

A Combined Noise Reduction and Partial Volume Estimation Method for Image Quantitation

John Chiverton, *Student Member, IEEE*, Kevin Wells, and Mike Partridge

Abstract—The partial volume effect is a corrupting artifact that affects nuclear imaging data such as PET and SPECT data, manifest as a blurring action on the resultant image data. This artifact is a result of the image acquisition process, where voxels in the PET or SPECT images are typically composed of a mixture of activity concentrations. This prevents accurate localization and quantitation of the target region activity. A further well-known image artifact found in most types of signal and image data is additive noise which is caused by limited photon count statistics for PET or SPECT imaging data. This work presents a novel methodology for statistically combining image noise reduction and partial volume estimation with particular application to low contrast to noise ratio image data, e.g. image data with poor target localization.

Each possible partial volume mixture is modeled as a Gaussian distribution and neighborhood statistical information is also incorporated in the form of the voxel neighborhood intensity mean, which has previously been shown to also be Gaussian distributed. This leads to an analytical solution of the optimal expected mean (thus minimizing the mean square loss), providing an equation that can iteratively and adaptively reduce the noise in each image voxel. Once the noise is reduced a further step that estimates the partial volume mixtures using an adaptive Markov Chain Monte Carlo method is found to improve the partial volume estimates in comparison to existing partial volume estimation techniques without a noise reduction step.

I. INTRODUCTION

THE partial volume effect is a corrupting artifact that affects nuclear imaging data such as PET and SPECT. This artifact is a result of the image acquisition process, where, the relatively large voxels in the PET or SPECT images are typically composed of a mixture of activity concentrations. The partial volume effect manifests as a blurring action on the resultant image data. This prevents accurate localization and quantitation of the activity information being imaged. A further well-known image artifact found in most types of signal and image data is additive noise. The accurate quantitation of activity is therefore affected by the partial volume effect *and* image noise. The result of the acquisition process and the objective of the methodology described in this paper, namely noise reduction is illustrated in figure 1.

Manuscript received November 15, 2006. This work was supported, in part, by the UK Engineering and Physical Sciences Research Council.

J. Chiverton and K. Wells are with the Centre for Vision, Speech and Signal Processing, School of Electronics and Physical Sciences, University of Surrey, GU2 7XH, UK. M. Partridge is with the Joint Department of Physics, Institute of Cancer Research and Royal Marsden NHS Trust, Royal Marsden Hospital, Sutton, SM2 5PT, UK.

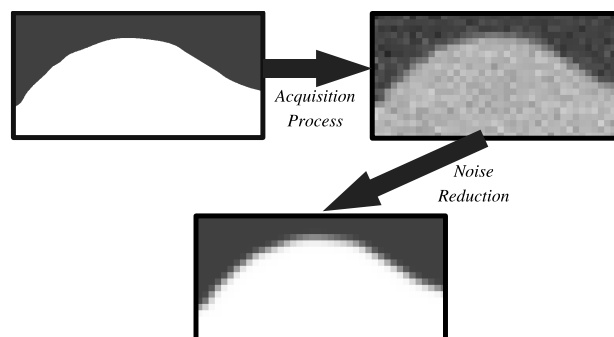


Fig. 1. Illustration of the effect of the image acquisition process and the resulting noise reduction on the acquired noisy imaging data.

This work presents a novel methodology for combining image noise reduction and partial volume estimation with particular application to low contrast to noise ratio image data. In contrast to other approaches, (e.g. [1], [2], [3], [4], [5], [6], [7]), no assumptions are made about the spatial distribution of the tracer or correspondence with prior anatomical information.

Often partial volume correction methods rely on a high-resolution anatomical data set which is segmented into regions that correspond to assumed contiguous regions of functional activity in the functional imaging modality. The segmented anatomical data sets are then convolved with a point spread function of the functional modality, the result of which is a set of mixing templates. These mixing templates in combination with the observed functional image intensities can be used to estimate the amount of physiological activity. However this assumes a correspondence with the prior anatomical information, which may not be valid and/or good quality (segmented) high resolution anatomical imaging data may not be available.

Other techniques have included the use of extra-vascular density images for partial volume correction in cardiac imaging [8], [9]. An extra-vascular density image is created by subtracting a blood pool image from a transmission image. The result of which quantifies tissue activity rather than the activity arising from the tissue *and* vascular components. The methodology presented in [8], [9] is interesting, particularly because of the absence of high resolution anatomical information, but the work is of limited use for non-cardiac imaging applications.

Previous work, such as [10], [11], [12] proposed partial volume analysis models that do not use prior high resolution anatomical information. These models rely on statistical anal-

ysis of the functional image intensities to determine relative quantities of activity concentrations. However, as with any methodology, the relative performance of the techniques are sensitive to the amount of noise in the imaging data, where low contrast to noise ratio imaging data prevents accurate activity concentration determination.

Often data samples may have what is often known as *context*. For spatially located data samples this context is space. Intuitively, if a data sample in a particular location is classified with a particular class label then any data samples located near to that data sample are likely to have a similar class label. This prior knowledge is often useful in improving the classification performance of otherwise ambiguous data samples in imaging data with low contrast to noise ratios.

Spatially derived contextual information is one possible source of information that can improve classification performance. This work demonstrates that the contextual information can be correctly modeled and incorporated into partial volume classification and builds on the work presented in [11], [12], [13]. An analytical approximation for the expected partial volume content of a voxel is derived providing a result that is intuitive and simple to calculate.

II. METHODOLOGY

A PET or SPECT image acquisition process produces imaging data that consists of finite sized voxels, $\{\omega\}$ with noise affected intensities, $g(\omega)$. The imaging data can also be considered to consist of n individual activity concentrations represented by a mean activity concentration vector, $\boldsymbol{\mu} = \{\mu_1 \mu_2 \dots \mu_n\}$ for the entire imaging data. Each voxel can then be considered to be composed of a linear combination of each of these mean activity concentrations, the proportion of which is controlled by elements from a *mixing* vector, $\boldsymbol{\alpha}(\omega) = \{\alpha_1(\omega) \alpha_2(\omega) \dots \alpha_n(\omega)\}$ where each mixing element is limited to values $\alpha_i(\omega) \in [0, 1] \forall i$ where 1 is representative of total voxel content and 0 is zero voxel content. So that each voxel can be described by a single inferred noiseless composite activity concentration, $\mu(\omega) = \alpha_1(\omega) \cdot \mu_1 + \alpha_2(\omega) \cdot \mu_2 + \dots + \alpha_n(\omega) \cdot \mu_n = \boldsymbol{\alpha}(\omega)^T \cdot \boldsymbol{\mu}$. Thus for a particular voxel, ω with intensity $g(\omega)$, the inferred intensity value in the absence of noise is $\mu(\omega)$.

Bayes theorem can be used to determine the probability density for a composite noiseless activity concentration value given a particular noisy intensity value:

$$p(\mu(\omega)|g(\omega)) = \frac{p(g(\omega)|\mu(\omega)) \cdot p(\mu(\omega))}{p(g(\omega))}, \quad (1)$$

where $p(g(\omega))$ is the overall marginal probability density for the measured intensity values, $g(\omega)$ and is given by $p(g(\omega)) = \int_{\mu(\omega)} p(g(\omega)|\mu(\omega)) \cdot p(\mu(\omega)) \cdot d\mu(\omega)$. The prior probability for the noiseless intensity values, $p(\mu(\omega))$ may take a form similar to that found for raw activity concentrations as described previously in e.g. [12], [14]. The intensity likelihood, $p(g(\omega)|\mu(\omega))$ is a description of the range of noisy image intensities that are likely to occur for any particular composite noiseless activity concentration, $\mu(\omega)$.

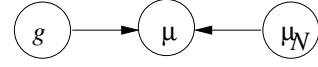


Fig. 2. Graphical illustration of the dependencies between the noisy intensity value, g , the composite noiseless activity concentration μ and the neighboring composite noiseless activity concentration, μ_N .

This formulation is, however, sensitive to the amount of noise found in the imaging data, especially for imaging data with a low contrast to noise ratio. Nevertheless this formulation can be extended to improve estimation accuracy by including additional information such as the composite noiseless activity concentration values of the neighboring voxels, described here with $\mu_N(\omega)$.

Therefore the posterior density can be extended (dropping the point specific notation):

$$p(\mu|g, \mu_N) = \frac{p(g|\mu, \mu_N) \cdot p(\mu, \mu_N)}{p(g, \mu_N)}.$$

Noting the independence of g from μ_N in the numerator due to the mutual dependence on μ , (see also figure 2):

$$\begin{aligned} p(\mu|g, \mu_N) &= \frac{p(g|\mu) \cdot p(\mu|\mu_N) \cdot p(\mu_N)}{p(g|\mu_N) \cdot p(\mu_N)} \\ &= \frac{p(g|\mu) \cdot p(\mu|\mu_N)}{p(g|\mu_N)}. \end{aligned} \quad (2)$$

This result is particularly convenient as will be seen shortly. The locally defined prior mixing distribution, $p(\mu|\mu_N)$ encapsulates the probabilistic information describing the range of probable composite activity concentrations, μ given the neighboring composite activity concentrations, μ_N .

A. Form of the Localized Mixing Prior

The neighborhood information may be succinctly described by the mean of the composite activity concentrations, $\mu(\omega_j)$ that are neighbors to the voxel of interest, ω_i :

$$\mu_N(\omega_i) = \frac{1}{|\mathcal{N}_{\omega_i}|} \sum_{\omega_j \in \mathcal{N}_{\omega_i}} \mu(\omega_j), \quad (3)$$

where $|\mathcal{N}_{\omega_i}|$ is the cardinality of the set of neighbors to point ω_i :

$$\mathcal{N}_{\omega_i} = \{\omega_j | \|\omega_i - \omega_j\| < \mathcal{C}_D\}. \quad (4)$$

i.e. \mathcal{N}_{ω_i} is a set of points located within a certain distance, \mathcal{C}_D , of ω_i .

The squared difference between the composite activity concentration for voxel ω_i and the mean of the neighboring voxels, ω_j results in a Gaussian form of the locally defined prior mixing distribution (omitting ω):

$$p(\mu|\mu_N) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_N^2}} \cdot \exp\left(-\frac{(\mu - \mu_N)^2}{2 \cdot \sigma_N^2}\right), \quad (5)$$

where σ_N^2 is a variance measure which determines the amount of information that is drawn from the neighboring voxels. A large value of this spatial variance measure results in little

information from being drawn from the neighboring voxels. The amount of spatial regularization, i.e. the value of σ_N^2 is controlled here by the magnitude of the locally defined image gradient, in common with [14]. The gradient magnitude is a well defined measure of change, where gradient kernels, such as a Sobel gradient kernel incorporate small amounts of smoothing, thus providing useful estimates of image change in a noisy environment.

B. Analytic Derivation of the Composite Activity Concentration Posterior Distribution

The probability density function of the image intensities, $p(g|\mu)$ may also be described by a Gaussian distribution, which, as will be seen shortly, is a convenient description of the intensities. However, due to the discrete photon counting action in the PET or SPECT image data acquisition process the imaging data are usually said to be governed by a Poisson random process which ultimately results in Poisson distributed image data. However, other non-Poisson type processes are also inherent to the image acquisition process, including the filtering action of the point spread function and steps in the image reconstruction process, which may reduce the validity of a Poisson description of the imaging data. A Gaussian probabilistic description may therefore be inaccurate, particularly for low count data, but for the purposes of this work it is deemed to be a sufficiently accurate description of the imaging data.

Thus the numerator of equation 2 consists of two Gaussian distributions: the intensity likelihood

$$p(g|\mu) = \frac{1}{\sqrt{2\pi}\sigma_\alpha^2} \exp\left(-\frac{(g-\mu)^2}{2\sigma_\alpha^2}\right), \quad (6)$$

and the locally defined composite activity concentration prior, $p(\mu|\mu_N)$ given by equation 5. The variance of the intensity likelihood is given by $\sigma_\alpha^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + \dots + \alpha_n^2\sigma_n^2$, due to the linear combination of the n Gaussian distributed noisy activity concentration intensities present in the imaging data, see e.g. [15], [12].

The denominator of equation 2 may also be described by a combination of these two distributions via a marginalization operation:

$$p(g|\mu_N) = \int_{\mu} p(g|\mu).p(\mu|\mu_N).d\mu. \quad (7)$$

Thus, the posterior distribution given by equation 2 may be described by (after canceling the common Gaussian normalizing factors):

$$p(\mu|g, \mu_N) = \frac{\exp\left(-\frac{1}{2}\left\{\frac{(g-\mu)^2}{\sigma_\alpha^2} + \frac{(\mu-\mu_N)^2}{\sigma_N^2}\right\}\right)}{\int_{\mu} \exp\left(-\frac{1}{2}\left\{\frac{(g-\mu)^2}{\sigma_\alpha^2} + \frac{(\mu-\mu_N)^2}{\sigma_N^2}\right\}\right).d\mu} \quad (8)$$

The common exponential term in the numerator and denominator may be manipulated

$$\begin{aligned} & \exp\left(-\left\{\frac{(g-\mu)^2}{2\sigma_\alpha^2} + \frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right\}\right) \\ &= \exp\left(-\left\{\frac{\sigma_N^2\cdot(g-\mu)^2 + \sigma_\alpha^2\cdot(\mu-\mu_N)^2}{2\sigma_\alpha^2\sigma_N^2}\right\}\right), \end{aligned}$$

multiplying out the bracket terms results in

$$= \exp\left(-\frac{\sigma_N^2\cdot g^2 - 2\cdot g\cdot\sigma_N^2\cdot\mu + \sigma_N^2\cdot\mu^2 + \sigma_\alpha^2\cdot\mu^2 - 2\sigma_\alpha^2\cdot\mu_N\cdot\mu + \sigma_\alpha^2\cdot\mu_N^2}{2\sigma_\alpha^2\sigma_N^2}\right). \quad (9)$$

Grouping the factors of μ results in

$$= \exp\left(-\frac{\mu^2\cdot(\sigma_N^2 + \sigma_\alpha^2) - \mu\cdot 2\cdot(\sigma_\alpha^2\mu_N + g\cdot\sigma_N^2) + \sigma_N^2\cdot g^2 - \sigma_\alpha^2\cdot\mu_N^2}{2\sigma_\alpha^2\sigma_N^2}\right) \quad (10)$$

and completing the square

$$= \exp\left(-\frac{(\sigma_N^2 + \sigma_\alpha^2)(\mu^2 - \mu\cdot 2\cdot\mu_K + \mu_K^2) - \mu_K^2\cdot(\sigma_N^2 + \sigma_\alpha^2) + \sigma_N^2\cdot g^2 - \sigma_\alpha^2\cdot\mu_N^2}{2\sigma_\alpha^2\sigma_N^2}\right); \quad (11)$$

where $\mu_K = (\sigma_\alpha^2\cdot\mu_N + g\cdot\sigma_N^2)/(\sigma_N^2 + \sigma_\alpha^2)$ so that:

$$= \exp\left(-\frac{(\mu - \mu_K)^2}{2\sigma_\alpha^2\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)}\right) \times \exp\left(-\frac{-\mu_K^2\cdot(\sigma_N^2 + \sigma_\alpha^2) + \sigma_N^2\cdot g^2 - \sigma_\alpha^2\cdot\mu_N^2}{2\sigma_\alpha^2\sigma_N^2}\right). \quad (12)$$

Thus

$$\begin{aligned} p(\mu|g, \mu_N) &= \frac{\exp\left(-\frac{(\mu - \mu_K)^2}{2\sigma_\alpha^2\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)}\right)}{\int_{\mu} \exp\left(-\frac{(\mu - \mu_K)^2}{2\sigma_\alpha^2\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)}\right).d\mu} \quad (13) \\ &= \frac{1}{\sqrt{2\pi}\sigma_\alpha^2\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)} \\ &\quad \times \exp\left(-\frac{(\mu - \mu_K)^2}{2\sigma_\alpha^2\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)}\right). \quad (14) \end{aligned}$$

So the posterior density of μ is Gaussian with mean, $\mu_K = (\sigma_\alpha^2\cdot\mu_N + g\cdot\sigma_N^2)/(\sigma_N^2 + \sigma_\alpha^2)$ and variance $\sigma_\alpha^2\cdot\sigma_N^2/(\sigma_N^2 + \sigma_\alpha^2)$. This result is particularly useful as an analytical optimal point estimate of μ is now possible due to the familiar properties of the Gaussian density. The expected value of μ which is equal to μ_K for this Gaussian density minimizes the squared error loss, $(\mu - \mu_{\text{est}})^2$. Therefore,

$$\mu_{\text{est}} = \mathbf{E}[\mu|g, \mu_N] = \frac{\sigma_\alpha^2\cdot\mu_N + g\cdot\sigma_N^2}{\sigma_N^2 + \sigma_\alpha^2}. \quad (15)$$

So that an optimal point estimate for the composite activity concentration, μ , is given by a combination of the mean of its neighboring values, $\mu_{\mathcal{N}}$, and the intensity of the voxel in question, g . The weighting of these values are the variances of the mixture, σ_{α}^2 and the variance in the neighborhood, $\sigma_{\mathcal{N}}^2$. If there is a large variance for the neighboring values, (signifying that the voxel is close to a class boundary) then more information would be inferred from the voxel's intensity. Similarly if there is large mixture variance for a particular voxel (in relation to the neighborhood variance) then more information would be inferred from the neighboring voxel values.

This result is intuitively appealing, simple to implement and quick to calculate.

C. Activity Inference with Optimal Composite Estimates

The expected composite value in equation 15 provides a route to inference of the estimated true intensity, but not the more interesting underlying mixture combinations, α . Nevertheless, the preceding framework could be used in isolation for a noise reduction strategy for image data. Therefore a further stage is required using the same statistical framework to infer the actual mixture combinations, given the optimal composite value, (letting $\mu_{\text{est}} = \mathbf{E}[\mu|g, \mu_{\mathcal{N}}]$):

$$p(\alpha|\mu_{\text{est}}) = \frac{p(\mu_{\text{est}}|\alpha) \cdot p(\alpha)}{p(\mu_{\text{est}})}. \quad (16)$$

Comparison can be made between equation 16 and equation 2. Both equations contain a global mixture prior and a likelihood governed by a Gaussian density, although the parameters for $p(\mu_{\text{est}}|\alpha)$ have to be calculated in a different manner.

Due to the normal distribution of the posterior of the composite value in equation 14, the optimal composite value is also normally distributed. Therefore the optimal composite value estimate is distributed according to

$$p(\mu_{\text{est}}|\alpha) = \frac{1}{\sqrt{2\pi} \cdot \sigma_K} \cdot \exp\left(-\frac{(\mu_{\text{est}} - \mu_{\alpha})^2}{2 \cdot \sigma_K^2}\right), \quad (17)$$

where it is assumed $\mu_{\alpha} = \alpha^T \cdot \boldsymbol{\mu}$ and from equation 14,

$$\sigma_K^2 = \frac{\{\sum \alpha_i^2 \cdot \sigma_i^2\} \cdot \sigma_{\mathcal{N}}^2}{\sum \alpha_i^2 \cdot \sigma_i^2 + \sigma_{\mathcal{N}}^2}. \quad (18)$$

The prior mixing distribution, $p(\alpha)$ can be given the same form as for work found in [12].

D. Implementation

1) *Posterior simulation:* The activity inference with optimal composite value estimates results in a difficult to realize posterior distribution (equation 16). Simulation of equation 16 is, however, possible via Markov Chain Monte Carlo where a Markov Chain of samples is generated with samples in the Markov Chain converging to samples from a stationary distribution, in this case equation 16. Markov Chain Monte Carlo simulation is therefore used here to generate samples

from this posterior distribution. A type of Markov Chain Monte Carlo approach known as the Independent Metropolis-Hastings is particularly useful as the prior distribution, $p(\alpha)$ may be used as the proposal distribution from which proposed random samples are drawn as possible candidates of the actual posterior distribution.

A global two class mixture prior probability was found in [14] to be well modeled by

$$p_{Gauss}(\alpha_{j,k}) = C \cdot \exp\left(\text{erf}^{-1}(2 \cdot \alpha_{j,k} - 1)^2\right), \quad (19)$$

where C is a normalizing constant and the mixing variable completely specifies the mixing for the two classes, j and k because $\alpha_{j,k} = \alpha_j = 1 - \alpha_k$. Equation 19 can not be easily used as a sampling density directly, therefore an alternative but similar density is used, known as the Beta density (for two classes, $n = 2$) or for $n > 2$ a Dirichlet density which is the Beta density generalized for $n > 2$. This is because the Beta density for particular parameter values can provide a reasonable fit to equation 19, as has been previously found by other authors for mixture distributions in non-medical images [16], [17], [18]. But more importantly, computationally efficient methods exist that enable sampling from the Beta and Dirichlet densities. Thus the global mixture prior, $p(\alpha)$ is given by a Dirichlet density with the form

$$p(\alpha) = \frac{\Gamma(\lambda_0)}{\prod_{i=1}^n \Gamma(\lambda_i)} \cdot \prod_{i=1}^n \alpha_i^{\lambda_i - 1} \quad (20)$$

where $\lambda_i \forall i$ are parameters of the Dirichlet density, values of which were selected based on the goodness of fit with equation 19.

2) *Iterative classification:* Initial values of $\alpha(\omega)$ have to be estimated for all ω , therefore an initial step, prior to noise reduction, is to provide an initial classification of $\alpha(\omega)$. The initial classification is determined by calculating the expectation of:

$$p(\alpha(\omega)|g(\omega)) = \frac{p(g(\omega)|\alpha(\omega)) \cdot p(\alpha(\omega))}{p(g(\omega))} \quad (21)$$

for all ω similar to the technique described in [12]. A number of iterations then follow, iteratively reducing the noise by calculating $\mu_{\text{est}}(\omega)$ using equation 15. After every noise reduction step updates of the mixture content for each voxel, $\alpha(\omega)$ is required. This is done via simulation of the mixture posterior distribution, $p(\alpha(\omega)|\mu_{\text{est}}(\omega))$, (equation 17) using the Independent Metropolis-Hastings algorithm, in conjunction with the calculation of the expectation of this simulated posterior distribution.

The noise reduction and activity inference steps are illustrated in figure 3.

III. CLASSIFICATION OF SIMULATED PARTIAL VOLUME DATA

A. Methods

Simulated partial volume data, illustrated in figure 4 consisting of a series of concentric spheroids were generated by the following:

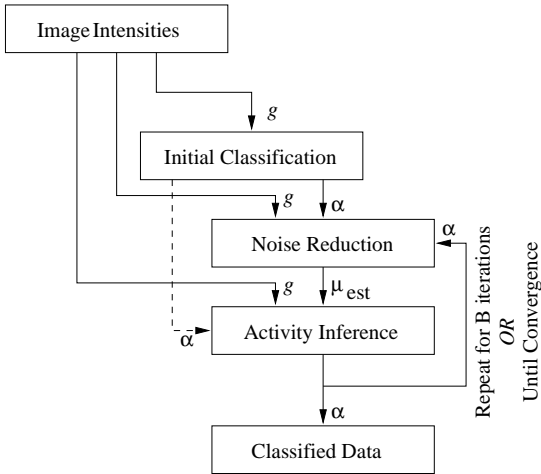


Fig. 3. Illustration of steps involved in the noise reduction and classification (activity inference) process. The iterative process of re-applying the noise reduction prior to activity inference is undertaken for B iterations or until some convergence criteria is met, such as very few changes in estimated activity levels. Dotted line between *Initial Classification* and *Activity Inference* steps signifies input for initial activity inference stage but not for any further iterations.

- 1) Generate high resolution data volume consisting of concentric spheroids with alternating composite value intensities;
- 2) Generate Gaussian distributed noise samples;
- 3) Add noise samples to voxels in high resolution data volume;
- 4) Convolve noisy high resolution data volume with simulated point spread function - simulating partial volume voxels;
- 5) Sub-sample high resolution data volume to produce a low resolution data volume with voxel dimensions equal to the full-width at half-maximum of the simulated point spread function.

The result of these steps is a data volume consisting of voxels that can be considered to contain pure and partial volume voxels, where voxels located on the boundary of each concentric spheroid will be partial volume voxels. An exemplar data slice from a simulated data volume is illustrated in figure 4.

A series of these simulations were performed with different noise parameters to simulate contrast to noise ratios of 2, 3, 4, 5, 6, 7, 8, 9, 10 and 20. These are plausible contrast to noise ratios that may be found in real imaging data.

The principal advantage of simulating the partial volume effect in this way is due to the availability of *ground truth* that can be generated by omitting the noise addition step and any other acquisition and or reconstruction related artifact. The result of which is a data volume with voxels that can be considered to be representative of the composite noiseless activity concentration parameter, μ , described by the prior probability density of equation 5. Similarly, the noisy data volume is described by the intensity likelihood of equation 6.

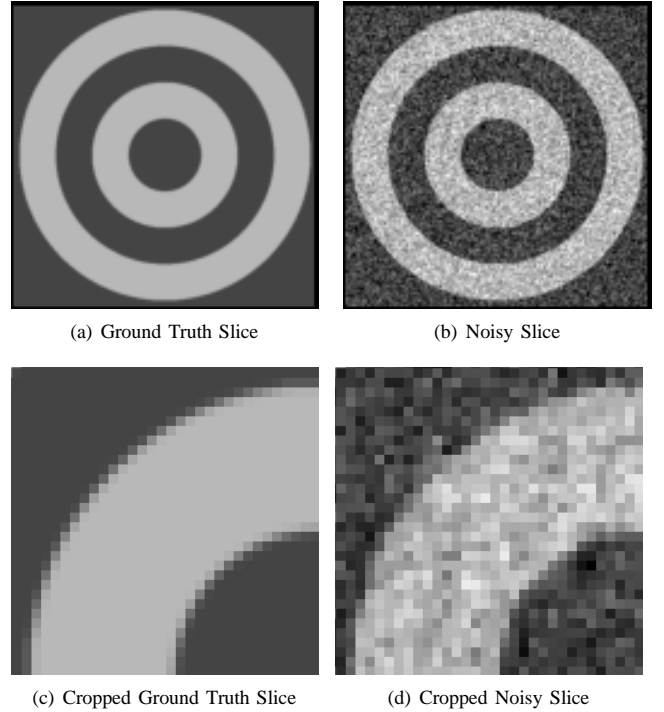


Fig. 4. Exemplar simulated partial volume data slices and cropped slices from a ground truth data volume (a) and (c) and from the corresponding noisy data volume (b) and (d) respectively. (c) and (d) illustrate the effect of the convolution of the simulated Gaussian point spread function resulting in simulated partial volume voxels. Notice the difficulty in which even a human observer might have in identifying partial volume voxels in the presence of noise in (d) when compared with (c). The contrast to noise ratio for this simulated partial volume data is 6.

The simulated partial volume data was then classified with the noise reduction step, as described in this work and also without the noise reduction step, similar to techniques such as [12], [17], [19]. The results of these classifications were then compared with the ground truth data using a root mean square error measure between the output of the classifiers and the ground truth data:

$$E_{\text{RMS}} = \frac{1}{\|\mu_1 - \mu_2\|} \sqrt{\frac{1}{N} \sum_{\forall \omega} (\mu_{\text{est}}(\omega) - \mu_{\alpha}(\omega))^2}, \quad (22)$$

where N is the number of voxels in the imaging data and μ_1 and μ_2 are the ground truth mean parameters of the imaging data, known *a priori* from the partial volume data simulation process.

B. Results

The exceptional benefit of including the noise reduction step in the classification can be seen in the result of classifying the lowest contrast simulated partial volume data in figure 5.

The classification performance difference between the classifier that utilized noise reduction and the classifier without noise reduction becomes less prominent for higher contrast to

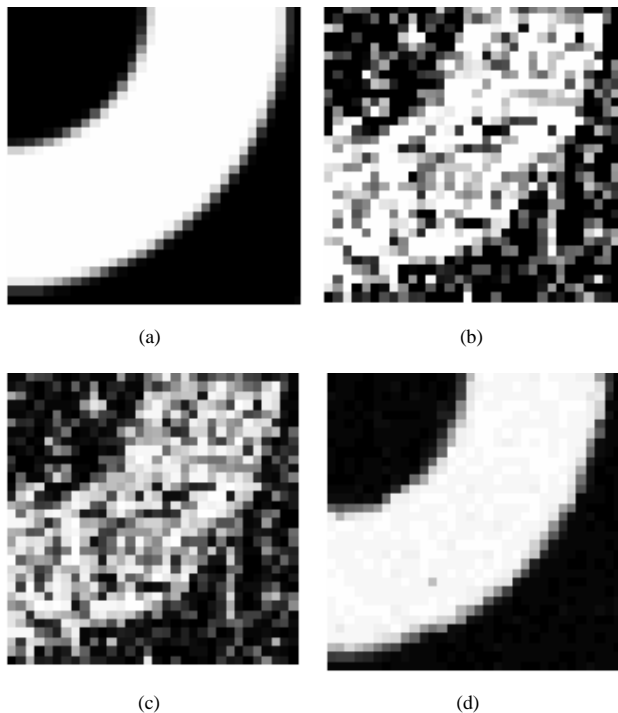


Fig. 5. Illustrative enlarged exemplar image slices for (a) ideal classified output i.e. ground truth; (b) CNR=2 unclassified noisy simulated partial volume data (means=980,1020 and standard deviations=20); (c) partial volume estimates classified without noise reduction; and (d) partial volume estimates classified with noise reduction. For display purposes window center=1000 and width=40 used in each image.

noise ratios, as illustrated with the voxel root mean square error classification results in figure 6.

These results help to illustrate that controlled noise reduction prior to classification can help to improve the classification rate. This simulated partial volume data, as with any type of simulation are simplifications of actual imaging problems. Therefore the noise reduction is now applied to real PET imaging data.

IV. CLASSIFICATION OF PET IMAGING DATA

A. Methods

A Phillips Gemini PET/CT scanner was used to scan an imaging phantom consisting of a series of varied diameter inserts with internal diameters of 4.09cm, 2.99cm, 2.02cm, 1.00cm and 0.465cm. The phantom was filled with 220MBq F-18 and the data acquired over a 15 minute acquisition time. The Row-Action Maximum Likelihood iterative reconstruction Algorithm [20], was used to reconstruct the imaging data in conjunction with CT based attenuation correction, together with corrections for randoms and scatter.

One iteration of the noise reduction and activity inference classification scheme was applied.

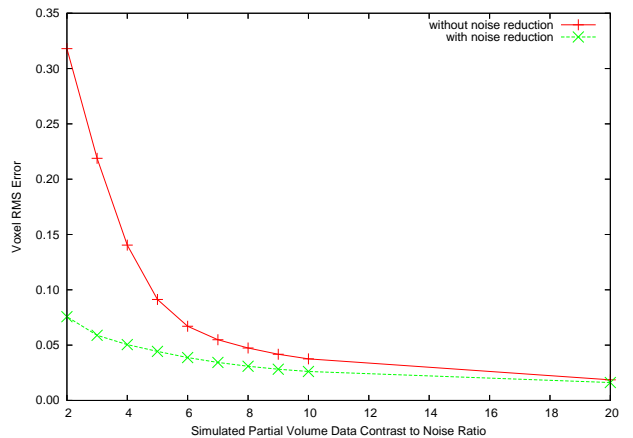


Fig. 6. Voxel root mean square error classification results for the noise reduced and non-noise reduced classifier outputs.

B. Results

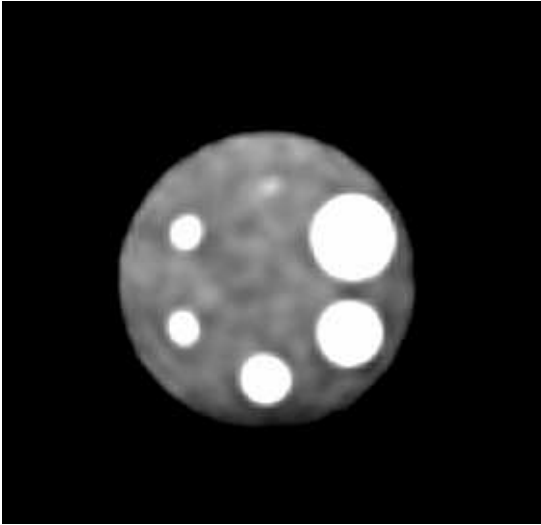
The results of applying one iteration of the noise reduction scheme can be seen with an exemplar image slice from the noisy PET imaging data compared with an exemplar slice from the noise reduced PET imaging data in figure 7. The warm backgrounds in this figure (7) clearly illustrates a reduction in the noise. Similarly, the histograms of the data sets illustrated in figure 7 can be seen in figure 8. The warm background and hot inserts have the characteristic peaks, but the peaks are accentuated for the noise reduced data histogram. This helps to illustrate a reduction in the variability of image intensities for these image regions. The partial volume components between the peaks are still present, but the quantity of a particular activity concentration for any voxel is more easily identified due to the reduction in the amount of noise.

V. DISCUSSION

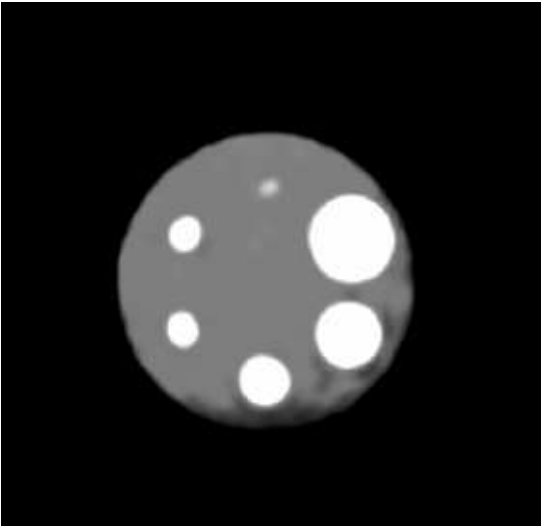
A noise reduction technique that is consistent with an underlying model of the partial volume effect has been described, implemented and tested on simulated and real PET imaging data. The work is a development of the work described in [11], [12], [13] and the results of classifying the simulated partial volume data in conjunction with noise reduction produces improved classification performance of the partial volume data. The subsequent partial volume classification is, however, dependent on simulation rather than on an analytic result (in contrast to the convenient analytic noise reduction step). Simulation is computationally expensive, however, if subsets of the imaging data are used (such as with region of interest analysis), then this computational expense will be small.

The benefits of noise reduction are most apparent for low contrast to noise ratio imaging data (contrast to noise ratio < 10). This is a logical conclusion as one might expect a noise reduction step to be most effective for data with greater amounts of noise.

Further work might include investigation of the necessity for the iterative nature of the noise reduction and partial volume



(a) Noisy PET Data



(b) De-noised PET Data

Fig. 7. Exemplar image slices from (a) PET phantom imaging data and (b) noise reduced PET phantom imaging data. The noise reduction appears to reduce the heterogeneity of the imaging data.

estimation steps. This iterative schema increases the computation time, especially when one considers that a simulation has to be undertaken for every voxel in every iteration. This is necessary as, particularly for low contrast to noise ratio data, the initial partial volume estimates, without noise reduction, are not accurate; see for example the non-noise reduction results presented in figure 6.

Other noise reduction techniques may also be of interest, such as simple linear filtering or perhaps rank order filters (see e.g. [21] and the references therein). These generic image processing techniques, however, may alter the image statistics in either a non-optimal or non-predictable manner. The application of the noise reduction step, (equation 15), is in

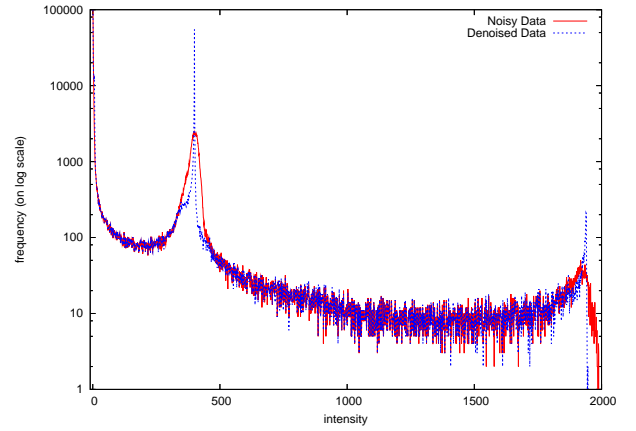


Fig. 8. Histograms for image data illustrated in figure 7. The effect of the noise reduction process appears to accentuate the peaks of the pure components whilst retaining the overall characteristics of the imaging data.

effect a type of filter, taking a proportion of the mean of the neighboring composite noiseless values and the actual voxel intensity value, in a controlled and predictable manner. This allows for the iterative nature of the noise reduction and partial volume estimation steps to remain theoretically consistent and to converge to a stable result after a number of iterations.

The formulation described in this work is similar to a continuous Markov random field for partial volume data. Some aspects of the formulations described here have also been used for partial volume data, albeit for magnetic resonance imaging data [13], [22], [23]. These formulations incorporate a Markov random field, conditioned on the underlying mixing components, described here by the mixture vector, α . This can be compared to the approach taken here which estimates optimal values of the underlying composite noiseless intensities, μ which is a function of α . This is considered here as a noise reduction step, which takes the mean of the neighboring composite values and it can be shown to be equivalent to a homogeneous Markov random field, where the action of $p(\mu|\mu_{\mathcal{N}})$ performs an adaptive spatial regularizing action in combination with the intensity likelihood, $p(g|\mu)$ (see equation 2). The principal advantage of the work described here is that the posterior distribution, $p(\mu|g, \mu_{\mathcal{N}})$ is fully available along with the expectation. The availability of the posterior distribution therefore provides a computationally efficient route to obtaining the optimal composite value, in contrast to alternative approaches that have to perform simulation, [13], [23] or arbitrary function minimization, [22].

VI. CONCLUSIONS

The work described here has provided a noise reduction scheme, specifically created to be consistent with an underlying model of the partial volume effect that does not require any high resolution imaging data. The solution to this noise reduction scheme is available analytically. The result of iteratively applying the noise reduction step in combination with a partial

volume estimation step is improved classification of simulated partial volume imaging data in comparison to partial volume estimation in the absence of the noise reduction step. The noise reduction step has also been seen to reduce the variability of the image intensities for real PET imaging data.

ACKNOWLEDGMENT

The authors would like to thank the Engineering and Physical Sciences Research Council, UK for partial funding of this work. The authors would also like to thank Jia Beier for assisting with the acquisition of the PET imaging phantom data.

REFERENCES

- [1] P. Herrero, J. Markham, D. W. Myears, C. J. Weinheimer, and S. R. Bergman, "Measurement of myocardial blood flow with positron emission tomography: Correction for count spillover and partial volume effects," *Math. Comput. Mod.*, vol. 11, p. 807812, 1988.
- [2] H. W. Mueller-Gartner, J. M. Links, J. L. Prince, R. N. Bryan, E. McVeigh, J. P. Leal, C. Davatzikos, and J. J. Frost, "Measurement of radiotracer concentration in brain gray matter using PET: MRI-based correction for partial volume effects," *J. Cereb. Blood Flow Metab.*, vol. 12, p. 571581, 1992.
- [3] O. G. Rousset, Y. Ma, G. C. Leger, A. H. Gjedde, and A. C. Evans, "Correction for partial volume effects in PET using MRI-based 3-D simulations of individual human brain metabolism," in *Quantification of Brain Function: Tracer Kinetics and Image Analysis in Brain PET*, K. Uemura, N. A. Lassen, T. Jones, and I. Kanno, Eds. New York: Elsevier, 1993, pp. 113–123.
- [4] O. G. Rousset, Y. Ma, S. Marengo, D. F. Wong, and A. C. Evans, "In vivo correction method for partial volume effects in positron emission tomography, accuracy and precision," in *Quantification of Brain Function using PET*, R. Meyers, V. Cunningham, D. Bailey, and T. Jones, Eds., 1996, pp. 158–165.
- [5] M. Frouin, C. Comtat, A. Reilhac, and M. Gregoire, "Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness," *J. Nucl. Med.*, vol. 43, pp. 1715–1726, 2002.
- [6] J. Aston, V. Cunningham, M. Asselin, A. Hammers, A. Evans, and R. Gunn, "Positron emission tomography partial volume correction: Estimation and algorithms," *J. Cereb. Blood Flow Metab.*, vol. 22, pp. 1019–1034, 2002.
- [7] M. Quarantelli, K. Berkouk, A. Prinster, B. Landeau, C. Svarer, L. Balkay, B. Alfano, A. Brunetti, J. Baron, and M. Salvatore, "Integrated software for the analysis of brain PET SPECT studies with partial-volume effect correction," *J. Nucl. Med.*, vol. 45, 2004.
- [8] H. Iida, C. Rhodes, R. deSilva, Y. Yamamoto, L. Araujo, A. Maseri, and T. Jones, "Myocardial tissue fraction-correction for partial volume effects and measure of tissue viability," *J. Nucl. Med.*, vol. 32, pp. 2169–2175, 1991.
- [9] R. Wassenaar, R. Beanlands, and R. deKemp, "Phantom studies investigating extravascular density for partial volume correction of 3-D PET FDG studies," *IEEE Trans. Med. Imag.*, vol. 51, pp. 68–71, 2004.
- [10] R. Muzic, C. Chen, and A. Nelson, "A method to correct for scatter, spillover, and partial volume effects in region of interest analysis in PET," *IEEE Trans. Med. Imag.*, vol. 17, pp. 202–213, 1998.
- [11] J. Chiverton and K. Wells, "A partial volume 3-D gradient magnitude model," in *Conf. Proc. Med. Image Und. Anal.* BMVA, Imperial College, London, UK, 2004, pp. 49–52.
- [12] K. Wells, J. Chiverton, M. Partridge, M. Barry, H. Kadhem, and R. Ott, "Quantifying the partial volume effect in PET using Benford's Law," *Submitted to IEEE Trans. Nucl. Sci.*
- [13] J. Chiverton and K. Wells, "A locally adaptive gradient controlled spatial regularization partial volume classifier," in *Proc. British Chapter of ISMRM*, University of Surrey, Guildford, UK, 2006.
- [14] —, "Mixture effects in FIR low-pass filtered signals," *IEEE Sig. Proc. Letters*, vol. 13, pp. 369–372, 2006.
- [15] H. Stark and J. Woods, *Probability and Random Processes with Applications to Signal Processing Third Edition*. USA: Prentice-Hall, 2002.
- [16] E. Hayman and J. Eklundh, "Statistical background subtraction for a mobile observer," in *Proc. Ninth IEEE ICCV 2003*, vol. 1, 2003, pp. 67–74.
- [17] A. Kitamoto and M. Takagi, "The area proportion distribution-relationship with the internal structure of mixels and its application to image classification," *Sys. Comps. Japan*, vol. 31, pp. 57–76, 2000.
- [18] Y. Wexler, A. Fitzgibbon, and A. Zisserman, "Bayesian estimation of layers from multiple images," in *Proc. ECCV, Copenhagen, Denmark, 2002*, 2002, pp. 487–501.
- [19] E. Vokurka, A. Herwadkar, and N. Thacker, "Using Bayesian tissue classification to improve the accuracy of Vestibular Schwannoma volume and growth measurement," *Amer. J. Neurorad.*, vol. 23, pp. 459–467, 2002.
- [20] J. Browne and A. De-Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, pp. 687–699, 1996.
- [21] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision Second Edition*. PWS, 1999.
- [22] H. Choi, D. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images—a mixel model," *IEEE Trans. Med. Imag.*, vol. 10, pp. 395–407, 1991.
- [23] M. Woolrich, T. Behrens, C. Beckmann, and S. Smith, "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data," *IEEE Trans. Med. Imag.*, vol. 24, pp. 1–11, 2005.