Classification of Material Mixtures in Volume Data for Visualization and Modeling

David H. Laidlaw, Kurt W. Fleischer, Alan H. Barr
California Institute of Technology, Pasadena, CA 91125

Abstract

Material classification is a key step in creating computer graphics models and images from volume data. We present a new algorithm for identifying the distribution of different material types in volumetric datasets such as those produced with Magnetic Resonance Imaging (MRI) or Computed Tomography (CT). The algorithm assumes that voxels can contain more than one material, e.g. both muscle and fat; we wish to compute the relative proportion of each material in the voxels.

Other classification methods have utilized Gaussian probability density functions to model the distribution of values within a dataset. These Gaussian basis functions work well for voxels with unmixed materials, but do not work well where the materials are mixed together. We extend this approach by deriving non-Gaussian "mixture" basis functions.

We treat a voxel as a volume, not as a single point. We use the distribution of values within each voxel-sized volume to identify materials within the voxel using a probabilistic approach.

The technique reduces the classification artifacts that occur along boundaries between materials. The technique is useful for making higher quality geometric models and renderings from volume data, and has the potential to make more accurate volume measurements. It also classifies noisy, low-resolution data well.

1 Introduction

The main motivation for this work is to make computer graphics models and images using volume measurements of real objects. Identifying different materials within these sampled datasets is an important step in this process (see Figure 1). By finding materials we wish to identify as parts, or to emphasize or deemphasize, we can better control a volume-rendered image [Levoy 88], a surface model [Lorensen and Cline 87], or a volume model created from the data.

Applications of these models and images include surgical planning and assistance, conventional computer animation, anatomical studies, and predictive modeling of complex biological shapes and behavior. Some aspects of our classification technique could also potentially be applied to medical imaging, with appropriate testing. With further development, the ideas may also apply to computer vision problems and to extracting mattes for digital optical special effects

Sources of sampled volume data are becoming more numerous and accessible. They include medical imaging techniques like Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), as well as astrophysical, meteorological and geophysical measurements. The computational sciences frequently produce sampled data as results, e.g. computational fluid dynamics (CFD) and finite element method (FEM) simulations. MRI is of particular interest to us because it is relatively non-invasive, and it measures chemical properties uniformly.

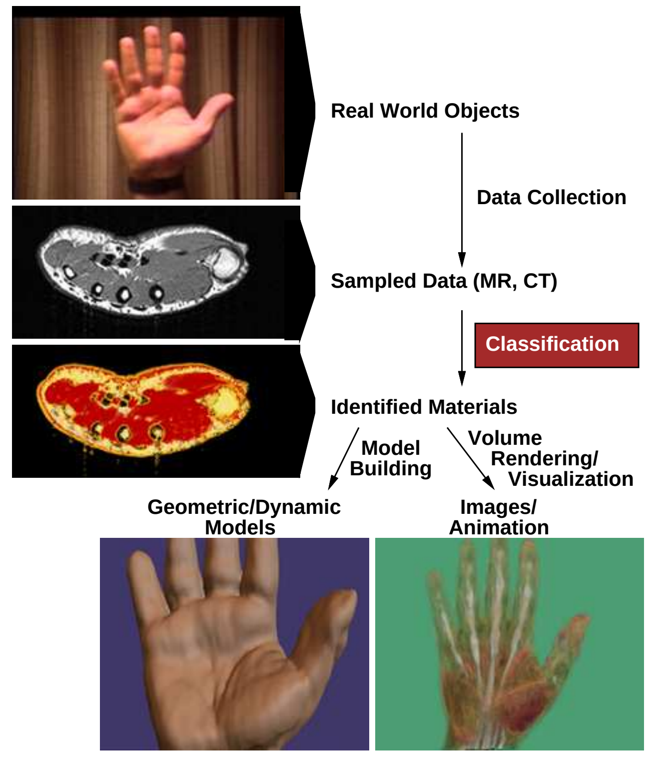
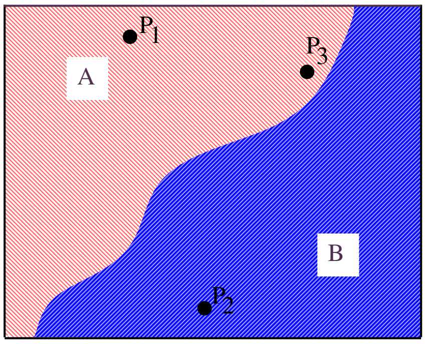


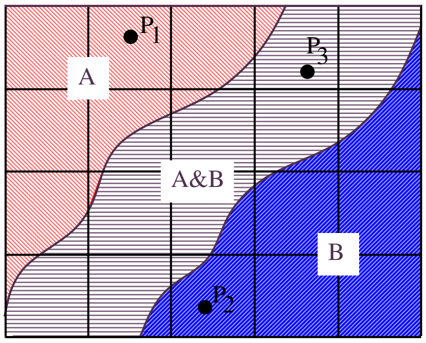
Figure 1: In this paper we focus on the classification step in the process of building models and visualizing sampled volume data. Our technique identifies materials in sampled volume data producing a new sampled volume dataset for each material.

Previous Work. There has been much work done on material classification in sampled datasets [Duda73], but many of the techniques introduce classification artifacts, particularly on boundaries between different materials. These artifacts, which tend to be jaggy-like stair-steps or additional surfaces, are particularly visible in computer graphics images and models.

Discrete statistical classification techniques are often used to identify a single class for each sample within a dataset [Vannier85], [Vannier88]. Each class contains samples representing a particular material. These techniques work well in regions where only one material is present, as in the interiors of single-material regions, but



(i) Real World Object



(ii) Sampled Data

Figure 2: We start from the assumption that in a real-world object each point is exactly one material, as in (i). The measurement process creates samples that mix materials together, from which we reconstruct a continuous, band-limited measurement function $\rho(x)$. 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 (ii) lies in the A&B region where materials A and B are mixed. The grid lines show how the regions may span voxels.

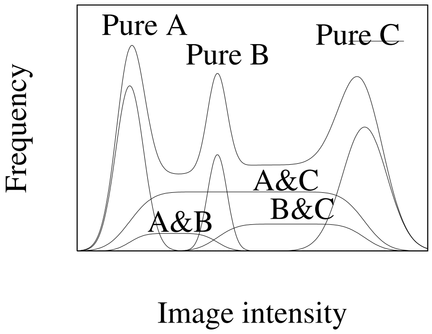
tend to fail near boundaries between these regions, since a given sample does not represent a single material there (see Figure 2).

A sample measures a combination of materials A simplifying assumption of some other techniques is that each sample represents a measurement of one material, rather than a combination of materials. Because the data collection process blends together measurements of more than one material at points near boundaries, this assumption is not always accurate (see Figure 2).

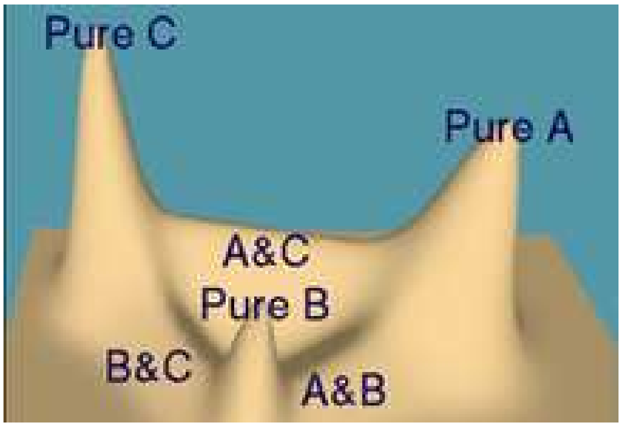
[Drebin88] mentions the need for mixture classification. He approximates the relative volume of materials represented by a sample as the probability that the sample is each material. As he points out, this works reasonably well for differentiating air, soft tissue, and bone in CT data, but not in general. In MR data the expected data value for one material may often be identical to the expected value for a mixture of two other materials. We address this problem further below.

Replacing point samples with histograms. [Choi91] presents a method that models each sample as representing a mixture of materials. This technique, like many others, classifies a region based on a single measurement within the region, effectively treating each voxel as a single point.

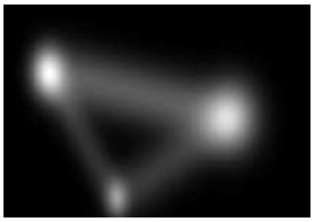
Instead, we treat each voxel as a volume. Ideally, we would like to measure the exact material at each point within the volume. The data collection process, however, discards high-frequency information in producing samples, and so we can only reconstruct a band-limited function $\rho(x)$ defined over the volume, from the samples. With the distribution of values from this function over each voxel-sized volume, we identify materials within the voxel prob-



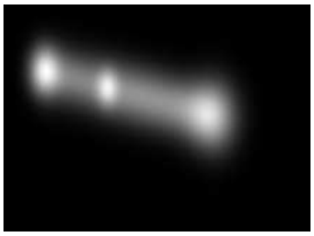
(i)



(ii)



(iii)



(iv)

Figure 3: Benefits of histograms of vector-valued data. We show histograms of an object with three materials. (i) is a histogram of scalar data and shows that material mean values are collinear; therefore, distinguishing among more than two materials is often ambiguous. (ii) represent a histogram of vector-valued data and shows that mean values often move away from collinearity in higher dimensions. (iii) is another representation of the same histogram. (iv) shows that the collinear problem can exist with vector-valued data.

abilistically. By using the reconstructed continuous measurement function $\rho(x)$ and not just a single measurement, we incorporate more information into the classification process and increase its accuracy.

Using vector-valued data. As with many other techniques, ours 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. First, they have improved signal-to-noise. Second, they frequently distinguish similar materials more effectively (see Figure 3).

In particular, the jump from scalar to 2-vector-valued data is very important. In scalar-valued datasets it is difficult to distinguish a mixture of two pure materials with values v_A and v_B from a pure material with some intermediate value such as $v_C = (v_A + v_B)/2$. This is because the three material values are collinear, as they must be for such a dataset.

With more measurement dimensions in the dataset, collinearity is less frequent for most combinations of materials, although Figure 3 illustrates that it can still occur. When it does occur, classification works as for scalar-valued data.

Our approach. In this paper we build on a statistical framework, using Bayesian probability theory and approximations of conditional and prior probabilities [Loredo] to calculate the probability of a particular mixture of materials given the histogram of a voxel-sized region. We then find the most likely mixture for the region.

In calculating histograms, we reconstruct the continuous function, band-limited to the Nyquist frequency, from the samples. In some cases this requires processing the samples to reduce aliasing artifacts, particularly for MRI data collected as slices.

We assume, as in Figure 2, that each voxel is a mixture of materials, with mixtures like A and B occurring where the band-limiting process blurs pure materials together. From this assumption we derive basis functions that model histograms for pure materials and for mixtures of two materials.

2 Overview

In this section we describe the classification problem that we solve, define terms, state assumptions we make about the data we classify, sketch the algorithm and its derivation, and give a roadmap for the remainder of the paper.

The input to our process is sampled measurement data, from which we reconstruct a continuous, band-limited function $\rho(x)$ that measures distinguishing properties of the underlying materials. The output is sampled data measuring the relative volume of each material. We call the output "material volume ratio densities" (see Figure 4).

Definitions. We refer to the coordinate system of the space of the object we are measuring as "spatial coordinates," and generally use $x \in X$ to refer to points. This space is n_x-dimensional, where n_x is 3 for volume data, but can be 2 for slices. Each measurement, which may be a scalar or vector, lies in "feature space," (see Figure 5) 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 2-valued vector data, etc.

The table in Appendix B may be useful for checking other definitions.

Assumptions. We make the following assumptions about the function $\rho(x) : \mathbf{R}^{n_x} \rightarrow \mathbf{R}^{n_v}$.

1. $\rho(x)$ produces the same value (modulo noise) for the same material wherever it may lie in the original object.
2. $\rho(x)$ is band-limited to the Nyquist frequency before sampling. If the band-limiting frequency is too high, then aliasing artifacts occur in the continuous function reconstructed from the samples. If the frequency is too low, then samples contain redundant information, voxel-sized regions will not contain as much information, and the classification process will degrade.
3. The noise in $\rho(x)$ is additive and normally distributed.
4. Different scalar-valued datasets that we treat as vector-valued must be aligned with one another so that a given measurement $\rho(x)$ returns a vector of values all representing the same location in the object.

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

Sketch of Derivation. As shown in Figure 2 we start with the assumption that each spatial location in the real world object is exactly one material, and that the measurement process mixes materials together as it band-limits the measurements to the Nyquist frequency of the sampling rate. From that assumption we will derive (in Section 3) an equation for a normalized histogram of data values within a region. This histogram function is a probability density function (PDF) that tells us the probability that a measurement will lie within a range of values in that region.

In Section 4 we create basis functions to model histograms. These basis functions are parameterized probability density functions for regions consisting of single materials and for regions consisting of mixtures of two materials. These mixtures are assumed to have been created by the band-limiting process accompanying sampling. The parameters represent the mean value c and variance s of a measurement.

Using Bayes' Theorem, the histogram of the entire dataset, our model basis functions, and a series of approximations, we derive an estimate of the most likely set of materials within an entire dataset (Section 5). Similarly, given the histogram of a voxel-sized region, we derive an estimate of the most likely density for each material in that region (Section 6).

We present results in Section 7, discussion in Section 8, and conclude in Section 9.

Sketch of Algorithm. The algorithm produces, as its end result, a sampled dataset containing estimates of material volume ratio densities. The process is illustrated in Figure 4. First, we collect and preprocess data to satisfy the assumptions listed above. Second, we calculate a histogram of the entire dataset, and fit parameterized material probability density functions to the histogram.

Using the fitted parameters, we process each voxel-sized region in the dataset as follows. We first calculate a histogram for the small region and find the combination of materials most closely fitting the histogram. Using the estimated parameters, we calculate material volume ratio densities for that small region.

3 Normalized Histograms

In this section we present the equation for a normalized histogram of a sampled dataset over a region. We will use this equation as a building block in several later sections, with regions that vary from the size of a single voxel to regions covering the entire dataset. We will also use this equation to derive basis functions that model

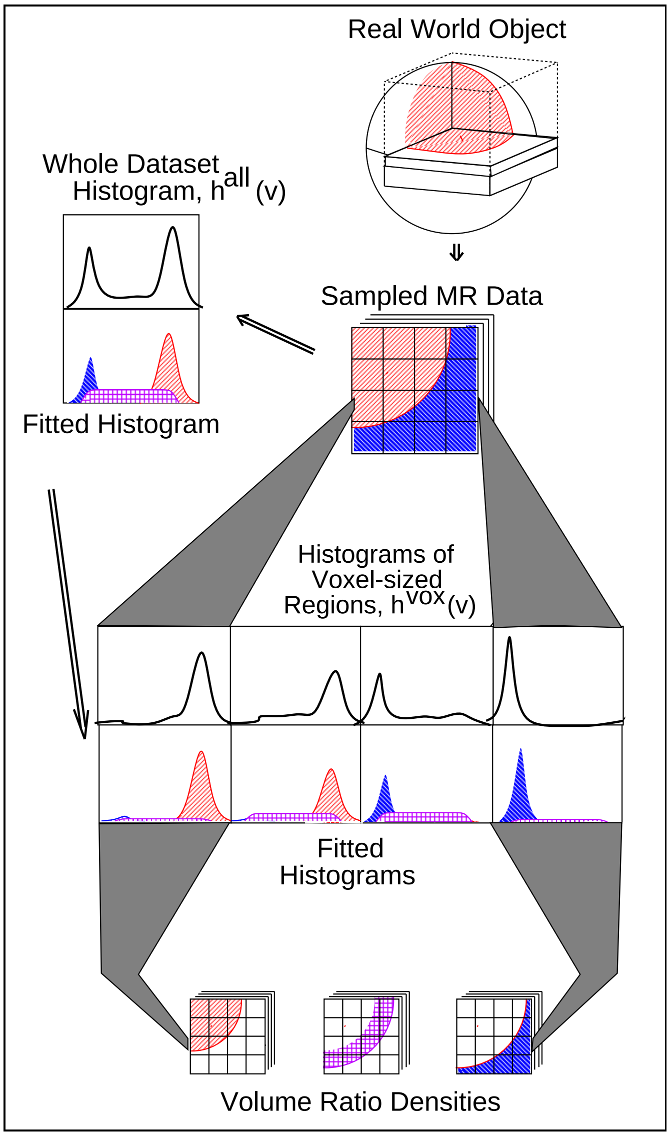


Figure 4: Steps in the classification process. We collect MR data, calculate a histogram of the entire dataset, $h^{\text{all}}(v)$, and use that to determine parameters of histogram-fitting basis functions. We then calculate histograms of each voxel-sized region, $h^{\text{vox}}(v)$, and identify the most likely mixture of materials for that region. The result is a sampled dataset of volume ratio densities.

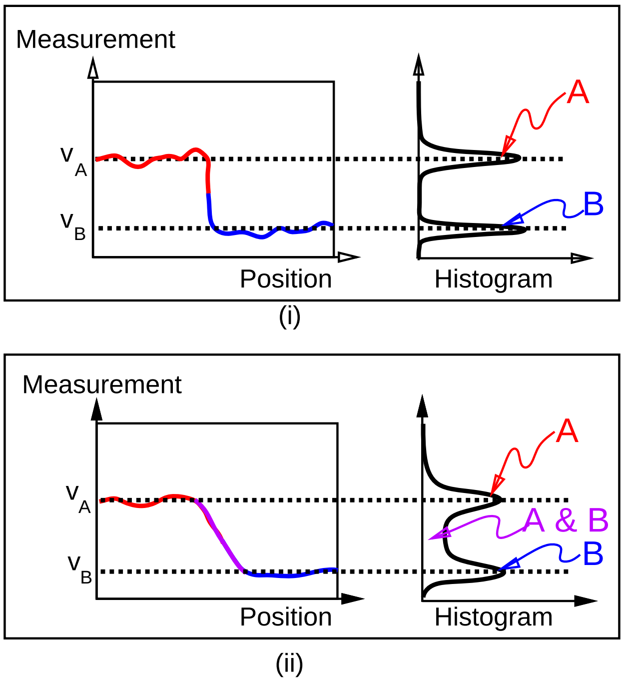


Figure 5: Noise and mixtures in histograms. The scalar data on the left represents measurements from a dataset containing two materials, A and B, such as that shown in Figure 2. One material has measurement values near v_A and the other near v_B. These values correspond to the gaussian-shaped peaks centered around v_A and v_B in the histograms, which are shown on their sides to emphasize the axis that they share. This axis is "feature space." In (i) we show a histogram of a function that has not been band-limited, but does have noise. In (ii) the function has been band-limited, and the measurement transition between v_A to v_B now appears in the histogram as the flat region between feature space values v_A and v_B.

histograms over regions containing single materials and regions containing mixtures of materials. Figure 5 shows an example of calculating a continuous histogram.

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 $\rho(x) = v$. We define histograms, $h^{\mathcal{R}}(v) : \mathbf{R}^{nv} \rightarrow \mathbf{R}$, as probability density functions (PDFs). This definition differs from others in that it represents a dataset as a continuous function $\rho(x)$, rather than by a finite set of samples. Thus the histograms over regions are also continuous functions:

$$h^{\mathcal{R}}(v) = \int \mathcal{R}(x)\delta(\rho(x) - v)dx \qquad (1)$$

This equation is the continuous analogue of a discrete histogram. $\mathcal{R}(x)$ is non-zero within the region of interest, and integrates to 1. We define $\mathcal{R}(x)$ to be constant in the region of interest making every spatial point contribute equally to the histogram $h^{\mathcal{R}}(v)$. Note also that $h^{\mathcal{R}}(v)$ integrates to 1, which is important for our interpretation as a PDF. $\delta()$ is the Dirac-delta function.

We use this equation both as a starting point for deriving material intensity PDFs, and also as a basis for calculating histograms of regions of our datasets. The derivations are shown in the following two sections. We will now discuss a few implementation considerations for calculating histograms.

Implementation Considerations. We calculate histograms in 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 we do not lose significant features in the histogram.

We calculate a histogram iteratively, first initializing the bins to zero. For each small region of the dataset we use the first terms of the Taylor series of $\rho(x)$ to create a linear approximation over that region. We then calculate a piecewise constant approximation of the histogram over that region, and add that to the bins. The histogram approximation is obtained by substituting the linearized version of $\rho(x)$ into Equation 1 and integrating that over the small regions. This takes into account correlation of values between different elements of vector data.

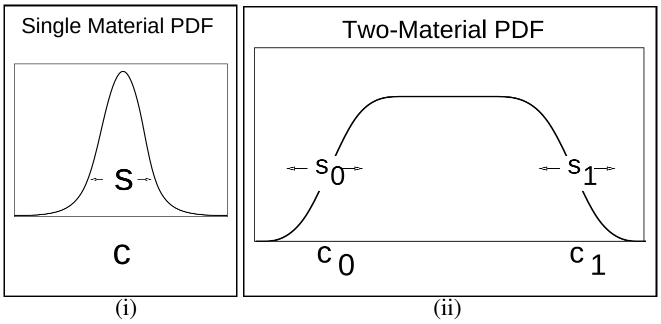


Figure 6: Parameters for a single material PDF, shown in (i) include c, the mean value for the material, and s, which measures the variance of the noise (see Equation 2). (ii) shows corresponding parameters for a two-material mixture basis function. s_0 and s_1 affect the slopes of the two-material PDF at either end. For vector-valued data c and s are vectors and are the mean values and variances of the noise for the two constituent materials (see Equation 3).

4 Histogram Basis Functions for Pure Materials and Mixtures

In this section we present definitions of basis functions that model histograms of pure materials and of mixtures of materials. These basis functions are PDFs that give the probability that a sample lies within a range of values given that it is a particular material or mixture. The parameters of the basis functions specify the expected value, c, and variance, s, of each material's measurements (see Figure 6).

We use Equation 1 to derive these basis functions, which we fit to the data. We then verify that the equations provide reasonable fits to typical MR data, which gives us confidence that our assumptions about the measurement function $\rho(x)$ were reasonable. The details of the derivations are in Appendix A.

For a single material, the PDF is a gaussian distribution:

$$f_{\text{single}}(v; c, s) = \pi^{-\frac{n_v}{2}} \exp(-((x-c)\cdot\gamma)^2), \tag{2}$$

Where $\gamma_i = 1/s_i$ for $i \in n_v$. We derive this equation by manipulating Equation 1 evaluated over a region of constant material, where the measurement function $\rho(x)$ is a constant value plus additive, normally distributed noise.

For mixtures along a boundary between two materials, we derive another equation similarly.

$$f_{\text{double}}(v; c, s) = \int_0^1 k_n((1-t)c_1 + tc_2 - v; s)dt \tag{3}$$

As with the single material, this derivation follows from Equation 1 evaluated over a region where two materials mix. In this case, we approximate the band-limiting filter of the data collection process with a box filter, and make the assumption that the variance of the additive noise is constant across the region. This basis function is a superposition of Gaussian distributions representing different amounts of the two constituent mixtures. k_n is the Gaussian distribution, t the relative quantity of the second material, c_1 and c_2 the expected values of the two materials, and s the variance of measurements.

The assumption of a box filter affects the shape of the resulting PDF. We derived similar equations for different filters (triangle, Gaussian, and Hamming), but chose the box filter derivation because we found it sufficiently accurate in practice and because the numerical tractability of the PDF in this case saved significant computation.

5 Estimating Material Intensity PDF Parameters

In this section we describe parameter estimation procedure for fitting material intensity PDFs to a dataset. For a given dataset we first calculate the histogram, $h^{\text{all}}(v)$, of the entire dataset.

We then combine an interactive process of specifying the number of materials and approximate feature-space locations for them with an automated optimization to estimate the parameters. 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 PDFs that describe the various materials and mixtures of interest in the dataset.

The optimization process estimates the relative volume of each material (vector α^{all}), the mean value (vector c) and variance (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. We define n_m as the number of pure materials in a dataset, and n_f as the number of material intensity PDFs. $n_f \geq n_m$, since n_f includes any material intensity PDFs for mixtures, as well as those for pure materials.

The optimization minimizes the function

$$\mathcal{E}(\alpha^{\text{all}}, c, s) = \int \left(\frac{q(v; \alpha^{\text{all}}, c, s)}{w(v)}\right)^2 dv \tag{4}$$

where:

$$q(v; \alpha^{\text{all}}, c, s) = h^{\text{all}}(v) - \sum_{j=1}^{n_f} \alpha_j^{\text{all}} f_j(v; c_j, s_j) \tag{5}$$

The function $w(v)$ is analogous to a variance at each point v in feature space, and gives the expected value of $|q(v)|$. We approximate $w(v)$ as a constant, and discuss it further in Section 8.

This equation is derived in Appendix B, using Bayesian probability theory with estimates of prior and conditional probabilities.

6 Classification

In this section we describe the process of classifying each voxel. This process is similar to that described in Section 5 for fitting the material PDFs to the entire dataset, but now we are doing it over small, voxel-sized regions. We use the previously computed material PDFs as fixed basis functions and no longer vary the mean vector c and variance s. The only parameters allowed to vary are the relative material volumes (vector α_j^{vox}), and an estimate of the local noise in the local region (vector $\bar{N}$) (see Equations 6 and 7).

Over large regions the noise is normally distributed, with zero mean. However for small regions the mean noise is generally non-zero due to the band-limiting introduced in the data collection process. We label this local mean voxel noise value $\bar{N}$. As derived in Appendix B the equation that we minimize is:

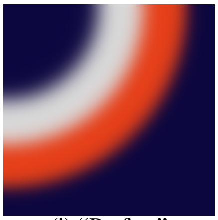
$$\mathcal{E}(\alpha^{\text{vox}}, \bar{N}) = \sum_{i=1}^{n_v} \left(\frac{\bar{N}_i}{\sigma_i}\right)^2 + \int \left(\frac{q(v; \alpha^{\text{vox}}, \bar{N})}{w(v)}\right)^2 dv \tag{6}$$

where

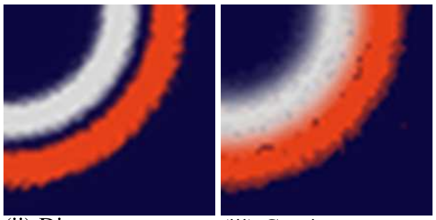
$$q(v; \alpha^{\text{vox}}, \bar{N}) = h^{\text{vox}}(v - \bar{N}) - \sum_{j=1}^{n_f} \alpha_j^{\text{vox}} f_j(v) \tag{7}$$

and subject to the constraints

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



(i) "Perfect" Classification



(ii) Discrete Classification (no mixtures)

(iii) Continuous, Multi-material Classification

Figure 7: Comparison of discrete, single-material classification (ii), and the new classification (iii). (i) is a reference for what "ideal" classification should produce. Note the band of background material in (ii) 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, 2-valued data of two concentric shells.

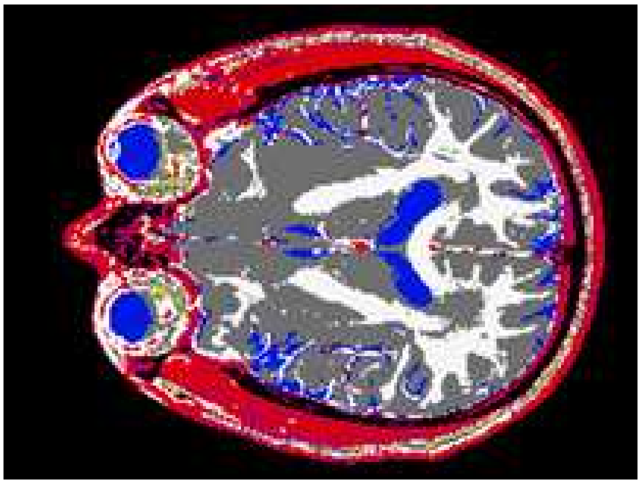


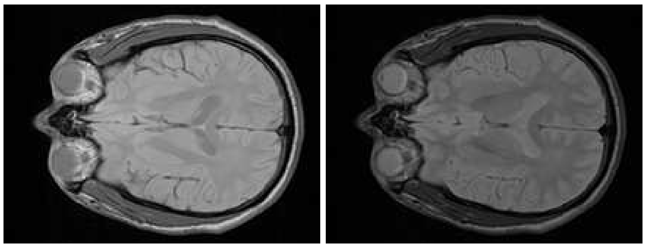
Figure 8: Discrete, single-material classification of the same slice shown in Figure 9.

Vector σ is the expected variance of the noise over the entire dataset. We estimate this as an average of the variances of the material intensity PDFs.

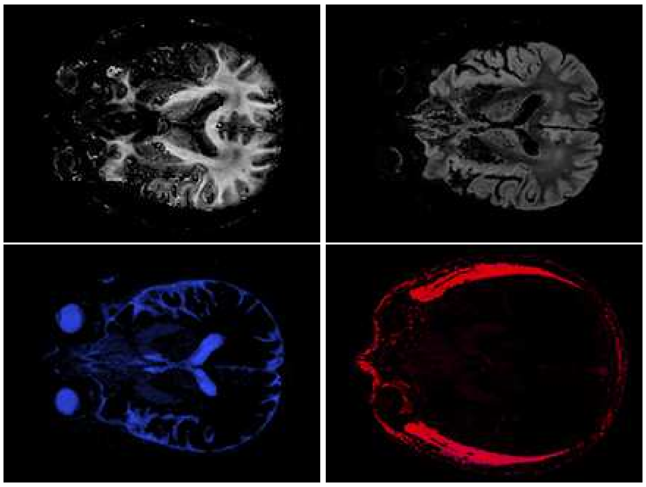
With vector α^{vox} for a given voxel-sized region and the mean value, vector $\bar{v}$, within that region, we solve for the amount of each pure material contributed by each mixture to the voxel. This is our output, the estimates of the amount of each pure material in the voxel-sized region.

7 Results

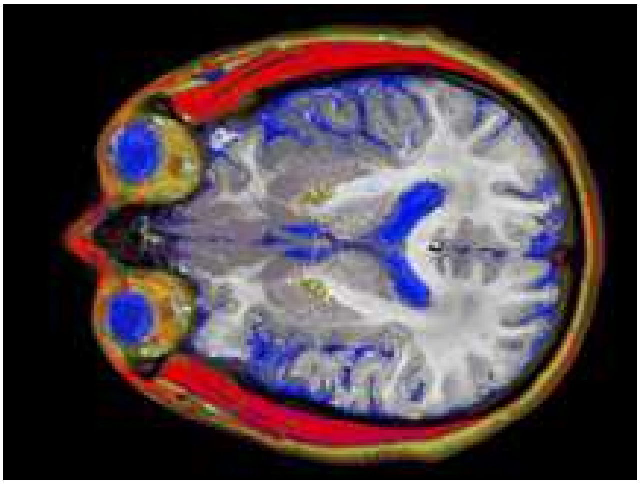
We have applied our new technique to several datasets. The following table lists the datasets, the MRI machine they were collected



(i) Original Data



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



(iii) Combined Classified Image

Figure 9: One slice of data from a human brain. (i) shows the original two-valued data, (ii) shows four of the identified materials, white matter, gray matter, cerebro-spinal fluid and muscle, separated out into separate images, and (iii) shows the 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 8.

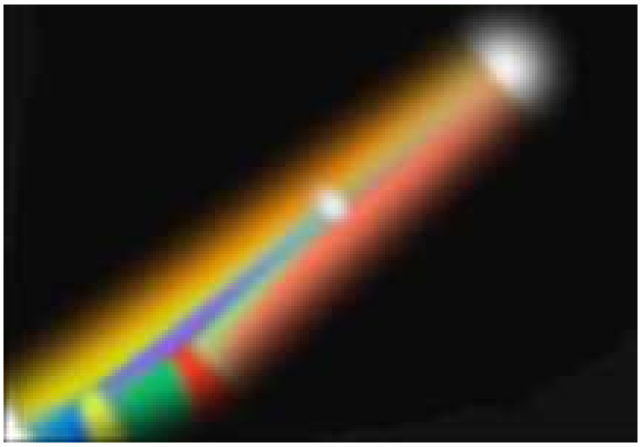


Figure 10: Basis functions fit to histogram. This figure illustrates the results of fitting basis functions to the histogram of the hand dataset. Bright dots are pure materials, while the lines connecting the dots are mixtures. The rightmost two white dots are pure fat and bone marrow in the hand. The lower yellow and red dot are pure skin and muscle, respectively. The mixture between muscle (red) and fat (white) is a salmon colored streak. The green streak between the red and yellow dots is a mixture of skin and muscle. These fitted basis fuctions were used to produce the classified data used in Figures 11 and 1.

Figure 11: 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.

on, the protocol used (with some collection parameters), the voxel size, and the figures in which each dataset appears. All datasets were collected with a spin-echo or fast spin-echo protocol, with one proton-weighted and one T_2-weighted acquisition.

Object	Machine	Voxel Size	Figs.
shells	simulated	1x1x10 mm	7
hand	GE 1.5T	0.7x0.7x3 mm	1, 11
brain	GE 1.5T	0.8x0.8x3 mm	8, 9

In figures 7, 8, and 9 we compare our technique with a probabilistic approach that uses pure materials only, and only a single measurement value per voxel. The new technique produces many fewer misclassified voxels, particularly in regions where materials are mixed due to filtering. In Figure 7(ii) and (iii) the difference is particularly noticeable where an incorrect layer of background material has been introduced between the white and red regions, where multiple materials are present in the same voxel. Figures 8 and 9(iii) also show comparative results between the two methods.

Models and volume rendered images, as in Figures 1 and 11 also benefit because less incorrect information is introduced into the classified datasets, and so the images and models more accurately depict the objects they are representing.

Implementation. Our implementation is written in C and C++ on Unix workstations. We use a sequential quadratic programming constrained optimization algorithm [NAG] 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. We have implemented this algorithm in parallel on these machines, and get a corresponding speedup on multiple machines.

8 Discussion

We have made several assumptions and approximations while developing and implementing this algorithm. This section will discuss some of the tradeoffs and suggest some possible directions to continue work.

Mixtures of Three or More Materials. We assume that each measurement contains values from at most two materials, although our approach easily extends to mixtures with more materials. We chose two-material mixtures because surfaces between boundaries of pure materials are one of the most important parts of computer graphics models. Voxels containing 3-material mixtures happen near lines where 3 materials meet, and are generally much less common.

Partial Mixtures. We note that the histograms $h^{\text{vox}}(v)$ for some voxel-sized regions are not ideally matched by a linear sum of basis functions. We address two problems here.

The first problem is that within a small region the assumption that we still have normally distributed noise is no longer valid. $\bar{N}$ models the fact that the noise no longer averages to zero, but we do not attempt to model the change in shape of the distribution as the region size shrinks.

The second problem is related. A small region may not contain the full range of values that the mixture of materials can produce. As a result, the histogram over that small region is not modeled ideally by a linear combination of pure material and mixture distributions. We are investigating additional parameters to reduce the range of the mixture intensity PDFs. This is an area for future research.

We postulate that these two effects weight the optimization process such that it tends to make $\bar{N}$ much larger than we expect. As a result, we have found 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 we would expect $w(v)$ to take on values equal to some small percentage of the maximum of $h^{\text{vox}}(v)$.

9 Conclusions

We have presented a new algorithm for classifying scalar and vector-valued volume data. Our algorithm uses a continuous reconstruction of the dataset, and we present a new histogram of this continuous data over regions. We derive intensity PDF basis functions for both pure materials and mixtures of materials due to the band-limiting portion of the data collection process. Our classification process uses a probabilistic approach to model histograms of voxel-sized regions with the material and mixture basis functions.

We demonstrated our technique on simulated and real data, and it correctly classifies many voxels containing multiple materials.

10 Acknowledgements

Thanks to Barbara Meier, David Kirk, John Snyder, Bena Currin and Mark Montague for reviewing early drafts and making suggestions. Thanks also to Allen Corcoran, Constance Villani, Cindy Ball and Eric Winfree for production help, and Jose Jimenez for the late-night MR sessions. Our data was collected in collaboration with the Huntington Magnetic Resonance Center in Pasadena.

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 and the National Institute of Mental Health 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 author(s) and do not necessarily reflect the views of the sponsoring agencies.

References

[Cline90] Cline, Harvey E., 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*, 14(6), November/December 1990, 1037-1045.

[Drebin88] Drebin, Robert, Loren Carpenter, and Pat Hanrahan, "Volume Rendering," *Computer Graphics* (Proc. SIGGRAPH),vol. 22, 1988, 65-74.

[Duda73] Duda, Richard P., and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[Loredo] Loredo, T.J., "From Laplace to Supernova SN1987A: Bayesian Inference in Astrophysics," in **Maximum Entropy and Bayesian Methods**, Dartmouth College 1989, ed: P. Fougere, Kluwer Academic Publishers, Denmark.

[ThesisXX] Author's MS thesis.

[Choi91] Choi, H. S, Haynor, D. R., and Kim, Y. M., "Partial Volume Tissue Classification of Multichannel Magnetic Resonance Images - A Mixel Model," *IEEE Transactions on Medical Images*, 10(3), 1991, 395-407.

[Keller 88] Keller, Paul, "Basic Principles of Magnetic Resonance Imaging," GE Medical Systems Report, Milwaukee, 1988.

[Levoy 88] Levoy, Marc, " Display of Surfaces from Volume Data," *IEEE Computer Graphics and Applications,* vol. 8(3), pp. 29-37, May, 1988.

[Lorensen and Cline 87] Lorensen, William, and Harvey Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," Computer Graphics (Proc. SIGGRAPH), vol. 21, pp. 163-169, 1987.

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

[Vannier85] Vannier, Michael W., Robert L. Butterfield, Douglas Jordan, William A. Murphy, Robert G. Levitt, Mokhtar Gado, "Multispectral Analysis of Magnetic Resonance Images," *Radiology*, 154, 1985, 221-224.

[Vannier88] Vannier, Michael W. Christopher M. Speidel, and Douglas L. Rickman, "Magnetic Resonance Imaging Multispectral Tissue Classification," Neural Information Processing Systems (NIPS), August 1988.

Appendices

A Derivation of Material PDFs

In this appendix we derive material PDFs which we use as basis functions (f_i) for fitting histograms. We derive two forms of basis functions: one for single, pure materials and another for two-material mixtures (which arise due to sampling). Here is Equation 1, the histogram equation:

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

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). Thus

$$h^{\mathcal{R}}(v) \quad = \quad k_n(v;s) * \int \mathcal{R}(x)\delta((\rho(x) - n(x;s)) - v)dx \tag{8}$$

A.1 Pure Materials

For a single pure material we assume that the measurement function has the form:

$$\rho_{\text{single}}(x) = c + n(x;s), \tag{9}$$

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

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

$$\begin{aligned} f_{\text{single}}(v;c,s) &= \int \mathcal{R}(x)\delta(\rho_{\text{single}}(x) - v)dx \\ &= \int \mathcal{R}(x)\delta(c + n(x;s) - v)dx \\ &= k_n(v;s) * \int \mathcal{R}(x)\delta(c - v)dx \\ &= \pi^{-\frac{n_v}{2}} \exp(-((c-v)\cdot\gamma)^2), \end{aligned} \tag{10}$$

Where $\gamma_i = 1/s_i$ for $i \in n_v$. Thus $f_{\text{single}}(v;c,s)$ is a gaussian distribution with mean c and variance s. We assume the noise is independent in each element of vector-valued data, which for MRI appears to be reasonable.

A.2 Mixtures

For a mixture of two pure materials, we assume the measurement function has the form:

$$\rho_{\text{double}}(x) = \ell_{\text{double}}(x; c_1, c_2) + n(x; s) \tag{11}$$

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 to c_2, the mean values for the two materials.

$$\begin{aligned} f_{\text{double}}(v; c, s) &= \int \mathcal{R}(x)\delta(\rho_{\text{double}}(x) - v)dx \\ &= \int \mathcal{R}(x)\delta(\ell_{\text{double}}(x; c_1, c_2) + n(x; s) - v)dx \\ &= k_n(v; s) * \int \mathcal{R}(x)\delta(\ell_{\text{double}}(x; c_1, c_2) - v)dx \\ &= \int_0^1 k_n(v; s) * \delta((1-t)c_1 + tc_2 - v)dt \\ &= \int_0^1 k_n((1-t)c_1 + tc_2 - v; s)dt \end{aligned} \tag{12}$$

B Derivation of Classification Parameter Estimation

In this appendix we derive the equations that we optimize to find material PDF parameters and to classify voxel-sized regions. We use Bayesian probability theory[Loredo] to derive an expression for the probability that a given histogram was produced by a particular set of parameter values in our model. We maximize an approximation to this "posterior probability" to estimate the best fit parameters.

$$\text{maximize } P(\text{ parameters } \mid \text{histogram }) \tag{13}$$

We use this optimization procedure for two purposes:

- **Find material PDF parameters**. Initially, we find parameters of basis functions to fit histograms of the entire dataset h^{all}. This gives us a set of basis functions which describe the pure materials and mixtures.
- **Classify voxel-sized regions**. We fit a weighted sum of the basis functions to the histogram of a voxel-sized region h^{vox}. This gives us our classification (in terms of the weights α).

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

B.1 Definitions

Term	Dimen-sionality	Definition
n_f	scalar	number of materials and mixtures
n_v	scalar	dimensions 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	$n_f \times n_v$	variance of material measurements (chosen by procedure discussed in Section 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

Probabilities (using Bayesian terminology [Loredo]):

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

B.2 Optimization

We perform the following optimization to find the best fit parameters:

$$\text{maximize } P(\alpha, c, s, \bar{N}|h) \tag{14}$$

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

B.3 Derivation of the posterior probability $P(\alpha, c, s, \bar{N}|h)$

We start with Bayes' Theorem, expressing the posterior probability in term of the likelihood, the prior probability, and the global likelihood.

$$P(\alpha, c, s, \bar{N}|h) = \frac{P(\alpha, c, s, \bar{N})P(h|\alpha, c, s, \bar{N})}{P(h)} \tag{15}$$

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

Prior Probabilities. We assume that α, c, s and $\bar{N}$ are independent, so

$$P(\alpha, c, s, \bar{N}) = P(\alpha)P(c, s)P(\bar{N}) \tag{16}$$

Because the elements of α represent relative volumes, we require that they sum to 1 and are positive.

$$P(\alpha) = \begin{cases} 0 & \text{if } \sum_{j=1}^{n_f} \alpha_j \neq 1 \\ 0 & \text{if } \alpha_j < 0 \text{ or } \alpha_j > 1 \\ p_1 & \text{(constant) otherwise} \end{cases} \tag{17}$$

We use a different assumption for $P(c, s)$ depending on which fit we are doing (h^{all} or h^{vox}). For fitting $h^{\text{all}}(v)$, we consider all values of c, s equally likely

$$P^{\text{all}}(c, s) = p_6 \tag{18}$$

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

$$P^{\text{vox}}(c, s) = \delta(c - c^0, s - s^0) \tag{19}$$

For a small region, we assume that the noise vector, $\bar{N}$, has normal distribution with variance σ.

$$P^{\text{vox}}(\bar{N}) = p_2 e^{-\sum_{i=1}^{n_v}(\frac{\bar{N}_i}{\sigma_i})^2} \tag{20}$$

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

Likelihood. We approximate the likelihood $P(h|\alpha, c, s, \bar{N})$ by analogy to a discrete normal distribution. We define $q(v)$ to measure the difference between the "expected," or "mean" histogram for particular $\alpha, c, s, \bar{N}$ and a given histogram $h(v)$

$$q(v; \alpha, c, s, \bar{N}) = h(v - \bar{N}) - \sum_{j=1}^{n_f} \alpha_j f_j(v; c, s) \tag{21}$$

Now we create a normal-distribution-like function. $w(v)$ is analogous to the variance of q at each point of feature space.

$$P(h|\alpha, c, s, \bar{N}) = p_3 e^{-\int (\frac{q(v;\alpha,c,s,\bar{N})}{w(v)})^2 dv} \tag{22}$$

Global Likelihood. Note that the denominator of Equation 15 is constant (it is essentially a normalization of the numerator):

$$\begin{aligned} P(h) &= \int P(\hat{\alpha}, \hat{c}, \hat{s}, \hat{N}) P(h|\hat{\alpha}, \hat{c}, \hat{s}, \hat{N}) d\hat{\alpha} d\hat{c} d\hat{s} d\hat{N} & (23) \\ &= p_4 & (24) \end{aligned}$$

B.3.1 Assembly

Using the approximations discussed above, we arrive at the following expression for the posterior probability:

$$P(\alpha, c, s, \bar{N}|h) = p_5 P(\alpha) P(c, s) e^{-\sum_{i=1}^{n_v} (\frac{\bar{N}_i}{\sigma_i})^2} e^{-\int (\frac{q(v;\alpha,c,s,\bar{N})}{w(v)})^2 dv} \quad (25)$$

For fitting h^{all}, the mean noise is assumed to be zero, so maximizing equation 25 is equivalent to minimizing $\mathcal{E}^{\text{all}}$ to find the free parameters $(\alpha^{\text{all}}, c, s)$:

$$\mathcal{E}^{\text{all}}(\alpha^{\text{all}}, c, s) = \int \left(\frac{q(v; \alpha^{\text{all}}, c, s)}{w(v)} \right)^2 dv \quad (26)$$

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

For fitting h^{vox}, the parameters c and s are fixed, so maximizing equation 25 is equivalent to minimizing $\mathcal{E}^{\text{vox}}$ to find the free parameters $(\alpha^{\text{vox}}, \bar{N})$:

$$\mathcal{E}^{\text{vox}}(\alpha^{\text{vox}}, \bar{N}) = \sum_{i=1}^{n_v} \left(\frac{\bar{N}_i}{\sigma_i} \right)^2 + \int \left(\frac{q(v; \alpha^{\text{vox}}, \bar{N})}{w(v)} \right)^2 dv \quad (27)$$

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

As stated in Equation 6, Section 6, Equation 27 is minimized to estimate volume ratio densities α^{vox} and the mean noise vector $\bar{N}$.