance of the differences in the immunology and cell biology of these entities.

This paper describes an *image-guided decision support* (IGDS) system designed to assist pathologists to discriminate among malignant lymphomas and chronic lymphocytic leukemia directly from microscopic specimens. The task of the system is to locate, retrieve and display cases with morphological profiles consistent to the case in question, and to suggest during each retrieval the most likely diagnosis based on majority logic. The ground truth of the cases recorded in the database is obtained a priori through immunophenotyping and is used to maximize the probability of correct classification.

Recent literature in diagnostic hematopathology ascribes much of the difficulty in rendering consistent diagnoses to subjective impressions of observers and shows that when morphologic cell classification is based upon computer-aided analysis, the level of objectivity and reproducibility improves [3]. However, only recently the potential of content-based retrieval systems for medical applications was recognized [31].

A survey of the state of the art of content-based image retrieval is presented in [1]. Technologies that cap-

ture, describe, and index the content of multimedia objects rely on methods from image analysis, pattern recognition, and database theory. A new family of information retrieval systems emerged, exploiting the richness of visual information and covering a large spectrum of applications [11, 13, 20, 23, 24]. These systems differ according to their degree of generality (general purpose versus domain specific), level of feature abstraction (primitive features versus logical features), overall dissimilarity measure used in retrieval ranking, database indexing procedure, level of user intervention (with or without relevance feedback), and evaluation methodology.

In contrast to general-purpose retrieval engines which use a subjective notion of similarity, the IGDS system is applied to a very specific problem. It has a high degree of content understanding [31] since elemental structures from the input image are localized and delineated (e.g., the leucocyte nucleus and cytoplasm). Being a system with a well-defined goal, its performance can be quantitatively evaluated and compared to the human expert results. The reason of this comparison, however, is only to assess the usefulness of the system. In a real analysis scenario, a lot of context information difficult to quantize is taken into account for the diagnosis and no technique can ever replace the pathologist and light microscopy. The IGDS system is designed as a tool to help physicians and technicians during routine screening and analysis. It is not intended to be an automatic cell classifier.

The user starts a typical IGDS session by loading the query image and selecting a rectangular region of interest (ROI) that contains cells which are either unidentifiable or are known to be key to the diagnosis. The elemental structures from the ROI (e.g., cell nuclei and cytoplasm areas) are then automatically delineated based on a fast color segmentation algorithm. By choosing a cell nucleus, the user initiates first the analysis of the nucleus attributes (shape, texture, area, and color), then the search in a remote database of digitized specimens. As a response, the system retrieves and displays the images from the database that are the closest matches to the query. The user can interactively review the retrieved cases including the associated clinical records and video clips.

Since the digitized specimens may have different statistical properties, the robustness of the color segmenter is of paramount importance. The employed segmenter is based on a recently proposed nonparametric cluster analysis technique which can process 10 000 color vectors in fractions of a second [7].

To access the database of digitized specimens, four visual attributes of the delineated cell nucleus are defined: shape, texture, area, and color. Medical literature frequently uses the first three of the above attributes to morphologically describe the appearance of malignant cells [2]. We characterize the nuclear shape through similarity-invariant Fourier descriptors [17]. Fourier invariants were recently proved to be superior to methods based on autoregressive models [15]. The uncertainty introduced by the segmentation process is taken into account to determine the number of harmonics which reliably represent the shape. The texture analysis is based on a multiresolution simultaneous autoregressive model (MRSAR) [21]. A 15-dimensional feature

vector and its covariance matrix are derived for each nucleus in the database.

The overall dissimilarity metric between two nuclei is defined as a linear combination of the normalized distances corresponding to each visual attribute. The weights are obtained off line by optimizing the probability of correct classification over the entire database. We found this metric to provide better results than the joint rank criterion expressed as the weighted sum of individual ranks.

The user interface of a retrieval system should provide information exchange in ways familiar and comfortable to the human [30]. A distinct feature of the IGDS system is its bimodal human-computer interaction. Queries can be formulated and refined and the retrievals can be browsed using speech recognition or graphical input. Audio feedback is provided by speech synthesis.

The paper is organized as follows. Experimental details regarding the database of ground truth cases are given in Sect. 2. Section 3 presents the architecture of the IGDS system. In Sect. 4 the segmentation algorithm is discussed. The shape descriptors are presented in Sect. 5. Section 6 describes the implementation of the texture analysis and also discusses the color attribute. Section 7 shows how the overall dissimilarity measure is defined and describes its optimization. The retrieval performance assessment of the system by cross-validation and comparison to human expert performance on the same database are presented in Sect. 8. The user interface of the system is outlined in Sect. 9.

## 2 Database of ground truth cases

To populate the current database, peripheral blood smears from 30 lymphoproliferative cases at Robert Wood Johnson University Hospital were air-dried, fixed with methanol, and stained with Wright Giemsa solution. One at a time, the stained specimens were examined by a certified hematopathologist using a Leica microscope, 40x/0.65 planachromatic objective while lymphoid cells and benign lymphocytes were identified and digitized. The imaging components of the system consist of an Intel-based workstation, high-resolution color video camera (Olympus OLY-750) and data acquisition board (Coreco Oculus). Blood specimens were immunophenotyped using a Coulter XL in order to independently confirm the classification of each of the cells into one of four categories: MCL, CLL, FCC, or normal (benign).

Immunophenotyping is the characterization of white blood cells by determining the cell surface antigens they bear. The cells are isolated and incubated with fluorescently tagged antibodies directed against specific cell surface antigens. Then, they pass through the flow cytometer past a laser beam. When the cells meet the laser beam, they emit fluorescent signals in proportion to the amount of the specific cell surface antigen they have, and a computer calculates the percentage of cells expressing each antigen.

The current image database consists of 66 MCL, 98 CLL, 38 FCC, and 59 benign cells. A typical image has about  $450 \times 350$  true-color pixels. Individuals using the IGDS system would need to either standardize the image acquisition according to the above procedure or use the IGDS to generate their own "gold standard" database.

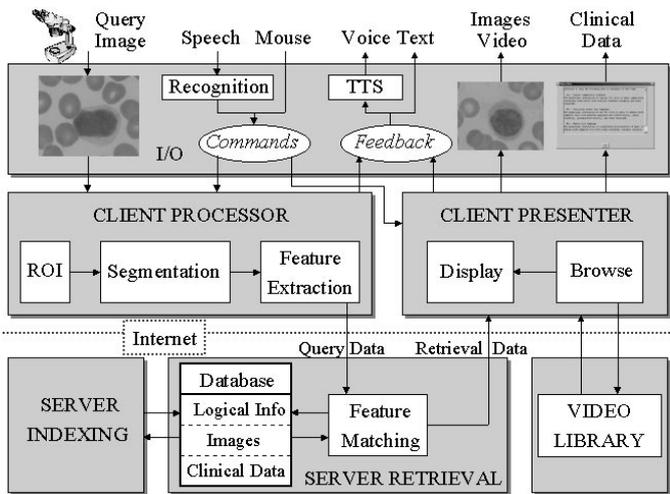


Fig. 1. Architecture of the IGDS system

### 3 System architecture

The decision support system has a client-server platform-independent architecture implemented in Java (Fig. 1). The client part is intended to be used in small hospitals and laboratories to access through the Internet the database at the server site.

The client I/O module loads the query image from a local or remote microscope [8] and saves the retrieved information. A fusion agent capable of multimodal inputs interprets the speech or graphical commands, calls the appropriate method, and gives voice feedback to the user based on a TTS (text-to-speech) component.

The client processor contains the query formation tools, performing the user-guided ROI selection, color segmentation and feature extraction. The query vector is submitted to the server through a serializable object. The serialization mechanism of Java provides an automatic framework for the transport of object collections from one machine to another. Based on the retrieved data, the client presenter communicates the suggested classification to the user, and allows the browsing of cases of interest, including their associated clinical records and video clips.

The IGDS server is composed of two parts: the retrieval and indexing modules. The retrieval process is multi-threaded, simultaneous accesses to the database being authorized. During feature matching, the query data and logical information in the database are compared to derive a ranking of the retrievals.

The database indexing is performed off line. A module similar to the client processor is used for the analysis and registration of the incoming cases, with ground truth established through immunophenotyping. Then, the weights of the dissimilarity measure are re-learned to account for the new entries in the database.

### 4 Segmentation algorithm

The IGDS segmenter is based on nonparametric analysis of the  $L^*u^*v^*$  color vectors obtained from the input image. The algorithm detects color clusters and delineates their borders

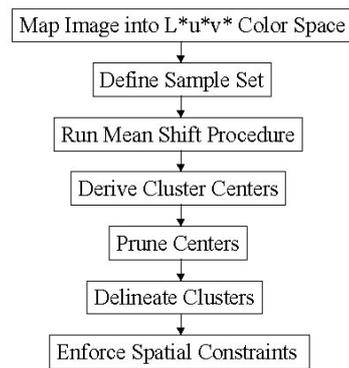


Fig. 2. Flowchart of the segmentation process

based on the gradient ascent *mean shift* procedure. It randomly tessellates the space with search windows and moves the windows till convergence at the nearest mode of the underlying probability distribution. The nonparametric, robust nature of the color histogram analysis allows accurate and stable recovery of the main homogeneous regions in the image. A short description of the segmentation is given below and is illustrated by the flowchart in Fig. 2. For more details, including proof of convergence of the mean shift procedure, see [7].

First, the RGB input vectors are converted into  $L^*u^*v^*$  vectors following a nonlinear transformation. The rationale of using the  $L^*u^*v^*$  color space is that perceived color differences in this space are measured by Euclidean distances [32, Sect. 3.3.9].

A set of  $m$  points  $\mathbf{x}_1 \dots \mathbf{x}_m$  called the *sample set* is then randomly selected from the data. Distance and density constraints are imposed on the points retained in the sample set, automatically fixing its cardinality. The distance between any two neighbors should not be smaller than  $h$ , the radius of a searching sphere  $S_h(\mathbf{x})$ , and the sample points should not lie in sparsely populated regions. A region is sparsely populated whenever the number of points inside the sphere is below a threshold  $T_1$ .

Next, the mean shift procedure is applied to each point in the sample set. The mean shift vector at the point  $\mathbf{x}$  is defined as [14, p. 534]

$$M_h(\mathbf{x}) = \frac{1}{n_{\mathbf{x}}} \sum_{\mathbf{x}_i \in S_h(\mathbf{x})} \mathbf{x}_i - \mathbf{x}, \quad (1)$$

where  $n_{\mathbf{x}}$  is the number of data points contained in the searching sphere  $S_h(\mathbf{x})$ . It can be shown that the vector (Eq. 1) has the direction of the gradient density estimate when this estimate is obtained with the Epanechnikov kernel [7]

$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2) (1 - \mathbf{x}^T \mathbf{x}) & \text{if } \mathbf{x}^T \mathbf{x} < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$c_d$  being the volume of the unit sphere in the  $d$ -dimensional space. Pointing towards the direction of maximum increase in the density, recursive computation of the mean shift vector defines a path leading to the nearest mode of the density.

The  $m$  points of convergence resulted by applying the mean shift to each point in the sample set are called *cluster center candidates*. Since a local plateau in the color space

can prematurely stop the mean shift iterations, each cluster center candidate is perturbed by a random vector of small norm and the mean shift procedure is let to converge again. The computation of the mean shift vectors is based on the entire data set, therefore, the quality of the density gradient estimate is not diminished by the use of sampling.

The candidates are then pruned to obtain the cluster centers  $\mathbf{y}_1 \dots \mathbf{y}_p$ , with  $p \leq m$ . Any subset of cluster center candidates which are sufficiently close to each other (for any given point in the subset, there is at least another point in the subset at a distance less than  $h$ ), defines a *cluster center*. The cluster center is the mean of the candidates in the subset.

The presence of a valley between each pair  $(\mathbf{y}_i, \mathbf{y}_j)$  of cluster centers is tested next (see [7] for the testing procedure). If no valley was found, the cluster center of lower density ( $\mathbf{y}_i$  or  $\mathbf{y}_j$ ) is removed from the set of cluster centers.

Cluster delineation has two stages. First, each sample point is allocated to a cluster center based on the history of its initial window. Then, each data point is classified according to the majority of its  $k$ -nearest sample points.

Finally, spatial constraints are enforced to validate homogeneous regions in the image. Small connected components containing less than  $T_2$  pixels are removed, and region growing is performed to allocate the unclassified pixels.

Three parameters control the segmentation: the searching sphere radius  $h$ , the threshold  $T_1$  which imposes the density constraint, and the threshold  $T_2$  which determines the minimum connected component size. With the default parameter values  $h = 4$ ,  $T_1 = 50$ , and  $T_2 = 1000$ , the system is working satisfactorily for most images in the database. The default settings are also used for the segmentation of all images presented in this paper. Examples of nucleus delineation performed by the segmentation module are given in Fig. 3.

## 5 Similarity-invariant shape descriptors

The analysis of the shape of the delineated nucleus is based on Fourier descriptors which are made invariant to changes in location, orientation and scale, that is, *similarity invariant*. Several representations are possible using an arc length  $s$  parametrization of the chain-encoded contour. The cumulative angular function  $\theta(s)$ , the centroidal distance  $R(s)$ , and the complex function of the coordinates  $u(s) = x(s) + jy(s)$  are examples of such representations.

We followed the approach of Kuhl and Giardina [17] and expanded the functions  $x(s)$  and  $y(s)$  separately to obtain the *elliptic Fourier descriptors* (EFD). The EFDs corresponding to the  $n$ th harmonic of a contour composed of  $K$  points are given by

$$\begin{aligned} a_n &= \frac{S}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta x_i}{\Delta s_i} \left[ \cos \frac{2n\pi s_i}{S} - \cos \frac{2n\pi s_{i-1}}{S} \right], \\ b_n &= \frac{S}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta x_i}{\Delta s_i} \left[ \sin \frac{2n\pi s_i}{S} - \sin \frac{2n\pi s_{i-1}}{S} \right], \\ c_n &= \frac{S}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta y_i}{\Delta s_i} \left[ \cos \frac{2n\pi s_i}{S} - \cos \frac{2n\pi s_{i-1}}{S} \right], \end{aligned}$$

$$d_n = \frac{S}{2n^2\pi^2} \sum_{i=1}^K \frac{\Delta y_i}{\Delta s_i} \left[ \sin \frac{2n\pi s_i}{S} - \sin \frac{2n\pi s_{i-1}}{S} \right], \quad (3)$$

where

$$s_i = \sum_{j=1}^i \Delta s_j, \quad S = \sum_{i=1}^K \Delta s_i, \quad (4)$$

$$\Delta s_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}, \quad (5)$$

$$\Delta x_i = (x_i - x_{i-1}), \quad \Delta y_i = (y_i - y_{i-1}), \quad (6)$$

$\Delta x_i$  and  $\Delta y_i$  representing the changes in the  $x$  and  $y$  projections of the chain code as the  $i$ th contour point is traversed.

By contrast to the use of the cumulative angular function where the truncation of the Fourier series can yield open curves, the curve reconstructed from the EFDs is always closed [25]. The EFDs have a straightforward geometric interpretation, the closed contour being represented as a composition in proper phase relationship of ellipses, called harmonic loci. The larger the number of ellipses involved, the more accurate the representation becomes.

Rotation invariance is obtained by compensating for the arbitrary position of the starting point on the contour and for the arbitrary orientation of the contour. Hence, two rotations are necessary to achieve the invariance. When the first harmonic locus is an ellipse, the rotations are defined relative to the semi-major axis of the locus, and produce two related representations of the curve

$$\begin{aligned} \begin{bmatrix} a_n^{(1)} & b_n^{(1)} \\ c_n^{(1)} & d_n^{(1)} \end{bmatrix} &= \begin{bmatrix} \cos \psi_1 & \sin \psi_1 \\ \sin \psi_1 & \cos \psi_1 \end{bmatrix} \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \\ &\times \begin{bmatrix} \cos n\theta_1 & -\sin n\theta_1 \\ \sin n\theta_1 & \cos n\theta_1 \end{bmatrix} \end{aligned} \quad (7)$$

and

$$\begin{bmatrix} a_n^{(2)} & b_n^{(2)} \\ c_n^{(2)} & d_n^{(2)} \end{bmatrix} = (-1)^{n+1} \begin{bmatrix} a_n^{(1)} & b_n^{(1)} \\ c_n^{(1)} & d_n^{(1)} \end{bmatrix}. \quad (8)$$

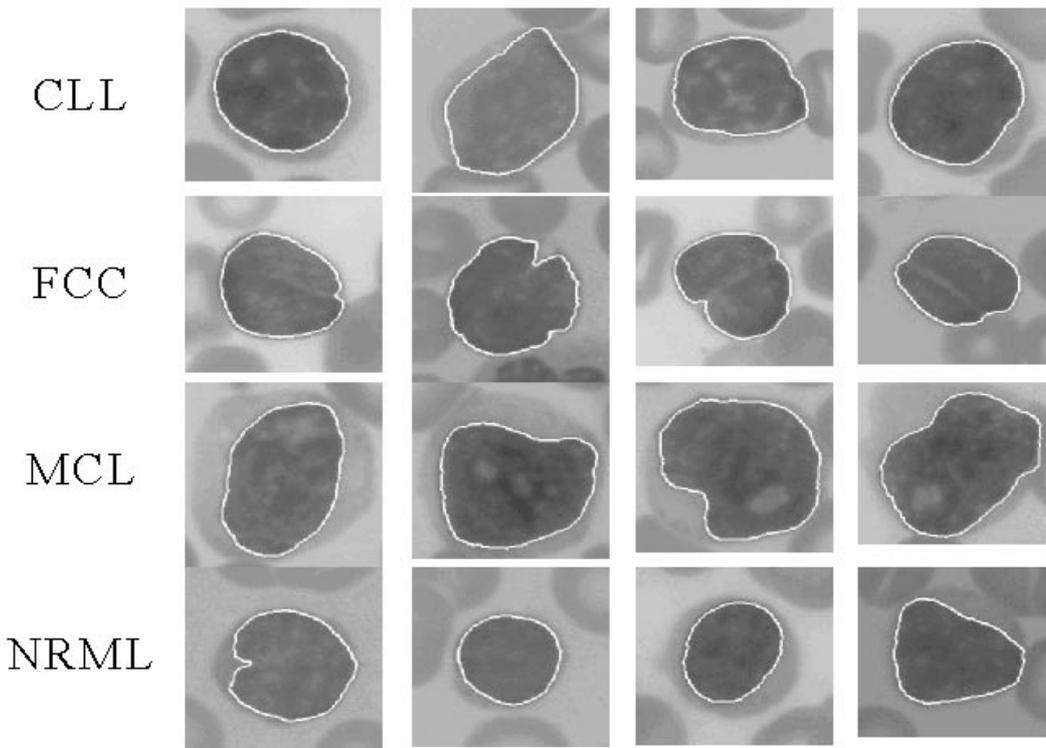
Expressions for the axial rotation  $\psi_1$  and starting point displacement  $\theta_1$  relative to the first semi-major axis are derived in [17].

If the first harmonic locus is circular, the rotations are made with respect to the line defined by the centroid of the contour and the point on the contour most distant from the centroid. Since the most distant point can be nonunique,  $k$  related representations can result, corresponding to  $k$  sets of Fourier coefficients

$$\begin{aligned} \begin{bmatrix} a_n^{(p)} & b_n^{(p)} \\ c_n^{(p)} & d_n^{(p)} \end{bmatrix} &= \begin{bmatrix} \cos \psi_p & \sin \psi_p \\ \sin \psi_p & \cos \psi_p \end{bmatrix} \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \\ &\times \begin{bmatrix} \cos n\theta_p & -\sin n\theta_p \\ \sin n\theta_p & \cos n\theta_p \end{bmatrix}, \end{aligned} \quad (9)$$

with  $p = 1, \dots, k$ , where the axial rotation  $\psi_p$  and starting point displacement  $\theta_p$  are defined relative to the  $p$ th most distant point.

Other Fourier-based methods for shape representation achieve rotation invariance by discarding the phase information of all coefficients [15, 20]. However, this can lead to misleading classifications, since the phase plays an important role in the contour representation. Note that a recently proposed set of Fourier invariants based on the discrete Fourier



**Fig. 3.** Segmentation of the nucleus of various cells. The cells in each row belong to the same class: CLL, FCC, MCL, and Normal, respectively

transform, also considers as reference the semi-major axis of the first harmonic locus to compensate for different rotations and starting points [29].

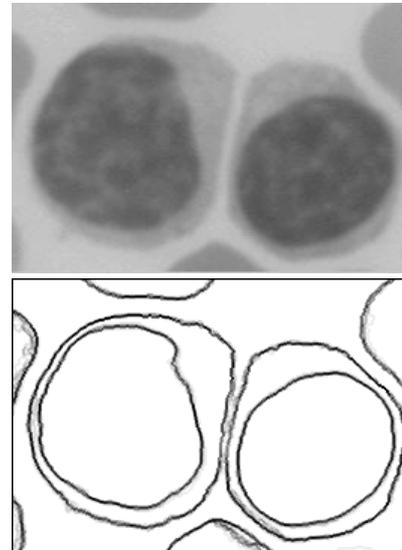
Scale invariance is obtained by normalizing each Fourier coefficient by the magnitude of the semi-major axis, when the first harmonic locus is elliptic, and by the magnitude of the radius, when the first harmonic locus is circular. To obtain translation invariance, the bias terms are removed from the Fourier series.

### 5.1 Accuracy of shape representation

The number of harmonics which reliably represent the shape of the nucleus is closely related to the uncertainty introduced by prior processing stages. The segmentation process is global and any change in the region of interest selected by the user may have effect on the nucleus delineation. Also, due to its probabilistic nature (i.e., random tessellation of the color space), the segmentation produces slightly different results when repeatedly applied to the same image.

Figure 4 shows at the bottom the result of superimposing the contours obtained by segmenting 25 times the image from the top. The darker a pixel in the contour image, the more stable is the contour passing through that pixel. One can see that the least stable regions are between the two cells. The stability of the delineated contour was shown to be a good measure of the confidence in segmentation [6].

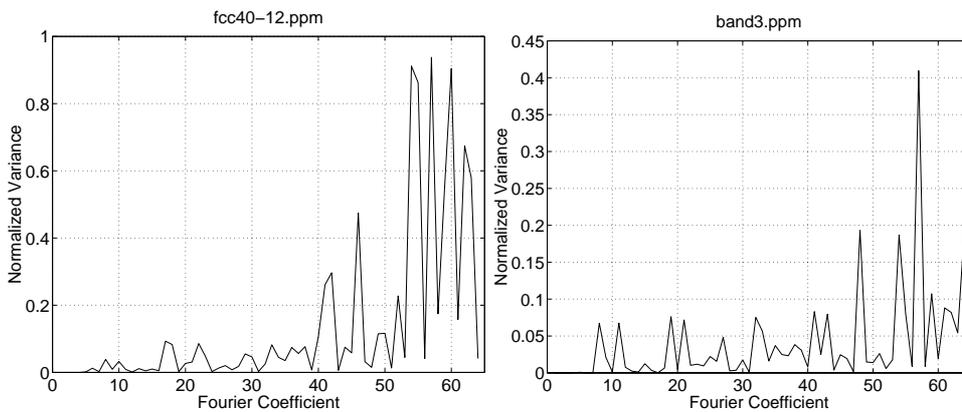
To estimate the influence of this uncertainty on the Fourier coefficients, experiments with several images were conducted and the normalized variance (variance over the squared mean) of each coefficient was computed. For a given image, a user delineated 25 times the ROI (a leukocyte). The region was then segmented and the first 16 harmonics (64 co-



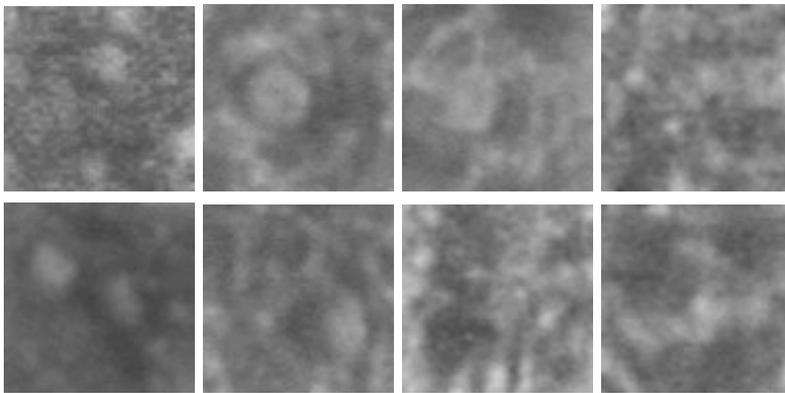
**Fig. 4.** Example of segmentation stability. A region of interest from the image (top) has been segmented 25 times and the resulting contours were superimposed (bottom). The regions between two cells are the least stable

efficients) were determined for the nucleus. Typical results, the normalized variances of the Fourier coefficients of the nucleus from two images are presented in Fig. 5. It can be concluded that the segmentation is sufficiently stable for the use of the first 10 harmonics (40 coefficients). Consequently, we compare a query contour with a reference contour in the database by computing the Euclidean distance between the corresponding 40-dimensional vectors of Fourier invariants

$$D_1 = \sqrt{(\mathbf{f}_{\text{query}} - \mathbf{f}_{\text{reference}})^T (\mathbf{f}_{\text{query}} - \mathbf{f}_{\text{reference}})} . \quad (10)$$



**Fig. 5.** Normalized variances of the first 64 Fourier coefficients corresponding to two images (see text for details)



**Fig. 6.** Examples of texture ( $80 \times 80$  pixels) from inside the nuclear border. The gray level dynamic range was enlarged to improve reproduction quality

## 6 Texture, area, and color metrics

Other attributes of interest are the texture of the nucleus, area of the nucleus, as well as the color associated with it by the segmentation process.

### 6.1 Texture

We describe the nuclear texture based on the MRSAR model [21] which assumes that the image data is randomly structured. For the current database, the MRSAR model provides an accurate description of the nuclear texture. The image data inside the nuclear border (Fig. 6) is indeed relatively unstructured, characterized by random patterns and with no presence of periodicity or directionality.

Note that being a representation of the chromatin density, the nuclear texture may present in certain clinical phases organized patterns. These cases would require a more complex treatment of the texture, such as the Wold-based texture model presented in [18]. According to the Wold formulation, an image (regarded as a homogeneous random field) is decomposed into mutually orthogonal subfields having perceptual properties that can be described as periodicity, directionality, and randomness.

The MRSAR model that we use characterizes only the random subfield of the Wold representation. It is a second-order noncausal model described by five parameters at each resolution level. A symmetric MRSAR is applied to the  $L^*$  component of the  $L^*u^*v^*$  image data. The pixel value  $L^*(\mathbf{x})$  at a certain location  $\mathbf{x}$  is assumed to linearly depend on the

neighboring pixel values  $L^*(\mathbf{y})$  and a zero-mean additive independent Gaussian noise term  $\epsilon(\mathbf{x})$ :

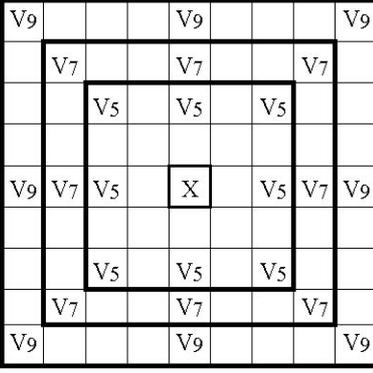
$$L^*(\mathbf{x}) = \mu + \sum_{\mathbf{y} \in \mathcal{Z}} \theta(\mathbf{y})L^*(\mathbf{y}) + \epsilon(\mathbf{x}) . \quad (11)$$

In Eq. 11,  $\mu$  is the bias dependent on the mean value of  $L^*$ ,  $\mathcal{Z}$  is the set of neighbors of pixel at location  $\mathbf{x}$ , and  $\theta(\mathbf{y})$  with  $\mathbf{y} \in \mathcal{Z}$  are the model parameters.

Figure 7 shows how are the neighbors defined for a window size of  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$ . The model being symmetric, we have  $\theta(\mathbf{y}) = \theta(-\mathbf{y})$ , hence, for a given neighborhood, four parameters are estimated through least squares. Thus, the model parameters and the estimation error define a 5-dimensional feature vector. The procedure is repeated for the three chosen window sizes and the vectors are concatenated.

In [18, 27], it was shown that the MRSAR features computed with  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$  neighborhoods provide the best overall retrieval performance over the entire Brodatz database. While the textures inside the nuclei are different from the Brodatz ones, the same neighborhoods are used here to form a 15-dimensional multiresolution feature vector.

To estimate the model parameters,  $21 \times 21$  overlapping windows moving every two pixels in both horizontal and vertical directions are used and for each window a multiresolution feature vector is obtained. The mean vector  $\mathbf{t}$  and the covariance matrix  $\Sigma$  over all windows inside a given cell nucleus are the MRSAR features associated with that nucleus.



**Fig. 7.** The set of neighbors  $\mathcal{N}$  used by the MRSAR model with window size  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$

Thus, the texture dissimilarity has to be measured by the distance between two multivariate distributions with known mean vectors and covariance matrices. We use the Mahalanobis distance between the MRSAR feature vectors to express this dissimilarity

$$D_2 = \sqrt{(\mathbf{t}_{\text{query}} - \mathbf{t}_{\text{reference}})^{\top} \Sigma_{\text{reference}}^{-1} (\mathbf{t}_{\text{query}} - \mathbf{t}_{\text{reference}})} \quad (12)$$

where  $\Sigma_{\text{reference}}^{-1}$  represents the inverse of the covariance matrix of  $\mathbf{t}_{\text{reference}}$ . For each entry in the database,  $\Sigma_{\text{reference}}^{-1}$  is obtained and stored off line for each indexed nucleus. Note that for the current database, the use of  $\Sigma_{\text{reference}}$  for Mahalanobis computation resulted in better classification than that obtained with  $\Sigma_{\text{query}}$ .

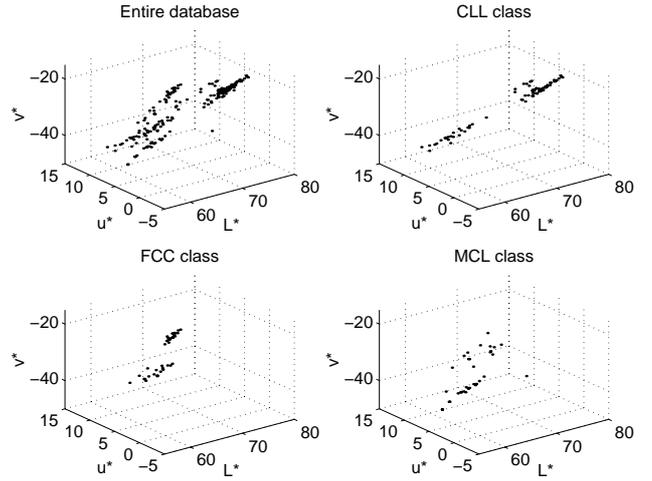
By taking into account not only the separation induced by different mean vectors but also the separation given by the difference in covariance matrices, the Bhattacharyya distance [14, p. 99] is theoretically better than the measure (Eq. 12). Its limited use for database search is due to the on-line matrix inversion required by the direct-distance computation. However, we recently showed [9] that the Bhattacharyya distance can be computed efficiently when most of the energy in the feature space is restricted to a low-dimensional subspace. The improved representation is to be implemented into the IGDS system.

By using the Mahalanobis distance, the assumption we make is that the covariance of the query  $\Sigma_{\text{query}}$  and the covariances of the references from the database  $\Sigma_{\text{reference}}$  are rather similar. For multivariate Gaussian data, the Mahalanobis distance (Eq. 12) becomes in this case equivalent to the Bhattacharyya distance.

## 6.2 Area

All the digitized specimens in the database having the same magnification, the nuclear area is computed as the number of pixels inside the delineated nucleus. The dissimilarity between two nuclei in terms of their areas is expressed as

$$D_3 = \sqrt{(a_{\text{query}} - a_{\text{reference}})^2} \quad (13)$$



**Fig. 8.** The color vectors of all nuclei in the database and those corresponding to the CLL, FCC, and MCL classes, respectively. Note that the color vectors are not clustered according to their class

## 6.3 Color

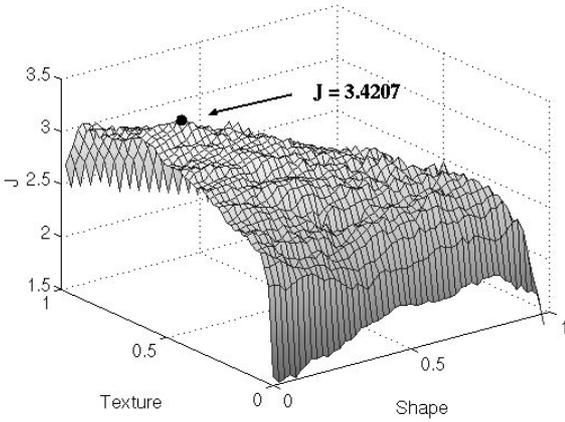
The nuclear color is specified as a 3D vector in the  $L^*u^*v^*$  space and is determined by the segmentation as the center of the associated color cluster. Since the colors of the nuclei in the database do not cluster as a function of the cell class (see Fig. 8), we handle color as a potential query attribute only. Therefore, the current implementation of the system uses the color attribute for nucleus separation from the background, but not for distinguishing among different cells.

## 7 Overall dissimilarity metric

The derivation of a dissimilarity metric which optimally combines the information carried by all the visual attributes of significance is a rather difficult problem [26]. A Bayesian solution is proposed in [19] where combinatorial search techniques are employed to minimize a classification metric, the cross-entropy. The method determines the most useful features for image classification, implicitly assuming that these features will perform well during retrieval. In [18], the individual ranks corresponding to each attribute derived from the query image are computed first, then the joint rank is obtained as the sum of the individual ranks weighted by posterior probabilities. The final similarity ordering of the database is formed by sorting images in the ascending order of their joint rank values. Various other information fusion techniques were proposed such as the intra- and inter-feature normalization [23], positive and negative relevance feedback [22], and PicHunter strategy [10] which employs a learned probabilistic model of human behavior to make better use of the feedback it obtains.

Recall that the current system uses 261 images corresponding to four cell categories whose ground truth was obtained through immunophenotyping. The suggested classification of the query image is based on the voting  $k$ NN rule [14, p. 305] among the classes of the closest  $k$  matches. That is,

$$k_i = \max\{k_1, \dots, k_4\} \rightarrow X \in \omega_i, \quad (14)$$



**Fig. 9.** Plot of the objective surface (resolution is 0.02 on each dimension). The downhill simplex converged in this case to the global maximum

where  $k_i$  is the number of neighbors from the class  $\omega_i$  ( $i = 1, \dots, 4$ ) among the  $k$ NNs, and  $k_1 + \dots + k_4 = k$ . In addition to the four original cell classes, the  $k$ NN rule may also produce a NO DECISION class, in the case when the value of  $i$  verifying (Eq. 14) is not unique.

The system performance is measured by the *confusion matrix*  $\mathbf{R}$  defined as having as element  $r_{j,i}$  the empirical probability of classification in class  $j$  when the query image belonged to class  $i$ ,  $P(j|i)$ . The criterion that should be maximized is the sum of conditional probabilities of correct decision

$$J = \sum_{j=1}^4 P(j|j) . \quad (15)$$

According to the Bayesian rule [16], an optimal decision is based on the ensemble statistics of all the extracted features. The shape, texture, and area, however, are visual attributes of different nature, being described respectively by a 40-dimensional vector, a 15-dimensional cluster, and a scalar. This heterogeneity of data makes it difficult to model its statistics. Hence, two suboptimal solutions for the dissimilarity metric are defined and tested below.

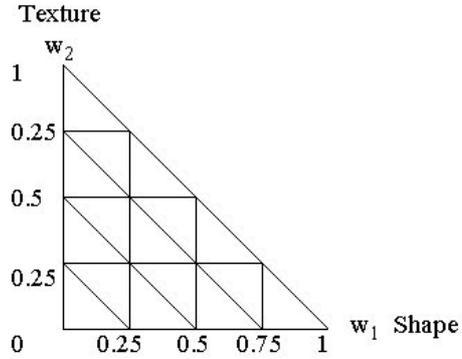
### 7.1 Weighted sum of distances

A relative simple solution is to express the dissimilarity as a linear combination of the distances corresponding to each query attribute. Thus, for three attributes,

$$D = \sum_{i=1}^3 w_i D_i , \quad (16)$$

where  $w_i$  represents the relevance of the  $i$ -th attribute and  $\sum_{i=1}^3 w_i = 1$ .

The best weights  $w_i$  are derived off line by employing the downhill simplex method [28, p. 408] with the objective function  $J$  (Eq. 15). A simplex in  $N$  dimensions consists of  $N + 1$  totally connected vertices. For example, in two dimensions, the simplex is a triangle, and in three dimensions, it is a tetrahedron. The optimization is based on a series of steps which reflect, expand, and contract the simplex such that it converges to a maximum of the objective function. As



**Fig. 10.** The 16 initial simplexes used for the initialization of the optimization procedure

**Table 1.** Best weights and the value of optimization criterion corresponding to the global maximum

Shape	Texture	Area	$J$
0.1140	0.5771	0.3089	3.4207

an advantage, the downhill simplex requires only function evaluations and not computation of derivatives.

To obtain the same order of magnitude for the individual distances in Eq. 16, they were normalized to the standard deviation calculated relative to the center of each class, excepting  $D_2$  which is a Mahalanobis distance and therefore has intrinsic normalization. The normalization ensures a numerically stable optimization procedure.

In Fig. 9, the objective surface as a function of the two independent weights is shown. Since the downhill simplex guarantees a local maximum only, we run the optimization 16 times with different initializations. A regular tessellation of the right-angled triangle defined by the values of  $w_1$  and  $w_2$  generated the 16 initial simplexes (Fig. 10). For convergence, about 40 iterations were needed for each trial. In Table 1, the best set of weights, obtained by running the optimization with seven retrievals over the entire database is shown. It corresponds to the highest obtained value ( $J = 3.4207$ ) of the objective function (Eq. 15).

### 7.2 Weighted sum of ranks

A second solution to derive the closest matches in the database is to compute first the individual ranks corresponding to each attribute and obtain the joint rank as a weighted sum of individual ranks:

$$O = \sum_{i=1}^3 w_i O_i . \quad (17)$$

Using the same tessellation as before, the downhill simplex was employed to find the best weights. This solution, however, yielded an objective function whose maximum ( $J = 3.3015$ ) is smaller than the one obtained by weighting distances. Hence, only the first solution will be considered further.

**Table 2.** Ten-fold cross-validated confusion matrix corresponding to the first seven retrievals

	CLL	FCC	MCL	NRML	NO DEC
CLL	<b>.8389</b>	.0200	.0711	<b>.0700</b>	.0000
FCC	.0250	<b>.9000</b>	.0000	<b>.0500</b>	.0250
MCL	.1357	.0143	<b>.8333</b>	<b>.0000</b>	.0167
NRML	.1333	.1200	.0000	<b>.7300</b>	.0167

**Table 3.** Confusion matrices describing the performance of three human experts

	CLL	FCC	MCL	NRML	NO DEC
CLL	<b>.5647</b>	.0352	.2117	<b>.1764</b>	.0117
FCC	.0285	<b>.9428</b>	.0000	<b>.0285</b>	.0000
MCL	.1538	.0769	<b>.5538</b>	<b>.1692</b>	.0461
NRML	.1228	.0000	.1053	<b>.7543</b>	.0175

	CLL	FCC	MCL	NRML	NO DEC
CLL	<b>.4000</b>	.0588	.1647	<b>.3765</b>	.0000
FCC	.0000	<b>1.000</b>	.0000	<b>.0000</b>	.0000
MCL	.0769	.0923	<b>.5538</b>	<b>.1692</b>	.0923
NRML	.0000	.0877	.1053	<b>.7719</b>	.0351

	CLL	FCC	MCL	NRML	NO DEC
CLL	<b>.4941</b>	.0235	.2118	<b>.2000</b>	.0471
FCC	.0000	<b>.8857</b>	.0857	<b>.0286</b>	.0000
MCL	.4308	.0154	<b>.3077</b>	<b>.0308</b>	.2154
NRML	.2000	.0364	.1455	<b>.3455</b>	.2727

## 8 Performance evaluation and comparisons

Ten-fold cross-validated classification [12, p. 238] was implemented to provide a more realistic estimation of the system performance. The data set was randomized and split into ten approximately equal test sets, each containing about 9 CLL, 3 FCC, 6 MCL, and 5 Normal cases. For the  $q$ -th test set, its complement was used to obtain the best weights through the downhill simplex method described above. The confusion matrix  $\mathbf{R}_q$  of the resulting classifier was then computed over the  $q$ -th test set for seven retrievals. The elements of the cross-validated confusion matrix (shown in Table 2) were defined as

$$P_{cv}(j|i) = \frac{1}{10} \sum_{q=1}^{10} P_q(j|i), \quad (18)$$

for  $i = 1 \dots 4$ , and  $j = 1 \dots 5$ .

As it can be seen, the system performance is satisfactory, especially when related to the current difficulties in differentiating among lymphoproliferative disorders based solely on morphological criteria [4].

The results of three human experts that classified the digitized specimens from the same database are presented in Table 3. The experts were shown one digitized specimen at a time on a high-resolution screen with no other distractor displayed.

By comparing Table 2 and Table 3, we observe that the human performance is slightly better for FCC and Normal cases, but it is worse for the CLL and MCL cases, both in terms of probabilities of correct decision (the marked diagonals) and probabilities of false negatives (the NRML column). The correlation between the human and machine results is also noteworthy. The classification of the FCC cells proved to be the easiest task, while the CLL and MCL cells resulted in similar levels of difficulty.

We note here that in a real classification scenario, the human expert uses a lot of context information including both patient data and additional data inferred from the digitized specimens. We therefore stress the *decision support* function of the IGDS system. The system is not intended to provide automatic identification of the disorder, but to assist the pathologist to improve its own analysis. The pathologist combines the objective classification suggested by the system with the context information to obtain a robust diagnostic decision.

## 9 User interface

A display capture of the user interface is shown in Fig. 11. The query image with the ROI is top-left, the delineated nucleus of the cell and the normalized shape of the nucleus recovered from 40 Fourier invariants are top-middle, and the eight retrieved images are at the bottom.

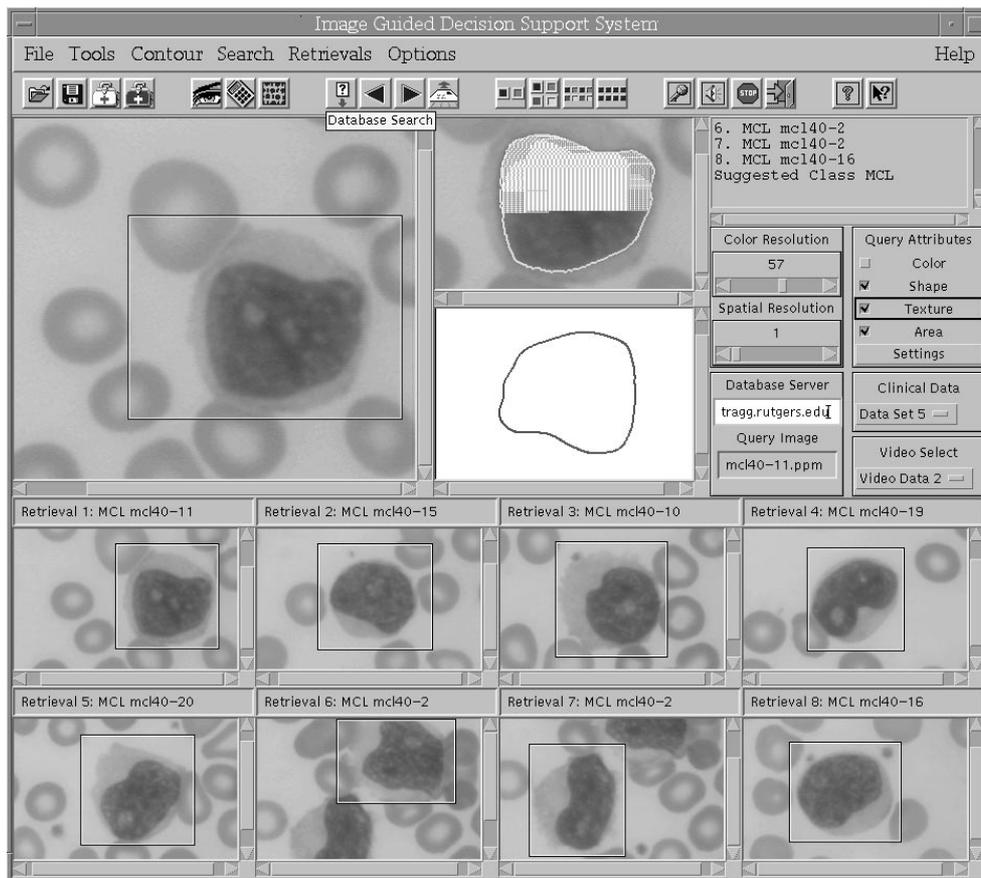
The user can modify the color resolution and spatial resolution of the segmentation, which are defined as the inverses of the segmentation parameters  $h$  and  $T_2$ , respectively. Access to the resolution parameters is only for experiments and maintenance, in normal operations of the system they are set by default. While the segmentation produces reliable results for almost all the images in the database, the system provides a user-handled contour correction tool based on cubic splines. It is also possible to select different query attributes, browse the retrievals, select a different scale for visualization, and display specific clinical data and video clips.

The design of the interface provides natural communication with the search engine. Most input commands can be formulated by voice or graphical input. Currently, the system employs a Microsoft speech recognizer engine with finite-state grammar. The use of a small, task-specific vocabulary results in very high recognition rate. The recognition is speaker-independent. Typical voice commands are: Open Image ##, Save Image ##, Segment the Image, Search the Database, Show 2 (4, 8) Retrievals, Show First (Next, Previous) Retrievals, Show Video, Clinical Data #. Examples of voice feedback are: Image ## Opened, Segmentation Completed, Analyzing Texture, Database Search Completed, Suggested Class: CLL (FCC, MCL, Normal).

## 10 Conclusion

We presented an image-understanding-based system that supports decision making in clinical pathology. Preliminary assessment of the IGDS system showed satisfactory performance. In addition, the retrieval is fast (sequential searching and ranking of the logical database takes about 50 ms on a Pentium II, 266 MHz), allowing the extension of the current database to thousands of images with no noticeable increase in the delay at the end user. The retrieval delay depends mostly on the bandwidth available for the client-server communication.

The focus of the research reported in this paper is primarily to reduce the number of false negatives during routine specimen screening by medical technologists. To realize a similar reduction in false negatives without the IGDS system, one would need to immunophenotype every specimen



**Fig. 11.** User interface of the IGDS system

that is flagged by the complete blood count device. We believe that such a policy would be more expensive than using the IGDS system.

We carefully selected this problem domain because of its clinical relevance and because the actual diagnosis can be independently and objectively determined through flow cytometry. Having an experimental framework where the ground truth is unambiguously established, we were able to formulate and validate a simple cellular model for expert system research in cytopathology. The establishment of new guidelines for visually characterizing lymphoproliferative disorders represents another goal of the IGDS-related research.

At present, the system features including the bimodal human-computer interaction are being evaluated in real retrieval scenarios at the Department of Pathology, UMDNJ-RWJ Medical School. A larger database is to be indexed, to perform a statistically more significant analysis of the system. Other query attributes are examined such as the ratio of the nuclear area over the cytoplasm area of the cell. A second operational mode of the system that would allow the automatic scanning and analysis of the entire microscopic specimen is also investigated.

Future information regarding the IGDS system can be found at <http://www.caip.rutgers.edu/~comanici/jretrieval.html>

Finally, we note that all the computational modules of the system are general and context independent (nonparametric segmentation, Fourier analysis, multiscale autoregressive modeling, multidimensional optimization). Therefore, with a

certain degree of tuning, other 2D application domains can be easily considered.

*Acknowledgements.* The authors would like to thank to Dr. Lauri Goodell and Dr. Pamela Kidd of the Department of Pathology, UMDNJ-RWJ Medical School, for providing the ground truth of the images in the database. Dorin Comaniciu and Peter Meer were supported by the NSF through the grant IRI-9530546 and IRI-9618854. David J. Foran was supported by the Whitaker Foundation, through the grant 98-0202. A short version of this paper appeared in IEEE Workshop on Applications of Computer Vision, October 1998, Princeton, N.J.

## References

1. Antani S, Kasturi R, Jain R (1998) Pattern Recognition Methods in Image and Video Databases: Past, Present and Future. In: Amin A, et al. (ed), *Advances in Pattern Recognition, Lecture Notes in Comp. Science*. Springer, Berlin Heidelberg New York, pp 31–53
2. Banks PM et al. (1992) Mantle Cell Lymphoma: A proposal for Unification of Morphologic, Immunologic, and Molecular Data. *Am. J. Surg. Pathology* 16:637–640
3. Bauman I, Nenninger R, Harms H, Zwierzina H, Wilms K, Feller AC, Meulen VT, Muller-Hermelink HK (1995) Image Analysis Detects Lineage-Specific Morphologic Markers in Leukemic Blast Cells. *Am J Clin Pathol* 105:23–30
4. Campo E, Jaffe E (1996) Mantle Cell Lymphoma. *Arch Pathol Lab Med* 120:12–14
5. Chan J, Banks PM, et al. (1995) A Revised European-American Classification of Lymphoid Neoplasms Proposed by the International Lymphoma Study Group. *Am J Clin Pathol* 103:543–560
6. Cho K, Meer P (1997) Image Segmentation from Consensus Information. *Comput Vision Image Understanding* 68:72–89

7. Comaniciu D, Meer P (1999) Distribution Free Decomposition of Multivariate Data. *Pattern Anal Appl* 2:22–30
8. Comaniciu D, Georgescu B, Meer P, Chen W, Foran D (1999) Decision Support System for Multiuser Remote Microscopy in Telepathology. In: *IEEE Symposium on Computer-Based Medical Systems, 1999*, Stamford, Conn., IEEE Computer Society Press, pp 150–155
9. Comaniciu D, Meer P, Xu K, Tyler D (1999) Retrieval Performance Improvement through Low Rank Corrections. In: *IEEE Workshop on Content-Based Access of Image and Video Libraries, 1999*, Fort Collins, Colorado, IEEE Computer Society Press, pp 50–54
10. Cox JJ, Miller ML, Minka TP, Yianilos PN (1998) An Optimized Interaction Strategy for Bayesian Relevance Feedback. In: *IEEE Conf. on Computer Vision and Pattern Recognition, 1998*, Santa Barbara, Calif., IEEE Computer Society Press, pp 553–558
11. Das M, Riseman EM, Draper BA (1997) FOCUS: Searching for Multicolored Objects in a Diverse Image Database. In: *IEEE Conf. on Computer Vision and Pattern Recognition, 1997*, San Juan, Puerto Rico, IEEE Computer Society Press, pp 756–761
12. Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*. Chapman & Hall, New York
13. Flickner M et al. (1995) Query by Image and Video Content: The QBIC System. *Computer* 9:23–31
14. Fukunaga K (1990) *Introduction to Statistical Pattern Recognition*, Second Edition. Academic Press, Boston, Mass.
15. Kauppinen H, Seppanen T, Pietikainen M (1995) An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification. *IEEE Trans Pattern Anal Mach Intell* 17:201–207
16. Kittler J, Hatem M, Duin RPW, Matas J (1998) On Combining Classifiers. *IEEE Trans Pattern Anal Mach Intell* 20:226–238
17. Kuhl FP, Giardina CR (1982) Elliptic Fourier Features of a Closed Contour. *Comput Graphics Image Process* 18:236–258
18. Liu F, Picard RW (1996) Periodicity, Directionality, and Randomness: World Features for Image Modeling and Retrieval. *IEEE Trans Pattern Anal Mach Intell* 18:722–733
19. Liu Y, Dellaert F (1998) A Classification Based Similarity Metric for 3D Image Retrieval. In: *IEEE Conf. on Computer Vision and Pattern Recognition, 1998*, Santa Barbara, Calif., IEEE Computer Society Press, pp 800–805
20. Ma WY, Manjunath BS (1997) NETRA: A toolbox for navigating large image databases. *IEEE Int. Conf. Image Processing, 1997*, Santa Barbara, Calif., IEEE Computer Society Press, pp 568–571
21. Mao J, Jain AK (1992) Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models. *Pattern Recognition* 25:173–188
22. Nastar C, Mitschke M, Meihac C (1998) Efficient Query Refinement for Image Retrieval. In: *IEEE Conf. on Computer Vision and Pattern Recognition, 1998*, Santa Barbara, Calif., IEEE Computer Society Press, pp 547–552
23. Ortega M, Rui Y, Chakrabarti K, Mehrotra S, Huang TS (1997) Supporting Similarity Queries in MARS. In: *Proc. ACM Multimedia, 1997*, Seattle, Wash., IEEE Computer Society Press, pp 403–413
24. Pentland A, Picard RW, Sclaroff S (1996) Photobook: Content-based manipulation of image databases. *Int J Comput Vision* 18:233–254
25. Persoon E, Fu KS (1998) Shape Discrimination using Fourier Descriptors. *IEEE Trans Syst Man Cybern* 7:170–179
26. Petrakis EGM, Faloutsos C (1997) Similarity Searching in Large Databases. *IEEE Trans Knowl Data Eng* 9:435–447
27. Picard RW, Kabir T, Liu F (1993) Real-time Recognition with the Entire Brodatz Texture Database. In: *IEEE Conf. on Computer Vision and Pattern Recognition, 1993*, New York, N.Y., IEEE Computer Society Press, pp 638–639
28. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical Recipes in C*, Second Edition. Cambridge University Press, Cambridge
29. Rui Y, She AC, Huang TS (1998) A Modified Fourier Descriptor for Shape Matching in MARS. In: Chang SK (ed) *Image Databases and Multimedia Search*, Series on Software Engineering and Knowledge Engineering Vol. 8. World Scientific Publishing, Singapore, pp 165–180
30. Sharma R, Pavlovic VI, Huang TS (1998) Toward Multimodal Human-Computer Interface. *Proc IEEE* 86(5):853–869
31. Tagare HD, Jaffe CC, Duncan J (1997) Medical Image Databases: A Content-based Retrieval Approach. *J Am Med Inf Assoc* 4:184–198
32. Wyszecki G, Stiles WS (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Second Edition. Wiley, New York



**Dorin Comaniciu** received the Dipl. Engr. and D.Sc. degrees in electrical engineering from the Polytechnic University of Bucharest, Romania, in 1988 and 1995, respectively. He is currently completing the Ph.D. degree in the Department of Electrical and Computer Engineering at Rutgers University, New Jersey. From 1988 to 1990 he was with ICE Felix Computers, Romania. Between 1991 and 1995 he was a teaching assistant at the Polytechnic University of Bucharest. He held research appointments in Germany and France. Since 1999 he is with Siemens Corporate Research, Princeton, New Jersey. His research interests include robust methods for computer vision, nonparametric analysis, content-based image/video retrieval, and data compression.



**Peter Meer** received the Dipl. Engn. degree from the Bucharest Polytechnic Institute, Bucharest, Romania in 1971, and the D.Sc. degree from the Technion, Israel Institute of Technology, Haifa, Israel, in 1986, both in electrical engineering. From 1971 to 1979 he was with the Computer Research Institute, Cluj, Romania, working on R&D of digital hardware. Between 1986 and 1990 he was Assistant Research Scientist at the Center for Automation Research, University of Maryland at College Park. In 1991 he joined the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ and is currently an

Associate Professor. He has held visiting appointments in Japan, Korea, Sweden, Israel and France and was on the organizing committees of several international workshops and conferences. He is an Associate Editor of the *IEEE Transaction on Pattern Analysis and Machine Intelligence*, and is member of the Editorial Board of the journal *Pattern Recognition*. He was coauthor of an award winning paper in *Pattern Recognition* in 1989, and the *IEEE Conference on Computer Vision and Pattern Recognition* in 1999. His research interest is in application of modern statistical methods to image understanding problems.



**David J. Foran** received the Bachelors degree in Zoology and Physics from Rutgers University in 1983 and worked as a junior scientist at Johnson & Johnson Research, Inc until 1986. In 1992 he earned a Ph.D. in Biomedical Engineering from the University of Medicine and Dentistry of New Jersey & Rutgers University and received one year post-doctoral training in computational biology and molecular imaging at the department of Biochemistry at UMDNJ-Robert Wood Johnson Medical School. He joined the faculty at RWJMS in 1994 where he is currently an Assistant Professor of Pathology & Radiology and the

Director of the Center for Biomedical Imaging & Informatics. Dr. Foran serves an Associate Editor for the *IEEE Transactions on Information Technology in Biomedicine*. His research interests include quantitative, biomedical imaging, computer-assisted diagnosis, and medical informatics.