

# IMAGE FUSION OF X-RAY AND ELECTRON TOMOGRAMS

*Yan Guo and Bernd Rieger*

Department of Imaging Physics, Delft University of Technology  
E-mail: {y.guo-3, b.rieger}@tudelft.nl

## ABSTRACT

With electron tomography, we can reconstruct a three-dimensional (3D) volume of a specimen from a series of its two-dimensional (2D) projection images on the nanoscale. In a scanning transmission electron microscope (STEM), element-specific maps and mass-contrast projections can be simultaneously acquired from the X-ray spectrometer and electron detector. The X-ray tomogram has high chemical specificity but low signal-to-noise ratio (SNR), while the electron tomogram has poor compositional information but high SNR. In this paper, we adopt and modify a regression-based image fusion algorithm to combine these two complementary modalities, so that the fused version would maintain both high chemical specificity and high SNR. We demonstrate that our method improves reconstruction quality on an experimental dataset of a core-shell nanoparticle. Specifically, it delivers tomograms with sharper edges and smoother fore- and background, and hence can enable easier and more accurate 3D characterization of such nanostructures.

**Index Terms**— Image fusion, partial least squares regression, cross-modality modeling

## 1. INTRODUCTION

As nanotechnology advances, electron tomography is becoming popular in materials science because it can examine the 3D structure of a specimen on the nanoscale [1]. In a transmission electron microscope, the sample under study is exposed to an electron beam and tilted to obtain 2D projection images at different angles. In tomography, these projections are called tilt-series, from which we can reconstruct a 3D tomogram of the sample via various algorithms [1]. In fact, tomography is not only broadly investigated and applied in 3D imaging in electron microscopy but also in other fields, such as medical imaging where X-rays are used to image tissues [2]. So far, dozens of reconstruction algorithms have been proposed [2–5], and filtered backprojection (FBP) and simultaneous iterative reconstruction technique (SIRT) are the most

commonly adopted. Nevertheless, tomographic reconstruction remains an ill-posed inverse problem for two reasons: insufficient number of projections during data acquisition, and inevitable noise in the measurement such as shot noise [2]. Therefore, we have to select suitable algorithms for different datasets. For instance, in the presence of limited projections and noise, FBP aggravates thin streaks in the reconstruction, whereas SIRT does not. However, we still need to choose a proper number of iterations for SIRT (typically between 20 and 30) to balance the reconstruction quality and SNR [4].

In STEM, different imaging modalities can provide us complementary information about a specimen. For example, projections formed by a high-angle annular dark-field (HAADF) detector are at atomic resolution with high SNR but not element-specific [6]. Conversely, elemental maps generated by the energy dispersive X-ray spectrometer (EDS) are rich in chemical information but suffer from low SNR [7]. In order to integrate the strengths of these two complementary modalities into one reconstruction, Zhong *et al.* proposed HAADF-EDS bimodal tomography [5], which introduces a weighting factor  $\alpha$  to balance the residue terms of HAADF-STEM and EDS-STEM. The choice of  $\alpha$  depends on the noise level of the EDS data and influences the reconstruction quality [5]. However, we cannot determine the “best” value of  $\alpha$  in advance. It is currently chosen by comparing reconstructions computed over the whole range of  $\alpha \in (0, 1)$  to a hand-segmented ground truth. Since this is impractical in the semiconductor industry, algorithms that do not require a hand-segmentation are desirable.

In order to benefit from different reconstruction algorithms and the two aforementioned imaging modalities, one can employ image fusion. It fuses multiple input images into a single output, so that the composite version would be more comprehensive for human and/or machine perception than any source image alone. From low to high, fusion can occur at three different levels: pixel, feature, and decision level [8–10]. Compared to the latter two, pixel-level fusion that directly uses the available information from respective source images without extracting features nor constructing classifiers is the most broadly developed. State-of-the-art image fusion techniques are summarized in [11]. Multi-sensor image fusion is one of the active subfields where source images originate from different sensors, such as infrared and

This work is partially supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research, and partially by the Ministry of Economic Affairs, Agriculture and Innovation under project number 13314.

optical cameras [12], or computed tomography (CT) and magnetic resonance imaging (MRI) scanners [13, 14]. It has been extensively applied in diverse scenarios, including remote sensing [15], medical diagnosis [16] and surveillance [17, 18]. In [19], van de Plas *et al.* extended the concept of multi-sensor image fusion to study protein, peptide, lipid and drug distributions in tissues. With partial least squares regression, they built a cross-modality model to fuse data obtained from mass spectrometry and optical microscopy. Surprisingly, this concept has still not been widely applied to electron tomography.

In this paper, we aim to fuse X-ray and electron tomograms for reconstructing nanomaterials with neither a ground truth nor a tuning parameter. Specifically, we modify the regression-based cross-modality modeling of van de Plas [19] to fuse EDS and HAADF reconstructions of a core-shell nanoparticle consisting of gold and silver. Compared with HAADF-EDS bimodal tomography (HEBT) [5], our fusion algorithm enables reconstructions with sharper edges and smoother fore- and background, and achieves higher Pearson coefficients between the reconstruction and a hand-segmented ground truth both for Au and Ag. To begin with, we explain our three-step algorithm in Section 2, and compare it to FBP, SIRT and HEBT qualitatively and quantitatively in Section 3. In Section 4, we summarize our work and discuss possible future extensions.

## 2. METHOD

Our specimen is a core-shell nanoparticle that contains gold (Au) in the inner shell and silver (Ag) in the outer [5]. It is projected at 31 tilt angles, ranging from  $-75^\circ$  to  $+75^\circ$  with an increment of  $5^\circ$  between the consecutive projections. At each angle, one HAADF projection and one X-ray spectrum image are recorded. Then, the latter is deconvolved into two element-specific channels, one for Au and the other Ag. This provides us three tilt-series datasets at hand.

We use six different reconstruction algorithms for fusion: filtered backprojection (FBP), simultaneous iterative reconstruction technique (SIRT), simultaneous algebraic reconstruction technique (SART) [2], conjugate gradient least squares (CGLS) algorithm, maximum likelihood reconstruction for emission tomography (EM) [3], and total variation minimization (TV-min) reconstruction technique [4]. As a result, we have six image volumes reconstructed by the six aforementioned algorithms for each tilt-series dataset. Throughout this paper, we use ASTRA toolbox, which is an open source platform under the GPLv3 license for 3D image reconstruction in tomography [20].

In this section, we take the HAADF, EDS-Au and EDS-Ag reconstructions to fuse a more accurate image for Au. Fusing Ag follows the same principle. Note that we are only working in 2D at present but will extend the algorithm to 3D in future. As depicted in Fig. 1, building a fusion model

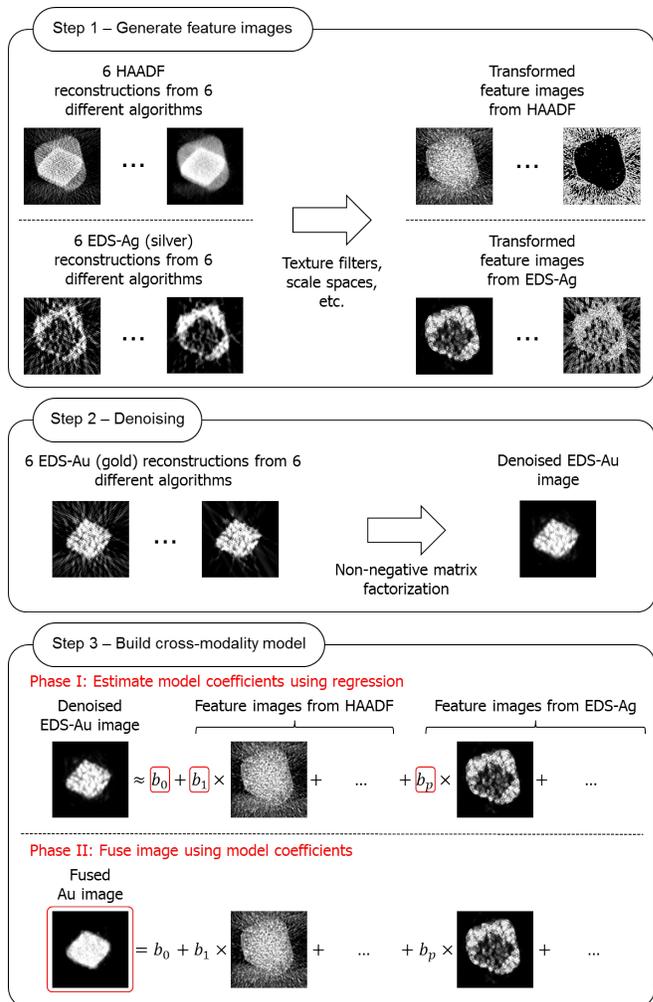


Fig. 1: Three-step fusion algorithm. Details in Section 2.

for Au has three steps (note the difference between Au and Ag at different positions): for each  $z$ -slice, (i) transform the  $xy$ -slice of HAADF (or EDS-Ag) reconstruction into many feature images; (ii) denoise EDS-Au image via non-negative matrix factorization; (iii) build a cross-modality model between the feature images and denoised EDS-Au image, and apply it for fusion.

### 2.1. Generate feature images

To build the cross-modality model, we need to collect enough relevant information by generating more feature images from the original twelve HAADF and EDS-Ag reconstructions (six HAADF and six EDS-Ag, from six algorithms). At this step, each  $xy$ -slice is processed by several texture filters, such as an entropy filter [21]. Table 1 lists all the filters and their parameters, with which we expand the number of feature images from 12 to 72. One can also extend the filtering operation to various scale spaces. We choose a two-level Gaussian scale

space, and hence have 144 feature images in total.

**Table 1:** Filters and related parameters in step (i)

Name	Parameter
Local entropy	$3 \times 3$ neighborhood
Local range	$3 \times 3$ neighborhood
Local standard deviation	$3 \times 3$ neighborhood
Local variance	$3 \times 3$ neighborhood
Gaussian gradient magnitude	Standard deviation $\sigma = 1$

## 2.2. Denoising

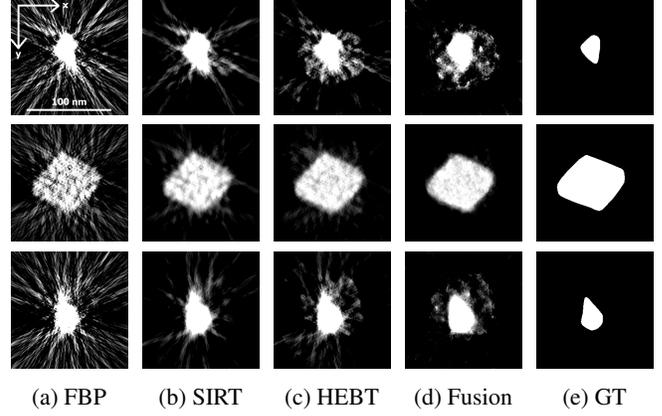
For each value of  $z$ , we have six EDS-Au  $xy$ -slices. Although they are computed using six different reconstruction algorithms, they share common patterns. To extract the most dominant structure among these slices and reduce their pixel-specific variations, we perform non-negative matrix factorization (NMF) over all the slices to obtain a denoised EDS-Au image [22]. Compared with principal component analysis (PCA), NMF only allows additive, not subtractive, combinations due to its non-negativity constraints [22]. Since all intensity values in the denoised image are inherently non-negative, we choose NMF rather than PCA.

## 2.3. Build the cross-modality model

Step (iii) contains two phases. First, we adopt partial least squares (PLS) regression [19] to build a cross-modality model between the feature images generated in step (i) and the denoised EDS-Au image in step (ii). Then, we apply it to fuse the final Au image. Similar to the ordinary least squares (LS) regression, PLS regression also constructs a linear model

$$\mathbf{y} = b_0 + b_1\mathbf{x}_1 + \cdots + b_P\mathbf{x}_P + \varepsilon \quad (1)$$

to specify the (linear) relationship between a set of predictor variables  $\mathbf{x}_p$ ,  $p = 1, \dots, P$  and one response variable  $\mathbf{y}$  [23]. In Eq. (1),  $b_0$  is the intercept,  $b_p$ ,  $p = 1, \dots, P$  are slopes, and  $\varepsilon$  is an error term. However, different from LS that directly establishes a linear regression model in the original data space, PLS first performs PCA to project both the predictor and response variables to  $N_{\text{comp}}$  components in another space [24]. As a result, even if the correlation among predictor variables is high, it can still guarantee stable results with low variability whereas LS cannot. For our purpose, we take each (vectorized) feature image as one predictor variable  $\mathbf{x}_p$  and the denoised EDS-Au image as  $\mathbf{y}$ , and solve this regression problem by `plsregress()` in MATLAB. Because the variance explained in response variable  $\mathbf{y}$  increases with the number of PLS components  $N_{\text{comp}}$ , we set  $N_{\text{comp}}$  to its maximum value  $N_{\text{comp}} = P$  [19]. Once we find all coefficients  $b_p$ ,  $p = 0, \dots, P$ , we fuse the Au image as



**Fig. 2:** Au (gold)  $xy$ -slices of (a) FBP, (b) SIRT with 30 iterations, (c) HEBT with 100 iterations and the associated optimal weighting factor, and (d) fusion compared to (e) hand-segmented ground truth (GT) at  $z = 80$  (top), 150 (middle) and 220 (bottom). The size of reconstruction volume is  $300 \times 300 \times 300$ . For better visualization, we perform percentile contrast stretching from 0 to 95%.

$b_0 + b_1\mathbf{x}_1 + \cdots + b_P\mathbf{x}_P$ , which is guaranteed to be the closest to the denoised EDS-Au image.

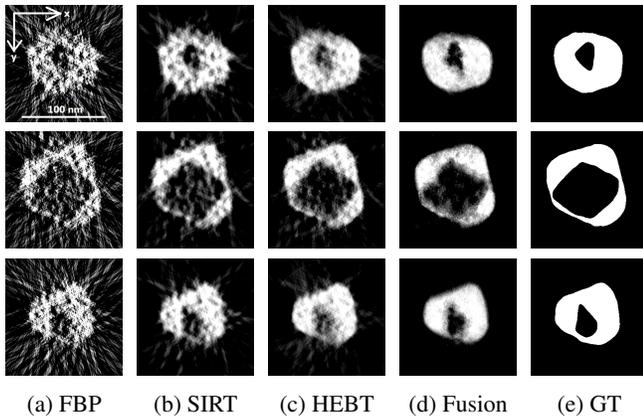
## 3. RESULTS

In this section, we compare our fusion algorithm to the most popular algorithms FBP and SIRT, and the newly proposed HAADF-EDS bimodal tomography (HEBT) [5]. FBP and SIRT are performed on EDS-STEM data to achieve an element-specific reconstruction. For HEBT, we set the number of iterations to 100, and choose the weighting factor  $\alpha$  by computing reconstructions for the whole range of  $\alpha \in (0, 1)$  and comparing them to a hand-segmented ground truth with Pearson coefficient [25]. This coefficient measures the similarity between two images  $f_1$  and  $f_2$  as

$$\text{PC} = \frac{\sum_i (f_{1,i} - \bar{f}_1)(f_{2,i} - \bar{f}_2)}{\sqrt{\sum_i (f_{1,i} - \bar{f}_1)^2 \sum_i (f_{2,i} - \bar{f}_2)^2}} \quad (2)$$

in which  $f_{1,i}$  and  $f_{2,i}$  are the intensity values of  $i$ -th pixel,  $\bar{f}_1$  and  $\bar{f}_2$  the average intensities over all pixels in  $f_1$  and  $f_2$ , respectively. Although HEBT can take one  $\alpha$  value as the input and simultaneously deliver Au and Ag reconstructions as the output, we separately determine the optimal  $\alpha$  for Au and Ag. The size of reconstruction volume is  $300 \times 300 \times 300$ . Note that we only consider  $xy$ -slices where  $z \in [80, 220]$ , because they are the least influenced by boundary artifacts.

Fig. 2 and Fig. 3 depict  $xy$ -slices of Au and Ag reconstructions at  $z = 80, 150$  and  $220$ , which are generated by FBP, SIRT, HEBT and our fusion algorithm, respectively. Binary images in the last column are the hand-segmented



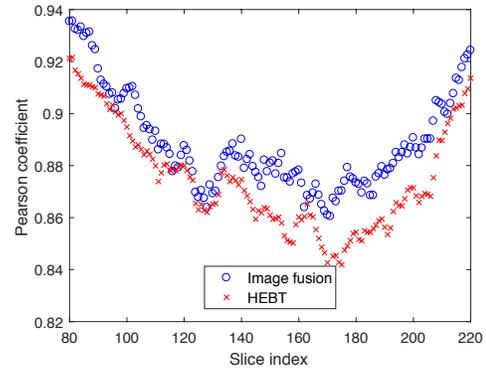
**Fig. 3:** Ag (silver)  $xy$ -slices of (a) FBP, (b) SIRT with 30 iterations, (c) HEBT with 100 iterations and the associated optimal weighting factor, and (d) fusion compared to (e) hand-segmented ground truth (GT) at  $z = 80$  (top), 150 (middle) and 220 (bottom). For better visualization, we perform percentile contrast stretching from 0 to 95%.

ground truth with homogeneous intensity. Since the number of projections in our EDS tilt-series datasets is low (only 31 in total), images reconstructed by FBP suffer from severe star-shaped thin streaks in the background. In this case, SIRT effectively suppresses such artifacts by setting up algebraic equations for unknown reconstructions in terms of the projection data. Compared with FBP and SIRT, tomograms delivered from HEBT show the smoothest foreground and the most continuous boundaries (see Fig. 3(b) and 3(c)). However, even with the optimal weighting factor, cross-element contamination where Au leaks into the background of Ag reconstruction (and vice versa) still exists. In comparison to HEBT, our fusion algorithm not only achieves a better separation between Au and Ag with sharper and clearer edges, but also more homogeneous fore- and background. Such improvement would make characterization of core-shell nanomaterials, like 3D metrology and surface roughness measurement, easier and more accurate. Note that neither a ground truth nor a tuning parameter is required in our fusion process.

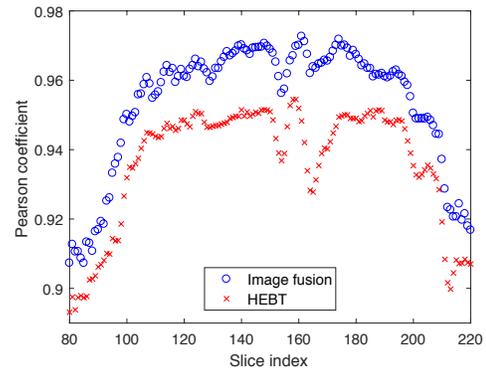
We also quantitatively compare our algorithm to HEBT for all slices between 80 and 220 by measuring the similarity between the reconstruction image and ground truth using the Pearson coefficient defined in Eq. (2). It can be seen in Fig. 4 that our algorithm outperforms HEBT, as it achieves higher Pearson coefficients both for Au and Ag reconstructions.

#### 4. CONCLUSION AND FUTURE WORK

In this paper, we adopt and modify a regression-based image fusion algorithm to build a cross-modality model between the EDS and HAADF reconstructions of a core-shell nanoparti-



(a) Au (gold)



(b) Ag (silver)

**Fig. 4:** Comparison of Pearson coefficients versus slice index for (a) Au and (b) Ag. Reconstructions generated by fusion algorithm, and by HEBT with 100 iterations and its optimal weighting factor.

cle consisting of Au and Ag. Compared with the conventional FBP and SIRT, and the more advanced bimodal tomography, our algorithm guarantees output with sharper edges and smoother fore- and background. As for the future work, we consider testing the robustness of our fusion algorithm for core-shell nanomaterials whose atomic number of the inner shell is close to the outer. Moreover, we will also extend the current framework to 3D (filtering and denoising) and/or to characterize more complex nanostructures such as multi-elemental semiconductor devices.

#### 5. REFERENCES

- [1] P. A. Midgley and M. Weyland, “3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography,” *Ultramicroscopy*, vol. 96, pp. 413–431, 2003.
- [2] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.

- [3] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. M1-1, pp. 113–121, 1982.
- [4] B. Goris et al., "Electron tomography based on a total variation minimization reconstruction technique," *Ultramicroscopy*, vol. 113, pp. 120–130, 2012.
- [5] Z. Zhong et al., "A bimodal tomographic reconstruction technique combining EDS-STEM and HAADF-STEM," *Ultramicroscopy*, vol. 174, pp. 35–45, 2017.
- [6] C. Kübel et al., "Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications," *Microscopy and Microanalysis*, vol. 11, pp. 378–400, 2005.
- [7] T. J. A. Slater et al., "STEM-EDX tomography of bimetallic nanoparticles: a methodological investigation," *Ultramicroscopy*, vol. 162, pp. 61–73, 2016.
- [8] S. Li et al., "Pixel-level image fusion: a survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, 2017.
- [9] F. A. Al-Wassai, N. V. Kalyankar, and A. A. Al-Zaky, "Multisensor images fusion based on feature-level," *International Journal of Latest Technology in Engineering, Management and Applied Science*, vol. 1, pp. 124–138, 2012.
- [10] F. T. Mahmoudi, F. Samadzadegan, and P. Reinartz, "Object recognition based on the context aware decision-level fusion in multiviews imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, pp. 12–22, 2015.
- [11] A. Dogra, B. Goyal, and S. Agrawal, "From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications," *IEEE Access*, vol. 5, pp. 16040–16067, 2017.
- [12] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.
- [13] D. W. Townsend, T. Beyer, and T. M. Blodgett, "PET/CT scanners: a hardware approach to image fusion," *Seminars in Nuclear Medicine*, vol. 33, pp. 193–204, 2003.
- [14] M. S. Judenhofer et al., "Simultaneous PET-MRI: a new approach for functional and morphological imaging," *Nature Medicine*, vol. 14, pp. 459–465, 2008.
- [15] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75–89, 2016.
- [16] A. P. James and B. V. Dasarathy, "Medical image fusion: a survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.
- [17] Z. Zhou et al., "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Information Fusion*, vol. 30, pp. 15–26, 2016.
- [18] A. C. Muller and S. Narayanan, "Cognitively-engineered multisensor image fusion for military applications," *Information Fusion*, vol. 10, pp. 137–149, 2009.
- [19] R. van de Plas et al., "Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping," *Nature Methods*, vol. 12, pp. 366–374, 2015.
- [20] W. van Aarle et al., "The ASTRA toolbox: a platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
- [21] C. Luengo et al., *DIPimage: a scientific image processing toolbox for MATLAB*, Quantitative Imaging Group, Delft University of Technology, 1999.
- [22] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [23] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109–130, 2001.
- [24] R. D. Tobias, "An introduction to partial least squares regression," in *Proceedings of the Twentieth Annual SAS Users Group International Conference*, Cary, USA, 1995, pp. 1250–1257.
- [25] E. M. M. Manders, F. J. Verbeek, and J. A. Aten, "Measurement of co-localization of objects in dual-color confocal images," *Journal of Microscopy*, vol. 169, pp. 375–382, 1993.