Ultramicroscopy 184 (2018) 90–97

Contents lists available at ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Reduced electron exposure for energy-dispersive spectroscopy using dynamic sampling

Yan Zhang[a], G. M. Dilshan Godaliyadda[a,b], Nicola Ferrier[c], Emine B. Gulsoy[d], Charles A. Bouman[b], Charudatta Phatak[a,*]

[a] Materials Science Division, Argonne National Laboratory, 9700 Cass Ave, Lemont, IL 60439 United States
[b] ECE Department, Purdue University, 465 Northwestern Ave, West Lafayette, IN 47907 United States
[c] Mathematics and Computer Science Division, Argonne National Laboratory, 9700 Cass Ave, Lemont, IL 60439 United States
[d] Department of Materials Science and Engineering, Northwestern University, 2220 Campus Drive, Evanston, IL 60208 United States

ARTICLE INFO

Article history:
Received 23 June 2017
Revised 17 October 2017
Accepted 20 October 2017
Available online 23 October 2017

Keywords:
Scanning electron microscopy (SEM)
Energy dispersive spectroscopy (EDS)
Dynamic sampling
SLADS
Neural networks
Dose reduction

ABSTRACT

Analytical electron microscopy and spectroscopy of biological specimens, polymers, and other beam sensitive materials has been a challenging area due to irradiation damage. There is a pressing need to develop novel imaging and spectroscopic imaging methods that will minimize such sample damage as well as reduce the data acquisition time. The latter is useful for high-throughput analysis of materials structure and chemistry. In this work, we present a novel machine learning based method for dynamic sparse sampling of EDS data using a scanning electron microscope. Our method, based on the supervised learning approach for dynamic sampling algorithm and neural networks based classification of EDS data, allows a dramatic reduction in the total sampling of up to 90%, while maintaining the fidelity of the reconstructed elemental maps and spectroscopic data. We believe this approach will enable imaging and elemental mapping of materials that would otherwise be inaccessible to these analysis techniques.

1. Introduction

Analytical electron microscopy based on energy dispersive X-ray spectroscopy (EDS) is a very versatile and successful technique for exploring elemental composition in microanalysis from the sub-nanometer scale to the micron scale [1–3]. Modern scanning electron microscopes (SEM) equipped with EDS detectors are routinely used for qualitative, semi-quantitative or quantitative elemental mapping of various materials ranging from inorganic to organic, and including biological specimens. Although EDS allows us to identify the elemental composition at a given location with high accuracy, each spot measurement can take anywhere from 0.1 to 10 s to acquire. As a result, if one wants to acquire EDS maps on a rectilinear grid with 256×256 grid points, the total imaging time could be on the order of tens to hundreds of hours. Furthermore, during the acquisition process, the sample gets exposed to a highly focused electron beam that can result in unwanted radiation damage such as knock-on damage, radiolysis, sample charging or heating. Organic and biological specimens are more prone to such damage due to electrostatic charging. Therefore minimizing the total radiation exposure of the sample is also of critical importance. One approach to solve this problem is to sample the rectilinear grid sparsely. However, it is critical that elemental composition maps reconstructed from these samples are accurate. Hence the selection of the measurement locations is of critical importance.

Sparse Sampling techniques in the literature fall into two main categories – Static Sampling and Dynamic Sampling (DS). In Static Sampling the measurement locations are predetermined. Such methods include object independent static sampling methods such as Random Sampling strategies [4] and Low-discrepancy Sampling strategies [5], and sampling methods based on a model of the object being sampled such as those described in [6,7]. In Dynamic Sampling, previous measurements are used to determine the next measurement or measurements. Hence, DS methods have the potential to find a sparse set of measurements that will allow for a high-fidelity reconstruction of the underlying sample. DS methods in the literature include dynamic compressive sensing methods [8,9] which are meant for unconstrained measurements, application specific DS methods [10–12], and point-wise DS methods [13–15]. In this paper, we use the dynamic sampling method described in [15], Supervised Learning Approach for Dynamic Sampling (SLADS). SLADS is designed for point-wise measurement schemes, and is both fast and accurate, making it an ideal candidate for EDS mapping.

* Corresponding author.
E-mail address: cd@anl.gov (C. Phatak).

https://doi.org/10.1016/j.ultramic.2017.10.015
0304-3991/

In SLADS, each measurement is assumed to be scalar valued, but each EDS measurement, or spectrum, is a vector, containing the electron counts for different energies. Therefore, in order to apply SLADS for EDS, we need to extend SLADS to vector quantities or convert the EDS spectra into scalar values. In particular, we need to classify every measured spectrum as pure noise or as one of L different phases. To determine whether a spectrum is pure noise, we use a Neural Network Regression (NNR) Model [16]. For the classification step we use Convolutional Neural Networks (CNNs).

Classification is a classical and popular machine learning problem in computer science for which many well-established models and algorithms are available. Examples include logistic regression and Support Vector Machines (SVM) which have been proven very accurate for binary classification [17]. Artificial neural networks, previously known as multilayer perceptron, have recently gained popularity for multi-class classification particularly because of CNNs [18,19] that introduced the concept of deep learning. The CNNs architecture has convolution layers and sub-sampling layers that extract features from input data before they reach fully connected layers, which are identical to traditional neural networks. CNNs-based classification has shown impressive results for natural images, such as those in the ImageNet challenge dataset [20], the handwritten digits (MNIST) dataset [21] and the CIFAR-10 dataset [22]. CNNs are also becoming popular in scientific and medical research, in areas such as tomography, magnetic resonance imaging, genomics, protein structure prediction etc. [23–26]. It is because of the proven success of CNNs that we chose to use one for EDS classification.

In this paper, we first introduce the theory for SLADS and for detection and classification of EDS spectra. Then, we show results from four SLADS experiments performed on EDS data. In particular, we show experiments on a 2-phase sample measured at two different resolutions and experiments on a 4-phase sample measured at two different resolutions. We also evaluate the performance of our classifier.

2. Theoretical methods

In this section we introduce the theory behind dynamic sampling as well as how we adapt it for EDS.

2.1. SLADS Dynamic sampling

Supervised learning approach for Dynamic Sampling (SLADS) was developed by Godaliyadda et al. [15,27,28]. The goal of dynamic sampling, in general, is to find the measurement which, when added to the existing dataset, has the greatest effect on the expected reduction in distortion (ERD). It is important to note that in this section we assume, as in the SLADS framework, that every measurement is a scalar quantity. We later elaborate how we generalize SLADS for EDS, where measurements are vectors.

First, we define the image of the underlying object we wish to measure as $X \in \mathbb{R}^N$, and the value of location s as X_s. Now assume we have already measured k pixels from this image. Then we can construct a measurement vector,

$$Y^{(k)} = \begin{bmatrix} s^{(1)}, X_{s^{(1)}} \\ \vdots \\ s^{(k)}, X_{s^{(k)}} \end{bmatrix}.$$

Using $Y^{(k)}$ we can then reconstruct an image $\hat{X}^{(k)}$.

Second, we define the distortion between the ground-truth X and the reconstruction $\hat{X}^{(k)}$ as $D\left(X, \hat{X}^{(k)}\right)$. Here $D\left(X, \hat{X}^{(k)}\right)$ can be any metric that accurately quantifies the difference between X and $\hat{X}^{(k)}$. For example, if we have a labeled image, where each label corresponds to a different phase, then,

$$D\left(X, \hat{X}^{(k)}\right) = \sum_{i=1}^{N} I\left(X_i, \hat{X}_i^{(k)}\right), \tag{1}$$

where I is an indicator function defined as

$$I\left(X_i, \hat{X}_i^{(k)}\right) = \begin{cases} 0 & X_i = \hat{X}_i^{(k)} \\ 1 & X_i \neq \hat{X}_i^{(k)}. \end{cases} \tag{2}$$

Assume we measure pixel location s, where $s \in \{\Omega \setminus \mathcal{S}\}$, where Ω is the set containing indices of all pixels, and $\mathcal{S}$ is the set containing pixel locations of all measured pixels. Then we can define the reduction in distortion (RD) that results from measuring s as,

$$R^{(k;s)} = D(X, \hat{X}^{(k)}) - D(X, \hat{X}^{(k;s)}) \ . \tag{3}$$

Ideally we would like to take the next measurement at the pixel that maximizes the RD. However, because we do not know X, i.e. the ground-truth, the pixel that maximizes the expected reduction in distortion (ERD) is measured in the SLADS framework instead. The ERD is defined as,

$$\tilde{R}^{(k;s)} = \mathbb{E}\left[R^{(k;s)} | Y^{(k)}\right] \ . \tag{4}$$

Hence, in SLADS the goal is to measure the location,

$$s^{(k+1)} = \arg\max_{s \in \Omega} \left\{\tilde{R}^{(k;s)}\right\}. \tag{5}$$

In SLADS the relationship between the measurements and the ERD for any unmeasured location s is assumed to be given by,

$$\mathbb{E}\left[R^{(k;s)} | Y^{(k)}\right] = \hat{\theta} V_s^{(k)}. \tag{6}$$

Here, $V_s^{(k)}$ is a $t \times 1$ feature vector extracted for location s and $\hat{\theta}$ is $1 \times t$ vector that is computed in training.

To compute $\hat{\theta}$ we use the procedure described in [15,27]. First, we select M images that are similar to the image of the object we intend to measure. Then, for an image $m \in M$, we first select u_1 number of pixels at random as measurements, and designate the remaining pixels as unmeasured pixels. Then, for every unmeasured pixel s, we can extract a feature vector V_s and compute the RD in Eq. (3) by computing the reconstruction before, $\hat{X}^{(k)}$, and reconstruction after, $\hat{X}^{(k;s)}$, pixel s is included in the measurements. However, with this method, we need to compute 2 reconstructions for each unmeasured pixel, and for the $N - u$ unmeasured pixels, we need to compute $N - u + 1$ reconstructions. To address this problem, an approximation to the RD is introduced in [15,27], which reduces the number of reconstructions to 1 per training image m and measurement selection u.

$$\tilde{R}^{(s)} = \sum_{r \in \Omega} h_{s,r} D\left(X_r, \hat{X}_r\right) \ , \tag{7}$$

where

$$h_{s,r} = \exp\left\{-\frac{1}{2\sigma_s^2} \|r - s\|^2\right\} \tag{8}$$

and $\|r - s\|$ is the Euclidean distance between pixels r and s, and σ_s is given by

$$\sigma_s = \frac{\min_{t \in \mathcal{S}} \|s - t\|}{c}, \tag{9}$$

where $\mathcal{S}$ is the set of measured locations. It is important to note that we have removed the superscript k in these and the proceeding equations because when building the training database, we extract entries with different initial measurements i.e. different values of k. The procedures for estimating this parameter an empirical validation of this approximation are also detailed in [15,27]. Now, from different random selections $u_1, u_2, \ldots u_h$ for each image $m \in M$, we extract a features vector and the corresponding RD for every unmeasured pixel, to form,

$$\mathbf{R} = \begin{bmatrix} \tilde{R}^{(s_1)} \\ \vdots \\ \tilde{R}^{(s_n)} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} V_{s_1} \\ \vdots \\ V_{s_n} \end{bmatrix}. \tag{10}$$