Published in final edited form as:
IEEE Trans Med Imaging. 2012 December ; 31(12): 2194–2204. doi:10.1109/TMI.2012.2211378.

Penalized Likelihood PET Image Reconstruction using Patch-based Edge-preserving Regularization

Guobao Wang [Senior Member, IEEE] and **Jinyi Qi [Senior Member, IEEE]**
Department of Biomedical Engineering, University of California, Davis, CA 95616, USA

Guobao Wang: gbwang@ucdavis.edu; Jinyi Qi: qi@ucdavis.edu

Abstract

Iterative image reconstruction for positron emission tomography (PET) can improve image quality by using spatial regularization that penalizes image intensity difference between neighboring pixels. The most commonly used quadratic penalty often over-smoothes edges and fine features in reconstructed images. Non-quadratic penalties can preserve edges but often introduce piece-wise constant blocky artifacts and the results are also sensitive to the hyper-parameter that controls the shape of the penalty function. This paper presents a patch-based regularization for iterative image reconstruction that uses neighborhood patches instead of individual pixels in computing the non-quadratic penalty. The new regularization is more robust than the conventional pixel-based regularization in differentiating sharp edges from random fluctuations due to noise. An optimization transfer algorithm is developed for the penalized maximum likelihood estimation. Each iteration of the algorithm can be implemented in three simple steps: an EM-like image update, an image smoothing and a pixel-by-pixel image fusion. Computer simulations show that the proposed patch-based regularization can achieve higher contrast recovery for small objects without increasing background variation compared with the quadratic regularization. The reconstruction is also more robust to the hyper-parameter than conventional pixel-based non-quadratic regularizations. The proposed regularization method has been applied to real 3D PET data.

Index Terms

Image reconstruction; penalized maximum likelihood; patch regularization; positron emission tomography

I. Introduction

Iterative reconstruction techniques have been increasingly used in positron emission tomography (PET) to improve image quality [1]. Comparing to analytical reconstruction (e.g. filtered backprojection), iterative reconstruction can make use of noise statistics, accurate system modeling, and image prior knowledge. Maximum likelihood (ML) reconstruction method estimates image from projections by maximizing the log likelihood of PET data. However, a true ML solution can be very noisy, so some form of regularization is necessary to stabilize the image estimation. It is commonly achieved by either terminating the iteration before convergence or by using a penalty function to encourage spatially smooth images. Here we focus on the latter approach. An ideal regularization function

should be able to preserve two conflicting properties of PET images: spatial smoothness within a region and sharp transition at region boundaries. However, the success has been limited with existing regularization functions. The most commonly used quadratic penalty often over-smoothes edges and small objects in reconstructed images. Non-quadratic penalties can preserve edges but often introduce piece-wise constant blocky artifacts and the results can be sensitive to the value of the hyper-parameter that controls the shape of the penalty function.

We postulate that the main reason for the instability of traditional non-quadratic penalties is that the image roughness is calculated based on the intensity difference between neighboring pixels. When an image is noisy, pixel intensity differences are not reliable in distinguishing real edges from noisy fluctuation. To overcome this problem, we propose a patch-based regularization that uses neighborhood patches instead of individual pixels in measuring image roughness. Since it compares the similarity between patches, the patch-based regularization is expected to be more robust than pixel-based regularizations. In addition to being edge preserving, patch-based regularization can also preserve features or texture within patches, which may have important applications beyond PET image reconstruction.

The patch-based regularization presented in this paper is closely related to the nonlocal regularization that has been studied in the context of image restoration [2]–[8] and image reconstruction [9]–[12]. However, existing nonlocal regularization methods either require a pre-known reference image for constructing the weight function or involve a nonconvex optimization (see Section III for more discussions). In image reconstruction, a good reference image is not directly available before reconstruction. A nonconvex optimization often suffers from unstable image estimation because of multiple local solutions. In comparison, the proposed regularization does not require a reference image before reconstruction and has a convex objective function, which allows us to derive an optimization transfer algorithm that guarantees convergence to the penalized likelihood solution. Part of this work was previously presented at the 2011 International Symposium of Biomedical Imaging [13]. Yang and Jacob also presented a similar work at the same conference but focused on nonconvex penalty functions [14] and applications in magnetic resonance imaging (MRI).

II. Penalized Likelihood Reconstruction

PET data $\boldsymbol{y} = \{y_i\}$ can be well modeled as a collection of independent Poisson random variables with the log likelihood function as

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n_i} y_i \log \bar{y}_i - \bar{y}_i. \quad (1)$$

The expected data $\bar{\boldsymbol{y}} = \{\bar{y}_i\}$ is related to the unknown image $\boldsymbol{x}$ through an affine transform

$$\bar{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{r} \quad (2)$$

where $\boldsymbol{P} = \{p_{ij}\} \in IR^{n_i \times n_j}$ is the system matrix with p_{ij} denoting the probability of detecting an event originated at pixel j by detector pair i, $\boldsymbol{r}$ accounts for background events such as randoms and scatters. n_i is the total number of detector pairs and n_j is the total number of pixels in image.

Penalized likelihood (PL) reconstruction (or equivalently maximum *a posteriori*, MAP) estimates the unknown image by maximizing a penalized likelihood function

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}>0} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) = L(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x}) \quad (3)$$

where $U(\boldsymbol{x})$ is an image roughness penalty. Conventionally the image roughness is measured based on the intensity difference between neighboring pixels

$$U(\boldsymbol{x}) = \frac{1}{4}\sum_{j=1}^{n_j}\sum_{k\in\mathcal{N}_j} w_{jk}\psi(x_j - x_k) \quad (4)$$

where $\psi(t)$ is the penalty function and w_{jk} is the weighting factor related to the distance between pixel j and pixel k in the neighborhood N_j. The regularization parameter β controls the trade-off between data fidelity and spatial smoothness. When β goes to zero, the reconstructed image approaches the ML estimate.

A common choice of $\psi(t)$ in PET image reconstruction is the quadratic function

$$\psi(t) = \frac{1}{2}t^2. \quad (5)$$

A disadvantage of the quadratic regularization is that it may over-smooth edges and small objects when a large β is used to smooth out noise in large regions. A penalty function that can preserve edges is the absolute value function

$$\psi(t) = |t \quad (6)$$

which is not differentiable at zero. A similar function but with continuous second-order derivatives is the Lange function [15]

$$\psi(t) = \delta\left(\frac{|t|}{\delta} - \log\left(1 + \frac{|t|}{\delta}\right)\right). \quad (7)$$

which approximates the quadratic function when $|t| \ll \delta$ and approaches the absolute function for $|t| \gg \delta$. Other examples of non-quadratic convex penalty functions include the hyperbola function $\sqrt{t^2+\delta^2}$, the Huber function [16] and the l_p penalty ($p \geq 1$) [17]. Non-convex penalty functions (e.g. [18], [19]) have also been proposed to form even sharper edges in reconstructed images. One problem with non-quadratic penalty functions is that

reconstructed images are often very sensitive to the hyper-parameter δ and there has been no practical method to select the optimum value.

III. Patch-based Regularization

A. Definition

We propose to use a patch associated with each pixel in calculating the image roughness between neighboring pixels j and k. Here a patch of a pixel is defined as a square region centered at that pixel and all patches in an image share the same size. The new roughness function based on patches is defined as

$$U(\boldsymbol{x})=\frac{1}{4}\sum_{j=1}^{n_j}\sum_{k\in\mathcal{N}_j}\psi\left(\left|\left|\boldsymbol{f}_j(\boldsymbol{x})-\boldsymbol{f}_k(\boldsymbol{x})\right|\right|_{\boldsymbol{h}}\right) \quad (8)$$

where $\boldsymbol{f}_j(\boldsymbol{x})$ is a feature vector consisting of intensity values of all pixels in the patch centered at pixel j. The patch-based distance between pixel j and k is measured by

$$\left|\left|\boldsymbol{f}_j(\boldsymbol{x})-\boldsymbol{f}_k(\boldsymbol{x})\right|\right|_{\boldsymbol{h}}=\sqrt{\sum_{l=1}^{n_l}h_l(x_{j_l}-x_{k_l})^2} \quad (9)$$

where j_l denotes the l-th pixel in the patch of pixel j and k_l denotes the l-th pixel in the patch of pixel k, both having the same geometric relationship with respect to their center pixels. n_l is the total number of pixels in a patch. h_l is a positive weighting factor equal to the normalized inverse spatial distance between pixel j_l and pixel j with

$$\sum_{l=1}^{n_l}h_l=1. \quad (10)$$

The pixel-based regularization in (4) can be considered as a special case of the patch-based regularization with $n_l = 1$.

To obtain a convex penalized likelihood function, we require the penalty functions to satisfy the following conditions:

1. The function $\psi(t)$ is symmetric and differentiable everywhere;
2. The first-order derivative

$$\dot{\psi}(t)\triangleq\frac{d\psi(t)}{dt} \quad (11)$$

   is nondecreasing (and hence $\psi(t)$ is convex);
3. The curvature

$$w^{\psi}(t) \triangleq \frac{\dot{\psi}(t)}{t} \quad (12)$$

is nonincreasing for $t \geq 0$, and $0 < w^{\psi}(0) < +\infty$.

The above conditions are the same as those for the pixel-based convex regularizations [16], [20]. The penalty $\psi(\|\boldsymbol{f}_j(\boldsymbol{x}) - \boldsymbol{f}_k(\boldsymbol{x})\|_{\boldsymbol{h}})$ is convex with respect to $\boldsymbol{x}$ because it is a nondecreasing convex function of the convex norm function $\|\boldsymbol{f}_j(\boldsymbol{x}) - \boldsymbol{f}_k(\boldsymbol{x})\|_{\boldsymbol{h}}$ of $\boldsymbol{x}$ ([21], p84). Therefore, the patch-based penalty given in (8) is convex for any $\psi(t)$ that satisfies the above conditions. Examples of $\psi(t)$ include the quadratic function, Huber function, Lange function, and hyperbola function. Note that when $\psi(t)$ is the quadratic function, the patch-based regularization is equivalent to the pixel-based quadratic regularization with a scaled β.

B. Relation to other work

Recently, nonlocal regularization has been proposed for image restoration, motivated by the work of nonlocal means filter [22], [23]. Buades *et al* [2] proposed the regularization model

$$U(\boldsymbol{x}) = \sum_{j=1}^{n_j} (x_j - NL_j(\boldsymbol{x}|\tilde{\boldsymbol{x}}))^2 \quad (13)$$

where $\boldsymbol{x}$ is a reference image and $NL_j(\boldsymbol{x}|\boldsymbol{x})$ is the nonlocal means defined by

$$NL_j(\boldsymbol{x}|\tilde{\boldsymbol{x}}) = \frac{1}{\sum_k w_{jk}(\tilde{\boldsymbol{x}})} \sum_k w_{jk}(\tilde{\boldsymbol{x}}) x_k. \quad (14)$$

The weight function $w_{jk}(\boldsymbol{x})$ is often calculated on the reference image $\boldsymbol{x}$ using patches

$$w_{jk}(\tilde{\boldsymbol{x}}) = \exp\left(-\frac{\left|\left|\boldsymbol{f}_j(\tilde{\boldsymbol{x}}) - \boldsymbol{f}_k(\tilde{\boldsymbol{x}})\right|\right|_{\boldsymbol{h}}}{\delta^2}\right) \quad (15)$$

where δ is a filter parameter.

Kindermann *et al* [5] proposed a general form for nonlocal regularization

$$U(\boldsymbol{x}) = \sum_{j=1}^{n_j} \sum_k w_{jk}(\tilde{\boldsymbol{x}}) \psi(|x_j - x_k|), \quad (16)$$

in which nonlocal means filters can be interpreted as one step of fixed-point iteration when the Yaroslavsky function $\psi(t) = 1 - \exp(-t^2)$ and patch-based distance are employed. Note that the regularization form defined in (16) is quite similar to the penalty function given in (4) but with the weights calculated based on a reference image.

A major problem with the nonlocal Yaroslavsky functional is that the corresponding optimization problem is hard to solve because the objective function is nonconvex. Inspired by the spectral graph, Gilboa and Osher [3] proposed to use the quadratic function $\psi(t) = t^2$ to keep the objective function convex. Later, nonlocal total variation (TV) regularization was proposed for better preservation of structures than the nonlocal quadratic regularization [4], [6]. A general form of the penalty function is given by Elmoataz *et al* [8],

$$U(\boldsymbol{x})=\frac{1}{p}\sum_{j=1}^{n_j}\left(\sum_k w_{jk}(\tilde{\boldsymbol{x}})(x_j-x_k)^2\right)^{p/2} \quad (17)$$

with $p = 1$ for the nonlocal TV regularization. Note that (17) reduces to the regular TV penalty [24] when $w_{jk}(\tilde{\boldsymbol{x}}) = 1$ and the neighboring pixels consist of the first-order nearest neighbors. The nonlocal TV model has been applied to image reconstruction for computed tomography (CT) [12] and for MRI [11].

All the regularization models stated above rely on the weights $w_{jk}(\tilde{\boldsymbol{x}})$ that are calculated on a reference image $\tilde{\boldsymbol{x}}$. The reference image is usually required to be as close as possible to the true image $\boldsymbol{x}$ so that they share the same structures. This might not be a problem for image denoising because a noisy image can be used directly as the reference image. However, for tomographic image reconstruction, a good reference image is not readily available before image reconstruction, although people have used filtered backprojection reconstruction as the reference image [12].

Alternatively, the weights in nonlocal regularizations can be computed on the unknown image $\boldsymbol{x}$ and estimated simultaneously to avoid the need of a prior image $\tilde{\boldsymbol{x}}$ and to improve the estimation precision [7], [9]. However, this leads to a complicated nonconvex optimization problem. Empirical iterative implementations are often used, which calculate the weights based on the current image estimate and then update the image assuming that the weights are fixed constants [6], [9]–[11]. However, no proof of global convergence has been established for this type of algorithms.

It is worth noting that the weights in (16) can also be determined without using patches and it has long history in Bayesian image reconstruction. One early example is the line site model where the weights were adaptively determined to preserve edges [25]–[27]. Anatomical CT or MRI images can also be used to define the line sites in PET or SPECT image reconstruction [28]–[31]. Another example is the level set based regularization which estimates the weights using nonlocal boundary information [32]. Limitations of these approaches include either the challenging nonconvex optimization or the requirement of a perfectly aligned anatomical image.

In comparison, the regularization model presented in this paper has a convex objective function and does not require a prior reference image. In the next section, we shall present a deterministic optimization algorithm that guarantees global convergence.

IV. Proposed Optimization Transfer Algorithm

A. Optimization transfer

We develop an optimization transfer algorithm for the PL reconstruction using the patch-based regularization. The basic idea of optimization transfer [33] is to construct a surrogate function $Q(\boldsymbol{x}; \boldsymbol{x}^n)$ of the image $\boldsymbol{x}$ at the nth iteration which minorizes the original objective function $\Phi(\boldsymbol{x})$ by satisfying the following two conditions:

$$Q(\boldsymbol{x};\boldsymbol{x}^n)-Q(\boldsymbol{x}^n;\boldsymbol{x}^n) \leq \Phi(\boldsymbol{x})-\Phi(\boldsymbol{x}^n), \quad (18)$$

$$\nabla Q(\boldsymbol{x}^n;\boldsymbol{x}^n)=\nabla\Phi(\boldsymbol{x}^n). \quad (19)$$

where ∇ denotes the gradient with respect to $\boldsymbol{x}$. Then the maximization of $\Phi(\boldsymbol{x})$ is transferred into maximizing $Q(\boldsymbol{x}; \boldsymbol{x}^n)$

$$\boldsymbol{x}^{n+1}=\arg\max_{\boldsymbol{x}} Q(\boldsymbol{x};\boldsymbol{x}^n). \quad (20)$$

The surrogate function $Q(\boldsymbol{x}; \boldsymbol{x}^n)$ is usually easier to optimize than the original objective function by design. The new update $\boldsymbol{x}^{n+1}$ increases the original penalized likelihood monotonically

$$\Phi(\boldsymbol{x}^{n+1}) \geq \Phi(\boldsymbol{x}^n), \quad (21)$$

and this minorization-maximization procedure guarantees the convergence. The well-known expectation maximization (EM) algorithm [34] is a special case of the optimization transfer algorithms [33].

B. Likelihood surrogate

For the log-likelihood function, we borrow the surrogate function from the ML EM (expectation maximization) algorithm [33], [34]:

$$Q_L(\boldsymbol{x};\boldsymbol{x}^n)=\sum_{j=1}^{n_j} p_j \left(\hat{x}_{j,\mathrm{EM}}^{n+1}\log x_j - x_j\right) \quad (22)$$

where

$$p_j=\sum_{i=1}^{n_i} p_{ij} \quad (23)$$

and $\hat{x}_{j,\mathrm{EM}}^{n+1}$ is given by

$$\hat{x}_{j,\mathrm{EM}}^{n+1}=\frac{x_j^n}{p_j}\sum_{i=1}^{n_i}p_{ij}\frac{y_i}{\overline{y}_i^n} \quad (24)$$

where the expected projection at iteration n is

$$\boldsymbol{y}^{-n}=\boldsymbol{P}\boldsymbol{x}^n+\boldsymbol{r}. \quad (25)$$

From the derivation of the ML EM algorithm [33], we know that the surrogate function satisfies

$$Q_{L}(\boldsymbol{x};\boldsymbol{x}^n)-Q_{L}(\boldsymbol{x}^n;\boldsymbol{x}^n)\le L(\boldsymbol{y}|\boldsymbol{x})-L(\boldsymbol{y}|\boldsymbol{x}^n), \quad (26)$$

$$\nabla Q_{L}(\boldsymbol{x}^n;\boldsymbol{x}^n)=\nabla L(\boldsymbol{y}|\boldsymbol{x}^n). \quad (27)$$

C. Penalty surrogate

For a penalty function $\psi(t)$ satisfying the three conditions described in Section III-A, the following quadratic surrogate function can be constructed [16], [20]:

$$\begin{aligned}\psi(t)&\le\psi(t^n)+\dot{\psi}(t^n)(t-t^n)+\tfrac{1}{2}w^{\psi}(t^n)(t-t^n)^2\\&=\tfrac{1}{2}w^{\psi}(t^n)t^2+c(t^n)\end{aligned} \quad (28)$$

where $\dot{\psi}(t)$ and $w^{\psi}(t)$ are defined in Eq. (11) and (12), respectively, and $c(t^n)\equiv\psi(t^n)-\frac{1}{2}w^{\psi}(t^n)(t^n)^2$ is a function of the known estimate t^n but is independent of the unknown t. Replacing t and t^n in (28) by $t=\|\boldsymbol{f}_j(\boldsymbol{x})-\boldsymbol{f}_k(\boldsymbol{x})\|_{\boldsymbol{h}}$ and $t^n=\|\boldsymbol{f}_j(\boldsymbol{x}^n)-\boldsymbol{f}_k(\boldsymbol{x}^n)\|_{\boldsymbol{h}}$, we get the following surrogate function $Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)$ at iteration n for the patch-based penalty term $U(\boldsymbol{x})$ in (8):

$$Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)=\frac{1}{8}\sum_{j=1}^{n_j}\sum_{k\in\mathcal{N}_j}w_{j,k}^{\psi}(\boldsymbol{x}^n)\left|\left|\boldsymbol{f}_j(\boldsymbol{x})-\boldsymbol{f}_k(\boldsymbol{x})\right|\right|_{\boldsymbol{h}}^2 \quad (29)$$

where the term related to $c(t^n)$ is omitted and

$$w_{j,k}^{\psi}(\boldsymbol{x}^n)\triangleq w^{\psi}(\left|\left|\boldsymbol{f}_j(\boldsymbol{x}^n)-\boldsymbol{f}_k(\boldsymbol{x}^n)\right|\right|_{\boldsymbol{h}}). \quad (30)$$

It is easy to show

$$Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)-Q_U^a(\boldsymbol{x}^n;\boldsymbol{x}^n)\ge U(\boldsymbol{x})-U(\boldsymbol{x}^n), \quad (31)$$

$$\nabla Q_U^a(\boldsymbol{x}^n;\boldsymbol{x}^n)=\nabla U(\boldsymbol{x}^n). \quad (32)$$

By expanding the quadratic term $\left|\left|\boldsymbol{f}_j(\boldsymbol{x}^n)-\boldsymbol{f}_k(\boldsymbol{x}^n)\right|\right|_{\boldsymbol{h}}^2$ in (29) for each neighborhood and combining together the $j-k$ pairs from different neighborhoods, the surrogate $Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)$ can be simplified to

$$Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)=\frac{1}{8}\sum_{j=1}^{n_j}\sum_{k\in\mathcal{N}_j} w_{jk}(\boldsymbol{x}^n)\,(x_j-x_k)^2 \quad (33)$$

where

$$w_{jk}(\boldsymbol{x}^n)=\sum_{l=1}^{n_l} h_l w_{j_l,k_l}^{\psi}(\boldsymbol{x}^n) \quad (34)$$

with j_l and k_l denoting the l-th pixels in the patch $\boldsymbol{f}_j(\boldsymbol{x})$ and patch $\boldsymbol{f}_k(\boldsymbol{x})$, respectively. Equation (33) is in the same form as the nonlocal quadratic regularization, but with the weight $w_{jk}(\boldsymbol{x}^n)$ adaptively determined by the penalty function and the current estimate at each iteration.

To obtain a closed-form update equation, we derive a separable surrogate function for the surrogate $Q_U^a(\boldsymbol{x};\boldsymbol{x}^n)$ using the De Pierro's decoupling rule [35]:

$$(x_j-x_k)^2 \le 2\left(x_j-\frac{x_j^n+x_k^n}{2}\right)^2+2\left(x_k-\frac{x_j^n+x_k^n}{2}\right)^2. \quad (35)$$

Applying the above inequality to Eq. (33) and combining all the terms involving x_j, we obtain the following separable surrogate function $Q_U^b(\boldsymbol{x};\boldsymbol{x}^n)$

$$Q_U^b(\boldsymbol{x};\boldsymbol{x}^n)=\frac{1}{2}\sum_{j=1}^{n_j} w_j^n\left(x_j-\hat{x}_{j,\mathrm{Reg}}^{n+1}\right)^2 \quad (36)$$

where the pixel-wise weight w_j^n is

$$w_j^n=\sum_{k\in\mathcal{N}_j} w_{jk}(\boldsymbol{x}^n) \quad (37)$$

and the intermediate image $\hat{x}_{j,\mathrm{Reg}}^{n+1}$ is calculated by

$$\hat{x}_{j,\mathrm{Reg}}^{n+1}=\frac{1}{2w_j^n}\sum_{k\in\mathcal{N}_j} w_{jk}(\boldsymbol{x}^n)\,(x_k^n+x_j^n) \quad (38)$$

which corresponds to a local smoothing operation. A constant that is independent of $\boldsymbol{x}$ has been omitted in $Q_U^b(\boldsymbol{x};\boldsymbol{x}^n)$. It is easy to prove that

$$Q_U^b(\boldsymbol{x};\boldsymbol{x}^n)-Q_U^b(\boldsymbol{x}^n;\boldsymbol{x}^n) \geq U(\boldsymbol{x})-U(\boldsymbol{x}^n), \quad (39)$$

$$\nabla Q_U^b(\boldsymbol{x}^n;\boldsymbol{x}^n)=\nabla U(\boldsymbol{x}^n). \quad (40)$$

D. Proposed algorithm

The optimization of the original objective function $\Phi(\boldsymbol{x})$ is now transferred to the optimization of the combined surrogate function. The image is updated by

$$\boldsymbol{x}^{n+1}=\arg\max_{\boldsymbol{x}>0} Q_L(\boldsymbol{x};\boldsymbol{x}^n)-\beta Q_U^b(\boldsymbol{x},\boldsymbol{x}^n), \quad (41)$$

which can be performed pixel-by-pixel

$$x_j^{n+1}=\arg\max_{x_j\geq 0} q_j(x_j) \quad (42)$$

$$q_j(x_j)=\left(\hat{x}_{j,\mathrm{EM}}^{n+1}\log x_j-x_j\right)-\frac{\beta_j^n}{2}\left(x_j-\hat{x}_{j,\mathrm{Reg}}^{n+1}\right)^2 \quad (43)$$

where

$$\beta_j^n=\beta w_j^n/p_j. \quad (44)$$

Solving the quadratic equation derived from the Karush-Kuhn-Tucker (KKT) condition, we get the PL image estimate at iteration $(n + 1)$,

$$x_j^{n+1}=\frac{2\hat{x}_{j,\mathrm{EM}}^{n+1}}{\sqrt{\left(1-\beta_j^n\hat{x}_{j,\mathrm{Reg}}^{n+1}\right)^2+4\beta_j^n\hat{x}_{j,\mathrm{EM}}^{n+1}}+\left(1-\beta_j^n\hat{x}_{j,\mathrm{Reg}}^{n+1}\right)}. \quad (45)$$

The above update equation reduces to the ML-EM formula when $\beta = 0$.

This optimization transfer algorithm guarantees a monotonic convergence to the global solution when the penalty function satisfies the three conditions given in Section III-A. The whole algorithm is summarized in Algorithm table 1.

E. Comparison with pixel-based regularization

The proposed algorithm can be applied to both patch-based regularization ($n_l > 1$) and pixel-based regularization ($n_l = 1$). The only difference between pixel-based regularization and patch-based regularization is the value of the weight $w_{jk}(\boldsymbol{x})$. We use a noise-free step-edge image and a noisy image shown in figure 1 to demonstrate the benefit of the patch-based regularization. We pick an edge pixel as the jth pixel and compute the weights $w_{jk}(\boldsymbol{x})$ for the

pixels in a 7 × 7 neighborhood of pixel j, where $\boldsymbol{x}$ îs either the noise-free step-edge image or the noisy image. The Lange function is used here as the penalty function, with the curvature $w^{\psi}(t)=\frac{1}{|t|+\delta}$. The left column of figure 2 shows the images of the pixel-based weights. For the noise free image, the resulting weights are as expected with the pixels on the right hand side of the image having higher weights than those on the left hand side. However, weights are unstable for the noisy image and are also sensitive to the hyper-parameter δ that controls the shape of penalty function. For comparison, the weights calculated using 3 × 3 patches are shown in the right column of figure 2. Higher values are concentrated along the step edge because only these pixels share a similar patch pattern to the central pixel. The weights are fairly robust to the image noise and insensitive to the change of δ value.

The robustness of the patch-based weight comes from two factors: (1) the patch distance is computed based on n_l pixels, so it is less sensitive to noise than pixel distance; (2) the weight $w_{jk}(\boldsymbol{x}_n)$ in (34) is a spatial average of $w^{\psi}_{j_l,k_l}(\boldsymbol{x}^n)$. We expect that the patch-regularization will result in more robust image reconstruction.

While the weight $\{w_{jk}(\boldsymbol{x})\}$ given in (34) is derived for the optimization transfer algorithm, it plays a critical role in preserving edges at convergence. To illustrate this point, let $\boldsymbol{x}^*$ be the solution at convergence. The KKT condition states that $\boldsymbol{x}^*$ must satisfy

$$\nabla L(\boldsymbol{y}|\boldsymbol{x}^*)-\beta\nabla U(\boldsymbol{x}^*)=0. \quad (46)$$

Substituting the matched gradient condition (32) into the above equation results in

$$\nabla L(\boldsymbol{y}|\boldsymbol{x}^*)=\beta\nabla Q^a_U(\boldsymbol{x}^*;\boldsymbol{x}^*). \quad (47)$$

For each pixel j, we have

$$[\nabla L(\boldsymbol{y}|\boldsymbol{x}^*)]_j=\frac{\beta}{2}\sum_{k\in\mathcal{N}_j} w_{jk}(\boldsymbol{x}^*)\,(x^*_j-x^*_k). \quad (48)$$

Equation (48) shows that the penalty of the pixel intensity difference between pixel j and pixel k is weighted by $w_{jk}(\boldsymbol{x}^*)$ at convergence. Therefore, the patch-based regularization can preserve edges by adaptively computing the weighting factor $w_{jk}(\boldsymbol{x}^*)$, although the edges may not be as sharp as in pixel-based regularization because of the averaging operation in (34). The quadratic regularization does not preserve edges because the corresponding $w_{jk}(\boldsymbol{x}^*)$ is always unity.

V. Computer Simulation

A. Setup

Computer simulations were conducted to validate the proposed patch-based regularization algorithm. We simulated a GE Discovery ST PET scanner in 2D mode [36]. The scanner has 420 crystals per ring with each crystal having a cross section of 6.3 × 6.3 mm^2. A brain

phantom (Fig. 3(a)) was used to simulate PET tracer distribution in the gray matter, white matter and a tumor. The noise-free image was forward projected to generate a noise-free sinogram. A uniform background of 25% total true coincidences was added to simulate the random and scatter fraction in 2D [36]. Independent Poisson noise was then introduced to generate 100 realizations, each having an expected total number of events equal to 500k. Images were reconstructed by the ML EM algorithm and the penalized likelihood method with three different regularizations: the quadratic penalty, the pixel-based Lange penalty, and the proposed patch-based Lange penalty. The noisefree background sinogram was included in the forward model (2) for scatter and random correction. Two hundred iterations were used in each reconstruction. All images were represented by 111 × 111 pixels with a pixel size of 3 × 3 mm^2. The patch size and the neighborhood size were both 3 × 3 pixels unless noted otherwise.

B. Comparison with pixel-based regularizations

The images reconstructed by different approaches are shown in Fig. 3(b)–(i) and the corresponding intensity profiles along the line through the tumor are shown in Fig. 4. Fig. 3(b) and (c) show the images reconstructed by the ML-EM with 20 iterations and the PL reconstruction using the quadratic regularization with $\beta = 0.1$, respectively. Both methods oversmoothed the tumor in order to control the noise in the brain region. Figures 3(d)–(f) show the images reconstructed using the pixel-based edge-preserving Lange regularization with $\beta = 0.2$ but different hyper-parameter values $\delta = 0.1, 0.01, 0.001$. Edges of the grey matter and tumor are preserved in the reconstructions. However, the resulting images are sensitive to the value of the hyper-parameter δ. In comparison, figures 3(g)–(i) show the reconstructions using the patch-based Lange regularization with the same set of parameters. The tumor target is well preserved and the results are insensitive to the change of the hyper-parameter δ. This is because that the patch distance has a much smaller variance than the pixel distance and the spatial average in the weight calculation substantially reduces the chance of generating isolated noisy pixels. As long as δ is less than the minimum patch distance of true edges, the final reconstruction is insensitive to the exact value of δ. This greatly simplifies the selection of the hyperparameter in the edge-preserving regularization.

Quantitatively we compare the ensemble mean of tumor contrast recovery coefficient (CRC) versus the standard deviation of background noise as a function of the smoothing parameter β between the quadratic regularization and patch regularization in figure 5. The CRC of the ith reconstructed image is calculated by

$$CRC_i = CR_i / CR_0,\ CR_i = |\overline{S}_i - \overline{B}_i| / \overline{B}_i \quad (49)$$

where $CR_0 = 3$ is the true contrast, $\bar{S}_i$ denotes the mean activity of the tumor region and $\bar{B}_i$ denotes the mean activity of the white matter region (background). Here the true tumor and white matter regions are used. The results clearly indicate that the patch-based regularization can achieve higher contrast recovery (better performance) than the quadratic regularization at any given background noise level.

The comparison between the patch-based regularization and the pixel-based Lange regularization is plotted in figure 6. Each group contains four curves corresponding to different hyper-parameter values δ = 1.0, 0.1, 0.01, 0.0001. As δ changes, the performance of the pixel-based Lange regularization changes dramatically, indicating the hyper-parameter δ has to be selected carefully for the pixel-based regularization. In comparison, the performance of the patch-based Lange regularization is insensitive to the change of δ value and achieves higher CRC at any given background SD level, which demonstrates that the patch-based regularization is more robust than the pixel-based regularization for image reconstruction.

C. Effects of the neighborhood size and patch size

The computational cost of the patch regularization depends heavily on the patch size and neighborhood size. For fast computation, smaller patch and neighborhood are preferred. In addition, large patches may not be able to identify small image features and hence cannot preserve the corresponding edges. Here we study the effect of the neighborhood size and patch size on image quality.

Figure 7 shows the tumor contrast recovery versus the background SD curves for five different neighborhood sizes (3×3, 5×5, 7×7, 9×9, and 11×11). The patch size was 3×3 and δ = 0.01. The 3 ×3 neighborhood was found to have the best performance amongst all the cases studied, except at very low standard deviation levels. The loss of performance with large neighborhood window can be due to the accumulation of small weights of non-similar patches, which biases the results [37].

The effect of the patch size on the tumor contrast recovery is shown in Figure 8. The neighborhood window was set to 3×3 and the hyper-parameter δ was 0.01. Except for the patch size H = 1, which reduces to the pixel-based regularization, there are little difference in the performance among the other three patch sizes (3 × 3, 5 × 5, 7 × 7).

The effects of neighborhood window size and patch size were also investigated at a lower count level of 150k events and 25% background. The results are shown in Figure 9. Except with an increase of background noise, the plots are similar to those at 500k count level. Overall, 3 × 3 pixels appears to be a good choice for the neighborhood size and patch size.

VI. Application to Real 3D Data

We applied the proposed algorithm to a 3D nonhuman primate data set acquired on a microPET P4 scanner [38]. The radiotracer is ^{11}C-SCH 23390 which binds to dopamine D1 receptors in the brain [39]. The total number of prompt events is 3.6M and number of delayed events is 1.3M. Randoms were pre-corrected by the delayed window technique. A transmission scan was performed to estimate the attenuation map. Scatter sinogram was estimated by using the single-scatter simulation method [40]. The estimated scatter fraction was 24%. The estimated scatter sinogram and attenuation factors were incorporated into the forward model of the PL reconstruction. All images were reconstructed using an array of 128 ×128 ×63 voxels with a voxel size of 1.0 × 1.0 ×1.2 mm^3.

Figure 10 shows the images reconstructed using the pixel-based quadratic regularization and the patch-based Lange regularization. The pixel-based Lange regularization was not shown here because the result highly depends on the hyper-parameter δ. Within each group, the left column shows the transverse, coronal and sagittal slices through the striatum regions, while the right column shows the slices through the eye glands. These regions were chosen because of their high uptake of the PET tracer. In the patch regularization, both the neighborhood and patch size were $3 \times 3 \times 3$ voxels. The hyper-parameter δ in the Lange function was set to 0.001. The regularization parameters were $\beta = 2 \times 10^4$ for the quadratic regularization and $\beta = 1 \times 10^4$ for the patch regularization. They were chosen to match the background noise level.

Uptakes in regions of interest were quantified using the AMIDE software [41] to compare the quadratic regularization and the patch regularization. Ten spherical ROIs of 8 mm in diameter were randomly placed in the background regions of the brain as shown in figure 11. Another four ROIs were drawn in the striatum and gland regions, which have higher uptake. The mean and standard deviation of the activities inside each ROI are shown in figure 12. As expected, the two regularization methods have similar quantitative results for the background ROIs. However, the patch regularization produces higher uptake in the striatum and gland regions than the quadratic regularization. In particular, the average intensity of the two gland ROIs (3 and 4) obtained by the patch regularization is more than 30% higher than that obtained by the quadratic regularization. The higher contrast recovery of the patch regularization is consistent with the expectation and simulation results. However, because no ground truth is available, we cannot perform any conclusive comparison for the real data.

VII. Conclusion

In this paper, we have proposed a patch-based regularization method for PET image reconstruction and developed an optimization algorithm to find the penalized likelihood solution. The proposed algorithm is validated by computer simulations. The results show that the patch-based regularization can achieve better quantitative performance than the commonly used quadratic regularization and the traditional pixel-based non-quadratic regularization. The application to the real primate brain data also showed promising results. The full benefit of the proposed algorithm for real PET studies will be evaluated in future.

Acknowledgments

The authors thank Dr. Karen Bales, Katie Hinde, and Nicole Maninger of the California National Primate Research Center for kindly sharing the real primate data which were collected under NIH grant numbers HD053555 to Karen Bales, RR00169 to the CNPRC, and a grant from the Good Nature Institute to Karen Bales. The authors also thank Dr. Douglas Rowland in the Center of Molecular and Genomic Imaging at UC Davis for his assistance in processing the real data.

This work was supported by Grant Number R01EB000194 from the National Institute Of Biomedical Imaging And Bioengineering. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Biomedical Imaging And Bioengineering or the National Institutes of Health.

References

1. Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography. Physics in Medicine and Biology. 2006; 51(15):R541–578. [PubMed: 16861768]
2. Buades, A.; Coll, B.; Morel, JM. Technical Report. Vol. 22. CMLA; 2006. Image enhancement by non-local reverse heat equation.
3. Gilboa G, Osher S. Nonlocal linear image regularization and supervised segmentation. SIAM Multiscale Modeling and Simulation. 2007; 6(2):595–630.
4. Gilboa G, Osher S. Nonlocal operators with applications to image processing. SIAM Multiscale Modeling and Simulation. 2008; 7(3):1005–1028.
5. Kindermann S, Osher S, Jones P. Deblurring and denoising of images by nonlocal functionals. SIAM Multiscale Modeling and Simulation. 2005; 4(4):1091–1115.
6. Zhang X, Burger M, Bresson X, Osher S. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM Journal of Imaging Sciences. 2010; 3(3):253–276.
7. Peyre G, Bougleux S, Cohen L. Non-local regularization of inverse problem. Inverse Problems and Imaging. 2011; 5(2):511–530.
8. Elmoataz A, Lezoray O, Bougleux S. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. IEEE Transactions on Image Processing. 2008; 17(7):1047–1060. [PubMed: 18586614]
9. Chen Y, Ma J, Feng Q, et al. Nonlocal prior Bayesian tomographic reconstruction. Journal of Mathematical Imaging and Vision. 2008; 30(2):133–146.
10. Adluru G, Tasdizen T, Schabel MC, DiBella EVR. Reconstruction of 3D dynamic contrast-enhanced magnetic resonance imaging using nonlocal means. Journal of Magnetic Resonance Imaging. 2010; 32(5):1217–1227. [PubMed: 21031528]
11. Liang D, Wang H, Chang Y, Ying L. Sensitivity encoding reconstruction with nonlocal total variation regularization. Magnetic Resonance in Medicine. 2011; 65(5):1384–1392. [PubMed: 21500265]
12. Lou Y, Zhang X, Osher S, Bertozzi A. Image recovery via nonlocal operators. SIAM Journal of Scientific Computing. 2010; 42(2):185–197.
13. Wang, G.; Qi, J. Patch-based regularization for iterative PET image reconstruction. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; April 2011; p. 1508-1511.
14. Yang, Z.; Jacob, M. A unified energy minimization framework for nonlocal regularization. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; April 2011; p. 1150-1153.
15. Lange K. Convergence of EM image reconstruction algorithms with Gibbs smoothing. IEEE Trans Med Imaging. 1990; 9(4):439–446. [PubMed: 18222791]
16. Huber, PJ. Robust Statistics. John Wiley & Sons; New York, NY, USA: 1981.
17. Bouman C, Sauer K. A generalized Gaussian image model for edge-preserving MAP estimation. IEEE Transactions on Image Processing. 2002; 2(3):296–310. [PubMed: 18296219]
18. Geman, S.; McClure, D. Bayesian image analysis: an application to single photon emission tomography. Proceedings of the Statistical Computing Section of the American Statistical Association; Washington, DC, USA. 1985. p. 12-18.
19. Alenius S, Ruotsalainen U. Bayesian image reconstruction for emission tomography based on median root prior. Europian Journal of Nuclear Medicine. 1997; 24(3):258–265.
20. Erdogan H, Fessler JA. Monotonic algorithms for transmission tomography. IEEE Transactions on Medical Imaging. Apr; 1999 18(9):801–814. [PubMed: 10571385]
21. Boyd, S.; Vandenberghe, L. Convex Optimization. Cambridge University Press; 2004.
22. Efros AA, Leung TK. Texture synthesis by non-parametric sampling. Proceedings of the International Conference on Computer Vision. 1999; 2:1022–1038.
23. Buades A, Coll B, Morel JM. On image denoising methods. SIAM Multiscale Modeling and Simulation. 2005; 4(2):490–530.

24. Rudin LI, Osher S, Fatemi E. Nonlinear totalvariation based noise removal algorithms. Physica D: Nonlinear Phenomena. 1992; 60(1–4):259–268.
25. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984; 6(6):721–741. [PubMed: 22499653]
26. Johnson VE, Wong WH, Chen CT. Image restoration using Gibbs priors: boundary modeling, treatment of blurring, and selection of hyperparameter. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1991; 13(5):413–425.
27. Kao CM, Pan X, Chen CT, Wong WH. Image restoration and reconstruction with a Bayesian approach. Medical Physics. 1998; 25(5):600–613. [PubMed: 9608469]
28. Leahy R, Yan X. Incorporation of anatomical MR data for improved functional imaging with PET. Information Processing in Medical Imaging, Lecture Notes in Computer Science. 1991; 511:105–120.
29. Chen CT, Ouyang X, Wong WH, Hu X, Johnson VE. Sensor fusion in image reconstruction. IEEE Transactions on Nuclear Science. 1993; 38(2):687–692.
30. Gindi G, Lee M, Rangarajan A, Zubal IG. Bayesian reconstruction of functional images using anatomical information as priors. IEEE Transactions on Medical Imaging. 1993; 12(4):670–680. [PubMed: 18218461]
31. Ouyang X, Wong WH, Johnson VE, Hu X, Chen CT. Incorporation of correlated structural images in PET image reconstruction. IEEE Transactions on Medical Imaging. 1994; 13(4):627–640. [PubMed: 18218541]
32. Yu DF, Fessler JA. Edge-preserving tomographic reconstruction with nonlocal regularization. IEEE Transactions on Medical Imaging. 2002; 21(2):159–173. [PubMed: 11929103]
33. Lange K, Hunter DR, Yang I. Optimization transfer using surrogate objective functions. Journal of Computational and Graphical Statistics. 2000; 9(1):1–20.
34. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B. 1977; 39(1):1–38.
35. De Pierro AR. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Transactions on Medical Imaging. 1995; 14(1):132–137. [PubMed: 18215817]
36. Mawlawi O, Podoloff DA, Kohlmyer S, et al. Performance Characteristics of a Newly Developed PET/CT Scanner Using NEMA Standards in 2D and 3D Modes. Journal of Nuclear Medicine. 2004; 45(10):1734–1742. [PubMed: 15471842]
37. Salmon J. On two parameters for denoising with non-local means. IEEE Signal Processing Letters. 2010; 17(3):269–272.
38. Tai YC, Chatziioannou A, Siegel S, Young J, Newport D, Goble RN, Nutt RE, Cherry SR. Performance evaluation of the microPET P4: a PET system dedicated to animal imaging. Physics in Medicine and Biology. 2001; 46(7):1845–1862. [PubMed: 11474929]
39. Farde L, Halldin C, Sone-Elander S, Sedvall G. PET analysis of human dopamine receptor subtypes using ^{11}C-SCH 23390 and ^{11}C-raclopride. Psychopharmacology. 1987; 92(3):278–284. [PubMed: 2957716]
40. Watson CC. New, faster, image-based scatter correction for 3D PET. IEEE Transactions on Nuclear Science. 2000; 47(4):1587–1594.
41. Loening AM, Gambhir S. AMIDE: a free software tool for multimodality medical image analysis. Molecular Imaging. 2003; 2(3):131–137. [PubMed: 14649056]

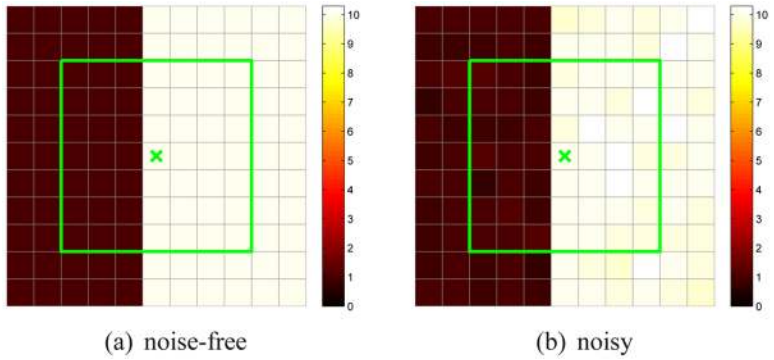

Fig. 1.
The noise-free and noisy edge images. Pixel *j* is marked by '×'.

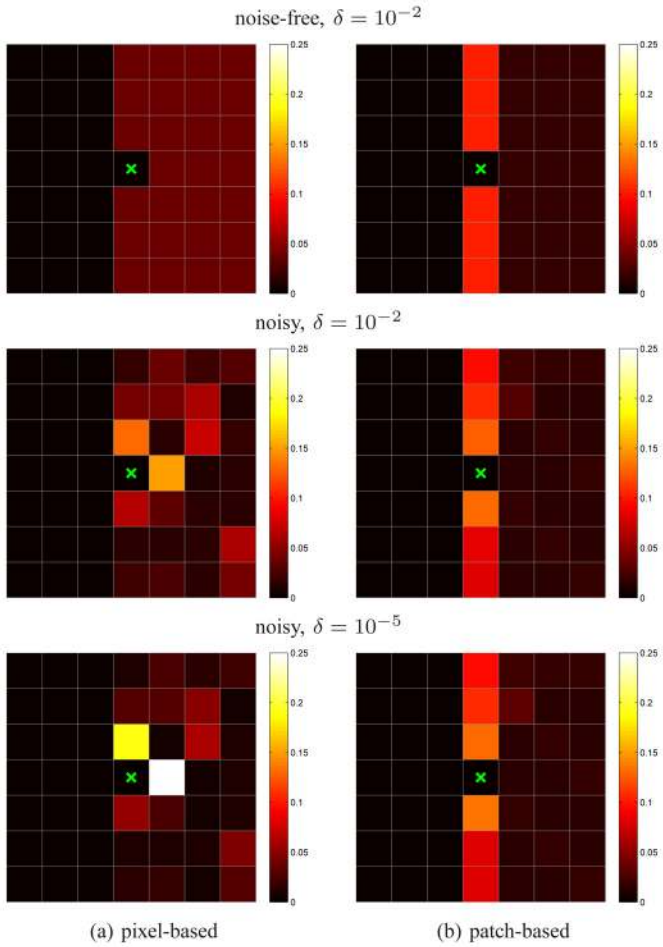

Fig. 2.

Images of the weights w_{jk} in the pixel-based regularization and patch-based regularization with different δ values for the noise-free and noisy step images shown in Fig. 1. The jth pixel is marked by '×'.

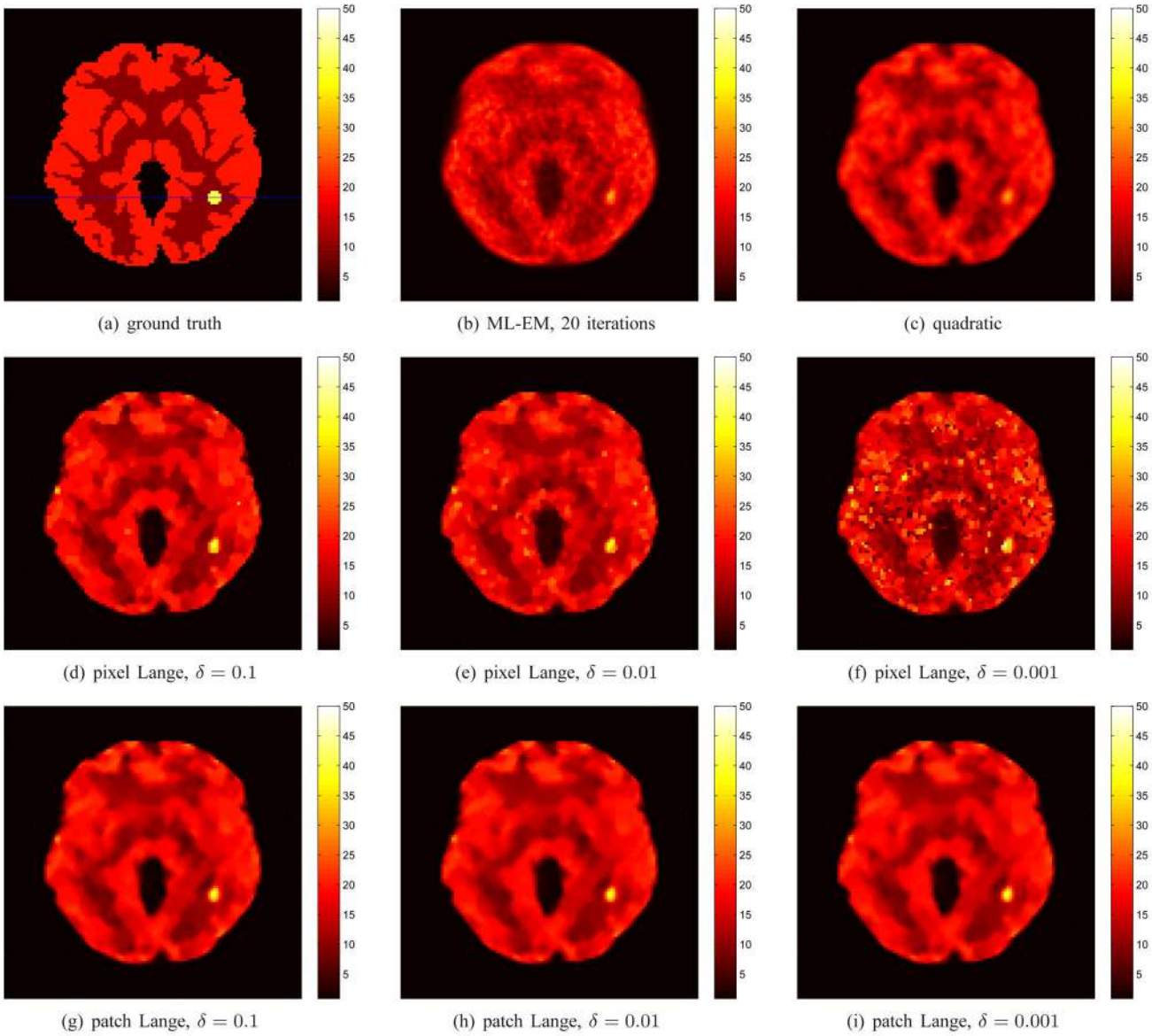

Fig. 3.

(a) The original phantom image with a marker line going through the tumor spot horizontally. (b–i) Reconstructed images by the ML-EM with 20 iterations (b), quadratic regularization with $\beta = 0.1$ (c), pixel-based Lange regularization (d–f), and patch-based Lange regularization (g–i). The smoothing regularization parameter β is 0.2 for the pixel and patch Lange regularizations.

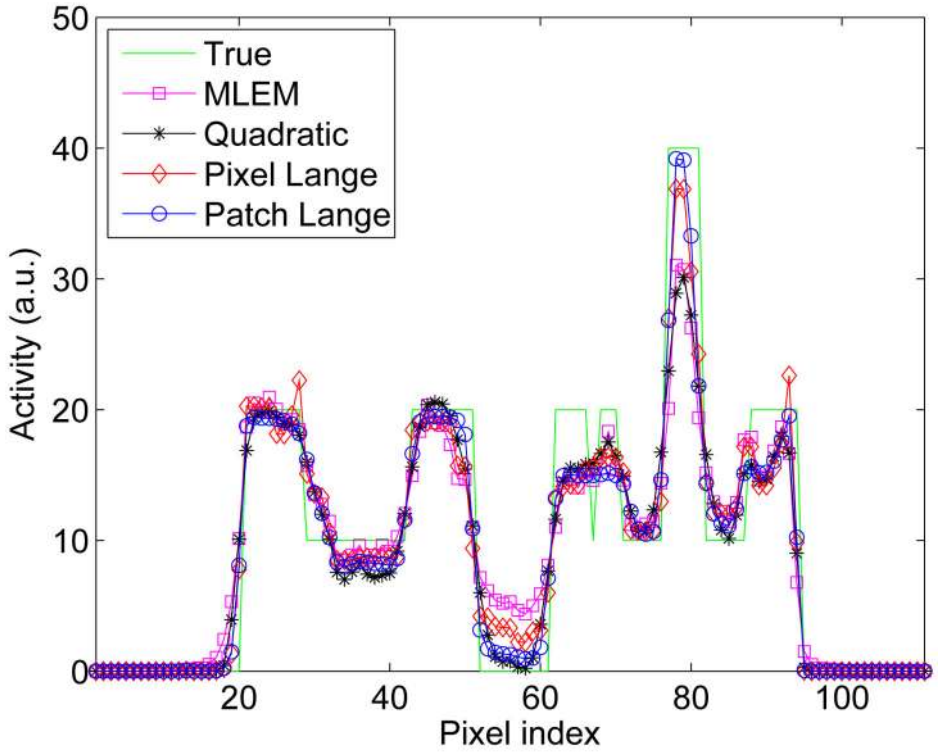

Fig. 4.

Intensity profiles along the horizontal line through the tumor as indicated in Fig. 3(a). The reconstruction methods and corresponding parameters are the MLEM algorithm with 20 iterations, the quadratic regularization ($\beta = 0.1$), the pixel Lange regularization ($\beta = 0.2$, $\delta = 0.01$), and the patch Lange regularization ($\beta = 0.2$, $\delta = 0.01$).

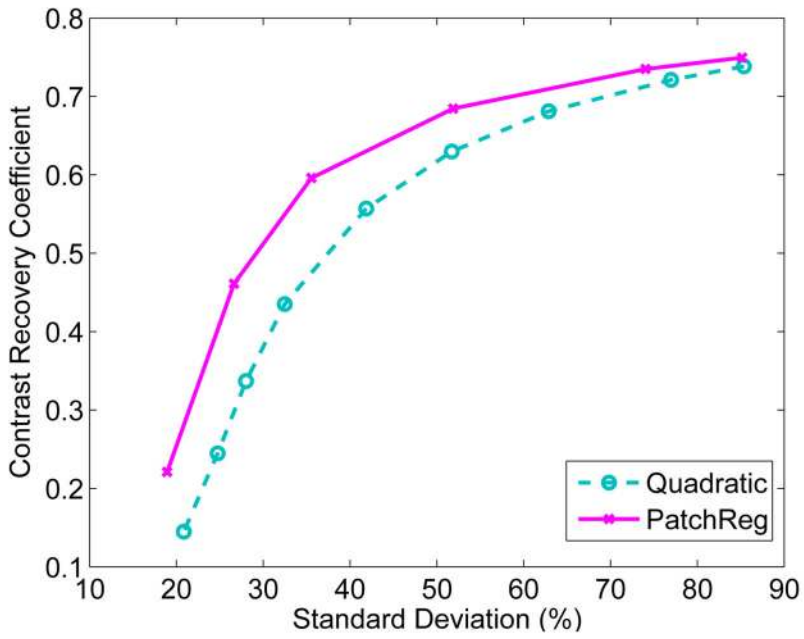

Fig. 5.

The tumor contrast recovery versus background standard deviation curves for the quadratic regularization and patch-based regularization. The curves are plotted by varying the regularization parameter β.

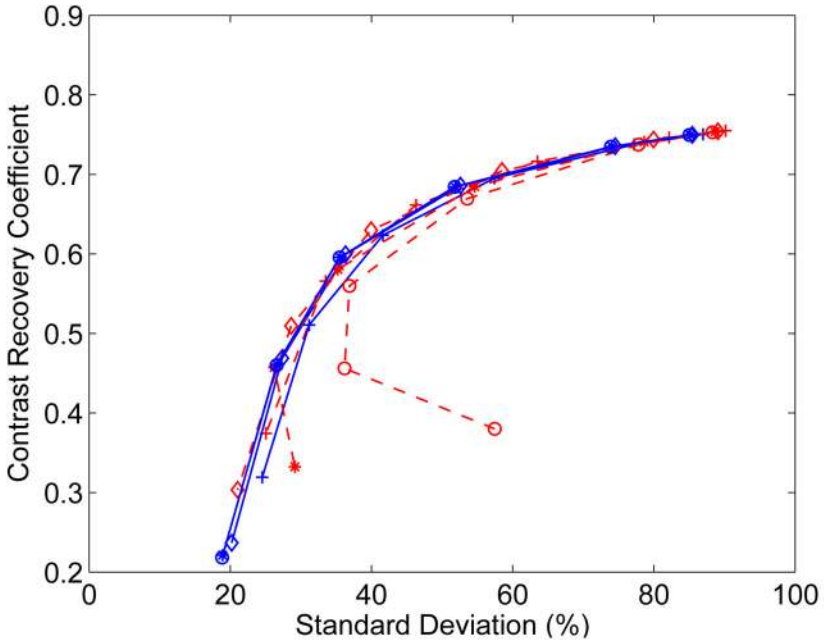

Fig. 6.

The tumor contrast recovery versus background standard deviation curves of the pixel-based (dash lines) and patch-based (solid lines) Lange regularizations. The curves are plotted by varying the smoothing parameter β for different hyper-parameter δ values: $\delta = 1.0$ (+),$\delta =$ 0.1 (◇), $\delta = 0.01$ (*),$\delta = 0.001$ (○).

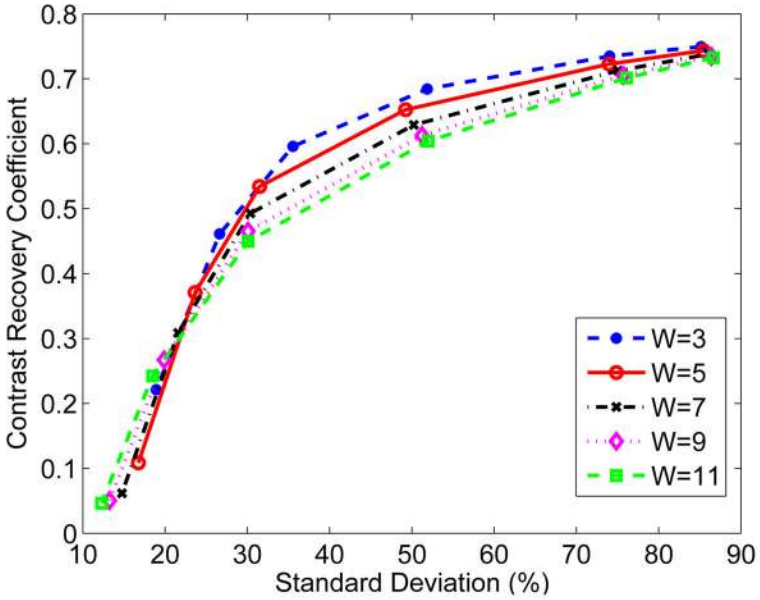

Fig. 7.

The tumor contrast recovery versus background standard deviation curves of the patch regularization with different neighborhood sizes. Curves are plotted by varying the regularization parameter β.

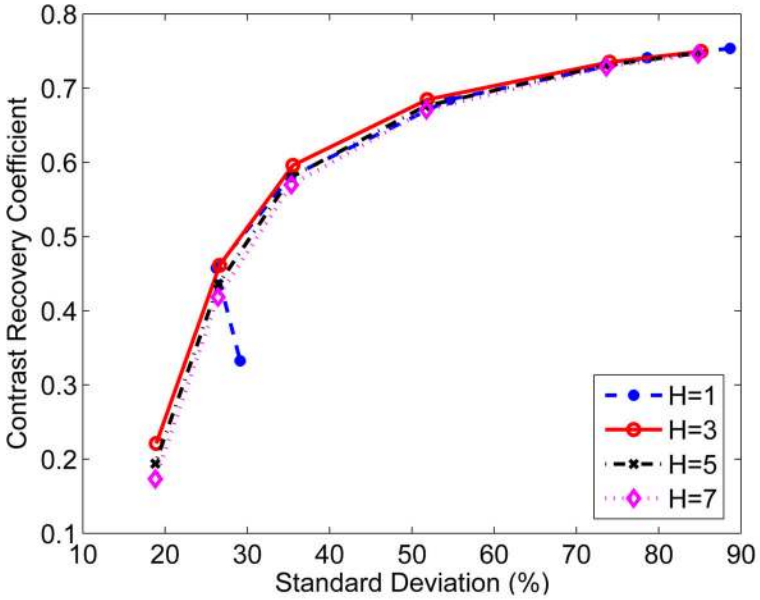

Fig. 8.

The tumor contrast recovery versus background standard deviation curves of the patch regularization with different patch sizes. Curves are plotted by varying the regularization parameter β.

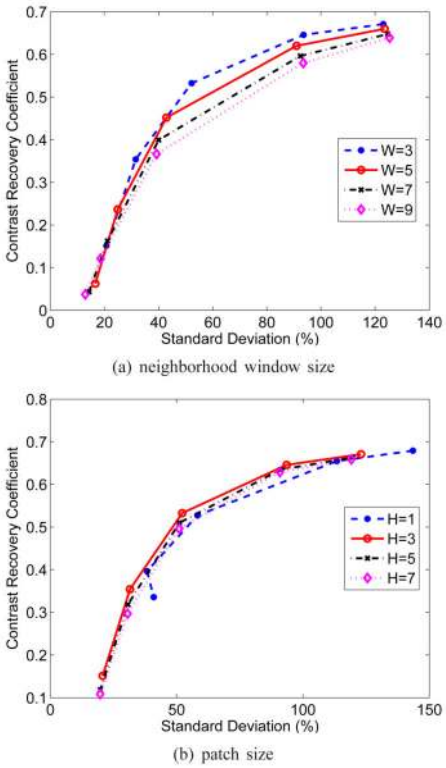

Fig. 9.

Effect of (a) neighborhood window size *W* and (b) patch size *H* on the tumor contrast recovery versus background noise curve of the patch-based regularization at 150k count level.

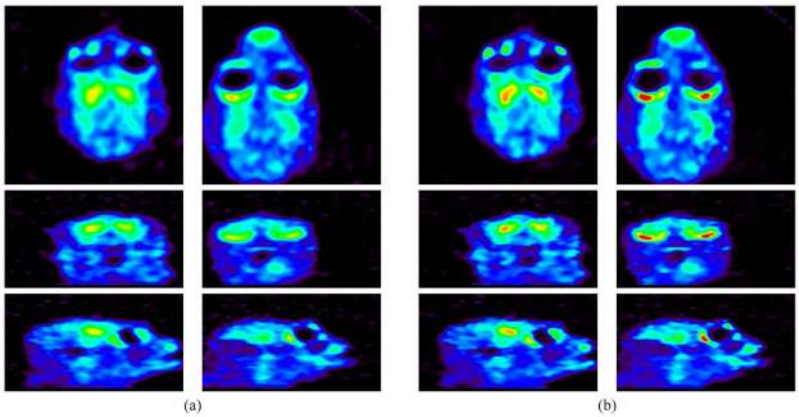

Fig. 10.

Reconstructed images of the real 3D primate data using (a) the quadratic regularization and (b) the patch regularization. Transverse, coronal and sagittal slices are shown in row 1, 2 and 3, respectively. The left column in (a) and (b) shows the slices through the striatum regions and the right column shows the slices through the gland regions.

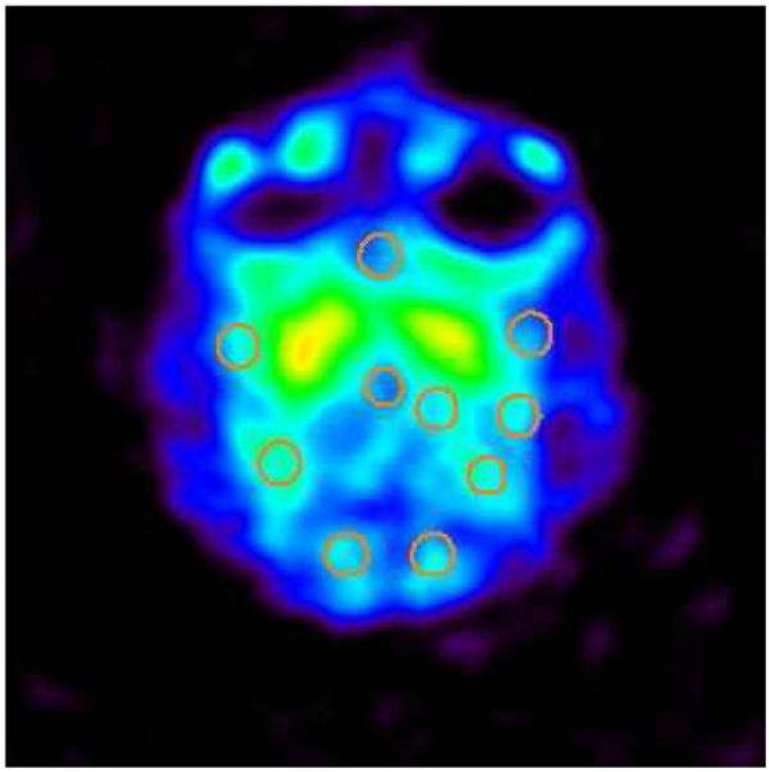

Fig. 11.
The ten spherical ROIs drawn in the brain background.

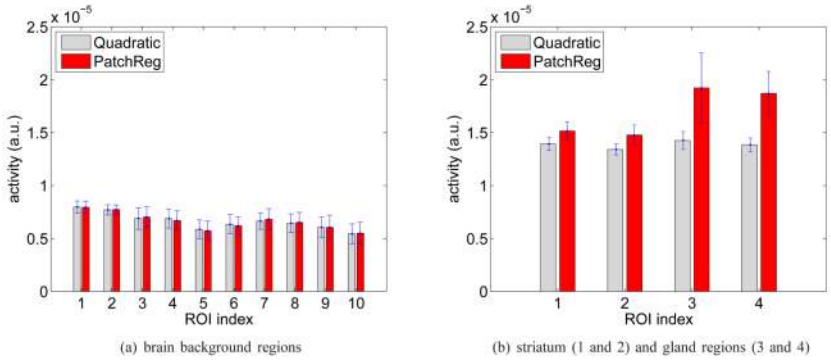

Fig. 12.

Comparison of the mean and standard deviation of the activities inside (a) the background ROIs and (b) the high-uptake ROIs.

Algorithm 1

The optimization transfer algorithm for PET image reconstruction using patch regularization

1: Initialize parameters: maximum iteration number MaxIter, regularization parameter β;

2: Initialize image: $x_j^1 = 1$

3: **for** n = 1 to MaxIter **do**

4: EM image update from sinogram $\{y_i\}$:

$$\hat{x}_{j,\mathrm{EM}}^{n+1} = \frac{x_j^n}{p_j} \sum_{i=1}^{n_i} p_{ij} \frac{y_i}{\overline{y}_i^n}$$

where $p_j = \Sigma_i p_{ij}$ and $\overline{y}_i^n$ is the expected projection calculated by (25);

5: Image smoothing:

$$\hat{x}_{j,\mathrm{Reg}}^{n+1} = \frac{1}{2w_j^n} \sum_{k \in \mathcal{N}_j} w_{jk}(\boldsymbol{x}^n)\,(x_k^n + x_j^n)$$

where the weight $w_{jk}(\boldsymbol{x}^n)$ is calculated by (34) and $w_j^n = \sum_{k \in \mathcal{N}_j} w_{jk}(\boldsymbol{x}^n)$

6: Pixel-by-pixel image fusion:

$$x_j^{n+1} = \frac{2\hat{x}_{j,\mathrm{EM}}^{n+1}}{\sqrt{\left(1 - \beta_j^n \hat{x}_{j,\mathrm{Reg}}^{n+1}\right)^2 + 4\beta_j^n \hat{x}_{j,\mathrm{EM}}^{n+1}} + \left(1 - \beta_j^n \hat{x}_{j,\mathrm{Reg}}^{n+1}\right)}$$

where β_j^n is defined by Eq. (44).

7: **end for**

8: **return** The image estimate $\boldsymbol{x}^{n+1}$