High Temporal-Resolution Dynamic PET Image Reconstruction Using A New Spatiotemporal Kernel Method

Guobao Wang

***Abstract*— Current clinical dynamic PET has an effective temporal resolution of 5-10 seconds, which can be adequate for traditional compartmental modeling but is inadequate for exploiting the benefit of more advanced tracer kinetic modeling. There is a need to improve dynamic PET to allow 1-second temporal sampling. However, reconstruction of these short-time frames from tomographic data is extremely challenging as the count level of each frame is very low and high noise presents in both spatial and temporal domains. Previously the kernel framework has been demonstrated as a statistically efficient approach to utilizing image prior for low-count image reconstruction. Nevertheless, the existing kernel methods mainly explore spatial correlations in the data and only have a limited ability in suppressing temporal noise. In this paper, we propose a new kernel method which extends the previous spatial kernel method to the general spatiotemporal domain. The new kernelized model encodes the spatiotemporal correlations obtained from image prior information and is incorporated into the PET forward projection model to improve the maximum likelihood (ML) image reconstruction. Computer simulations show that the proposed approach can achieve effective noise reduction in both spatial and temporal domains and outperform the spatial kernel method and conventional ML reconstruction method for improving high temporal-resolution dynamic PET imaging.**

***Index Terms*— High temporal resolution, dynamic PET, image reconstruction, maximum likelihood, image prior, kernel method, spatiotemporal correlation**

I. Introduction

Dynamic positron emission tomography (PET) can monitor spatiotemporal distribution of a radiotracer in human body. With tracer kinetic modeling, dynamic PET is capable of quantifying physiologically or biochemically important parameters in regions of interest or voxelwise to detect disease status and characterize severity. Traditionally compartmental modeling is used for kinetic analysis of dynamic PET data. Other advanced tracer kinetic models [1] such as the distributed-parameter model and its approximations are considered closer to the physiological process than compartmental models. However, those models have not been well explored in dynamic PET because the effective temporal resolution of clinical dynamic PET has been limited to 5-10 seconds with already-compromised image quality [2].

We aim to improve the effective temporal resolution of clinical dynamic PET to 1-2 seconds to value the use of advanced kinetic modeling in PET. To achieve the high temporal resolution, short scan durations must be used, which however results in very low counting statistics in the dynamic frames. Image reconstruction from the low-count projection data is extremely challenging because tomography is ill-posed and high noise exists in tomographic measurement.

In PET, incorporation of image prior information into image reconstruction has become a popular means to improve the quality of reconstructed images [3]. The prior information can be either local smoothness of neighboring pixels or obtained from a co-registered anatomical MRI or CT image. While most of existing reconstruction methods employ an explicit regularization form to incorporate image prior and can be complex for practical implementation, the recent kernel method [4]–[6] encodes image prior information in the forward model of PET image reconstruction and requires no explicit regularization. The kernel method is easier to implement and can be more efficient and better improve image reconstruction than regularization-based methods [4], [5].

Existing kernel methods [4]–[6] mainly explore spatial correlations of image pixels to improve image quality in the spatial domain. These spatial kernels, however, have a limited ability in suppressing noise in the temporal domain. In high temporal-resolution dynamic PET imaging, noise variation in the temporal domain can be very severe because many short-time frames are used. It is therefore desirable to include temporal prior knowledge in the kernel method to suppress temporal noise. In this paper, we extend the spatial kernel method to a spatiotemporal kernel method that allows both spatial and temporal correlations to be encoded in the kernel matrix. We propose a separable spatiotemporal kernel to make the method more computationally tractable and easier use. The new method is expected to achieve substantial noise reduction in temporal domain in addition to the enhancement on image quality by existing spatial kernels.

II. Theory

A. Dynamic PET Image Reconstruction

For a time frame m, we denote the PET image intensity value at pixel j by $x_{j,m}$ and the measurement in detector pair i by $y_{i,m}$. The expectation of the dynamic projection data $\bar{\boldsymbol{y}} = \{\bar{y}_{i,m}\}$ is related to the unknown dynamic image $\boldsymbol{x} = \{x_{j,m}\}$ through

$$\bar{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{r} \tag{1}$$

where $\boldsymbol{P}$ is the detection probability matrix for dynamic PET and $\boldsymbol{r}$ is the expectation of dynamic random and scattered

GB Wang is with the Department of Radiology, School of Medicine, University of California at Davis, Sacramento, CA 95817, USA. Email: gbwang@ucdavis.edu

DOI: 10.12059/Fully3D.2017-11-3103002

events [3].

Dynamic PET projection measurement $\boldsymbol{y} = \{y_{i,m}\}$ can be well modeled as independent Poisson random variables with the log-likelihood function [3],

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n_i}\sum_{m=1}^{M} y_{i,m}\log\bar{y}_{i,m} - \bar{y}_{i,m} - \log y_{i,m}!, \quad (2)$$

where n_i is the total number of detector pairs and M is the total number of time frames. The maximum likelihood (ML) estimate of the dynamic image $\boldsymbol{x}$ is found by maximizing the Poisson log-likelihood,

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\geq\boldsymbol{0}} L(\boldsymbol{y}|\boldsymbol{x}). \quad (3)$$

The expectation-maximization (EM) algorithm [7] with the following iterative update

$$\boldsymbol{x}^{n+1} = \frac{\boldsymbol{x}^n}{\boldsymbol{P}^T\boldsymbol{1}_N}\cdot\big(\boldsymbol{P}^T\frac{\boldsymbol{y}}{\boldsymbol{P}\boldsymbol{x}^n+\boldsymbol{r}}\big), \quad (4)$$

is often the choice to find the solution, where $\boldsymbol{1}_N$ is a vector of length $N = n_i \times M$ with all elements being 1, n denotes iteration number and the superscript "T" denotes matrix transpose. The vector multiplication and division are element-wise operations.

B. The Spatiotemporal Kernel Method

The kernel method for tomographic image reconstruction [4] was inspired by the kernel methods for classification and regression in machine learning. Different from the kernel methods in machine learning, the kernel method for image reconstruction has unknown "label" values and the available data for kernel coefficient estimation is the tomographic projection data. Previously the kernel method [4] was derived for frame-by-frame spatial image reconstruction, here we adapt the expressions for spatiotemporal reconstruction.

In machine learning language, the image intensity $x_{j,m}$ at pixel j in time frame m is the "label" value. For each spatiotemporal location, a set of features are identified to form the feature vector $\boldsymbol{f}_{j,m}$, which is also called a "data point" in machine learning. A mapping function $\boldsymbol{\phi}(\boldsymbol{f}_{j,m})$ is then used to transform the data points $\{\boldsymbol{f}_{j,m}\}$ into a feature space of very-high dimension $\{\boldsymbol{\phi}(\boldsymbol{f}_{j,m})\}$. By doing this, the "label" value $x_{j,m}$ can be better described as a linear function in the high-dimensional feature space,

$$x_{j,m} = \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{f}_{j,m}) \quad (5)$$

where $\boldsymbol{w}$ is a weight vector which also sits in the transformed space:

$$\boldsymbol{w} = \sum_{j'=1}^{n_j}\sum_{m'=1}^{M}\alpha_{j',m'}\boldsymbol{\phi}(\boldsymbol{f}_{j',m'}) \quad (6)$$

with $\boldsymbol{\alpha}$ being the coefficient vector. n_j is the number of pixels in image. By substituting (6) into (5), the kernel representation for the image intensity at pixel j and in time frame m is written as

$$x_{j,m} = \sum_{j'=1}^{n_j}\sum_{m'=1}^{M}\alpha_{j',m'}\boldsymbol{\phi}(\boldsymbol{f}_{j',m'})^T\boldsymbol{\phi}(\boldsymbol{f}_{j,m}) \quad (7)$$

$$= \sum_{j'=1}^{n_j}\sum_{m'=1}^{M}\alpha_{j',m'}\kappa(\boldsymbol{f}_{j,m},\boldsymbol{f}_{j',m'}), \quad (8)$$

where

$$\kappa(\boldsymbol{f}_{j,m},\boldsymbol{f}_{j',m'}) \triangleq \boldsymbol{\phi}(\boldsymbol{f}_{j,m})^T\boldsymbol{\phi}(\boldsymbol{f}_{j',m'}) \quad (9)$$

is a kernel defined by the kernel function $\kappa(\cdot,\cdot)$ (e.g. radial Gaussian function). The mapping function $\boldsymbol{\phi}$ is now implicitly defined by the kernel and not required to be known. The image intensity $x_{j,m}$ at pixel j in time frame m is thus described as a linear function in the kernel space but is nonlinear in the original space of the data points $\{\boldsymbol{f}_{j,m}\}$. With $\boldsymbol{x}$ denoting the dynamic image and $\boldsymbol{K}$ the spatiotemporal kernel matrix, The equivalent matrix-vector form of (8) is

$$\boldsymbol{x} = \boldsymbol{K}\boldsymbol{\alpha}. \quad (10)$$

where $\boldsymbol{\alpha} \triangleq \{\alpha_{j,m}\}$ denotes the kernel coefficient vector.

Substituting the kernelized image model (10) into the standard PET forward projection model (1), we obtain the following kernelized forward projection model for dynamic PET image reconstruction:

$$\bar{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{K}\boldsymbol{\alpha} + \boldsymbol{r}. \quad (11)$$

The advantage of using this kernelized model (11) is that image prior knowledge can be incorporated in the forward projection to improve the reconstruction of low-count scans.

A kernelized EM algorithm can be easily derived [4]. The EM update of $\boldsymbol{\alpha}$ at iteration $(n+1)$ is

$$\boldsymbol{\alpha}^{n+1} = \frac{\boldsymbol{\alpha}^n}{\boldsymbol{K}^T\boldsymbol{P}^T\boldsymbol{1}_N}\cdot\big(\boldsymbol{K}^T\boldsymbol{P}^T\frac{\boldsymbol{y}}{\boldsymbol{P}\boldsymbol{K}\boldsymbol{\alpha}^n+\boldsymbol{r}}\big). \quad (12)$$

Once the coefficient image $\boldsymbol{\alpha}$ is estimated, the reconstructed dynamic PET image is calculated by

$$\hat{\boldsymbol{x}} = \boldsymbol{K}\hat{\boldsymbol{\alpha}}. \quad (13)$$

C. Separable Spatiotemporal Kernel

The kernel matrix $\boldsymbol{K}$ encodes image prior information based on the feature vectors $\{\boldsymbol{f}_{j,m}\}$. For each pixel j in time frame m, we identify a set of features to form $\boldsymbol{f}_{j,m}$,

$$\boldsymbol{f}_{j,m} = [(\boldsymbol{f}_j^s)^T, (\boldsymbol{f}_m^t)^T]^T, \quad (14)$$

where the vector consists of two components. $\boldsymbol{f}_j^s$ is the vector for exploring spatial correlations between pixels and $\boldsymbol{f}_j^t$ is for exploring temporal correlations between frames.

We further define the spatiotemporal kernel function $\kappa(\boldsymbol{f}_{j,m},\boldsymbol{f}_{j',m'})$ to be spatially and temporally separable, i.e.

$$\kappa(\boldsymbol{f}_{j,m},\boldsymbol{f}_{j',m'}) = \kappa_s(\boldsymbol{f}_j^s,\boldsymbol{f}_{j'}^s)\kappa_t(\boldsymbol{f}_m^t,\boldsymbol{f}_{m'}^t) \quad (15)$$

where $\kappa_s(\cdot,\cdot)$ denotes the kernel function for calculating spatial correlations and $\kappa_t(\cdot,\cdot)$ is for calculating temporal correlations.

Thus the overall spatiotemporal kernel matrix $\boldsymbol{K}$ is decoupled into a spatial kernel matrix $\boldsymbol{K}_s \in I\!R^{n_j \times n_j}$ and a temporal kernel matrix $\boldsymbol{K}_t \in I\!R^{M \times M}$,

$$\boldsymbol{K} = \boldsymbol{K}_t \otimes \boldsymbol{K}_s, \tag{16}$$

where $\otimes$ represents the Kronecker product.

Derivation of the spatial kernel matrix $\boldsymbol{K}_s$ has been developed in our previous work [4]. $\boldsymbol{K}_s$ is often formed as a sparse matrix based on image prior data. For obtaining image prior in dynamic PET, an effective and efficient means is to use composite frames [4]. For example, an one-hour dynamic FDG-PET scan can be first rebinned into three composite frames, each with 20 minutes. From the reconstructed composite images, three time activity points are obtained at each pixel j and used as the feature vector $\boldsymbol{f}_j^s$ to construct the kernel matrix.

The spatial kernel method [4] is a special example of the spatiotemporal kernel method with the temporal kernel matrix set to the identity matrix $\boldsymbol{K}_t = \boldsymbol{I}_M$. In this paper, we explore the role of $\boldsymbol{K}_t$ in the context of high temporal-resolution dynamic PET imaging. The (m, m')th element of $\boldsymbol{K}_t$ is obtained by comparing the feature vectors of the frames m and m':

$$\kappa_t(\boldsymbol{f}_m^t, \boldsymbol{f}_{m'}^t) = \begin{cases} \exp\left(-\frac{||\boldsymbol{f}_m^t - \boldsymbol{f}_{m'}^t||^2}{2\sigma_t^2}\right), & |\tau_m - \tau_m'| < d_t, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

where τ_m denotes the middle time point of frame m and d_t is the width of time window for neighborhood construction. σ_t is a parameter to adjust the weight calculation.

The simplest form of the temporal feature $\boldsymbol{f}_m^t$ is probably

$$\boldsymbol{f}_m^t \triangleq \tau_m \tag{18}$$

by which $\boldsymbol{K}_t$ becomes a shift-invariant Gaussian smoothing kernel. This type of kernel can smooth out noise but may also over-smooth sharp signals in temporal domain. To make the temporal kernel more adaptive to time varying data, we propose to use the whole sinogram of each frame as the feature vector to capture temporal correlations between frames, i.e.

$$\boldsymbol{f}_m^t \triangleq \tilde{\boldsymbol{y}}_m \tag{19}$$

where $\tilde{\boldsymbol{y}}_m$ denotes a smoothed version of the raw sinogram $\boldsymbol{y}_m$ of frame m.

As both $\boldsymbol{K}_s$ and $\boldsymbol{K}_t$ are very sparse, inclusion of the kernel matrix $\boldsymbol{K}$ in the projection model does not add a significant computational cost in the reconstruction.

III. Simulation Results

Dynamic PET scans were simulated for a GE DST whole-body PET scanner using a Zubal head phantom (Figure 1). The scanning schedule consisted of 91 time frames over 60 minutes: 60×1 s, 12×5 s, 3×40 s, 4×60 s, 4×180 s and 8×300 s. Regional time activity curves were assigned to different brain regions. Dynamic activity images were first forward projected to generate noise-free sinograms. A 20% uniform background was included to simulate random and scattered events. Poisson noise was then generated with 50 million expected events over 60 minutes.

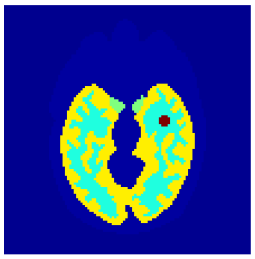


Fig. 1. The digital Zubal phantom used in the simulation studies is composed of gray matter, white matter and a tumor (15mm in diameter).

The noisy sinograms were reconstructed independently by four different image reconstruction methods: the traditional MLEM method, spatial kernelized EM (KEM-S), and new spatiotemporal kernel method with the Gaussian smoothing temporal kernel (KEM-ST-G, defined using Eq. (18)) and the spatiotemporal kernel method with the data-driven temporal kernel (KEM-ST-D, defined using Eq. (19)). The spatial kernel matrix was constructed using the k nearest neighbor (kNN) approach described in [4] and four composite images (1×5-minute, 1×15- minute, 2×20-minute) which were obtained by recombining the full 60-minute dynamic data. The time window used for temporal kernel matrix was set to $d_t = 15$s. Signal-to-noise ratio (SNR) was used as the figure of merit to evaluate image quality.

As our goal is to compare the reconstruction methods for high temporal resolution dynamic PET imaging, here we only show the results for the first one minute (60 1-second frames) of the dynamic scan. This early phase has very fast kinetics and is challenging to reconstruct.

Figure 2 shows the true activity image and reconstructed images by the four different methods for the 1-s time frame at $t = 25$s. We also rebinned the 1-s dynamic frames to generate 10-s frames to compare the effect of scan durations. The reconstruction of the 10-s frame centered at $t = 25$s by MLEM is shown in Figure 2 (f). As expected, the spatial kernel method (KEM-S) achieved a significant SNR increase compared with the MLEM method. By incorporating temporal correlations, the spatiotemporal kernel methods (KEM-ST-G and KEM-ST-D) further improved the image quality with higher SNR values than KEM-S. The MLEM of the 10-s frame has a much higher SNR than the MLEM of the 1-s frame. The SNR improvement, however, is still much lower than all the KEM reconstructions.

Figure 3 shows the plots of image SNR of all the 60 1-s time frames for different methods. The SNR of each frame is maximized over iteration numbers in different methods. The three KEM methods outperformed the MLEM reconstruction for all 1-s frames and also had higher SNR than the MLEM of 10-s frames. Compared with the spatial kernel method KEM-S, the spatiotemporal kernel methods KEM-ST-G and KEM-ST-D improved the late frames. This is because the TACs of these frames are lower and more flat than early frames (see Figure 4). Incorporation of temporal correlations thus became beneficial and achieved noise reduction. KEM-ST-D and KEM-ST-G had similar performance given the two temporal kernels were close to each other in these late frames.

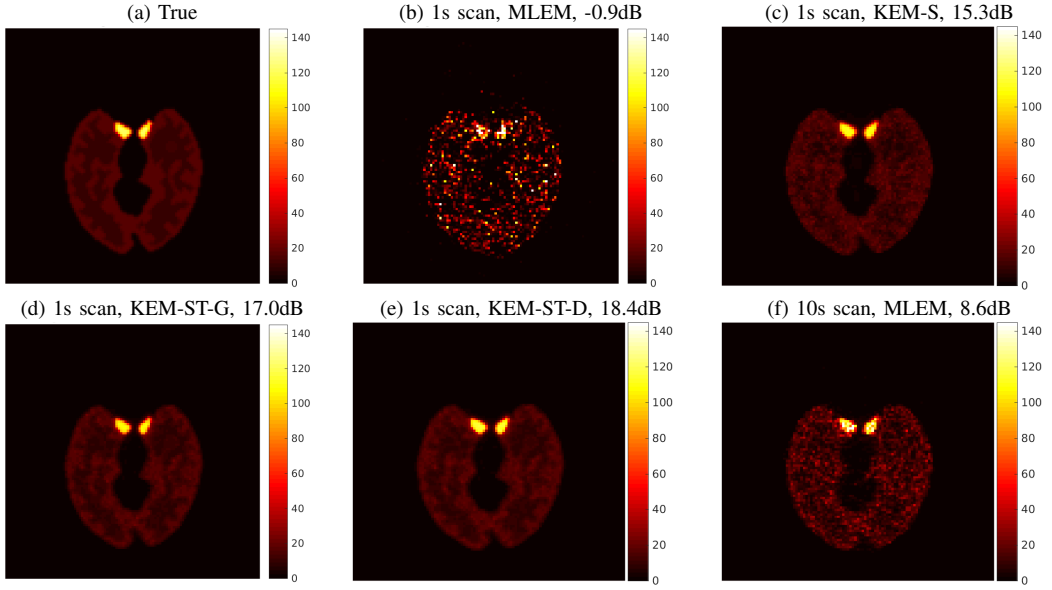


Fig. 2. True activity image at $t = 25$s and comparison of different reconstructions at iteration 50. (1) True image of the 1-s time frame at $t = 25$s; (b-e) reconstructed images by the four different methods: (b) MLEM, (c) spatial kernel method (KEM-S), (d) spatiotemporal kernel method with the time-invariant Gaussian smooth kernel (KEM-ST-G), (e) spatiotemporal kernel method with a data-driven temporal kernel matrix (KEM-ST-D), and (f) MLEM reconstruction of an extended 10-s scan at $t = 25$s. Image SNR in dB is included with each image.

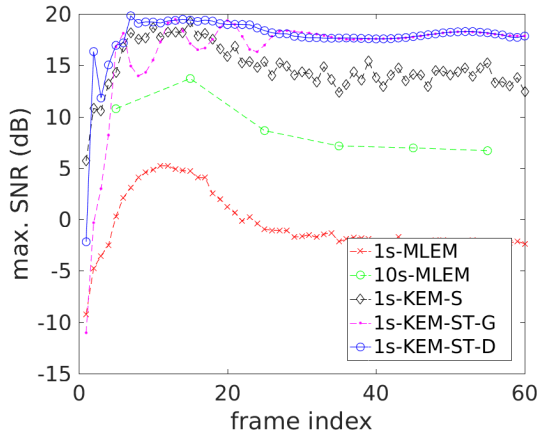


Fig. 3. Plots of Image SNR of 60 1-s time frames in different methods.

In the early frames where activity change is fast, use of the time invariant temporal kernel in the KEM-ST-G method over-smoothed the temporal signals. Thus KEM-ST-G had even lower SNR than KEM-S. In comparison, KEM-ST-D had data-adaptive temporal kernels and achieved the best overall results.

Figure 4 demonstrates the MLEM and KEM reconstructions of high temporal resolution (1s) TACs for the blood and tumor regions. The low-resolution (10s) TACs by MLEM were also included. The 1-s TACs by MLEM captured finer temporal changes but generally had higher noise than the 10-s reconstructions, especially in the tumor region. The noise in the blood region was relatively small because its activity was much higher than other regions. The spatial kernel method KEM-S was unable to suppress temporal noise, though it had achieved a significant noise reduction in spatial domain. In comparison, the new spatiotemporal kernel method (KEM-ST-D) overcame the limitation of MLEM and KEM-S and achieved high temporal resolution and low noise in the reconstructed TACs.

IV. Conclusion

In this paper, we proposed a spatiotemporal kernel method to incorporate both spatial and temporal prior information into the kernel framework for dynamic PET image reconstruction. The spatiotemporal kernel is separable in the spatial and temporal domains and thus can be easily and efficiently implemented. We conducted a computer simulation to validate the method and compared different temporal kernel matrices. The results demonstrated the new kernel method outperformed existing methods and achieved high temporal-resolution while maintaining noise at a low level. Future work will include comparisons with regularization-based reconstruction methods and application of the method to patient data.

References

[1] St Lawrence KS, Lee TY, "An Adiabatic Approximation to the Tissue Homogeneity Model for Water Exchange in the Brain: I. Theoretical Derivation," *Journal of Cerebral Blood Flow & Metabolism*, vol. 18, no. 12, pp. 1365-1377, 1998.

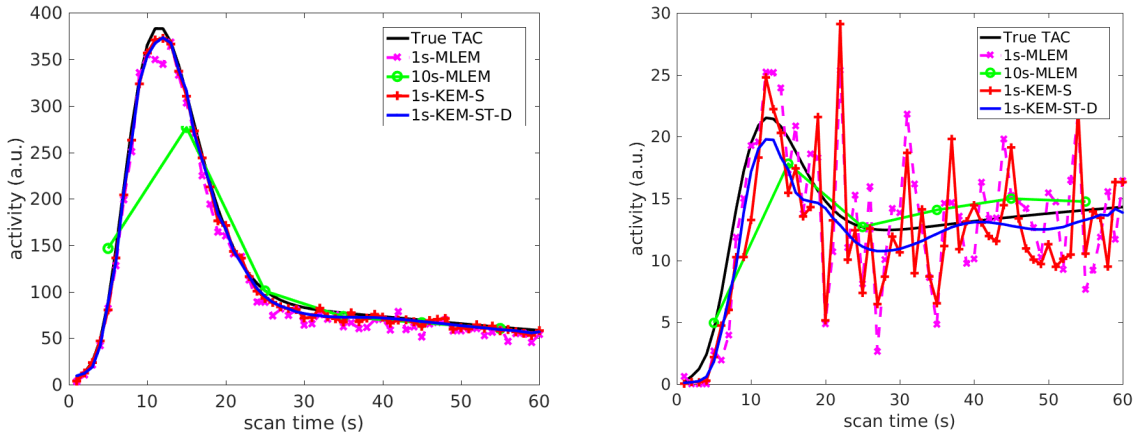
[2] Muzic RF, Saidel GM, "Distributed versus compartment models for PET receptor studies," *IEEETrans. Med. Imag.*, vol. 22, no. 1, pp. 11-21, 2003.

[3] J. Qi and R.M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Physics in Medicine and Biology*, vol. 51, no. 15, pp. R541-R578, 2006.

[4] Wang G, and Qi J, "PET Image Reconstruction Using Kernel Method," *IEEETrans. Med. Imag.*, vol. 34, no. 1, pp. 61-71, 2015

[5] Hutchcroft W, Wang GB , Chen K, Catana C, Qi J, "Anatomically-aided PET reconstruction using the kernel method, Physics in Medicine and Biology," *Physics in Medicine and Biology*, 61(18): 6668-6683, 2016.

[6] Novosad P and Reader AJ, "MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions," *Physics in Medicine and Biology*, 61(12): 46244645, 2016



(a) Blood region (b) Tumor region

Fig. 4. Time activity curves reconstructed by MLEM and KEM. (a) Blood region, (b) tumor region.

[7] L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction in positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. MI-1,pp. 113-122, 1982.

V. Acknowledgment

This work is supported in part by NIH R21 HL 131385 and AHA BGIA 25780046.