Super-Energy-Resolution Material Decomposition for Spectral Photon-Counting CT Using Pixel-Wise Learning

BINGQING XIE1, YUEMIN ZHU1, PEI NIU1, TING SU2, FENG YANG3, LIHUI WANG4, PIERRE-ANTOINE RODESCH5, LOIC BOUSSEL5, PHILIPPE DOUEK5, and PHILIPPE DUVAUCHELLE6

1 Univ Lyon, INSA Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France
2 Research Center for Medical Artificial Intelligence, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
3 Beijing Jiaotong Univ, School of Computer and Information Technology, Beijing, China
4 Guizhou Univ, Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, School of Computer Science and Technology, Guiyang, China
5 Univ Lyon, CREATIS ; CNRS; Inserm; INSA-Lyon; Radiology Department, Hospices Civils de Lyon, Lyon, France
6 Univ Lyon, INSA Lyon, Laboratoire Vibrations Acoustique, F-69621 Villeurbanne, France

Corresponding author: Yuemin Zhu (e-mail: yuemin.zhu@creatis.insa-lyon.fr).

This work is funded partially by the Program PHC-Cai Yuanpei 2018 (No. 41400TC), the International Research Project METISLAB, the Committee of Science and Technology of Shanghai (No. 19510711200), and European Union Horizon 2020 (No. 643694).

ABSTRACT Spectral photon-counting CT offers novel potentialities to achieve quantitative decomposition of material components, in comparison with traditional energy-integrating CT or dual-energy CT. Nonetheless, achieving accurate material decomposition, especially for low-concentration materials, is still extremely challenging for current sCT, due to restricted energy resolution stemming from the trade-off between the number of energy bins and undesired factors such as quantum noise. We propose to improve material decomposition by introducing the notion of super-energy-resolution in sCT. The super-energy-resolution material decomposition consists in learning the relationship between simulation and physical phantoms in image domain. To this end, a coupled dictionary learning method is utilized to learn such relationship in a pixel-wise way. The results on both physical phantoms and in vivo data showed that for the same decomposition method using lasso regularization, the proposed super-energy-resolution method achieves much higher decomposition accuracy and detection ability in contrast to traditional image-domain decomposition method using L1-norm regularization.

INDEX TERMS X-ray CT, Material decomposition, Photon-counting detector, Super energy resolution.

I. INTRODUCTION

SPECTRAL photon-counting X-ray computed tomography (sCT) provides additional X-ray spectrum information about material components in an organ or tissue, due to its ability of separating photons into different energy bins by using photon-counting detector (PCD). This ability makes possible efficient material decomposition of which the objective is to quantitatively measure different materials existed in a pixel.

It is a huge challenge to obtain high accuracy of material decomposition in clinical applications, especially for low-concentration materials [1], [2]. This is linked to the principle itself of sCT. With respect to conventional CT or dual-energy CT (DECT) to some extent, sCT divides the whole energy range into so-called energy bins. More there are energy bins, the higher energy resolution (narrower energy-bin width) and more spectral
information. However, more energy bins mean at the same time less photons in each energy bin and consequently higher quantum noise. Many aspects can influence the accuracy of material decomposition, including exposure time, energy bin width, acceptable flux rate, etc. Although there is complex interdependency between the parameters, we will focus on energy bin that is a basic parameter for spectral CT in contrast to conventional CT. In other words, energy resolution of sCT is limited by a trade-off between the total number of energy bins and the image quality at each bin, which restricts the accuracy of material decomposition and ability of detecting low-concentration materials.

Existing material decomposition methods can be divided into three categories: projection-domain [3]–[6], image-domain [7]–[12] and one-step material decomposition [13]–[15], according to the sequence of material decomposition and image reconstruction, i.e. before, after or during. Most decomposition methods of all three ways try to improve performance of material decomposition by exploiting the high correlation between spectral and spatial features embedded in the multi-energy images. However, almost no work can be found in the literature to enhance energy resolution of sCT by software. Actually, for DECT, there is an imaging method that aims to enhance the visibility of contrast agent by reconstructing the image at one chosen computed energy, called virtual monochromatic image (mono-E or VMS) [16]. The performance of mono-E from DECT is determined by the material decomposition that is processed only on actual images, which is noisy and depends on the chosen computed energy.

In this paper, we propose to introduce the notion of super-energy-resolution (SER) into sCT in order to improve the accuracy of material decomposition in image domain. The idea is to learn the relationship between spectral features of simulated data and those of physical data, then applying it to other physical data. To do it, theoretical sCT images containing abundant spectral energy information, designated as SER images, are simulated according to physical phantoms. The mapping between actual sCT images and SER images are learned by utilizing a coupled dictionary learning (CDL) method in a pixel-wise way. SER images synthesized based on the learned mapping are then used for material decomposition in image domain.

II. MODEL AND METHOD

Common models of image-domain material decomposition are firstly shown in this section. Then, SER for sCT is introduced and modeled. At last, the proposed SER-based material decomposition is described.

A. MODELS OF IMAGE-DOMAIN MATERIAL DECOMPOSITION

Tomography reconstruction, which is a preliminary step for image-domain material decomposition methods, is separately processed at each individual bin. Reconstruction at the i-th energy bin tries to achieve the spatial image, i.e. the linear attenuation coefficients (LAC) μ(\vec{x}, i) at pixel \( \vec{x} \):

\[
\mu(\vec{x}, i) = \arg\min_\mu \mathcal{D}(\mathcal{C}(s_i(u)), \mu(\vec{x}, i)) + R, \tag{1}
\]

where \( s_i(u) \) represents the recorded photons along the u-th ray in the i-th energy bin, \( \mathcal{C} \) the mapping function between \( s_i(u) \) and LAC, \( \mathcal{D} \) the data-fidelity term, and \( R \) the regularization. The detailed relationship between \( \mu \) and \( s \), i.e. \( \mathcal{C} \) in equation (1), is described by:

\[
\tilde{s}_i(u) = \int_{\mathbb{R}} n_0(E) d\tilde{\ell}(E) e^{-\tilde{\ell}(\vec{x}, E)} \mu(\vec{x}, E) d\ell, \tag{2}
\]

where \( \tilde{s}_i(u) \) denotes the mean value of \( s_i(u) \), \( n_0(E) \) the incident photon fluence from X-ray source, \( L(u) \) the u-th ray, \( \mu(\vec{x}, E) \) the LAC at energy \( E \), and \( d\tilde{\ell}(E) \) the detector absorption efficiency or detector response function that can give the probability of a photon of energy \( E \) to be measured within bin \( i \) [3]. For X-ray CT images, noise level is negatively related to the square root of the total number of detected photons. In other words, sCT suffers from stronger quantum noise for narrower bin.

Material decomposition is then performed on the reconstructed multi-energy images:

\[
\mu(\vec{x}, i) = \sum_{\alpha=1}^{M} \mu_{\text{ma}}(i) \rho_{\alpha}(\vec{x}), \quad i = 1, \ldots, B, \tag{3}
\]

where subscript \( \alpha \) denotes the \( \alpha \)-th basis material, \( \mu_{\text{ma}}(i) \) the calculated i-th energy-bin effective mass attenuation coefficient (MAC), \( B \) the number of energy bins, \( M \) the number of basis materials, and \( \rho_{\alpha}(\vec{x}) \) the mass density at pixel \( \vec{x} \). We rewrite image-domain material decomposition into its matrix form as:

\[
Y = MX + N, \tag{4}
\]

where \( Y \in \mathbb{R}^{B \times N_p} \) designates the reconstructed multi-energy spatial images of \( \mu \) with \( N_p \) designating the number of pixels or voxels, \( M \in \mathbb{R}^{B \times M} \) the decomposition matrix of \( \mu_m \), \( X \in \mathbb{R}^{M \times N_p} \) the decomposed basis material images of mass densities \( \rho \) and \( N \) the noise. Consequently the aforementioned trade-off between the number of energy bins and noise level leads to limited energy resolution in sCT. In practice, the number of energy bins of sCT is generally scheduled by manufacturers, e.g. 5 bins for Phillips sCT prototype [17].
B. SUPER-ENERGY-RESOLUTION IMAGING IN SCT

We propose to improve the energy resolution by means of fully exploiting both simulation data and physical data. To do it, the notion of SER is introduced in sCT. SER images are virtual multi-energy images with much more bins but noise-reduced compared to the actual sCT images. Fig. 1 illustrates the relationship between actual sCT, denoted by low-energy-resolution (LER) imaging, and SER imaging by respectively plotting the multi-energy values of corresponding pixels in the LER and SER images. For one pixel in the reconstructed LER images (e.g. yellow asterisks in the left of Fig. 1, designated as \( y_L \in \mathcal{R}^{B_L} \)), it has 5 values for 5 energy bins, i.e. \( B_L = 5 \). While for the corresponding SER pixel (yellow points in the right of Fig. 1, designated as \( y_S \in \mathcal{R}^{B_S} \)), it can have much more values, e.g. 50 values for \( B_S = 50 \) (30 - 80 keV). Note that the choice of 50 bins corresponds to the most precise \( \mu_m \) we can get from NIST [18]. That is, \( \mu \) is sampled every 1 keV.

![Figure 1. Linear attenuation coefficients: LER (left) and SER (right). Curves with different colors denote different pixels; corresponding LER and SER pixels have the same color.](image)

The proposed SER imaging is achieved by introducing the assumption that if containing the same basis materials and concentrations, the mapping between LER and SER images is determined and can be learned. We realize it by learning the relationship between physical data and simulation data and then applying it to other physical data, as illustrated in Fig. 2. The physical data are measured and reconstructed as LER images, while the corresponding SER versions are produced by simulation with same shapes and components but more bins. Then, LER and SER spectral features are extracted and their relationship will be learned by a mapping which can be utilized to synthesize SER images for other in vivo data.

C. SER IMAGE SYNTHESIS BASED ON COUPLED DICTIONARY LEARNING

We learn the mapping between LER and SER spectral features by utilizing CDL method in a pixel-wise way.

1) Pixel-wise training strategy

It is inefficient to directly learn the mapping from the whole LER and SER images or patches. In general, learning morphological features of certain organ or more is another challenge that demands abundant images with various spatial information for training. It will be expensive to produce these kinds of physical data currently. Meanwhile, complex material components can make the learning more difficult.

We solve it by training the mapping from LER and SER pixel pairs, rather than the whole images or patches. With this strategy, the training samples only contain spectral information. Correspondingly, the physical phantoms used in our work can have simple shapes, but abundant material components.

Meanwhile, to cope with the difficulty of extracting features from noisy LER pixels, we enhance the learning by adding conventional CT images. The conventional CT images of physical phantoms are measured on the same sCT system but reconstructed with full spectrum (i.e. a single energy reconstruction), thus having lower noise level compared to with single energy bin. The conventional CT and sCT pixels are then concatenated together as one LER sample for training. As a result, now \( y_L \in \mathcal{R}^{(B_L+1)} \) is a combined LER pixel.

2) Learn the mapping by CDL

CDL was used in the field of super spatial resolution [19], [20]. In the theory of dictionary learning, features of a pixel \( y \) can be represented as a sparse combination of atoms inside an over-complete dictionary \( D \). That is, \( y = Da \), where \( a \) is the sparse codes with very few nonzero entries:

\[
\text{argmin}_{a} \|a\|_0 \quad \text{s.t.} \quad \|y - Da\|_2^2 \leq \epsilon, \tag{5}
\]

where \( \|a\|_0 \) designates \( L_0 \) norm of \( a \). Generally, solving optimization problem in (5) is NP-hard, and it can be instead solved using \( L_1 \) norm, as:

\[
\text{argmin}_{a} \|y - Da\|_2^2 + \lambda \|a\|_1, \tag{6}
\]

where \( \lambda \) denotes the Lagrange multiplier.

LER and SER pixel sets can then be separately represented using a LER dictionary \( D_L \) and a SER dictionary \( D_S \):

\[
\begin{align*}
&\text{argmin}_{\Lambda_L} \|Y_L - D_L \Lambda_L\|_2^2 + \lambda_L \|\Lambda_L\|_1, \\
&\text{argmin}_{\Lambda_S} \|Y_S - D_S \Lambda_S\|_2^2 + \lambda_S \|\Lambda_S\|_1. \tag{7}
\end{align*}
\]

where \( Y_L \in \mathcal{R}^{(B_L+1)\times N_L} \) and \( Y_S \in \mathcal{R}^{B_S\times N_L} \) respectively designate the LER and SER pixel sets with \( N_L \) pixels in the training set, \( D_L \in \mathcal{R}^{(B_L+1)\times N_D} \) and \( D_S \in \mathcal{R}^{B_S\times N_D} \) the LER and SER dictionaries containing \( N_D \) atoms, \( \Lambda_L \) and \( \Lambda_S \) the corresponding sparse codes (i.e. sets of \( a \) in equation (5) and (6)), and \( \|Y - DA\|_2^2 \) the data fidelity.
term determined by Frobenius-Norm \( \| \cdot \|_F \). Then, the connection between LER and SER features is established by jointly training \( D_L \) and \( D_S \), formulated as:

\[
\begin{align*}
\argmin_{\Lambda, D} & \quad \| \hat{Y} - \bar{D}\Lambda \|_F^2 + \lambda \| \bar{\Lambda} \|_1, \\
s.t. & \quad \hat{Y} = [ Y_S \quad Y_L ], \bar{D} = [ D_S \quad D_L ], \bar{\Lambda} = [ \Lambda_S \quad \Lambda_L ].
\end{align*}
\] (8)

Meanwhile, a transform matrix is utilized to decrease the impacts of dramatic gap between simulation and physical data:

\[
\Lambda_S = W\Lambda_L,
\] (9)

where \( W \) denotes the transform matrix that was firstly introduced in super spatial resolution for stable cross-style image synthesis [20]. The enhanced training model then becomes:

\[
\begin{align*}
\argmin_{D, \alpha, W} & \quad \| Y_S - D_S\alpha_S \|_F^2 + \| Y_L - D_L\alpha_L \|_F^2 \\
& + \gamma \| \alpha_S - W\alpha_L \|_F^2 + \lambda_W \| W \|_F^2 + \lambda_S \| \alpha_S \|_1 + \lambda_L \| \alpha_L \|_1.
\end{align*}
\] (10)

We solve the above model by alternating-direction-method-of-multipliers (ADMM), which alternately updates coupled dictionaries \( \bar{D} \), sparse codes \( \bar{\alpha} \) and transform matrix \( W \).

It should be noted that atoms in the dictionaries learned using a pixel-wise strategy are energy-dependent, in contrary to spatial features in traditional super-spatial-resolution applications. Figure 3 illustrates the difference between super spatial resolution and super energy resolution by showing the learned features in practice. Atoms learned from a spatial image are morphological features, while for SER, atoms in the dictionary are energy-dependent and look more like \( \mu \) curves.

After training, we can readily synthesize SER pixel \( \hat{y}_S \) from its LER version \( y_L \) with the learned coupled dictionaries and transform matrix, by solving the following optimization:

\[
\begin{align*}
\hat{y}_S(y_L \rightarrow \hat{y}_S) : \\
\argmin_{\hat{y}_S} & \quad \| \hat{y}_S - D_S\alpha_S \|_F^2 + \| y_L - D_L\alpha_L \|_F^2 \\
& + \gamma \| \alpha_S - W\alpha_L \|_F^2 + \lambda_S \| \alpha_S \|_1 + \lambda_L \| \alpha_L \|_1.
\end{align*}
\] (11)

D. MATERIAL DECOMPOSITION USING SER IMAGES

The synthesized SER images are then utilized for material decomposition in image domain.

For existing sCT, decomposition matrix \( M \) is generally initialized by the aforementioned \( \mu_m \) [7] according to:

\[
\mu_{ma}(i) = \frac{\int_{E_i} n_0(E) d_i(E) \mu_{ma}(E) dE}{\int_{E_i} n_0(E) d_i(E) dE}, \quad i = 1, \ldots, B,
\] (12)

where \( \mu_{ma}(E) \) designates the theoretical MAC value and \( E_i \) the width of the \( i \)-th bin. Note that \( \mu_m(i) \) is an estimate of the true value of \( \mu_{ma}(i) \) corresponding to the \( i \)-th bin. As a result, thinner bins can produce more precise \( \mu_m(i) \), i.e. more precise \( M \). A more precise decomposition matrix suitable for the SER images is then needed. For LER, \( \mu_m \) of each basis material (one
column of $M$ in equation (4) designated as $M_L$) is an effective value at $B_L$ energy bins. In contrast, $M$ for SER (designated as $M_S$) has much more energy bins, e.g. $B_S = 50$. Both $M_L$ and $M_S$ have the same material components, e.g. iodine (I), gadolinium (Gd) and water in our experiment. Each column of $M_S$ is a more precise value directly retrieved from NIST, as illustrated in Fig. 4.

To assess the decomposition of the synthesized SER images, we choose a common image-domain decomposition algorithm using $L_1$ norm (lasso) regularization. Mathematically, we formulate the SER-based material decomposition using lasso as:

$$\hat{x} = \arg\min_x \|y_S - M_Sx\|_F^2 + \lambda \|x\|_1. \quad (13)$$

III. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETTING AND EVALUATION METRICS

The experiments were carried on real data from a Philips sCT prototype [21]–[23]. This medical approved prototype has a 500mm FOV and a pixel pitch of 0.274mm at iso-center. The photon-counting detector has a 1848x64 grid. However we used a 2x8 binning mode to obtain 924x8 pixels with a size of 1x2mm. The experiments were performed at tube voltage 120 kVp and current 220 mA under rotation time of 1s with a small focal spot. The volumetric CT dose index (CTDIvol) was 22.2 mGy. The system acquires 2400 projections per rotation. It has 5 energy thresholds set to 30, 51, 62, 72, 81 keV. These values have been selected to optimize the Gadolinium k-edge imaging and are the same for all the physical data in the present work. This configuration leads to 5 energy bins: #1[30-51keV], #2[51-62keV], #3[62-72keV], #4[72-81keV] and #5[81-120keV].

Reconstructed images from one single bin are shown in Fig. 5a. Fig. 5b depicts the cylindrical inserts of the physical phantom. The upper image in Fig. 5a is for training, while the lower two regions-of-interests (ROIs) of phantoms (i.e. all the areas inside red contours) are for training and testing, respectively. The reconstructed images for training and testing have different noise levels due to the presence of extensional fat rings added to the testing phantom exterior. Removable tubes containing different dilutions of iodine or gadolinium are inserted in the ROIs, as illustrated in Fig. 5b. The number refers to the concentration of each tube which covers from 0.1 mg/cc to 15 mg/cc. In our experiments, both iodine and gadolinium were measured for training and testing. Meanwhile, we use another in vivo dataset to further evaluate the proposed method. In our experiments, a rabbit was injected by gadolinium dilution, the image of which 5 and 15 minutes after the injection are respectively given in Fig. 6. For quantitative evaluation, tubes with known concentrations of gadolinium (2, 5, 10 mg/cc) are also placed beside the rabbit. The image of 5 min after injection is a little blurred due to movement.

All the spatial images in physical datasets were separately reconstructed at each energy bin via a common reconstruction method: conjugate gradient method (with Reconstruction Toolkit (RTK) [24]). Note that the decomposition method of Philips prototype is projection-domain based, while our proposed SER-based method is performed in image domain. What’s more, our reconstruction software (RTK) is also different from that used by Philips. In other words, we apply our own algorithms on prepped photon count
data from the scanner, whereas other correction steps from the Philips image chain (e.g. proper handling of bad detectors) are not used. Hence the results will be different from the Philips’ reconstruction results. The detector bin response function $d_i(E)$ in model (2) was provided by Philips. All phantoms were reconstructed into 9 slices and the size of each slice is 900*900. We learned the coupled dictionaries composed of 1024 ($N_D$) atoms from 400K ($N_I$) random LER-SER pixel pairs. On a machine equipped with Nvidia Tesla V100 and Intel I9 CPU, our method took 18h for training and 5 minutes for testing of one slice.

Both SER images synthesis and material decomposition were quantitatively evaluated. Normalized Root Mean Square Error (NRMSE) (i.e. $\text{NRMSE} = \text{RMSE}/\bar{x}$ for the image $x$) and peak signal-to-noise ratio (PSNR) were utilized to assess the quality of synthesized SER images. Note that NRMSE and PSNR were respectively calculated on the whole images of iodine and gadolinium, in order to assess the total synthesis performance of all concentrations. The smaller the NRMSE and the higher the PSNR, the more accurate the synthesis.

Mean and standard error of the mean (SEM) were calculated for each dilution to assess decomposition accuracy. Meanwhile, linear regression between the nominal (y-axial) and measured concentrations (using the calculated mean values, x-axial) was performed for all dilutions, as: $y = ax + b$, where $a$ and $b$ represent the slope and y-intercept, respectively. To further assess the detection ability of contrast agents, we also calculated the limit of detection (LOD) for both iodine and gadolinium. The LOD was described by the Clinical and Laboratory Standards Institute [25] as:

$$\text{LOB} = \bar{x}_{\text{blank}} + 1.645\delta_{\text{blank}},$$
$$\text{LOD} = \frac{\text{LOB} - b}{a},$$

where $\bar{x}_{\text{blank}}$ denotes mean water signal detected in the images of decomposed contrast agents, $\delta_{\text{blank}}$ the standard deviation of the measurement, and $\text{LOB}$ the limit of blank. LOB denotes the highest apparent contrast agent concentration expected to be found when replicates of a sample contain no contrast agent, which was calculated using tubes of water in each phantom. LOD is taken as the lowest concentration of a contrast agent in a sample that can be detected.

### B. TRAINING

NRMSE and PSNR (dB) of synthesized SER images at each energy bin are illustrated in Fig. 7. SER images of iodine have relatively higher accuracy than gadolinium, e.g. smaller NRMSE and higher PSNR, although both of them have good synthesis performance. Note that low-energy bins have relatively larger NRMSE and smaller PSNR, especially for bins of k-edge (i.e. energy bins containing k-edge). Also, we need to recover more details at the first two bins: 30 - 50 keV and 51 - 61 keV.

**FIGURE 7.** The (a) NRMSE and (b) PSNR of synthesized SER images for I and Gd.

The material decomposition results based on synthesized SER images are shown in Fig. 8. Visually, the proposed SER-based decomposition presents good detection ability according to the edge-preserving performance in training. It can substantially preserve edge information even for low-concentration iodine and gadolinium e.g. 0.5 mg/cc, as illustrated in the ROIs. Note that all the ROIs are shown with grayscale images in order to highlight morphological information.

To quantitatively assess the decomposition, linear regressions of iodine and gadolinium for all the concentrations are illustrated in Fig. 9. Note that there is no additional scatter correction for all the results of linear regression in this work. Both iodine and gadolinium have clearly high accuracy with respect to theoretical concentrations. The correlations between the measured and prepared concentrations are strongly linear for all dilutions (all $R^2 \geq 0.95$). The slopes are close to 1 for both iodine ($a = 0.96$) and gadolinium ($a = 0.97$). The y-intercept values are very low, e.g. $b = -0.26$ for iodine and $-0.38$ for gadolinium. Additionally, the decomposition is very reliable because both iodine and gadolinium are always underestimated and their SEMs
FIGURE 8. The decomposition results of SER based method on physical phantoms in training. Note that ROIs are shown in grayscale to highlight morphological information. Labels in ROIs denote the dilution concentration (mg/cc).

FIGURE 9. The linear-regression results of iodine and gadolinium for SER-based material decomposition in training: linear regression model, R-squared ($R^2$) and standard error of the mean (SEM).

are small.

Table 1 lists LOD for both iodine and gadolinium, to further assess the detection limit. Both iodine and gadolinium can be detected for concentrations lower than 0.5 mg/cc in training. More specifically, the proposed SER method has better detection ability for iodine than for gadolinium.

TABLE 1. The LOD of different material decomposition methods on physical phantoms.

| Material | LER | SER |
|----------|-----|-----|
| Iodine   | 2.00| 0.34| 0.36|
| Gadolinium | 4.81| 0.45| 0.84|

C. TESTING

The LER-based material decomposition method using the same lasso regularization was compared to the proposed SER-based method. Both physical phantoms with more noise and artifacts and in vivo data were tested.

Fig. 11 shows the NRMSE and PSNR (dB) of synthesized SER images at each energy bin for testing. As in the case of training, SER images of iodine have relatively higher accuracy than gadolinium, e.g. smaller NRMSE and higher PSNR. Compared to training, although testing has relatively higher NRMSE and smaller RSNR, the difference is small.

The material decomposition results of physical phantoms with more noise and artifacts are shown in Fig. 10. In contrast to LER method, the proposed SER method shows better performance in quantification and detection compared to LER method. Firstly, the proposed SER method gives much better morphological accuracy in terms of edge-preserving performance and much less noise, even for low-concentration materials, e.g.
0.5 mg/cc for both iodine and gadolinium. The edges are substantially preserved by SER method. While for LER method, the selected areas of low-concentration materials are overwhelmed by noise and cannot be detected.

Secondly, more quantitatively as shown in Fig. 12, linear regressions of both I and Gd illustrate that SER-based method has much smaller errors. SER method presents more precise slope (for iodine: $a_{LER} = 0.71$ and $a_{SER} = 0.87$; for gadolinium: $a_{LER} = 1.24$ and $a_{SER} = 0.94$) and y-intercept value (for iodine: $b_{LER} = 1.71$ and $b_{SER} = -0.08$; for gadolinium: $b_{LER} = 0.92$ and $b_{SER} = -0.51$). Meanwhile, SER method is more reliable with smaller SEMs and almost all decomposed values always exhibit the same underestimation trend. As can be observed in Table 1 that lists LOD values of different methods for both iodine and gadolinium, our SER method has much stronger detection ability for both iodine and gadolinium compared to LER method by producing smaller LOD values in testing. It has better detection ability for iodine than for gadolinium, as in the case of training.

Meanwhile, the decomposition results of different methods on in vivo data are given in Fig. 13. We identify calcium (Ca) inside spine as the basis material iodine in terms of their similar MAC under our energy-thresholds arrangement. Clearly, LER method has less accurate separation between gadolinium and iodine and has more noise for the two basis materials. In contrast, the SER method has better detection ability for both iodine (in the spine) and gadolinium (in tubes and organs). Mean and SEM of gadolinium in the tubes are listed in Table 2. Obviously, SER method leads to more accurate decomposition for gadolinium of concentrations covering from 10 mg/cc to 2 mg/cc.

**IV. DISCUSSION**

Our main goal is to improve the energy resolution of sCT in order to improve the k-edge quantification accuracy and detection ability. To achieve that we introduced the notion of SER in sCT to achieve SER imaging, by fully exploiting abundant spectral features embed-
ded in both simulation and physical data. Preliminary results revealed that the proposed SER method enables accurate decomposition and high detection ability even for low-concentration materials.

The SER-based material decomposition method is reliable. Firstly, even for noisier conditions in training, the proposed SER method still enables high decomposition accuracy and detection ability. In our experiments, we trained the dictionaries on the phantoms with less noise and artifacts, but tested on the ones with stronger noise and more artifacts. However, the results still show better decomposition accuracy and detection ability of the proposed method compared to the LER method, especially for iodine. Secondly, accurate decomposition of in vivo data polluted by movement (5 min after the injection) also shows the reliability of our method. Moreover, the decomposition results of both iodine and gadolinium keep stable trend, although always underestimated for almost all concentrations.

In this work, we did not consider the mixtures of multiple elements in the experiments. Nevertheless, materials detection mainly relies on the subsequent decomposition method if the input is SER images. At the same time, the ability of the decomposition method with L1 norm to detect mixtures has been demonstrated in [26], [27]. Hence, our method would work for the mixtures of interested elements (I or Gd).

Concerning the partial volume effect, which arises when more than one material type occurs in a voxel (or pixel) having a limited spatial resolution, it is a very interesting and pertinent issue in sCT. In our method, material decomposition is processed pixel by pixel, which implies that only material category and density are accounted for and spatial neighborhood is not involved. As a result, partial volume effect would be minimized. Nevertheless, it would be interesting to study how a decomposed material quantitatively changes with different spatial resolutions, different types of mixtures, different material proportions in the voxel, etc.

The interest of projection-domain and one-step methods lies in the fact that they generate directly material images from projection data without needing to first reconstruct spatial images. In contrast, image-domain decomposition has its own advantages of being able to both exploit spatial images and generate material images. On the other hand, projection data is not always available or not practical to use in clinical practice because of commercial confidentiality or practical constraints. In short, image-domain methods are hardware-independent and thus provide much flexibility than the other 2 methods.

Each single synthesized SER image is a virtual monochromatic image (mono-E). The mono-E image form DECT is generally reconstructed from a pair of decomposed basis material images (e.g., a soft-tissue-like basis material and a contrast agent). The performance of mono-E is determined by the material decomposition that is processed only on actual images (LER images for sCT in the present study). In contrast, SER

| TABLE 2. The mean and standard error of the mean (SEM) of each tube in in vivo data (15 min after injection of gadolinium). |
| True Conc. (mg/cc) | LER mean±SEM | SER mean±SEM |
|-------------------|--------------|--------------|
| 10                | 9.2 ± 0.05   | 10.0 ± 0.05  |
| 5                 | 3.3 ± 0.01   | 4.4 ± 0.01   |
| 2                 | 0.6 ± 0.01   | 1.9 ± 0.01   |

FIGURE 13. The decomposition results of different methods on in vivo data in testing: (a) 5 min after the injection of contrast agents; (b) 15 min after. Note that calcium in spine is treated as iodine, and the basis material gadolinium exists in tubes and organs.
images are directly synthesized using spectral features learned from both simulation and physical data. As a result, our method could contribute to more accurate mono-E imaging.

V. CONCLUSION

We have proposed a SER imaging method to improve image-domain material decomposition in sCT. The method is based on generating SER images from LER images. This is achieved through learning mapping between simulation and physical data in a pixel-wise way and applying it to other physical data. The results on both physical phantoms and in vivo data showed that the SER-based material decomposition method shows clearly higher accurate material decomposition and detection ability compared to the LER-based method. In the future work, the performance of other image-domain decomposition methods using SER images will be investigated. It would be also interesting to compare image-domain decomposition with projection-domain methods such as that used by Philips sCT prototype.

ACKNOWLEDGMENT

We thank Dr. Klaus Erhard and Dr. Yoad Yagil for the careful reading of the manuscript and many helpful comments.

REFERENCES

[1] M. C. Jacobsen et al., “Dual-Energy CT: Lower Limits of Iodine Detection and Quantification,” in Radiology, vol. 292, no. 2, pp. 414–419, 2019.

[2] Xie, Bingqing, et al. “Image-domain material decomposition in spectral photon-counting CT for medical applications (Doctoral dissertation, Université de Lyon) 2020.

[3] J. P. Schomika et al., “Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography,” in Physics in Medicine and Biology, vol. 53, no. 15, pp. 4031–4047, 2008.

[4] N. Ducros, J. F. P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin, “Regularization of nonlinear decomposition of spectral x-ray projection images,” in Medical Physics, vol. 44, no. 9, pp. e174–e187, Sep. 2017.

[5] T. Su, V. Kafandeljian, P. Duvachele, and Y. Zhu, “A spectral X-ray CT simulation study for quantitative determination of iron,” in Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 594, 39–46, 2018.

[6] Y. Zhang, X. Mou, G. Wang, and H. Yu, “Tensor-Based Dictionary Learning for Spectral CT Reconstruction,” in IEEE Transactions on Medical Imaging, vol. 36, no. 1, pp. 142–154, Jan. 2017.

[7] Q. Le Huy and S. Molloi, “Least squares parameter estimation methods for material decomposition with energy discriminating detectors,” in Medical Physics, vol. 38, no. 1, pp. 245–255, Jan. 2011.

[8] C. J. Bateman et al., “MARS-MD rejection based image domain material decomposition,” Tech. Rep., 2018.

[9] W. Xia et al., “Spectral CT ReconstructionASSIST: Aided by Self-Similarity in Image-Spectral Tensors,” in IEEE Transactions on Computational Imaging, vol. 5, no. 3, pp. 420–436, Sep. 2019.

[10] Z. Chen and L. Li, “Robust multimaterial decomposition of spectral CT using convolutional neural networks,” in Optical Engineering, vol. 58, no. 01, 2019.

[11] W. Wu, P. Chen, V. V. Vardhanahalli, W. Wu, and H. Yu, “Improved Material Decomposition with a Two-Step Regularization for Spectral CT,” in IEEE Access, vol. 7, no. October, pp. 158770–158781, 2019.

[12] Y. Zhang, X. Mou, G. Wang and H. Yu, “Tensor-based dictionary learning for spectral CT reconstruction.” in IEEE transactions on medical imaging, 36.1, 142-154, 2016.

[13] R. Foygel Barber, E. Y. Sidky, T. Gilat Schmidt, and X. Pan, “An algorithm for constrained one-step inversion of spectral CT data,” in Physics in Medicine and Biology, vol. 61, no. 10, pp. 3734–3818, 2016.

[14] K. Mechlen et al., “Joint Statistical Iterative Material Image Reconstruction for Spectral Computed Tomography Using a Semi-Empirical Forward Model,” in IEEE Transactions on Medical Imaging, vol. 37, no. 1, pp. 68–80, Jan 2018.

[15] C. Mory, B. Sixou, S. Si-Mohamed, L. Bousel, and S. Rit, “Comparison of five one-step reconstruction algorithms for spectral CT,” in Physics in Medicine & Biology, vol. 63, no. 23, p. 235001, Nov. 2018.

[16] I. Yama, T. Nakaura, A. Iyama, K. Katahira, and Y. Yamashita, “The Usefulness of Dual-Layer Spectral Computed Tomography for Myelography: A Case Report and Review of the Literature,” in Case Reports in Orthopedics, vol. 2018, pp. 1–4, 2018.

[17] S. Si-Mohamed, D. Bar-Ness, M. Sigovan, T. Tatar-Leitman, D. P. Cormode, P. C. Naha, and P. Douek, “Multicolour imaging with spectral photon-counting CT: a phantom study,” in European Radiology Experimental, vol. 2, no. 1, p. 34, dec 2018.

[18] E. Saloman and J. Hubbell, “X-ray attenuation coefficients (total cross sections): Comparison of the experimental data base with the recommended values of Henke and the theoretical values of Scofield for energies between 0.1-100 keV,” Tech. Rep., Jul 1986.

[19] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” in IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2861–2873, 2010.

[20] S. Wang, D. Zhang, Y. Liang, and Q. Pan, “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2216–2223, 2012.

[21] N. R. van der Werf, S. Si-Mohamed, P. A. Rodesch, R. W. van Hamersvelt, M. J. W. Greuter, S. Boccalini, and P. Douek, “Coronary calcium scoring potential of large field-of-view spectral photon-counting CT: a phantom study,” European radiology, 1-11 2021.

[22] S. Si-Mohamed, S. Boccalini, P. A. Rodesch, R. Dessouky, F. Lahoud, T. Broussaud, and P. Douek, “Feasibility of lung imaging with a large field-of-view spectral photon-counting CT system. Diagnostic and Interventional Imaging, 102(5), 305-312, 2021.

[23] S. Si-Mohamed, J. Greffier, J. Mialhe, S. Boccalini, P. A. Rodesch, A. Vuillod, and P. Douek, “Comparison of image quality between spectral photon-counting CT and dual-layer CT for the evaluation of lung nodules: a phantom study.” European Radiology, 1-9, 2021.

[24] D. J. Metcalfe, “Direct reconstruction from projections,” in Medical Physics, vol. 19, no. 11, pp. 2861–2873, 2010.

[25] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, “The Reconstruction Toolkit (RTK),” in Medical Physics, vol. 40, no. 12, pp. 235001, Nov 2018.

[26] D. G. Delprete, “CLSI EP17-A: Proposed harmonized methodology for use of total prostate-specific antigen assays,” in Journal of Clinical Chemistry and Clinical Biology, vol. 58, no. 12, pp. 1441–1453, 2020.

[27] J. E. Plewa, “Spectral CT: A new imaging modality,” in Journal of Applied Clinical Medical Physics, vol. 21, no. 1, pp. 16–22, 2020.

[28] J. E. Plewa, “Spectral CT: A new imaging modality,” in Journal of Applied Clinical Medical Physics, vol. 21, no. 1, pp. 16–22, 2020.

[29] S. Si-Mohamed, J. Greffier, J. Mialhe, S. Boccalini, P. A. Rodesch, and S. Vuillod, “Comparison of image quality between spectral photon-counting CT and dual-layer CT for the evaluation of lung nodules: a phantom study.” European Radiology, 1-9, 2021.

[30] D. J. Metcalfe, “Direct reconstruction from projections,” in Medical Physics, vol. 19, no. 11, pp. 2861–2873, 2010.

[31] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, “The Reconstruction Toolkit (RTK),” in Medical Physics, vol. 40, no. 12, pp. 235001, Nov 2018.
BINGQING XIE received the B.S. degree in physics, from Shandong University in 2013 and the M.S. degree in microelectronics from University of Chinese Academy of Sciences. He received his Ph.D degree from CREATIS, INSA Lyon in 2020, under supervision of Prof. Yuemin Zhu. His research interests include medical image processing, especially spectral CT image processing, and deep learning.

LIHUI WANG was born in Jilin, China in 1982. She received the B.S. degree in electronic science and technology from Harbin Engineering University, Harbin, in 2006 and the M.S. degree in optics from the same university in 2008, and From 2013 to 2017, she was an Assistant Professor with the Computer Science and Technology Department, the Ph.D. degree in image and system from INSA Lyon, Lyon, France in 2013. Guizhou University, Guiyang. Since 2017 she has been associate professor in the Key Lab of Intelligent Medical Image Analysis and Precise Diagnosis, Guiyang, Guizhou. She is the author of more than 30 articles. Her research interests include medical imaging, image processing, pattern recognition, and machine learning.

YUEMIN ZHU is a permanent Research Director of the CNRS (Centre National de la Recherche Scientifique) of France. He received his B.Sc. degree in telecommunication in 1982 from Huazhong University of Science and Technology (HUST), China, M.Sc. degree in 1984 and Ph.D. in 1988 from the INSA, Lyon, France. He obtained the “Habilitation à Diriger des Recherches” in 1993, France. His current research interests include image representation, reconstruction, correction, denoising, registration, visualization, and fusion. And his current research projects focus on human cardiac diffusion tensor imaging and human cardiac polarize light imaging. He has so far published more than 260 research papers, including 105 journal articles, 2 book chapters, 12 invited conference keynotes, 147 conference proceedings, etc. Due to his research contribution, he has received twice the "French Prize of Scientifique Excellence" ("Prime d’Excellence Scientifique"), respectively in 2010 and 2014. He is currently the Editor of Journal of Medical Engineering, and invited reviewer of more than 20 international journals.

PHILIPPE DUVACHELLE is an associate professor in CREATIS, INSA Lyon. He is also an expert CND in INSA Lyon. His research field involves clinicians, basic science researchers; imaging companies and pharmaceutical companies.

PEI NIU received her Ph.D degree from CREATIS, INSA Lyon in 2020. Her research interests include medical image processing, deep learning and pattern recognition. She is now working in Huawei on computer vision, including neural network searching, and optical character recognition.

TIING SU received her Ph.D degree from LVA, INSA Lyon in 2018. Her research interests include medical image processing, signal processing and artificial intelligence. She is now an associate researcher at Research Center for Medical Artificial Intelligence, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

LOIC BOUSSEL is a professor in CREATIS, INSA Lyon. He also works for Radiology department, Hospices Civils de Lyon, France. His work focuses on medical imaging processing and its applications, especially on spectral CT.

PHILIPPE DOUEK is a professor in CREATIS, INSA Lyon. He also works for Radiology department, Hospices Civils de Lyon, France. He is Chairman of a consortium of 10 Departments of Radiology and Nuclear Medicine in Lyon Hospital (HCL), France. He has more than 15 years of imaging research, with many multidisciplinary projects in this field, as well as international collaboration (NIH, UCSF). He has strong experience in coordination of preclinical and clinical projects in imaging research involving clinicians, basic science researchers; imaging companies and pharmaceutical companies.

FENG YANG received her Ph.D degree from CREATIS, INSA Lyon. She is now an associate professor in Beijing Jiaotong University. Her research interests include image segmentation, registration, intelligent image processing, especially deep learning combined with image processing algorithms for disease prediction and classification. She is also now working with National Library of Medicine, National Institute of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA.