

# Unsupervised learning assisted secondary electron hyperspectral imaging for high-throughput cheminformatics analysis of materials

Jingqiong Zhang<sup>1</sup>, James Nohl<sup>2</sup>, Nicholas Farr<sup>3</sup>, Cornelia Rodenburg<sup>3</sup>, Lyudmila Mihaylova<sup>1</sup>

<sup>1</sup>Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, UK, <sup>2</sup>Department of Materials Science and Engineering, The Faraday Institution, The University of Sheffield, Sheffield, UK,

<sup>3</sup>Department of Materials Science and Engineering, The University of Sheffield, Sheffield, UK

## Background

As a powerful instrumentation tool, scanning electron microscopy (SEM) has been increasingly used to perform analysis on a wide range of minerals, metals, biological specimens, nanostructured materials, polymers, composites, and electronic components [1]. By integrating hyperspectral imaging techniques into SEM, secondary electron hyperspectral imaging (SEHI) technology is able to offer detailed and comprehensive spatial-spectral information of the materials, making it versatile for analysing chemical properties and surface morphology at micro- to nano-scale [2].

However, effectively visualizing spatial-spectral information can still be challenging, especially when changes in spectra are subtle due to spectral mixing or occur only in a very small percentage of the area analysed. To address this problem, we propose a novel analytical workflow for SEHI data using unsupervised learning, which can automatically identify chemical bonds or elements present in the imaged materials and additionally segment the materials surfaces into corresponding chemical groups.

## Methods

The proposed automated analytical workflow includes the following steps:

- (1) Microscopy image data is processed by traversing the whole field of view of the image, through small block-based or pixel-wise methods. For block-based processing, the entire image is divided into smaller, predetermined units (e.g., blocks of 3\*3 pixels).
- (2) For each small block, the peaks in the corresponding spectral curve, also known as spectral peaks, are identified and gathered. This allows for the collection of the overall distribution of all spectral peaks across the image.
- (3) The distribution of spectral peaks is learned by using unsupervised clustering approaches. In this work, the Gaussian mixture model (GMM) approach is adopted to perform probabilistic clustering. The centroid of each GMM component reflects the location of the corresponding spectral peak,

which can be used to deduce the associated chemical bonds or elements in the material sample.

(4) Image blocks that fall into the same cluster are then identified.

Accordingly, the spectra of these image blocks from the same cluster are extracted. These extracted spectral signatures then act as reference spectra, or like "endmembers" in spectral unmixing processes.

(5) By evaluating spectral similarity using spectral angle mapper (SAM), the image regions, sharing similar spectral properties with these reference spectra obtained in (4), are distinguished.

## Results

We implemented this framework into analysing a complex metal alloy (palladium & silver, PdAg) and carbon film, imaged using a Helios Nanolab G3 UC microscope [3]. Firstly, the raw hyperspectral image slices are registered through a template-matching algorithm [4]. By dividing the entire image into smaller blocks and identifying the localized spectral peaks from these blocks, we obtain the distribution of spectral peaks. To figure out the predominant spectral peaks within this distribution, unsupervised clustering by the GMM is applied. As shown in Fig.1, the GMM outcomes reveal 5 components with peak locations at 0.83, 1.98, 3.57, 4.75, and 5.60 eV, respectively. According to the literature, the spectral peaks at 0.83 and 1.98 eV are likely attributed to metals Pd and Ag [3]. The peaks observed in 3-6 eV range are thought to be linked to the contributions from  $sp^2$ -like,  $\alpha$ -CH and  $sp^3$ -like carbon bond types. Intuitively, image blocks that fall into the same cluster, primarily contribute to a particular spectral peak. Thus, the reference spectra, or endmembers-like spectra, can be extracted from the image blocks belonging to the same cluster. The SAM is utilized to assess spectral similarity against these reference spectra for image segmentation.

## Conclusion

Conventional and manual microscopy data analysis methods, due to their limitations in processing efficiency and accuracy, could hinder the applications of SEHI in advanced characterization of materials. Machine learning-based approaches, especially unsupervised learning, offer promising automated analytical solutions for tackling these challenges effectively. The automated analytical workflow proposed here well identifies the chemical bonds and elements present in the imaged materials. The unsupervised clustering method, GMM, perform well in modelling the overall distribution of all spectral peaks and uncovering the chemical bonding types. According to the clustering outcomes, the materials surfaces can then be segmented into the corresponding chemical groups. This novel workflow can facilitate comprehensive cheminformatics analysis of materials, particularly complex carbon material systems.

**Graphic:**

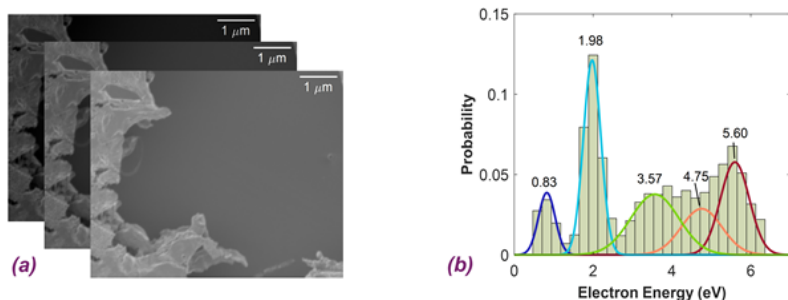


Fig1. (a) Raw SEHI data, collected with a thin PdAg metal alloy and carbon film. (b) Probability histogram with GMM results showing the distribution of all spectral peaks derived from (a).

**Keywords:**

SEHI, clustering, material cheminformatics, segmentation

**Reference:**

- [1] M. Abd Mutalib et al., Membrane characterization, 2017, 161-179. DOI: 10.1016/B978-0-444-63776-5.00009-7
- [2] J. Nohl et al., Micron, 2022 (156), 103234. DOI: 10.1016/j.micron.2022.103234
- [3] K. J. Abrams et al., Advanced Science, 2019 (19), 1900719. DOI: 10.1002/adv.201900719
- [4] J. Nohl et al., Materials Today Advances, 2023 (19), 100413. DOI: 10.1016/j.mtadv.2023.100413