A STATISTICAL METHOD FOR EFFICIENT SEGMENTATION OF MR IMAGERY*

J. KAUFHOLD†
Department of Biomedical Engineering, Boston University
44 Cummington St., Boston, MA 02215, USA

M. SCHNEIDER and A. S. WILLSKY
Stochastic Systems Group, Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, USA

W. C. KARL
Departments of Electrical and Computer Engineering and Biomedical Engineering
Boston University, 44 Cummington St., Boston, MA 02215, USA

Magnetic resonance imaging (MRI) has become a widely used research and clinical tool in the study of the human brain. The ability to robustly and accurately quantify repeatable morphological measures from such data is aided by the ability to accurately segment the MRI data set into homogeneous regions such as gray matter, white matter, and cerebro spinal fluid. The large amount of data associated with typical MRI scans makes completely manual segmentation prohibitive on a large scale. In this paper an efficient approach to the segmentation of such MR imagery is presented. The approach uses an estimation-theoretic interpretation of the segmentation problem to develop a computationally efficient, statistically-based recursive technique for its solution. Being statistically based, the method also provides associated measures of uncertainty of the resulting estimates, which are extremely important both for evaluation of the estimates as well as their combination with other sources of information.

Keywords: MR imaging, medical imaging, segmentation, estimation, Kalman filter, anisotropic diffusion, edge-preserving, smoothing variational methods, calculus of variations.

1. INTRODUCTION

Magnetic resonance imaging (MRI) has become a widely used research and clinical tool in the study of the human brain. The size and shape (i.e. the morphology) of the various structures of the brain which can be obtained from such imagery correlate with various developmental differences, disease states, and injuries. In Ref. 1 independent measures of brain volumes, shapes, and positions are readily correlated with behavioral as well as physiological measures. For instance, Ref. 2 suggests that Rett syndrome, a progressive disorder associated with the regression of psychomotor development and slow brain growth to age one, could be diagnosed by precise volumetric analysis of cortical and nuclear structures over time. Also, in Ref. 3, it was shown that the female brain is approximately 10% smaller than a male brain but the disproportionate variation in the volumes of male and female brain

*This work was supported by the National Institutes of Health under Grant NINDS 1 R01 NS34189, by the Air Force Office of Scientific Research under Grant F49620-96-1-0028, by ONR under Grant N00014-91-J-1004, by a National Science Foundation Graduate Research Fellowship, and a Whitaker Foundation Graduate Research Fellowship.
†Corresponding author. E-mail: johnk@bu.edu

International Journal of Pattern Recognition and Artificial Intelligence Vol. 11 No. 8 (1997) 1213–1231
© World Scientific Publishing Company

regions and substructures reported was not consistent with a simple allometric explanation for the sexual dimorphism. Shenton *et al.*[4] measured post mortem left temporal lobe volumes in schizophrenics. Certain anatomical differences from normals in these volumes correlated with schizophrenia. In Ref. 5, clinically relevant frontal brain lesion volumes were measured morphologically by MR imaging and validated by measuring CT blood flow in the same volume. These are only a few examples which strongly suggest the extreme importance of the ability to robustly, accurately quantify such repeatable morphological measurements.

Accurate and reproducible quantification depends on the accurate segmentation and classification of the MRI brain data set into constituent functional parts. In particular, this overall goal is aided by the segmentation of the raw MRI data into homogeneous regions such as gray matter, white matter, and cerebro spinal fluid (CSF). These regions and their boundaries have typically been delineated through a labor intensive procedure by specially trained technicians, at great expense in time, money, and effort and with the greater variability that comes with human intervention.[6] The incorporation of automation into the segmentation process has been shown to reduce variability across technicians and across morphological structure measurements.[7] Such results have prompted greater interest in increasing the automation of this component of the analysis process, which is the long term focus of this work.

One promising approach that has been taken to automating such segmentation problems in the computer vision literature is the use of a variational or cost-functional framework.[8–14] In these approaches to segmentation an energy functional captures the desired properties of the resulting segmentation, such as smoothness within homogeneous regions, preservation of boundaries between homogeneous regions, etc. The minimum energy the functional can attain given the observed image is chosen as the segmentation. The advantage of such methods is the ease with which various desirable effects can be directly incorporated into the energy functional and reflected in the resulting segmentation. Unfortunately, obtaining the minimum of these energy functions leads to large and computationally taxing optimization problems. In addition, while these deterministic methods can provide estimates of the underlying components and their boundaries, they do not provide information about the uncertainty in these estimates. Such measures of uncertainty are important both for direct evaluation of the estimates themselves, as well as providing the information necessary to fuse these estimates with other sources of knowledge, such as probabilistic atlases, images from other modalities, etc. Further, while the addition of many terms capturing various effects is conceptually simple, the specification of the corresponding weights or the importance given to these terms can be challenging.

Alternative statistical formulations of such segmentation problems based on Markov random field (MRF) prior models,[15–17] while also providing visually desirable solutions, lead to computationally taxing optimization problems. Indeed, the resulting optimization problems are very similar to the variationally derived ones. Furthermore, while these Bayesian approaches conceptually provide a frame-

work for the calculation of measures of uncertainty (i.e. estimation error variances), in practice, such measures are virtually never obtained due to the great computational cost involved. Thus current methods for the automation of MRI segmentation are computationally costly and, further, do not provide rational measures of uncertainty.

To overcome the above limitations of existing methods, in this paper we present a statistically-based recursive approach to the segmentation of MR imagery which is computationally efficient and provides measures of uncertainty of the resulting estimates. The approach is based on an interpretation of the segmentation problem as an equivalent recursive estimation problem. This enables the use of efficient, near-optimal recursive filtering methods for its solution. As a result, rather than solving the large overall segmentation problem all at once, the problem is decomposed into a sequence of smaller estimation problems, yielding computational efficiency in addition to uncertainty measures.

The paper is organized as follows. In Sec. 2 we present a summary of the variationally-based image segmentation on which we build. We then present an interpretation of the segmentation problem as an estimation problem in Sec. 3. In Sec. 4 we show how this problem may be solved recursively and thus efficiently. In Sec. 5 we show examples of brain segmentations using our approach. We then present our conclusions in Sec. 6.

2. A VARIATIONAL FORMULATION OF THE SEGMENTATION PROBLEM

2.1. Problem Statement

We base our approach to the segmentation of MR imagery on a variational formulation of the segmentation problem presented in Refs. 8, 9 and 12. This formulation simultaneously produces a segmented image estimate $\widehat{f}$ and edge map estimate $\widehat{s}$ as the minimizers of the following energy functional:

$$E(f,s) = \iint_{\Omega} \underbrace{\mu(g-f)^2}_{\text{Data Fidelity}} + \underbrace{\lambda(1-s)^2\,|\nabla f|^2}_{\text{Smoothness}} + \underbrace{\nu\left(\rho\,|\nabla s|^2 + \frac{1}{\rho}s^2\right)}_{\text{Edge Penalty}} dx\,dy \qquad (1)$$

where g denotes the observed image data, f denotes the piecewise smooth approximating field, s denotes the corresponding continuous edge strength image, Ω is the image domain and μ, λ, ν, and ρ are scalar weights specifying the relative importance of the terms. The edge image s can be viewed as an indication of the edge or boundary strength at each location in the observed image, and ranges in value between 0 and 1. There are three types of terms in Eq. (1) — a data fidelity term, a smoothness term, and edge penalty terms. These terms ensure that the pair $\widehat{f}$, $\widehat{s}$ that minimizes (1) has the properties that $\widehat{f}$ is close to the data g, that $\widehat{f}$ is nearly flat except where s is close to 1 (indicating the presence of an edge), that the edge field $\widehat{s}$ itself is smooth, and that the number of edges (i.e. the size of $\widehat{s}$) is not too great, preventing over-segmentation. Note that as $\rho \to 0$ it is shown in Ref. 12 that the solution of (1) approaches the binary edge formulation of Ref. 10.

In practice, (1) is solved by alternatively fixing s and minimizing with respect to f and then fixing f and minimizing with respect to s, as shown in Fig. 1. This corresponds to alternatively estimating the values of homogeneous image regions assuming that the location of the boundaries between them are known and then estimating the location of the boundaries assuming the values within the smooth, homogeneous regions are known. In particular, with the edge field s considered fixed, the following energy functional is minimized for the segmented image values f:

$$E_s(f) = \iint_{\Omega} \mu(g-f)^2 + \lambda(1-s)^2 |\nabla f|^2 \, d\Omega . \tag{2}$$

Note, in this step a penalty, $|\nabla f|^2$, enforces smoothness on the resulting estimated field f *except* where the edge field s is close to 1.

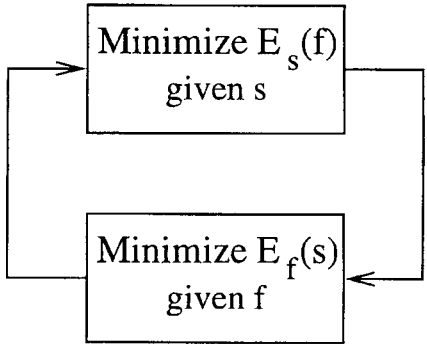


Fig. 1. Segmentation algorithm.

For the other step, the image f is considered fixed and the minimum of the functional with respect to s is sought. This step can be shown to be equivalent to minimizing the following energy functional for s [18]:

$$E_f(s) = \iint_{\Omega} \left(\lambda |\nabla f|^2 + \frac{\nu}{\rho}\right) \left[\frac{\lambda |\nabla f|^2}{\lambda |\nabla f|^2 + \frac{\nu}{\rho}} - s\right]^2 + \nu\rho |\nabla s|^2 \, d\Omega . \tag{3}$$

The second term is a penalty that will impose smoothness on the resulting edge field everywhere, while the first term contains an equivalent derived "observation" of the edge field based on a normalized gradient of the image f. In particular, note that this observation, $\lambda |\nabla f|^2 / (\lambda |\nabla f|^2 + \frac{\nu}{\rho})$, has values in the range [0, 1] and will be close to 1 where the gradient of f is large.

2.2. Discretization

In practice, we work on a discrete grid of sample points and approximate the gradient operator with a finite difference scheme. In particular, we assume that both

f and s are sampled on the same $n \times n$ row-column grid containing a total of n^2 points. It will prove convenient to define the n-vectors $\mathbf{f}(j)$ and $\mathbf{s}(j)$, obtained by stacking the sampled image values $(f)_{ij}$ and sampled edge field values $(s)_{ij}$ in column j, respectively, as illustrated for the f field in Fig. 2. We also define $\mathbf{f}$ and $\mathbf{s}$ as the overall n^2-vector of these sampled points, obtained by stacking the $\mathbf{f}(j)$ and $\mathbf{s}(j)$, respectively. Similarly, we define $\mathbf{g}(j)$ to be the n^2-vector obtained by stacking the sampled observation values $(g)_{ij}$ in column j and define $\mathbf{h}(j)$ to be the n^2-vector obtained by stacking the sampled values of the s-field "observation" $\lambda|\nabla f|^2/\left(\lambda|\nabla f|^2 + \nu/\rho\right)$ in column j. In particular, the latter samples are given by $\mathbf{h}(j)_i = ([(f)_{ij} - (f)_{i,j-1}]^2 + [(f)_{ij} - (f)_{i-1,j}]^2)/([(f)_{ij} - (f)_{i,j-1}]^2 + [(f)_{ij} - (f)_{i-1,j}]^2 + \frac{\nu}{\lambda\rho})$ and serve as a normalized gradient observation, lying in the range $[0, 1]$.

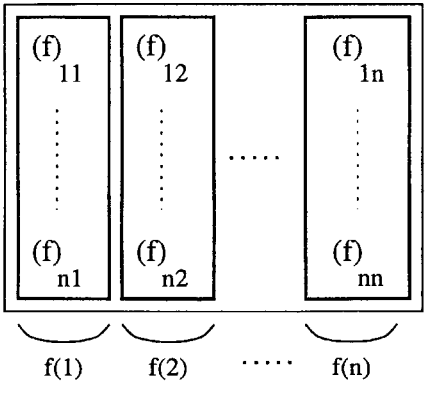


Fig. 2. Sampled values.

With this notation, a discrete version of the field energy term $E_{\mathbf{s}}(\mathbf{f})$ given in (2) may be obtained as[a]:

$$E_{\mathbf{s}}(\mathbf{f}) = \sum_{j=1}^{n} \|\mathbf{g}(j) - \mathbf{f}(j)\|^2_{\mu I} + \|D_r \mathbf{f}(j)\|^2_{V(j)} + \|\mathbf{f}(j) - \mathbf{f}(j-1)\|^2_{V(j)} \tag{4}$$

where $\|\mathbf{x}\|^2_M = \mathbf{x}^T M \mathbf{x}$ and D_r is a first-order difference operator along the rows of each column given by:

$$D_r = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \tag{5}$$

The weighting matrices $V(j)$ are diagonal and capture the spatially varying, edge-dependent weights on column j of the gradient term in (2). In particular, $V(j)_{ii} = \lambda[1 - (s)_{ij}]^2$.

[a]For clarity of presentation we ignore the edge effects in the discrete formulations of (4) and (5). Such edge effects only increase the notational complexity of the development.

Similar to (4), a discrete version of the edge energy term $E_{\mathbf{f}}(\mathbf{s})$ given in (3) may be obtained as:

$$E_{\mathbf{f}}(\mathbf{s}) = \sum_{j=1}^{n} \|\mathbf{h}(j) - \mathbf{s}(j)\|^2_{W(j)} + \|D_r\mathbf{s}(j)\|^2_{\nu\rho I} + \|\mathbf{s}(j) - \mathbf{s}(j-1)\|^2_{\nu\rho I} \qquad (6)$$

where D_r is the row-derivative operator defined in (5) and the weighting matrices $W(j)$ are diagonal and capture the spatially varying weights on column j of the data term in (3). In particular, $W(j)_{ii} = \lambda[(f)_{ij} - (f)_{i,j-1}]^2 + \lambda[(f)_{ij} - (f)_{i-1,j}]^2 + \frac{\nu}{\rho}$.

The solution to the overall segmentation problem is then obtained by alternately minimizing each of the two quadratic energies $E_{\mathbf{s}}(\mathbf{f})$ in (4) and $E_{\mathbf{f}}(\mathbf{s})$ in (6), as shown in Fig. 1. Each of these minimizations is a large, computationally intensive, optimization problem, which must be performed repeatedly to obtain a segmentation. Direct solution of the system of equations defining the minimum of (4) or (6) at any stage of the algorithm would involve $O(n^6)$ calculations and, since n is typically in the range of 10^2–10^3, would be prohibitively expensive. In practice, therefore, the solution is approximated to an arbitrary degree of accuracy via an iterative method, such as conjugate gradient[19] or multigrid.[20] Even with such methods, solution of the large set of equations is computationally challenging. Further, only the *estimate* of the image field $\widehat{\mathbf{f}}$ or the edge process $\widehat{\mathbf{s}}$ is provided, with no measure of reliability.

To avoid these difficulties we use an interpretation of the problems (4) and (6) in an estimation theoretic context.[18] Such an interpretation of this segmentation problem allows a variety of statistically-motivated approaches to be taken. For example, in Ref. 18 a novel multiresolution estimation approach is used. In this work, however, we interpret the underlying spatial estimation problem as an equivalent *dynamic* estimation problem. Based on this interpretation we are able to formulate an efficient, statistically-based recursive solution. This statistically-based approach provides not only estimates of the fields themselves, but also corresponding measures of uncertainty. For simplicity we focus on the problem of processing two-dimensional MRI slices, i.e. images. The development to follow, however, carries over naturally to the case of three (or even higher) dimensional fields (i.e. volumes), where the computational gains should be even greater. Such results are the focus of present research.

3. AN ESTIMATION-THEORETIC INTERPRETATION

We now provide an estimation-theoretic interpretation of the quadratic minimization problems (4) and (6) as equivalent recursive estimation problems. To this end, note that (4) and (6) have the following common form:

$$E(\mathbf{x}) = \sum_{i=1}^{n} \|\mathbf{d}(j) - \mathbf{x}(j)\|^2_{R_1^{-1}(j)} + \|D_r\mathbf{x}(j)\|^2_{R_2^{-1}(j)} + \|\mathbf{x}(j) - \mathbf{x}(j-1)\|^2_{Q^{-1}(j)} \qquad (7)$$

where the weighting matrices $R_1(j)$, $R_2(j)$, and $Q(j)$ are diagonal, and the vector $\mathbf{x}(j)$ contains the elements from column j of the overall unknown desired image vector $\mathbf{x} = [\mathbf{x}^T(1)|\cdots|\mathbf{x}^T(n)]^T$.

Now, it is straightforward to show[21–23] that the value of $\hat{\mathbf{x}}$ or, as an equivalent, $\hat{\mathbf{x}}(j)$ for each j, that minimizes the quadratic energy (7) is the same as the maximum-likelihood estimate of $\mathbf{x}(k)$ for each k based on the following set of observations and constraints:

$$\mathbf{x}(j) = \mathbf{x}(j-1) + \mathbf{q}(j-1), \qquad \mathbf{q}(j) \sim \mathcal{N}(0, Q(j)) \tag{8}$$

$$\mathbf{y}(j) = C(j)\,\mathbf{x}(j) + \mathbf{r}(j), \qquad \mathbf{r}(j) \sim \mathcal{N}(0, R(j)) \tag{9}$$

for $j = 1, \ldots, n$, where $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, P)$ denotes a Gaussian random vector with mean $\mathbf{m}$ and covariance P, the initial condition is given by $\mathbf{x}(0) = \mathbf{0}$ and $Q(0) = \lim_{\alpha \to \infty} \alpha I$, and the variables $\mathbf{y}(j)$, $C(j)$ and $R(j)$ are defined as:

$$\mathbf{y}(j) = \left[\frac{\mathbf{d}(j)}{\mathbf{0}}\right], \quad C(j) = \left[\frac{I}{D_r}\right], \quad R(j) = \left[\begin{array}{c|c} R_1(j) & \mathbf{0} \\ \hline \mathbf{0} & R_2(j) \end{array}\right]. \tag{10}$$

Given the form of (8) and (9) we can see that another equivalent interpretation of these equations specifies a dynamic estimation problem for the $\mathbf{x}(j)$. In particular, the components $\hat{\mathbf{x}}(j)$ of the minimizer $\hat{\mathbf{x}}$ of (7) are *precisely the same* as the series of smoothed estimates of $\mathbf{x}(j)$ based on the dynamic equation (8) and the observation (9) for $1 \leq j \leq n$. The advantage of the formulation (8) and (9) is that we may use efficient, recursive Kalman filtering based techniques for its solution. In effect, by recursively processing $\mathbf{x}(j)$ we are recursively estimating the columns of the segmented image. In effect, we trade solving a single large problem (estimating the entire image) for solving a series of smaller problems (estimating the values in a column). In addition, these estimation-theoretic methods provide not only estimates of the field, but also corresponding measures of uncertainty. Finally, the statistical smoothing view of (8) and (9) suggests a statistical interpretation of the parameters of the original formulation, which can provide a rational guide to their selection in practice.

4. A RECURSIVE FILTER-BASED SOLUTION

In this section we show how to efficiently solve (8) and (9). First note that what is needed is the optimal estimate $\hat{\mathbf{x}}(j)$ for each j based on *all* the data $\mathbf{y}(k)$, $k = 1, \ldots, n$. This is known as the optimal *smoothed* estimate.[24] There are a number of efficient ways to generate this estimate, as described in Refs. 25–27. In this work we use the Mayne-Fraser two-filter form of the optimal smoother.[25,26] In this approach, the smoothed estimates and associated error variances are obtained by combining the outputs of two independent Kalman filters based on (8) and (9) — one starting at column $j = 1$ and running forward (i.e. left to right) and the other starting at column $j = n$ and running backward (i.e. right to left), as depicted in Fig. 3. Since each filter is independent, they may be run in parallel for greater throughput. The separate forward and backward estimates are then combined by weighting them by the inverse of their respective error variances, as detailed in Refs. 25 and 26. The key, then, is to find an efficient method for performing the independent filtering steps, which we focus on next. In particular, we focus on

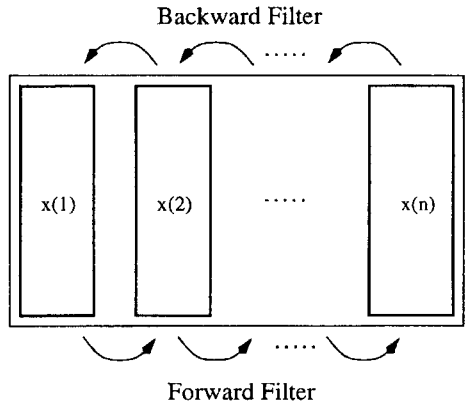


Fig. 3. Recursive segmentation via recursive smoothing.

the forward filtering step. The backward step is equivalent and the weighted averaging operation straightforward.

4.1. An Information Form Filter

To perform the filtering based on (8) and (9) we will use an efficient version of the *information form* of the Kalman filter developed in Ref. 22. In particular, an optimal information-form filtering algorithm for the system (8)—(10) is provided by the following recursive algorithm[22]:

Information Filter Algorithm:

Prediction:

$$\overline{L}(j) = Q^{-1}(j) - Q^{-1}(j)\left(Q^{-1}(j) + \widetilde{L}(j-1)\right)^{-1} Q^{-1}(j), \tag{11}$$

$$\overline{\mathbf{x}}(j) = \widetilde{\mathbf{x}}(j-1), \tag{12}$$

$$\overline{\mathbf{z}}(j) = \overline{L}(j)\overline{\mathbf{x}}(j) \tag{13}$$

Update:

$$\widetilde{L}(j) = \overline{L}(j) + R_1^{-1}(j) + D_r^T R_2^{-1}(j) D_r, \tag{14}$$

$$\widetilde{\mathbf{z}}(j) = \overline{\mathbf{z}}(j) + R_1^{-1}\mathbf{d}(j) \tag{15}$$

$$\widetilde{L}(j)\widetilde{\mathbf{x}}(j) = \widetilde{\mathbf{z}}(j) \tag{16}$$

where D_r is the derivative operator defined in (5) and note that $R_1(j)$, $R_2(j)$, and $Q(j)$ are diagonal matrices. The matrices $L(j)$ appearing in the above algorithm are the *information matrices* corresponding to the estimate, and are equal to the inverse of the estimation error covariance for the filtered estimate of column j. The filtered solution to (8)–(10) provided by the use of (11)–(16) is exact and, when optimally combined with the corresponding backward filtered estimates provides an estimate

which is the same as the minimizer of (7). However, the above algorithm and interpretation break the problem into a series of smaller recursive steps and also provide measures of uncertainty for the estimates through the associated estimation error variances, which are the diagonal elements of $L^{-1}(j)$.

For the sake of computational efficiency we go even further and approximate this exact solution. In particular, as done in Ref. 22, we use a two term series expansion of the matrix inverse in parentheses and replace the exact expression (11) for the predicted information matrix with:

$$\overline{L}(j) \approx Q^{-1}(j) - Q^{-1}(j)\left(\Lambda^{-1}(j) - \Lambda^{-1}(j)\Omega(j)\Lambda^{-1}(j)\right)Q^{-1}(j) \tag{17}$$

where $\Lambda(j)$ is the diagonal part of $Q^{-1}(j) + \widetilde{L}(j-1)$ and $\Omega(j)$ is the remaining, off-diagonal part. This simple approximation to (11) yields results close to the exact solution of the variational formulation in most cases. If greater fidelity to the original variational formulation is required additional terms may certainly be added to the above approximate expansion, at the cost of more computation, larger models, and greater storage requirements. The fidelity of such approximations will depend on quantities such as the specific choice of parameters for the original variational problem along with the structure of the image. Such relationships form a topic of the current investigation.

Finally, note that calculation of the updated estimate in (16) requires the solution of a system of equations. While this system of equations is much smaller than that appearing in the original problem, it can still be a computationally intensive task. Fortunately, the structure of this segmentation problem coupled with the approximation (17) combine to ensure that $\widetilde{L}(j)$ will remain sparse and banded (in particular, tridiagonal). In this case, the inversion operation implied in (16) can itself be efficiently performed via iterative methods, such as multigrid[20] or preconditioned conjugate gradient.[19] In our present implementation we use the conjugate gradient method preconditioned with the diagonal of $\widetilde{L}(j)$.

4.2. The Smoothed Estimate

Once the separate forward and backward filtered estimates are found using (8) and (9) together with the efficient variant of (11)–(16), the individual estimates are then combined to obtain the resulting smoothed estimate $\widehat{\mathbf{x}}(j)$ together with its uncertainty. Let $\widehat{\mathbf{x}}_f(j)$ and $\widehat{\mathbf{x}}_b(j)$ denote the output of the forward and backward filtering operations, respectively, and let $L_f(j)$ and $L_b(j)$ denote the information matrices corresponding to these filtered estimates. Then the overall smoothed estimate at column j, $\widehat{\mathbf{x}}(j)$, is obtained as the solution of:

$$\left[L_f(j) + L_b(j)\right]\widehat{\mathbf{x}}(j) = L_f(j)\widehat{\mathbf{x}}_f(j) + L_b(j)\widehat{\mathbf{x}}_b(j) \tag{18}$$

which we again solve using the diagonally preconditioned conjugate gradient technique. The corresponding error variances are given by the diagonal elements of $\left(L_f(j) + L_b(j)\right)^{-1}$, which we approximate by the reciprocal of the corresponding

diagonal elements of $L_f(j) + L_b(j)$. These error variances give us a measure of the uncertainty of the corresponding estimates.

5. EXAMPLES

In this section, we present numerical examples of segmented MR imagery using the techniques we have described. We use the global segmentation algorithm shown in Fig. 1 with the energy terms $E_{\mathbf{s}}(\mathbf{f})$ and $E_{\mathbf{f}}(\mathbf{s})$ described in (4) and (6), respectively. We initialize the $\mathbf{s}$ field estimate to zeros and terminate the iteration when the percent change in $E(f, s)$ falls below a threshold. For the experiments shown here we used a 2% change as the indication of convergence. This structure is common to all the examples. What will differ is how the subproblems of minimization of $E_{\mathbf{s}}(\mathbf{f})$ or $E_{\mathbf{f}}(\mathbf{s})$ are performed. In particular, we compare straightforward solution of (4) and (6) to the proposed recursive estimation-theoretic based technique. In both cases note that the formulation of (1) assumes that the value of the edge field s lies in the range $[0, 1]$ while our solutions to the subproblem (6) contain no such constraint. To ensure the global iterations are well defined we simply truncate the edge function to the range $[0, 1-\epsilon]$ for some small ϵ at each iteration if necessary. In practice, the solution to the unconstrained problem almost always lie in the required range anyway.

5.1. Comparison of Estimates Obtained with Direct and Recursive Approaches

Here we compare images segmented using the algorithm of Fig. 1 and exact solution of the normal equations to those obtained using the proposed efficient recursive technique. For these experiments we used parameter values of $\mu = 1$, $\lambda = 3$, $\nu = 1200$, and $\rho = \frac{1}{2}$. In Fig. 4 we show the original proton density weighted MR image. In Fig. 5, the two segmentations are shown. On the left is the solution

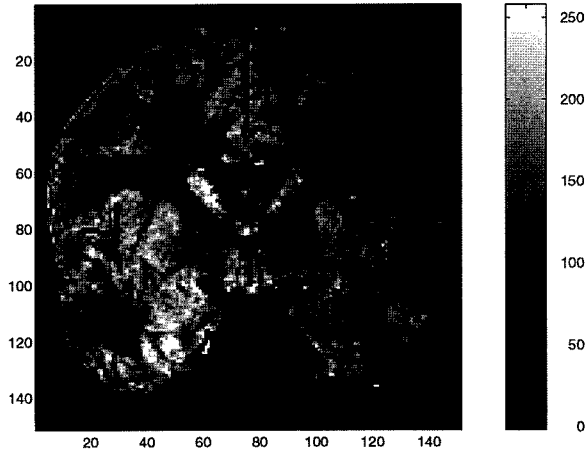
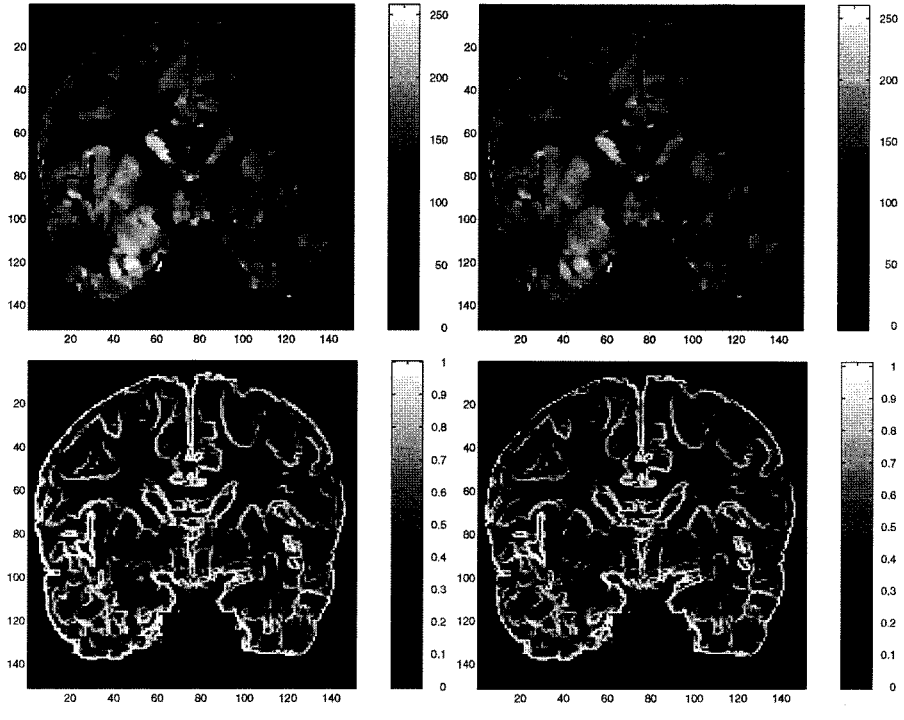


Fig. 4. Original PD weighted image.

obtained through direct minimization of the energy functionals (4) and (6) via exact solution of the corresponding normal equations or, the equivalent, Euler equations. On the right is the corresponding solution for both **f** and **s** obtained by using the near optimal recursive approach. The images show that the recursive technique produces results that are almost identical to the direct solutions. To emphasize this point, in Fig. 6 we show a histogram of the differences between the field estimates obtained by the two techniques. These histograms show the relative frequency (i.e. the histogram area equals one) of the corresponding errors as a percent of the corresponding full scale value. As can be seen all the values are clustered around zero. Both segmentations appear to do well in capturing the main boundaries between gray and white matter despite the presence of a significant gradient in absolute intensity across the image. In Fig. 8 we show the result of thinning the thresholded edge field obtained in Fig. 5.



(a) Direct solution (b) Recursive solution.

Fig. 5. Comparison of answer obtained by direct solution of the normal equations and our recursive estimation-based approach. The top figures are the **f** field estimates and the bottom figures are the corresponding **s** field estimates.

5.2. Uncertainty Estimates

While the segmentations obtained by direct energy minimization and our recursive approach are quite similar, since our Kalman-filter-based technique is derived

from an estimation-theoretic interpretation of the segmentation problem, we also generate associated error variance information, and thus measures of segmentation uncertainty. Such measures are important not only for the direct evaluation of the estimates themselves, but also for the fusion of such estimates with other sources of information, which we demonstrate later. While many approaches have been proposed for the segmentation of such MR imagery, few have addressed this issue of error statistics. In Fig. 9 we show the error standard deviation fields corresponding to the estimates in Fig. 5. On the left is the standard deviation field corresponding to the smoothed field while on the right is the standard deviation field corresponding to the edge field. As can be seen, the error standard deviations for the field itself increase near edges while those of the edge field decrease near edges, as we would expect. In areas where the edges are weak, the corresponding error field indicates low reliability. These error measures are obtained as a by product of the processing in the recursive technique, and are thus obtained for free.

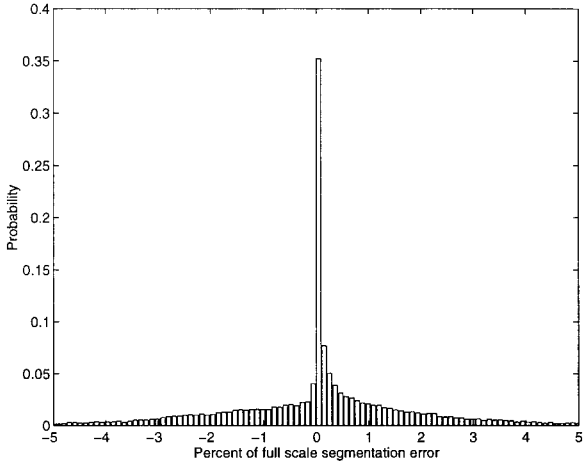


Fig. 6. Histogram of difference between field estimates based on direct solution of the normal equations and our recursive estimation-based approach.

5.3. Computational Cost

We compare the computational cost of solving the segmentation problem represented in Fig. 1 both through direct minimization of the energies and by our recursive method. The direct method finds the minimum of the energies (4) and (6) at each iteration by explicitly solving the resulting system of the so-called normal or Euler equations. In practice, such large systems of equations are usually solved via iterative techniques, such as preconditioned conjugate gradient. Thus, for our comparison here we find the minimum of the energies (4) and (6) at each stage using the conjugate gradient technique preconditioned with the diagonal of the corresponding normal equations.[19] Our recursive approach has already been described.

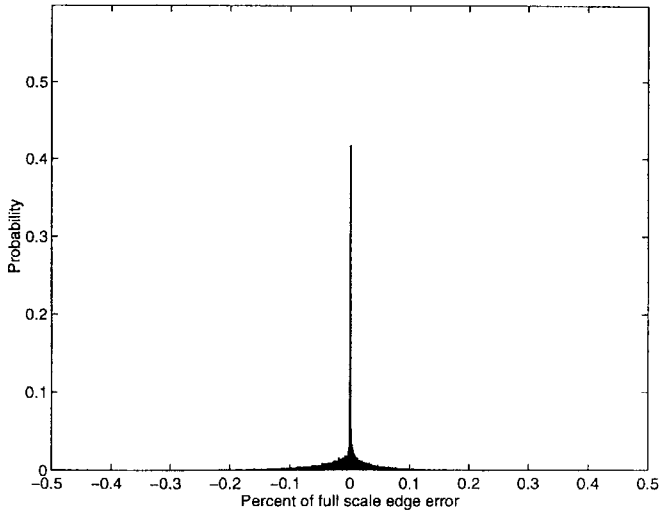


Fig. 7. Histogram of difference between edge estimates based on direct solution of the normal equations and our recursive estimation-based approach.

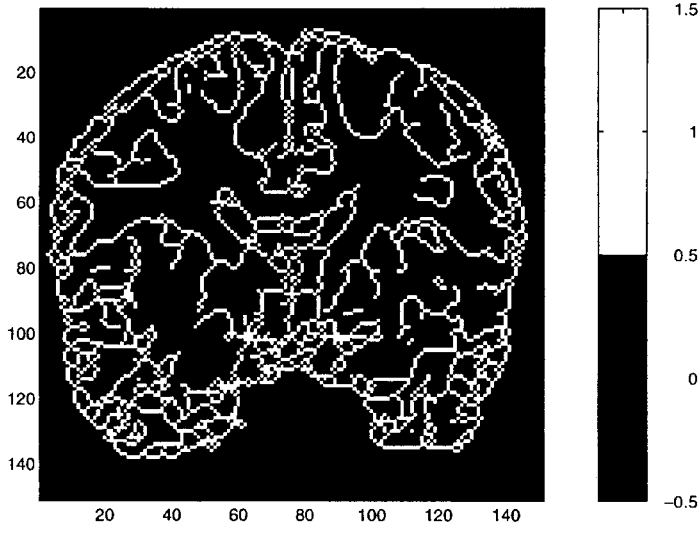
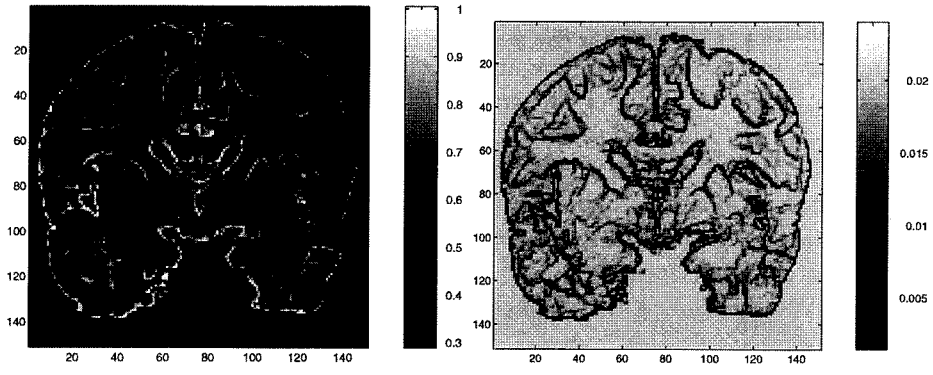


Fig. 8. Thinned edge field.

First we consider only the field estimates. In this case, both techniques took 4 global iterations to converge. Our recursive approach required approximately 20% fewer computations per iteration compared to the direct minimization approach. These solutions were obtained in a few minutes on a SPARC 20 using unoptimized MATLAB code.



(a) Field estimate error standard deviation. (b) Edge estimate error standard deviation.

Fig. 9. Field and edge estimate error standard deviations.

Next we compare the cost of calculating *both* the estimates of $\mathbf{f}$ and $\mathbf{s}$ together with the corresponding error measures (i.e. the error variances). For the direct approach, this calculation essentially corresponds to finding the diagonal elements of the *inverse* of the large matrix arising in the normal equations of the optimization problem which requires inverting the matrix and is extremely costly. In contrast, for our recursive approach these error measures are obtained as a by product of the processing and are thus effectively free. In particular, in computing both these quantities the direct approach required a factor of approximately 10^4 more computations than our recursive method.

5.4. Fusion of Multichannel Data

Here we demonstrate the use of the error statistics provided by our technique for the fusion of edge information from multichannel data, in particular from registered T_1, T_2 and PD imagery. Our goal here is to demonstrate the potential usefulness of such uncertainty information through a prototype fusion problem.

There are at least two ways our approach can accomplish this fusion of edge information. The first method is based on the direct use of the three channels in a single overall variational formulation, as proposed in Ref. 8. In this case all three channels are simultaneously used to directly obtain a *single* edge field. This single edge map is used as the basis for the generation of a piecewise smooth estimate of the intensity of each separate channel. Our approach to this method also provides the associated uncertainty measure for the aggregate edge map and individual channel estimates, and thus provides one rational means to obtain such overall measures for multichannel data. In particular, note that the measures obtained for each channel will reflect the effects of the information in all the channels. Since this approach is a direct extension of that in Ref. 8 we will not pursue it here.

An alternative approach to the fusion of the edge or segmentation information from such multichannel data is to perform a segmentation on each channel sepa-

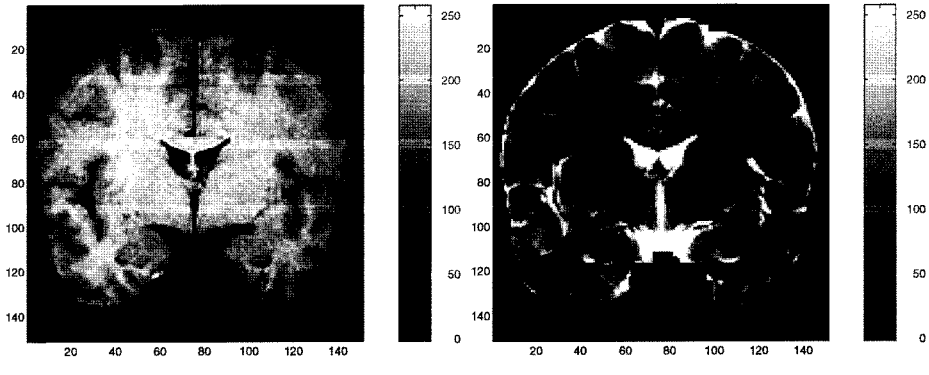
rately and then use the resulting uncertainty measures obtained for each edge image to subsequently fuse the individual edge maps. In effect, this approach treats the separate edge field estimates from each channel as independent observations of the edge field with uncertainty provided by the associated error variance measure. Such an approach serves as a paradigm of a variety of similar fusion problems, in which it is desired to combine information obtained from a variety of sources. In particular, suppose at a given image pixel location (i, j), each channel provides a corresponding edge field estimate at that pixel, denoted by s_{T_1}, s_{T_2}, s_{PD}, respectively. In addition, suppose that $\sigma_{T_1}^2$, $\sigma_{T_2}^2$, σ_{PD}^2 are the corresponding error variances associated with that pixel value as, for example, provided by our technique. If these edge estimates are uncorrelated from pixel to pixel and from channel to channel, then the optimal linear estimate of the overall edge field value at that pixel is provided by the quantity:

$$s = \frac{\sigma_{T_1}^{-2} s_{T_1} + \sigma_{T_2}^{-2} s_{T_2} + \sigma_{PD}^{-2} s_{PD}}{\sigma_{T_1}^{-2} + \sigma_{T_2}^{-2} + \sigma_{PD}^{-2}} \tag{19}$$

and the associated overall error variance σ^2 for the combined edge field estimate at that pixel is given by:

$$\sigma^2 = \frac{1}{\sigma_{T_1}^{-2} + \sigma_{T_2}^{-2} + \sigma_{PD}^{-2}} \cdot \tag{20}$$

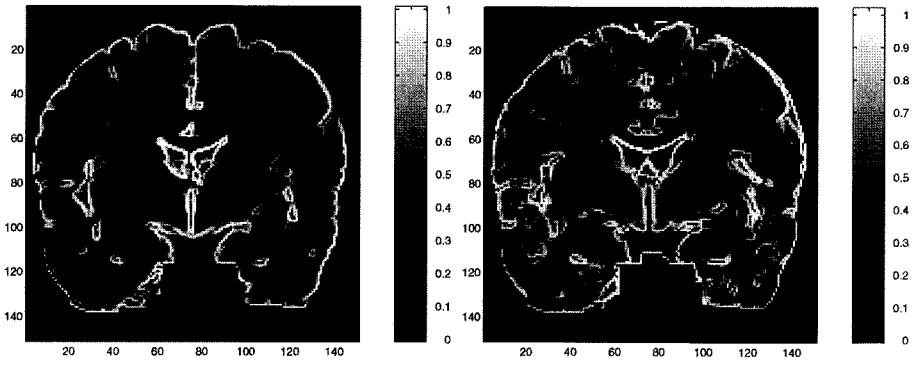
As stated above, this assumes that the edge estimates obtained for each channel correspond to uncorrelated random variables. While this will not be the case, in general, the formulas (19) and (20) still provide a rational way to combine such data. Note that if other sources of edge information are available with associated uncertainty measures, they may also be included in this framework.



(a) T1 weighted image (b) T2 weighted image

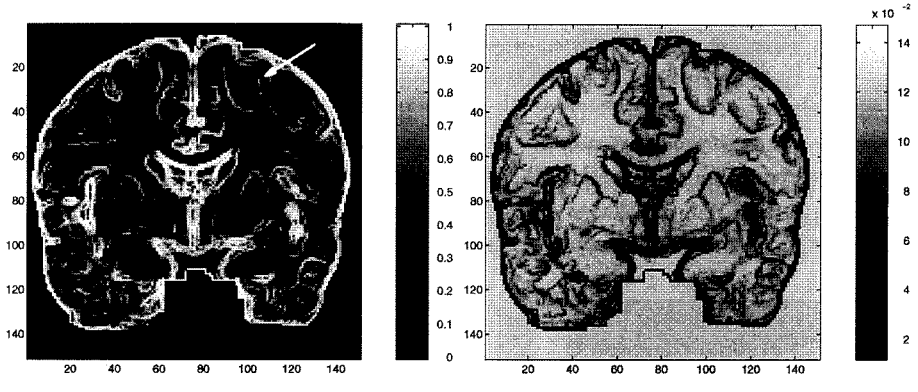
Fig. 10. T1 and T2 weighted images.

Figure 10 shows T1 and T2 weighted MRI images that correspond to the PD weighted image in Fig. 4. We generated separate edge estimates for each of these images using our technique. For the T1 image we used the parameters $\mu = 1$, $\lambda = 3$, $\rho = \frac{1}{2}$ and $\nu = 1000$, while for the T2 image we used the parameters $\mu = 1$, $\lambda = 3$,



(a) T1 weighted edge field (b) T2 weighted edge field

Fig. 11. T1 and T2 weighted edge estimates.



(a) Fused edge image (b) Corresponding fused uncertainty image

Fig. 12. Fusion of edge estimates and standard deviation of estimation error.

$\rho = \frac{1}{2}$ and $\nu = 1000$. In Fig. 11 the resulting estimated edge fields for the T1 and T2 images are shown. The edge field for the PD weighted image is shown in Fig. 5(b). We then combined these individual edge estimates and corresponding uncertainty measures using the formulae in (19) and (20). The resulting fused edge image is shown in Fig. 12(b). Note that it contains edge information that does not exist in the edge image of any single channel. For instance, in the fused edge estimate, the boundary indicated by the arrow has been reinforced, indicating the incorporation of collateral edge information, since the same boundary in the PD edge estimate is incomplete. Note also how the estimation error, shown in Fig. 12(b) decreases as more information is added to the fused estimate of the edge field and more confidence in its value is developed.

6. CONCLUSIONS

In this paper we have presented a recursive approach to the segmentation of MR imagery. This approach was based on an estimation-theoretic interpretation of the segmentation problem and while efficient, provides solutions indistinguishable from the direct solution of an equivalent variational problem. In addition to its efficiency, being statistically based, the method also provides associated measures of uncertainty of the resulting estimates. Such measures are critical not only for the evaluation of the segmentation but also for the subsequent stages of the processing. An illustration of the use of such information was demonstrated through the fusion of edge information obtained from registered multichannel (PD, T1, and T2) imagery. This example demonstrated how the collateral information present in such data can be rationally combined when such measures of uncertainty are available. We are in the process of extending this approach to the 3-D case, which while providing greater computational challenges, should also provide greater gains. A systematic study of the robustness of this approach in a clinical setting together with the usefulness and interpretation of the associated uncertainty measures that are generated is also currently being undertaken.

REFERENCES

1. V. S. Caviness, P. Filipek and D. Kennedy, "Magnetic resonance technology in human brain science: A blueprint for a program based upon morphometry", *Brain Development* **11** (1989) 1–13.
2. V. S. Caviness, P. Filipek and D. Kennedy, "Quantitative magnetic resonance imaging and studies of degenerative diseases of the developing human brain", *Brain Development* **14**, Suppl:S80–85 (1992).
3. P. Filipek, C. Richelme, D. Kennedy, J. Rademacher, D. Pitcher, S. Zidel and V. S. Caviness, "Morphometric analysis of the brain in developmental language disorders and autism", *Ann. Neurol.* **32** (1992) 475.
4. M. Shenton, R. Kikinis, F. Jolesz, S. Pollak, M. LeMay, C. Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman, *et al.*, "Abnormalities of the left temporal lobe and thought disorder in schizophrenia", *New England J. Med.* **327**, 9 (1992) 604–612.
5. Maeder *et al.*, "Volumes of chromis traumatic frontal brain lesions measured by MR imaging and cbf tomography", *Acta Radiologica* **32** (1991).
6. D. Kennedy, P. Filipek and V. Caviness, "Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging", *IEEE Trans. Med. Imag.* **8**, 1 (1989) 1–7.
7. R. Kikinis, M. Shenton, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. Guttmann, R. McCarley, W. Lorensen, H. Cline *et al.*, "Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging", *J. Magn. Reson. Imag.* **2**, 6 (1992) 619–629.
8. H. H. Pien and J. M. Gauch, "Variational segmentation of multi-channel MRI images", in *IEEE Int. Conf. Imag. Process.*, Austin, Texas, November 1994.
9. J. Shah, H. Pien and J. Gauch, "Recovery of surfaces with discontinuities by fusing shading and range data within a variational framework", *IEEE Trans. Imag. Process.* **5**, 8 (1996).
10. D. Mumford and J. Shah, "Boundary detection by minimizing functionals, I", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1985, pp. 22–26.

11. D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems", Technical Report CICS-P-88, Center for Intelligent Control Systems, MIT, 1988.
12. J. Shah, "Segmentation by nonlinear diffusion, II", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1992, pp. 644–647.
13. A. Blake and A. Zisserman, *Visual Reconstructions*, MIT Press, Cambridge, MA, 1987.
14. P. Perona and J. Malik, "Scale-space edge detection using anisotropic diffusion", *IEEE Trans. Patt. Anal. Mach. Intell.* **12**, 7 (1990) 629–639.
15. H. Derin, H. Elliot, R. Cristi and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields", *IEEE Trans. Patt. Anal. Mach. Intell.* **6** (1984) 707–720.
16. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. Patt. Anal. Mach. Intell.* **6** (1984) 721–741.
17. R. Azencott and B. Chalmond, "Nuclear magnetic resonance imagery restoration by Markov modelization and annealing", Onzi'eme Colloque sur le traitement du signal et des images, Nice du 1er, vol. 2, pp. 555–556.
18. M. Schneider, P. Fieguth, W. C. Karl and A. S. Willsky, "Multiscale statistical methods for the segmentation of signals and images", *IEEE Trans. Imag. Process.* (1996), submitted.
19. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Press, 1989.
20. W. Briggs, *A Multigrid Tutorial*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1987.
21. F. L. Lewis, *Optimal Estimation*, John Wiley & Sons, New York, 1986.
22. T. M. Chin, W. C. Karl and A. S. Willsky, "Sequential filtering for multi-frame visual reconstruction", SI on Multidimensional Signal Processing, *Signal Process.* **28**, 3 (1992) 311–333.
23. R. Szeliski, *Bayesian Modeling of Uncertainty in Low-level Vision*, Kluwer Academic Publishers, Norwell, Massachusetts, 1989.
24. H. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley and Sons, New York, NY, 1968.
25. A. Gelb, *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.
26. D. Fraser and J. Potter, "The optimum linear smoother as a combination of two optimum linear filters", *IEEE Trans. Autom. Control* **7**, 8 (1969) 387–390.
27. H. E. Rauch, F. Tung and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems", *AIAA J.* **3**, 8 (1965) 1445–1450.

John P. Kaufhold received the B.S. degree Summa Cum Laude in biomedical engineering in 1995 from Boston University, Boston Massachusetts. In 1995, he was awarded a Whitaker Foundation Predoctoral Fellowship, which supports him in the Ph.D. program in Boston University's Biomedical Engineering Department.

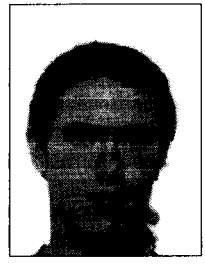
His interests are in medical image processing, especially stochastic model-based estimation, segmentation and semi-automated medical image interpretation. He is especially interested in applications to magnetic resonance imaging and transmission light microscopy. He is a member of the Biomedical Engineering Society, the Engineering Honor Society Tau Beta Pi, and the Biomedical Engineering Honor Society Alpha Eta Mu Beta.

Alan S. Willsky received both the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology in 1969 and 1973 respectively. He joined the M.I.T. faculty in 1973 and his present position is Professor of electrical engineering.

From 1974 to 1981, Dr. Willsky served as an Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. He is also a founder and member of the board of directors of Alphatech, Inc. In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. In 1988 he was made a Distinguished Member of the IEEE Control Systems Society.

Dr. Willsky is the author of the research monograph Digital Signal Processing and Control and Estimation Theory and is co-author of the undergraduate text Signals and Systems. He was awarded the 1979 Alfred Nobel Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE.

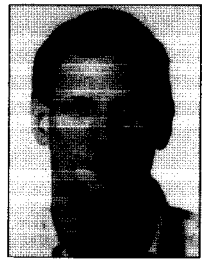
His research interests are in the development and application of advanced methods of estimation and statistical signal and image processing.



Michael K. Schneider received the BSE and MS degrees in electrical engineering from Princeton University in 1994 and the Massachusetts Institute of Technology in 1996, respectively.

He is currently pursuing a PhD in electrical engineering at the Massachusetts Institute of Technology.

His research interests include statistical signal and image processing and numerical linear algebra.



William C. Karl received the Ph.D. degree in electrical engineering and computer science in 1991 from the Massachusetts Institute of Technology, Cambridge, where he also received the S.M., E.E., and S.B. degrees. He held the position of Staff Research Scientist with the Brown-Harvard-M.I.T. Center for Intelligent Control Systems and the M.I.T. Laboratory for Information and Decision Systems from 1992 to 1994. He joined the faculty of Boston University in 1995, where he is currently Assistant Professor of electrical and computer engineering. Since January 1996 he has also held a joint appointment in the department of Biomedical Engineering.

Prof. Karl's research interests are in the areas of multidimensional and multiscale signal and image processing and estimation, geometric estimation and medical signal and image processing. He is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.