

# Noise Filtering and Testing Illustrated Using a Multi-Dimensional Partial Volume Model of MR Data

N.A. Thacker, M. Pokrić, D.C. Williamson.

Imaging Science and Biomedical Engineering

School of Medicine

University of Manchester

Manchester M13 9PT, UK

neil.thacker@man.ac.uk

## Abstract

One of the most common problems in image analysis is the estimation and removal of noise or other artefacts using spatial filters. Common techniques include Gaussian, Median and Anisotropic Filtering. Though these techniques are quite common they must be used with great care on medical data, as it is very easy to introduce artifact into images due to spatial smoothing. The use of such techniques is further restricted by the absence of a 'gold standard' data against which to test the behaviour of the filters. Following a general discussion of the equivalence of filtering techniques to likelihood based estimation using an assumed model, this paper describes an approach to noise filtering in multi-dimensional data using a partial volume data density model. The resulting data sets can then be taken as a gold standard for spatial filtering techniques which use the information from single images. We demonstrate equivalence between the results from this analysis and techniques for performance characterisation which do not require a 'gold standard'.

## 1 Introduction

A common problem across the field of computer vision is the need to define effective methods for the characterisation of algorithms [4, 6]. By characterisation we mean that we need a quantitative way of assessing the performance of an algorithm which makes it possible to decide if the data generated is suitable for subsequent tasks. Despite over 30 years of research in this field this issue is still largely unsolved in many classes of problems. Recent advances include the use of ROC curves for feature detection evaluation [1]. It is tempting to think that most problems should be solved by identifying a "gold standard" against which to measure performance. However, in many real tasks it cannot be obtained and alternative approaches have to be used to measure and characterise algorithm performance [8, 11]. The work in this paper illustrates an approach to construct a "gold

standard” for Magnetic Resonance (MR) images and at the same time compare the results to alternative methods. We hope to illustrate, using the noise filtering as an example, that in some cases characterisations which would initially appear to require a ”gold standard” can, with some ingenuity, be done without one.

Noise filtering of any data involves the assumption of a specific image generation mechanism (or image model). The process of Gaussian smoothing can be interpreted as consistent with a likelihood estimation of the central value within a region. This is done on the assumption that the data within this region can be described by some functional model with the expected grey level residuals being drawn from a Gaussian noise distribution with variance inversely proportional to the spatial Gaussian weighting term. The specific form of the assumed model is best understood by considering a more simplistic problem first. Gaussian filtering removes high spatial frequency content from the structure of images. A less destructive approach to noise filtering is based on the concept of anisotropic filtering, where the data is preferentially smoothed along a direction selected in order to minimise the loss of spatial structure in the image. One particular variant of this we call *Tangential Filtering*. Here, the tangential direction to the local image slope is computed and the data is smoothed by taking the weighted average of points along this line situated on either side of the central value. Averaging of multiple values in this way assumes that the data can locally be fitted to a 1D line and selection of the tangential direction results in the least destructive impact on edge structure. In fact any anti-symmetric function will result in an appropriate central estimate, and this can be taken as the most general assumption for the underlying image model at each point. We can also see that for an average of two points the noise in the resulting image should be reduced by typically a factor of  $\sqrt{2}$  of the original image noise <sup>1</sup>. Returning to Gaussian filtering, we can interpret this process as an averaging of multiple estimates of the central value for any pair of pixels with equal weight in the Gaussian kernel. The class of functions for which this would be an appropriate model would include Cartesian polynomials (expanded as a function of shifts  $(x, y)$ ) from the central location  $(x_0, y_0)$ , either with no even terms, or at least exact cancellation of the magnitudes of even power coefficients. For example:

$$I(x - x_0, y - y_0) = a + bx + cy + dxy + ex^2 - ey^2 + \dots \quad (1)$$

Though the only global function for which the model would work correctly at every image location would be an inclined plane. When described in this way it is easy to see the oversimplicity of this model in comparison to the structures found in real medical images. It is this which results in the characteristic problems with Gaussian filtering of image smoothing and the loss of sharp edge structures.

As a direct contrast to spatial image filtering, which assumes specific forms of spatial correlation between grey level values, multi-dimensional tissue segmentation algorithms rely instead upon correlations between multiple measurements of the same physical location (voxel) using different imaging modalities. The typical approach involves building a model not of spatial structure but of grey level density distribution. A common form of noise removal which makes use of grey-level density distribution is the median filter, which can be considered as a bootstrapped likelihood estimator. While many authors have concentrated on the use of grey level density estimation for tissue *labelling*, estimates of tissue *volume proportion* [9] can be used to predict the original image content for each

---

<sup>1</sup> $\sqrt{3}$  if we also include the central value in the estimate

modality in the case of zero noise. The process is simply one of using the estimated model parameters to generate the expected grey level values, using the linear equations implicit in the segmentation. Thus multi-spectral segmentation techniques can be used as the basis for noise filtering.

Many readers may find the interpretation of filtering methods from a perspective of a model based estimation process unusual. However, one advantage of this approach is that explicit identification of the assumed model makes it possible to consider testing the conformity of the data under analysis to the model. This is something which has not generally been tested for spatial filtering approaches. We explain below how this can be used to prevent noise removal in un-modelled parts of the data, i.e. pathology. In addition, we will demonstrate how the results from multi-spectral noise filtering are usable as a gold standard for spatial filtering techniques.

## 2 Methods

We will assess the stability of the selected filtering schemes first using a Monte-Carlo approach. Here a small quantity of noise is added to the input image and the relative change in output grey-level values is measured. However, this technique is not sufficient as an evaluation, as it will not measure failure of the implicit filtering model. By interpreting each noise filtering technique as an estimation process we suggest that it is possible to assess the validity of the filtering method by observing the number of grey-level values which are changed by more than 3 standard deviations (S.D.) of the estimated image noise, i.e. a residual outlier measurement (ROM). To estimate the image noise to set the scale of this measurement, we use a technique we call Local Noise Estimation (LNE). This is based upon the observation that high order derivatives are heavily corrupted by image noise. On the assumption of uniform independent noise we can use error propagation to predict the expected level of noise on any order of spatial derivative. By measuring the width of the distribution of derivatives around the peak at zero we can get an estimate of the original image noise by scaling with the appropriate error propagation factor. We use second derivatives which means that the technique is also consistent with measuring the deviation from a purely linear spatial grey-level model.<sup>2</sup>

The Monte-Carlo assessment and the ROM measure are complementary aspects of performance, the first giving an indication of the best case noise filter on the assumption that the filter model is adequate and the second giving an indication of how often this model is inappropriate. Importantly, neither technique requires a “gold standard” to test against. Ideally, we wish to show that these techniques give a complete characterisation of the algorithm performance and for this we will need a surrogate “gold standard”. We choose to use for this the Multi-Spectral filtering method which will now be described.

Multi-dimensional Gaussian distributions are used to model the effects of both inherent tissue variability and measurement noise for pure tissues. A multi-variate Gaussian distribution for multi-dimensional data  $\mathbf{g}$  for each pure tissue  $t$  is defined as:

$$d_t(\mathbf{g}) = \alpha_t e^{-\frac{1}{2}(\mathbf{g}-\mathbf{m}_t)^T C_t (\mathbf{g}-\mathbf{m}_t)} \quad (2)$$

---

<sup>2</sup>Although this method will estimate the image noise well (within a few percent) for data with spatially uncorrelated noise, we cannot use this technique to estimate noise following spatial filtering.

where:

- $\mathbf{m}_t$  is a mean tissue vector
- $C_t$  is inverse of a covariance matrix
- $\alpha_t$  is a constant which gives unit normalisation

Using an assumption of a linear image formation process, whereby the total intensity level at each voxel results from summation of different tissue fractions present at that particular voxel, the partial volume distribution can be thought of as being composed of two triangular distributions convolved with a Gaussian ( $T_{ts}(\mathbf{g}) + T_{st}(\mathbf{g})$ ). Where  $T_{ts}(\mathbf{g})$  is the local density estimate for tissue  $t$  generated by a partial voluming process with tissue  $s$ . Assuming that tissue variability is more significant than the measurement processes, multi-dimensional partial volume distributions can be modelled along the line between two pure tissue means  $\mathbf{m}_t$  and  $\mathbf{m}_s$ :

$$d_{ts}(\mathbf{g}) = \beta_{ts} T_{ts}(h) e^{-(\mathbf{g}-\mathbf{h}\cdot\mathbf{g}/|\mathbf{h}|)^T C_h (\mathbf{g}-\mathbf{h}\cdot\mathbf{g}/|\mathbf{h}|)} \quad (3)$$

where the parameters for the given data  $\mathbf{g}$  are:

- $h$  is a fractional distance between two centres of distribution [ $0 < h < 1$ ]
- $h = (\mathbf{g} - \mathbf{m}_s) C_h (\mathbf{m}_t - \mathbf{m}_s) / |(\mathbf{M}_t - \mathbf{m}_s) C_h (\mathbf{m}_t - \mathbf{m}_s)|$
- $C_h$  is a covariance matrix:  $C_h = C_t h + C_s (1 - h)$
- $T_{ts}(h)$  is the 1D partial volume distribution between pure tissues  $t$  and  $s$ .
- $\beta_{ts}$  is a constant which gives unit normalisation

The 1D partial volume distribution,  $T_{ts}$ , obtained by convolving a triangular distribution normalised to  $\frac{1}{2}$  with Gaussian distribution normalised to 1, takes the following form [13]:

$$T_{ts}(x) = -\frac{kx+c}{2} \left\{ \text{erf}\left(\frac{x-b}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x-a}{\sigma\sqrt{2}}\right) \right\} - \frac{M\sigma}{\sqrt{2\pi}} \left\{ \exp^{-\frac{(x-b)^2}{2\sigma^2}} - \exp^{-\frac{(x-a)^2}{2\sigma^2}} \right\}. \quad (4)$$

where:

- $x$  is a grey level value calculated as a normal projection of vector  $\mathbf{g}$  onto line between two distribution means
- $k$  and  $c$  are the slope and intercept of the line which forms the triangle
- $a$  and  $b$  are the start and end points of an interval at which the triangular distribution has a non-zero value
- $\sigma$  is a standard deviation of a Gaussian function

The overall partial volume distribution is calculated as a product of a Gaussian function of the normal distance ( $\mathbf{g} - \mathbf{h}\cdot\mathbf{g}/|\mathbf{h}|$ ) from two distribution centres and the 1D partial volume distribution  $T_{ts}(h)$ . Examples of the types of distributions obtained from the model parameters of two images for three pure tissues and their partial volumes are shown in Figure 1. It can be seen that pure tissue distribution models take the form of elliptical features, while the partial volumes are shown as elongated structures between centres of distributions.

Parameters of the model can be iteratively estimated using the Expectation Maximisation (EM) approach [5, 10]. EM is used to estimate the parameters by maximising the

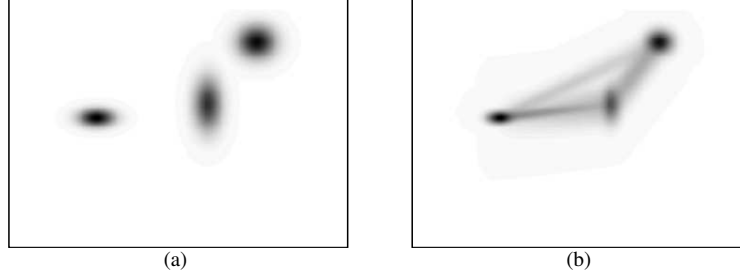


Figure 1: An example of distributions generated from the model for two images (a) Pure tissue distributions (b) Combined distributions of pure tissues and partial volumes between centres of pure tissues

likelihood of the data distribution. This involves first getting from the likelihood distributions defined above to a probability of a given tissue proportion given the data  $P(t|\mathbf{g}_v)$ . The conditional probability of a grey level being due to a certain mechanism  $n$  (either a pure or mixture tissue component) can be calculated using Bayes theory, as follows:

$$P(n|\mathbf{g}) = \frac{d_n(\mathbf{g})f_n}{\sum_t(f_O + d_t(\mathbf{g})f_t) + \sum_t \sum_s d_{ts}(\mathbf{g})f_{ts}} \quad (5)$$

where  $f_n$ ,  $f_O$ ,  $f_t$  and  $f_{ts}$  are effectively "priors", expressed here as frequencies (i.e. number of voxels) which belong to a particular tissue type, pure tissues or partial volumes. Unknown tissues are accounted for in the Bayesian formulation by including a fixed extra term  $f_O$  for infrequently occurring outlier data [2] in total probability which enables separation of pathological tissues.

Noise free estimates of the individual image grey-levels  $g'_i$  can be calculated using

$$g'_i = g_i P(O|\mathbf{g}) + \sum_t g_{it} P(t|\mathbf{g}) + \sum_t \sum_s g_{it} P(ts|\mathbf{g}) \quad (6)$$

Where  $g_{it}$  is the expected pure tissue grey level for image  $i$ . This formulation implicitly reverts to the original image grey-level for outlier data (large  $P(O|\mathbf{g})$ ). In addition, failure to model individual voxels can be identified by checking for consistency between the filtered image and original. In particular, any reconstructed grey-level value which differs from the original by more than 3 standard deviation of the image noise can be said to be inconsistent with the model. This value can then be replaced with the original value in order to preserve all significant information present in the original image.

Finally, in order to use the results from the multi-spectral technique as gold standard there are a few issues which must be addressed. The first is that multi-spectral filtering is expected to eliminate spatially distributed errors, such as field in-homogeneity. Although these are expected to be small in our data (in comparison to intrinsic image noise) they also vary between acquisitions and will bias comparisons of residual distributions. In order to eliminate the majority of these effects we construct our pseudo-gold standard by removing a smooth estimate of local difference between the multi-spectral reconstruction and the original image, (constructed using a 5 pixel S.D. Gaussian kernel). Secondly, the multi-spectral noise filtering process removes noise in a tissue dependant manner (also removing genuine tissue variability which cannot be interpreted as a partial volume process), so that residual difference distributions are not simple Gaussians. Therefore we fit all residual

distributions within 3 S.D. of the estimated image noise with the sum of two Gaussians with the same mean but individual widths and normalisations (parameters A and B in Table 2). It is the results of this fitting process which we wish to reconcile with the results from Monte-Carlo and ROM quantification.

### 3 Results

Original and reconstructed images following co-registration and partial volume analysis are shown in Figure 2.

The noise level ( $\sigma_{original}$ ) in each image was estimated using the LNE technique. Tangential filtering was applied by averaging over three pixels (one central and two either side). Gaussian filtering was done using a spatial filter with S.D. of 1 pixel. Median filtering was performed over the local neighbourhood of 9 pixels. The Monte-Carlo stability analysis estimates of the fraction of noise remaining after filtering and the number of voxels lying beyond 3 S.D. of the model (ROM) are listed in Table 1.

	LNE	Median	Gaussian	Tangential	Multi-Spectral
<b>IRTSE</b>	58.76	0.64 (1559)	0.27 (2405)	0.66 (698)	0.22 (1689)
<b>VE(PD)</b>	64.06	0.66 (1466)	0.26 (3127)	0.68 (530)	0.20 (1804)
<b>VE(T2)</b>	58.2	0.63 (1287)	0.26 (1909)	0.69 (426)	0.17 (938)
<b>FLAIR</b>	52.4	0.63 (1934)	0.27 (3827)	0.69 (966)	0.13 (4971)

Table 1: Monte-Carlo and ROM Estimates

Following partial volume filtering the reconstructed images were corrected for low frequency spatial noise as described in the methods section. The resulting images were then used as the basis for evaluation of spatial filtering techniques applied to the original images (Table 2).

	Original					Median Filter				
	Mean	A	$\sigma_1$	B	$\sigma_2$	Mean	A	$\sigma_1$	B	$\sigma_2$
<b>IRTSE</b>	-4.6	0.47	44.5	0.53	75.3	0.39	0.39	21.9	0.61	63.7
<b>VE(PD)</b>	-4.9	0.54	37.5	0.46	90.6	-7.7	0.60	25.7	0.40	81.5
<b>VE(T2)</b>	-7.0	0.34	23.9	0.66	62.9	-7.1	0.58	21.7	0.44	61.5
<b>FLAIR</b>	0.27	0.50	37.1	0.50	93.9	0.15	0.49	20.7	0.51	74.0
	Gaussian Smoothing					Tangential Smoothing				
	Mean	A	$\sigma_1$	B	$\sigma_2$	Mean	A	$\sigma_1$	B	$\sigma_2$
<b>IRTSE</b>	1.1	0.35	14.5	0.65	65.3	-1.1	0.46	29.3	0.54	65.4
<b>VE(PD)</b>	0.77	0.36	10.6	0.64	52.2	4.3	0.51	24.2	0.49	75.1
<b>VE(T2)</b>	1.12	0.35	8.6	0.65	43.7	4.4	0.41	17.2	0.59	49.7
<b>FLAIR</b>	0.79	0.38	10.6	0.62	56.2	0.2	0.43	19.7	0.57	68.1

Table 2: Quantitative Performance of Spatial Filtering

Though the fits to the double Gaussian are less stable than those from the Monte-Carlo the results confirm that for the central narrow component the widths of the residual distribution vary by amounts consistent with the prediction. In addition the proportion of the secondary Gaussian components and secondary peak widths give an estimate of

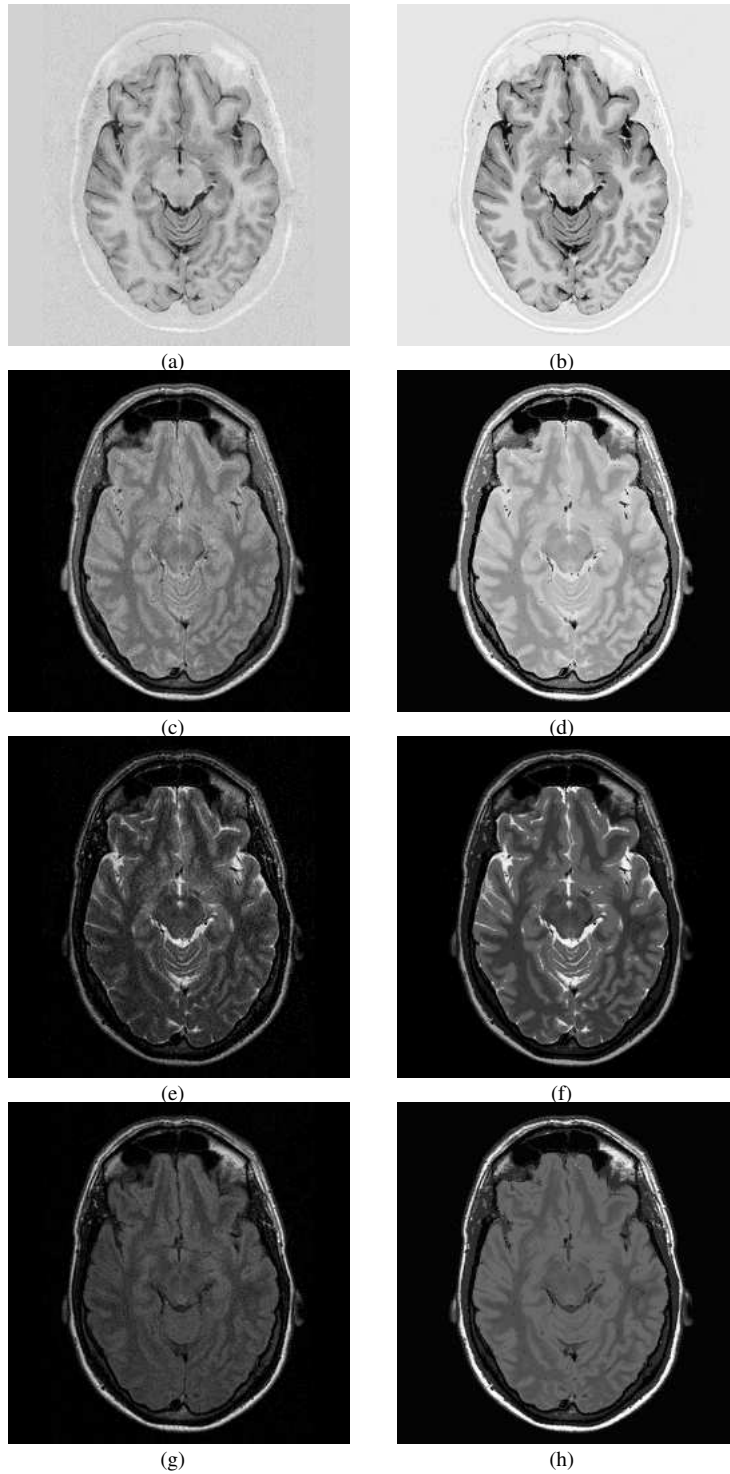


Figure 2: (a) Inversion Recovery Turbo Spin Echo (b) Filtered IRTSE (c) Variable Echo (PD) (d) Filtered VE(PD) (e) Variable Echo (T2) (f) Filtered VE(T2) (g) FLAIR (h) Filtered FLAIR

the number of failures. While the Gaussian filter reduces the width of both the primary and the secondary Gaussian distribution, the proportion of data in the broader part of the distribution has also increased. The proportion of data in the secondary Gaussian remains the same for tangential smoothing and median filtering as for the original image. However, there is additional broadening of the secondary distribution for the median filter, as predicted by the ROM data. The data are in general reconcilable with ROM estimates, with the one exception being the lack of evidence for significant outlier proportion in multi-spectral filtering of FLAIR images. We believe that this is most likely due to flow artefact in the FLAIR images which has then been removed from the reference image by low frequency residual subtraction. A summary of the results for all filtering techniques are given in Figure 3 by plotting ROM (in voxels) vs Monte-Carlo (the fraction of noise remaining after filtering) estimates. On this graph we see that in general multi-spectral filtering is no more destructive to image contents than tangential smoothing or median filtering but removes more image noise than Gaussian smoothing. Therefore, such image reconstruction techniques may be a useful way of getting the most from multiple MR acquisitions. Software and test data are available on the web ([www.tina-vision.net](http://www.tina-vision.net)).

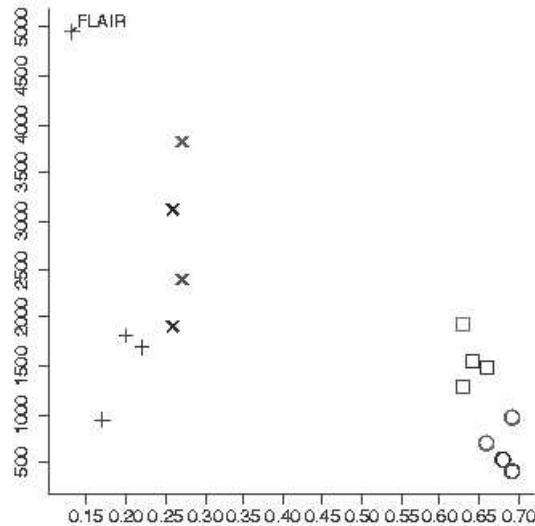


Figure 3: Residual Outlier Measure (ROM) vs Monte-Carlo Stability: Multi-spectral Filtering +; Gaussian Smoothing X; Median filtering □; Tangential Smoothing O.

## 4 Discussion and Conclusions

Multi-spectral filtering can be considered as a regression onto the lines joining pairs of pure tissue locations in the multi-dimensional grey level space, followed by a weighting with pure tissue values according to the Bayesian priors. Any voxels composed of pure tissues of appropriate mean values will therefore have the noise on each grey level removed in such a way as to make the grey level value more consistent with the estimated



position along this partial volume line (Figure 4). The results of such an analysis can be also interpreted as a method of data fusion, where data from alternative modalities are combined in order to improve the data from each. The results demonstrate that such an approach does not produce the loss of high spatial frequency structure inherent in even the most careful spatial filtering schemes.

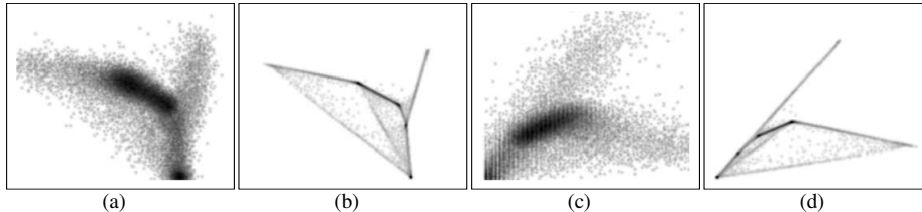


Figure 4: Grey Level Distributions before (a) and after (b) Partial Volume Noise Filtering for IRTSE and VE(PD) images, and before (c) and after (d) Partial Volume Noise Filtering for VE(T2) and FLAIR images.

There is a subtle but important difference between using models based upon spatial distribution and those based upon partial volume behaviour for MR analysis. The former can only be determined from example data and there can never be a spatial model which will be appropriate for the contents of all biological images. Grey level density models however, have statistical characteristics which are purely determined by the acquisition process (i.e. the underlying physics of the measurement process). We might therefore expect that if we knew enough about the image formation process for a particular imaging protocol we may be able to construct a model which is true for all images from a particular acquisition containing equivalent tissue types. This approach to filtering may therefore be regarded as a gold standard for testing of spatial filtering techniques. In comparison to other mechanisms for the generation of test data, the most commonly used technique is probably the BRAINWEB [3, 7] simulation. This can be thought of as equivalent to the final stages of the reconstruction process presented here, except that their volume estimation process is based upon a "fuzzy" minimum distance classifier, not explicit partial volume estimation, and of course they have one fixed model. Our approach raises the possibility of generating personalised models with checks on the validity of the reconstruction process.

This paper has shown how techniques generally used for tissue segmentation can be used to provide noise filtering of multi-spectral images based upon an analysis of partial volume structure. We have used this data to corroborate the use of performance measures which do not require a 'gold standard'. We conclude that the combination of an ROM and a Monte-Carlo stability analysis are sufficient to predict the important characteristics of a noise filtering scheme. The Monte-Carlo analysis might be expected to give results which are the same regardless of image contents for well designed algorithms. However, the ROM measure must vary according to scene contents and must therefore be interpreted as a "scenario evaluation" [12].

## Acknowledgements

Presented work was part of EC funded project "Performance Characterisation in Computer Vision" IST-1999-14159.

## References

- [1] K. W. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical roc curves. In *Computer Vision and Pattern Recognition '99*, pages 1354–1359, Ft. Collins, CO, USA, 1999.
- [2] P.A. Bromiley, N.A. Thacker, and M.L.J. Scott et al. Bayesian and non-bayesian probabilistic models for medical image analysis. *Image and Vision Computing*, 21(10):851–864, 2003.
- [3] D. L. Collins, C. J. Holmes, and A. C. Evans. Automatic 3d model-based neuro-anatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.
- [4] P. Courtney and N. A. Thacker. *Imaging and Vision Systems: Theory, Assessment and Applications*. NOVA Science Books, 2001.
- [5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via em algorithm. *Journal of the Royal Society*, 39:1–38, 1977.
- [6] R.M. Haralick. Overview: Computer vision performance characterization. In *Proceedings of the ARPA Image Understanding Workshop*, pages 663–665, Monterey, USA, 1994.
- [7] R. K. S. Kwan, A. C. Evans, and G. B. Pike. An extensible mri simulator for post-processing evaluation. In *Proceedings of the 4th International Conference on Visualization in Biomedical Computing, VBC '96*, volume 1131, pages 135–140, 1996.
- [8] Y. G. Leclerc, Q. T. Luong, and P. Fua. Measuring the self-consistency of stereo algorithms. In *Proceedings of the European Conference on Computer Vision*, pages 282–298, Dublin, Ireland, 2000.
- [9] M. Pokrić, N. A. Thacker, and A. Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001*, pages 1293–1294. Springer-Verlag, 2001.
- [10] M. Pokrić, N. A. Thacker, M. L. J. Scott, and A. Jackson. Multi-dimensional medical image segmentation with partial voluming. In *Proceedings of Medical Image Understanding and Analysis*, pages 77–80. BMVA, 2001.
- [11] V. Ramesh and R. M. Haralick. Random perturbation models and performance characterization in computer vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 521–527, Urbana-Champaign, USA, 1992.
- [12] N. A. Thacker. Using quantitative statistics for construction of machine vision systems. In *Proceedings of SPIE: Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision*, volume 4877, pages 1–15. SPIE, 2003.
- [13] D. C. Williamson, N. A. Thacker, S. R. Williams, and M. Pokrić. Partial volume tissue segmentation using grey-level gradient. In *Proceedings of Medical Image Understanding and Analysis*, pages 17–20. BMVA, 2002.