Graph-Based Mumford–Shah Segmentation of Dynamic PET With Application to Input Function Estimation

Brian J. Parker and (David) Dagan Feng, *Fellow, IEEE*

***Abstract*—A graph-theoretic three-dimensional (3-D) segmentation algorithm based on Mumford–Shah energy minimization is applied to the segmentation of brain ^{18}F-fluoro-deoxyglucose (FDG) dynamic positron emission tomography data for the automated extraction of tissues with distinct time activity curves (TACs), and, in particular, extraction of the internal carotid arteries and venous sinuses for the noninvasive estimation of the input arterial TAC. Preprocessing by principal component analysis (PCA) and a Mahalanobis distance metric provide segmentation based on distinct TAC shape rather than simply activity levels. Evaluations on simulation and clinical FDG brain positron emission tomography (PET) studies demonstrate that differing tissue types can be accurately demarcated with superior performance to k-means clustering approaches, and, in particular, the internal carotids and venous sinuses can be robustly segmented in clinical brain dynamic PET datasets, allowing for the fully automatic noninvasive estimation of the arterial input curve.**

***Index Terms*—Brain modeling, covariance analysis, graph theory, image segmentation, Mumford–Shah, pattern recognition, positron emission tomography.**

I. Introduction

DYNAMIC functional imaging techniques such as dynamic positron emission tomography (PET) are used for the in vivo quantitative measurement of time-varying biochemical processes. In such studies, differing tissue types show distinctive activity levels over time after injection of radio-tracer, due to differing transfer rates between the various physiological compartments within the tissues, and so segmenting these PET data sets into their component tissue types based on the distinctive shapes of their time activity curves (TACs) can provide additional diagnostic information.

In particular, we aim to extract brain vasculature automatically from brain PET data sets. A principal application for this is image-derived input function estimation without requiring blood sampling; the input blood activity curve is needed for accurately quantifying PET studies by the use of compartmental models. Previous work has demonstrated the feasibility of input curve estimation from regions-of-interest (ROIs) placed over the internal carotids [1]–[3], but these required careful manual placement of the ROIs.

Manuscript received November 15, 2003; revised August 13, 2004. This work was supported in part by ARC Grants.

B. J. Parker is with the School of Information Technologies, Faculty of Science, University of Sydney, Australia (e-mail: brianp@cs.usyd.edu.au).

D. Feng is with the School of Information Technologies, University of Sydney, Australia, and also with the Electronic and Information Engineering Department, Hong Kong Polytechnic University (e-mail: feng@it.usyd.edu.au).

Digital Object Identifier 10.1109/TNS.2004.843133

PET is a challenging modality to segment due to its high noise and low resolution. The large number of image segmentation schemes described in the literature can be divided into the broad classes of: edge detection-based, region-similarity (e.g., k-means, seeded region-growing, split and merge), active surface (e.g., "snakes"), and stochastic (e.g., Markov Random Field) schemes [4]. Numerous clustering and segmentation algorithms have been previously applied to dynamic PET for the general problem of segmenting based on TACs, the most studied being variations of k-means or other clustering algorithms, primarily for improving quantification [5]–[9]. These schemes can suffer from poor demarcation in high-noise environments such as PET due to the lack of an explicit region compactness criterion in the segmentation (or indeed even local adjacency information for k-means clustering approaches), and can have seeding problems that can cause them to miss small features such as vasculature.

Earlier, ad hoc, algorithms for data and image segmentation did not impose an explicit ordering on the quality of the segmentations. By contrast, in more recent energy-minimization approaches, an explicit energy, or cost, is defined by a functional over the set of possible segmentations and the segmentation problem is treated as an optimization problem based on this objective function. The Mumford–Shah functional [10] (described in the sequel) is a cost that combines both a region-based error term and an edge-based compactness term, and energy minimization using this functional can be considered as a most general segmentation approach, with most other schemes such as edge-based, seeded region-growing, and the various active surface algorithms being viewed as specializations within this framework, as demonstrated by Morel and Solimini [11] and Shah [12]. In this paper, we apply and evaluate the Mumford–Shah segmentation approach to the problem of PET functional tissue segmentation.

In Section II, we describe a graph-based Mumford–Shah energy-minimizing segmentation algorithm [13]. This algorithm has several advantages over existing segmentation approaches when applied to dynamic PET.

- Unseeded—no initial seed selection is necessary and so the segmentation performance is fully deterministic.
- Fully automatic with no algorithm parameters except the segmentation scale.
- Provides a multiscale segmentation, which is useful for exploratory data analysis, and varying applications, e.g., a segmentation at a coarse scale can be used for figure/ground separation, and a fine segmentation can

be used for the extraction of small features such as vasculature.
- Regularization by region compactness lowers the sensitivity to noise.

Another important aspect of the problem is the use of a distance metric in the segmentation algorithm that will effectively demarcate differing tissue types. The aim is to distinguish primarily between TACs that have differing overall shape, as opposed to segmentation based on differing activity level. Most commonly, some variant of Euclidean distance has been used in clustering approaches, e.g., [6], but this metric is less sensitive in discriminating between differing TAC shapes and, hence, differing tissue types, e.g., it can be confounded by areas of gray matter with varying relative activities. In [9], the cross correlation between TACs was used with a clustering approach and found to be more effective than Euclidean distance for segmentation on TAC shape. The approach investigated in this paper is to first preprocess the data set by a noise-normalized principal component analysis (PCA) and then use a Mahalanobis metric on the PCA channels.

In Section III-A, we empirically evaluate the hypothesis that the region compactness term of the Mumford–Shah formulation, in combination with a PCA/Mahalanobis distance metric, leads to more accurate functional segmentation by comparison with a k-means clustering algorithm, which lacks such a region compactness criterion. In Section III-B, we experimentally compare our approach to k-means for segmenting of a distinct tissue type (epileptic focus), and in Section III, we experimentally demonstrate that the proposed algorithm can correctly localize and demarcate the vasculature in clinical brain PET studies.

In Section IV, the results are presented and discussed.

II. Method

An overview of the approach used is the following.

Step 1: A PCA of the original dynamic PET data is used to reduce the data set to the significant information.

Step 2: The PCA channels are segmented using an energy-minimization segmentation algorithm, with an eigenvalue-weighted distance used as a similarity metric.

Step 3: The vasculature is detected by pattern recognition based on the moments of the segment TACs.

Step 4: The input curve is estimated from the maximum and mean TACs over the internal carotid, other vasculature, and surrounding tissue ROIs.

We now describe these steps in detail.

Step 1: PCA: Dynamic PET studies produce a series of three-dimensional (3-D) images over time (frames) with the shape of the time-series of activity data for each voxel (the TAC) depending on the tissue type of the voxel. Hence, dynamic PET data is a vector-valued function, with a time series of activity data at each three-dimensional spatial voxel. The three-dimensional segmentation algorithm to be described in Step 2 can be applied to arbitrary vector lengths and so could be applied directly to the original dataset, but the early frames in dynamic PET are dominated by noise and each frame is highly correlated with the surrounding frames. Therefore, we use a PCA to reduce the number of vector components of the data before segmentation from the n original frames to m PCA channels. (e.g., from $n = 22$ original frames down to $m = 5$ PCA channels in the experiments).

PCA uses a transformation matrix formed from the eigenvectors of the covariance matrix of the vector data to decorrelate the data. As described in [14], to help prevent the noise in each frame from being confused with signal by the PCA, before the PCA the overall noise variance in each frame needs to be standardized. Given the sinogram data, expressed in units of counts/sec, for the approximately Poisson noise of PET data this noise-normalization can be performed by dividing each frame by its standard deviation, $\sqrt{N}/\Delta T$, where N is the total count of the detections per frame, and ΔT is the frame duration. This noise normalization scheme is compared empirically with others in [15].

The PCA can be applied either to the reconstructed images, or, optionally, directly to the sinograms, before image reconstruction [16]. Applying the PCA directly to the sinograms can provide some computational advantage in this application as it requires a full image reconstruction on only the small number of PCA channels, and the later estimation of the input curves can then be done with a reconstruction of the original frames limited to the ROIs over the vasculature and surrounding tissue. By contrast, performing the PCA on the reconstructed images requires a large number (e.g., 22) of full image reconstructions which can be an issue when using computationally intensive iterative reconstruction schemes such as EM. In the clinical experiments, the PCA is applied in the sinogram-domain. (Note: EM requires nonnegative values, so each PCA channel is shifted to have zero minimum before reconstruction).

Step 2: Segment PCA Data: Optimizing the Mumford–Shah energy exactly is known to be intractible, and so heuristic solutions are used to generate a good approximation to the global minimum—when discretized over a grid of voxels, this can be treated as a combinatorial optimization problem. Koepfler *et al.* [17] presented an efficient algorithm for the solution of the Mumford–Shah functional using a pure region-merging approach using only local merge operations, where neighbors of minimal energy difference are merged. Unfortunately, this algorithm uses an iterative approach that requires the user to explicitly provide as a parameter an increasing lambda schedule used to determine a multiscale segmentation. The segmentation results are sensitive to this schedule of parameters, and no technique to automate this is described. In Redding *et al.* [18], a so-called "full-lambda schedule" variant of [17] is described which avoids the need to provide an explicit schedule of lambda parameters and instead needs only a single stopping parameter. In [13], we described a graph-theoretic formulation of this segmentation scheme of Redding *et al.* applied to large MRI datasets. In this study, we apply this graph-based formulation to the Poisson-distributed high-noise environment of the PET modality.

A. Mumford–Shah Segmentation Theory Background

Let the PCA channels of the dataset from Step 1 be treated as a 3-D vector-valued function g with m PCA channel components, defined over a 3-D grid of voxels Ω. Then the segmentation is

obtained by minimizing the piecewise constant Mumford–Shah cost functional

$$M(K) = \int_{\Omega} \|u - g\|^2 + \lambda \cdot l(K) \tag{1}$$

where K is the set of boundary faces defining the 3-D segmentation, with total surface area $l(K)$; u is the mean over each segment induced by K; λ is a regularization parameter; and the distance metric $\|\cdot\|^2$ is a Mahalanobis distance, discussed further in the next section. Intuitively, the first term limits the error in dataset approximation resulting from segmentation, and λ controls the fineness of the segmentation by limiting the total boundary surface area of the segmentation. Thus, by varying λ a multiscale segmentation of the correct depth for a given application can be achieved. Also, this regularization, by favoring compact surface descriptions, ensures that this approach works well with noisy datasets such as PET.

To find an approximate solution for the minimum energy segmentation, the optimization heuristic strategy of [18]—the "full-λ-schedule" segmentation heuristic—is used. Assuming a region-merging strategy, the change in cost when two adjacent segment pairs (O_i, O_j) are merged is

$$M\left(K \setminus \partial(O_i, O_j)\right) - M(K) = \frac{|O_i| \cdot |O_j|}{|O_i| + |O_j|} \cdot \|u_i - u_j\|^2 - \lambda \cdot \ell\left(\partial(O_i, O_j)\right) \tag{2}$$

where $|O_i|$ is the volume of segment O_i; u_i is the mean of g over segment O_i; $l(\partial(O_i, O_j))$ is the area of the shared boundary between the segments $\mathrm{O_i}$ and $\mathrm{O_j}$; and $\|\overline{u}_i - \overline{u}_j\|^2$ is the similarity metric between the mean values of the segments.

In Koepfler *et al.* [17], a greedy heuristic is used whereby for each preselected lambda value, adjacent regions giving a locally maximal decrease in energy by (2) are iteratively merged until all negative energy merges are completed. In [18], it is noted that, from (2), the decision to merge O_i and O_j occurs when λ has the value

$$t_{i,j} = \frac{\frac{|O_i||O_j|}{|O_i|+|O_j|}\|\overline{u}_i - \overline{u}_j\|^2}{l\left(\partial(O_i, O_j)\right)} \tag{3}$$

and so the full-λ-schedule segmentation algorithm iteratively performs a greedy search over all adjacent segment pairs for the pair with the smallest value of $t_{i,j}$, and then merges the corresponding segments O_i and O_j. As this process iterates, the current $t_{i,j}$ will trend upwards, and the monotonically increasing upper bound on this sequence of $t_{i,j}$ values defines a schedule of lambda values so that a progression of segmentations is produced, from an initially fine segmentation corresponding to a Mumford–Shah solution with a small λ parameter, to progressively coarser segmentations.

B. Graph-Based Segmentation Algorithm Design

By formulating this approach as a graph-based problem, the list-based data structures described in [18] are simplified and improved—in particular, an adjacency list graph representation is used, with a Fibonacci heap to effect the ordering of the edges. This improved graph theoretic formulation is described here.

Let the dataset be represented by a simple, undirected, weighted graph $G(V, E)$, using the dual representation whereby vertices V represent segments, and the edges E represent the boundaries between segments. Each vertex V_i maintains the associated segment volume and mean of g over the segment, $(\overline{u}_i, |V_i|)$, and the edge weights are set to the surface area between the corresponding segments, $l(\partial(V_i, V_j))$. The following algorithm can then be used to approximate the minimum for functional (1).

1) Populate graph G with an initial trivial segmentation of the PET data with each voxel as a separate segment (vertex), with the six direct neighbors separated by boundaries with weights initialized to the area of the corresponding voxel face (for anisotropically sampled data such as that produced by typical PET scanners this area will differ between axes). (Also note that, optionally, a subpixel accurate segmentation can be obtained by first using a high-quality interpolation such as tri-cubic to up-sample the original data to $2\times$ or$4\times$ the original resolution.)
2) Perform the region merging step as an edge contraction on the graph G by removing the segment boundary E_{ij} from G that has the minimum $t_{i,j}$ as calculated by (3), and merging the two vertices/segments V_i and V_j to form merged segment V_{ij}, where $|V_{ij}| = |V_i| + |V_j|$, and $\overline{u}_{ij}$ is the volume-weighted average of $\overline{u}_i$ and $\overline{u}_j$, i.e., $\overline{u}_{ij} = (\overline{u}_i \cdot |V_i| + \overline{u}_j \cdot |V_j|)/(|V_i| + |V_j|)$. This step can be efficiently performed by maintaining an ordering on the edges with a priority queue implementation.
3) If any multiple edges have been introduced by the edge contraction, merge the repeated edges to again make G simple, setting the area of the merged boundary to the combined areas of the two constituent boundaries.
4) Recalculate, and update the priority queue with, the new merging costs $t_{k,ij}$ for all boundaries $E_{k,ij}$ incident on segment V_{ij}. This causes the priority queue to reorder, efficiently maintaining the boundaries in priority order on $t_{i,j}$, allowing later merging operations to be performed quickly.
5) Repeat Steps 2–4 until all segments are merged, to generate a complete multiscale hierarchy of potential segmentations.
6) Select the best final segmentation level for the given application. As the iterations proceed, the $t_{i,j}$ of (3) trends upwards, but is punctuated by local minima representing the collapse of internal structure to form larger segments. The monotonically increasing sequence of $t_{i,j}$ values excluding these minima defines the "full-lambda schedule" of [18], and together define a multiscale segmentation, from fine to coarse. A simple image-based heuristic is used to choose the final segmentation level—the segmentation preceding the Nth last local minimum is used, where N is determined empirically (e.g., $\mathrm{N} = 31$ for the simulation experiments). The internal carotids are adequately segmented throughout a range of the final segmentations and so the precise ordinal position used is not critical and this procedure has been found to be robust

for a given image type, across varying patient datasets and absolute activity levels.

To efficiently implement the algorithm described previously, graph G is implemented using an adjacency list representation, with a Fibonacci heap (or similar priority queue) used to store and maintain the order of the edges E_{ij}. Access to the priority queue's minimum element is O(log N), so it can be shown, with minor changes to the algorithm analysis in [18], that the expected worst case complexity is O(N log N), where N is the total number of voxels. A further dynamic extension of the algorithm [13] can be used to further limit the computational complexity for large datasets such as whole body scans. This extension was not used in the experiments described in this paper.

The distance metric used in the merge operations is a Mahalanobis distance between the PCA channel vectors

$$||\overline{u}_i - \overline{u}_j||^2 = (\overline{u_i} - \overline{u}_j)^T C_y^{-1} (\overline{u_i} - \overline{u}_j) \tag{4}$$

where C_y is the covariance matrix of the PCA channel vectors and is a diagonal matrix of ordered eigenvalues from the principal component transform; therefore the inverse in this case is simply the reciprocal of the diagonal elements. The effect of this standardized distance is to perform a segmentation based primarily on differing TAC shape, and not simply on activity levels.

Step 3: Internal Carotid Detection: Segments corresponding to brain vasculature are detected by a pattern recognition stage: the following moment-based test of the mean TAC of each segment is used to distinguish blood segments based on shape. The early peak of the blood curve leads to a large negative value for this shape descriptor, and so thresholding on this leads to a robust test for vasculature. For a given time activity vector $\mathrm{TAC} = \mathrm{TAC(t)}$, we require

$$\frac{\left(\sum_t (t - \overline{t}) \cdot TAC(t)^2\right)}{\left(\sum_t TAC(t)\right)^2} < \text{threshold} \tag{5}$$

where $\overline{t} = \sum_t t \cdot TAC(t) / \sum_t TAC(t)$ and the threshold is determined empirically for a given dynamic PET protocol.

Note that such a pattern recognition technique is more robust than alternative schemes such as thresholding of segments based on the energy in the second PCA channel, as the signal in the second and higher channels can vary with the absolute activity levels of a given dataset—enabling such pattern recognition techniques is a major advantage of an underlying robust segmentation.

The internal carotid segments are recognized from the venous segments by the central location of the former (currently a coarse bounding box in the central location is used to select the segments, but more sophisticated pattern recognition schemes could be used here).

A morphological dilation operation is then used to sample the tissue surrounding the vasculature ROIs to be used in spill-over correction in later steps, i.e., several iterations of morphological dilation of the vasculature segment followed by subtraction of the original segment produces an ROI over a ring of tissue around the vasculature (of approximately the same width as the internal carotid segments).

Step 4: Input Function Estimation: Ideally, having segmented out the vasculature, one could directly read off the arterial input curve by averaging over the internal carotid ROIs for each frame. However there are several factors that can limit the quantitative accuracy of PET for such small ROIs if not otherwise taken into account. These include: a) spill-over and partial recovery, b) partial volume effects, c) whole blood versus plasma activities, d) coarse PET sampling schedules, and e) patient movement.

These issues are now discussed in detail.

a) Spill-over and partial recovery: This is the most significant factor. It is well-known that it can be difficult to accurately quantify small regions of interest in PET as they suffer from a partial recovery coefficient and spill-over effects from adjacent tissues due to the point spread function of the imaging process [19]. The size of the recovery coefficient depends upon the reconstruction algorithm used: filtered-back projection (FBP) is known to give worse recovery coefficients compared with iterative expectation maximization–maximum likelihood (EM-ML) reconstruction algorithms. It is also known that the maximum over a region of interest gives a more accurate quantification compared with the mean, due to the approximately Gaussian shape of the blurred ROI, but at the expense of increased noise [19], [20].

The internal carotids are small regions of interest that are close to the limit of what the ECAT 951R scanner used in the experiments can give accurate quantitative results for, therefore this is a significant issue for input function estimation, and so an EM algorithm is used for the image reconstruction in the clinical experiments in this paper.

There are several approaches that have been previously used to estimate the recovery and spill-over coefficients for the mean of the internal carotid ROIs: in [1], a fit to several late venous blood samples is used; [2] and [3] used phantom studies to determine the appropriate recovery and spill-over factors. These approaches could potentially be used with the current algorithm, but an alternative approach is evaluated in this paper.

Define $\mathrm{I_{mean}}$ as the mean over the internal carotid ROI, $\mathrm{T_{mean}}$ as the mean over the tissue ROI surrounding the internal carotid, $\mathrm{I_{max}}$ as the maximum over the internal carotid ROI, and $\mathrm{T_{max}}$ as the maximum over the tissue ROI surrounding the internal carotid (in practice, the mean of the top 5% of data points is used for $\mathrm{I_{max}}$ and $\mathrm{T_{max}}$).

As in [1], the mean over the internal carotid is assumed to consist of some fraction of the true input curve superimposed with the spill-over activity from the surrounding tissue

$$I_{\mathrm{mean}}(t) = RC \cdot I_{\mathrm{true}}(t) + SP \cdot T_{\mathrm{mean}}(t) \tag{6}$$

where RC is the recovery coefficient and SP is the spillover from adjacent tissues. If I_{true} is known from blood sampling, then RC and SP can be directly computed using a (linear) least squares fit of I_{mean} and T_{mean} to I_{true} (the "fit-to-known" procedure). However, in clinical application I_{true} is, of course, unknown. In this case, as discussed previously, the maximum over the internal carotid ROI, $\mathrm{I_{max}}$, can be used as an initial quantitatively accurate but noisy estimate, $\hat{I}_{\mathrm{true}}$, to the true arterial input curve estimate, as it is not substantially affected by partial recovery effects (when using an EM reconstruction—this was found to

not be the case with FBP). A least squares fit against $\hat{I}_{\text{true}}$ as a linear fitting problem can then be used to estimate the recovery coefficient, RC, and spill-over coefficient, SP, and therefore a final, low noise, estimate of I_{true} can be computed from I_{mean}, T_{mean}, RC and SP (the "fit-to-maximum" procedure). (Note 1: it was found empirically that in the least squares fit to $\hat{I}_{\text{true}}$, a weighting to favor the latter portion of the curve improved the estimation results as the initial arterial peak shape matched less well with the peripherally-taken arterialized-venous blood samples—a weighting of 5.0 for the last 8 samples was empirically found to work well across the datasets studied, after training on the first two datasets.)

(Note 2: One complication in determining $\mathrm{I}_{\max}$ is that it is possible for the surrounding tissue activity to exceed the blood activity toward the end of a one hour scan, in which case the maximum over the internal carotid ROI will be increased due to spillover of activity from the surrounding tissue. Thus, if $T_{mean} > I_{mean}$ at the end of the scan, assuming that $I_{\max}/T_{\max} \approx I_{mean}/T_{mean}$, $\mathrm{I}_{\max}$ is approximately corrected to $I_{\max} \cdot I_{mean}/T_{mean}$).

Note that, as in [1], if late venous samples are available then they could be included in the least-squares fit to improve the estimation.

b) Partial volume effects: Voxels that contain two or more tissues can give inaccurate quantification, and this especially affects ROIs lying predominantly in the transverse direction. In the context of vasculature segmentation, including the venous transverse sinus can lead to inaccurate quantification (in a preliminary study [21], the bias noted in one patient dataset was subsequently found to be due to the transverse sinus being included in that dataset due to a slight vertical misalignment of the patient in the gantry). In this paper, only the internal carotids were used to determine the max. blood curve to avoid transverse sinus effects.

c) Arterial versus plasma activities: The activity measured over the internal carotid is of whole blood, whereas the measured input function from blood sampling is plasma activity (which is the form required for compartmental modeling). In [22], it is demonstrated that in humans the red blood cell and plasma activities equilibrate quickly and so it was concluded that there is minimal error in equating the whole blood and plasma activities in humans. However, in [3] and [23] careful measurements were used to correct for the small difference between whole blood and plasma activities (an average of 9% difference was found in [23]). Such corrections were not used in the clinical experiments in this paper, but could potentially be incorporated to further improve the input curve estimations.

d) Limited PET sampling schedule: A fine PET sampling schedule is needed to faithfully capture the input function. Alternatively, list mode acquisition could be used so that an optimal schedule could be designed after the study. In the data for the clinical experiments, an infusion over 3 min rather than a bolus injection of the tracer $^{18}\mathrm{F}$-fluoro-deoxyglucose (FDG) was used, which assists in the curve shape being adequately sampled by the relatively coarse sampling schedule used.

e) Patient movement: Patient movement needs to be restricted during the scan, or alternatively inter-frame registration performed, when dealing with small features such as internal carotids. The patient data used in the clinical experiments avoided significant patient motion during the study.

TABLE I
MODEL PARAMETERS FOR SIMULATION TISSUES

Tissue	k1	k2	k3	k4
Gray matter	0.1060	0.4108	0.0868	0.0065
White matter	0.1028	0.9949	0.1863	0.0062
Epileptic	0.0785	0.2229	0.0324	0.0006

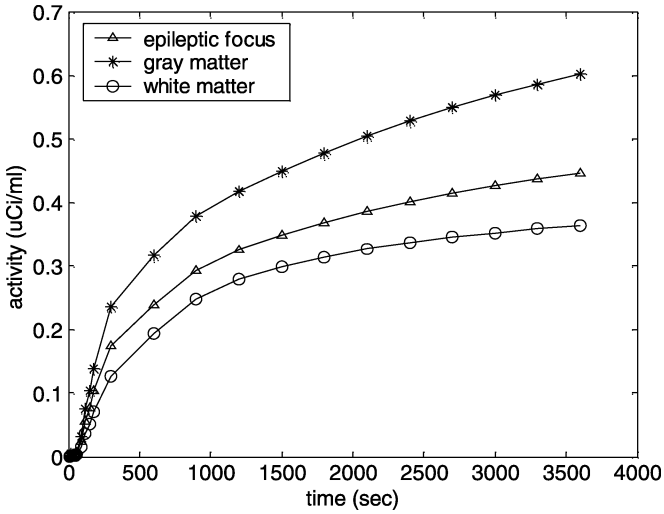

Fig. 1. TACs of simulated tissues.

III. EXPERIMENTAL DATA AND METHOD

To evaluate the algorithm, we first performed simulation studies to investigate the general performance of the segmentation based on TAC shape compared with k-means clustering. We then used four clinical datasets to evaluate the algorithm specifically for vasculature segmentation performance.

A. Simulation Studies

To investigate the ability of the segmentation algorithm to distinguish tissues of differing TAC shapes in the presence of noise, and to compare its performance with clustering approaches, a simulation study was performed. The hypothesis tested is that the Mumford–Shah model, by incorporating region compactness criteria, which clustering approaches ignore, will perform with better demarcation in the presence of noise.

Five slices of the brain Zubal phantom were used to simulate a dynamic FDG PET brain study [24]. The activities in gray matter, white matter and an epileptic focus were generated using a three-compartment FDG model using clinically derived model parameters published in [25] (see Table I).

As in [6], the activities of the thalamus and caudate body were set to scaled versions of the gray matter tissue TAC ($\times$ 1.2)—this model was found to closely match clinical samples. The epileptic focus TAC was assigned to the temporal lobes, and to an artificial focus inserted into the upper right white matter. The plasma time-activity curve used was obtained from a previous clinical study. An optimal sampling schedule [25] of 248, 761, 1681, 2997, and 3600 s was used to generate the tissue TACs.

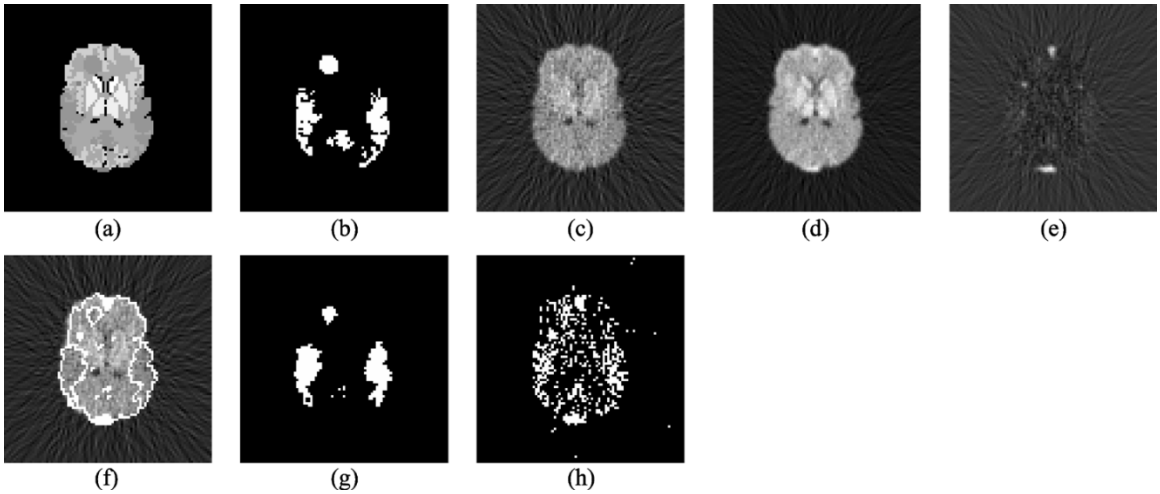

Fig. 2. (a)–(h) Simulation study results.

The tissue TACs were assigned to each brain region and a dynamic sequence of sinograms was generated by forward projecting the images into 3.13 mm bins on a 192×256 angle sinogram (no additional blur was modeled to ensure the ground truth remained valid). Attenuation was modeled according to each tissue type, and appropriate Poisson noise was then added to each sinogram frame such that noise simulated the system sensitivity of the ECAT PET scanner used in the clinical experiments (3×10^7 total detections over all frames). The first five PCA channels of the sinograms were reconstructed in this case using filtered-back projection (FBP) with a Shepp-Logan filter at 0.5 cycles/pixel cutoff and $\mathrm{zoom} = 1.0$.

The phantom was then segmented by the described algorithm and for comparison was also clustered by a k-means clustering algorithm using the same Mahalanobis distance metric and a stopping level of five clusters for the k-means (determined empirically as giving the best performance). The k-means used a randomized initial seeding.

Note that in FDG brain PET, gray matter, white matter, basal ganglia and thalamus have approximately the same TAC shape and so the experimental objective is for these tissues to form a single segment/cluster (the similar TAC shape of these tissues was, independently, noted in [9]). Vasculature and the epileptic focus tissues have substantially differing shape and so should result in distinct segments/clusters. The noise-free tissue TAC curves used are shown in Fig. 1—note that the epilepsy curve is of only subtly different shape and so is a sensitive test of segmentation performance.

The segmentation algorithm was applied and the segments/clusters with the maximal intersection with the truth epileptic regions were extracted; this models an optimal pattern recognition stage for selecting the segments of interest.

The performance variables used in this experiment were the true positive pixel fraction which is defined as the number of pixels in the detected segments that overlie the truth epileptic region, divided by the total pixel area of the truth epileptic region, and the false positive pixel fraction which is the number of pixels in the matching segments that do not overlie the truth epileptic region divided by the area of brain tissue excluding the truth epileptic region.

TABLE II
SEGMENTATION TRUE AND FALSE POSITIVES

	True positive pixel %	False positive pixel %
M-S Segmentation	74	10
K-means Clustering	58	26

B. Clinical Studies

To evaluate the segmentation performance on realistic clinical data, in this experiment we aim to segment the vasculature (internal carotid and venous sinuses). As well as being a primary aim for the application of input function estimation, this is a sensitive test of segmentation performance, as the vasculature is a tissue type for which an objective truth exists—the TAC from arterial blood sampling—with which to validate the segmentation demarcation and localization.

Four clinical brain FDG datasets from a previous quantitative dynamic PET study were analyzed. They were acquired on a SIEMENS ECAT 951R PET scanner (BGO detectors—110 kcps/uCi/ml system sensitivity, spatial resolution of 5.4 mm FWHM on-axis, plane spacing 3.375 mm) in two-dimensional (2-D) mode at the PET and Nuclear Medicine Department, Royal Prince Alfred Hospital, Sydney. (Five datasets were initially acquired, but one dataset was rejected for this study as the blood samples were missing early values due to operator error).

A dynamic sequence of 22 frames was acquired over 60 min, consisting of 6×10 s, 4×30 s, 1×120 s, and 11×300 s intervals. The patients were injected with approximately 400 MBq of FDG, infused at a constant rate over a 3-min period using an automated injection pump. Blood samples were venous, taken after heating the arms to promote arterio-venous shunting. Such peripheral arterialized-venous blood samples have been previously validated as an adequate replacement for peripheral arterial sampling for quantification purposes [26]. The blood was sampled at, approximately, the following schedule: 12×30 s, 2×2 min, 2×5 min, 1×10 min, 2×15 min. The plasma activity was measured in a calibrated well counter. The sinogram data were corrected for attenuation, randoms and differences in detector efficiency (no scatter correction was performed). Data

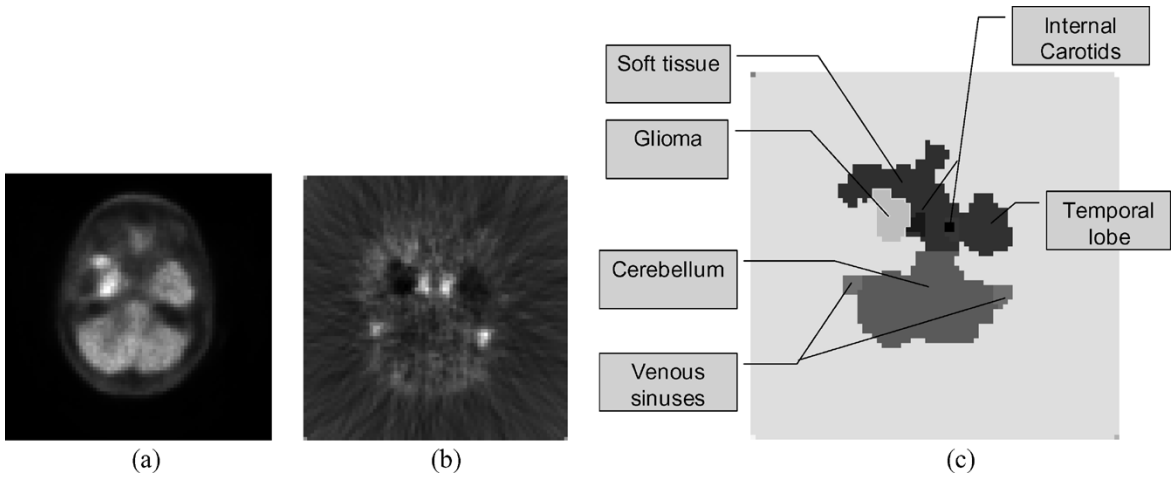

Fig. 3. Clinical functional segmentation results. (a) PCA channel 1. (b) PCA channel 2. (c) Clinical tissue segmentation result showing functional segmentation into distinct tissue types.

was decay-corrected to the start of the study. All image reconstructions were performed using a standard ML-EM reconstruction (with no form of regularization) with 32 iterations, with zoom factor 1.0, and over the full field of view.

The input curves were estimated using the procedure described in Section II. The recovery and spill-over coefficients were estimated by both (a) fitting to the known input curve ("fit-to-known"), and (b) fitting to the maximum curve ("fit-to-maximum") as described in Section II, Step 4.

The primary figure-of-merit used to evaluate the correct localization and demarcation of the vasculature segmentation was the normalized root mean squared error (NRMS) between the known input curve and the fit-to-known estimated input curve

$$NRMS(x) = \sqrt{\frac{\sum_1^n \left(\hat{x}(i) - x(i)\right)^2}{\sum_1^n \left(x(i)\right)^2}}. \tag{7}$$

Using the fit-to-known procedure removes as an experimental variable the errors introduced in the recovery and spillover estimation procedure, and so gives a more specific test of pure segmentation performance.

As an indication of the accuracy that may be obtained in practice, where the true input curve is not known, the NRMS error of the fit-to-maximum curve is also presented. Also, the effect of the errors in the input curve estimate on local cerebral metabolic rate of glucose (LCMRGlc) estimation was evaluated by using the estimated input curves to estimate the LCMRGlc from gray and white matter TACs generated with a three-compartment model using the known true blood input curve [27], [28].

IV. Results and Discussion

A. Simulation Study

Fig. 2(a) shows the final frame of the central plane of the phantom data with no noise added. Fig. 2(b) shows the regions that have had the epilepsy TAC assigned—the temporal lobes and a small focus in the upper right quadrant. Fig. 2(c) shows the final frame of the central plane after noise is added (approximately clinical levels). Fig. 2(d) and (e) shows PCA channels one and two: note that the superior sagittal sinus is clearly visible in the second channel, and the epilepsy ROIs are faintly visible.

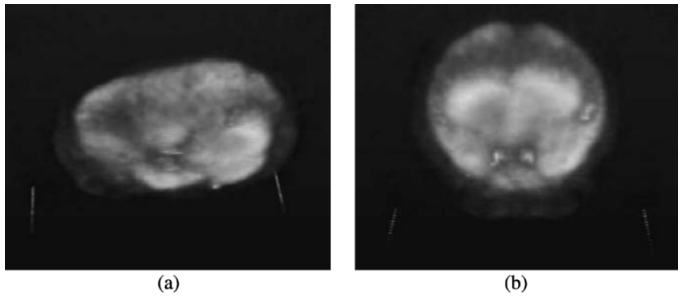

Fig. 4. Clinical data PCA preprocessing results. (a) Left latero-inferior view. (b) Postero-inferior view.

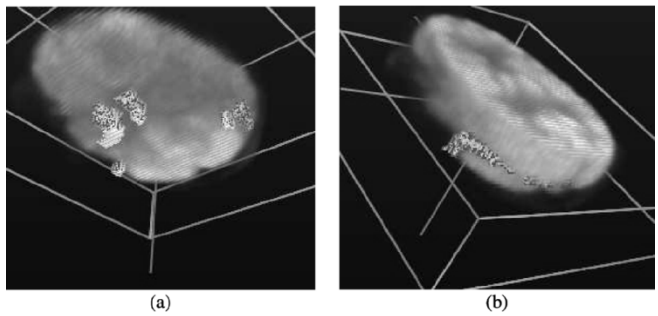

Fig. 5. Clinical data vasculature segmentation results. (a) Left inferior view. (b) Left posterior view.

Fig. 2(f) shows the segmentation results using the Mumford–Shah segmentation algorithm (segments outlined in white); note that the gray and white matter, basal ganglia and thalamus, which have TACs that are (approximately) scaled versions of each other, are, correctly, not segmented out, and only the epileptic focus and vasculature which have differing TAC shapes are segmented out. Fig. 2(g) shows the final detected segments corresponding to the epileptic regions.

Fig. 2(h) shows the corresponding clusters matching the epileptic regions using a k-means clustering algorithm (using the same Mahalanobis metric as the segmentation algorithm).

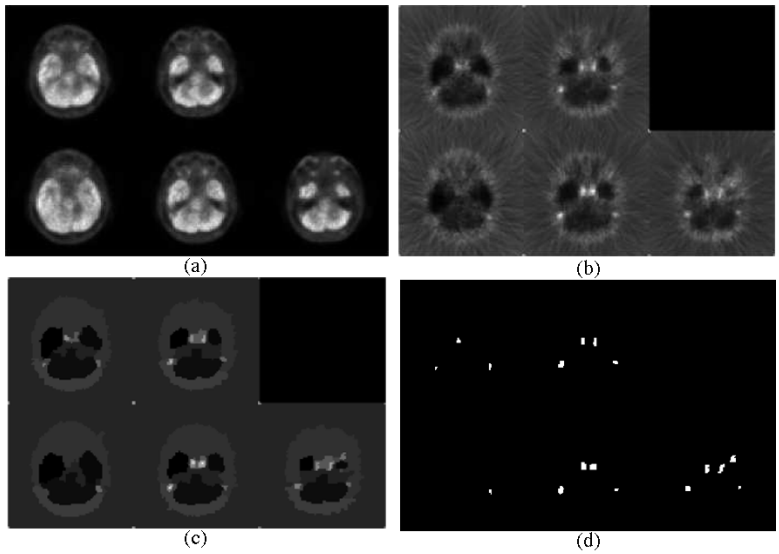

Fig. 6. (a) Clinical data PCA channel 1: lower 5 planes. (b) PCA channel 2: lower 5 planes. (c) Segmentation results: channel 2. (d) Detected vasculature segments.

In this high-noise environment, the k-means performs poorly and in particular suffers from poor demarcation. By contrast, the Mumford–Shah segmentation algorithm which incorporates region adjacency information and favors compact regions generates well demarcated results even in the presence of severe noise. It was also noted that the results for the k-means varied widely between runs due to different initial cluster seeds and sometimes could fail to segment out the vasculature, whereas the Mumford–Shah segmentation was fully deterministic and stable.

For these segmentation results, the percentage of true and false positive pixels for the segments/clusters corresponding to the epilepsy regions are shown in Table II: the Mumford–Shah formulation has a lower false positive pixel fraction and greater true positive pixel fraction.

B. Clinical Study

The internal carotids and venous sinuses were successfully segmented in all four datasets, by visual inspection.

Fig. 3 shows one plane of the segmentation of one patient dataset at the base of the brain. Fig. 3(a) shows the (noise-normalized) PCA channel 1. Fig. 3(b) shows the PCA channel 2. Note that the internal carotids and venous sinuses are visible. Fig. 3(c) shows the segmentation result on this dataset—the various tissues with substantially different TAC curve shapes are successfully distinguished by the segmentation algorithm. The cerebellum, arterial and venous vasculature and glioma have substantially different TACs compared with gray and white matter and so are separately segmented out. (Note that the soft tissue is distorted due to previous excision surgery on the glioma).

Fig. 4(a) and (b) shows frames from a real-time 3-D volume visualization of the PCA results of one clinical dataset. The first PCA channel is shown in gray-scale and the second channel is shown in red shades. Note that most of the vasculature signal is separated into the second and higher channels: the internal carotids and venous sinuses are clearly visible. Fig. 5(a) and (b) displays a 3-D surface-rendering of the final segmentation results of one dataset: the internal carotids and transverse and sigmoid sinuses were successfully segmented.

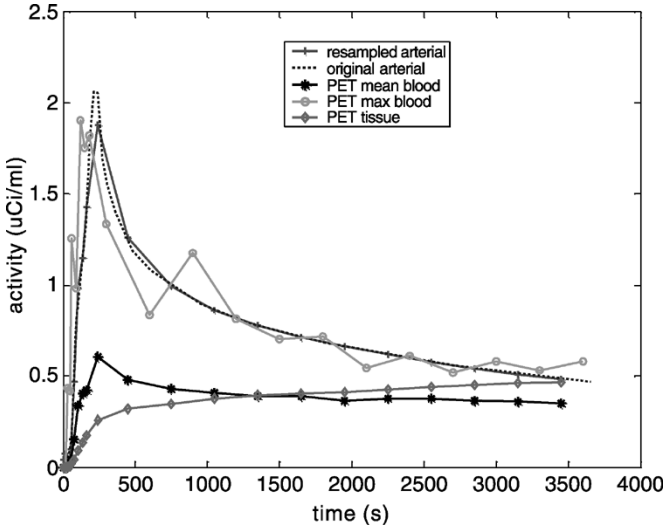

Fig. 7. Patient 1 measured and estimated input functions.

Fig. 6 shows the segmentation results of the lower five planes in one patient dataset. Fig. 6(a) and (b) shows PCA channels 1 and 2. Fig. 6(c) shows the segmentation result as an approximation image of channel 2, where each segment is replaced by its mean value in channel 2. Fig. 6(d) shows the segments that correspond to the (arterial and venous) vasculature.

Fig. 7 shows the input function estimation results for patient one. In the legend, the "original arterial" curve plots the original measured (arterialized-venous) blood activity in uCi/ml versus time in seconds.

Because the PET sampling schedule used in the studied datasets is coarser than that of the blood samples, the measured input curve needed to be resampled (by integration) to the PET sampling schedule to allow a valid comparison with the extracted blood curve—this is the curve "resampled arterial" in the legend. Note that a finer PET sampling schedule would

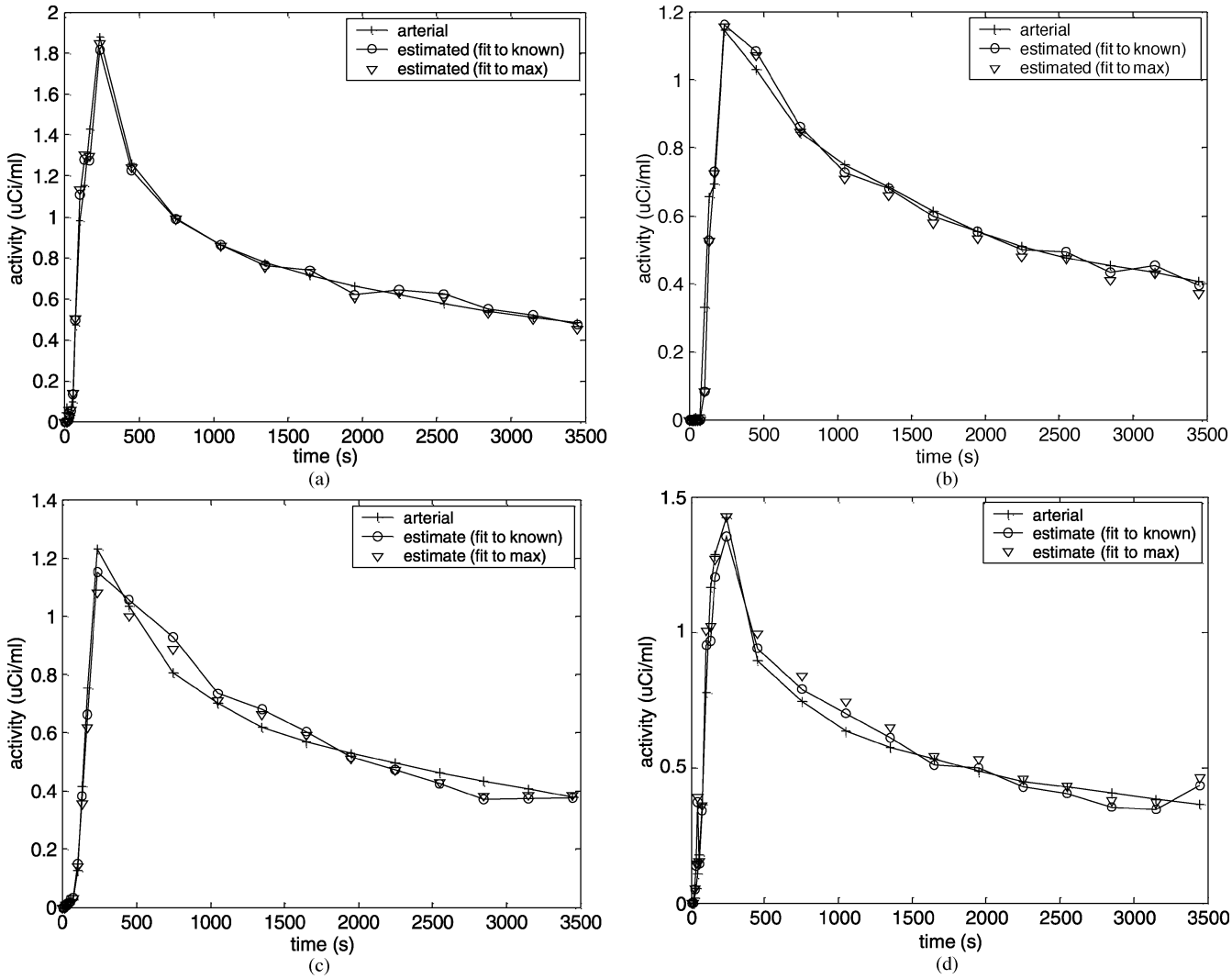

Fig. 8. (a) Patient 1 estimated input curves. (b) Patient 2 estimated input curves. (c) Patient 3 estimated input curves. (d) Patient 4 estimated input curves.

allow this curve to approach the measured blood curve arbitrarily closely.

"PET max blood" shows the maximum (5%) over the internal carotid ROIs. This gives a noisy curve with a recovery coefficient close to 1.0. "PET mean blood" plots the mean over the vasculature ROIs—it was found that using both the internal carotid and venous sinus ROIs gave a better estimate of the measured plasma input curve than using the internal carotid alone. This is likely due to arterialized-venous blood samples being used as the comparison standard; in practice, using the internal carotids alone may be expected to give a more accurate estimate of the true arterial input function. The recovery coefficient for the mean curve is substantially <1.0. It was also noted that the peak of the maximum curve precedes that of the mean curve in time (on the order of 1 min), presumably due to the first circulatory passes being detected in the maximum curve, but not in the mean; the peripheral vasculature is known to filter out small signals that are visible in direct internal carotid sampling due to mixing and other haemodynamic effects [29]. "PET tissue" plots the mean tissue TAC of the tissue surrounding the internal carotids.

TABLE III
NRMS ERROR % OF ESTIMATED INPUT CURVES

Patient Number	NRMS error % (fit to known)	NRMS error % (fit to max)	NRMS error % (fit to known- arterial)
1	7.4	7.6	9.7
2	11.1	11.4	12.2
3	8.4	9.7	8.3
4	14.0	15.1	16.8

Fig. 8(a)–(d) shows the estimated input functions for patients 1 to 4, respectively. The estimate fitted to the known true input curve, which gives an indication of the best possible input curve estimation result, is marked with circles. The curve marked with triangles shows the estimated input curve created by a fitting of the mean and tissue curves using the maximum curve to estimate the recovery coefficient and spill-over fraction.

The NRMS error (%) of the estimated input curves for the four datasets is shown in Table III. The first column shows the fit-to-known results when both arterial and venous vasculature is used in the estimation. The results using fit-to-known estimation were sufficiently close to indicate that correct vasculature segmentations had been obtained. The estimate is less accurate, as expected, with fit-to-max estimates due to the additional error

TABLE IV
RELATIVE ERROR OF ESTIMATED METABOLIC RATE OF GLUCOSE USING ESTIMATED INPUT CURVES

Pt.	Error % (gray) (fit to known)	Error % (white) (fit to known)	Error % (gray) (fit to max)	Error % (white) (fit to max)	Resampling Error % (gray)	Resampling Error % (white)
1	0.89	1.43	0.92	0.77	2.73	2.70
2	7.18	3.69	6.71	9.83	5.7	2.7
3	10.18	1.96	6.61	1.44	0.38	0.35
4	9.98	4.80	15.03	10.24	2.29	2.83

TABLE V
ESTIMATED RECOVERY AND SPILL OVER COEFFICIENTS (FIT-TO-KNOWN)

Pt	RC	Spillover
1	0.27	0.48
2	0.28	0.38
3	0.22	0.41
4	0.25	0.47

introduced by recovery and spill-over coefficient estimation. It is expected that a finer PET sampling schedule would improve these results. The final column shows the fit-to-known estimate when only the internal carotids are used for the estimate and not the venous sinuses; as expected this does not fit the arterialized-venous samples as well.

Table IV shows the percentage error when using the extracted blood curves to estimate the glucose metabolic rate for simulated (noise-free) gray and white matter TACs generated using the true blood curve. Columns 1 and 2 show the LCMRGlc estimation errors for the fit-to-known input curve estimate. Columns 3 and 4 show the errors for the fit-to-max input curve. The maximum error in quantification was 15%—the average error for fit-to-maximum was 6%. Columns 5 and 6 show the proportion of the error in quantification due to the resampling of the blood curve to the limited PET sampling schedule used in the datasets—the worst error due to this effect was only 6%.

Table V shows the recovery and spill-over coefficients computed for the four patient datasets, using the fit-to-known procedure. Note that for this small sample of patients that the spill-over fraction is fairly consistent, due to the automated, reproducible sampling of the tissue surrounding the internal carotids; in [1] it was noted that the spill-over coefficients varied greatly due to the manual selection of the tissue ROI. This suggests that, for a given PET scanner and protocol, when using this technique the use of a fixed spill-over coefficient for the input function estimation may be practical.

V. CONCLUSION

We have demonstrated that the Mumford–Shah segmentation model and its heuristic optimization by the described graph-theoretic segmentation algorithm combined with a PCA data reduction and Mahalanobis distance metric is effective in segmenting dynamic PET datasets into functionally distinct tissues, and, in particular, is effective in segmenting brain vasculature. The Mumford–Shah formulation, with its regularization on region compactness, was shown to be more effective in functionally segmenting dynamic PET than simple clustering approaches.

We evaluated its application to image-derived input function estimation, and have validated the segmentation correctness for this purpose by comparing the extracted input curve with a measured arterialized-venous input curve. Also, our results suggest that, with EM reconstruction, the maximum over the internal carotid ROIs can be used to calibrate the mean curves and so avoid or limit the late venous samples required by the method of [1].

The metabolic rate of glucose estimation results demonstrate the feasibility of this approach for completely noninvasive input function estimation, and the current results may be sufficiently accurate for some applications; however, the limiting factor for this technique appears to be the quantitative accuracy of PET for small ROIs rather than in the delineation of the vasculature itself, and further improvement in these results could be expected by the use of more recent PET scanners with higher sensitivity and resolution, and further refinements of the Step 4: input function estimation.

This technique should be applicable to the extraction of vasculature in other regions for input function estimation and visualization purposes, e.g., abdominal PET for aortic and hepatic vasculature segmentation; for such larger vessels more accurate input curve estimation could be expected. Also, this technique could be applied to visualizing other tissue types characterized by differing TACs.

ACKNOWLEDGMENT

The authors would like to thank the staff of the Department of Nuclear Medicine, Royal Prince Alfred Hospital, for providing the PET clinical datasets and for useful discussions and comments.

REFERENCES

[1] K. Chen, D. Bandy, E. Reiman, S. C. Huang, M. Lawson, D. Feng, L. Yun, and A. Palant, "Noninvasive quantification of the cerebral metabolic rate for glucose using positron emission tomography, ^{18}F-Fluoro-2-Deoxyglucose, the Patlak method, and an image-derived input function," *J. Cereb. Blood Flow Metab.*, vol. 18, pp. 716–723, 1998.

[2] J. Litton, "Input function in PET brain studies using MR-defined arteries," *J. Comput. Assist. Tomog.*, vol. 21, no. 6, pp. 907–909, 1997.

[3] R. Trebossen *et al.*, "Quantifying cerebral PET with FDG using dynamic internal carotid arteries imaging," in *Proc. IEEE Conf. Rec. Nuclear Science Symp. Medical Imaging*, vol. 3, 1998, pp. 1733–1736.

[4] R. Gonzalez and R. Woods, *Digital Image Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[5] J. Ashburner, J. Haslan, C. Taylor, V. J. Cunningham, and T. Jones, "A cluster analysis approach for the characterization of dynamic PET data," in *Quantification of Brain Function Using PET*, R. Myers, V. Cunningham, D. Bailey, and T. Jones, Eds. San Diego, CA: Academic, 1996, pp. 301–306.

[6] K. P. Wong, D. Feng, S. Meikle, and M. Fulham, "Segmentation of dynamic PET images using cluster analysis," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 200–207, Feb. 2002.

[7] J. Chen, S. Gunn, M. Nixon, and R. Gunn, "Markov random field models for segmentation of PET images," in *IPMI2001, LNCS 2082*. Berlin, Germany: Springer-Verlag, 2001, pp. 468–474.

[8] Y. Kimura, M. Senda, and N. M. Alpert, "Fast formation of statistically reliable FDG parametric images based on clustering and principal components," *Phys. Med. Biol.*, vol. 47, pp. 455–468, 2002.
[9] J. Brankov, N. Galatsanos, Y. Yang, and M. Wernick, "Segmentation of dynamic PET or fMRI images based on a similarity metric," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 5, pp. 1410–1414, Oct. 2003.
[10] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Image Understanding*, S. Ullman and W. Richards, Eds. Norwood, NJ: Ablex, 1988.
[11] J. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Cambridge, MA: Birkhauser, 1995.
[12] J. Shah, "A common framework for curve evolution, segmentation and anisotropic diffusion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1996, pp. 136–142.
[13] B. J. Parker and D. Feng, "Large three-dimensional data set segmentation using graph-theoretic energy-minimization approach," in *Proc. SPIE Medical Imaging*, vol. 5032, San Diego, CA, Feb. 2003, pp. 491–499.
[14] F. Pederson, M. Bergstrom, E. Bengtsson, and B. Langstrom, "Principal component analysis of dynamic positron emission tomography images," *Eur. J. Nucl. Med.*, vol. 21, pp. 1285–1292, 1994.
[15] M. Samal, M. Karny, H. Benali, W. Backfrieder, A. Todd-Pokropek, and H. Bergmann, "Experimental comparison of data transformation procedures for analysis of principal components," *Phys. Med. Biol.*, vol. 44, pp. 2821–2834, 1999.
[16] Z. Chen, B. J. Parker, and D. Feng, "Temporal processing of dynamic positron emission tomography via principal component analysis in the sinogram domain," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2612–2619, Oct. 2004.
[17] G. Koepfler, C. Lopez, and J. M. Morel, "A multiscale algorithm for image segmentation by variational method," *SIAM J. Numer. Anal.*, vol. 31, no. 1, pp. 282–299, 1994.
[18] N. J. Redding, D. J. Crisp, D. H. Tang, and G. N. Newsam, "An efficient algorithm for Mumford-Shah segmentation and its application to SAR imagery," in *Proc. Conf. Digital Image Computing Techniques and Applications (DICTA '99)*, 1999, pp. 35–41.
[19] E. Hoffman, S. C. Huang, and M. Phelps, "Quantitation in positron emission computed tomography: 1. Effects of object size," *J. Comput. Assist. Tomog.*, vol. 3, no. 3, pp. 299–308, Jun. 1979.
[20] J. W. Keyes, "SUV: standard uptake or silly useless value?," *J. Nucl. Med.*, vol. 36, no. 10, pp. 1837–1839, Oct. 1995.
[21] B. J. Parker and D. Feng, "Graph-based energy-minimization segmentation and PCA applied to internal carotid extraction in neurological PET," in *Proc. IEEE Conf. Rec. Nuclear Science Symposium and Medical Imaging*, vol. 4, Portland, OR, Oct. 2003, pp. 2613–2617.
[22] S. S. Gambhir, M. Schwaiger, and S. C. Huang *et al.*, "Simple noninvasive quantification method for measuring myocardial glucose utilization in humans employing positron emission tomography and fluorine-18-deoxyglucose," *J. Nucl. Med.*, vol. 30, pp. 359–366, 1989.
[23] T. Ohtake, N. Kosaka, T. Watanabe, and I. Yokoyama *et al.*, "Noninvasive method to obtain input function for measuring tissue glucose utilization of thoracic and abdominal organs," *J. Nucl. Med.*, vol. 32, no. 7, July 1991.
[24] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gingi, and P. B. Hoffer, "Computerized three-dimensional segmented human anatomy," *Med. Phys.*, vol. 21, pp. 299–302, 1994.
[25] X. Li, D. Feng, and K. Chen, "Optimal image sampling schedule: a new effective way to reduce dynamic image storage space and functional image processing time," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 710–719, Oct. 1996.
[26] M. Phelps, S.-C. Huang, E. J. Hoffman, C. Selin, L. Sokoloff, and D. E. Kuhl, "Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-decoxy-D-glucose: validation of method," *Ann. Neurol.*, vol. 6, pp. 371–388, 1979.
[27] R. A. Hawkins, M. E. Phelps, and S. C. Huang, "Effects of temporal sampling, glucose metabolic rates, and disruptions of the blood-brain barrier on the FDG model with and without a vascular compartment: studies in human brain tumors with PET," *J. Cereb. Blood Flow Metab.*, vol. 6, pp. 170–183, 1996.
[28] D. D. Feng, D. Ho, K. Chen, L.-C. Wu, J.-K. Wang, R.-S. Liu, and S.-H. Yeh, "An evaluation of the algorithms for determining local cerebral metabolic rates of glucose using positron emission tomography dynamic data," *IEEE Trans. Med. Imag.*, vol. 14, no. 4, Dec. 1995.
[29] J. Litton and L. Eriksson, "Transcutaneous measurement of the arterial input function in positron emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 2, pp. 627–628, Apr. 1990.