

Bimodal System for Interactive Indexing and Retrieval of Pathology Images

Dorin Comaniciu *
Department of ECE
Rutgers University
Piscataway, NJ 08855
comanici@caip.rutgers.edu

Peter Meer *
Department of ECE
Rutgers University
Piscataway, NJ 08855
meer@caip.rutgers.edu

David Foran
Biomedical Imaging
UMDNJ-RWJ
Piscataway, NJ 08854
djf@pleiad.umdj.edu

Attila Medl
CAIP Center
Rutgers University
Piscataway, NJ 08855
medl@caip.rutgers.edu

Abstract

*The prototype of a system to assist the physicians in differential diagnosis of lymphoproliferative disorders of blood cells from digitized specimens is presented. The user selects the region of interest (ROI) in the image which is then analyzed with a fast, robust color segmenter. Queries in a database of validated cases can be formulated in terms of shape (similarity invariant Fourier descriptors), texture (multiresolution simultaneous autoregressive model), color ($L^*u^*v^*$ space), and area, derived from the delineated ROI. The uncertainty of the segmentation process (obtained through a numerical method) determines the accuracy of shape description (number of Fourier harmonics). Ten-fold cross-validated classification over a database of 261 color 640×480 images was implemented to assess the system performance. The ground truth was obtained through immunophenotyping by flow cytometry. To provide a natural man-machine interface, most input commands are bimodal: either using the mouse or by voice. A speech synthesizer provides feedback to the user. All the employed computational modules are context independent and thus the same system can be used in a large variety of application domains.*

1 Introduction

In this paper we describe a task-oriented indexing and retrieval system which exploits the informational content of digitized specimens in order to assist pathologists to discriminate among Chronic Lymphocytic Leukemia (CLL), Follicular Center Cell Lymphoma (FCC), Mantle Cell Lymphoma (MCL), and healthy leukocytes (NORMAL) (see Figure 1). The MCL was only recently (1992) identified and is often misdiagnosed as CLL or FCC [1]. Its timely, accurate diagnosis is of extreme importance since MCL typically has a more aggressive clinical course than CLL or FCC.

In contrast to general purpose retrieval engines [4, 6, 10, 12], our system is applied to a very specific problem. In the medical practice such a system should only provide reliable support but not the final decision [15]. While it is possible to automatically select the region of interest (ROI) in an image, the nucleus of a leukocyte cell, to obey the medi-

cal tradition, this task is left to the user. The technician will draw a rectangular ROI containing the cell of interest. The ROI is analyzed automatically with a color segmentation algorithm. The digitized specimens have different contrasts and noise levels, therefore the robustness of the segmenter is of paramount importance. The employed segmenter is based on a recent nonparametric cluster analysis technique [3] which can process a 10000 point 3-D data set in only a few seconds.

To access the database of digitized specimens four attributes of the delineated cell nucleus are defined: shape, texture, color, and area. Similarity invariant Fourier descriptors [8] are used to characterize the shape, with the stability of the segmentation process being taken into account to determine the number of significant harmonics. The texture analysis module uses the multiresolution simultaneous autoregressive model [11].

A distinct feature of our retrieval system is its bimodal human-computer interaction, which allows natural communication with the search engine. Queries can be formulated or refined by speech or mouse input. A fusion agent interprets the commands, calls the appropriate method, and gives voice feedback to the user.

The paper is organized as follows. In Section 2 the segmentation algorithm is discussed. The shape descriptors are presented in Section 3. Section 4 describes the implementation of the texture analysis and also discusses the color attribute. Section 5 shows how the overall similarity measure is defined and presents the cross-validated retrieval performance. The user interface of the system is outlined in Section 6.

2 Segmentation Algorithm

A fast nonparametric clustering method is employed in the perceptually uniform $L^*u^*v^*$ space to detect clusters and delineate the cluster borders. Each cluster corresponds to homogeneous regions in the image. The resulting analysis completely describes the color space, and allows the design of a robust and stable color segmentation algorithm. The steps of the algorithm are described below. For more details see [3].

*Supported by the NSF under the grant IRI-9530546.

1. Map the image into the color space. Transforming the RGB vectors into $L^*u^*v^*$ vectors a perceptually uniform color space is obtained.

2. Define the sample set. A set of m points $X_1 \dots X_m$ called the *sample set* is randomly selected from the data. Two constraints are imposed on the points retained in the sample set. 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. The latter condition is required to avoid convergence problems for the mean shift procedure. A region is sparsely populated whenever the number of points inside the sphere is below a threshold T_1 . Note that the distance and density constraints automatically determine the size m of the sample set.

3. Apply the mean shift procedure. A set containing m *cluster center candidates* is defined by the points of convergence of the m mean shift procedures. Since the computation of the mean shift vectors is based on the entire data set, the quality of the density gradient estimate is not diminished by the use of sampling.

4. Test the mean shift convergence. A local plateau in the color space can prematurely stop the mean shift iterations. Therefore, each cluster center candidate is perturbed by a random vector of small norm and the mean shift procedure is let to converge again.

5. Derive the cluster centers $Y_1 \dots Y_p$. 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 such that their distance is less than h), defines a *cluster center*. The cluster center is the mean of the cluster center candidates in the subset. Note that $p \leq m$.

6. Validate the cluster centers. Between any two cluster centers Y_i and Y_j a significant valley should occur in the underlying density. The existence of the valley is tested for each pair (Y_i, Y_j) . The testing procedure is described in [3]. If no valley was found between Y_i and Y_j , the cluster center of lower density (Y_i or Y_j) is removed from the set of cluster centers.

7. Delineate the clusters. At this stage each sample point is associated with a cluster center. To allocate the data points a k -nearest neighbor technique is employed, i.e., each data point belongs to the cluster defined by the majority of its k -nearest sample points.

8. Enforce spatial constraints 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.

The segmentation makes use of three parameters: 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. The results described in this paper were obtained with $h = 4$, $T_1 = 50$

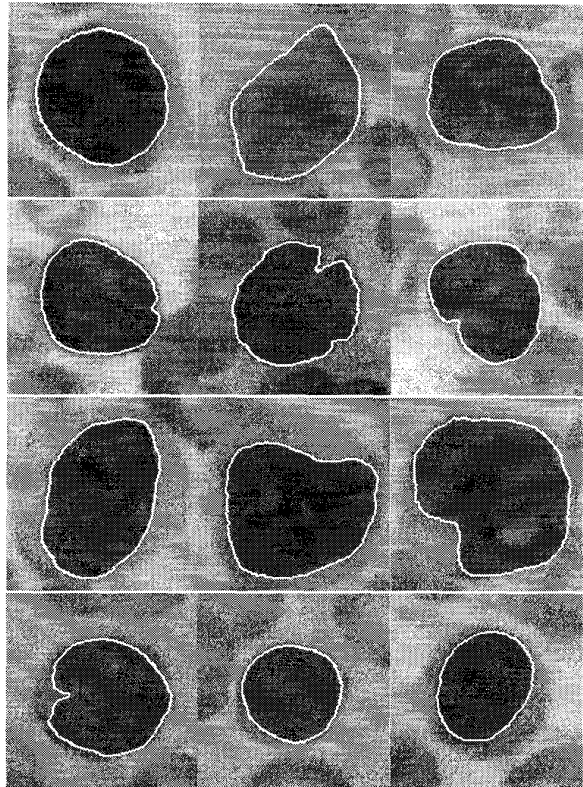


Figure 1. Segmentation of the nucleus of various cells. The cells in each row belong to the same class: CLL, FCC, MCL, and NORMAL, respectively.

and $T_2 = 1000$. Examples of nucleus delineation performed by the segmentation module are given in Figure 1.

3 Similarity Invariant Shape Descriptors

The shape analysis module implements Fourier descriptors which are made invariant to changes in location, orientation and scale, that is, *similarity invariant*. Fourier invariants were recently proved to be superior to methods based on autoregressive models [7].

The system employs the method described in [8] which uses an arc length s parameterization of the contour and the complex function of the coordinates $u(s) = x(s) + jy(s)$, where the functions $x(s)$ and $y(s)$ are separately expanded. There exists a straightforward geometric interpretation of this approach, the closed contour being represented as a composition of ellipses, called harmonic loci.

To obtain Fourier coefficients which are rotation invariant, one must compensate for the arbitrary position of the starting point on the contour, and for the arbitrary orientation of the contour. Two rotations achieve the invariance. When the first harmonic locus is an ellipse, the rotations are defined relative to the semi-major axis of the locus, resulting in two

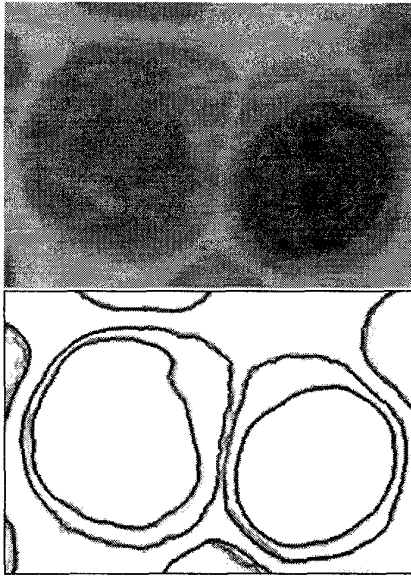


Figure 2. Example of segmentation stability. A region of interest from the image *cll40-2* (top) has been segmented 25 times and the resulting contours were superimposed (bottom). The darker a pixel in the contour image, the more stable is the contour passing through that pixel. The regions between two cells are the least stable.

related representations of the curve. 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.

Note that the Fourier representations in [7, 10] achieve rotation invariance by discarding the phase information of all coefficients. However, since misleading classifications can be obtained by neglecting the phase, we use the rotation invariant representation described above, which preserves the phase of the coefficients.

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. Translation invariance results by simply removing the DC terms from the Fourier series.

3.1 Number of Reliable Harmonics

To optimize the accuracy of shape description we take into account the stability of the segmentation when choosing the number of harmonics which reliably represent the shape. Due to its probabilistic nature the segmentation algorithm produces slightly different results when repeatedly

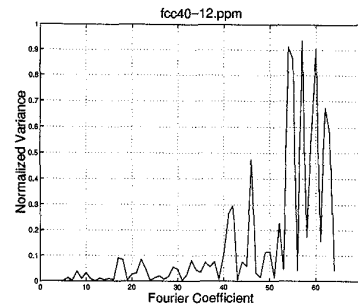


Figure 3. Normalized variances of the first 64 Fourier coefficients corresponding to image *fcc40-12*.

applied to the same image (see Figure 2). The stability of the delineated contour is a good measure of the confidence in the segmentation [2].

To estimate the influence of the segmentation 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 region of interest (a leukocyte). The region was then segmented and the first 16 harmonics (64 coefficients) were determined for the nucleus. A typical result, the normalized variances of the Fourier coefficients of the nucleus from the image *fcc40-12* are presented in Figure 3. It can be concluded that the segmentation is sufficiently stable for the use of the first 10 harmonics (40 coefficients). Consequently, we compare two shapes by computing the Euclidean distance between the first 40 Fourier invariants.

4 Texture and Color Analysis

Texture analysis of the cell nucleus is based on the multiscale simultaneous autoregressive (MRSAR) model [11]. We apply a symmetric MRSAR model to the L^* component of the $L^*u^*v^*$ image data. The pixel value at a certain location is expressed as a linear combination of the neighboring pixel values and an additive independent Gaussian noise term. Four model parameters (the weights associated with each of the neighboring pixel values) are estimated through least squares. The model parameters and the estimation error define a 5-dimensional feature vector for the given neighborhood. The multiresolution feature vector is obtained by varying the neighborhood size and concatenating the obtained features.

In [9, 13] it was shown that the MRSAR features computed with 5×5 , 7×7 , and 9×9 neighborhoods provide the best overall retrieval performance over the entire Brodatz database. While the textures inside the nuclei (see Figure 4) are different from the Brodatz ones, the same features are used here to form a 15-dimensional multiresolution feature vector. The covariance matrix of the feature vectors

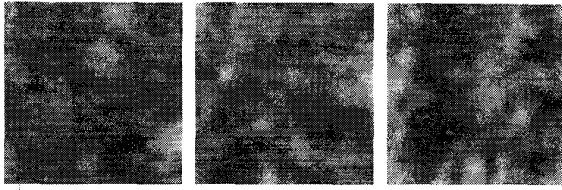


Figure 4. Examples of nucleus texture. The gray level dynamic range was enlarged to improve reproduction quality.

within each cell nucleus is also computed, and the distance between two nuclei in terms of their texture is given by the Mahalanobis distance of their MRSAR feature vectors.

The nucleus color is specified as a 3-D vector in the $L^*u^*v^*$ color space and is determined by the segmentation module. However, we handle color as a potential query attribute only. To validate the query by color, the formation and capturing of the stained cellular material should follow a standardized procedure. Since the images in our database were captured under different conditions, the colors of the nuclei tend to cluster according to the capturing conditions and not as a function of the cell class (see Figure 5). Therefore, the current implementation of the system uses the color attribute for nucleus separation from the background, but not for distinguishing among different cells.

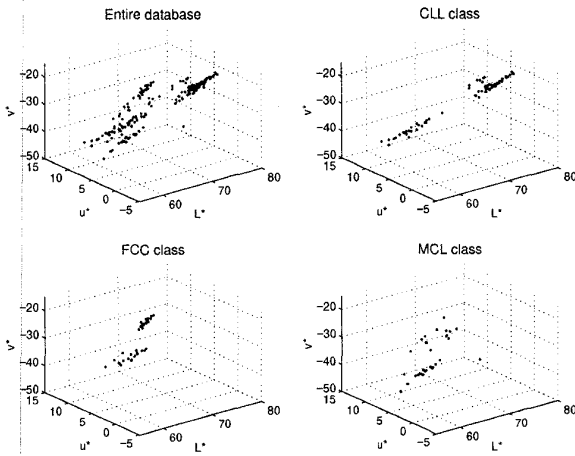


Figure 5. The color vectors of all nuclei in the database and those corresponding to the CLL, FCC, and MCL classes, respectively. Note that the color vectors are not clustered according to their class. For example, the data in class FCC was captured under two different conditions and this fact is clearly represented in the distribution of the color vectors.

5 Overall Similarity Metric and Classification Performance

The retrieval system is currently using a database containing 98 CLL, 38 FCC, 66 MCL, and 59 NORMAL, a total of 261 images. The ground truth was obtained for each specimen by an independent chemical procedure, immunophenotyping by flow cytometry. A number of k images are retrieved (closest matches) in response to a given query. The suggested classification of the query image is based on majority voting among the classes of the first n retrievals. This procedure may also produce a NO DECISION class, in addition to the four original cell classes.

The system performance is given by the *confusion matrix* \mathbf{R} defined as having the element r_{ij} equal to the empirical probability of classification in class j when the query image belonged to class i . For optimum performance the sum of probabilities of correct decision

$$J = \sum_{i=1}^4 r_{ii} \quad (1)$$

should be maximized.

Since the shape, texture, and area are attributes of different nature the definition of overall similarity metric is a complex issue. A relative simple solution is to express the similarity as a linear combination of the normalized distances corresponding to each query attribute. Thus, for 3 attributes

$$D = \sum_{i=1}^3 w_i D_i, \quad (2)$$

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

The optimal weights w_i were obtained off-line by employing the downhill simplex method [14, p. 408] with the objective function J (1). The optimization space is 2-dimensional (there are only 2 independent weights). As an advantage, the downhill simplex requires only function evaluations and not computation of derivatives. It performs an iterative search for the optimal weights by allowing the initial simplex (a triangle in 2-D) to be modified and moved such that to converge to a maximum of the objective function. Since the method guarantees a local maximum only, we run the optimization 16 times with different initializations. A tessellation of the triangle defined by the values of w_0 and w_1 yielded the 16 initial simplexes. In Table 1 two sets of weights, obtained by running the optimization with $n = 7$ retrievals (the query image excluded) over the entire database are shown. They correspond to the two highest obtained values of the objective function (1).

Ten-fold cross-validated classification [5, p. 238] was implemented to provide a more realistic estimation of the system performance. The data set was randomized and split into 10 approximately equal test sets, each containing about 9 CLL, 3 FCC, 6 MCL, and 5 NORMAL cases. For the

Table 1. Optimal weights and the objective function produced by the downhill simplex running over the entire database.

Shape	Texture	Area	J
0.1140	0.5771	0.3089	3.4207
0.1992	0.3399	0.4609	3.3530

k -th test set its complement was used to obtain the optimal weights through the downhill simplex method described above. The confusion matrix $\mathbf{R}^{(k)}$ of the resulting classifier was then computed over the k -th test set for $n = 7$ retrievals. The elements of the cross-validated confusion matrix \mathbf{R}_{cv} were defined as

$$r_{ij} = \frac{1}{10} \sum_{k=1}^{10} r_{ij}^{(k)}, \quad (3)$$

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

Table 2 shows the cross-validated confusion matrix \mathbf{R}_{cv} . As can be seen the system performance is very promising, especially when related to the current difficulties in differentiating among lymphoproliferative disorders based solely on morphologic criteria [1].

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

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

6 User Interface

A capture of the user interface is showed in Figure 6. The design of the interface provides natural, bimodal communication with the search engine. Most input commands can be formulated by voice or using the mouse. A fusion agent capable of multimodal inputs interprets the commands, calls the appropriate method, and gives voice feedback to the user. Currently the system employs a speech recognizer engine with finite-state grammar and a restricted task-specific vocabulary. The recognition is speaker-independent. The system has a client-server architecture. Server access is multithreaded, Internet based, and employs Java sockets.

7 Conclusion

At present the system performance is being evaluated in real retrieval scenarios by comparing the accuracy and time required to render diagnosis using our system and using conventional screening methods. Other query attributes are also being investigated such as the ratio of the cytoplasm area over the nucleus area. Since all the computational modules

are context independent other 2-D application domains can be easily considered.

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

Acknowledgment

The authors would like to thank to Dr. Lauri Goodell and Dr. Pamela Kidd of Department of Pathology, UMDNJ-RWJ Medical School, for providing the ground truth of the images in the database.

References

- [1] E. Campo and E. Jaffe. Mantle Cell Lymphoma. *Arch. Pathol. Lab. Med.*, 120:12–14, 1996.
- [2] K. Cho and P. Meer. Image Segmentation from Consensus Information. *Comp. Vis. and Image Understanding*, 68:72–89, 1997.
- [3] D. Comaniciu and P. Meer. Distribution Free Decomposition of Multivariate Data. *Second Int'l Workshop on Statistical Techniques in Pattern Recognition*, To appear, Sydney, Australia 1998.
- [4] M. Das, E. Riseman, and B. Draper. FOCUS: Searching for Multi-colored Objects in a Diverse Image Database. *IEEE Conf. on Comp. Vis. and Pattern Recognition*, pages 756–761, Puerto Rico 1997.
- [5] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
- [6] M. Flickner et al. Query by Image and Video Content: The QBIC System. *Computer*, 28(9):23–31, 1995.
- [7] H. Kauppinen, T. Seppanen, and M. Pietikainen. An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification. *IEEE Trans. Pattern Anal. Machine Intell.*, 17:201–207, 1995.
- [8] F. Kuhl and C. Giardina. Elliptic Fourier Features of a Closed Contour. *Comp. Graphics Image Process.*, 18:236–258, 1982.
- [9] F. Liu and R. Picard. Periodicity, Directionality, and Randomness: Wold Features for Image Modeling and Retrieval. *IEEE Trans. Pattern Anal. Machine Intell.*, 18:722–733, 1996.
- [10] W. Ma and B. Manjunath. NETRA: A Toolbox for Navigating Large Image Databases. *IEEE Int'l Conf. Image Process.*, 1:568–571, Santa Barbara 1997.
- [11] J. Mao and A. Jain. Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models. *Pattern Recognition*, 25:173–188, 1992.
- [12] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-Based Manipulation of Image Databases. *Int'l. J. of Comp. Vis.*, 18:233–254, 1996.
- [13] R. Picard, T. Kabir, and F. Liu. Real-time Recognition with the Entire Brodatz Texture Database. *IEEE Conf. on Comp. Vis. and Pattern Recognition*, pages 638–639, New York 1993.
- [14] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. University Press, Cambridge, 1992.
- [15] H. Tagare, C. Jaffe, and J. Duncan. Medical Image Databases: A Content-based Retrieval Approach. *J. of the American Medical Inform. Assoc.*, 4:184–198, 1997.

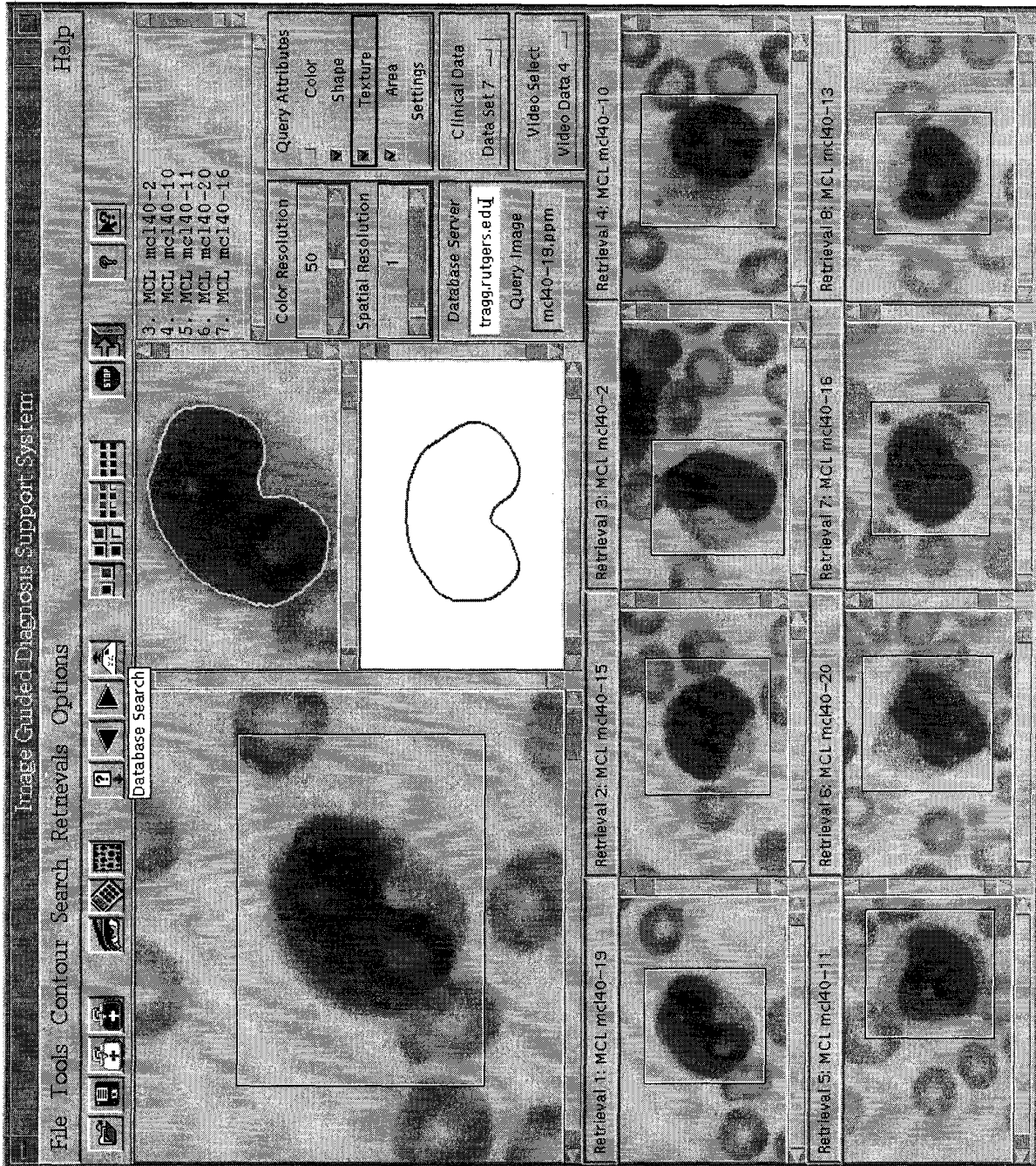


Figure 6. A screen capture of the system. The query image *mcl40-19* with the region of interest is top-left, the delineated nucleus of the cell and the normalized shape of the nucleus recovered from 40 Fourier invariants are top-middle, and the eight retrieved images are at the bottom. The user can 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.