Chapter 12

Partial-Volume Bayesian Classification using Voxel Histograms

[1]

David H. Laidlaw
Department of Computer Science
Brown University
Providence, RI 02912
dhl@cs.brown.edu

Kurt W. Fleischer
Pixar Animation Studios
Richmond, CA 94804
kurt@pixar.com

Alan H. Barr
Computer Graphics Laboratory
Department of Computer Science
Division of Engineering and Applied Science
California Institute of Technology
Pasadena, CA 91125
barr@gg.caltech.edu.

The distribution of different material types can be identified in volumetric datasets such as those produced with Magnetic Resonance Imaging (MRI) or Computed Tomography (CT). By allowing mixtures of materials and treating voxels as regions, the technique presented in this chapter reduces errors that other classification techniques can create along boundaries between materials and is

[1]Based on "Partial-Volume Bayesian Classification of Material Mixtures in MR Volume Data using Voxel Histograms" by David H. Laidlaw, Kurt W. Fleischer, and Alan H. Barr which appeared in IEEE Transactions on Medical Imaging, Vol. 17, No. 1, pp 74–86.

particularly useful for creating accurate geometric models and renderings from volume data. It also has the potential to make volume measurements more accurately and classifies noisy, low-resolution data well.

There are two unusual aspects to the approach. First, it uses the assumption that, due to partial-volume effects, or blurring, voxels can contain more than one material, e.g., both muscle and fat; it computes the relative proportion of each material in the voxels. Second, the approach incorporates information from neighboring voxels into the classification process by reconstructing a continuous function, $\rho(x)$, from the samples and then looking at the distribution of values that $\rho(x)$ takes on within the region of a voxel. This distribution of values is represented by a histogram taken over the region of the voxel; the mixture of materials that those values measure is identified within the voxel using a probabilistic Bayesian approach that matches the histogram by finding the mixture of materials within each voxel most likely to have created the histogram. The size of regions that are classified is chosen to match the spacing of the samples because the spacing is intrinsically related to the minimum feature size that the reconstructed continuous function can represent.

12.1 Introduction

Identifying different materials within sampled datasets can be an important step in understanding the geometry, anatomy, or pathology of a subject. By accurately locating different materials, individual parts can be identified and their size and shape measured. The spatial location of materials can also be used to selectively visualize parts of the data, thus better controlling a volume-rendered image [1], a surface model [2], or a volume model created from the data, and making visible otherwise obscured or subtle features. Classification is a key step towards understanding such geometry,

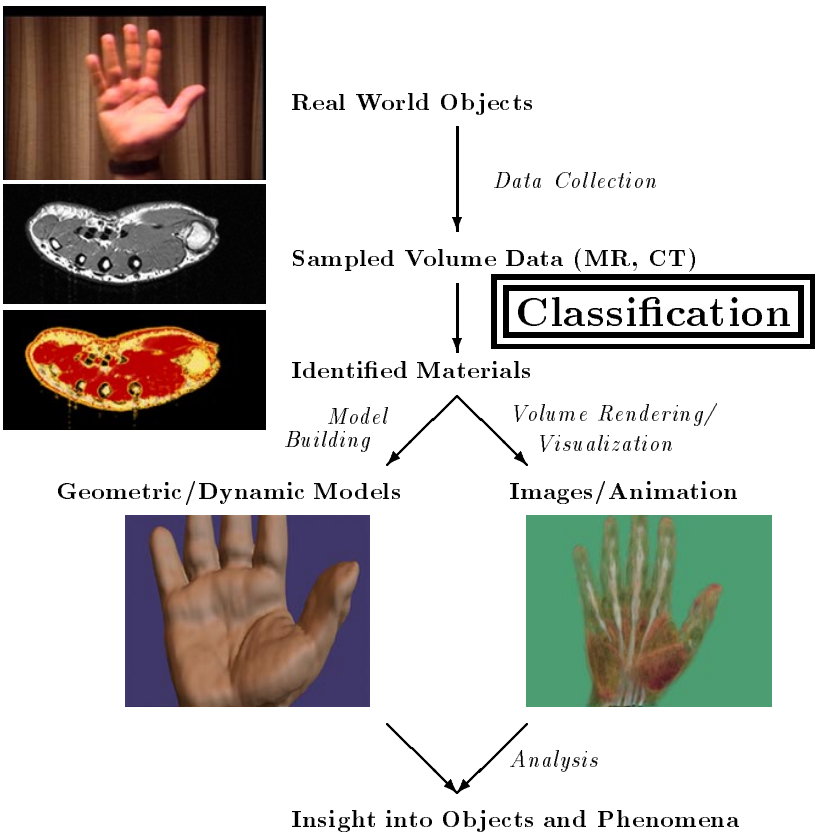
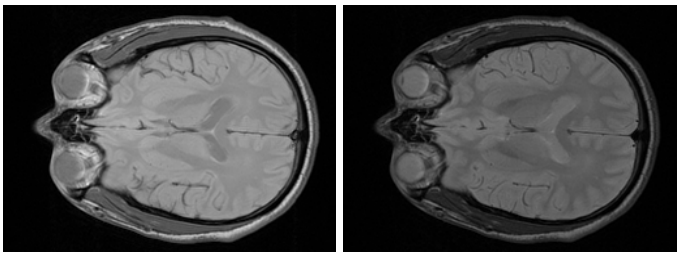
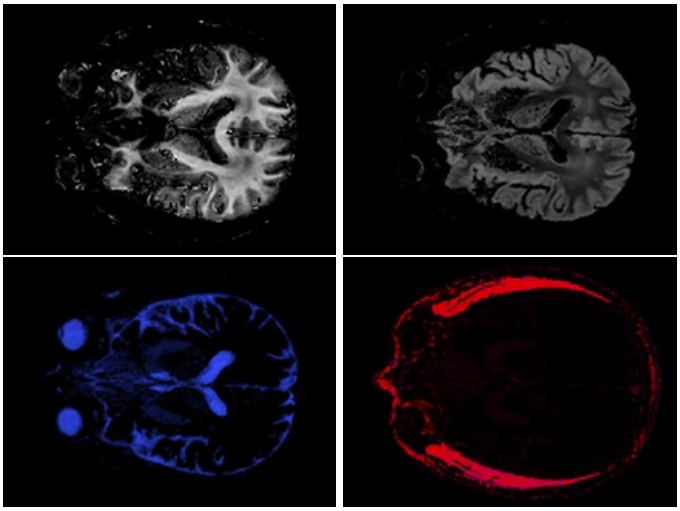


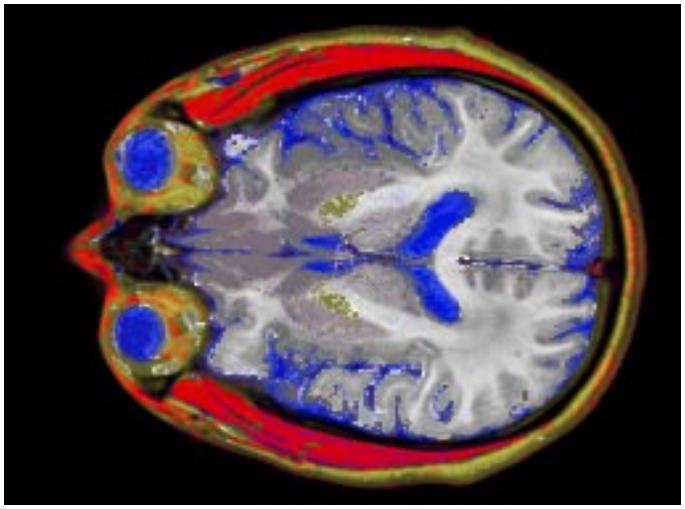
Figure 12.1: Classification is a key step in the process of the process of visualizing and extracting geometric information from sampled volume data. For accurate geometric results, some constraints on the classification accuracy must be met.



(a) Original Data



(b) Results of Algorithm
Classified White Matter (white), Gray Matter (gray)
Cerebro-Spinal Fluid (blue), Muscle (red)



(c) Combined Classified Image

Figure 12.2: One slice of data from a human brain. (a) The original two-valued MRI data. (b) Four of the identified materials, white matter, gray matter, cerebro-spinal fluid, and muscle, separated out into separate images. (c) Overlaid results of the new classification mapped to different colors. Note the smooth boundaries where materials meet and the much lower incidence of misclassified samples than in Figure 12.5.

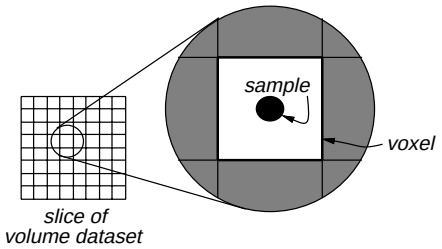


Figure 12.3: Definitions: a *sample* is a scalar or vector valued element of a 2-D or 3-D dataset; a *voxel* is the region surrounding a sample.

as shown in Figure 12.1. Figure 12.2 shows an example of classified MRI data; each color represents a single material identified within the data.

Applications of classified images and geometric models derived from them include surgical planning and assistance, diagnostic medical imaging, conventional computer animation, anatomical studies, and predictive modeling of complex biological shapes and behavior.

12.1.1 A partial-volume classification approach using voxel histograms.

Bayesian probability theory can be used to estimate the highest-probability combination of materials within each voxel-sized region. The estimation is based on the histogram of data values within the region. The posterior probability, which is maximized, is based on conditional and prior probabilities derived from the assumptions about what is being measured and how the measurement process works [3]. With this information the materials contained within each voxel can be identified based on the sample values for the voxel and its neighbors. Each voxel is treated as a region (see Figure 12.3), not as a single point. The sampling theorem [4] allows the reconstruction of a continuous function, $\rho(x)$, from the samples. All of the values that $\rho(x)$ takes on within a voxel are then represent by a histogram of $\rho(x)$ taken over the voxel. Figure 12.4(a) shows samples,

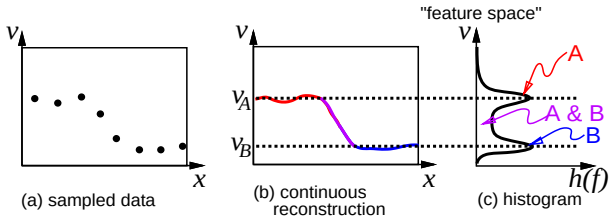


Figure 12.4: Continuous histograms. The scalar data in (a) and (b) represent measurements from a dataset containing two materials, A and B, as shown in Figure 12.6. One material has measurement values near v_A and the other near v_B. These values correspond to the Gaussian-shaped peaks centered around v_A and v_B in the histograms, which are shown on their sides to emphasize the axis that they share. This shared axis is "feature space."

Figure 12.4(b) shows the function $\rho(x)$ reconstructed from the samples, and Figure 12.4(c) shows a continuous histogram calculated from $\rho(x)$.

Each voxel is assumed to be a mixture of materials, with mixtures occurring where partial-volume effects occur, i.e., where the band-limiting process blurs measurements of pure materials together. From this assumption, basis functions are derived that model histograms of voxels containing a pure material and of voxels containing a mixture of two materials. Linear combinations of these basis histograms are fit to each voxel, and the most likely combination of materials is chosen probabilistically.

The regions that are classified could be smaller or larger than voxels. Smaller regions would include less information, and so the context for the classification would be reduced and accuracy would suffer. Larger regions would contain more complicated geometry because the features that could be represented would be smaller than the region. Again, accuracy would suffer. Because the spacing of sample values is intrinsically related to the minimum feature size that the reconstructed continuous function, $\rho(x)$, can represent, that spacing is a natural choice for the size of regions to be classified.

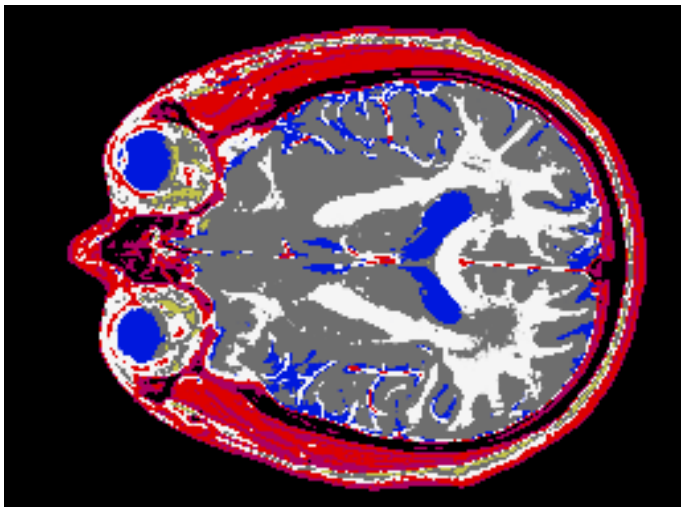


Figure 12.5: Discrete, maximum-likelihood (DML) classification of the same brain data shown in Figure 12.2. This existing method assigns each voxel to a single material class. The class is identified here by its color: gray for gray matter, blue for CSF/fluid, white for white matter, red for muscle. Note the jagged boundaries between materials within the brain and the layer of misclassified white matter outside of the skull. See Section 12.7 for more detail.

12.1.2 Related work.

Many researchers have worked on identifying the locations of materials in sampled datasets [5], [6], [7], [8]. [9] gives an extensive review of the segmentation of MRI data. However, existing algorithms still do not take full advantage of all the information in sampled images; there remains room for improvement. Many of these algorithms generate artifacts like those shown in Figure 12.5, an example of data classified with a maximum-likelihood technique based on sample values. These techniques work well in regions where a voxel contains only a single material, but tend to break down at boundaries between materials. In Figure 12.5 note the introduction of both stair-step artifacts, as shown between gray matter and white matter within the brain, and thin layers of misclassified voxels, as shown by the white matter between the skull and the skin. Both types of artifacts can be ascribed to the partial-volume effects ignored by the segmentation algorithms and to the assignment of discrete material types to each voxel.

[10] presents a technique that uses *a priori* information about brain anatomy to avoid the layers of misclassified voxels. However, this work still produces a classification where each voxel is assigned to a single, discrete material; results continue to exhibit stair-step artifacts. It is also very dependent on brain anatomy information for its accuracy; broader applicability is not clear.

[11] demonstrates that accounting for mixtures of materials within a voxel can reduce both types of artifacts, and approximates the relative volume of each material represented by a sample as the probability that the sample is that material. Their technique works well for differentiating air, soft tissue, and bone in CT data, but not for differentiating materials in MR data, where the measured data value for one material is often identical to the measured value for a mixture of two other materials.

[12] and [13] avoid partial-volume artifacts by taking linear combinations of components of vector measurements. An advantage of their techniques is that the linear operations they perform preserve the partial-volume mixtures within each sample value, and so partial-volume artifacts are not created. A disadvantage is that the linear operations are not as flexible as non-linear operations, and so either more data must be acquired or classification results will not be as accurate.

[14] and [15] address the partial-volume issue by identifying combinations of materials for each sample value. As with many other approaches to identifying mixtures, these techniques use only a single measurement taken within a voxel to represent its contents. Without the additional information available within each voxel region, these classification algorithms are limited in their accuracy.

[16] derives a distribution of data values taken on for partial volume mixtures of two materials. The technique described here shares the distribution that they derive. Their application of the

distribution, however, fits a histogram of an entire dataset and then quantifies material amounts over the entire volume. In contrast with this work, they represent each voxel with a single measurement for classification purposes, and do not calculate histograms over single voxels.

[17] presents an interesting approach to partial-volume imaging that makes similar assumptions about the underlying geometry being measured and about the measurement process. The results of their algorithm are a material assignment for each sub-voxel of the dataset. Taken collectively, these multiple sub-voxel results provide a measure of the mixtures of materials within a voxel but arrive at it in a very different manner than is done here. This work has been applied to satellite imaging data, and so their results are difficult to compare with medical imaging results, but aspects of both may combine well.

[18] gives an overview of the technique presented below in the context of the Human Brain Project, and [19] gives a complete description. [20] describes an imaging protocol for acquiring MRI data from solids and applies the classification technique to the extraction of a geometric model from MRI data of a human tooth (see Figure 12.11).

12.2 Overview

This section describes the classification problem, defines terms, states assumptions made about the imaging data to be classified, and sketch the algorithm and its derivation. Sections 12.3–12.6 give more information on each part of the process, with detailed derivations in Sections 12.8 and 12.9. Section 12.7 shows results of the application of the algorithm to simulated MR data and to real MR data of a human brain, hand, and tooth. Some limitations and future extensions are discussed in in Section 12.10 and a summary and conclusion made in Section 12.11.

12.2.1 Problem statement.

The input to the classification process is sampled measurement data, from which a continuous, band-limited function, $\rho(x)$, can be reconstructed. $\rho(x)$ measures distinguishing properties of the underlying materials. The output is sampled data measuring the relative volume of each material in each voxel.

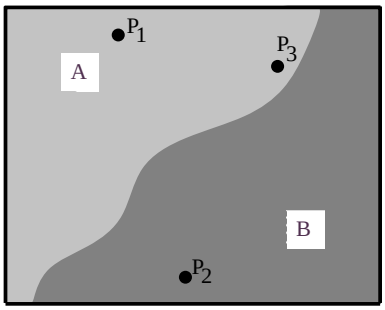
12.2.2 Definitions.

The coordinate system of the space containing the object to be measured is referred to as "spatial coordinates," and points are generally labeled $x \in X$. This space is n-dimensional, where n is three for volume data, can be two for slices, and is one for the example in Figure 12.4. Each measurement, which may be a scalar or vector, lies in "feature space" (see Figure 12.4), with points frequently denoted as $v \in V$. Feature space is n_v-dimensional, where n_v is one for scalar-valued data, two for two-valued vector data, etc. Tables 12.5 and 12.6 in Section 12.9 contain these and other definitions.

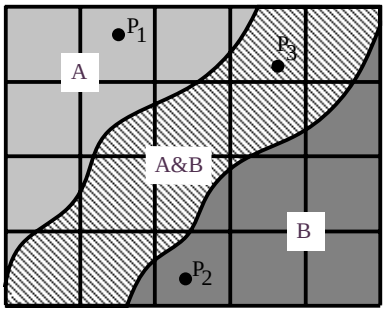
12.2.3 Assumptions.

The following set of assumptions are made about the measured objects and the measurement process.

1: **Discrete materials.** The first assumption is that materials within the objects to be measured are discrete at the sampling resolution. Boundaries need not be aligned with the sampling grid. Figure 12.6(a) shows an object with two materials. This assumption is made because



(a) Real World Object



(b) Sampled Data

Figure 12.6: Partial-volume effects. The derivation of the classification technique starts from the assumption that in a real-world object each point is exactly one material, as in (a). The measurement process creates samples that mix materials together; from the samples a continuous, band-limited measurement function, $\rho(x)$, is reconstructed. Points P_1 and P_2 lie inside regions of a single material. Point P_3 lies near a boundary between materials, and so in (b) lies in the A&B region where materials A and B are mixed. The grid lines show sample spacing and illustrate how the regions may span voxels.

the technique is geared toward finding boundaries between materials, and because its input is sampled data, where information about detail finer than the sampling rate is blurred.

This assumption does not preclude homogeneous combinations of sub-materials that can be treated as a single material at the sampling resolution. For example, muscle may contain some water, and yet be treated as a separate material from water. This assumption is not satisfied where materials gradually transition from one to another over many samples or are not relatively uniformly mixed; however, the algorithm appears to degrade gracefully even in these cases.

2: **Normally-distributed noise.** The second assumption is that noise from the measurement process is added to each discrete sample and that the noise is normally distributed. A different variance in the noise for each material is assumed. This assumption is not strictly satisfied for MRI data, but seems to be satisfied sufficiently to classify data well. Note that the sample values with noise added are interpolated to reconstruct the continuous function, $\rho(x)$. The effect of this band-limited noise is discussed further in Section 12.6.

3: **Sampling theorem is satisfied.** The third assumption is that the sampled datasets satisfy the sampling theorem [4]. The sampling theorem states that if a sufficiently band-limited function is point sampled, the function can be exactly reconstruct from the samples, as demonstrated in Figure 12.4(b). The band-limiting creates smooth transitions in $\rho(x)$ as it traverses boundaries where otherwise $\rho(x)$ would be discontinuous. The intermediate region of Figure 12.6(b) shows a sampling grid and the effect of sampling that satisfies the sampling theorem. Partial-volume mixing of measurements occurs in the region labeled “A & B.”

Multi-slice MRI acquisitions do not satisfy this assumption in the through-plane direction. For these datasets the data can be interpolated only within each plane.

4: **Linear mixtures.** Each voxel measurement is a linear combination of pure material measurements and measurements of their pair-wise mixtures created by band limiting (see Figure 12.6).

5: **Uniform tissue measurements.** Measurements of the same material have the same expected value and variance throughout a dataset.

6: **Box filtering for voxel histograms.** The spatial measurement kernel, or point-spread function, can be approximated by a box filter for the purpose of deriving histogram basis functions.

7: **Materials identifiable in histogram of entire dataset.** The signatures for each material and mixture must be identifiable in a histogram of the entire dataset.

For many types of medical imaging data, including MRI and CT, these assumptions hold reasonably well, or can be satisfied sufficiently with preprocessing [21]. Other types of sampled data, e.g., ultrasound, and video or film images with lighting and shading, violate these assumptions, thus the technique described here does not apply directly to them.

12.2.4 Sketch of derivation.

Histograms represent the values taken on by $\rho(x)$ over various spatial regions. Section 12.3 describes the histogram equation for a normalized histogram of data values within a region. Section 12.4 described how the histogram equation can be used to create basis functions that model

histograms taken over small, voxel-sized regions. These basis functions model histograms for regions consisting of single materials and for regions consisting of mixtures of two materials. Using Bayes' Theorem, the histogram of an entire dataset, the histogram model basis functions, and a series of approximations, Section 12.5 derives an estimate of the most likely set of materials within an entire dataset. Similarly, given the histogram of a voxel-sized region, Section 12.6 derives an estimate of the most likely density for each material in that voxel. The classification process is illustrated in Figure 12.7.

12.3 Normalized histograms

This section presents the equation for a normalized histogram of a sampled dataset over a region. This equation will be used as a building block in several later sections, with regions that vary from the size of a single voxel to the size of the entire dataset. It will also be used to derive basis functions that model histograms over regions containing single materials and regions containing mixtures of materials.

For a given region in spatial coordinates, specified by $\mathcal{R}$, the histogram $h^{\mathcal{R}}(v)$ specifies the relative portion of that region where $\rho(x) = v$, as shown in Figure 12.4. Because a dataset can be treated as a continuous function over space, histograms, $h^{\mathcal{R}}(v) : \mathbf{R}^{n_v} \rightarrow \mathbf{R}$, are also continuous functions:

$$h^{\mathcal{R}}(v) = \int \mathcal{R}(x)\delta(\rho(x) - v)dx \tag{12.1}$$

Equation 12.1 is the continuous analogue of a discrete histogram. $\mathcal{R}(x)$ is non-zero within the

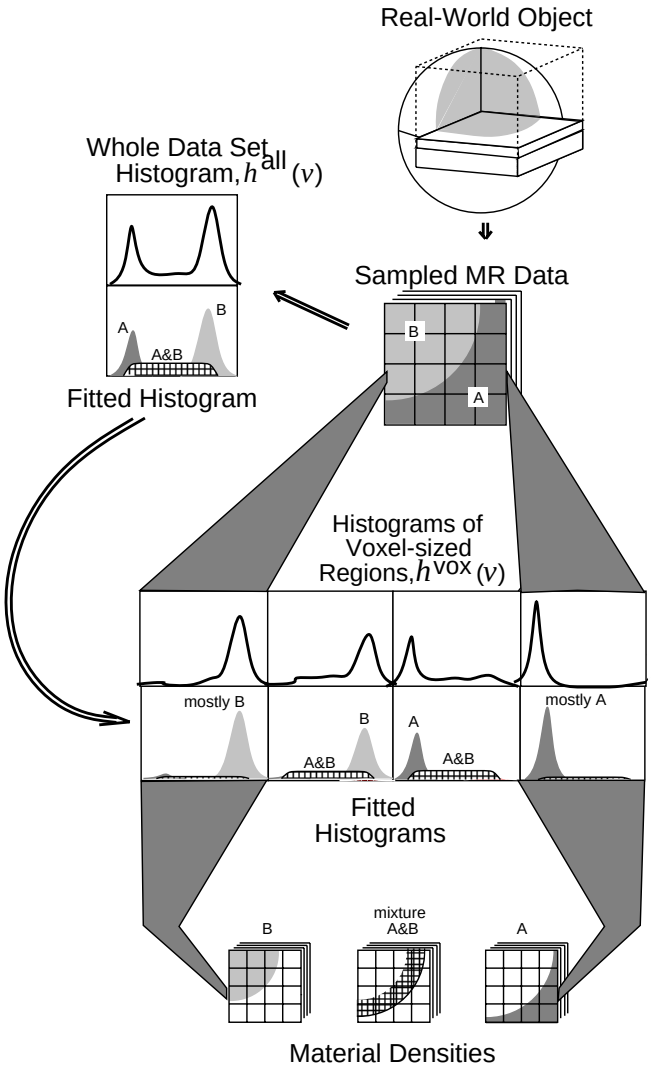


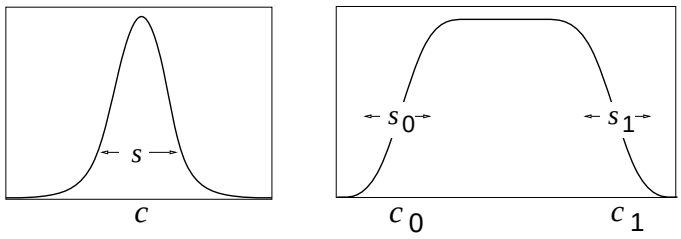
Figure 12.7: The classification process. MR data is collected, a histogram of the entire dataset, $h^{\mathrm{all}}(v)$, calculated and used to determine parameters of histogram-fitting basis functions. One basis function represents each pure material and one represents each mixture in the dataset. histograms are then calculated for each voxel-sized region, $h^{\mathrm{vox}}(v)$, and used to identify the most likely mixture of materials for that region. The result is a sampled dataset of material densities within each voxel.

region of interest and integrates to one. $\mathcal{R}(x)$ is set constant in the region of interest, making every spatial point contribute equally to the histogram $h^{\mathcal{R}}(v)$, but $\mathcal{R}(x)$ can be considered a weighting function that takes on values other than zero and one to more smoothly transition between adjacent regions. Note also that $h^{\mathcal{R}}(v)$ integrates to one, which means that it can be treated as a probability density function, or PDF. δ is the Dirac-delta function.

12.3.1 Computing voxel histograms.

Histograms can be calculated in constant-sized rectangular "bins," sized such that the width of a bin is smaller than the standard deviation of the noise within the dataset. This ensures that significant features are not lost in the histogram.

The bins are first initialized to zero. Each voxel is subdivided into sub-voxels, usually 4 for 2-D data or 8 for 3-D data, and $\rho(x)$ and its derivative evaluated at the center of each sub-voxel. $\rho(x)$ is interpolated from the discrete data using a tri-cubic B-spline basis [22] that approximates a Gaussian. Thus, function and derivative evaluations can be made not only at sample locations, but anywhere between samples as well. From the function value and the derivative, Equation 12.1 is used to calculate the contribution of a linear approximation of $\rho(x)$ over the sub-voxel to each histogram bin, accumulating the contributions from all sub-voxels. This provides a more-accurate histogram than would be obtained by evaluating only the function values at the same number of points.



(a) (b)

Figure 12.8: Parameters for histogram basis function. (a) Single-material histogram parameters include c, the mean value for the material, and s, which measures the standard deviation of measurements (see Equation 12.2). (b) Corresponding parameters for a two-material mixture basis function. s_0 and s_1 affect the slopes of the two-material histogram basis function at either end. For vector-valued data, c and s are vectors and are the mean values and standard deviations of the noise for the two constituent materials (see Equation 12.3).

12.4 Histogram basis functions for pure materials and mixtures

This section describes basis functions that model histograms of regions consisting of pure materials and regions consisting of pairwise mixtures of materials. Other voxel contents are also possible and are discussed in Section 12.10. The parameters of the basis functions specify the expected value, c, and standard deviation, s, of each material's measurements (see Figure 12.8).

Equation 12.1 can be used to derive these basis functions which are subsequently fit to histograms of the data. The equations provide reasonable fits to typical MR data, which gives confidence that the assumptions about the measurement function, $\rho(x)$, are reasonable. The details of the derivations are in Section 12.8.

For a single material, the histogram basis function is a Gaussian distribution:

$$f_{\text{single}}(v; c, s) = \left(\prod_{i=1}^{n_v} \frac{1}{s_i \sqrt{2\pi}} \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{n_v} \left(\frac{v_i - c_i}{s_i} \right)^2 \right), \tag{12.2}$$

where c is the vector-valued mean, s the vector-valued standard deviation, and $v_i, c_i,$ and s_i scalar components of $v, c,$ and s, respectively. This equation is derived by manipulating Equation 12.1 evaluated over a region of constant material, where the measurement function, $\rho(x)$, is a constant value plus additive, normally-distributed noise. Because the noise in different channels of multi-valued MRI images is not correlated, the general vector-valued normal distribution reduces to this equation with zero co-variances.

For mixtures along a boundary between two materials, another equation can be derived similarly:

$$f_{\text{double}}(v; c, s) = \int_0^1 k_n((1-t)c_1 + tc_2 - v; s)dt \tag{12.3}$$

As with the single material case, this derivation follows from Equation 12.1 evaluated over a region where two materials mix. In this case, the band-limiting filter that causes partial-volume effects is approximated with a box filter and an assumption is made that the variance of the additive noise is constant across the region. This basis function is a superposition of normal distributions representing different amounts of the two constituent pure materials. k_n is the normal distribution, centered at zero, t the relative quantity of the second material, c (comprised of c_1 and c_2) the expected values of the two materials, and s the standard deviation of measurements.

The assumption of a box filter affects the shape of the resulting histogram basis function. Similar equations for different filters (triangle, Gaussian, and Hamming) can also be derived, but box filter is sufficiently accurate in practice and is numerically more efficient.

12.5 Estimating histogram basis function parameters

This section describes parameter-estimation procedures for fitting histogram basis functions to a histogram of an entire dataset. For a given dataset the histogram, $h^{\text{all}}(v)$, is first calculated over the entire dataset. The second step combines an interactive process of specifying the number of materials and approximate feature-space locations for them with an automated optimization [21] to refine the parameter estimates. Under some circumstances, users may wish to group materials with similar measurements into a single "material," whereas in other cases they may wish the materials to be separate. The result of this process is a set of parameterized histogram basis functions, together with values for their parameters. The parameters describe the various materials and mixtures of interest in the dataset. Figure 12.9 shows the results of fitting a histogram. Each colored region represents one distribution, with the labeled spot-shaped regions representing pure materials and connecting shapes representing mixtures.

To fit a group of histogram basis functions to a histogram, as in Figure 12.9, the optimization process estimates the relative volume of each pure material or mixture (vector α^{all}), and the mean value (vector c) and standard deviation (vector s) of measurements of each material. The process is derived from the assumption that all values were produced by pure materials and two-material mixtures. n_m is the number of pure materials in a dataset, and n_f the number of histogram basis functions. Note that $n_f \geq n_m$, since n_f includes any basis functions for mixtures, as well as those for pure materials.

The optimization minimizes the function

$$\mathcal{E}(\alpha^{\text{all}}, c, s) = \frac{1}{2} \int \left(\frac{q(v; \alpha^{\text{all}}, c, s)}{w(v)} \right)^2 dv, \tag{12.4}$$

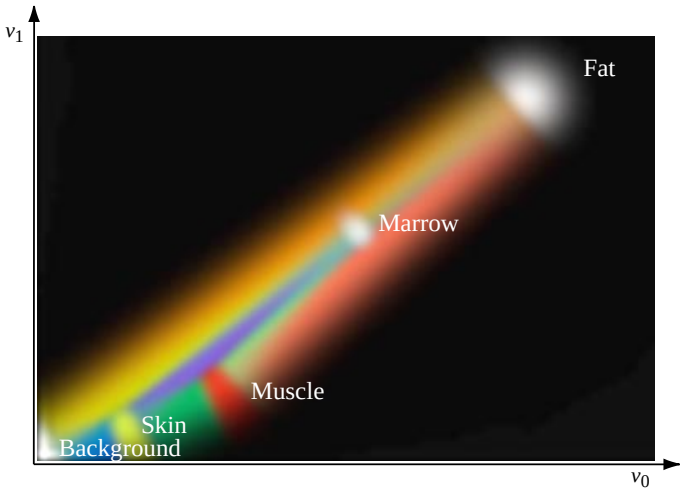


Figure 12.9: Basis functions fit to histogram of entire dataset. This figure illustrates the results of fitting basis functions to the histogram of the hand dataset. The five labeled circular regions represent the distribution of data values for pure materials, while the colored regions connecting them represent the distribution of data values for mixtures. The mixture between muscle (red) and fat (white), for example, is a salmon-colored streak. The green streak between the red and yellow dots is a mixture of skin and muscle. These fitted basis functions were used to produce the classified data used in Figure 12.12.

with respect to $\alpha^{\text{all}}, c,$ and s, where:

$$q(v; \alpha^{\text{all}}, c, s) = h^{\text{all}}(v) - \sum_{j=1}^{n_f} \alpha_j^{\text{all}} f_j(v; c_j, s_j) \tag{12.5}$$

Note that f_j may be a pure or a mixture basis function and that its parameter c_j will be a single feature-space point for a pure material or a pair for a mixture. The function $w(v)$ is analogous to a standard deviation at each point, v, in feature space, and gives the expected value of $|q(v)|$. $w(v)$ can be approximated as a constant; it is discussed further in Section 12.10.

Equations 12.4 and 12.5 are derived in Section 12.9 using Bayesian probability theory with estimates of prior and conditional probabilities.

12.6 Classification

This section describes the process of classifying each voxel. This process is similar to that described in Section 12.5 for fitting the histogram basis functions to the entire dataset histogram, but now histograms taken over small, voxel-sized regions are being fitted. The previously computed histogram basis functions calculated from the entire dataset histogram are used. The mean vector, c, and standard deviation, s are no longer varied. The only parameters allowed to vary are the relative material volumes (vector α^{vox}), and an estimate of the local noise in the local region (vector $\bar{N}$) (see Equations 12.6 and 12.7).

Over large regions including many voxels, the noise in $\rho(x)$ is normally distributed, with zero mean; however, for voxel regions the noise mean is generally non-zero. This is because normally distributed noise is added to each sample value, not to each point of $\rho(x)$. When the samples are used to reconstruct $\rho(x)$, the values $\rho(x)$ takes on near a particular sample tend to be similar, and so have a non-zero mean. The local mean voxel noise value is labeled $\bar{N}$. As derived in Section 12.9, the equation that is minimized, with respect to α^{vox} and $\bar{N}$, is:

$$\mathcal{E}(\alpha^{\text{vox}}, \bar{N}) = \frac{1}{2}\sum_{i=1}^{n_v}\left(\frac{\bar{N}_i}{\sigma_i}\right)^2 + \frac{1}{2}\int\left(\frac{q(v;\alpha^{\text{vox}}, \bar{N})}{w(v)}\right)^2 dv \tag{12.6}$$

where

$$q(v;\alpha^{\text{vox}}, \bar{N}) = h^{\text{vox}}(v - \bar{N}) - \sum_{j=1}^{n_f}\alpha_j^{\text{vox}} f_j(v), \tag{12.7}$$

the minimization is subject to the constraints

$$0 \le \alpha_j^{\text{vox}} \le 1, \text{ and } \sum_{j=1}^{n_f}\alpha_j^{\text{vox}} = 1,$$

and vector σ is the standard deviation of the noise over the entire dataset. For MR data the standard deviations in the signals for different materials are reasonably similar, and σ is estimated to be an average of the standard deviations of the histogram basis functions.

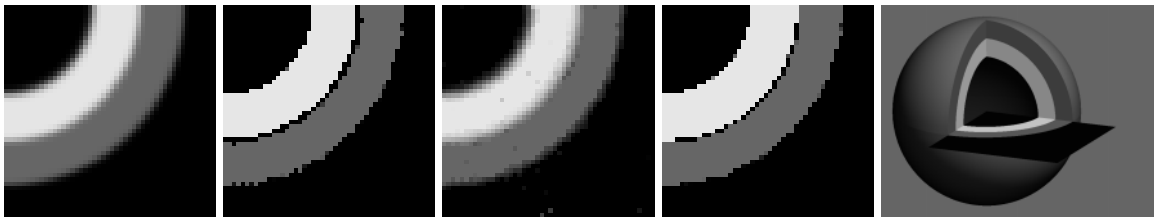
With optimal vector α^{vox} for a given voxel-sized region and the mean value, vector $\bar{v}$, within that region, the amount of each pure material contributed by each mixture to the voxel is estimated. This is the output, estimates of the amount of each pure material in the voxel-sized region.

$$\bar{v} = \int \left(h^{\text{vox}}(v) - \sum_{i=1}^{n_m} \alpha_i f_{\text{single}}(v) \right) dv \tag{12.8}$$

$\bar{v}$ contains the mean signature of the portion of the histogram that arises only from regions with partial-volume effects. The algorithm determines how much of each pure component of pairwise mixture materials would be needed to generate $\bar{v}$, given the amount of each mixture that α^{vox} indicates is in the voxel. t_k represents this relative amount for mixture k, with $t_k = 0$ indicating that the mixture is comprised of only the first pure component, $t_k = 1$ indicating that it is comprised of only its second component, and intermediate values of t_k indicating intermediate mixtures. The t_k values are calculated by minimizing the following equation with respect to t, subject to the constraint $0 \leq t_k \leq 1$.

$$\mathcal{E}_{\bar{v}}(t) = \left(\bar{v} - \sum_{k=n_m+1}^{n_f} \alpha_k \left(t_k c_{ka} + (1 - t_k) c_{kb} \right) \right)^2 \tag{12.9}$$

Vector c_{ka} is the mean value for the first pure material component of mixture k, and vector c_{kb} the mean value for the second component. The total amount of each material is the amount of pure material added to the t_k-weighted portion of each mixture.



(a) ground truth (b) DML (c) PVB (d) PPVC (e) Slice Geometry

Figure 12.10: Comparison of DML classification (b), the PVB classification (c), and PPVC classification (d). (a) is a reference for what "ideal" classification should produce. Note the band of dark background material in (b) and (d) between the two curved regions. This band is incorrectly classified, and could lead to errors in models or images produced from the classified data. The original dataset is simulated, two-valued data of two concentric shells, as shown in (e), with SNR of 14.2.

12.7 Results

This section shows the results of voxel-histogram classification applied to both simulated and collected MRI datasets. When results can be verified and conditions are controlled, as shown with the classification of simulated data, the algorithm comes very close to "ground truth," or perfect classification. The results based on collected data illustrate that the algorithm works well on real data, with a geometric model of a tooth showing boundaries between materials, a section of a human brain showing classification results mapped on to colors, and a volume-rendered image of a human hand showing complex geometric relationships between different tissues.

The Partial Volume Bayesian algorithm (PVB) described in this chapter is compared with four other algorithms. The first, DML (Discrete Maximum Likelihood), assigns each voxel or sample to a single material using a Maximum Likelihood algorithm. The second, PPVC (Probabilistic Partial Volume Classifier), is described in [23]. The third is a Mixel classifier [14] and the fourth is eigenimage filtering (Eigen)[12].

PVB significantly reduces artifacts introduced by the other techniques at boundaries between

materials. Figure 12.10 compares performance of PVB, DML and PPVC on simulated data. PVB produces many fewer misclassified voxels, particularly in regions where materials are mixed due to partial-volume effects. In Figure 12.10(b) and (d) the difference is particularly noticeable where an incorrect layer of dark background material has been introduced between the two lighter regions, and where jagged boundaries occur between each pair of materials. In both cases this is caused by partial-volume effects, where multiple materials are present in the same voxel.

Table 12.1 shows comparative RMS error results for the PPVC, Eigen, and PVB simulated data results, and also compares PPVC with the Mixel algorithm. Signal-to-noise ratio (SNR) for the data used in PPVC/Eigen/PVB comparison was 14.2. SNR for the data used in PPVC/Mixel comparison was 21.6. Despite lower SNR, PPVC/PVB RMS error improvement is approximately double that of the PPVC/Mixel improvement. RMS error is defined as $\sqrt{\sum_x (\alpha(x) - p(x))^2}$, where $\alpha(x)$ is classified data and $p(x)$ is ground truth. The sum is made only over voxels that contain multiple materials.

Table 12.2 shows similar comparative results for volume measurements made between PPVC and PVB on simulated data, and between PPVC and Mixel on real data. Volume measurements made with PVB are significantly more accurate that those made with PPVC, and the PPVC to PVB improvement is better than the PPVC to Mixel improvement. Table 12.3 compares noise levels in PVB results and Eigen results. The noise level for the PVB results is about 25% of the level for the Eigen results.

Figures 12.2 and 12.5 also show comparative results between PVB and DML. Note that the same artifacts shown in Figure 12.10 occur with real data and are reduced by the technique described here.

Table 12.1: Comparative RMS error for the four algorithms: PVB, PPVC, Mixel, and Eigen. The PPVC/Eigen/PVB comparison is from the simulated data test case illustrated in Figure 12.10, SNR=14.2. The PPVC/Mixel comparison is taken from Figures 7 and 8 in [14], SNR=21.6. PVB, in the presence of more noise, reduces the PPVC RMS error to approximately half that of the Mixel algorithm.

	RMS Error			Improvement Ratio
	PPVC	Eigen	PVB	PPVC/PVB
Background	20%	11.7%	6.5%	3.09
Outer	25%	24.1%	4.3%	5.79
Inner	20%	9.8%	6.5%	3.04
	PPVC	Mixel		PPVC/Mixel
Background	16%	9.5%		1.68
Tumor	21%	13.5%		1.56
White matter	37%	16.0%		2.31
Gray matter	36%	17.0%		2.11
CSF	18%	13.0%		1.38
All other	20%	10.0%		2.00

Table 12.2: Comparative volume measurement error for four algorithms (PVB, PPVC, Mixel, and Eigen). The PPVC/Eigen/PVB comparison is from the simulated data test case illustrated in Figure 12.10, SNR=14.2. Note that the Eigen results are based on 3-valued data while the other algorithms used 2-valued data. The PPVC/Mixel comparison is taken from Figure 9 and Table V in [14], SNR=21.6.

PPVC	Eigen	PVB	PPVC	Mixel
2.2%	-0.021%	0.004%	5.6%	1.6%
-5.3%	0.266%	-0.452%	44.1%	7.0%
0.3%	-0.164%	0.146%		

Table 12.3: Comparison of voxel histogram classification (PVB) with eigenimage filtering (Eigen) for voxels having no partial volume effects. Desired signatures should be mapped to 1.0 and undesired signatures to 0.0. Note that the PVB classification has consistently smaller standard deviations – the Eigen results have noise levels 2-4 times higher despite having 3-valued data to work with instead of the 2-valued data PVB was given.

	Eigen (3-valued data)		PVB (2-valued data)	
	mean	Std. Dev.	mean	Std. Dev.
desired signatures				
Material 1	1.0113	0.241	0.9946	0.064
Material 2	0.9989	0.124	0.9926	0.077
Background	0.9986	0.113	0.9976	0.038
undesired signatures				
Material 1	-0.0039	0.240	0.0013	0.017
Material 2	-0.0006	0.100	0.0002	0.004
Background	0.0016	0.117	0.0065	0.027

Table 12.4: MRI dataset sources, acquisition parameters, and figure references.

Object	Machine	Voxel Size mm	$T_R/T_{E_1}/T_{E_2}$ s/ms/ms	Figs.
shells	simulated	$1.9^2 \times 3$	N/A	12.10
brain	GE	$0.94^2 \times 3$	$2/25/50$	12.2, 12.5
hand	GE	$0.7^2 \times 3$	$2/23/50$	12.12
tooth	Bruker	0.312^3	$15/0.080$	12.11

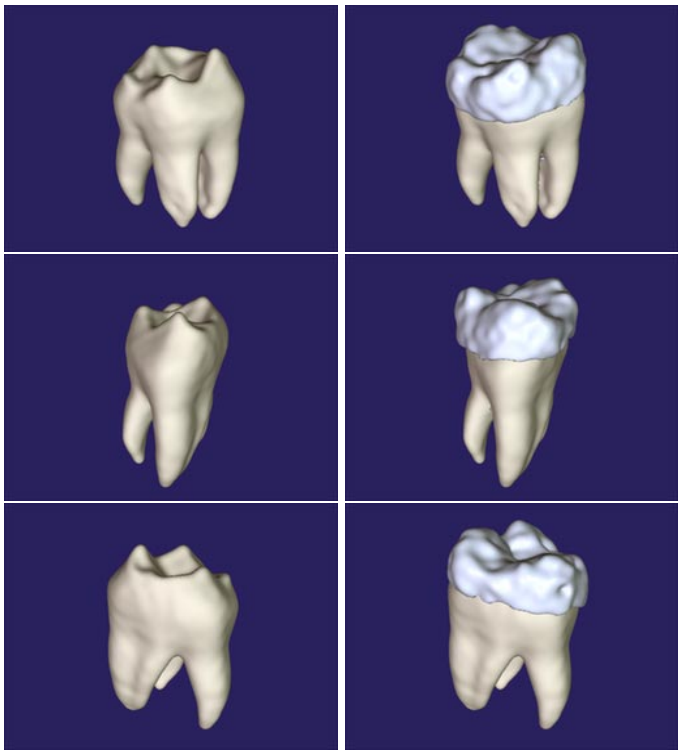


Figure 12.11: A geometric model of tooth dentine and enamel created by collecting MRI data samples using a technique that images hard solid materials [20] and classifying dentine, enamel, and air in the volume data with the PVB algorithm. Polygonal isosurfaces define the bounding surfaces of the dentine and enamel. The enamel-dentine boundary, shown in the left images, is difficult to examine non-invasively using any other technique.

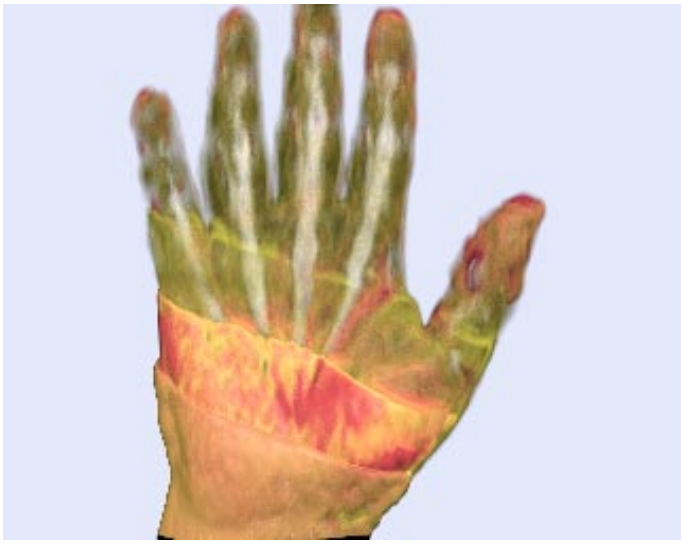


Figure 12.12: A volume-rendering image of a human hand dataset. The opacity of different materials is decreased above cutting planes to show details of the classification process within the hand.

Models and volume-rendered images, as shown in Figures 12.11 and 12.12, benefit from the PVB technique because less incorrect information is introduced into the classified datasets, thus the images and models more accurately depict the objects they are representing. Models and images such as these are particularly sensitive to errors at geometric boundaries because they illustrate the underlying geometries.

Table 12.4 lists the datasets, the MRI machine they were collected on, some collection parameters, the voxel size, and the figures in which each dataset appears. The GE machine was a 1.5T Signa. The Bruker machine was an 11.7T AMX500. Acquired data were collected with a spin-echo or fast spin-echo protocol, with one proton-weighted and one T_2-weighted acquisition. The tooth was acquired with a technique described in [20]. Preprocessing was only performed on data used for the hand example (Figure 12.12). For this case each axial slice was multiplied by a constant and then offset by another to compensate for intensity falloff as a function of the distance from the center of the RF coil. The constants were chosen to make the mean values of user-identified material regions consistent from slice to slice.

12.8 Derivation of histogram basis functions

This section derives parameterized model histograms that are used as basis functions, f_i, for fitting histograms of data. Two forms of basis functions are derived: one for single, pure materials; another for two-material mixtures that arise due to partial-volume effects in sampling. Equation 12.1, the histogram equation, is:

$$h^{\mathcal{R}}(v) = \int \mathcal{R}(x)\delta(\rho(x) - v)dx$$

and measures a histogram of the function $\rho(x)$ over a region defined by $\mathcal{R}(x)$. x ranges over spatial locations, and v over feature space. Note that if $\rho(x)$ contains additive noise, $n(x;s)$, with a particular distribution, $k_n(v;s)$, then the histogram of ρ with noise is the convolution of $k_n(v;s)$ with $\rho(x) - n(x;s)$ (i.e, $\rho(x)$ without noise). $k_n(v;s)$ is, in general, a normal distribution. Thus,

$$h^{\mathcal{R}}(v) = k_n(v;s) * \int \mathcal{R}(x)\delta((\rho(x) - n(x;s)) - v)dx \tag{12.10}$$

12.8.1 Pure Materials

For a single pure material it is assumed that the measurement function has the form:

$$\rho_{\text{single}}(x) = c + n(x;s) \tag{12.11}$$

where c is the constant expected value of a measurement of the pure material, and s is the standard deviation of additive, normally-distributed noise.

The basis function used to fit the histogram of the measurements of a pure material is

$$\begin{aligned}
f_{\text{single}}(v;c,s) = & \\
& \int \mathcal{R}(x)\delta(\rho_{\text{single}}(x) - v)dx \\
= & \int \mathcal{R}(x)\delta(c + n(x;s) - v)dx \\
= & k_n(v;s) * \int \mathcal{R}(x)\delta(c - v)dx \\
= & k_n(v;s) * \left(\delta(c - v)\int \mathcal{R}(x)dx\right) \\
= & k_n(v;s) * \delta(c - v)
\end{aligned}$$

$$
\begin{aligned}
&= k_n(v - c; s) \\
&= \left(\prod_{i=1}^{n_v} \frac{1}{s_i \sqrt{2\pi}} \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{n_v} \left(\frac{v_i - c_i}{s_i} \right)^2 \right)
\end{aligned} \tag{12.12}
$$

Thus, $f_{\text{single}}(v; c, s)$ is a Gaussian distribution with mean c and standard deviation s. v_i, c_i, and s_i are scalar components of $v, c,$ and s. The noise is assumed to be independent in each element of vector-valued data, which for MRI appears to be reasonable.

12.8.2 Mixtures

For a mixture of two pure materials, the measurement function is assumed to have the form:

$$\rho_{\text{double}}(x) = \ell_{\text{double}}(x; c_1, c_2) + n(x; s) \tag{12.13}$$

where ℓ_{double} approximates the band-limiting filtering process, a convolution with a box filter, by interpolating the values within the region of mixtures linearly between c_1 and c_2, the mean values for the two materials.

$$\ell_{\text{double}} = (1 - t)c_1 + tc_2 \tag{12.14}$$

$$
\begin{aligned}
&f_{\text{double}}(v; c, s) = \\
&\quad \int \mathcal{R}(x) \delta(\rho_{\text{double}}(x) - v) dx \\
&= \int \mathcal{R}(x) \delta(\ell_{\text{double}}(x; c_1, c_2) + n(x; s) - v) dx
\end{aligned}
$$

$$
\begin{aligned}
&= k_n(v;s) * \int \mathcal{R}(x)\delta(\ell_{\text{double}}(x;c_1,c_2) - v)dx \\
&= \int_0^1 k_n(v;s) * \delta((1-t)c_1 + tc_2 - v)dt \\
&= \int_0^1 k_n((1-t)c_1 + tc_2 - v;s)dt
\end{aligned} \tag{12.15}
$$

12.9 Derivation of classification parameter estimation

The following two sections can be safely skipped on a first reading. They present detailed derivations and information helpful for implementing the algorithm or for creating an analogous one.

This section contains a derivation of the equations that are used to find model histogram parameters and to classify voxel-sized regions. Bayesian probability theory [3] is employed to derive an expression for the probability that a given histogram was produced by a particular set of parameter values in the model. This "posterior probability" is maximized to estimate the best-fit parameters.

$$\text{maximize}\, P(\text{ parameters } | \text{ histogram }) \tag{12.16}$$

The optimization procedure is used for two purposes:

- **Find model histogram parameters**. Initially, it is used to estimate parameters of basis functions to fit histograms of the entire dataset h^{all}. This results in a set of basis functions that describes histograms of voxels containing pure materials or pairwise mixtures.

- **Classify voxel-sized regions**. Subsequently, the optimization procedure is used to fit a weighted sum of the basis functions to the histogram of a voxel-sized region h^{vox}. This produces the classification results (in terms of the weights α).

Table 12.5: Probabilities, using Bayesian terminology from [3].

$P(\alpha, c, s, N\|h)$	posterior probability (maximized)
$P(\alpha, c, s, \bar{N})$	prior probability
$P(h\|\alpha, c, s, \bar{N})$	likelihood
$P(h)$	global likelihood

Table 12.6: Definitions of terms used in the derivations.

Term	Dim.	Definition
n_m	scalar	number of pure materials
n_f	scalar	number of pure materials & mixtures
n_v	scalar	dim. of measurement (feature space)
α	n_f	relative volume of each mixture and material within the region
c	$n_f \times n_v$	mean of material measurements for each material
s	$n_f \times n_v$	standard deviation of material measurements (chosen by procedure discussed in Section 12.5) for each material
$\bar{N}$	n_v	mean value of noise over the region
p_{1-6}	scalars	arbitrary constants
$h^{\text{all}}(v)$	$\mathbf{R}^{n_v} \to \mathbf{R}$	histogram of an entire dataset
$h^{\text{vox}}(v)$	$\mathbf{R}^{n_v} \to \mathbf{R}$	histogram of a tiny, voxel-sized region

The posterior probabilities P^{all} and P^{vox} share many common terms. In the following derivation they are distinguished only where necessary, using P where their definitions coincide.

12.9.1 Definitions

Table 12.5 lists Bayesian probability terminology as used in [3] and in the derivations. Table 12.6 defines additional terms used in this section.

12.9.2 Optimization

The following optimization is performed to find the best-fit parameters:

$$\text{maximize } P(\alpha, c, s, \bar{N}|h) \tag{12.17}$$

With $P \equiv P^{\text{all}}$, the histogram basis function parameters c, s, α^{all}, are fit to the histogram of an entire dataset, $h^{\text{all}}(v)$. With $P \equiv P^{\text{vox}}$, the parameters α^{vox}, $\bar{N}$ are fit to classify the histogram of a voxel-sized region, $h^{\text{vox}}(v)$.

12.9.3 Derivation of the posterior probability, $P(\alpha, c, s, \bar{N}|h)$

The derivation begins with Bayes' Theorem, expressing the posterior probability in term of the likelihood, the prior probability, and the global likelihood.

$$P(\alpha, c, s, \bar{N}|h) = \frac{P(\alpha, c, s, \bar{N})P(h|\alpha, c, s, \bar{N})}{P(h)} \tag{12.18}$$

Each of the terms on the right side is approximated below, using p_{1-6} to denote positive constants (which can be ignored during the optimization process).

Prior probabilities.

It is assumed that α, c, s, and $\bar{N}$ are independent, so

$$P(\alpha, c, s, \bar{N}) = P(\alpha)P(c, s)P(\bar{N}) \tag{12.19}$$

Because the elements of α represent relative volumes, they are constrained to sum to 1 and are all positive.

$$P(\alpha) = \begin{cases} 0 & \text{if } \sum_{j=1}^{n_f} \alpha_j \neq 1 \\ 0 & \text{if } \alpha_j < 0 \text{ or } \alpha_j > 1 \\ p_1 & (\text{constant}) \text{ otherwise} \end{cases} \tag{12.20}$$

A different assumption is used for $P(c, s)$ depending on which fit is being done (h^{all} or h^{vox}). For fitting $h^{\text{all}}(v)$, all values of c, s are considered equally likely:

$$P^{\text{all}}(c, s) = p_6 \tag{12.21}$$

For fitting h^{vox}, the means and standard deviations, c, s, are fixed at c^0, s^0 (the values determined by the earlier fit to the entire data set):

$$P^{\text{vox}}(c, s) = \delta(c - c^0, s - s^0) \tag{12.22}$$

For a small region, it is assumed that the mean noise vector, $\bar{N}$, has normal distribution with standard deviation σ:

$$P^{\text{vox}}(\bar{N}) = p_2 e^{-\frac{1}{2}\sum_{i=1}^{n_v}\left(\frac{\bar{N}_i}{\sigma_i}\right)^2} \tag{12.23}$$

For a large region, the mean noise vector, $\bar{N}$, should be very close to zero; hence, $P^{\text{all}}(\bar{N})$ will be a delta function centered at $\bar{N} = 0$.

Likelihood.

The likelihood, $P(h|\alpha, c, s, \bar{N})$, is approximated by analogy to a discrete normal distribution. $q(v)$ is defined as the difference between the "expected" or "mean" histogram for particular $\alpha, c, s, \bar{N}$ and a given histogram $h(v)$:

$$q(v; \alpha, c, s, \bar{N}) = h(v - \bar{N}) - \sum_{j=1}^{n_f} \alpha_j f_j(v; c, s) \tag{12.24}$$

Now a normal-distribution-like function is created. $w(v)$ is analogous to the standard deviation of q at each point of feature space:

$$P(h|\alpha, c, s, \bar{N}) = p_3 e^{-\frac{1}{2}\int \left(\frac{q(v;\alpha,c,s,\bar{N})}{w(v)}\right)^2 dv} \tag{12.25}$$

Global likelihood.

Note that the denominator of Equation 12.18 is a constant normalization of the numerator:

$$P(h) = \int P(\hat{\alpha}, \hat{c}, \hat{s}, \hat{N}) P(h|\hat{\alpha}, \hat{c}, \hat{s}, \hat{N}) d\hat{\alpha} d\hat{c} d\hat{s} d\hat{N} \tag{12.26}$$

$$= p_4 \tag{12.27}$$

Assembly

Using the approximations discussed above, the following expression for the posterior probability can be calculated:

$$P(\alpha, c, s, \bar{N}|h) =$$

$$p_5 P(\alpha) P(c,s) \exp\left(-\frac{1}{2}\sum_{i=1}^{n_v}\left(\frac{\bar{N}_i}{\sigma_i}\right)^2\right) \exp\left(-\frac{1}{2}\int\left(\frac{q(v;\alpha,c,s,\bar{N})}{w(v)}\right)^2 dv\right) \tag{12.28}$$

For fitting h^{all}, the mean noise is assumed to be zero, so maximizing equation 12.28 is equivalent to minimizing $\mathcal{E}^{\text{all}}$ to find the free parameters $(\alpha^{\text{all}}, c, s)$:

$$\mathcal{E}^{\text{all}}(\alpha^{\text{all}}, c, s) = \frac{1}{2}\int\left(\frac{q(v;\alpha^{\text{all}}, c, s)}{w(v)}\right)^2 dv \tag{12.29}$$

subject to $P(\alpha^{\text{all}}) \neq 0$. Because both $P(\alpha^{\text{all}})$ and $P^{\text{all}}(c,s)$ are constant valued in that region, they are not included.

For fitting h^{vox}, the parameters c and s are fixed, so maximizing equation 12.28 is equivalent to minimizing $\mathcal{E}^{\text{vox}}$ to find the free parameters $(\alpha^{\text{vox}}, \bar{N})$:

$$\mathcal{E}^{\text{vox}}(\alpha^{\text{vox}}, \bar{N}) = \frac{1}{2}\sum_{i=1}^{n_v}\left(\frac{\bar{N}_i}{\sigma_i}\right)^2 + \frac{1}{2}\int\left(\frac{q(v;\alpha^{\text{vox}}, \bar{N})}{w(v)}\right)^2 dv \tag{12.30}$$

subject to $P(\alpha^{\text{vox}}) \neq 0$.

As stated in Equation 12.6, Section 12.6, Equation 12.30 is minimized to estimate relative material volumes, α^{vox}, and the mean noise vector, $\bar{N}$.

12.10 Discussion

Several assumptions and approximations were made while developing and implementing this algorithm. This section will discuss some of the tradeoffs, suggest some possible directions for further work, and consider some related issues.

12.10.1 Mixtures of three or more materials.

It was assumed that each measurement contains values from at most two materials. Two-material mixtures were chosen based on a dimensionality argument. In an object that consists of regions of pure materials, as shown in Figure 12.6, voxels containing one material will be most prevalent because they correspond to volumes. Voxels containing two materials will be next most prevalent, because they correspond to surfaces where two materials meet. As such, they are the first choice to model after those containing a single material. The approach can be extended in a straightforward manner to handle the three-material case as well as cases with other less-frequent geometries, such as skin, tubes, or points where four materials meet. This extension could be useful for identifying sub-voxel-sized geometry within sampled data, thus extending the resolution.

12.10.2 Mixtures of materials within an object.

Based on the assumptions, voxels only contain mixtures of materials when those mixtures are caused by partial-volume effects. These assumptions are not true in many cases. By relaxing them and then introducing varying concentrations of given materials within an object, one could derive histogram basis functions parameterized by the concentrations and could fit them to measured data. The derivation would be substantially similar to that presented here.

12.10.3 Benefits of vector-valued data.

As with many other techniques, what is described here works on vector-valued volume data, in which each material has a characteristic vector value rather than a characteristic scalar value. Vector-valued datasets have a number of advantages and generally give better classification results. Such datasets have improved SNR and frequently distinguish similar materials more effectively (see Figure 12.13).

12.10.4 Partial mixtures.

Note that the histograms, $h^{\text{vox}}(v)$, for some voxel-sized regions are not ideally matched by a linear sum of basis functions. There are two possible sources of this mismatch.

The first source is the assumption that within a small region there is still normally distributed noise. $\bar{N}$ models the fact that the noise no longer averages to zero, but there is no attempt to model the change in shape of the distribution as the region size shrinks.

The second source is related. A small region may not contain the full range of values that the mixture of materials can produce. The range of values is dependent on the bandwidth of the sampling kernel function. As a result, the histogram over that small region is not modeled ideally by a linear combination of pure material and mixture distributions. Other model histogram basis functions with additional parameters can better match histograms [18], [19]. Modeling the histogram shape as a function of the distance of a voxel from a boundary between materials is likely to address both of these effects and give a result with a physical interpretation that will make geometric model extraction more justifiable and the resulting models more accurate.

These two effects weight the optimization process such that it tends to make $\bar{N}$ much larger

than expected. As a result, experience shows that setting $w(v)$ to approximately 30 times the maximum value in $h^{\text{vox}}(v)$ gives good classification results. Smaller values tend to allow $\bar{N}$ to move too much, and larger values hold it constant. Without these problems $w(v)$ should take on values equal to some small percentage of the maximum of $h^{\text{vox}}(v)$.

12.10.5 Non-uniform spatial intensities.

Spatial intensity in MRI datasets can vary due to inhomogeneities in the RF or gradient fields. It is assumed that they are small enough to be negligible for this algorithm, but it would be possible to incorporate them into the histogram basis functions by making the parameter c vary spatially.

12.10.6 Quantitative comparison with other algorithms

Because of the lack of a "gold standard" against which classification algorithms can be measured, it is difficult to compare the technique described here with others. Each technique presents a set of results from some application area, and so anecdotal comparisons can be made, but quantitative comparisons require reimplementing other algorithms. Work in generating a standard would greatly assist in the search for effective and accurate classification techniques. The voxel histogram technique appears to achieve a given level of accuracy with fewer vector elements than the eigenimages of [12] or the classification results of [14], which use 3-valued data. Their results are visually similar to the voxel histogram results, and underscore the need for quantitative comparison. Because neighboring sample values are interpolated, a given accuracy can be achieved with two-valued or even scalar data, while their technique is likely to require more vector components.

[13] shows good results for a human brain dataset, but their technique may be less robust in the presence of material mixture signatures that overlap, a situation their examples do not include.

12.10.7 Implementation.

The examples were calculated using an implementation in C and C++ on Unix workstations. It uses a sequential-quadratic-programming (SQP) constrained-optimization algorithm [24] to fit h^{vox} for each voxel-sized region, and a quasi-Newton optimization algorithm for fitting h^{all}. The algorithm classifies approximately 10 voxels per second on a single HP9000/730, IBM RS6000/550E, or DEC Alpha AXP 3000 Model 500 workstation.

12.11 Conclusions

The algorithm described in this chapter for classifying scalar- and vector-valued volume data produces more-accurate results than existing techniques in many cases, particularly at boundaries between materials. The improvements arise because: 1) a continuous function is reconstructed from the samples, 2) histograms taken over voxel-sized regions are used to represent the contents of the voxels, 3) sub-voxel partial-volume effects caused by the band-limiting nature of the acquisition process are incorporated into the model, and 4) a Bayesian classification approach is used. The technique correctly classifies many voxels containing multiple materials in the examples of both simulated and real data. It also enables the creation of more-accurate geometric models and images. Because the technique correctly classifies voxels containing multiple materials, it works well on low-resolution data, where such voxels are more prevalent. The examples also illustrate that it works well on noisy data ($\mathrm{SNR} < 15$).

The construction of a continuous function is based on the sampling theorem, and while it does not introduce new information, it provides classification algorithms a richer context for the information. It incorporates neighbor information into the classification process for a voxel in a natural and mathematically rigorous way and thereby greatly increases classification accuracy. In addition, because the operations that can be safely performed directly on sampled data are so limited, treating the data as a continuous function helps to avoid introducing artifacts.

Histograms are a natural choice for representing voxel contents for a number of reasons. First, they generalize single measurements to measurements over a region, allowing classification concepts that apply to single measurements to be generalized. Second, the histograms can be calculated easily. Third, the histograms capture information about neighboring voxels; this increases the information content over single measurements and improves classification results. Fourth, histograms are orientation independent; orientation independence reduces the number of parameters in the classification process hence simplifying and accelerating it.

Partial-volume effects are a nemesis of classification algorithms, which traditionally have drawn from techniques that classify isolated measurements. These techniques do not take into account the related nature of spatially-correlated measurements. Many attempts have been made to model partial-volume effects, and this work continues that trend, with results that suggest that continued study is warranted.

The Bayesian approach described is a useful formalism for capturing the assumptions and information gleaned from the continuous representation of the sample values, the histograms calculated from them, and the partial-volume effects of imaging. Together, these allow a generalization of many sample-based classification techniques.

12.12 Acknowledgments

Many thanks to Matthew Avalos, who has been instrumental in implementation. Thanks to Barbara Meier, David Kirk, John Snyder, Bena Currin, and Mark Montague for reviewing early drafts and making suggestions. The data shown was collected in collaboration with the Huntington Magnetic Resonance Center with the cooperation of Brian Ross and Jose Jimenez, and at the Caltech Biological Imaging Center jointly with Pratik Ghosh and Russell Jacobs.

This work was supported in part by grants from Apple, DEC, Hewlett Packard, and IBM. Additional support was provided by NSF (ASC-89-20219) as part of the NSF/ARPA STC for Computer Graphics and Scientific Visualization, by the DOE (DE-FG03-92ER25134) as part of the Center for Research in Computational Biology, by the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Science Foundation as part of the Human Brain Project, and by the Beckman Institute Foundation. All opinions, findings, conclusions, or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

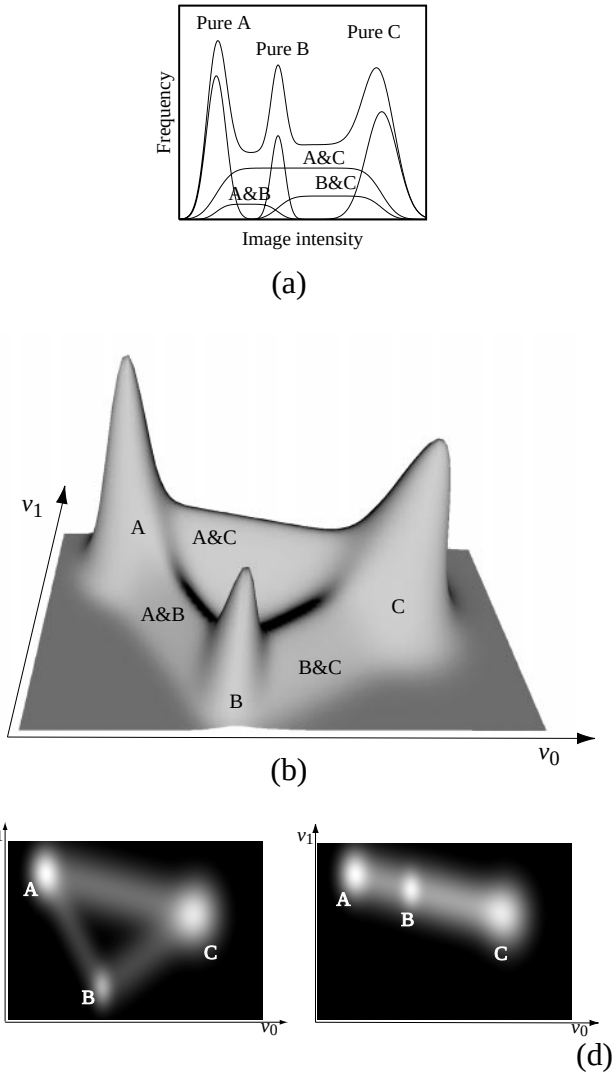


Figure 12.13: Benefits of histograms of vector-valued data. These figures show histograms of an object consisting of three materials. (a) This histogram of scalar data shows that material mean values are collinear. Distinguishing among more than two materials is often ambiguous. (b) and (c) are two representations of histograms of vector-valued data and show that mean values often move away from collinearity in higher dimensions, and so materials are easier to distinguish. High/bright locations indicate more-common (v_0, v_1) data values. While less likely, (d) shows that the collinearity problem can exist with vector-valued data.

Bibliography

[1] Marc Levoy, "Display of surfaces from volume data," *IEEE Computer Graphics and Applications*, vol. 8, no. 3, pp. 29–37, May 1988.

[2] William E. Lorensen and Harvey E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Computer Graphics (SIGGRAPH '87 Proceedings)*, Maureen C. Stone, Ed., July 1987, vol. 21, pp. 163–169.

[3] T. J. Loredo, "From Laplace to supernova SN1987A: Bayesian inference in astrophysics," in *Maximum Entropy and Bayesian Methods*, P. Fougere, Ed. Kluwer Academic Publishers, Denmark, 1989.

[4] Alan V. Oppenheim, Alan S. Willsky, and Ian T. Young, *Signals and Systems*, Prentice-Hall, Inc., New Jersey, 1983.

[5] Michael W. Vannier, Robert L. Butterfield, Douglas Jordan, William A. Murphy, Robert G. Levitt, and Mokhtar Gado, "Multispectral analysis of magnetic resonance images," *Radiology*, vol. 154, pp. 221–224, 1985.

[6] Michael W. Vannier, Christopher M. Speidel, and Douglas L. Rickman, "Magnetic resonance imaging multispectral tissue classification," in *Proc. Neural Information Processing Systems (NIPS)*, Aug. 1988.

[7] Harvey E. Cline, William E. Lorensen, Ron Kikinis, and Ferenc Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity," *Journal of Computer Assisted Tomography*, vol. 14, no. 6, pp. 1037–1045, Nov., Dec. 1990.

[8] Richard P. Duda and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[9] L. P. Clarke, R. P Velthuizen, M. A. Camacho, J. J Neine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger, "MRI segmentation: Methods and applications," *Magnetic Resonance Imaging*, vol. 13, no. 3, pp. 343–368, 1995.

[10] Marc Joliot and Bernard M. Mazoyer, "Three-dimensional segmentation and interpolation of magnetic resonance brain images," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, pp. 269–277, 1993.

[11] Robert A. Drebin, Loren Carpenter, and Pat Hanrahan, "Volume rendering," in *Computer Graphics (SIGGRAPH '88 Proceedings)*, John Dill, Ed., Aug. 1988, vol. 22, pp. 65–74.

[12] Joe P. Windham, Mahmoud A. Abd-Allah, David A. Reimann, Jerry W. Froelich, and Allan M. Haggar, "Eigenimage filtering in MR imaging," *Journal of Computer-Assisted Tomography*, vol. 12, no. 1, pp. 1–9, 1988.

[13] Yi-Hsuan Kao, James A. Sorenson, and Stefan S. Winkler, "MR image segmentation using vector decompsition and probability techniques: A general model and its application to dual-echo images," *Magnetic Resonance in Medicine*, vol. 35, pp. 114–125, 1996.

[14] H. S. Choi, D. R. Haynor, and Y. M. Kim, "Partial volume tissue classification of multichannel magnetic resonance images — a mixel model," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 395–407, 1991.

[15] Derek R. Ney, Elliot K. Fishman, Donna Magid, and Robert A. Drebin, "Volumetric rendering of computed tomography data: Principles and techniques," *IEEE Computer Graphics and Applications*, vol. 10, no. 2, pp. 24–32, Mar. 1990.

[16] Peter Santago and H. Donald Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 566–574, september 1993.

[17] Zhenyu Wu, Hsiao-Wen Chung, and Felix W. Wehrli, "A bayesian approach to subvoxel tissue classification in NMR microscopic images of trabecular bone," *Journal of Computer Assisted Tomography*, vol. 12, no. 1, pp. 1–9, 1988.

[18] David H. Laidlaw, Alan H. Barr, and Russell E. Jacobs, "Goal-directed brain micro-imaging," in *Neuroinformatics: An Overview of the Human Brain Project*, Steven H. Koslow and Michael F. Huerta, Eds., vol. 1, chapter 6. Kluwer, Feb. 1997.

[19] David H. Laidlaw, *Geometric Model Extraction from Magnetic Resonance Volume Data*, Ph.D. thesis, California Institute of Technology, 1995.

[20] Pratik R. Ghosh, David H. Laidlaw, Kurt W. Fleischer, Alan H. Barr, and Russell E. Jacobs, "Pure phase-encoded MRI and classification of solids," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 616–620, 1995.

[21] David H. Laidlaw, "Tissue classification of magnetic resonance volume data," Master's thesis, California Institute of Technology, 1992.

[22] Richard Bartels, John Beatty, and Brian Barsky, *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling*, Morgan Kaufmann Publishers, Palo Alto, CA, 1987.

[23] H. S. Choi, D. R. Haynor, and Y. M. Kim, "Multivariate tissue classification of MRI images for 3-d volume reconstruction – a statistical approach," *SPIE Medical Imaging III: Image Processing*, vol. 1092, pp. 183–193, 1989.

[24] NAG, *NAG Fortran Library*, Numerical Algorithms Group, 1400 Opus Place, Suite 200, Downers Grove, Illinois 60515, 1993.