Tina Memo No. 2000-005
Internal Report, used for MSc teaching. Used as the basis for; Voxel based Analysis of Tissue Volume from MRI Data, British Journal of Radiology, 77, 114-125, 2004.

Tutorial: Structural MRI Analysis Using Volumetric Voxel Analysis.

N.A.Thacker and M. Pokrić.

Last updated
7 / 5 / 2000

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Introduction

When attempting to segment data in an image, though we may be unaware of it, we are really asking a question which has no perfect answer. In common with many other data analysis tasks the best thing that we can determine is only the probability of a particular interpretation. It is a simple fact that in order for a program to segment a data set well it must use statistical principles.

Many segmentation methods work by applying calculations which at first sight appear to have nothing in common with either probability theory or statistics, but if we wish to explain the best approaches to MR image segmentation we must be able to relate these techniques to the statistical models that they are based upon. The most direct piece of information which we can obtain from data is in the form of a conditional probability. $P(C|D)$ is the probability of the interpretation C given the data D. Given such probabilities for each pixel in an image we can segment regions or locate the boundaries between tissues. If the method (or algorithm) used to determine these probabilities is appropriate, then the regions and boundaries determined in this manner will be optimal. That is, they will (by definition) have extracted all of the useful information relating to the problem from the data. Determining that an algorithm is appropriate amounts to being able to confirm that the assumptions underlying the statistical approach are valid. To do this we must first know what these assumptions are. Areas of algorithmic research which give rise to image processing algorithms are fundamentally linked to matching assumptions to data sets. Though it is possible to develop good algorithms blindly (by guesswork and testing) it is always better to apply a statistical methodology, systematically testing the effects of any assumption on the result.

The consequence of all of this is that there is no technique that can be guaranteed to work on any data set, that is a "magic bullet". For any method to work it must be applied to data which falls within the range of behaviour for which it has been designed. Different algorithms have varying ranges of applicability. Algorithms which make the most assumptions regarding the data often have very limited use in comparison to those which take into account a broader range of data characteristics. Though algorithms which make a large number of assumptions can often be simple it is not necessarily true that complex algorithms will always perform better. The extra complexity must be used wisely and for good reason. Extra complexity can just as easily result in unreliability as in improved results. These are the issues that algorithm designers consider when developing a new technique.

The following sections explain the various approaches to tissue segmentation algorithms which deliver either boundaries or labelled regions. We start with the simplest and the assumptions upon which they are based and work gradually towards a more general solution for use on a broad range of data. Technical information is given in detail in maths boxes for those who are mathematically literate.

Simple Image Segmentation.

The first thing we need to know about MR and CT data is that to a very good approximation the greylevel values in an image can be assumed to be formed by a linear process. This means that the contribution to the signal in any pixels is simply proportional to the relative fractions of each tissue within the voxel [3].

The expression for the signal intensities of pure tissues in an inversion recovery spin-echo (IRSE) sequence follows directly from the Bloch equations and is

$$S \;=\; N(H)(1 - 2e^{(-TI/T_1)} \;+\; 2e^{(-(TR-\tau)/T_1)} \;-\; e^{(-TR/T_1)})e^{(-TE/T_2)}$$

where $N(H)$ is the spin density and TE is the echo time. The modern equivalent to (IRSE) is the inversion recovery turbo spin-echo (IRTSE) sequence. The expression for the signal intensities for pure tissues is

$$S \;=\; N(H)(1 - 2e^{(-TI/T_1)} \;+\; 2e^{(-(TR-T_s N_f)/T_1)} \;-\; e^{(-TR/T_1)})e^{(-TE_{eff}/T_2)}$$

where T_s is echo spacing and N_f is the factor number of the TSE train.
The linear dependency of expressions on $N(H)$ are typical for MR sequences with the consequence that the grey level within any voxel g_v can be written as the linear contribution from a set of partial volumes

$$g_v \;=\; p_1 G_1 \;+\; p_2 G_2 \;+\; p_3 G_3 \;+\; ... \;+\; p_N G_N$$

with

$$p_1 \;+\; p_2 \;+\; p_3 \;+\; ... \;+\; p_N \;=\; 1$$

where p_n is the nth partial volume and G_n the mean grey level for pure tissue.

Box 1: Image Formation in MR.

Armed with this assumption it is possible to justify simple approaches to tissue boundary identification, such as thresholding. This is because a given greylevel corresponds directly to particular fractional proportions of a given pair of tissues. A fraction of 50 percent of two tissues defines the most likely location for a boundary. This is used as the basis for many visualisation techniques which require the identification of surfaces [4].

Such an algorithm will deal adequately with identification of tissue boundaries provided that we are always looking for the boundary between the same two tissues and that there are no processes during image formation which invalidate our assumption. Unfortunately, both of these requirements are generally not met by the majority of images.

The most likely failiure in the assumptions behind the thresholding approach is that the data is not simply composed of two tissues. A technique for boundary location which can work with multiple tissues is based closely on the way that humans percieve image data. The idea involves identifying the boundaries between otherwise homogenous regions by locating the positions of maximum contrast or "edges". Edge detectors are more common in the field of machine vision than medical image analysis [5] and are generally based on the idea of computing the local spatial derivative of an image after smoothing. Taking the peak of this derivative defines the maximum transition point between the tisues which is generally also the 50 percent probability boundary between any two adjacent tissues.

Detection of edge boundaries in 2D is generally performed in a manner similar to the following;

- Convolution of the image with a spatial noise reduction filter, (eg: a Gaussian) $I' \;=\; I \otimes G(x, y)$.
- Calculation of local image gradients $\delta_x = (\partial I/\partial x)$ (ie: $I' \otimes (-1, 0, 1)$) , $\delta_y = (\partial I/\partial y)$ (ie: $I' \otimes (-1, 0, 1)^T$).
- Calculation of an edge strength $E(x, y) \;=\; \delta_x^2 \;+\; \delta_y^2$
- Identification of edges as those voxels with edge strengths above a statistical threshold $(E(x, y) \;>\; k)$ and less than no more than two of its neighbours (implying simple linear connectivity).

Box 2: A simple edge detector.

While this technique will work well on sharp boundaries the more slowly varying partial volume regions which occur in MR data can cause problems with such an approach.

Accounting for the partial volume behaviour of multiple tissues can be compensated for but may require more information than is present in a single image. In fact obtaining the direct solution for the proportion of each of N tissues within a voxel is an exercise in linear algebra and requires N-1 images.

The three linear equations for the grey level value in two images and the total proportion constraint can be solved for each tissue of three tissues (eg: c,w,g) within each voxel v as follows:

$$p_{cv} = \frac{g_{1v}(G_{2w} - G_{2g}) - g_{2v}(G_{1w} - G_{1g}) - (G_{1g}G_{2w} - G_{2g}G_{1w})}{(G_{1c} - G_{1g})(G_{2w} - G_{2g}) - (G_{2c} - G_{2g})(G_{1w} - G_{1g})}$$

$$p_{gv} = \frac{g_{1v}(G_{2c} - G_{2w}) - g_{2v}(G_{1c} - G_{1w}) - (G_{1w}G_{2c} - G_{2w}G_{1c})}{(G_{1g} - G_{1w})(G_{2c} - G_{2w}) - (G_{2g} - G_{2w})(G_{1c} - G_{1w})}$$

and

$$p_{wv} = 1 - p_{cv} - p_{gv}$$

Box 3: Direct estimation of partial volume fractions using linear algebra.

Such an approach will deliver unbiased estimates of tissue proportion ([1, 8]). However, it can only deliver correct estimates for the tissues within the model, meaning it cannot deal with unexpected (or pathological) behaviour. From a medical standpoint this is equivalent to saying that it can only deal with normal tissues. Further, the assumption of a pure linear model is equivalent to the statistical assumption of noise free data. The consequence of this is that estimates of tissue proportion are noisy and values can be outside the physical range of 0-100 percent. Dealing with both of these problems requires a more overtly statistical approach to data analysis.

What we must do is apply the methodology of probability theory directly to the modelling of data. This involves constructing a likelihood model for each tissue component present in the data. A common approach involves modelling only the pure tissue distributions, but in order to account for partial volume effects partial volume distributions must also be modelled [11] (Figure 1). The various parameters in the density model must be determined using an optimisation algorithm to minimise the difference between the model and the data (The simplex algorithm [6] and Expectation Maximisation [7]. Estimation of relative tissue probabilities can then be made by the direct use of Bayes theory.

The total probability of getting a particular set of grey level values (g) within a region of the image comprised of entirely three tissues (1-3) and two sets of partial volumes can be written as;

$$P_{tot}(g) = f_a P_1(g) + f_b P_2(g) + f_c P_3(g) + f_d P_{12}(g) + f_d P_{21}(g) + f_e P_{23}(g) + f_e P_{32}(g)$$

The separate components for the likelihood of each grey level value from pure tissues can be written as;

$$P(g|1) = f_a P_1(g) + f_d P_{12}(g)$$

$$P(g|2) = f_b P_2(g) + f_d P_{21}(g) + f_e P_{23}(g)$$

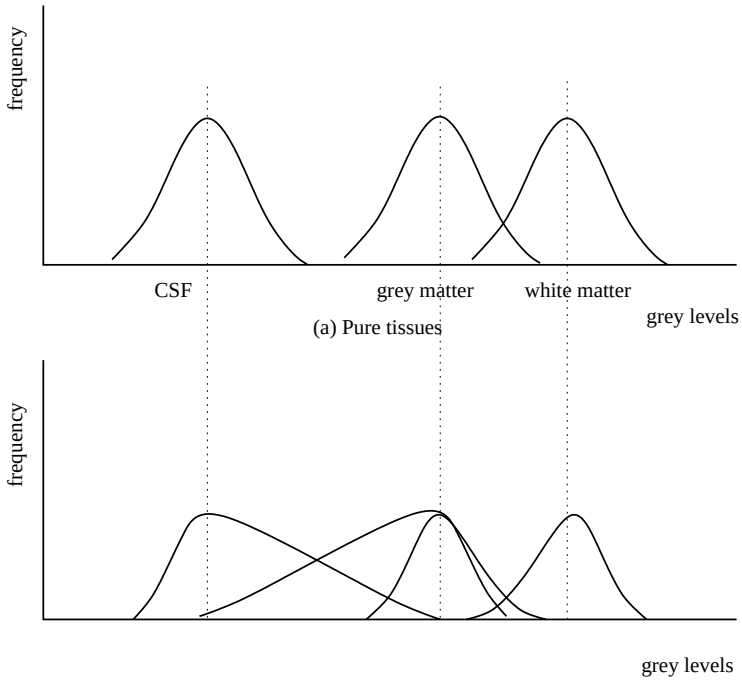
$$P(g|3) = f_c P_3(g) + f_e P_{32}(g)$$

We can now use Bayes theory to compute the conditional probability of a grey level of a particular value being due to one of the three classes n as.

$$P(n|g) = P(g|n)/P_{tot}(g)$$

Box 4: Bayes classification of grey level vaues for three tissues.

The delivered probabilities are exactly those described in the introduction. If the derived frequency model is an accurate representation of the data then the result delivered by this technique is the optimal solution to the problem of segmenting data on the basis of voxel grey levels. Understanding the performance of such an algorithm thus changes from asking; How well does this algorithm work?, to; How well does the data conform to the assumptions used in this algorithm? This second question is one that we may be in a better position to answer on a new data set. The probability labelling technique will work with multiple tissues on a single image provided that the grey level distributions do not overlap significantly. Overlapping tissues can be eliminated by the use of multiple images, as ambigious regions in the data can be separated with additional information. However, this does involve a slightly more complicated analysis in order to determine all of the parameters in the multi-dimensional model. Extension of this technique to deal with pathological (unmodelled) tissues can be incorporated by allowing an additional category for infrequently occuring data [9].



(a) Pure tissues

(b) Partial volume tissues

Figure 1: Probability distributions for brain tissues. *Pure tissues have Gaussian distributions and partial volume distributions for paired tissue combinations take the form of a triangular distribution convolved with a Gaussian which is intended to model the response function of the measurement system.*

For multi-spectral data $\mathbf{g}$ we must define a multi-variate distribution for each pure tissue t.

$$P_t(\mathbf{g}) \;=\; \alpha_t e^{-(\mathbf{g}-\mathbf{G_t})^T C_t (\mathbf{g}-\mathbf{G_t})}$$

Where G_t is the mean tissue vector and C_t its covariance and α chosen to give unit normalisation. Partial volume distributions can be modelled along the line between two pure tissue means G_t G_s.

$$P_{ts}(\mathbf{g}) \;=\; \beta_{ts} P_{ts}(h) e^{-(\mathbf{g}-\mathbf{h}.\mathbf{g}/|h|)^T C_h (\mathbf{g}-\mathbf{h}.\mathbf{g}/|h|)}$$

with $\mathbf{h} \;=\; (\mathbf{g}-\mathbf{G_s})C_h(\mathbf{G_t}-\mathbf{G_s})/|(\mathbf{G_t}-\mathbf{G_s})C_h(\mathbf{G_t}-\mathbf{G_s})|$, $C_h \;=\; C_t\mathbf{h} \;+\; C_s(1-\mathbf{h})$ and $P_{ts}(h)$ is the 1D partial volume distribution such as used in Figure 1. Parameters for the model can be iteratively estimated by taking weighted averages over the selected volume V using a process generally referred to as Expectation Maximisation (EM).

$$f_t \;=\; \sum_v^V P(t|\mathbf{g_v}) \qquad , \qquad f_{ts} \;=\; f_{st} \;=\; \frac{1}{2}\sum_v^V P(ts|\mathbf{g_v}) \;+\; P(st|\mathbf{g_v})$$

$$G'_t = \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v)\mathbf{g}_v \qquad , \qquad {C'}_t^{-1} \;=\; \frac{1}{V}\sum_v^V P(t|\mathbf{g}_v)(\mathbf{g}_v-\mathbf{G_t})\otimes(\mathbf{g}_v-\mathbf{G_t})^T$$

Unknown tissues are included in the Bayesian formulation by including a fixed extra term f_o for outlying data points in $P_{tot}(\mathbf{g})$.

Box 5: Multispectral modelling with an unknown tissue.

Dealing with Data In-Homogenaeity.

Having dealt with the issues of partial volumes, noise and pathological data the only remaining problem which is likely to be encountered in real data is that the assumption of a pure linear model for the image formation process is not correct due to spatially varying nominal tissue data values. This can occur for one of two reasons, either the tissue itself has physical properties that genuinely vary depending upon the location, or the values appear to change due to inhomogenaeity in the measurement system. The former of these we will return to below. The latter factor can be corrected if we can correct for the spatially varying sensitivity (or gain) of the system. This will be in the form of a multiplicative correction image. Of the algorithms described above the only one which could potentially deal with such a problem without modification is the edge based approach, which operate in a purely local fashion.

There are two approaches to determining a gain correction. The first involves building a low parameter model for the expected gain correction into the solution for the label probabilities. These parameters can then be adjusted via an optimisation process which minimises the variances of the pure tissues [2]. This approach has several drawbacks. Firstly we may not know the correct parametric function for a given correction image, it would be very easy to assume a functional form which did not match the true characteristics. Secondly, the approach cannot work well if there are regions of unmodelled tissue in the data, so pathalogical data is once again excluded. Finally, determination of a set of parameters will need to be achieved via an iterative process if it is to stand any chance of making use of robust statistical assumptions. Iterative processes on data sets of this size are both slow and unreliable.

The alternative approach involves assuming that the image is composed of homogenous regions of tissue separated by partial volume boundaries. Provided that we can detect partial volume regions (using for example a local contrast or derivative estimate), we can estimate the local relative gain change across a voxel in any uniform region. Applying the method to image slices from a 3D volume there will be two relative gain change images, one for the horizontal and the other for vertical changes. There will be some regions where we have no estimates, and some relative gain estimates will be more accurate than others depending upon the local signal to noise. However, spatial gain variations are expected to be locally smooth and we can make direct use of this information. By smoothing the local gain change by an amount less than the expected level of smoothness in the data we can fill in the missing data and increase the accuracy and stability of local estimates. This must be done using an appropriate statistical calculation which also takes account measurement accuracy. The relative horizontal and vertical gain changes determined can then be used to compute (via integration) an estimate of the original gain variation image.

Calculation of a smooth correction image can be perfomed as follows;

- estimation of local image noise
- estimation of local image relative gradients $\Delta_x = (\partial I/\partial x)/I$, $\Delta_y = (\partial I/\partial y)/I$ and variances σ_x^2 σ_y^2.
- Maximum Likelihood estimation of smoothed local derivative using statistical averaging with a stability term for missing data w_{reg} (points with large gradient relative to noise) which assumes no image slope 0_{reg}.

$$\Delta_x^{ML}(x,y) \;=\; \frac{S \otimes (\Delta_x(x,y)/\sigma_x^2(x,y) + 0_{reg})}{S \otimes (1/\sigma_x^2(x,y) + w_{reg})}$$

- Integration of these derivatives along any path L from l_0 to $l \;=\; (x,y)$ can be witten as

$$\int_{l_0}^{l} \Delta_l^{ML}(l)dl \;=\; \int_{l_0}^{l} \frac{\partial F}{F_l} \;=\; [log(F_l)]_{l_0}^{l}$$

  defining the starting point as unity gain gives the relative gain factor to that point $F(xy)$

$$F(x,y) \;=\; F_l \;=\; exp(\int_{l_0}^{l} \Delta_l^{ML}(l)dl)$$

The regularisation for missing data requires us to iterate the algorithm a few times for data with large regions of indeterminate slope.

Box 6: Gain correction using local slope.

In comparison to the iterative parametric approach, this technique is simple, fast, reliable (non-iterative) and does not have to assume a particular parametric form for the gain variation [10]. Yet it can also deal with unmodelled data provided that it is composed of homogenous regions, or regions that have high spatial derivative (and therefore get excluded as partial volumes). For large regions of slight inhomogenaiety there will be a distortion of the local estimated gain variation. If we are not modelling this tissue explicitly this has no consequence in terms of quantitative measurement.

Returning now to the issue of tissue variability, the direct estimate of gain variation from local changes will of course also attempt to model, and correct for, gradual changes across tissues. The level to which they will manage to do this will depend upon the assumed scale of image smoothness. Large regional changes in tissue are likely to be corrected just as though they were due to the measurement process. However, this is not a problem if we are going to assume that all tissues can be modelled using a single mean value in the subsequent algorithms.

Segmentation of Multiple Tissues and Boundaries.

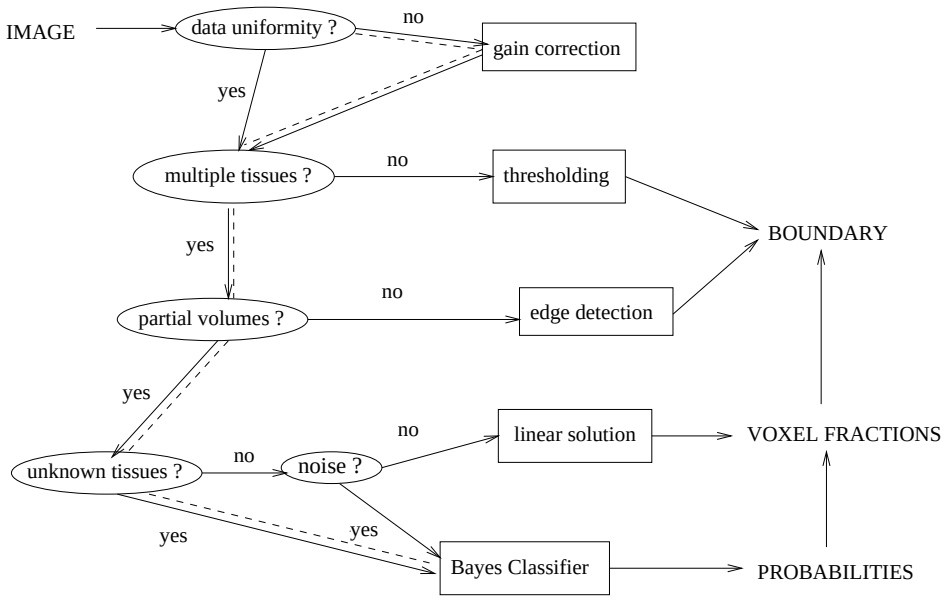


Figure 2: Tissue segmentation strategy.

The overall strategy for selecting a segmentation method based upon the characteristics of the data is shown diagramatically in figure 2. Answers to each of the questions in the bubbles determines the correct approach for a particular data set. One component of algorithm evaluation can therefore be done at the theoretical level of what variabilities we are taking account of in the model. The above diagram specifies the tests which need to be done in order to confirm that an algorithmic approach is at least appropriate. Clearly, formal evaluation is then also necessary for any subset of algoriths which may be identified as suitable in order to establish the "best".

The most general approach to tissue segmentation, which deals with all of the common problems found in typical data sets, is shown by the dotted curve and comprises gain correction and Bayes estimation of conditional probabilities. As this method can be implemented as a quick and reliable algorithm this is the preferred method for general image segmentation.

References

1. H.Rusinek, et. al. Alzheimer's Disease: measuring loss of cerebral grey matter with MR imaging. Radiology, 178(1), p109-14, 1991.
2. R. Guillemaud and J.M.Brady, Estimating the bias Field of MR Images, IEEE TRANS. on Medical Imaging, 16(3), 238-251, 1997.
3. S.Webb, The Physics of Medical Imaging, Medical Science Series, Pub. Adam Hilger ,Bristol, Philadelphia and New York, 1988.
4. W.E.Lorensen, H.E.Cline, Marching Cubes: A High Resolution 3D Surface Construction Algorithm, SIGGRAPH '87, ACM Computer Graphics, 21(4), 163-169, 1987.
5. Canny, J.F., 'A Computation Approach to Edge Detection', IEEE PAMI, 8, pp. 679-698, 1986.
6. W.H.Press B.P.Flannery S.A.Teukolsky W.T.Vetterling, Numerical Recipes in C, Cambridge University Press 1988.
7. W.M.Wells, E.L.Grimson, R.Kikinis et al. Adaptive segmentation of MRI data. IEEE Trans. on Medical Imaging, 15(4), 429-442, 1996.
8. N.A.Thacker, A.Jackson, X.P.Zhu and K.L.Li, 'Accuracy of Tissue Volume Estimation in MR Images ', Proc. MIUA , 137-140, Leeds, 1998 .
9. K.Fukenaga, 'Introduction to Statistical Pattern Recognition', 2ed, Academic Press, San Diego (1990).
10. E.Vokurka, N.Thacker, A.Jackson, "A Fast Model Independant Method for Automatic Correction of Correction of Intensity Non-Uniformity in MRI Data." JMRI, 10, 4, 550-562, 1999.
11. Laidlaw DH, Fleischer KW, Barr AH, Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms, IEEE Trans. Med. Imag., vol. 17, no. 1, 74-86, Feb. 1998.
12. URL: www.niac.man.ac.uk