Published in final edited form as:

Phys Med Biol. 2010 November 7; 55(21): 6549–6569. doi:10.1088/0031-9155/55/21/014.

Segmentation of Mouse Dynamic PET Images Using a Multiphase Level Set Method

Jinxu Cheng-Liao and Jinyi Qi

Department of Biomedical Engineering, University of California, Davis, CA 95616, USA

Jinyi Qi: qi@ucdavis.edu

Abstract

Image segmentation plays an important role in medical diagnosis. Here we propose an image segmentation method for 4-dimension mouse dynamic PET images. We consider that voxels inside each organ have similar time activity curves. The use of tracer dynamic information allows us to separate regions that have similar integrated activities in a static image but with different temporal responses. We develop a multiphase level set method that utilizes both the spatial and temporal information in a dynamic PET data set. Different weighting factors are assigned to each image frame based on the noise level and activity difference among organs of interest. We used a weighted absolute difference function in the data matching term to increase the robustness of the estimate and to avoid over-partition of regions with high contrast. We validated the proposed method using computer simulated dynamic PET data, as well as real mouse data from a microPET scanner, and compared the results with those of a dynamic clustering method. The results show that the proposed method results in smoother segments with less number of misclassified voxels.

1. Introduction

Positron emission tomography (PET) can measure physiologic functions of organs *in vivo* using radioactive tracers tagged with positron emitters (Phelps, 2004). Three-dimensional (3-D) images of the radiotracer distribution can be obtained using either an analytical or iterative reconstruction algorithm. One advantage of PET is its capability of measuring dynamic changes of radiotracer distribution over a period of time. These dynamic four-dimensional (4-D) PET images include multiple frames of different duration to catch the change of tracer distribution. By segmenting the dynamic image into regions of interest (ROIs), time activity curves (TACs) of different tissues/organs can be extracted and be further analyzed to estimate kinetic parameters.

For dynamic PET, ROIs are often segmented using a single image frame or the sum of a few frames because segmenting dynamic images on each individual frame is impractical and the resulting segments can vary from frame to frame. Many segmentation methods have been developed for 2-D and 3-D images (Pham et al., 2000; Pal and Pal, 1993), such as thresholding (Sahoo et al., 1988; Lee et al., 1998), clustering (Pappas, 1992; Bezdek et al., 1993), and deformable models (Terzopoulos et al., 1987; McInerney and Terzopoulos, 1996; Cohen, 1991; Cohen and Cohen, 1993). However, single frame approaches discard the temporal information and thus reduce the accuracy of the segmentation. Some researchers used clustering technique to segment the 4D data (Wong et al., 2002; Guo et al., 2003; Kim et al., 2006; Brankov et al., 2003; Kimura et al., 2002; Koivistoinen et al., 2004). Wong (Wong et al., 2002) proposed a clustering segmentation algorithm with weighting factors inversely proportional to the noise in each frame. Kimura *et al.* (Kimura et al., 2002)

clustered voxels based on the principal components. O'Sullivan proposed a mixture model (O'Sullivan, 2006) that expressed a voxel-level TAC as a weighted combination of scaled sub-TACs while imposing a local constraint on the number of sub-TACs. Brankov *et al.* (Brankov et al., 2003) developed a similarity component analysis method based on the shape of the TAC rather than the absolute activity. However, these methods did not consider spatial relationship among voxels. Kim *et al.* (Kim et al., 2006) combined the k-means method with a region growing method to use both spatial and temporal information. Parker and Feng (Parker and Feng, 2005) used principal component analysis (PCA) to segment brain dynamic PET image based on TAC shape, while minimizing the surface area to spatially regularize the ROI. Most of these methods were designed for brain images where regions are of similar size, but they are inefficient for handling whole body images where region sizes are quite different. Recently, a local means analysis method was proposed to segment rodent whole-body images by using a hierarchical linkage algorithm with a seed growing method (Maroy et al., 2008, 2010).

Level set method (LSM), originally proposed by Sethian and Osher (Osher and Sethian, 1988), is an effective implicit representation of evolving curves and surfaces. It handles topology changes naturally which reduces the requirement for initialization. Generally, there are two approaches using LSM in image segmentation, boundary-based (Malladi and Sethian, 1995) and object-based (Chan and L. Vese, 2001; Vese and Chan, 2002; Gao and Bui, 2005). Abudfadel *et al.* (Abudfadel et al., 2008) and Kohlberger *et al.* (Kohlberger et al., 2006) developed methods to segment 4D cardiac data that use a shape prior to regularize the result, but these methods segment only one organ at a time. To segment images with more than two regions, one would need to use more than one level set function (LSF) (Vese and Chan, 2002) or a hierarchical approach (Gao and Bui, 2005). Chan and Vese (Vese and Chan, 2002) developed a multi-phase approach using $\log_2$ PLSFs to represent P classes with complex topologies and automatically remove the problem of vacuum.

In this paper, we develop a segmentation method that utilizes both the spatial and temporal information in dynamic PET data based on the piece-wise constant multiphase LSM. This method is capable of segmenting a 4-D dynamic PET image into P classes with $\log_2 P$ level set functions. The classes are considered time-invariant and each is identified by its unique time activity curve. Note that the original images are not necessary to be piece-wise constant and there can be variation in the time activity curves among the voxels inside each region. While there are physiological motions during a PET scan, it has been found that the effect of respiratory motions on image quality is relatively small in rodent imaging and effect of cardiac motion can be eliminated by gating (Yang et al., 2005). Our approach is similar to that presented by Chan *et al.* in (Chan et al., 2000) where a two-phase level set algorithm was applied to vector-value images, but we use an absolute difference data term and develop a new weighting scheme for dynamic PET segmentation.

In a PET scan, activity contrast between organs varies from frame to frame because of changes in tracer distribution over time. However, the commonly used weighting factors that are chosen based on the overall noise in each frame (Wong et al., 2002) ignore the change in activity. For organs that are only separable in early noisy frames, the noise based weighting scheme tends to group them to a single region. To improve the segmentation accuracy of such organs, we developed a feature weighting scheme where the weights are calculated not only based on the noise variance, but also with consideration of activity separation. Another uniqueness of our approach is that we use the absolute difference in the data matching term instead of the classic least squares used in most existing methods (Vese and Chan, 2002; Chan et al., 2000; Yezzi, 2002; Samson et al., 2000; Wong et al., 2002). The reason is that the variation of intensities among different organs in PET images is high. For example, the activity in myocardium or bladder can be tens times greater than that in other soft tissues.

The limited spatial resolution of PET systems often blurs these organ boundaries. Using a quadratic error function would result in many regions on the blurred boundary surrounding the organs with high activity, which make it difficult to segment other low activity regions. While our method is developed for rodent dynamic PET data, the method can also be applied to coregistered images from dual modality imaging systems, such as PET/CT scanners, where both anatomical and functional information can be used simultaneously in the segmentation process.

2. Methods

2.1. Modified Mumford-Shah Model

Dynamic PET contains a time series of images representing tracer concentration at different time points. Let $I_{i,f}$ be the image intensity of the i^{th} voxel at the f^{th} frame with $f = 1, \dots, F$, and ψ_{ji} be the value of the j^{th} level set function at the i^{th} voxel. Let us also introduce a matrix C , where $C_{p,f}$ represents the mean activity in class p at frame f . The objective function $Q(C, \psi_1, \dots, \psi_m)$ is defined as a modified Mumford-Shah model:

$$Q(C, \psi_1, \dots, \psi_m) = \mu \sum_{j=1}^m \sum_i |\nabla H(\psi_{ji})| + \sum_{f=1}^F w_f \sum_{p=1}^P \sum_i |I_{i,f} - C_{p,f}| \chi_{p,i}. \quad (1)$$

The first term in (1) is a regularization term that minimizes the total area of the zero level set surface, where $H(x)$ is the Heaviside function with a global support defined as (Vese and Chan, 2002)

$$H(x) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan\left(\frac{x}{\varepsilon}\right) \right), \quad (2)$$

where ε is the parameter determining the shape of H .

The second term is a data matching term that calculates the weighted sum of the absolute intensity difference between the original image and an estimated image. The characteristic function $\xi_{p,i}$ is defined based on the sign of level set functions. w_f is the weighting factor for frame f and $P = 2^m$ is the maximum number of possible phases/classes[‡]. μ is the regularization parameter that balances between the data matching term and the regularization term. The absolute difference is used here to increase the robustness of the estimate as compared to the least squares function used in (Chan and L. Vese, 2001; Yezzi, 2002; Samson et al., 2000).

Given a dynamic image sequence, the number of LSFs are determined by the number of regions to be extracted. If there are n regions, $m = \text{ceil}(\log_2 n)$ LSFs are usually needed and the image can be segmented into 2^m classes. The characteristic function of each class is defined by the combination of the sign of the LSFs. For two LSFs, the characteristic functions are defined as

$$\chi_1 = H(\psi_1)H(\psi_2) \quad (3)$$

[‡]An image can have less than P regions because some regions defined by the multiphase LSFs can be empty.

$$\chi_2 = H(\psi_1)(1 - H(\psi_2)) \quad (4)$$

$$\chi_3 = (1 - H(\psi_1))H(\psi_2) \quad (5)$$

$$\chi_4 = (1 - H(\psi_1))(1 - H(\psi_2)). \quad (6)$$

Because the classes are considered as time-invariant, the level set functions ψ_{ji} and the corresponding characteristic $\xi_{p,i}$ are independent of the frames.

2.2. Update Formulas

At each time step, $\mathbf{C}$ and ψ_j are updated by an alternating optimization algorithm to minimize the objective function $Q(\mathbf{C}, \psi_1, \dots, \psi_m)$. We first fix the level set functions ψ_j and minimize Q with respect to $\mathbf{C}$, which results in

$$C_{p,f} = I_{q,f}, \quad (7)$$

where $I_{q,f}$ is the smallest voxel value such that the following condition holds

$$\sum_{i: I_{i,f} < I_{q,f}} \chi_{p,i} \geq \sum_{j: I_{j,f} > I_{q,f}} \chi_{p,j}. \quad (8)$$

We then fix the intensity matrix $\mathbf{C}$ and minimize the energy function Q by updating the LSFs in series. A steepest descent algorithm is used to update each ψ_j . If two level set functions are used, the update equations are

$$\frac{d\psi_1}{dt} = \delta(\psi_1) \left\{ \mu \operatorname{div} \left(\frac{\nabla \psi_1}{|\nabla \psi_1|} \right) - \sum_{f=1}^F w_f [(d_{1f} - d_{3f})H(\psi_2) + (d_{2f} - d_{4f})(1 - H(\psi_2))] \right\} \quad (9)$$

and

$$\frac{d\psi_2}{dt} = \delta(\psi_2) \left\{ \mu \operatorname{div} \left(\frac{\nabla \psi_2}{|\nabla \psi_2|} \right) - \sum_{f=1}^F w_f [(d_{1f} - d_{2f})H(\psi_1) + (d_{3f} - d_{4f})(1 - H(\psi_1))] \right\}, \quad (10)$$

where

$$d_{pf} = |I_{i,f} - C_{p,f}|, p=1, \dots, 4. \quad (11)$$

and

$$\delta(x) = H'(x) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + x^2}. \quad (12)$$

$\delta(x)$ controls the effective region of the LSF update and works similarly as the narrow band method (Malladi and Sethian, 1995), with ε determining the width of the narrow band.

2.3. Weights Determination

Dynamic PET uses time frames with varying durations to capture the initial fast changes in tracer distribution while minimizing the data size. The frame durations in a dynamic PET scan range from a few seconds to several minutes. As a result, the noise level varies significantly from frame to frame. The raw PET data can be modeled as Poisson random variables (Maitra and O'Sullivan, 1998), whose signal-to-noise ratio (SNR) is proportional to the square root of the total number of events. For MAP reconstructed images, the overall noise variance in each frame is about proportional to the data variance in the frame. If raw sinogram data are available, we can estimate the spatially variant noise variance based on the Fisher information matrix (Qi and Leahy, 1999). Here we assume only image files are available for segmentation, so we weight each frame by its global noise level. The weighting factor is just the inverse of the standard deviation of noise in each frame. Considering the Poisson statistics, the variance of a decay corrected and scan duration normalized image is

$$\text{var}(f) = \frac{N(f)}{A^2(f)} \quad (13)$$

where $N(f)$ is the total number of events in the f th frame and $A(f)$ is the area under the decay curve

$$A(f) = \int_{t_s(f)}^{t_e(f)} e^{-\lambda t} dt, \quad (14)$$

where $t_s(f)$ and $t_e(f)$ are the start and end times of the f th frame, respectively, and λ is the decay constant of the radionuclide. Thus, the noise-based weighting factor is

$$W_{\text{noise}}(f) = \frac{A(f)}{\sqrt{N(f)}}. \quad (15)$$

The above noise weighting scheme ignores changes of activity separation between frames, so the segmentation accuracy of organs that are only separable in early noisier frames can be limited. To improve the segmentation accuracy of such organs, we propose a feature weighting scheme by considering the difference between the organ TACs. After identifying targeted regions, we manually draw a small sphere with a diameter between 3 and 5 voxels in each region to estimate the mean TACs of the ROIs and surrounding background. Then a separation-to-noise ratio for class i in frame f is calculated by

$$S(i, f) = \frac{\sum_j [C(i, f) - C(j, f)]^2}{\text{var}(i, f)} \quad (16)$$

where $C(i, f)$ and $\text{var}(i, f)$ are the mean activity and the variance of ROI i in frame f , respectively. The above formula was motivated by the feature weighting methods developed for clustering and other segmentation methods (Wettschereck et al., 1997; Dy and Brodley, 2004; Modha and Spangler, 2003; Paredes and Vidal, 2000; Tahir et al., 2007). If only one organ is of interest, then $C(j, f)$ is the activity of the background and the weight is the square root of $S(i, f)$ for l^1 (absolute difference) data matching term. When more than one organ are considered, the feature-based weight is set to

$$W_{\text{feature}}(f) = \sqrt{\frac{\sum_i S(i, f)}{\sum_f S(i, f)}}. \quad (17)$$

In the experiment section, we will compare results with these two sets of weights, W_{noise} and W_{feature} . Note that when ℓ^2 (squared difference) data term is used, the weighting factors in (15) and (17) should be squared.

2.4. Handling of ε and Time Step dt

As apparent in (12), ε controls the range of the “narrow band” around a boundary where the level set functions will be updated at each iteration. On one hand, a large ε includes more voxels in the narrow band and can speed up the front propagation of the level set functions. This helps to attract a level set function to the true region at early iterations. However, a large ε also blurs the characteristic functions ξ_p and lowers the accuracy of the estimates of mean activities. A small ε reduces the overlap between adjacent regions and allows more precise localization of a boundary. Therefore, we start with a large ε (e.g. 0.2) for fast initial convergence and gradually reduce its value in subsequent iterations (e.g. divide ε by 1.05 after each iteration) to fine-tune the segmentations, similar to a multi-resolution approach.

The step size dt also affects the convergence speed of the algorithm. A small dt will guarantee the descend of the objective function. However, a small dt also means long computational time. To solve this problem, we used a simple method to adjust dt during iterations. At each iteration, if the new LSFs decrease the objective function, dt will be multiplied by a greater-than-one constant (e.g. 1.1) to speed up convergence. However, if the objective function is increased, the level set function will be rolled back, and dt will be divided by two until the objective function is decreased.

2.5. Post-processing

For LSM with more than one LSF, regions are defined by the combination of the signs of LSFs, so some regions can be surrounded by a very thin band caused by small differences between two LSFs. These surrounding regions cannot be eliminated by increasing the smoothing parameter μ in the multiphase LSM because the smoothing term operates on each LSF separately. Moreover, small regions with moderate activities, such as the spleen and tumor, are usually grouped with bigger regions because the data term prefers larger regions. Therefore, we use the following post-processing step to fine-tune selected regions to eliminate the surrounding thin-band regions and to extract small regions.

We first define the foreground and background of a targeted region. For a large region, such as the liver and lungs, we defined the foreground as the resulting region from the previous step and background as the voxels outside the region but within a distance of 5 voxels from the region boundary. For a small organ that is assigned to a large region, we use a sphere with a diameter of 5 voxels inside the organ as the foreground and the attached large region as the background. We then calculate a set of feature weighting factors based on the separation-to-noise ratio between the foreground and the background and initialize a new LSF as the shortest distance function to the foreground boundary. Finally, we evolve this LSF using the feature weighting to fine-tune the boundary of the targeted region, while keeping other regions intact.

3. Experimental Validations

3.1. Computer Simulation

We simulated a dynamic scan of a mouse with [^{18}F]FDG using the MOBY phantom (Segars et al., 2004). We focused on the slices spanning from shoulder to bladder. The phantom was represented by $80 \times 80 \times 80$ (sagittal $\times$ coronal $\times$ transverse) images, with $0.4 \times 0.4 \times 0.8 \text{ mm}^3$ voxels. The voxel size is the typical value that is used in microPET images. We added two tumors extracted from a real mouse scan to the phantom, each having a size of 319 and 724 voxels, respectively. Figure 1 shows the digital phantom with regions marked by different color. The voxels in navy are partial volume voxels containing more than one type of tissues. We included the heart beating motion in the simulation, but excluded respiratory motions because its effect on image quality is relatively small in rodent imaging (Yang et al., 2005).

Time activity curves obtained from a real mouse scan were assigned to the organs (shown in Figure 2). The real data contain 28 frames with sampling schedule as $1 \times 1\text{s}$, $7 \times 2\text{s}$, $3 \times 6\text{s}$, $1 \times 20\text{s}$, $1 \times 40\text{s}$, $1 \times 80\text{s}$, $9 \times 180\text{s}$, and $5 \times 360\text{s}$. We forward projected the noise-free images with modeling of the photon attenuation effect. A Gaussian kernel with a zero mean and covariance of 0.1 was used to blur the sinograms. The resulting sinograms were scaled to the same count level as the real data (500 million events in 60 minutes) (Namba et al., 2006), and then Poisson noise was added. The noisy projection data were reconstructed using a maximum *a posteriori* (MAP) algorithm (Qi et al., 1998) to form dynamic PET images. The MAP algorithm provides high-resolution images by accurately modeling the PET system response. The smoothing parameter was selected so that the reconstructed images have similar resolution as real mouse images. A sample reconstruction is shown in Figure 3. In addition, we generated data with 100M and 50M events to study the performance of the proposed method under high noise situations. We expected that the proposed method to be robust with respect to noise level because of its spatial smoothing component.

For the feature-based weighting, the targeted organs were the lungs and liver. We manually drew a spherical ROI with a diameter of 5 voxels in each organ to estimate the mean activity and variance, and calculated the feature weighting factor as described in Section 2.3. The background region used in the calculation is the region surrounding the lungs. Examples of both weighting factors are plotted in Figure 4 for a dynamic PET scan with 500M events. The feature weighting scheme assigns more weight to early frames where the difference between the targeted organs are large.

We used three level set functions to segment the dynamic PET images. The parameter μ was manually adjusted such that the segmentation results do not contain small isolated regions with less than 5 voxels and the organ boundaries were not over-smoothed. We found that the segmentation results were robust to small changes in μ , so we kept μ fixed for all datasets. We chose the dynamic clustering method (Wong et al., 2002) for comparison. We used the “kmean” function in MATLAB with two distance metrics, “Euclidean” (ℓ^2) and “city” (ℓ^1), to study the effect of data matching term. To apply different weighting factors in the k-means method, images were multiplied by the corresponding weighting factors before clustering.

For the level set method, we used simple geometries (ellipsoid, cylinder, etc) as the initial (Figure 5). Each region was manually placed at the corresponding organ in the phantom image. Because LSM can handle changes in topology, it does not require all the organs to be drawn in the initial image. For example, we did not include the tumors and spleen in the initial. Median activity of each initial region can be calculated from (7). When we used feature weighting, the initial mean activity for the targeted regions were replaced by the

sample mean activity. For the k-means, different initials may lead to different results. We thus performed 10 runs with random initials and picked the best result for comparison. We also performed the k-means clustering by initializing with the same LSM initial.

For both methods, the number of organs can affect the accuracy of segmentation. To eliminate the effect of insufficient number of classes, we first used a short phantom by cutting out the bladder region (the portion above the black line in Figure 5) so that 8 classes should be enough to segment the phantom for both methods. We then applied both methods to the full phantom to study the performance when the maximum number of classes is insufficient.

To compare the segmentation results with the original phantom, we isolated all the disjointed regions automatically using the MATLAB function “bwlabeln” and matched them to the original organs. The accuracy of the segmentation was measured by two quantities, the true positive fraction (TPF) (sensitivity) and the false positive fraction (FPF). True/false was defined by the original images. Because of the limited spatial resolution in PET images, partial volume effect appears in all regions. There are two ways to characterize whether a voxel belongs to a region. One is to only consider the voxels fully belongs to the region. The other is to include partial volume voxels. Here, we defined “true” as a voxel belonging to the region completely, and “false” as a voxel being completely outside the region. Positive and negative were determined by the segmentation result, with positive for the voxels that were assigned to the region and negative otherwise. For small regions, such as kidneys, the original FPF would be very small due to the large number of false voxels. Thus, we used the total number of true voxels instead of the total number of false voxels to computed FPF*. The TPF and FPF* were calculated as

$$TPF = \frac{n_{tp}}{n_t} \times 100\% \quad (18)$$

$$FPF^* = \frac{n_{fp}}{n_t} \times 100\% \quad (19)$$

where n_{tp} and n_{fp} are the number of true-positive voxels and false-positive voxels, respectively, and n_t is the total number of true voxels.

We also computed the Dice coefficient, which is defined as

$$S = \frac{2n_{tp}}{n_t + n_p} \times 100\% \quad (20)$$

where n_p is the total number of positive voxels. Note that while partial volume voxels are not included in the TPF and FPF*, they are included in the Dice coefficient.

3.2. Real Mouse Data Experiments

Real mouse data were also used to demonstrate the performance of the proposed method. Two tumor-bearing mice were injected with about 200 μ Ci of 18 F¹⁸FDG and scanned for one hour in a microPET scanner (Siemens preclinical systems, Knoxville, TN). Data were binned into 28 nonuniform frames following the same sampling schedule as in the computer simulation study and were reconstructed using the MAP algorithm (Qi et al., 1998). We applied the proposed method to the dynamic PET images and compared the results to those from the dynamic clustering approaches with both l^1 and l^2 distance functions. Noise-based

weighting (W1) and feature weighting (W2) were applied to both the proposed method and dynamic clustering. To calculate the feature weighting factors, three spheres were manually selected in the lungs, liver, and background tissues. Three LSFs were used in this experiment and eight clusters were used in the dynamic clustering method. Figure 6 and 7 show the reconstructed images at different frames for the two datasets. The lungs and heart had high activity in frame 5 but very noisy. Liver activity was high in frame 13 with moderate noise. In frame 28, both lungs and liver had low activity and it was hard to distinguish them. The cyan color region inside the lungs in coronal slice 34 at frame 28 is mostly caused by the spill-over effect from the heart.

To obtain the LSF initial, we first identify the position of the heart, kidneys and bladder in each mouse image. Then we applied an affine transformation on the initial image in Figure 5 to match the positions of the above identified targets. With this simple procedure, we do not need to manually draw an initial image for each individual mouse image.

For quantitative evaluation of the segmentation results, we manually delineated the regions of the lungs, liver, kidneys, and spleen in both mouse images and used the manual segmentation as the ground truth. We compared the Dice coefficient for the LSM and the k-means segmentations. We also computed the average TACs in the segmented liver and lung regions, respectively, and compared them with those from the manual segmentation. The relative root mean squared error (RRMSE) was calculated by

$$\text{RRMSE} = \frac{\sqrt{\sum_f [C_{\text{seg}}(f) - C_{\text{ref}}(f)]^2 [t_e(f) - t_s(f)]/T}}{\sum_f C_{\text{ref}}(f) [t_e(f) - t_s(f)]/T} \times 100\% \quad (21)$$

where $C_{\text{seg}}(f)$ and $C_{\text{ref}}(f)$ are the estimated TAC and the reference TAC at frame f , respectively, $t_s(f)$ and $t_e(f)$ are the start time and end time of frame f , respectively, and T is the total scan duration.

4. Results

4.1. Computer Simulation Results: Short Phantom

4.1.1. Visual assessment—Figure 8 shows coronal slice 36 of the original phantom and resulting segments of k-means and LSM with noise weighting scheme (W1) (top row) and feature weighting scheme (W2) (bottom row). Comparing all the results with W1, regions from the LSM were the closest to the true regions. Among all the k-means results, I^1 data term with LSM initial was the best at separating the lungs from other organs, as all other k-means results had the lungs assigned to the same class as the intestine. Kidney regions from all methods had similar results. With the feature weighting scheme, some lung voxels were separated from other regions in the k-means with I^2 data term. However, because the noise variance in the lung region was high in early frames, the boundaries of the lungs were not smooth for the k-means with I^2 data term. The lung region from the k-means with I^1 data term and LSM initial in (k) were closer to the true region than that from the noise-based weighting shown in (e). The LSM results had higher quality and were similar to those with the noise-based weighting.

For low count images, we compare the best results from the k-means (I^1 data term with LSM initial) with the LSM results in Figure 9. The k-means results were visually worse than those from the high count data (Figure 8). The lung region in the k-means results were separated into two classes and it was noisier with the feature weighting because of high noise in the early frames. Part of the lung voxels were assigned to the same class as the intestine because

the lung activity was very close to that of the intestine in the later frames. In comparison, regions from the LSM were visually similar to the results from high count images.

4.1.2. Quantitative comparison—Here we compared the TPF and FPF* of the two methods in Table 1. We used the best k-means results (I^1 data term and LSM initial). However, as we can see from the images in Figure 8, the k-means method classified the liver and the tissue surrounding the kidneys into a single region and assigned the intestine to the same class of the lungs. Such misclassification can be easily corrected by constraining each organ to a certain range of the transaxial slices based on the mouse anatomy. For the lungs, we used the slices above the liver. Without this constraint, the FPF* of the lungs would be more than 1000%. We also combined the two classes for the lung region in the k-means results for a better match to the truth. For the liver, we constrained the slices to be above the kidneys and below the heart. For the spleen region, we used the slices above the bottom of the kidneys and below the liver region. This post-processing procedure substantially improved the k-means results.

According to the results in Table 1, the LSM results were better than the k-means results under both weighting schemes for the lung and liver regions. The feature-based weighting improved the segmentation accuracy of the targeted organs compared with the noise-based weighting for both methods. For the lung region, most LSM results had higher TPF and lower FPF* than the k-means results. The FPF* values of the LSM were less than 5% while some k-means results were over 40%. For the liver region, LSM results were more accurate than the k-means with the FPF* being less than 0.2%, while the FPF* of k-means was up to 6%. For the spleen region, the FPF* of the LSM was less than 5.5% while the FPF* of the k-means method was more than 150%.

Table 2 shows the standard deviation of TPF and FPF* for the LSM calculated from the segmentations of 10 independent noisy reconstructions at 100M events level. All the standard deviations are quite small, which shows that LSM segmentation is fairly robust to different noisy realizations.

4.2. Computer Simulation: Long Phantom

We applied both methods with the same settings to the full phantom which has 10 regions in total. Figure 10 shows the results from the proposed method and compared with the best results from the k-means method. With a limited number of clusters, dynamic clustering failed to separate the lungs and intestine with both weights at all noise levels. Up to four regions were assigned to the heart and bladder due to the spill over effect. The LSM results in similar region definitions as those obtained with the short phantom. While there are only $2^m = 8$ classes available with three LSFs, the post processing step can separate disjointed regions within the same class and thus can result in more than 2^m total segmented organs.

4.3. Comparison of the Dice Coefficient

Table 3 compares the Dice coefficients of the segmentation results of the two methods for the different regions of the short (S) and long (L) phantoms. The k-means results are with the I^1 data term and LSM initial (best case). For kidneys and tumors region, both methods had similar performance. For the lungs and liver region, the Dice coefficients of the LSM were greater than those of the k-means method for all cases. The improvement for the lung region is more than 13% in most cases. For the spleen region, the dice coefficients of the LSM were above 66% while the results of the k-means were less than 30% for all cases.

4.4. Real Mouse Data Results

4.4.1. Visual assessment—Figure 11 shows the segmentation results of the two methods when applied to the real mouse dynamic PET images. For the k-means method, we only show the results with the LSM initials, because they are similar to the best results from 10 random initials. Visually the level set method produced better segmentation results than the k-means method with either weighting scheme, especially for the lungs and liver regions. With the noise-based weighting, the k-means method could not separate the lungs from the liver, while two organs were partially separated with the feature-based weighting. For the level set method, the feature weighting scheme also results in more appropriate and cleaner segmentation for the lungs and liver.

4.4.2. Quantitative comparison—Using the manual segmentation as the ground truth, we computed the Dice coefficients of the LSM and k-means segmentations. The results are shown in Table 4. Only the k-means results were with I^1 data term because the k-means results with I^2 data term were much worse. We performed the same postprocessing procedure for the k-means result as that for the phantom data. The LSM obtained greater Dice coefficients than the k-means for nearly all cases, especially for the liver and spleen. The Dice coefficients of the spleen region from the LSM were between 50% and 70% while the Dice coefficients of the k-means were all less than 10%. The results also show that the feature weighting scheme substantially improved segmentation accuracy of the lung region with a more than 20% increase in the Dice coefficient.

In addition, we computed the TACs of the segmented liver and lung regions for mouse #1 and compared them with the reference TACs obtained from the manual segmentation in Figure 12. The relative root mean squared errors (RRMSE) between the estimated TACs and the reference TACs are shown in Table 5. These results show that the TACs obtained by the LSM methods were much closer to the reference TACs than those obtained by the k-means method with a substantial reduction in the RRMSE. The reduction factor is about 16 (and 1.5) for the lung region with noise weighting (and feature weighting), and is more than 3 for the liver region with either weighting. For both methods the feature weighting scheme (W2) resulted in more accurate TACs with much less mean squared errors than the noise weighting scheme (W1).

5. Discussion

We proposed a level set method with the absolute difference data term to segment mouse dynamic PET images. The spatial regularization imposed by the level set functions encourages each segmented region to have a smooth boundary. Therefore, the resulting regions from the proposed method have a more regular shape than those obtained from the k-means method (Wong et al., 2002). The LMA method (Maroy et al., 2008) used the active contour approach to achieve a similar spatial regularization. For the MOBY phantom, the quality factor (as defined by equation (9) in (Maroy et al., 2008)) of our segmentation results is 73% on average for the organs shown in Table 3 (long phantom), and the average quality factor reported in (Maroy et al., 2008) is 72%. Unfortunately, there are many differences in the simulation setup, including using different PET scanners, different image reconstruction methods, different time activity curves, different organs of interest, etc, that prevent any direct comparison of the results. Our results of the k-means method are also better than those reported in (Maroy et al., 2008). This is most likely due to the post-processing that we performed to constrain each organ to be within a certain transaxial slices based on the mouse anatomy and also to the better resolution of the microPET scanner and partial volume correction in the MAP reconstruction method.

A feature weighting scheme has been developed to improve the segmentation accuracy of certain organs by considering the separation-to-noise ratio in each image frame. It is calculated based on the TACs in manually selected regions inside targeted organs. Thus, it is not a fully automatic procedure, but it can improve the segmentation accuracy of certain organs. For example, we observed a more than 20% increase in the Dice coefficient and more than 3-fold reduction in the RRMSE of the estimated TAC for the lung region by using the feature weighting scheme in the real mouse experiment.

The initial regions of the proposed method also need to be manually defined. For the phantom and real mouse studies, we used a fixed initial image with an affine coregistration procedure to obtain the initial for each mouse image to be segmented. This geometrically defined initial improves the segmentation accuracy by setting a good starting point of the energy function to avoid being trapped in a less favorable local minimum of the energy function.

In the real data experiment, the liver region in mouse #2 was connected to other adjacent tissues for all methods because they shared very similar TACs. Figure 13 shows the TACs of the liver region and the adjacent tissues that being assigned to the same class. Note the difference between the TACs is very small at later frames. To separate adjacent organs with similar TACs, prior information about the organ shapes can be incorporated into the segmentation method. We are currently working along this direction.

6. Conclusion

We have developed a segmentation method that utilizes both spatial and temporal information in dynamic PET data based on a piece-wise constant multi-phase level set method. A weighted absolute difference is used in the data matching term to handle the unequal noise in different image frames and to improve the robustness of the estimates. We also developed a feature weighting scheme for more accurate segmentation of targeted regions. The proposed method was validated by using the MOBY phantom with TACs obtained from a real mouse scan, and was also applied to real mouse data. We compared the proposed approach with the k-means method with l^2 and l^1 data term. We studied the effect of different data terms and the effect of different weighting schemes. Quantitative evaluation from the MOBY phantom data showed that the proposed method achieved higher accuracy than the k-means method. The superior performance of the proposed method is also confirmed by the real mouse data experiment.

Acknowledgments

We would like to thank Dr. William Paul Segars at Duke University for kindly providing the MOBY phantom used in the simulation study and the Center for Molecular and Genomic Imaging at UC Davis for providing the real mouse PET data. We would like to thank anonymous reviewers for their insightful comments and suggestions. We also thank Nannan Cao and Michel Tohme for proofreading early versions of the paper.

This work is supported by the National Institutes of Health under grant no. R01EB000194.

References

- Abufadel, Amer; Yezzi, Tony; Schafer, Ronald. 4D segmentation of cardiac data using active surfaces with spatiotemporal shape priors. *Applied Pattern Recognition, Studies in Computational Intelligence*. 2008; 91/2008:77–100.
- Bezdek JC, Hall LO, Clarke LP. Review of MR image segmentation techniques using pattern recognition. *Medical Physics*. 1993; 20:1033–1048. [PubMed: 8413011]
- Brankov JG, Galatsanos NP, Yang YY, Wernick MN. Segmentation of dynamic PET or fMRI images based on a similarity metric. *IEEE Transactions on Nuclear Science*. 2003; 50(5):1410–1414.

- Chan T, Vese L. Active contours without edges. *IEEE Transation on Image Processing*. 2001; 10:266–277.
- Chan T, Sandberg BY, Vese L. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*. 2000; 11:130–141.
- Cohen L. On active contour models and balloons. *Computer vision, Graphics, and Image Processing*. 1991; 53:211–218.
- Cohen L, Cohen I. Finite element methods for active contour models and balloons for 2D and 3D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1993; 15(11):1131–1147.
- Jennifer G, Carla E. Brodley. Feature selection for unsupervised learning. *Journal of Machine Learning Research*. 2004; 5:845–889.
- Gao S, Bui TD. Image segmentation and selective smoothing by using mumford-shah model. *IEEE Transation on Image Processing*. 2005; 14(10):1537–1549.
- Guo HB, Renaut R, Chen KW, Reiman E. Clustering huge data sets for parametric PET imaging. *Biosystems*. 2003; 71(1–2):81–92. [PubMed: 14568209]
- Kim J, Cai WD, Feng DG, Eberl S. Segmentation of VOI from multidimensional dynamic PET images by integrating spatial and temporal features. *IEEE Transactions on Information Technology in Biomedicine*. 2006; 10(4):637–646. [PubMed: 17044397]
- Kimura Y, Senda M, Alpert NM. Fast formation of statistically reliable FDG parametric images based on clustering and principal components. *Physics in Medicine and Biology*. 2002; 47(3):455–468. [PubMed: 11848122]
- Kohlberger, Timo; Cremers, Daniel; Rousson, Mikal; Ramaraj, Ramamani; Funka-Lea, Gareth. 4D shape priors for a level set segmentation of the left myocardium in SPECT sequences. *Medical Image Computing and Computer-Assisted Intervention*. 2006; LNCS 4190:92–100. [PubMed: 17354878]
- Koivistoinen, H.; Tohka, J.; Ruotsalainen, U. Comparison of pattern classification methods in segmentation of dynamic PET brain images. *Signal Processing Symposium, 2004. NORSIG 2004. Proceedings of the 6th Nordic; 2004*. p. 73–76.
- Lee C, Hun S, Ketter T, Unser M. Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR image. *Computers in Biology and Medicine*. 1998; 28:2437–2446.
- Maitra, Ranjan; O'Sullivan, Finbarr. Variability assessment in positron emission tomography and related generalized deconvolution models. *Journal of the American Statistical Association*. 1998; 93(444):1340–1355.
- Malladi R, Sethian JA. Shape modeling with front propagation: a level set approach. *IEEE Transation On Pattern analysis and Machine Intelligence*. 1995; 17(2):158–176.
- Maroy R, Boisgard RL, Comat C, Frouin V, Cathier P, Duchesnay E, Dolle F, Nielsen PE, Trebossen R, Tavitian B. Segmentation of rodent whole-body dynamic PET images: An unsupervised method based on voxel dynamics. *IEEE Transactions on Medical Imaging*. 2008; 27(3):342–354. [PubMed: 18334430]
- Maroy R, Comtat C, Fontyn Y, Jan S, Trbossen R. Quantitative organ time activity curve extraction from rodent PET images without anatomical prior. *Medical Physics*. 2010; 37:1507–1518. [PubMed: 20443471]
- McInerney T, Terzopoulos D. Deformable models in medical image analysis: A survey. *Medical Image Analysis*. 1996; 1(2):91–108. [PubMed: 9873923]
- Modha, Dharmendra S.; Spangler, W Scott. Feature weighting in k-means clustering. *Machine Learning*. 2003; V52(3):217–237.
- Namba R, Young L, Abbey C, Kim L, Borowsky A, Qi J, Cherry S, Tepper C, Cardiff R, Gregg J. Rapamycin inhibits growth of premalignant and malignant mammary lesions in a mouse model of DCIS. *Clinical Cancer Research*. 2006; 12(8):2613–2621. [PubMed: 16638874]
- Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulation. *Journal of Computational Physics*. 1988; 79:12–49.
- O'Sullivan F. Locally constrained mixture representation of dynamic imaging data from PET and MR studies. *Biostatistics*. 2006; 7(2):318–338. [PubMed: 16361274]
- Pal NR, Pal S. A review on image segmentation techniques. *Pattern Reconition*. 1993; 26:1277–1294.

- Pappas TN. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*. 1992; 40:901–914.
- Paredes, Roberto; Vidal, Enrique. A class-dependent weighted dissimilarity measure for nearest neighbor classification problems. *Pattern Recognition Letters*. 2000; 21(12):1027–1036.
- Parker BJ, Feng DG. Graph-based Mumford-Shah segmentation of dynamic PET with application to input function estimation. *IEEE Transactions on Nuclear Science*. 2005; 52(1):79–89.
- Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. *Annual Review Biomedical Engineering*. 2000; 02:315–337.
- Phelps, ME. *PET: Molecular Imaging and Its Biological Applications*. Springer; 2004.
- Qi J, Leahy RM, Cherry SR, Chatziioannou A, Farquhar TH. High resolution 3D bayesian image reconstruction using the microPET small animal scanner. *Physics in Medicine and Biology*. 1998; 43(4):1001–1013. [PubMed: 9572523]
- Qi, Jinyi; Leahy, RM. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *Medical Imaging, IEEE Transactions on*. 1999; 18(4):293–305.
- Sahoo PK, Soltani S, Wong AKC. A survey of thresholding techniques. *Computer vision, Graphics, and Image Processing*. 1988; 41:233–260.
- Samson C, Blanc-Feraud L, Aubert G, Zerubia J. A level set model for image classification. *International Journal of Computer Vision*. 2000; 40(3):187–197.
- Segars WP, Tsui BMW, Frey EC, Johnson GA, Berr SS. Development of a 4-D digital mouse phantom for molecular imaging research. *Molecular Imaging and Biology*. 2004; 6(3):149–159. [PubMed: 15193249]
- Tahir, Muhammad A.; Bouridane, Ahmed; Kurugollu, Fatih. Simultaneous feature selection and feature weighting using hybrid tabu search/k-nearest neighbor classifier. *Pattern Recognition Letters*. 2007; 28(4):438–446.
- Terzopoulos D, Platt J, Barr A, Fleischer K. Elastically deformable models. *Computer Graphics*. 1987; 21:205–214.
- Vese L, Chan T. Multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*. 2002; 50(3):271–293.
- Wettschereck, Dietrich; Aha, David W.; Mohri, Taka. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*. 1997; V11(1):273–314.
- Wong KP, Feng DG, Meikle SR, Fulham MJ. Segmentation of dynamic PET images using cluster analysis. *IEEE Transactions on Nuclear Science*. 2002; 49(1):200–207.
- Yang, Yongfeng; Rendig, Stephen; Siegel, Stefan; Newport, Danny F.; Cherry, Simon R. Cardiac PET imaging in mice with simultaneous cardiac and respiratory gating. *Physics in Medicine and Biology*. 2005; 50(13):2979–2989. [PubMed: 15972975]
- Yezzi A. A fully global approach to image segmentation via coupled curve evolution equations. *Journal of Visual Communication and Image Representation*. 2002; 13:195–216.

Figure 1.

The digital phantom: coronal slices 36 (left) and 49 (right). From the top to bottom, the regions are the lungs (mid blue), myocardium (orange, right figure only), liver (purple), stomach (dark red), spleen (red, left figure only), kidneys (yellow, left figure only), intestine (cyan), bladder (orange, right figure only), and tumors (red, right figure only). The voxels in navy are partial volume voxels.

Figure 2.
TACs for the regions in the simulation phantom.

Figure 3. Reconstructed noisy phantom images in coronal slices 36 (top row) and 49 (bottom row) at frame 5 (left), 13 (middle) and 28 (right), respectively. The total count is 500M.

Figure 4.
Plot of weighting factors for different frames.

Figure 5. Initial regions of the LSM for the phantom data. Left: coronal slice 36; right: coronal slice 49. The short phantom is the portion above the black line.

Figure 6. Real mouse images (mouse #1) at different frames. Coronal slices 34 (top) and 48 (bottom).

Figure 7.
Real mouse images (mouse #2) at different frames. Coronal slices 34 (top) and 48 (bottom).

Figure 8.

True regions and resulting segments of slice 36 using the noise (top) and feature (bottom) weighting, 500M events.

(a) True regions (b) k-means W1 (c) k-means W2 (d) LSM W1 (e) LSM W2

Figure 9.
True regions and resulting segments of slice 36, 100M events.

Figure 10. True regions and resulting segments of slice 36 (top) and 49 (top) for the long phantom, 100M events.

(a) k-means, l^2 , W1 (b) k-means, l^1 , W1 (c) LSM, W1 (d) k-means, l^2 , W2 (e) k-means, l^1 , W2 (f) LSM, W2

(g) k-means, l^2 , W1 (h) k-means, l^1 , W1 (i) LSM, W1 (j) k-means, l^2 , W2 (k) k-means, l^1 , W2 (l) LSM, W2

Figure 11.

Surface rendering of the segmentation results of the real mouse data. Top row: animal #1; Bottom row: animal #2. In each image, the top red region is the heart, the black region is the lungs, the cyan region is the liver, yellow regions are the kidneys, and the bottom red region is the bladder.

Figure 12.
TACs of the lung and liver regions obtained by different segmentation methods for mouse #1.

Figure 13.
TACs of the liver and the adjacent tissues of mouse #2.

Table 1

TPF and FPF* of each region (short phantom) (%)

Region (#)	count		k-means W1	k-means W2	LSM W1	LSM W2
Lungs (2023)	500M	TPF	90.4	92.7	87.4	89.0
		FPF*	51.2	6.5	0.9	0.7
	100M	TPF	88.8	90.2	85.7	88.3
		FPF*	34.4	31.0	1.2	1.1
	50M	TPF	72.7	74.1	82.3	77.8
		FPF*	83.3	94.3	1.8	0.6
Liver (15465)	500M	TPF	91.5	99.5	98.3	99.4
		FPF*	0.5	0.5	0.2	0.2
	100M	TPF	95.9	96.2	99.2	99.3
		FPF*	0.6	0.2	0.2	0.1
	50M	TPF	98.9	98.5	97.2	98.1
		FPF*	6.4	5.9	0.2	0.2
Kidneys (2147)	500M	TPF	99.4	99.4	99.7	99.4
		FPF*	0.0	0.0	0.0	0.0
	100M	TPF	99.9	99.9	99.9	99.8
		FPF*	0.0	0.0	0.0	0.0
	50M	TPF	1.1	1.1	99.7	99.7
		FPF*	0	0	0	0
Spleen (429)	500M	TPF	99.8	94.2	93.3	94.2
		FPF*	240	1400	0	0
	100M	TPF	99.7	84.3	77.8	73.1
		FPF*	160	250	1.7	2.2
	50M	TPF	78.9	71.2	63.9	61.7
		FPF*	300	1800	3.5	5.2

Table 2

Standard Deviation of the TPF and FPF* from 10 independent noisy reconstructions (short phantom).

Region (True vox.)	TPF (%)		FPF * (%)	
	W1	W2	W1	W2
Lungs (2023)	85.7±1.1	88.3±1.6	1.2±0.6	1.1±0.5
Liver (15465)	99.2±0.1	99.3±0.1	0.2±0.2	0.1±0.0
Kidneys (2147)	99.9±0.0	99.8±0.0	0.0±0.0	0.0±0.0
Spleens (429)	77.8 ± 2.4	73.1± 1.1	1.7 ± 1.1	2.15 ± 0.6

Table 3

Dice coefficient (%) of each region in the phantom.

Region	count	k-means W1	k-means W2	LSM W1	LSM W2
Lungs	500M	58.8	70.7	73.3	74.1
	100M (S)	55.4 ± 0.2	57.2 ± 6.9	75.1 ± 1.3	77.4 ± 0.7
	100M (L)	62.1	56.9	74.5	77.8
Liver	50M	60.9	57.1	75.6	74.3
	500M	87.5	87.6	94.5	92.9
	100M (S)	90.3 ± 0.1	90.4 ± 1.2	92.5 ± 0.1	92.8 ± 0.2
	100M (L)	85.3	87.1	92.2	92.8
	50M	84.6	86.9	91.4	91.3
	500M	93.5	93.3	92.5	93.5
Kidneys	100M (S)	91.6 ± 0	91.6 ± 0	90.8 ± 0	91.1 ± 0
	100M (L)	96.3	93.8	90.1	91.3
	50M	90.3	90.7	91.1	91.3
Spleen	500M	25.7	7.6	84.1	82.2
	100M (S)	29.7 ± 9.5	21.6 ± 14.5	77.2 ± 1.7	73.6 ± 0.8
	100M (L)	23.8	23.8	83.3	79.6
Tumor	50M	7.1	9.2	66.9	70.9
	100M(L)	87.1	89.4	91.4	92.5

Table 4

Dice coefficient (%) of each region for the real mouse data.

Region	mouse	k-means, I^1 , W1	k-means, I^1 , W2	LSM W1	LSM W2
Lungs	#1	52.7	70.3	45.3	73.8
	#2	28.5	70.1	30.3	72.2
Liver	#1	78.3	66.8	83.8	83.2
	#2	65.8	65.1	72.2	75.2
Kidneys	#1	87.1	88.7	87.6	88.3
	#2	56.3	87.6	72.2	85.5
Spleen	#1	8.4	7.0	54.3	71.1
	#2	4.0	2.9	51.9	51.6

Table 5

The relative root mean squared errors (RRMSE) between the estimated TACs and the reference TACs for the liver and lung regions (%).

Region	K-means, l^1 , W1	K-means, l^1 , W2	LSM W1	LSM W2
Lung	229	5.2	13.6	3.5
Liver	14.7	8.1	4.7	2.5