Gap compensation during PET image reconstruction by constrained, total variation minimization

Seonmin Ahn[b)]
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

Soo Mee Kim[b)] and Jungah Son
Department of Nuclear Medicine, Seoul National University, Seoul 151-742, Korea and Institute of Radiation Medicine, Medical Research Center, and Interdisciplinary Programs in Radiation Applied Life Science Major, Seoul National University, Seoul 151-742, Korea

Dong Soo Lee
Department of Nuclear Medicine, Seoul National University, Seoul 151-742, Korea; Institute of Radiation Medicine, Medical Research Center, and Interdisciplinary Programs in Radiation Applied Life Science Major, Seoul National University, Seoul 151-742, Korea; and Department of WCU Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul 151-742, Korea

Jae Sung Lee[a)]
Department of Nuclear Medicine, Seoul National University, Seoul 151-742, Korea; Institute of Radiation Medicine, Medical Research Center, and Interdisciplinary Programs in Radiation Applied Life Science Major, Seoul National University, Seoul 151-742, Korea; and Departments of Biomedical Sciences and WCU Brain and Cognitive Sciences, Seoul National University, Seoul 151-742, Korea

(Received 25 August 2011; revised 8 December 2011; accepted for publication 9 December 2011; published 11 January 2012)

Purpose: Positron emission tomography (PET) is a noninvasive molecular imaging tool with various clinical and preclinical applications. The polygonal structure of small-diameter PET scanners that are designed for specific purposes can lead to gaps between the detector modules and result in loss of PET data during measurement. In the current study, the authors applied the compressed sensing (CS)-based total variation (TV) minimization method to PET image reconstructions to reduce the artifacts caused by gaps in small-diameter PET systems.
Methods: The first step in each iteration estimates whether an image is consistent with the measured PET data using the existing common reconstruction algorithms (ART, OSEM, and RAMLA). The second step recovers sparsity in the gradient domain of the image by minimizing the TV of an estimated image. The authors evaluated the gap-compensable reconstruction algorithms with uniform disk and Shepp-Logan phantoms by simulating sinograms which contained Poisson random noise and a data loss due to detector gaps. In addition, these methods were applied to real high resolution research tomography (HRRT)-like sinograms of human brain and uniform phantom. A comparison with other methods for gap compensation prior to or during image reconstruction was also made. Quantitative evaluations were performed by computing the uniformity, root mean squared error, and difference between the reconstructed images of nongapped and gapped sinograms.
Results: The simulation results showed that the gap-compensable methods incorporating TV minimization could control gap artifacts, as well as Poisson random noise. In particular, OSEM-TV and RAMLA-TV showed distinct potential via the properties of convergence and robustness to different noise levels and gap angle.
Conclusions: A TV minimization strategy incorporated into commonly used PET reconstruction algorithms was useful for reducing the occurrence of artifacts due to gaps between detector modules in small-diameter PET scanners. © *2012 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3673775]

Key words: gap compensation, TV minimization, compressed sensing, positron emission tomography (PET), image reconstruction

I. INTRODUCTION

Positron emission tomography (PET) provides relevant information about biochemical processes at the molecular level. Consequently, the use of PET has increased remarkably over the past 10 years in various clinical and preclinical fields of application such as disease diagnosis, therapeutic response monitoring, and drug discovery.[1–4] In addition, increasing demand for imaging systems with high spatial resolution and sensitivity have led to the development of PET scanners with small ring diameters dedicated to small animals or specific organs, such as the brain and breast.[5–8] Commonly, small-diameter PET systems are constructed with multiple (flat) detector modules arranged in a polygonal structure. However, the polygonal shape of the

small-diameter PET system creates undesirable space or gaps between the detector modules and results in a measurement loss of PET data. Because of the considerable influence of the data loss on the quality of reconstructed images, compensation methods are required that can reduce the distortions due to gaps between the PET detector modules.

Various methods have been proposed to compensate for the data loss from detector gaps prior to or during the image reconstruction. One type of approach is the gap filling method which estimates the missing data from the measured data prior to image reconstruction, based on several different algorithms. These algorithms include bilinear interpolation, model-based methods, and filtering methods in different domains, such as the constrained Fourier space method and the discrete cosine transform (DCT) domain filter.[9–11] Van Velden *et al*. compared these gap filling methods by assessing gap filling effects for high resolution research tomography (HRRT) system.[12] Another approach is the use of statistical image reconstruction, such as maximum *a posteriori* (MAP), which could be defined by the combination of likelihood and prior probability.[13,14] The main idea in this approach is that the prior probability used in the MAP reconstruction could reduce the gap distortion in the reconstructed images.

In the current study, we suggest another approach for gap compensation, which finds solutions with incomplete data sets based on compressed sensing (CS) theory. The CS, also known as compressive sampling, is a new method with which to precisely reconstruct a signal by utilizing fewer samples than is generally thought necessary.[15,16] To realize this, CS relies primarily on the sparsity of the signal, which is also commonly observed in the medical images. For instance, the Shepp-Logan phantom in the image domain [Fig. 1(a)], which may imitate brain structures, has sparsity in the gradient magnitude domain of the image as shown in Fig. 1(b).[17,18] In Fig. 1(a), the intensities of most pixels are not zero. On the other hand, the transform of the image into gradient magnitude domain [Fig. 1(b)] leads to much increased number of zero pixels.

When measured data are not sufficient to reconstruct an image accurately, the reconstructed image could contain undesirable artifacts. According to the CS theory, however, if we recover the sparsity in a domain other than the image domain, then the reconstructed image could have higher than expected accuracy although incomplete measured data were used. Therefore, in this study we applied the CS-based total variation (TV) minimization method, which was originally suggested for cone-beam computed tomography (CT) reconstruction,[17] to PET image reconstructions to reduce the artifacts arising from the detector gaps in small-diameter PET systems. We incorporated the TV minimization in several common PET reconstruction algorithms and evaluated the gain in image quality achieved by the application of TV minimization.

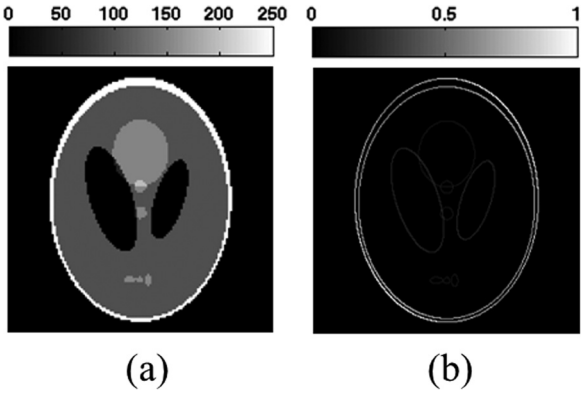

FIG. 1. (a) Shepp-Logan phantom and (b) magnitude of the image gradient of (a)

In the following sections, the proposed gap-compensable reconstruction method, experimental datasets, and evaluation methods are presented. Special focus is made on the evaluation of improvement in intensity uniformity, quantitative accuracy, and convergence properties using the numerical and real experimental datasets. Section IV summarizes the key findings that were observed throughout the study.

II. MATERIALS AND METHODS

II.A. Gap-compensable reconstruction methods

Common reconstruction algorithms for small-diameter PET systems yield undesirable artifacts in the reconstructed images if they are applied without any compensation procedures. To reduce artifacts and recover more accurate images, we propose the application of the constrained, TV minimization method to these common reconstruction algorithms. The constrained, TV minimization method derives from the compressed sensing (CS) theory. The CS theory mainly relies on the recovery of the sparsity of a target signal to reconstruct an accurate original signal from incomplete datasets. Since medical images have the sparsity in the gradient magnitude domain [Fig. 1(b)],[17,18] if this sparsity can be recovered, it will be possible to reconstruct more accurate images.

To recover the sparsity of an image in the gradient magnitude domain during image reconstruction, a solution must be found that minimizes TV of the estimate and is consistent with the measured data. The equation for this constrained, TV minimization reconstruction is of the form

$$\arg\min_f \|f\|_{TV} = \sum_x \sum_y |\nabla f(x,y)|, \tag{1}$$

$$\text{subject to } Mf = g, \tag{2}$$

where f is a reconstructed image, g is the measured PET data (sinogram), and M is the system matrix. This optimization problem can be solved iteratively by recasting it as second-order cone programs (SOCPs).[19]

Since the SOCPs require reformulation of the problem, Sidky *et al.*[17] introduced a separated two-step iterative method. The first step is to solve Eq. (2) separately, which enforces measured data to an image f. Then, the next step is to find an image f which minimizes TV of an estimate from the first step. The two separate steps in each iteration make it possible to simplify the optimization problem utilizing existing reconstruction algorithms. This method has significance as a TV-based algorithm that exploits the exact

reconstruction principle for linear system tomographic image reconstructions.

To solve the constrained, TV minimization problem by the separated two-step iterative methods, in this study, three commonly used image reconstruction algorithms were employed in the first step: the algebraic reconstruction technique (ART), ordered subsets expectation maximization (OSEM), and row-action maximum likelihood algorithm (RAMLA). Next, TV minimization by the gradient descent method was performed as the second step. Depending on the method used in the first step, these TV methods can variously be referred to as ART-TV, OSEM-TV or RAMLA-TV. The performances of these three TV methods were then compared to the conventional non-TV methods (ART, OSEM, and RAMLA).

Sections II A 1 and II A 2 provide a brief explanation of the methods used in each step of the TV-based reconstruction methods for detector gap compensation.

II.A.1. Reconstruction algorithms

Although the three reconstruction algorithms used in this study are well known, their mathematical formula are briefly described here so as to provide information about the parameters that are adaptively used according to the data.

ART has been commonly applied to CT data and is a direct iterative reconstruction algorithm used to solve the linear equations expressing the detection process.[20] In ART, the measurement $\boldsymbol{g}$ is expressed by the linear equations of the system coefficient matrix $\boldsymbol{M}$ and unknown source vector $\boldsymbol{f}$ as shown in Eq. (2). ART updates an image at each iteration n as

$$f_j^{(n,m+1)} = f_j^{(n,m)} + \lambda^{(n)} \left\{ M_{ij} \frac{g_i - \sum_k M_{ik} f_k^{(n,m)}}{\sum_j M_{ij}^2} \right\}, \tag{3}$$

where $\lambda^{(n)}$ is a relaxation parameter which is gradually decreased over iteration number and M_{ij} is a system matrix element representing the probability that a photon emitted from a source site (voxel) of j would be detected in measurement bin of i.

Assuming Poisson counting statistics, Eq. (2) can be expressed statistically as follows:

$$P_L(g|f) = \Pi_i \left[\frac{\langle M^i, f \rangle^{g_i}}{g_i!} \exp\left(-\langle M^i, f \rangle\right) \right], \tag{4}$$

where mean of measurement g_i is

$$\bar{g}_i = \langle M^i, f \rangle = \sum_j M_{ij} f_j. \tag{5}$$

The statistical optimization problem is then to find an estimate f when the likelihood probability of Eq. (4) reaches to maximum value. The maximum likelihood (ML) problem for PET measurement can therefore be solved using the famous EM statistical reconstruction algorithm, shown in the following equation:[21,22]

$$f_j^{n+1} = \frac{f_j^n}{\sum_i M_{ij}} \sum_i \frac{M_{ij} g_i}{\sum_k M_{ik} f_k^n}. \tag{6}$$

Due to the slow convergence rate of the EM algorithm, Hudson and Larkin[23] proposed the OSEM algorithm which could accelerate the convergence rate by performing similar procedures to the EM algorithm with subsets ($\{S_m\}$) of measurements as follows:

$$f_j^{(n,m+1)} = \frac{f_j^{(n,m)}}{\sum_{i \in S_m} M_{ij}} \sum_{i \in S_m} \frac{M_{ij} g_i}{\sum_k M_{ik} f_k^{(n,m)}}. \tag{7}$$

However, the OSEM is not a globally convergent algorithm, in which the closeness to the solution is dependent on the subset number that is used. Therefore, this method exhibits a convergence problem depending on the subset numbers. On the contrary, RAMLA is an accelerated algorithm overcoming the nonconvergence of OSEM as follows:[24]

$$f_j^{(n,m+1)} = f_j^{(n,m)} + \lambda^{(n)} f_j^{(n,m)} \sum_{i \in S_m} M_{ij} \left(\frac{g_i}{\sum_k M_{ij} f_k^{(n,m)}} - 1 \right). \tag{8}$$

In Eq. (8), the relaxation parameter $\lambda^{(n)}$ controls the convergence.

II.A.2. Total variation minimization

The second step is to minimize the total variation of the current image $\boldsymbol{f}^{(n,m+1)}$ which is obtained after the first step described in the previous section.

$$\min \|f\|_{TV}. \tag{9}$$

The difference from Eq. (1) is that Eq. (9) does not have additional conditions on the closeness to the solution. Thus, the second step becomes a straightforward unconstrained optimization problem which can be simply implemented by the gradient descent method. The gradient descent method decreases $||\boldsymbol{f}||_{\mathrm{TV}}$ in the direction of the negative gradient of $||\boldsymbol{f}||_{\mathrm{TV}}$ as follows:

$$f^{l+1} = f^l - \alpha d \frac{v}{\|v\|_2}, \tag{10}$$

where

$$\begin{aligned} v_{x,y} &= \frac{\partial \|f^l\|_{\mathrm{TV}}}{\partial f_{x,y}^l} \\ &= \frac{\left(f_{x,y}^l - f_{x-1,y}^l\right) + \left(f_{x,y}^l - f_{x,y-1}^l\right)}{\sqrt{\varepsilon + \left(f_{x,y}^l - f_{x-1,y}^l\right)^2 + \left(f_{x,y}^l - f_{x,y-1}^l\right)^2}} \\ &\quad + \frac{f_{x,y+1}^l - f_{x,y}^l}{\sqrt{\varepsilon + \left(f_{x,y+1}^l - f_{x-1,y+1}^l\right)^2 + \left(f_{x,y+1}^l - f_{x,y}^l\right)^2}} \\ &\quad + \frac{f_{x+1,y}^l - f_{x,y}^l}{\sqrt{\varepsilon + \left(f_{x+1,y}^l - f_{x,y}^l\right)^2 + \left(f_{x+1,y}^l - f_{x+1,y-1}^l\right)^2}}, \end{aligned} \tag{11}$$

and where ε is a small positive number, d represents the amount of change in the first reconstruction step and α is a

parameter to balance between the amount of change in the reconstruction step and TV minimization step. The $l\ (=0,\ldots, L-1)$ is the iteration number in the TV minimization step and $\boldsymbol{f}^0$ in Eq. (10) should be set to the current estimate, $\boldsymbol{f}^{(n,M-1)}$, obtained from the first reconstruction step. The estimate $\boldsymbol{f}^{L-1}$ after the TV minimization step is then used at the next iteration, $n+1$, as the initial estimate, $\boldsymbol{f}^{(n+1,0)}$, in the reconstruction step.

II.B. Experimental datasets

To explore the feasibility of current methods, numerical experiments simulated a PET scanner with a hexagonal geometry. The scanner has six gaps for which the angle is θ degree.

We used numerical disk (top) and Shepp-Logan (bottom) phantoms shown in Fig. 2(a). In Figs. 2(b)–2(e), each row shows the ideal (nongapped) and gapped sinograms of the disk and Shepp-Logan phantoms, respectively. The gapped sinograms on columns (c–e) were generated with the gap angle $\theta=5^\circ$, 10°, and 15°, respectively. In Fig. 2, we can see the data loss of an x-strip shape in the sinogram resulted from the gap angle (θ).

To simulate more practical situations, we also generated noisy sinograms of the Shepp-Logan phantom at three different noise levels. The noisy sinograms (levels 1–3) were generated as follows from the noiseless (denoted by level 0) sinogram in Fig. 2(b).

After scaling down the sinogram by dividing the data by the scale factor, we generated Poisson random noise in each voxel. The larger scaling factor for the sinogram, the relatively larger the Poisson random noise generated. For each noise level, the scale factor was $2^{\text{level}+2}$, so that the level 3 sinogram was most noisy. After generating the Poisson random noise, we rescaled each sinogram back to have the original total count. The noise level of 1–3 determined the scaling factors of 8, 16, and 32, and the noise quantities were 3, 4, and 6%, respectively, in terms of mean coefficient of variation. We also generated noisy and gapped sinograms. In this simulation, all the sinograms were acquired from 64 angular positions through 180° and were then reconstructed by the non-TV and TV-methods.

We also applied these methods to measured PET data (human brain imaged with F-18-FDG and a uniform cylindrical phantom with F-18) acquired using a clinical PET scanner (Siemens Biograph TruePoint TrueV PET/CT system) without any gaps. From this data, gaps of 5° degree were added to mimic the Siemens HRRT of eight detector banks following the procedures. First, the number of detector bins (x-axis in sinogram) was reduced ($168 \rightarrow 76$) in consideration of the different ring diameters of the Biograph and HRRT. We did not change the bin size (4.01 mm) and the angular sampling number because the measurement of the spatial resolution was not the aim of this study. Then, the gapped HRRT-like sinogram was generated by multiplying the nongapped sinogram by the gap mask of the HRRT. The nongapped and gapped HRRT-like sinograms were reconstructed using the non-TV and TV methods.

For the comparative experiment to other gap compensation methods (see II D in detail), a silicon multiplier (SiPM)-based small animal PET scanner,[25–27] which has been developed in our group for simultaneous PET/MR imaging studies, was considered. The scanner consists of eight detector modules with the diameter of 9 cm and the compensation methods were applied to the missing data due to the presence of gaps of 9.2° in a silicon multiplier. Both the binned and angular sampling numbers were 128.

II.C. Evaluation methods

For the quantitative evaluation, the uniformity of the inside circle and outside ring of the disk phantom were calculated separately to determine how each feature was affected by detector gaps. The uniformity u in each region of interest (ROI) was defined by

$$u = \left(1 - \frac{\text{standard deviation of ROI}}{\text{mean of ROI}}\right) \times 100\ (\%). \quad (12)$$

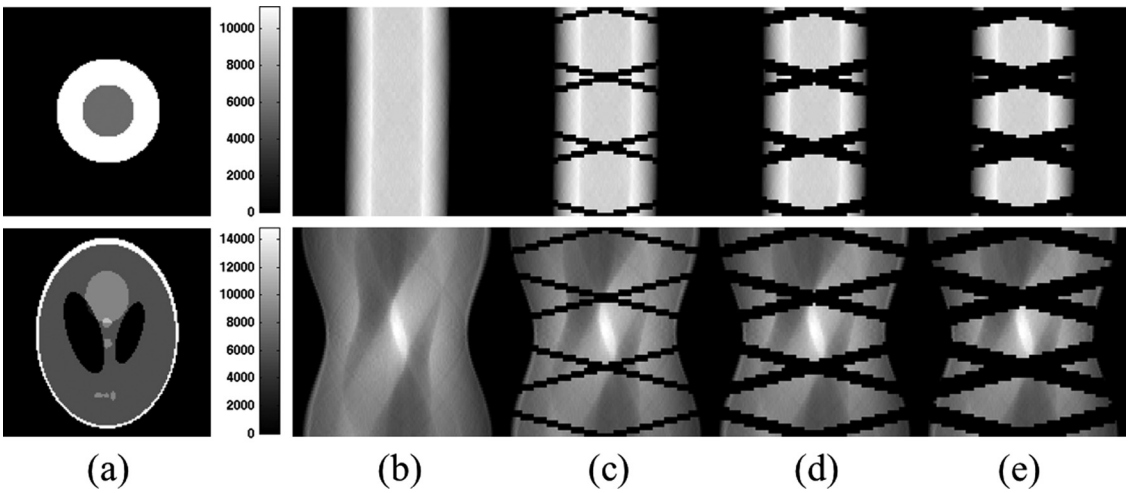

FIG. 2. First row: Sinograms of the disk phantom, second row: Sinograms of the Shepp-Logan phantom (a) from the PET simulator (b) without gaps between the detector modules, with gaps of angle (c) $\theta=5^\circ$, (d) $\theta=10^\circ$, and (e) $\theta=15^\circ$.

The more uniform images should have a uniformity u closer to 100%.

For the Shepp-Logan phantom, the percent root mean squared error (RMSE) was calculated as follows:

$$\%\text{RMSE}\ (f^{\text{rec}}) = \sqrt{\frac{\sum_x \sum_y \left[f^{\text{rec}}_{x,y} - f^{\text{org}}_{x,y}\right]^2}{\sum_x \sum_y \left[f^{\text{org}}_{x,y}\right]^2}} \times 100\ (\%). \tag{13}$$

For the percent RMSE calculations of the reconstructed image $\boldsymbol{f}^{\text{rec}}$, the original phantom image $\boldsymbol{f}^{\text{org}}$ was employed as the normalization factor.

Difference images were calculated between the reconstructed images of nongapped and gapped HRRT-like sinograms as well as the sum of the absolute difference.

II.D. Comparison to other methods

As mentioned previously, there are two approaches to compensate the gap artifacts in PET that have been attempted. Therefore, an empirical comparison was performed between the current methods and these prior approaches. Among these methods, the most advanced one in terms of the theoretical background and performances in the gap compensation was employed. One method is the DCT domain filtering, an example of the gap filling methods that are applied prior to image reconstruction, and the other method is the block sequential regularized expectation maximization (BSREM) method, which is an example of the MAP reconstruction algorithms.[11,14]

For this comparative experiment, both methods were applied to the missing data in a SiPM-based small animal PET scanner which consists of eight detector modules and has gap angles of 9.2°. For the DCT and BSREM methods, software code was used that was either provided by the authors of each of method, or was generated under close collaboration with them to minimize any bias in results due to an erroneous understanding or implementation of the algorithms.

For the DCT methods, the DCT domain filter was obtained from the gap mask of 9.2° in the SiPM PET detector ring. The gap mask has two specific patterns in the DCT domain, horizontal lines and a "V"-shape, as in Ref. 11. These patterns were utilized to filter out the DCT coefficient corresponding to the gap data (x-strips in the gapped sinogram). The gap region of the x-strips in the measured sinogram was filled iteratively by two successive operations in the DCT and sinogram domain. For a given iteration, an estimate of the nongapped sinogram was obtained by the DCT domain filter, followed by a taking of the inverse DCT transform. The gap data in the sinogram domain were extracted from the estimated nongapped sinogram.

The BSREM was originally proposed to solve MAP estimation problems in an accelerated manner,[28,29] and it was shown that BSREM algorithms with a convex-nonquadratic penalty function were effective in the compensation of gap in PET data.[13] In this study, we used the Bouman and Sauer prior function[30] for BSREM because this combination showed the best performance in preliminary studies. The regularization parameter of 0.5 was used in the BSREM.

For the comparison, the OSEM-TV and RAMLA-TV methods were used because they showed excellent robustness to statistical noise and the gap artifact in real datasets (see the Results section). The numerical Shepp-Logan phantom and real datasets (comprising a uniform phantom and human brain dataset) used in Sec. II B were also employed to evaluate these gap compensation methods. The size of the real data was scaled to fit within the diameter of the SiPM PET scanner.

III RESULTS

III.A. Numerical experiments

III.A.1. Disk phantom

The performance of the six different non-TV and TV reconstruction methods, ART, ART-TV, OSEM, OSEM-TV, RAMLA, and RAMLA-TV, were compared using the numerical disk phantom. In each reconstruction, 32 iterations were performed. For OSEM and RAMLA in the non-TV and TV reconstructions, the subset numbers were 8 and 64, respectively. Figures 3–5 show the results of the simulation for the disk phantom. The fixed relaxation parameter of 1 was used for each iteration of ART and ART-TV. For the TV-methods, Eq. (10) was iterated 20 times ($=L$) with the

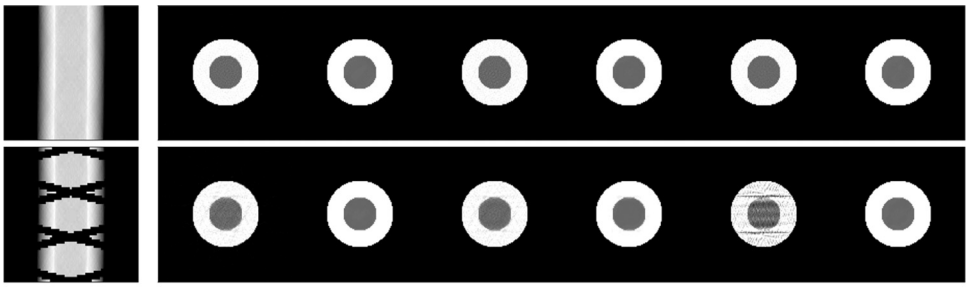

FIG. 3. Reconstructed images of the numerical disk phantom (b)–(g) from sinograms in column (a) using six different non-TV and TV reconstruction methods, (b) ART, (c) ART-TV, (d) OSEM, (e) OSEM-TV, (f) RAMLA, and (g) RAMLA-TV. The first and second rows show the data without and with gaps of angle $\theta = 10°$ between detector modules, respectively.

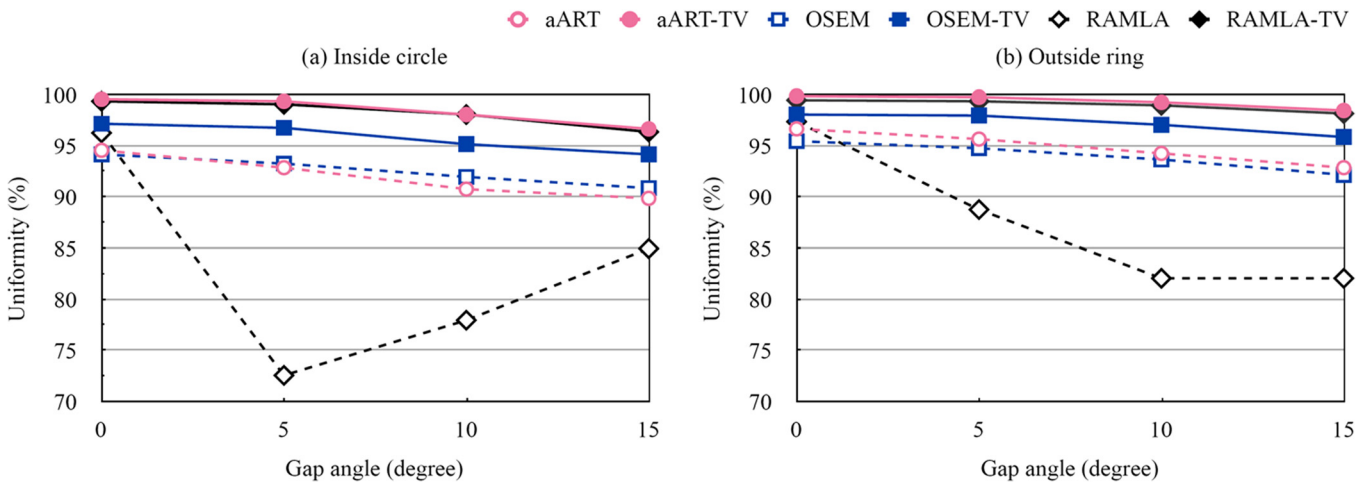

FIG. 4. For the disk phantom, (a) the inside circle and (b) outside ring uniformity graphs over gap angle for the six different non-TV and TV reconstruction methods (with the iteration of 32).

parameter α set to 0.2 in every step, as has been proposed previously.[31]

The first and second rows of Fig. 3(a) show the ideal (nongapped) and 10° gapped sinograms which were used for the reconstructions. Reconstructed images from the six reconstruction algorithms are illustrated in Figs. 3(b)–3(g). TV methods [(c) ART-TV, (e) OSEM-TV, and (g) RAMLA-TV] produced accurate images for ideal and gapped sinograms, while non-TV methods [(b) ART, (d) OSEM, and (f) RAMLA] for gapped sinogram resulted in small fluctuations at uniform regions, which caused the uniformity to decrease.

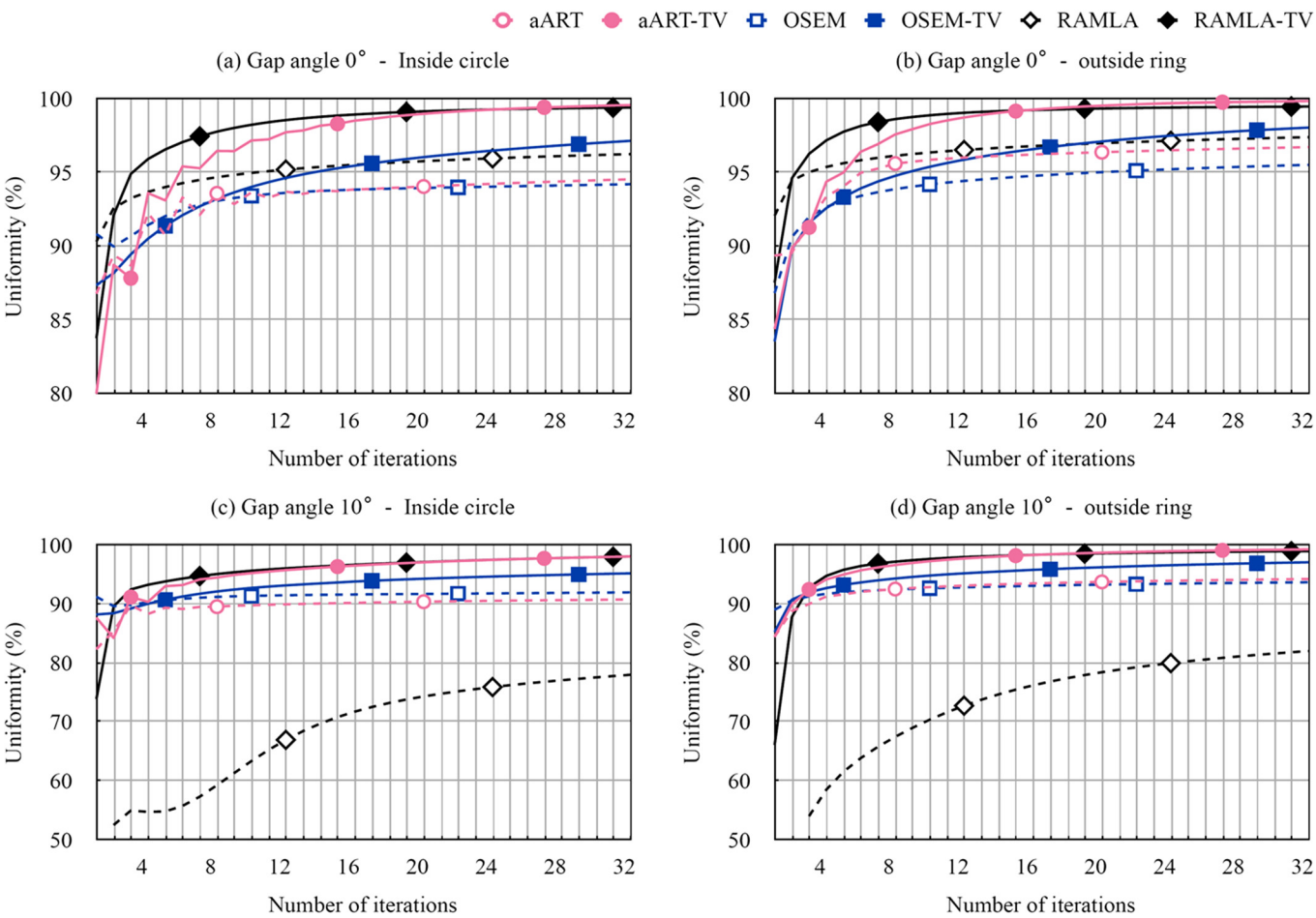

FIG. 5. Inside circle (a, c) and outside ring (b, d) uniformity graphs over the number of iterations for the six different non-TV and TV reconstruction methods using sinograms of gap angles 0° and 10°. (For OSEM and RAMLA in the non-TV and TV reconstructions, the subset numbers were 8 and 64, respectively.)

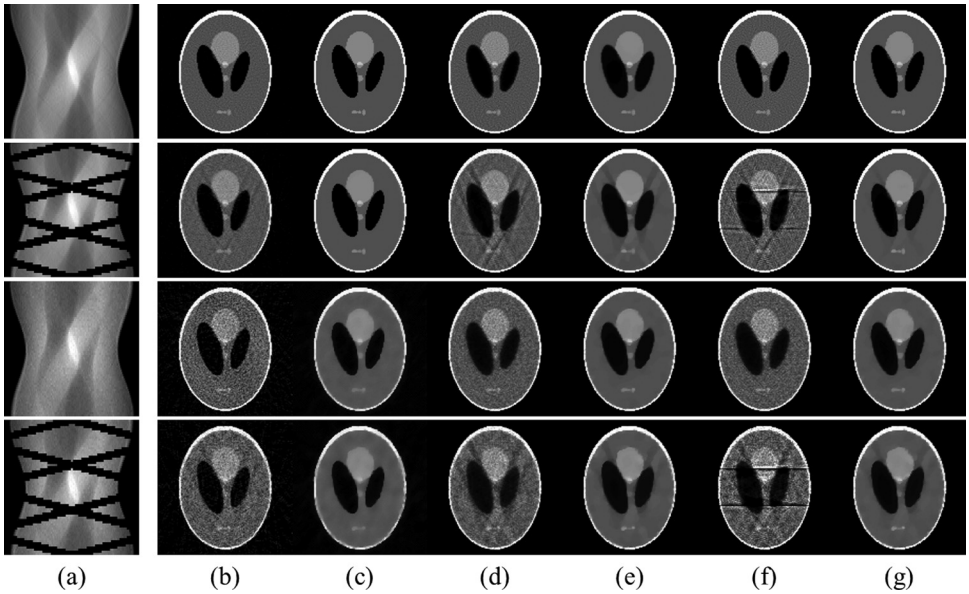

FIG. 6. Reconstructed images of the Shepp-Logan phantom (b)-(g) from sinograms in column (a) using six different non-TV and TV reconstruction methods, (b) ART, (c) ART-TV, (d) OSEM, (e) OSEM-TV, (f) RAMLA, and (g) RAMLA-TV. The first and second rows show the noiseless (noise level 0) data without and with gaps of angle $\theta = 10°$, respectively. The third and fourth rows show the noisy data (noise level 1).

As shown in Figs. 3 and 4, the uniformities from all TV methods were better than those from non-TV methods in both the inside circle and outside ring of the disk phantom. The results obtained from the gapped sinogram more clearly showed the better performance of the TV methods. While the artifacts due to data loss in the sinograms were visible in the reconstructions of the non-TV methods, the TV methods produced more accurate images that were indistinguishable

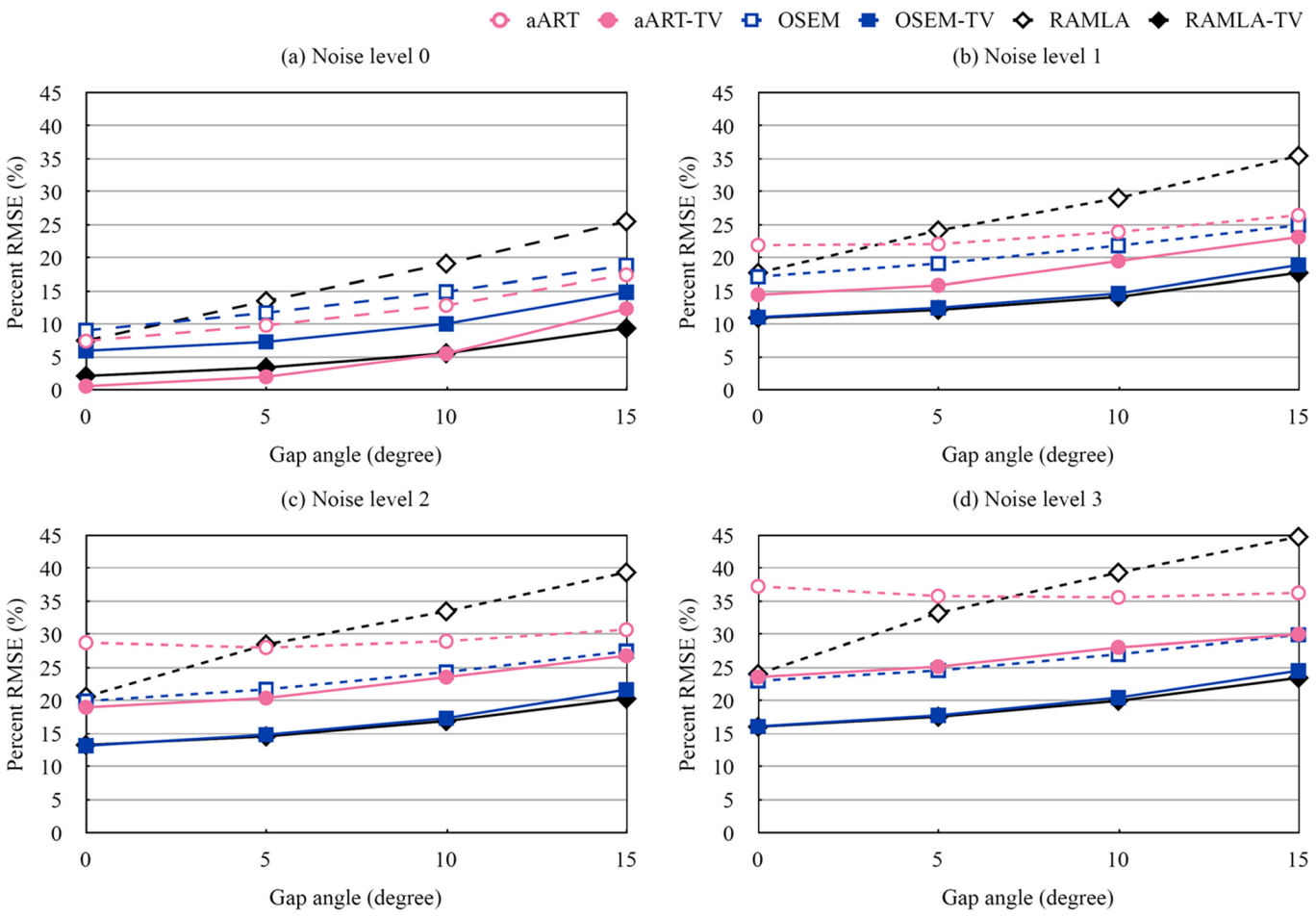

FIG. 7. For the Shepp-Logan phantom, percent RMSE graphs over gap angle at noise level (a) 0 (noiseless), (b) 1, (c) 2, and (d) 3.

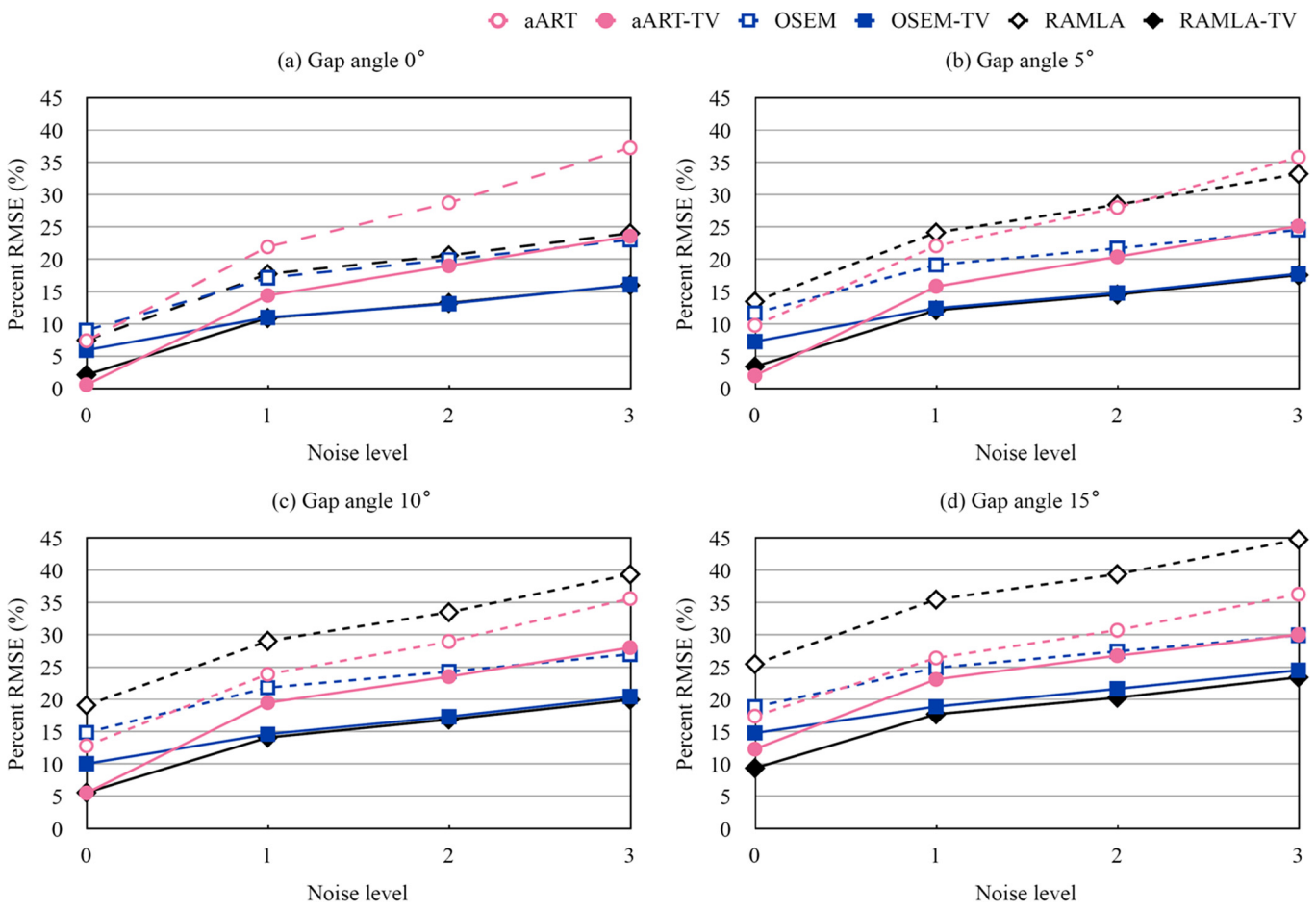

FIG. 8. For the Shepp-Logan phantom, percent RMSE graphs over noise level with the gap angle $\theta =$ (a) 0°, (b) 5°, (c) 10°, and (d) 15°.

from the original disk phantom. Only the OSEM-TV resulted in a slight blurring near the border between the inside circle and the outside ring [Fig. 3(e)]. The uniformities of the ART-TV and RAMLA-TV reconstructions were greater than 95% for both the inside circle and the outside ring even when the gap angle increased to 15° (Fig. 4).

In particular, the TV minimization step in RAMLA-TV succeeded in removing horizontal artifacts that appeared in the RAMLA reconstructions [Figs. 3(f) and 3(g)]. These horizontal artifacts were located near the center when the gap angle was small, and they spread apart from each other as the gap angle increased. Since the horizontal artifacts did not occur within the inside circle when the gap angle became large, the uniformity of the inside circle improved with the increasing gap angle (Fig. 4). Figure 5 shows the uniformity graphs over the number of iterations. For all reconstructions except for RAMLA, the uniformity almost converged after 16 iterations.

III.A.2. Shepp-Logan phantom

Using the Shepp-Logan phantom, six different non-TV and TV reconstruction methods were evaluated by the percent RMSE. The results for the Shepp-Logan phantom are summarized in Figs. 6–9. Each reconstruction was performed with 32 iterations, and the final result was chosen as the iteration with the lowest percent RMSE. The subset numbers in OSEM and RAMLA were the same as in the disk phantom. ART and ART-TV used a fixed relaxation parameter of 1. In all TV-methods, the TV minimization step was iterated 20 times with the parameter α set to 0.2.

The images in Fig. 6(a) depict the sinograms used in the reconstructions. The first row is the ideal sinogram without noise or detector gaps. The second and third rows show the sinogram deteriorated due to either 10° gaps or level 1 noise, respectively. The sinogram on the fourth row was affected by both detector gaps and noise, which would be the most realistic case. Figures 6(b)–6(g) show the reconstructed images using the six reconstruction methods. All TV methods [Figs. 6(c), 6(e), and 6(g)] recovered images that were qualitatively undistinguishable from the original Shepp-Logan phantom. Only the OSEM-TV [Fig. 6(e)] produced slightly smoothed edges. Conversely, the non-TV methods [Figs. 6(b), 6(d), and 6(f)] produced fluctuations of intensity over the reconstructed images.

The data loss from the detector gaps propagated into the reconstructed images except for those reconstructed by the ART-TV method and yielded diagonal artifacts (second and fourth rows in Fig. 6). As in the reconstructions of the disk phantom, horizontal artifacts were shown in RAMLA. However, the TV minimization step drastically reduced these artifacts.

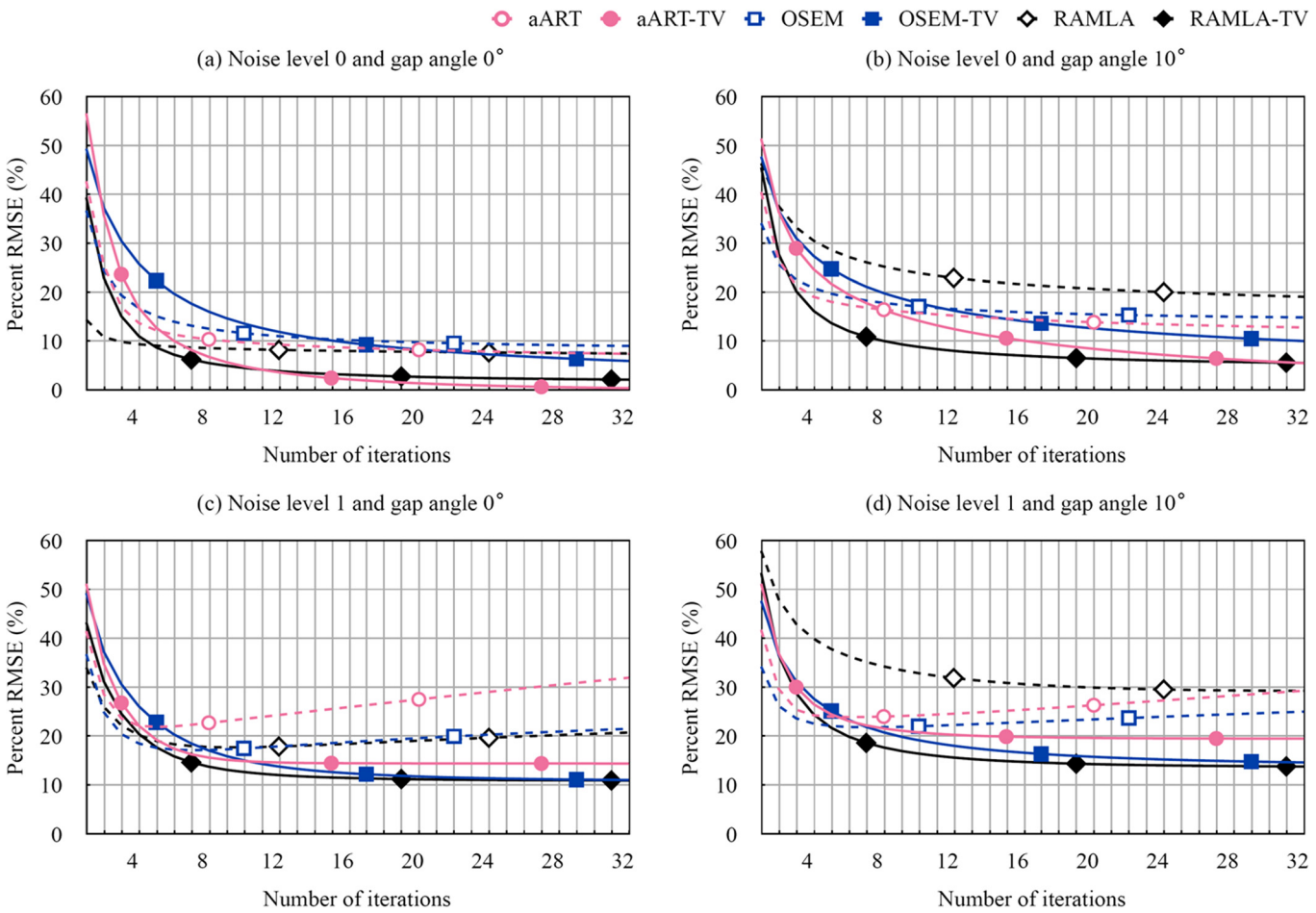

FIG. 9. For the Shepp-Logan phantom, percent RMSE graphs over the number of iterations using (a) the sinogram with gap angle $\theta = 0°$ and noise level 0 (noiseless), (b) $\theta = 10°$ and noise level 0, (c) $\theta = 0°$ and noise level 1, (d) $\theta = 10°$ and noise level 1. (For OSEM and RAMLA in the non-TV and TV reconstructions, the subset numbers were 8 and 64, respectively.)

All non-TV methods caused a propagation of Poisson random noise in reconstructed data (third and fourth rows in Fig. 6). Among these three reconstructions, ART produced the highest amplitude of noise, and the streak artifacts were propagated into the outside ring of the phantom. Although the application of TV minimization resulted in slightly smoothed edges of the images, the propagation of Poisson random noise in the data was effectively reduced.

Figures 7 and 8 illustrate the percent RMSE graphs as a function of gap angle and noise level. For each noise level, the way in which the percent RMSEs changed depending on gap angle are shown in Fig. 7. As in the reconstructed images, the data loss due to the detector gaps most severely affected the RAMLA reconstructions: The increasing percent RMSE as gap angle increased was larger than in other methods. Figure 8 shows how the percent RMSEs changed depending on the noise level for each gap angle. The TV methods created less percent RMSEs than the non-TV methods as the gap angle and the noise level increased. The increasing amount of percent RMSE as noise level increased for ART-TV seems to be larger than in the other two TV methods, which may indicate that OSEM-TV and RAMLA-TV are more robust to noise in the sinograms.

Figure 9 shows the percent RMSE graphs over the number of iterations. For the noiseless sinogram, the errors converged even if there existed the data loss in the sinogram as shown in Fig. 9(b). When the sinograms were contaminated by the Poisson random noise, the errors for the non-TV methods did not converge, while those for the TV methods converged as shown in Figs. 9(c) and 9(d). This indicates that the TV minimization step could control the propagation of Poisson noise into the reconstructions.

III.B. Evaluation using real measurement data

The HRRT-like sinograms of human brain and of the uniform cylindrical phantom were reconstructed with the non-TV and TV methods. The reconstructed images had the dimension of $128 \times 128 \times 109$ ($2.44 \times 2.44 \times 2.03$ mm^3). The HRRT-like sinograms were generated with gap angle of 5 degree. For ART and RAMLA, the initial relaxation parameters (λ_0) were set to 0.5 and 0.2, respectively, and these parameters gradually decreased over the iteration by $\lambda_0/(\text{iteration} + 1)$. OSEM with non-TV and TV methods were performed with a single subset. In the TV methods, the performance for gap compensation was evaluated with

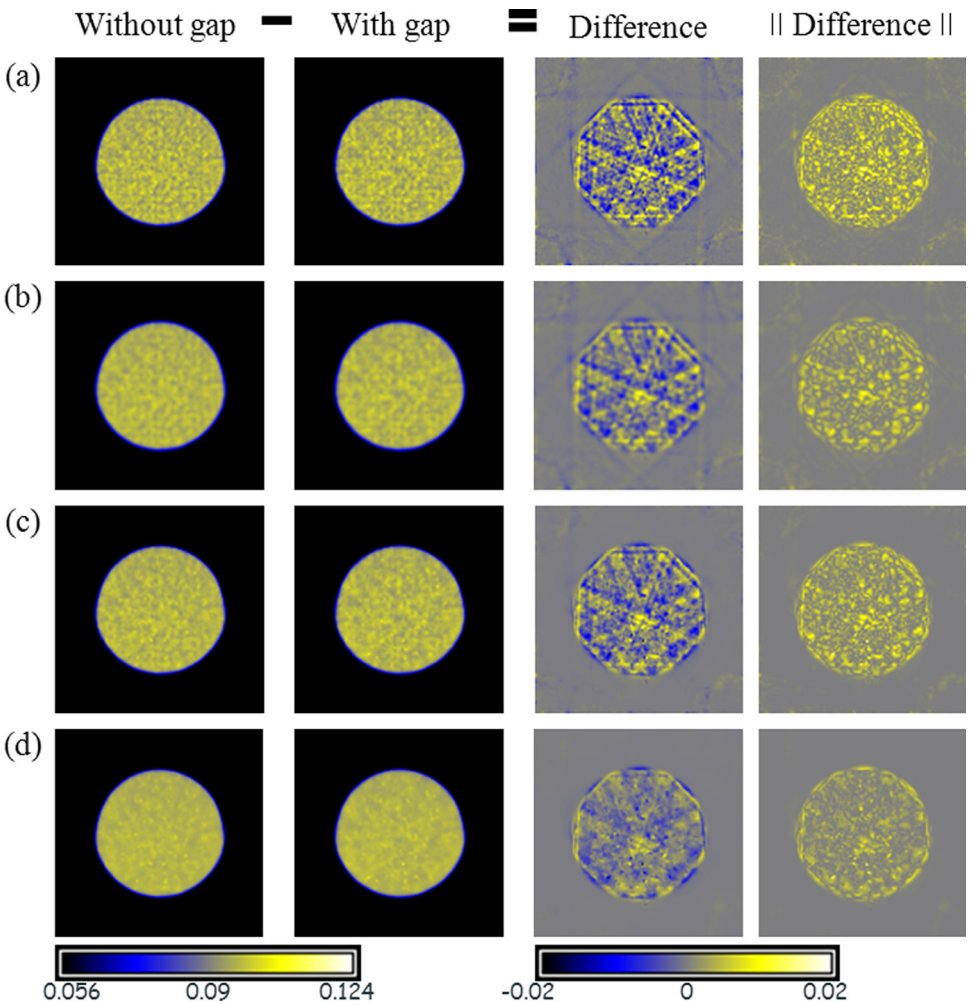

FIG. 10. Difference images between two reconstructions from nongapped and gapped PET data of uniform cylindrical phantom; (a) OSEM, (b) OSEM post-smoothed by Gaussian function (FWHM = 4.9 mm), OSEM-TVs with (c) $\alpha = 0.05$: $L = 10$ and (d) $\alpha = 0.2$: $L = 20$.

various values of α and L. The TV control parameter, α, was varied in the range of 0.05–0.2 with an interval of 0.05. The TV iteration number, L, was set to 10 and 20. The difference images between reconstructions from the nongapped and gapped PET data were then obtained.

Figure 10 representatively shows the transverse slices of OSEM and OSEM-TV reconstructions from nongapped and gapped sinograms of the uniform cylindrical phantom with different parameters (α and L) and their difference images. All non-TV reconstructions were postfiltered with a Gaussian smoothing kernel (FWHM = 4.9 mm). In Fig. 10, only the results of the non-TV and TV reconstructions of OSEM (64 iterations) with the lowest ($\alpha = 0.05$ and $L = 10$) and highest ($\alpha = 0.2$ and $L = 20$) parameters are demonstrated.

TABLE I. For real HRRT-like sinograms of uniform cylindrical phantom, sum of absolute difference between reconstructions of nongapped and gapped data which were resulted from non-TV and TV methods of ART, OSEM, and RAMLA.

	ART (6 iterations)	OSEM (64 iterations)	RAMLA (10 iterations)
Non-TV	26.27	11.31	13.9
Non-TV + Gaussian (4.9 mm)	18.57	7.88	8.9
TV ($\alpha = 0.05$, $L = 10$)	16.41	8.2	10.24
TV ($\alpha = 0.05$, $L = 20$)	13.14	6.66	8.98
TV ($\alpha = 0.1$, $L = 10$)	13.62	7.67	9.29
TV ($\alpha = 0.1$, $L = 20$)	10.57	6.19	7.85
TV ($\alpha = 0.15$, $L = 10$)	8.13	7.31	8.7
TV ($\alpha = 0.15$, $L = 20$)	9.68	5.74	7.01
TV ($\alpha = 0.2$, $L = 10$)	11.61	6.97	8.17
TV ($\alpha = 0.2$, $L = 20$)	9.33	5.35	6.35

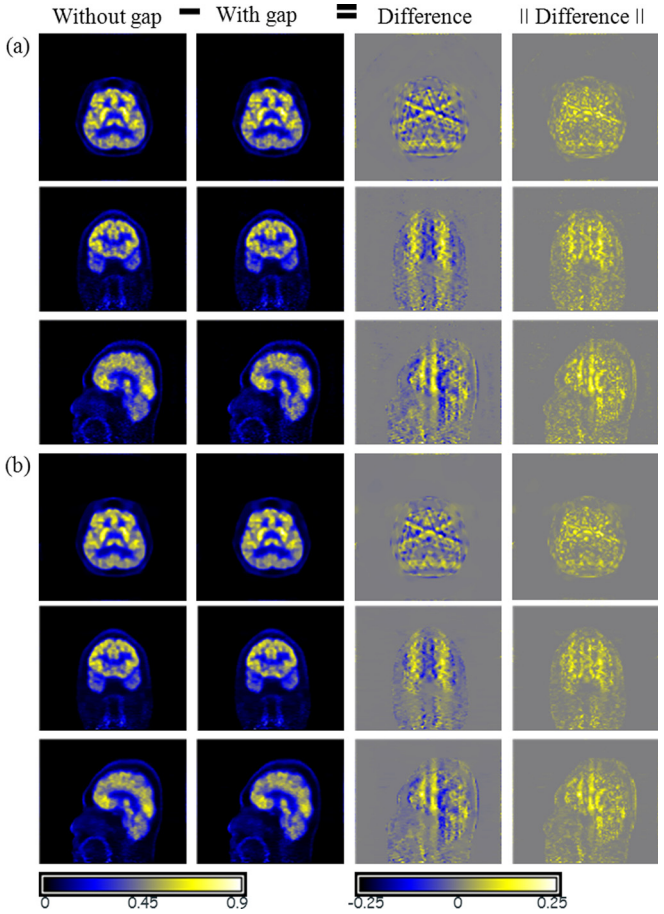

FIG. 11. Planes of (a) OSEM and (b) OSEM-TV reconstructions from nongapped and gapped PET data of human brain; transverse (first row), coronal (middle), sagittal (bottom) planes. (OSEM-TV used $\alpha = 0.15$ and $L = 10$.)

The difference images suggest that the higher TV parameter reduced more artifacts that were propagated from the missing data in the gapped sinogram.

For the uniform cylinder with HRRT-like gaps, Table I shows the sum of absolute difference images in the non-TV and TV methods of ART, OSEM, and RAMLA. All TV reconstructions reduced the absolute difference between images of nongapped and gapped data more than the non-TV reconstructions. The higher the TV parameters were chosen, the more suppressed were the quantitative difference between the nongapped and gapped sinograms. However, greater losses in image quality were observed such as a blocky pattern.

TABLE II. Comparison of gap compensation methods: %RMSE for noisy data of the Shepp-Logan phantom.

	EM	DCT	OSEM-TV	RAMLA-TV	BSREM
%RMSE	29.5	33.8	22.2	21.1	27.0

Figure 11 show the transverse, coronal, and sagittal planes of OSEM and OSEM-TV reconstructions of human brain data and the corresponding difference images, respectively. The TV parameters, α and L, were set to 0.15 and 10, respectively, for all TV reconstructions. Compared with non-TV reconstructions (a), TV reconstructions (b) effectively reduced the gap artifacts and the noise fluctuations in the reconstructed images.

III.C. Comparison to other methods

The first and second rows in Fig. 12 show the reconstructed images of noiseless and noisy sinograms with SiPM-PET gap for the Shepp-Logan phantom, respectively. The OSEM (a, f) of the gapped sinograms in the first column were compared with the aforementioned gap-compensation methods which were DCT domain filtering followed by OSEM (b, g), OSEM-TV (c, h), RAMLA-TV (d, i), and BSREM (e, j). In Fig. 12, TV reconstructions (OSEM-TV and RAMLA-TV) for the Shepp-Logan phantom yielded a lower percent RMSE than the other two compensation methods, the DCT filter and BSREM (Table II). The undercompensation of the missing sinogram data by the DCT filter resulted in other artifacts, such as the narrow stripe pattern (indicated by arrows) in the uniform region of the reconstructed images.

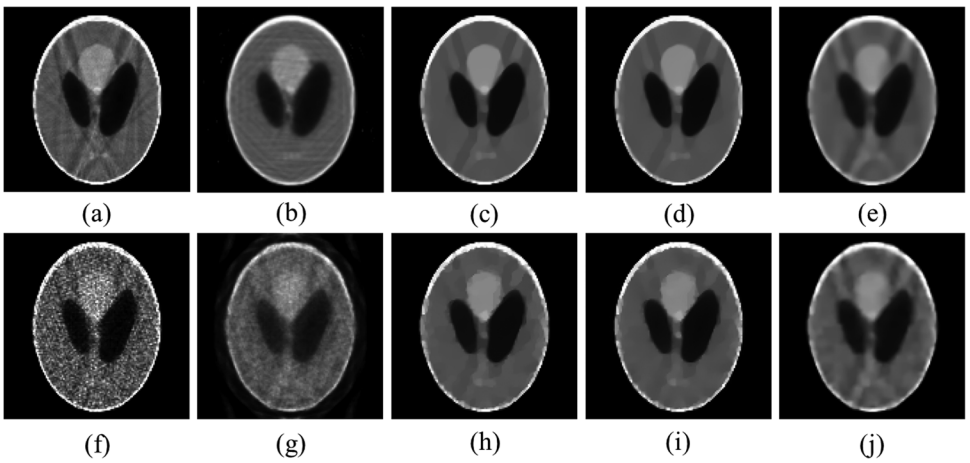

FIG. 12. For the noiseless (top) and noisy (bottom) sinograms of the Shepp-Logan phantom with SiPM-PET gap, the reconstructed images using OSEM (a, f), DCT filter followed by EM (b, g), OSEM-TV (c, h), RAMLA-TV (d, i), and BSREM with the Bouman and Sauer prior (e, j).

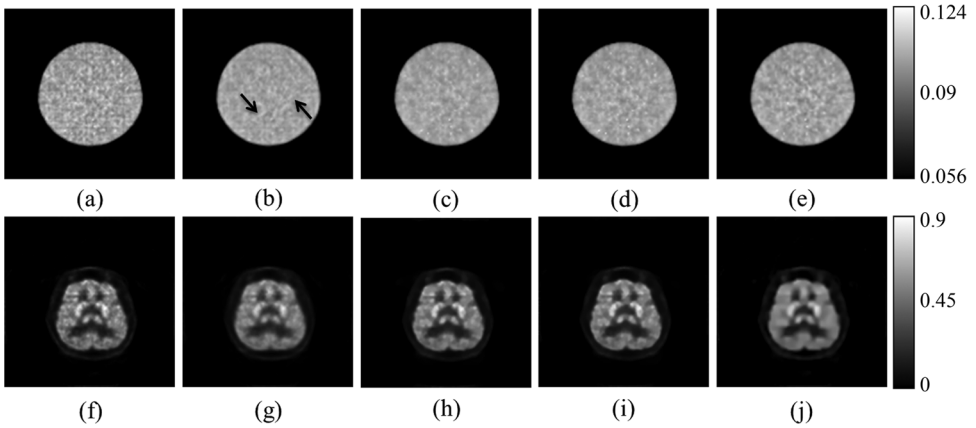

FIG. 13. For the measured data of the uniform cylindrical phantom (top) and human brain (bottom), the reconstructed images using EM (a, f), DCT filter followed by EM (b, g), OSEM-TV (c, h), RAMLA-TV (d, i), and BSREM with the Bouman and Sauer prior (e, j).

Figure 13 shows the reconstructions of measured data of the uniform cylindrical phantom and human brain used in Sec. III B. In the case of the uniform cylindrical phantom, as shown in the top row (a–e), several transverse slices were averaged. The bottom images (f–j) are for real brain data. In the image from the uniform phantom data, the DCT filter resulted in a stripe artifact although the other methods did not. In the case of real brain data, the BSREM created more blurred images than the other methods.

IV. DISCUSSIONS AND CONCLUSIONS

PET scanners with a small ring diameters are designed in a polygonal geometry, which causes gaps between detector modules.[32,33] These gaps lead to data loss in the measured data, thus the sinogram will contain data loss of the x-strip shape from the gaps. The data loss is especially significant if we employ flat panel photomultiplier tubes or multiple layers of scintillation crystals for depth of interaction measurement.[32,34–38] The conventional iterative reconstruction methods (without TV minimization) produced considerable artifacts in the reconstructed images from the gapped sinograms. To reduce the artifacts due to gaps, we applied the TV minimization to the most commonly used iterative reconstruction methods, such as ART, OSEM, and RAMLA. In the numerical experiments in which the disk phantom and the Shepp-Logan phantom were used, we found that the TV methods (ART-TV, OSEM-TV, and RAMLA-TV) produced more accurate images than the conventional non-TV reconstruction methods (ART, OSEM, and RAMLA). This was also confirmed in the analysis of the human brain data and uniform cylindrical phantom data.

The experiments using the disk phantom proved that the TV methods could improve image uniformity. The image uniformity was evaluated for two different regions (the inside circle and outside ring). For the TV methods, all of the reconstructed images yielded excellent uniformity (nearly 100%) and qualitative comparisons with the original image indicated that the methods produced images that were recoverable with very good uniformity (Fig. 4). Conversely, non-TV methods produced lower-quality images containing fluctuations in the uniform region. In particular, the images using RAMLA produced horizontal artifacts, which occurred differently depending on how the number of subsets in RAMLA was set (Fig. 3). However, RAMLA-TV resolved this problem regardless of the number of subsets, so we did not need to pay careful attention to the number of subsets in RAMLA-TV reconstructions.

The next numerical experiment with the Shepp-Logan phantom verified that the TV methods improved the image quality and the convergence rate in terms of percent RMSEs, even for the noisy and gapped sinograms. For the ideal sinogram without noise and gaps, TV methods recovered images which were qualitatively indistinguishable from the original phantom data. When noise and gaps affect the sinograms, the reconstructed images and percent RMSEs of the TV methods appeared differently, depending on the reconstruction algorithm used (ART, OSEM, and RAMLA). For the gapped sinograms, ART-TV resulted in near-exact images, while the artifacts were still propagated into the images when using OSEM-TV and RAMLA-TV (Fig. 6). On the other hand, OSEM-TV and RAMLA-TV showed more robust results to the noisy sinograms than ART-TV. For the same gap angle, ART-TV resulted in larger RMSE increase than OSEM-TV and RAMLA-TV as the noise level increased.

For the TV methods, the percent RMSEs converged after 16 iterations for all cases, while the non-TV methods tended to diverge after a few iterations due to the propagation of noise in the sinogram (Fig. 9). This would be a remarkable property of TV methods for application to practical imaging studies. When handling real data, no stopping criteria exist since we are not able to compute the percent RMSE due to the absence of knowledge about the ground truth. Also, we do not iterate too many times because excessive iterations often cause severe artifacts from noisy and gapped sinograms as shown in numerical experiments; however, this would not be problematic for the TV methods as we have shown that the TV-reconstructed images from the noisy and gapped sinograms almost converge after >16 iterations.

The sinograms from the HRRT-like system were generated from the real human brain and uniform cylindrical

phantom data acquired using clinical whole-body PET scanner by considering the diameter difference between two PET systems and gap size in HRRT. The TV methods were applied with two parameters, α and L. In the case of the uniform phantom, the larger TV parameters resulted in a more efficiently reduced gap artifact. However, the TV reconstructions of the gapped sinogram of brain with the large TV parameters ($\alpha = 0.2$ and $L = 20$) lead to undesirable blocky patterns in the image. Thus, the suitable TV parameters should be selected as a trade-off between the suppression of gap artifact and the noise correlation ("blocky patterns") in the reconstructed images. Further investigation is needed to develop a method that can effectively reduce the gap artifacts without generating the blocky patterns.

Three different gap compensation methods, DCT filter, TV methods, and BSREM, were compared. The gapped sinograms for the SiPM PET system were simulated using a Shepp-Logan phantom. The real data of a uniform cylindrical phantom and human brain were also used. Although all the compensation methods were effective at reducing gap artifacts in the reconstructed images, these artifacts were not completely removed, especially in the simulation data. This would be because the SiPM PET system contains more missing data than the HRRT system. In general, TV methods demonstrated better performance than DCT and BSREM.

The proposed gap-compensable image reconstruction method based on CS (ART-TV, OSEMTV and RAMLA-TV) effectively reduced artifacts from the detector gaps as well as Poisson random noise. OSEM-TV and RAMLA-TV in particular showed distinct potential via the properties of convergence and robustness to different noise levels and gap angle.

ACKNOWLEDGMENTS

This work was supported by grants from the Atomic Energy R&D Program (2008-2003852, 2010-0026012) and the WCU Program (R32-10142) through the KOSEF, which is funded by the Korean Ministry of Education, Science and Technology.

[a]Author to whom correspondence should be addressed. Electronic mail: jaes@snu.ac.kr; Telephone: 822-2072-2938; Fax: 822-745-2938.
[b]Co-first authors.

[1]M. E. Phelps, "Nuclear medicine, molecular imaging, and molecular medicine," J. Nucl. Med. **43**, 13N–14N (2002).
[2]S. S. Gambhir, "Molecular imaging of cancer with positron emission tomography," Nat. Rev. Cancer **2**, 683–693 (2002).
[3]M. G. Pomper and J. S. Lee, "Small animal imaging in drug development," Curr. Pharm. Des. **11**, 3247–3272 (2005).
[4]A. Zhu and H. Shim, "Current molecular imaging positron emitting radiotracers in oncology," Nucl. Med. Mol. Imaging **45**, 1–14 (2001).
[5]M. Schmand *et al.*, "Performance results of a new DOI detector block for a high resolution PET-LSO research tomograph HRRT," IEEE Trans. Nucl. Sci. **45**, 3000–3006 (1998).
[6]A. F. Chatziioannou, "PET scanners dedicated to molecular imaging of small animal models," Mol. Imaging Biol. **4**, 47–63 (2002).
[7]S. Surti and J. S. Karp, "Design considerations for a limited angle, dedicated breast, TOF PET scanner," Phys. Med. Biol. **53**, 2911–2921 (2008).
[8]J. S. Lee, "Technical advances in current PET and hybrid imaging systems," Open Nucl. Med. J. **2**, 192–208 (2010).
[9]H. de Jong, R. Boellaard, C. Knoess, M. Lenox, C. Michel, M. Casey, and A. Lammertsma, "Correction methods for missing data in sinograms of the HRRT PET scanner," IEEE Trans. Nucl. Sci. **50**, 1452–1456 (2003).
[10]J. Karp, G. Muehllehner, and R. Lewitt, "Constrained Fourier space method for compensation of missing data in emission computed tomography," IEEE Trans. Med. Imaging **7**, 21–25 (1988).
[11]U. Tuna, S. Peltonen, and U. Ruotsalainen, "Gap-filling for the high-resolution PET sinograms with a dedicated DCT-domain filter," IEEE Trans. Med. Imaging **29**, 830–839 (2010).
[12]F. Velden, R. Kloet, B Berckel, C. Molthoff, A. Lammertsma, and R. Boellaard, "Gap filling strategies for 3-D-FBP reconstructions of high-resolution research tomograph scans," IEEE Trans. Med. Imaging **27**, 934–942 (2008).
[13]P. Kinahan, J. Fessler, and J. Karp, "Statistical image reconstruction in PET with compensation for missing data," IEEE Trans. Nucl. Sci. **44**, 1552–1557 (1997).
[14]V.-G. Nguyen and S.-J. Lee, "Image reconstruction from limited-view projections by convex nonquadratic spline regularization," Opt. Eng. **49**, 037001 (2010).
[15]E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**, 489–509 (2006).
[16]D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).
[17]E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).
[18]X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" Inverse Probl. **25**, 1–36 (2009).
[19]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).
[20]R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," J. Theor. Biol. **29**, 471–481 (1970).
[21]K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," J. Comput. Assist. Tomogr. **8**, 306–316 (1984).
[22]L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging **1**, 113–122 (1982).
[23]H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," IEEE Trans. Med. Imaging **13**, 601–609 (1994).
[24]J. Browne and A. de Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," IEEE Trans. Med. Imaging **15**, 687–699 (1996).
[25]J. S. Lee, H. S. Yoon, G. B. Ko, S. I. Kwon, M. Ito, C. M. Lee, D. S. Lee, I. C. Song, and S. J. Hong, "Simultaneous PET/MR imaging with MR-compatible Geiger-mode APD PET system," *IEEE Med Imaging Conference PET-MR Workshop* (Knoxville, Tennessee, 2010).
[26]G. B. Ko, H. S. Yoon, S. I. Kwon, S. J. Hong, D. S. Lee, and J. S. Lee, "Development of FPGA-based coincidence units with veto function," Biomed Eng. Lett. 1, 27–31 (2011).
[27]H. S. Yoon, G. B. Ko, S. I. Kwon, C. M. Lee, M. Ito, I. C. Song, D. S. Lee, S. J. Hong, and J. S. Lee, "Initial results of simultaneous PET/MRI experiments with an MR-compatible silicon photomultiplier PET scanner," J. Nucl. Med. (in press).
[28]A. de Pierro and M. Yamagishi, "Fast EM-like methods for maximum "A Posteriori" estimates in emission tomography," IEEE Trans. Med. Imaging **20**, 280–288 (2001).
[29]S. Ahn and J. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," IEEE Trans. Med. Imaging **22**, 613–626 (2003).
[30]C. Bouman and K. Sauer, "Generalized Gaussian image model for edge preserving MAP estimation," IEEE Trans. Image Process. **2**, 296–310 (1993).
[31]E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," J. X-Ray Sci. Technol. **14**, 119–139 (2006).

[32]S. J. Hong *et al.*, "Concept verification of three-layer DOI detectors for animal PET," IEEE Trans. Nucl. Sci. **55**, 912–917 (2008).
[33]S. I. Kwon *et al.*, "Development of small-animal PET prototype using silicon photomultiplier (SiPM): Initial results of phantom and animal imaging studies," J. Nucl. Med. **52**, 572–579 (2011).
[34]M. Ito, J. S. Lee, M. J. Park, K. S. Sim, and S. J. Hong, "Design and simulation of a novel method for determining depth-of-interaction in a PET scintillation crystal array using a single-ended readout by a multi-anode PMT," Phys. Med. Biol. **55**, 3827–3841 (2010).
[35]M. Ito, S. J. Hong, and J. S. Lee, "Positron emission tomography (PET) detectors with depth-of-interaction (DOI) capability," Biomed. Eng. Lett. **1**, 70–81 (2011).
[36]T. Yamaya, N. Hagiwara, T. Obi, T. Tsuda, K. Kitamura, T. Hasegawa, H. Haneishi, N. Inadama, E. Yoshida, and H. Murayama, "Preliminary resolution performance of the prototype system for a 4-layer DOI-PET scanner: jPET-D4," IEEE Trans. Nucl. Sci. **53**, 1123–1128 (2006).
[37]S. Yamamoto, M. Honda, T. Oohashi, K. Shimizu, and M. Senda, "Development of a brain PET system, PET-Hat: A wearable PET system for brain research," IEEE Trans. Nucl. Sci. **58**, 668–673 (2011).
[38]C. M. Lee, S. I. Kwon, G. B. Ko, M. Ito, H. S. Yoon, D. S. Lee, S. J. Hong, and J. S. Lee, "A novel compensation method for the anode gain non-uniformity of multi-anode photomultiplier tubes," Phys. Med. Biol. **57**, 191–207 (2012).