Unsupervised Segmentation Based on Robust Estimation and Color Active Contour Models

Lin Yang, Peter Meer, *Senior Member, IEEE*, and David J. Foran, *Member, IEEE*

Abstract—One of the most commonly used clinical tests performed today is the routine evaluation of peripheral blood smears. In this paper, we investigate the design, development, and implementation of a robust color gradient vector flow (GVF) active contour model for performing segmentation, using a database of 1791 imaged cells. The algorithms developed for this research operate in Luv color space, and introduce a color gradient and L_2E robust estimation into the traditional GVF snake. The accuracy of the new model was compared with the segmentation results using a mean-shift approach, the traditional color GVF snake, and several other commonly used segmentation strategies. The *unsupervised* robust color snake with L_2E robust estimation was shown to provide results which were superior to the other *unsupervised* approaches, and was comparable with *supervised* segmentation, as judged by a panel of human experts.

Index Terms—Active contours, image segmentation, L_2E robust estimation, unsupervised segmentation.

I. INTRODUCTION

A. Clinical Background and System Description

A differential diagnosis is crucial for determining which medications are appropriate and what level of risk is justified. As new therapy options become available, it has become increasingly important to distinguish among subclasses of pathologies [40]. The subtle visible differences exhibited by some malignancies can give rise to a significant number of false negatives during routine screening. In such cases, the diagnosis can only be rendered after immunophenotyping and/or molecular or cytogenetic studies have been performed. Unfortunately, these studies occur too late in the diagnostic pathway to impact significantly on the frequency of false negatives. Passing specimens through a reliable image-based screening system, however, could potentially reduce costs and patient morbidity.

Usually, the cells of interest are the white blood cells (WBC) including neutrophils, eosinophils, monocytes, and lymphocytes. The main objective of our studies was to distinguish among different subgroups of lymphocytes. Three major well-defined entities of leukemia/lymphomas which can sometimes be confused with one another during routine microscopic evaluation are: B-cell chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL), and follicle-center cell lymphoma (FCC) [46], [47]. These are small lymphoid lymphomas that are known to be associated with bone marrow and/or peripheral blood involvement during the course of the disease or at the time of the diagnosis [48]. Depending on the clinical presentation, age, sex, and peripheral blood counts of patients, an accurate diagnosis may prove difficult when only a limited amount of specimen is available for ancillary studies, such as flow cytometry, karyotyping, fluorescence *in-situ* hybridization (FISH), polymerase chain reaction (PCR), or reverse transcriptase PCR (RT-PCR). The tremendous effort undertaken to subtype these disorders is driven by the medical community's desire to identify improved treatment protocols and therapy options [41], [42].

Recent literature [43]–[45] ascribes much of the difficulty in rendering consistent diagnoses to subjective impressions of observers, and shows that when morphological cell classification is based upon computer-aided analysis, both objectivity and reproducibility improve. An image-guided decision support (IGDS) system [16], [18] based on the mean-shift clustering algorithm [17] was recently developed to provide assistance to technicians and physicians when they are confronted with detecting and discriminating among CLL, MCL, FCC, and benign cells. The IGDS enables individuals to submit query imaged cells from local or remote computers and robotic microscopes to search engines, which automatically classify the query and retrieve digitized imaged cells and correlated data (protein/molecular studies) from the "ground-truth" databases which exhibit the most similar features. The most probable diagnosis is provided, based upon the majority logic of the ranked retrievals.

While the performance of the first generation of IGDS prototype showed promising results, it exhibited several limitations: 1) populating the "ground-truth" database was a tedious, time-consuming task. A more automated system would be required to facilitate data collection; 2) while the mean-shift segmentation algorithm used in the first generation of the IGDS system could accurately segment cells into multiple subregions, this approach, like many other low-level image-segmentation approaches, often over-segmented cells; 3) the original image-segmentation algorithm used in the first generation of IGDS only addressed the segmentation of the nucleus, although the cytoplasm also contains important information, which can be useful in improving classification performance.

In an attempt to improve the recognition accuracy and automate cellular segmentation, this paper focuses on developing

L. Yang is with the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854 USA and the Center for Biomedical Imaging and Informatics, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854 USA (e-mail: linyang@eden.rutgers.edu).

P. Meer is with the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854 USA.

D. J. Foran is with the Center for Biomedical Imaging and Informatics, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854 USA.
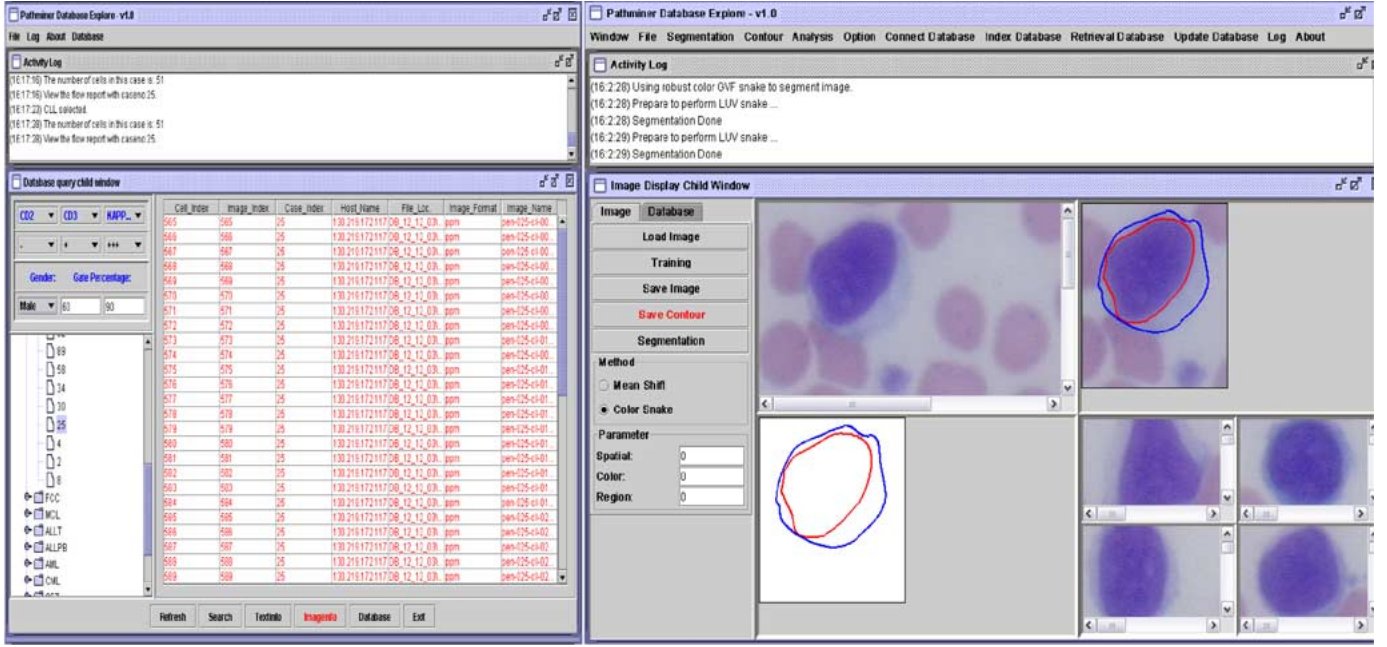
Fig. 1. User interface of the IGDS system. The left panel is the user interface of the new "ground-truth" image database system. The right panel shows the user interface of the new image segmentation and indexing system.

a reliable and *unsupervised* approach, which could facilitate accurate *unsupervised* segmentation of both the cytoplasm and the nucleus. A screen-shot of the new IGDS system implementing a robust color gradient vector flow (GVF) snake is shown in Fig. 1.

B. Image Segmentation

Image segmentation is the process of delineating an image into "homogeneous" regions, based on the similarity of pixel attributes. Active contour models, or snakes, have gained significant attention and become popular image-segmentation methods after their first introduction by Kass *et al.* [1]. Snakes are curves defined within the image's domain, and driven by the internal forces within the curve and the external forces derived from the image data. There are mainly two general types of active contour models described in the literature: parametric active contours [1] and geometric active contours [7], [8], [9], [49]. Parametric active contour models are widely used in many applications, including edge detection [1], object recognition [4], [5], shape modeling [2], [5], and motion tracking [4], [6], to mention only a few. The image gradients can be used as the external forces in parametric active contour models. Examples include the traditional snake [10], [11], the balloon snake [12], the pressure forces model [13], and the GVF model [14], [15]. Geometric active contour models, or geodesic snakes, were almost simultaneously proposed by Caselles *et al.* [9] and by Malladi *et al.* [49] to address the fact that parametric active contour models can not resolve topological changes. Geodesic snakes are based on the theory of curve evolution, and are numerically implemented using level sets. They can automatically manage topology changes in an image, and allow for multiple simultaneous boundary estimations. However, due to their computational complexity, their speed of convergence is slower than parametric snakes.

In our applications, the preprocessing steps of the IGDS system automatically selects regions of interest (ROI) [16], [18], which contains the object cell. In each ROI, there are no topological changes, because there is only one object of interest in each field. In addition, since the speed of convergence is an important issue in our application, parametric active contour models were selected in the development of the new IGDS system. The GVF snake often outperforms other gradient-based models, because of its insensitivity to initial positions and larger capture region. However, it is best suited for binary or gray-level images, as cited in the literature review above, rather than chromatically stained pathology specimens. Simply transforming color images into gray-level images without using color gradients, or using color gradients and then directly applying the GVF snake, does not provide satisfactory results. In this paper, we propose a robust color GVF snake model which combines robust estimation and color gradient. The proposed algorithm was shown to be effective in solving the real-world problem of segmenting stained pathology specimens.

The remainder of the paper is organized as follows. Section II introduces the requisite background information concerning active contour models. In Section III, we introduce the L_2E robust estimation and the newly developed robust color GVF snake. Section IV-A provides the experimental results of the L_2E robust estimation. Section IV-B compares the robust color GVF snake with the Fisher linear discriminant (FLD) approach, the original GVF snake, the watershed algorithm derived from mathematical morphology, and the mean-shift clustering algorithm used in the prototype IGDS subsystem. Section V provides the conclusion and future directions.

II. PARAMETRIC ACTIVE CONTOUR MODEL

A traditional snake is parametrically defined as $\mathbf{x}(s) = (x(s), y(s))$, where $x(s)$ and $y(s)$ are x and y coordinates

along the contour, and s represents the arc length with value in [0, 1], to minimize an energy function

$$E_{\text{snake}} = \int_0^1 \left(E_{\text{int}}(\mathbf{x}(s)) + E_{\text{ext}}(\mathbf{x}(s)) \right) ds \qquad (1)$$

where the first term represents the internal energy, defined as

$$E_{\text{int}}(\mathbf{x}(s)) = \frac{\left(\alpha \left| \mathbf{x}_s(s) \right|^2 + \beta \left| \mathbf{x}_{ss}(s) \right|^2 \right)}{2} \qquad (2)$$

where $\mathbf{x}_s(s)$ is the first derivative of $\mathbf{x}(s)$, and $\mathbf{x}_{ss}(s)$ is the second derivative of $\mathbf{x}(s)$ with respect to s. The external energy is defined as

$$E_{\text{ext}}(\mathbf{x}(s)) = -\gamma \left| \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\} \right| \qquad (3)$$

where f is the original image, and $G_{\sigma(x,y)}$ is the two-dimensional (2-D) Gaussian kernel with σ as standard deviation.

The goal of the computation is to find the local minima of E_{snake}, defined in (1). Based on the Euler–Lagrange principle, (1) has a minimum when

$$\alpha \mathbf{x}_{ss}(s) - \beta \mathbf{x}_{ssss}(s) - \nabla E_{\text{ext}}(\mathbf{x}(s)) = 0 \qquad (4)$$

where $\mathbf{x}_{ss}(s)$ and $\mathbf{x}_{ssss}(s)$ are the second and fourth derivatives of the curve with respect to the parameter s.

In order to find the solution for (4), the snake is made dynamic by defining $\mathbf{x}$ as the function of time t and s

$$\mathbf{x}_t(s,t) = \alpha \mathbf{x}_{ss}(s,t) - \beta \mathbf{x}_{ssss}(s,t) - \nabla E_{\text{ext}}(\mathbf{x}(s,t)). \qquad (5)$$

The solution to (5) can be achieved by solving the discrete equations iteratively.

To effectively lead the snakes into the concave region, Xu and Prince [14] proposed a new external force. According to the Helmholtz theorem [19], rewriting (5) and replacing the $-\nabla E_{\text{ext}}(\mathbf{x}(s))$ with Θ

$$\mathbf{x}_t(s,t) = \alpha \mathbf{x}_{ss}(s,t) - \beta \mathbf{x}_{ssss}(s,t) + \Theta \qquad (6)$$

where Θ is the GVF defined as $\Theta(x,y) = [u(x,y), v(x,y)]$ that minimizes the energy functional shown in (7) at the bottom of the page, where $\nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\}$ is the gradient of the input image $f(x,y)$ after Gaussian smoothing, with variance σ and mean $\mathbf{0}$. Equation (7) is dominated by $u_x^2 + u_y^2 + v_x^2 + v_y^2$ when $\left| \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\} \right|$ is small. When $\left| \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\} \right|$ is large, the second term dominates the integrand, and is minimized when $\Theta = \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\}$. This result keeps Θ nearly equal to the gradient of the edge when the snake is near the object, while enabling the snake to move toward the edges when it is far away from the object.

III. UNSUPERVISED IMAGE SEGMENTATION USING ROBUST COLOR GVF SNAKE

A. Color Gradient in Luv Color Space

In gray-level images, the gradient is defined as the first derivative of the image luminance. It has a high value in those regions exhibiting high luminance contrast. However, this strategy is not suitable for color images. Simply transforming color images into gray-level images by taking the average of three channels and applying the gray-level image gradient operator does not provide satisfactory results in our applications.

In this paper, we adopt the definition of gradients for color images [22]–[24]. In contrast to previous approaches, we define the color gradient in Luv color space rather than RGB color space, because Euclidean metrics and distances are perceptually uniform in Luv color space, which is not the case in RGB color space [23].

Let $\Gamma(x,y) : \mathbb{R}^3$ be a color image, based on classical Riemannian geometry results [21]. The L_2 norm can be written in matrix form

$$d\Gamma^2 = \begin{bmatrix} dx \\ dy \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} \qquad (8)$$

where $g_{11} = [\partial\Gamma/\partial x]^2$, $g_{12} = g_{21} = (\partial\Gamma/\partial x)(\partial\Gamma/\partial y)$, $g_{22} = [\partial\Gamma/\partial y]^2$.

The quadratic form (8) achieves its extrema changing rates in the directions of the eigenvectors of matrix $[g_{i,j}]$, $i = 1, 2$, $j = 1, 2$, and the changing magnitude is decided by its eigenvalues λ_+ and λ_-. In our approach, we select $\sqrt{\lambda_+ - \lambda_-}$ [23], [50] to define the color gradient

$$\nabla\Theta = \sqrt{\lambda_+ - \lambda_-} \qquad (9)$$

where

$$g_{11} = \begin{bmatrix} \left| \frac{\partial L}{\partial x} \right|^2 \\ \left| \frac{\partial u}{\partial x} \right|^2 \\ \left| \frac{\partial v}{\partial x} \right|^2 \end{bmatrix}, \quad g_{22} = \begin{bmatrix} \left| \frac{\partial L}{\partial y} \right|^2 \\ \left| \frac{\partial u}{\partial y} \right|^2 \\ \left| \frac{\partial v}{\partial y} \right|^2 \end{bmatrix} \qquad (10)$$

$$g_{12} = g_{21} = \begin{bmatrix} \left| \frac{\partial^2 L}{\partial x \partial y} \right| \\ \left| \frac{\partial^2 u}{\partial x \partial y} \right| \\ \left| \frac{\partial^2 v}{\partial x \partial y} \right| \end{bmatrix} \qquad (11)$$

where L, u, v correspond to the three channels in Luv color space.

B. L_2E Robust Estimation

In the previous section, we defined the color gradient in Luv color space in order to replace the gray level gradient of the original GVF snake. Although the capture range of the original GVF snake is large, it still may fail to find the edges of the object when given unsuitable initial positions. For example, to capture the boundary of a circular object, the initial snake must include the center of the circle [34]. Another potential difficulty relates to computational complexity. Given an initial location far away from the object, the original GVF snake may take a long time to converge to the edges of the object. Other challenges arise from the diversity and complexity of the images under study.

$$\Psi = \int \int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy + \left(\left| \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\} \right|^2 \cdot \left| \Theta - \nabla \left\{ G_{\sigma(x,y)} * f(x,y) \right\} \right|^2 \right) dx dy \qquad (7)$$

The current "ground-truth" database of the IGDS system contains 1791 imaged cells corresponding to 58 different cases, collected from three different institutions (University of Pennsylvania School of Medicine, University of Medicine and Dentistry of New Jersey, and City of Hope National Medical Center). There were obvious variations in the staining characteristics of specimens among the institutions which had been introduced by differences in manufacturers of dyes, automated stainer choices, and overall intensity variations. All of these variables led to variations in shadowing, shading, contrasts, and highlighting cues. In addition, many of the specimens also contained artifacts (flattened and crushed cells; debris) which had unintentionally been introduced during specimen preparation. Because of these challenges, it was necessary to introduce a robust nucleus and cytoplasm classifier which could reject "outliers" and provide accurate estimations of initial positions.

One of the common approaches used to resolve this sort of classification problem is the least-square (LS) approach. Define the image classification model as

$$\mathbf{g} = \mathbf{Fw} + \mathbf{e} \tag{12}$$

where $\mathbf{F}$ is the sample image, an $m \times 3$ matrix with $m = w \cdot h$ as the total pixels in the image, w is the width of the image, and h is the height of the image. Applying orthogonal principles, we obtain the LS solution

$$\mathbf{w} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{g}. \tag{13}$$

By choosing $\mathbf{g}$ as 1 and -1 to denote the background and the object, respectively, we provide an LS solution to the classification problem.

A total least square (TLS) performs better by minimizing the error between both $\mathbf{g}$, $\widehat{\mathbf{g}}$ and $\mathbf{F}$, $\widehat{\mathbf{F}}$. The weights $\mathbf{w}$ are given by the right singular vector corresponding to the smallest eigenvalue of the singular value decomposition (SVD) of $[\mathbf{F} \quad \mathbf{g}]$, which is the stacked data of input image $\mathbf{F}$ and output image $\mathbf{g}$. It can be shown [39, p. 383] that the solution is

$$\mathbf{w} = -\frac{\mathbf{v}_{n+1}(1:n)}{\mathbf{v}_{n+1}(n+1)} \tag{14}$$

where $\mathbf{v}_{n+1}$ corresponds to the right singular vector of the smallest singular value.

Because the L_2 norm is sensitive to outliers, neither LS nor TLS are robust estimators. In order to address the outliers and variabilities of the images, L_2E robust estimation and training have been proposed. By merging training and L_2E robust estimation with the color GVF snake, the segmentation accuracy improved significantly, while simultaneously saving computation time. In order to obtain a robust estimation of the initial locations of the nucleus and cytoplasm of cells without being
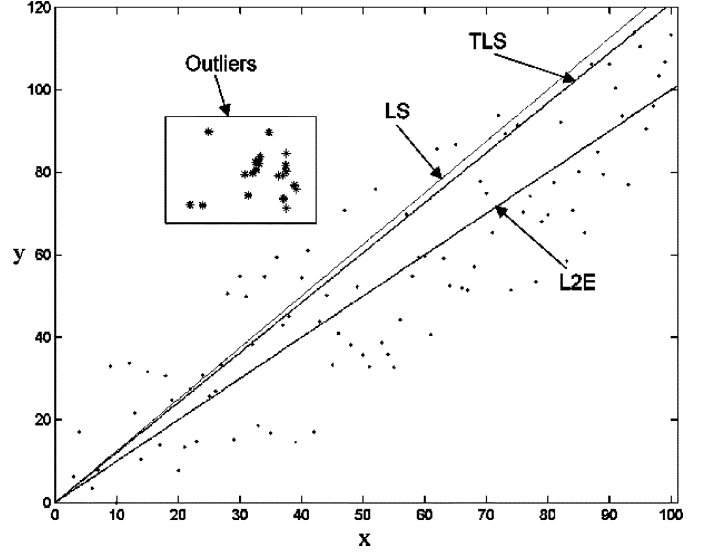

Fig. 2. Estimation results using LS, TLS, and L_2E for a linear regression problem of a dataset containing outliers.

affected by outliers, we adopt a statistically robust matching criteria based on the minimization of the integral squared error (ISE), also known as L_2E between a Gaussian model of the residual and the true density function of the residual. It was shown by Scott [25]–[27] that minimum distance estimators, including L_2E, are inherently robust without requiring the specification of tuning parameters.

Fig. 2 is a linear-regression example showing the capacity of L_2E robust estimation to reject outliers. One can easily see that the LS and TLS are perturbed by the outliers on the left upper side of Fig. 2, marked with a rectangle, while the L_2E-based fit has successfully rejected the outliers.

In (12), the residual $\mathbf{e}$ is assumed to be composed of independent, identically distributed random variables, whose distribution will be modeled by a Gaussian with mean $\mathbf{0}$ and variance σ. Our goal is to minimize the ISE or L_2E error measurement given by

$$\text{ISE}(\widehat{\boldsymbol{\theta}}) = E_{\widehat{\boldsymbol{\theta}}} \int [g(\mathbf{e}\,|\widehat{\boldsymbol{\theta}}) - h(\mathbf{e}\,|\boldsymbol{\theta})]^2 d\mathbf{e} \tag{15}$$

where $g(\cdot)$ is the Gaussian function modeling the density of the residual error, $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ are the estimation parameter and the true parameter, respectively, and h is the true unknown density of the residual error term. $E_{\widehat{\boldsymbol{\theta}}}(\cdot)$ denotes the estimate of the integral. Considering minimizing an estimate of ISE with respect to $\boldsymbol{\theta}$, see (16) at the bottom of the page.

The third term $h^2(\cdot)$ is independent of $\widehat{\boldsymbol{\theta}}$ and can be omitted from the minimization. The first term in the expansion is $E_{\widehat{\boldsymbol{\theta}}} \int g^2(\cdot)d\mathbf{e}$, and can be evaluated in closed form. The second

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\widehat{\boldsymbol{\theta}}} E_{\widehat{\boldsymbol{\theta}}} \int [g(\mathbf{e}|\widehat{\boldsymbol{\theta}}) - h(\mathbf{e}|\boldsymbol{\theta})]^2 d\mathbf{e}$$

$$= \arg\min_{\widehat{\boldsymbol{\theta}}} \left[E_{\widehat{\boldsymbol{\theta}}} \int g^2(\mathbf{e}|\widehat{\boldsymbol{\theta}})d\mathbf{e} - 2E_{\widehat{\boldsymbol{\theta}}} \int g(\mathbf{e}|\widehat{\boldsymbol{\theta}})h(\mathbf{e}|\boldsymbol{\theta})d\mathbf{e} + E_{\widehat{\boldsymbol{\theta}}} \int h^2(\mathbf{e}|\boldsymbol{\theta})d\mathbf{e} \right] \tag{16}$$

term is $-2E_{\widehat{\boldsymbol{\theta}}} \int g(\mathbf{e}|\widehat{\boldsymbol{\theta}})h(\mathbf{e}|\boldsymbol{\theta})d\mathbf{e}$, with $E_{\widehat{\boldsymbol{\theta}}}[g(\cdot)]$ being the expectation of $g(\cdot)$ with respect to $\widehat{\boldsymbol{\theta}}$. The following Rudemo unbiased estimator can be used for the second term [33]:

$$-\frac{2}{m}\sum_{i=1}^{m} g(e_i|\theta) \quad \text{for } i = 1, .., m \tag{17}$$

where $m = w \cdot h$ denotes the total number of pixels. Thus, the minimization using the L_2E criterion for normal density is given by

$$\widehat{\theta}_{L_2E} = \arg\min_{\theta=[\sigma,\mathbf{w}]} \left[\frac{1}{2\sqrt{\pi}\sigma} - \frac{2}{m}\sum_{i=1}^{m} \exp\left(-\frac{\|\mathbf{g}-\mathbf{Fw}\|^2}{2\sigma^2}\right) \right]. \tag{18}$$

The numerical solution is calculated by applying the gradient descent algorithm to (18). For readers who are interested in robust estimators, we should point out that L_2E robust estimation has many useful applications, such as image registration [28], computer latency and workload prediction [29], applied statistics [32], [33], and data mining [31].

C. Image-Segmentation Approaches

In this subsection, several state-of-the-art approaches, such as FLD classification, watershed image segmentation, and mean-shift robust clustering algorithm, are described, and the proposed robust color GVF snake is described in detail. Experiments comparing the performance of these algorithms with the robust color GVF snake follow in the next section.

1) Fisher Linear Discriminant: One of the common approaches in pattern recognition is FLD [59]. FLD can also be used in image-segmentation problems by treating segmentation as classification. If we form a linear combination of the pixels $P_{ij} \in \mathbf{F}$, we obtain a scalar dot product

$$\mathbf{g} = \mathbf{Fw} \tag{19}$$

and a corresponding set of $g_{ij} \in \mathbf{g}$ divided into subsets g_1 and g_2, which correspond to the objects of interests and background, respectively. FLD is the algorithm used to maximize the ratio of between-class scatter to within-class scatter. Suppose we have $\mathbf{p}$ as a 3-D vector that belongs to $\mathbf{F}$. For a binary classification problem, where $c = \{D_1, D_2\}$, the result can be immediately determined for $\mathbf{w}$

$$\mathbf{w} = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{20}$$

where

$$S_w = \sum_{i=1}^{2}\sum_{\mathbf{p}\in D_i}(\mathbf{p}-\mathbf{m}_i)(\mathbf{p}-\mathbf{m}_i)^T \tag{21}$$

where $\mathbf{m} = (1/n)\sum_{\mathbf{p}\in c}\mathbf{p}$ is the total mean vector, and $\mathbf{m}_i = (1/n_i)\sum_{\mathbf{x}\in D_i}\mathbf{x}$ is the ith within-class mean vector, and n_i denotes the number of pixels in class D_i. The result is the same as LS in this special case. For those who are interested in the details, we refer readers to [37, p. 243].

2) Watershed: Another popular image-segmentation algorithm is the watershed transform [51]. This algorithm is derived from the field of mathematical morphology. The term watershed refers to a ridge that divides areas drained by different river systems. A catchment basin is the geographical area draining into a river or reservoir. The gradients of the image are used to compute watershed transform, and therefore, the catchment basin boundaries are located on the maximum gradient points.

A good review of different definitions and implementations of the watershed transform can be found in [52]. In our experiments, we adopted the watershed definition based on topographical distance applied to discrete images. Let $\mathbf{F}$ be an original image. The *lower slope* LS($\mathbf{p}$) of $\mathbf{F}$ at a pixel $\mathbf{p}$, is defined as the maximal slope linking $\mathbf{p}$ to any of its neighbors of lower altitude

$$\text{LS}(\mathbf{p}) = \max_{q \in N_G(\mathbf{p})\cup\mathbf{p}} \frac{\mathbf{F}(\mathbf{p})-\mathbf{F}(\mathbf{q})}{d(\mathbf{p},\mathbf{q})} \tag{22}$$

where $N_G(\mathbf{p})$ is the set of neighbors of pixel $\mathbf{p}$ on the grid G, and $d(\mathbf{p},\mathbf{q})$ is the Euclidean distance between two pixels $\mathbf{p}$ and $\mathbf{q}$. The corresponding definition of *lower neighbors* and topographical distance can be found in [52].

The watershed in the original form often produces an over-segmentation of images. Many small catchment basins are produced, due to the large number of local minima in the input image. Several improved marker-based watershed algorithms [54]–[56] have been proposed to remedy this issue. The basic idea behind the marker-based segmentation is to transform the input image in such a way that the watershed of the transformed image corresponds to meaningful object boundaries. The technique for filtering the image is the marker imposition [57], and is based on the morphological reconstruction algorithm [58]. In our experiments, both the original watershed algorithm using the Vincent–Soille algorithm [53] and the marker-based watershed algorithm [57] are implemented for comparison with the robust color GVF snake.

3) Mean-Shift Robust Clustering: Mean-shift algorithm is a general nonparametric technique proposed for clustering of a complex multimodel feature space [17]. It randomly tessellates the space with search windows, and moves the windows until convergence is achieved at the nearest mode of the underlying probability distribution.

In our implementation, the RGB input vectors are first converted into Luv vectors through a nonlinear transformation to produce perceptually uniform Euclidean distances and metrics. A set of m points $x_1, \ldots, x_m$ called the sample set are then randomly selected from the data with the radius of a searching sphere $S_h(x)$. The mean-shift procedure is then applied to each point in the sample set. The mean-shift vector at the point $\mathbf{x}$ is defined as [38, p. 534]

$$M_h(\mathbf{x}) = \frac{1}{n_\mathbf{x}}\sum_{\mathbf{x}_i \in S_h(\mathbf{x})} \mathbf{x}_i - \mathbf{x} \tag{23}$$

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 has the direction of the gradient density estimate $\widehat{\nabla}f(\mathbf{x})$

$$M_h(\mathbf{x}) = \frac{h^2}{5}\frac{\widehat{\nabla}f(\mathbf{x})}{\widehat{f}(\mathbf{x})} \tag{24}$$

where the kernel is defined as

$$K_E(\mathbf{x}) = \begin{cases} \frac{5}{2}c_3^{-1}(1-\mathbf{x}^T\mathbf{x}), & \text{if } \mathbf{x}^T\mathbf{x} \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

where c_3 is the volume of the unit sphere in the 3-D space. Pointing toward 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. For more details about mean-shift algorithm, please refer to [17].

4) Robust Color GVF Snake: Now we describe the design and development of the robust color GVF snake. Rewriting (6) and replacing the original GVF vector field Θ with the new color GVF vector field Θ in Luv color space, we have

$$E_{\text{snake}} = \alpha v_{ss}(s) - \beta v_{ssss}(s) + \Theta \tag{26}$$

where Θ is the color GVF defined as $\Theta(x, y) = [u(x, y), v(x, y)]$ that minimizes the energy function defined in (7).

In the robust color GVF snake, ∇ in (7) is the Luv color gradient defined in (9). The two initial snake contour locations are obtained from L_2E robust estimation, and correspond to the boundaries of the nucleus and the cytoplasm, respectively. By applying calculus of variations, it can be shown that minimizing the integral in (7) is equal to solving the following equation:

$$\mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \tag{27}$$
$$\mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0. \tag{28}$$

The numerical solution to (27) and (28) can be solved by treating u and v as functions of time, and solving

$$
\begin{aligned}
u_t(x, y, t) = {} & \mu \nabla^2 u(x, y, t) \\
& - (u(x, y, t) - f_x(x, y))(f_x^2(x, y) + f_y^2(x, y))
\end{aligned}
\tag{29}
$$

$$
\begin{aligned}
v_t(x, y, t) = {} & \mu \nabla^2 v(x, y, t) \\
& - (v(x, y, t) - f_y(x, y))(f_x^2(x, y) + f_y^2(x, y)).
\end{aligned}
\tag{30}
$$

In order to set up an iterative solution, let the indexes i, j, and n corresponding to x, y, and t be derived from the iterative algorithm for (29) and (30).

The steps associated with the automatic segmentation, indexing, and archiving of the IGDS system using robust color GVF snake are as follows.

1) The pathologist selects an imaged cell from the data set for training.
2) The training image is mapped from RGB color space to Luv color space.
3) L_2E robust estimation is first used to separate the imaged specimen into object Y_o and background Y_b. The parameters are recorded as $w_{\text{cytoplasm}}$. The object Y_o contains the nucleus and cytoplasm of the cells. The cytoplasm of the cells, as a general rule, surrounds the nucleus; therefore, this step is actually the initial estimation of cytoplasm contour. The cytoplasm of the cell and the background are roughly separated in the same step. The L_2E robust estimation is then applied within the roughly delineated object region Y_o and separates it into cytoplasm C and nucleus N. This step roughly separates the cytoplasm and nucleus regions to denote the initial guess of the nuclear contour. The parameters are recorded as w_{nucleus}.
4) Both $w_{\text{cytoplasm}}$ and w_{nucleus} are applied to process all the cells from a given case, and the resulting contours are used as the initial positions to start the double robust color GVF snakes. Because the initial positions are close to the accurate positions, they can converge quickly and provide clear, inherently connected, and smooth contours.
5) In the current stage of development, the segmentation results are displayed for pathologists to review before actually becoming part of the "ground-truth" database. The cell image, the geometrical features of the cells, the mask of nucleus and cytoplasm, the texture features of the cells, and the cytoplasm/nucleus ratios are used to generate a record and indexed into the database for future query.

IV. EXPERIMENTAL RESULTS

In the experiments, pathologists were asked to confirm the segmentation results for a mixed set of 58 lymphoproliferative cases. The cell types included a mixed set of MCL, CLL, FCC, and normal (BENIGN). The imaged cells were collected from the Hospital of the University of Pennsylvania, Philadelphia, PA; Robert Wood Johnson University Hospital, New Brunswick, NJ; and City of Hope National Medical Center, Duarte, CA.

The slides were prepared using standard peripheral blood-collection techniques, wherein a drop of blood is placed on the glass slide and smeared in a thin film by using an automatic slide maker. The smear is then air dried and stained with Wright–Giemsa stain, using the usual staining protocol for hematology specimens with mixtures of basic (methylene blue derivatives) and acid dyes (usually eosin). According to the number of acid and basic groups present, cell components take up the dyes from the mixture in various proportions. Different cells were stained with different hues depending on their composition (in proteins, amino acids, enzymes, etc.). However, for a particular cell type, the staining quality is generally maintained.

The test platform for the experiments consisted of an Intel-based workstation interfaced with a high-resolution Olympus DP70 camera, equipped with 12-bit color depth on each color channel and 1.45 million pixel effective resolution. The system also includes a single 2/3-in charge-coupled device (CCD) digital camera, and an Olympus AX70 microscope equipped with a Prior 6-way robotic stage, motorized objective turret, and a magnification changer.

A. Experimental Results of L_2E Robust Estimation

In this section, we first present the comparative results of the L_2E robust estimation, LS and TLS. In every row of Fig. 3, the first image is the original image, and the second is the image with the estimated initial contours imposed on it. The next two images in the same row are the estimated contours of the nucleus and the cytoplasm. Fig. 3(a) shows the "ground-truth" contours delineated by human experts. The outer contour is the curve that surrounds the cytoplasm. The inner contour corresponds to the nuclear region. Fig. 3(b)–(d) show the initial positions of the contours calculated based on LS, TLS, and L_2E robust estimation, respectively.

Evaluated from the preprocessing results, the L_2E robust estimation provides the most accurate estimation results when
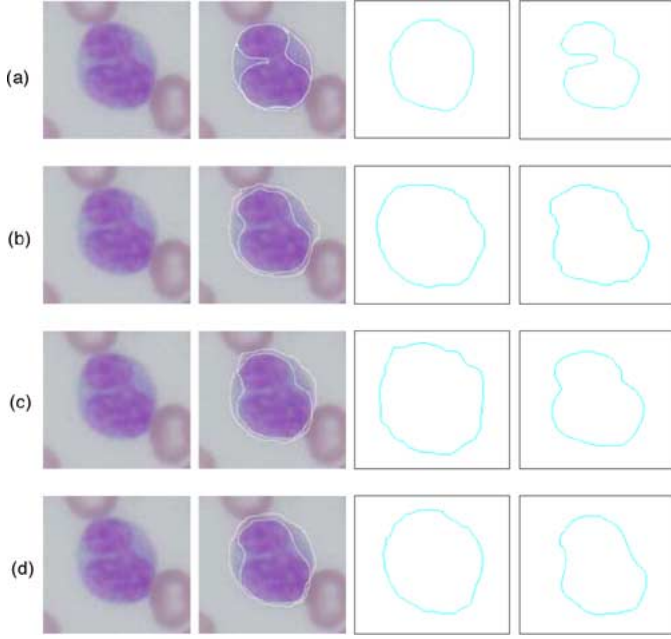
Fig. 3. Estimation results of the initial locations of the cytoplasm and nucleus of an FCC image. (a) "Ground-truth" contours delineated by human experts. (b) Estimation results using LS. (c) Estimation results using TLS. (d) Estimation results using L_2E robust estimation.
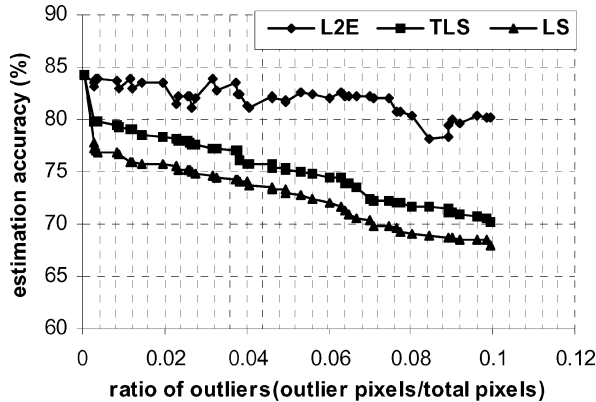

Fig. 4. Relationship between the estimation accuracy of LS, TLS, L_2E robust estimation, and the number of outliers in the image.

compared with LS and TLS. To quantify the difference between the true boundaries M and the estimation results S, the segmentation accuracy rates are defined as

$$\varepsilon = \frac{A(S) \cup A(M) - A(S) \cap A(M)}{A(M)} * 100\% \qquad (31)$$

where $A(M)$ and $A(S)$ are the areas enclosed within M and S. ε is the cytoplasm accuracy rate (CAR) or nucleus accuracy rate (NAR).

Fig. 4 illustrates the average error ration of CAR and NAR for LS, TLS, and L_2E robust estimation. If there are no outliers in the image, all methods exhibit the same results with about 85% estimation accuracy. As the ratio of outliers increases, L_2E robust estimation provides a stable and low error rate, which is more robust than LS and TLS.

The next set of experiments were performed to determine whether the number of training images impacted the estima-
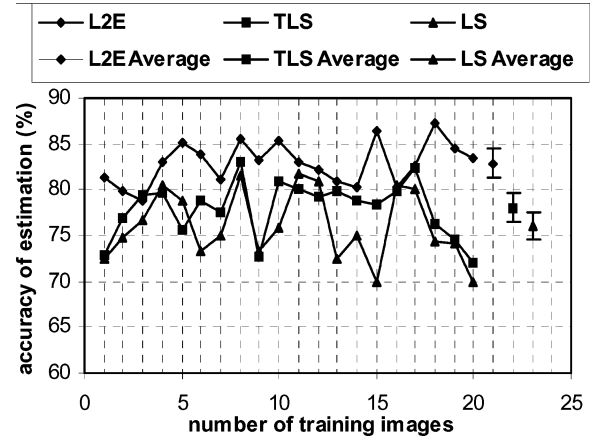

Fig. 5. Relationship between the estimation accuracy and the number of imaged cells used in the training set.

tion accuracy rates. Fig. 5 shows the relationship between the number of training image and average estimation accuracy rates. No significant benefits were observed by increasing the size of the training set. This is probably due to the fact that images originating from a given case exhibit similar imaging conditions and reflection properties. Based upon the result of the pilot study, pathologists were asked to choose only a single representative image from each case for training.

B. Segmentation Results of Robust Color GVF Snake

In this section, the experimental results are presented for studies comparing the performance of several state-of-the-art image-segmentation algorithms and the robust color GVF snake. In Fig. 6, the performance of the newly proposed robust color GVF snake is compared with the FLD, the watershed segmentation algorithms, the mean-shift clustering algorithm used in the prototype of the IGDS system, and the original GVF snake. Fig. 6(a) is an FCC cell image with the nucleus and the cytoplasm. Fig. 6(b) shows the segmentation results of FLD. The *unsupervised* segmentation results using the mean-shift algorithm was shown in Fig. 6(c). Fig. 6(d) shows the results of the original watershed algorithm. Fig. 6(e) shows the results of the improved marker-based watershed algorithm. The *supervised* segmentation results of mean-shift after region merging through human interventions are shown in Fig. 6(f). The segmentation result of running the original GVF snake in RGB color space after 175 iterations is shown in Fig. 6(g). The proposed robust color GVF snake operating in Luv color space converges to the objects in just 20 iterations, as shown in Fig. 6(h).

The results show that the traditional GVF snake, while quite useful for its original intended purpose, is not effective in segmenting the hematopathology images without modification. Using the traditional approach may lead the snake to converge to locations exhibiting highest luminance, i.e., the red blood cells, which should be classified as background for our purpose. At the same time, this approach requires a longer time to converge than other methods. FLD provides unacceptable segmentation results, because the linear discriminant is not suitable for separating nonlinear separable datasets. The
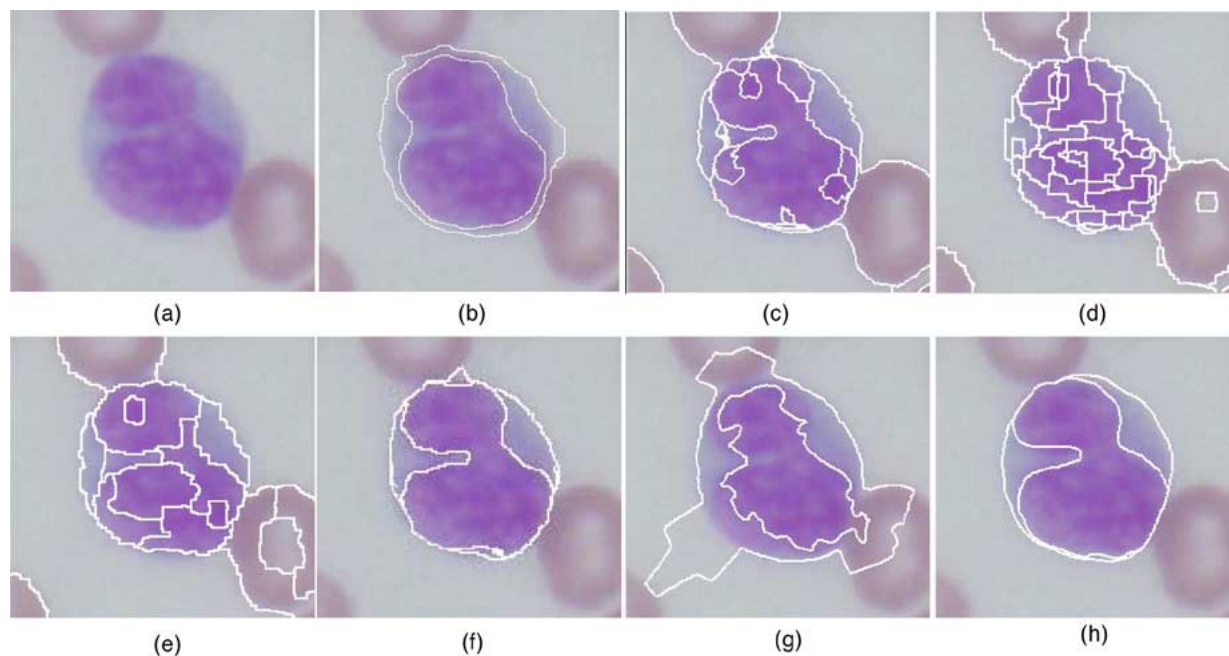
Fig. 6.   Segmentation results of (a) FCC applying (b) FLD. (c) Unsupervised mean-shift algorithm. (d) Unsupervised traditional watershed algorithm. (e) Unsupervised marker-based watershed algorithm. (f) Supervised mean-shift algorithm after region merging. (g) Original GVF snake. (h) Robust color GVF snake.

mean-shift algorithm segmented every subregion accurately, including even the most pronounced concave region. However, to achieve accurate results, as illustrated in Fig. 6(f), human intervention is required for manually adjusting parameters and merging the oversegmented regions. The original watershed algorithm oversegmented the image into a myriad of small regions, rendering its usage impractical for the application at hand. One of the main contributing factors leading to the oversegmentation is the fact that all of the images in our application exhibit high texture variations throughout both the nucleus and the cytoplasm. Marker-based watershed algorithms [54]–[56] have been proposed to resolve oversegmentation problems. In our experiments, it did reduce the number of oversegmented regions significantly, but still required region merging to provide useful results. Achieving unsupervised region merging is not trivial, because it is not easy to find a general criterion to decide which regions should be merged and which should not. For example, two separate small regions, even if they have similar pixel values, may actually belong to a single object or multiple objects. There is some recent research which focuses on the region-merging problem [57], [60]. Aside from the difficulty presented by oversegmentation, the contours delineated by the watershed approach are not as smooth as those produced using the robust color GVF snake. In our experiments, the robust color GVF snake provided the most accurate segmentation results, while maintaining smooth boundaries.

An advantage of using L_2E robust estimation is the significant improvement in the speed of convergence. The current system is running with Java sdk 1.4.3 on a PC with an Intel 2.4G processor and 512-MB RAM. It took 47 s to preprocess 100 640×480 images while automatically selecting the ROI, each with a dimension of around 180×180. ROI selection was the first step for every segmentation algorithm used in the experi-

ments. After this step, it took 43 s to segment both the nucleus and cytoplasm of the 100 imaged cells using the robust color GVF snake, compared with 245 s using original GVF snake, 73 s with mean-shift algorithm, 98 s using the marker-based watershed algorithm, and 28 s applying FLD. Although FLD provides the fastest speed, its segmentation results are unacceptable for our purpose. The robust color GVF snake provided the fastest speed of convergence among these methods, while providing accurate segmentation results.

Fig. 7 shows segmentation results using the robust color GVF snake to segment different types of cells. In each row, the first image is the original image and the second is the initial positions calculated by L_2E robust estimation. The next two images in the same row show iterative intermediate results (after 5 and 10 iterations, respectively), while the last image is the final output (after 20 iterations). Fig. 7(a) shows a typical FCC cell. While each disorder under study exhibited a spectrum of morphologies, FCC cells tend to exhibit pronounced cleaves in the nucleus and relatively small quantities of cytoplasm. Fig. 7(b) shows an MCL, which has a small cleave in the cell and almost no cytoplasm. Fig. 7(c) shows CLL, which is similar to the MCL example, but contains more cytoplasm. Fig. 7(d) shows a normal cell (BENIGN), which has a relatively larger cytoplasm and regularly shaped nucleus. Fig. 8 shows some results of more difficult cell cases using the robust color GVF snake, where (a) and (c) are original images, and (b) and (d) exhibit corresponding segmentation results. Some images exhibit very weak differences between cytoplasm and background. Some images have complex texture or even several different color regions in the cytoplasm.

In the next set of experiments, we compared the performance for all 58 cases within the "ground-truth" database: nine cases of FCC (total of 315 cell images), 18 cases of MCL (total of 543
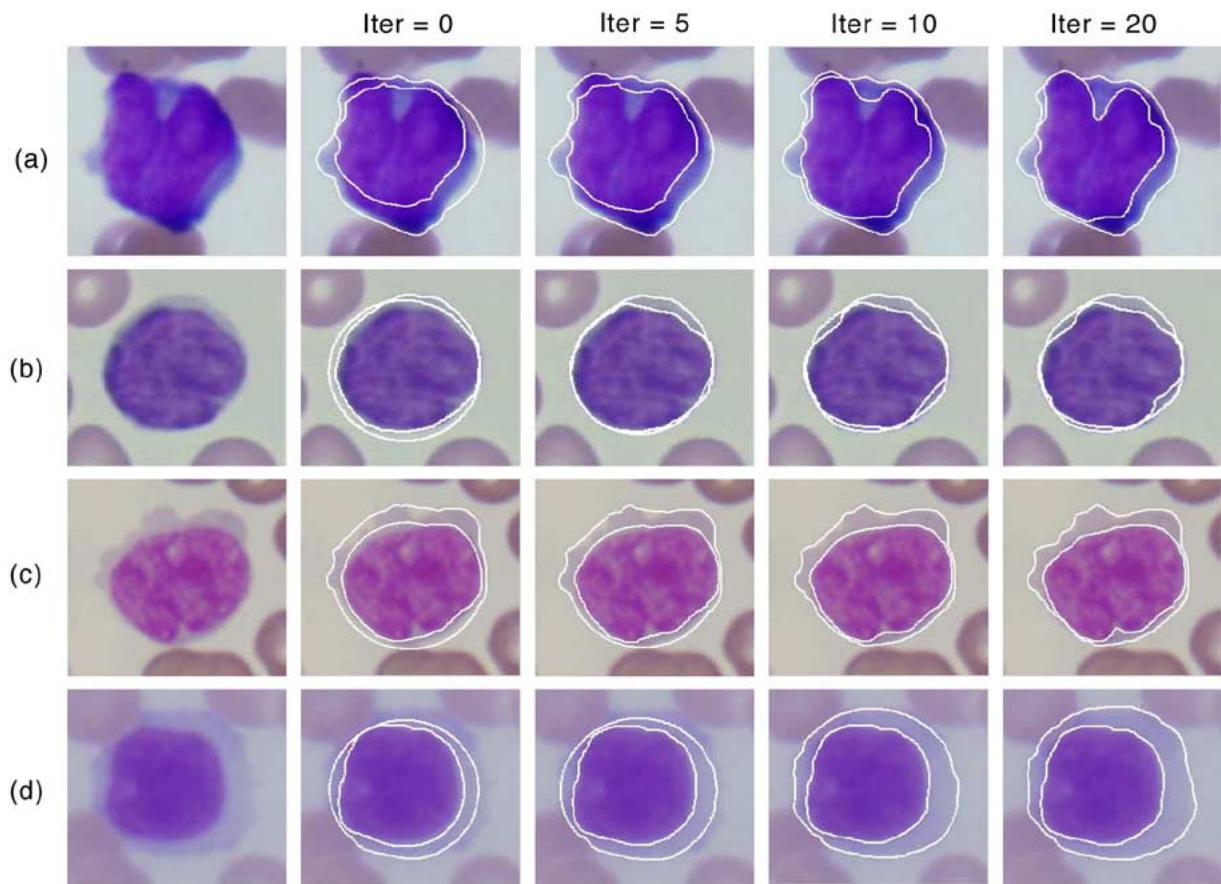
Fig. 7. Image-segmentation results applying robust color GVF snake on four different types of cells. (a) FCC. (b) MCL. (c) CLL. (d) Normal cells (BENIGN).
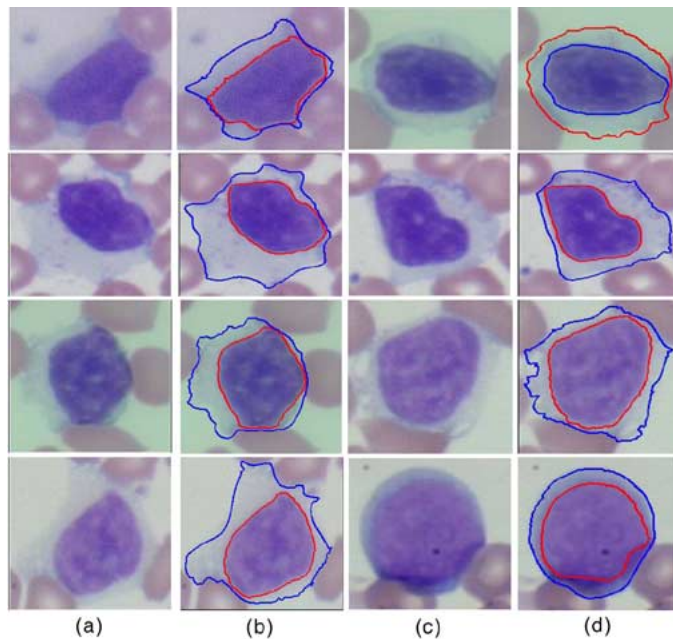

Fig. 8. Additional image-segmentation experimental results applying robust color GVF snake. (a) and (c) are the original images. (b) and (d) are the corresponding segmentation results.

Fig. 9. Data distribution of the "ground-truth" database with four cell types, 58 specimen cases, and 1791 images.

images), 15 cases of Benign (total of 536 cell images), and 16 cases of CLL (total of 397 cell images). The distribution of the cases in the database is listed in Fig. 9. We limited the comparative performance studies to one of the *supervised* algorithms
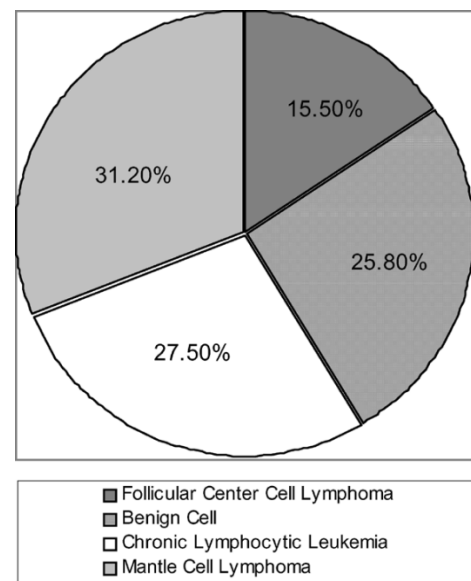
and *unsupervised* robust color GVF snake, in order to determine how closely the *unsupervised* approach compared with the *supervised* segmentation results. In the experiment, the *supervised* mean-shift segmentation results are obtained after postprocessing is performed to merge oversegmented subregions within the image.

TABLE I
COMPARATIVE SEGMENTATION RESULTS OF UNSUPERVISED ROBUST COLOR GVF SNAKE AND SUPERVISED MEAN-SHIFT ALGORITHM

Cell Type	Number of Cases	Number of Images	SNAR	SCAR	MNAR	MCAR
FCC	9	315	91.20	89.57	92.12	87.32
MCL	18	543	90.09	87.92	91.03	89.89
CLL	16	397	93.56	91.83	93.19	90.73
Benign	15	536	97.10	92.94	96.04	94.98

The nuclear segmentation accuracy ratio based on robust color GVF snake (SNAR).
The cytoplasm segmentation accuracy ratio based on robust color GVF snake (SCAR).
The nuclear segmentation accuracy ratio based on supervised mean-shift algorithm (MNAR).
The cytoplasm segmentation accuracy ratio based on supervised mean-shift algorithm (MCAR).
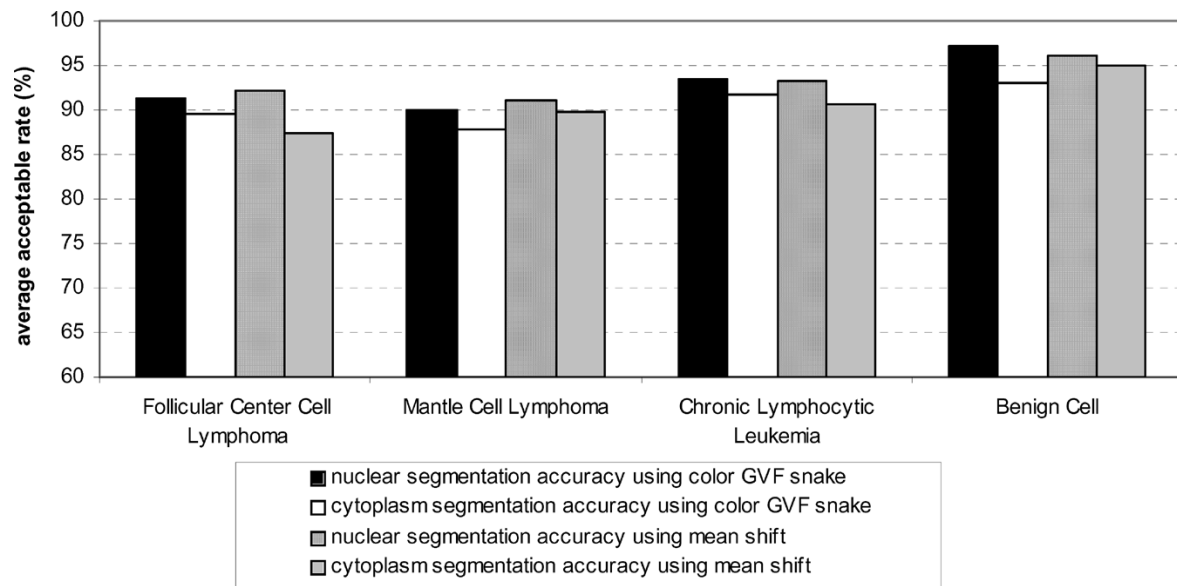

Fig. 10. Average segmentation-error rates of four different cell types, 58 specimen cases, and 1791 cell images in the "ground-truth" database using the robust color GVF snake and supervised mean-shift clustering algorithm, respectively.

Table I provides the error rates of the segmentation results obtained from the *unsupervised* robust color GVF snake and the *supervised* mean-shift algorithm. The database used for these studies contained around 100 images per case, with two cases having only 10 images. For each case, a single imaged cell was selected as the training image. The nuclear and the cytoplasm segmentation-accuracy ratios were calculated based on the remaining imaged cells in the same data set. MNAR is the mean-shift nucleus segmentation-accuracy ratio, and MCAR denotes the mean-shift cytoplasm segmentation-accuracy ratio. Similarly, SNAR and SCAR, defined in (31), are the nucleus and cytoplasm segmentation-accuracy ratios using the robust color GVF snake, respectively. The average of these four error measurements for all the cell images in the "ground-truth" database are graphically shown in Fig. 10.

From these sets of experiments, it was shown that the *unsupervised* robust color GVF snake provided comparable segmentation results with those achieved using the *supervised* mean-shift algorithm, which requires considerable human intervention. The *unsupervised* segmentation results of the color GVF snake were deemed satisfactory by a panel of experts, while significantly accelerating the rate of processing and decreasing the workload introduced by *supervised* approaches.

V. CONCLUSION

The results of the robust color GVF snake were obtained by exploiting prior knowledge of the specific application. In this particular microscopy application, there are two ROI: the nuclei and the cytoplasm, with the cytoplasm of the cells, as a general rule, surrounding the nucleus. Therefore, the robust color GVF snake, like many other high-level image-segmentation approaches, is tuned to provide a reliable means for segmenting imaged hematopathology specimens. The mean-shift algorithm and watershed segmentation algorithms, on the other hand, provide a general solution applicable to a wider range of image-segmentation problems, since they rely exclusively on pixel attributes. It should be emphasized that throughout our studies, the pixel-based strategies consistently gave rise to oversegmented images, which required considerable human intervention to achieve the desired results. Watershed algorithms do, however, offer the potential of resolving issues related to touching cells. We are currently investigating how best to incorporate this desirable feature while avoiding the stated oversegmentation problem. The watersnake [61] approach belongs to this class of algorithms, and by introducing the snake's smoothing energy into the watershed, it is able to provide wa-

tershed-based segmentation results exhibiting smooth contours with less oversegmented regions. Another set of interesting and effective approaches are active contours without edges [65] or its more general model, the region-based snakes [62]–[64]. Region-based snakes offer insensitivity to initialization and can be made computationally efficient for practical use. Combining region-based snakes and robust estimation is a potentially promising area for exploration that we plan to pursue as our experiments move toward the analysis of imaged tissues in which there are few distinct edges.

Although L_2E robust estimation is a new powerful tool which has attracted interest throughout the research community in statistics [30]–[32], to our knowledge, it has not been used in conjunction with active contours for image segmentation. By merging L_2E robust estimation with the GVF snake and using a Luv color gradient, we have developed a quick, robust approach which was shown to produce reliable results for detecting and delineating the boundaries of imaged cells. Since the robust color GVF snake was able to provide satisfactory performance, even when confronted with images exhibiting weak contrast and subtle edges, we have begun to investigate the use of parallel, robust color GVF snakes for evaluating tissue microarrays.

REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Computer Vis.*, vol. 1, pp. 321–331, 1987.

[2] D. Terzopoulos, "The computation of visible-surface representations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 4, pp. 417–438, Apr. 1988.

[3] T. McInerney and D. Terzopoulos, "A dymamic finite element surface model for segmentation and tracking in multidimensional medical images with applications on cardiac 4D image analysis," *Comput. Med. Imag. Graphics*, vol. 19, pp. 69–83, 1995.

[4] F. Leymarie and M. D. Levine, "Tracking deformable objects in the plane using an active contour model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 6, pp. 617–634, Jun. 1993.

[5] R. Durikovic, K. Kaneda, and H. Yamashita, "Dynamic contour: A texture approach and contour operations," *The Visual Computer*, vol. 11, pp. 277–289, 1995.

[6] D. Terzopoulos and R. Szeliski, "Tracking with Kalman snakes," in *Active Vision*, A. Blake and A. Yuille, Eds. Cambridge, MA: MIT Press, 1992, ch. 1, pp. 3–20.

[7] T. F. Cootes, C. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. 5th Eur. Conf. Computer Vis.*, vol. 2, H. Brukhardt and B. Neumann, Eds., Berlin, Germany, 1998, pp. 484–498.

[8] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," in *Proc. 8th Machine Vis. Conf.*, A. Clarke, Ed., Essex, U.K., Sep. 1997, pp. 110–119.

[9] V. Caselles, F. Catte, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numer. Math.*, vol. 66, pp. 1–32, Oct. 1993.

[10] A. A. Amini, T. E. Weymouth, and R. C. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 9, pp. 855–867, Sep. 1990.

[11] D. J. Williams and M. Shah, "A fast algorithm for active contoursand curvature estimation," *Computer Vis., Graph., Image Process.: Image Understanding Arch.*, vol. 55, no. 1, pp. 14–26, 1992.

[12] L. D. Cohen, "On active contour models and balloons," *Computer Vis., Graph., Image Process.: Image Understanding Arch.*, vol. 53, no. 2, pp. 211–218, Mar. 1991.

[13] L. D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-D and 3-D images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 11, pp. 1131–1147, Nov. 1993.

[14] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[15] C. Xu and J. L. Prince, "Generalized gradient vector flow external forces for active contours," *Int. J. Signal Process.*, vol. 71, no. 2, pp. 131–139, Dec. 1998.

[16] D. Comaniciu, P. Meer, and D. J. Foran, "Image-guided decision support system for pathology," *Machine Vis. Applic.*, vol. 11, no. 4, pp. 213–224, 1999.

[17] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[18] D. J. Foran, D. Comaniciu, P. Meer, and L. A. Goodell, "Computer-assisted discrimination among malignant lymphomas and leukemia using immunophentyping, intelligent image repositories and telemicroscopy," *IEEE Trans. Inf. Technol. Biomed.*, vol. 4, no. 4, pp. 265–273, Dec. 2000.

[19] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic, pp. 1084–2000.

[20] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York: Interscience, 1953, vol. 1.

[21] E. Kreyszig, *Differential Geometry*. New York: Dover, 1991.

[22] S. D. Zenzo, "A note on the gradient of a multi-image," *Computer Vis., Graph. Image Process.*, vol. 33, pp. 116–125, Jan. 1986.

[23] G. Sapiro and D. L. Ringach, "Anisotropic diffusion on multivalued images with applications to color filtering," *IEEE Trans. Image Process.*, vol. 5, no. 11, pp. 1582–1586, Nov. 1996.

[24] T. Gevers, "Adaptive image segmentation by combining photometric invariant region and edge information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 6, pp. 848–852, Jun. 2002.

[25] D. W. Scott, "Parametric modeling by minimum L_2 error," Rice Univ., Dept. Statist., Houston, TX, Tech. Rep. 98–3, 1998.

[26] ——, "Parametric statistical modeling by minimum integrated square error," *Technometrics*, vol. 43, pp. 274–285, Aug. 2001.

[27] D. W. Scott, "Remarks on fitting and interpreting mixture models," *Computer Sci. Statist.*, vol. 31, pp. 104–109, Jun. 1999.

[28] J. Liu, B. C. Vemuri, and J. L. Marroquin, "Local frequency representation for robust multi-modal image registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 5, pp. 462–469, May 2002.

[29] L. Yang, J. Liu, L. R. Welch, and C. D. Cavanaugh, "A robust QoS forecasting technique for dynamic distributed real-time testbed," in *Proc. IEEE Int. Workshop Computer Architect. Machine Perception*, vol. 1, May 2003, pp. 35–43.

[30] N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York: Wiley, 1998.

[31] D. Kim and B. Bodt, "Building a robust explanatory model from a large dataset," in *Proc. 9th U.S. Army Conf. Appl. Statist.*, vol. 2, Oct. 2003, pp. 98–108.

[32] B. J. Gajewski, L. R. Rilett, M. P. Dixon, and C. H. Spiegelman, "Robust estimation of origin-destination matrices," *J. Transport. Statist.*, vol. 5, pp. 35–49, 2002.

[33] M. Rudemo, "Empirical choice of histogram and kernel density estimators," *Scand. J. Statist.*, vol. 1, pp. 65–78, Sep. 1982.

[34] N. Ray, S. T. Acton, T. Altes, and E. E. de Lange, "MRI ventilation analysis by merging parametric active contours," in *Proc. IEEE Int. Conf. Image Process.*, 2001, pp. 861–864.

[35] W. J. J. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin, Germany: Springer, 1988.

[36] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 2000.

[37] R. O. Duda, P. E. Hart, and D. G. Stork, *Patten Classification*, 2nd ed. New York: Wiley, 2000.

[38] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

[39] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[40] J. Garcia-Conde and F. Cabanillas, "Mantal cell lymphoma: A lympho-proliferative disorders associated with aberrant function of the cell cycle," *Leukemia*, vol. 10, pp. s78–s83, 1996.

[41] G. Vadlamudi, "Leukemic phase of mantal cell lymphoma: Two case reports and review of the literature," *Arch. Pathology Lab. Med.*, vol. 120, pp. 35–40, 1996.

[42] J. Chan and P. M. Banks, "A revised European–American classification of lymphoid neoplasms proposed by the International Lymphoma Study Group–A summary version," *Amer. J. Clin. Pathology*, vol. 103, pp. 543–560, 1995.

[43] J. M. Bennett, D. Catovsky, and M. T. Daniel, "Proposals for the classification of acute leukemias," *Brit. J. Haemotol.*, vol. 33, pp. 451–458, 1976.

[44] I. Bauman, R. Nenninger, H. Harms, H. Zwierzina, K. Wilms, A. C. Feller, V. T. Meulen, and H. K. Muller-Hermelink, "Image analysis detects lineage–Specific morphologic markers in leukemic blast cells," *Amer. J. Clin. Pathology*, vol. 105, no. 1, pp. 23–30, 1995.

[45] S. W. Hur and J. Lillis, "Mongrel: Hybrid techniques for standard cell placement," in *Proc. IEEE Int. Conf. Computer-Aided Des.*, San Jose, CA, Nov. 5–9, 2000, pp. 165–170.

[46] J. M. Bennet, "The FAB/MIC/WHO proposals for the classification of the chronic lymphoid leukemias," *Rev. Clin. Exp. Hematol.*, vol. 6, no. 4, pp. 330–334, Dec. 2002.

[47] B. P. Nelson, D. Variakojis, and L. C. Peterson, "Leukemic phase of B-cell lymphomas mimicking chronic lymphocytic leukemia and variants at presentation modern pathology," *Modern Pathology*, vol. 15, no. 11, pp. 1111–1120, Nov. 2002.

[48] P. L. Cohen, P. J. Kurtin, K. A. Donovan, and C. A. Hanson, "Bone marrow and peripheral blood involvement in mantle cell lymphoma," *Brit. J. Haemotol.*, vol. 101, no. 2, pp. 302–310, May 1998.

[49] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 2, pp. 158–175, Feb. 1995.

[50] T. Gevers, S. Ghebreab, and A. W. M. Smeulders, "Color invariant snakes," in *Proc. 9th Brit. Machine Vis. Conf.*, P. Lewis and M. Nixon, Eds., 1998, pp. 659–670.

[51] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transform," in *Mathematical Morphology in Image Processing*, E. R. Dougherty, Ed. New York: Marcel Dekker, 1993, vol. 12, pp. 433–481.

[52] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, vol. 41, pp. 187–228, 2000.

[53] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[54] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical Morphology and its Applications to Image Processing*, J. Serra and P. Soille, Eds. Dordrecht, The Netherlands: Kluwer, 1994, pp. 69–76.

[55] A. N. Moga, "Parallel watershed algorithms for image segmentation," Ph.D. dissertation, Tampere Univ. Technol., Tampere, Finland, 1997.

[56] A. N. Moga and M. Gabbouj, "Parallel marker-based image segmentation with watershed transformation," *J. Parallel Distrib. Comput.*, vol. 51, no. 1, pp. 27–45, 1998.

[57] V. Grau, A. U. J. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 447–458, Apr. 2004.

[58] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Commun. Image Rep.*, vol. 1, no. 1, pp. 21–46, Sep. 1990.

[59] T. Cooke, "Two variations on Fisher's linear discriminant for pattern recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 2, pp. 268–273, Feb. 2002.

[60] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 11, pp. 1452–1458, Nov. 2004.

[61] H. T. Nguyen, M. Worring, and R. V. D. Boomgaard, "Watersnakes: Energy-driven watershed segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 3, pp. 330–342, Mar. 2003.

[62] T. F. Chan and L. A. Vese, "A level set algorithm for minimizing the Mumford–Shah functional in image processing," in *Proc. IEEE Workshop Variational, Level Set Methods*, 2001, pp. 161–171.

[63] R. P. Fedkiw, G. Sapiro, and C. Shu, "Shock capturing, level sets, and PDE based methods in computer vision and image processing: A review of Osher's contributions," *J. Computat. Phys.*, vol. 185, no. 2, pp. 309–341, Mar. 2003.

[64] S. Jehan-Besson, M. Gastaud, M. Barlaud, and G. Aubert, "Region-based active contours using geometrical and statistical features for image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Barcelona, Spain, 2003, pp. 643–646.

[65] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

Lin Yang received the B.S. degree in electrical and communication engineering in 1999, and the M.S. degree in signal and information processing in 2002, both from Xian Jiaotong University, Xian, P. R. China. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Rutgers University at New Brunswick, Piscataway, NJ.

His research interests include the development of robust image segmentation and registration, computer vision, statistical pattern recognition, content-based image retrieval, and intelligent large-scale multimedia databases.

Peter Meer (S'84–M'86–SM'95) received the Dipl. Eng. 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 an Assistant Research Scientist with 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 a Professor. He has held visiting appointments in Japan, Korea, Sweden, Israel, and France, and was on the organizing committees of numerous international workshops and conferences. His research interest is in application of modern statistical methods to image-understanding problems.

Dr. Meer was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE between 1998 and 2002, is a member of the Editorial Board of *Pattern Recognition*, and was a Guest Editor of *Computer Vision and Image Understanding* for a special issue on "Robust Statistical Techniques in Image Understanding." He is coauthor of an award-winning paper in *Pattern Recognition* in 1989, the Best Student Paper in the 1999 and the Best Paper in the 2000 IEEE Conferences on Computer Vision and Pattern Recognition.

David J. Foran (S'89–M'91) received the B.S. degree from Rutgers University, New Brunswick, NJ, in 1983, and the Ph.D. degree in biomedical engineering from the University of Medicine and Dentistry of New Jersey (UMDNJ), Rutgers University, Piscataway, NJ, in 1992.

He served as a Physics Instructor with New Jersey Institute of Technology, Newark, from 1984 to 1985, and was a Junior Scientist with Johnson & Johnson Research, Inc., North Brunswick, NJ, from 1986 to 1988. He received one year of postdoctoral training at the Department of Biochemistry at UMDNJ-Robert Wood Johnson Medical School (RWJMS) in 1993. He joined the faculty at RWJMS in 1994, where he is currently an Associate Professor of Pathology Radiology and the Director of the interdepartmental Center for Biomedical Imaging Informatics. He also serves as the Associate Director for Research for the university-wide Informatics Institute. He is a member of the graduate faculty in the program in Computational Molecular Biology and Genetics, and he is a Research Associate Professor at the Center for Advanced Information Processing, both at Rutgers University. His research interests include quantitative, biomedical imaging, computer-assisted diagnosis, and medical informatics.