

Iterative Clustering and Classification Analysis of Secondary Ion Mass Spectrometry Images

J. Lefman*** and C. Szakal**

* Oncogenomics Section, Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Gaithersburg, MD, 20877, USA.

** Surface and Microanalysis Science Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA.

Desorption-based mass spectrometry imaging methods such as secondary ion mass spectrometry (SIMS) are constantly evolving due to advances in hardware, data analysis protocols, sample preparation, and knowledge of the targeted specimen. However, despite recent advances, the sample targets have also become immensely more complicated, often with hundreds of chemically unique signatures per image. With a full mass spectrum from 0 to 500 m/z at every pixel within a 256 x 256 pixel image, data mining for chemically localized species can be almost impossible to process manually. To address this issue, we have developed an automated algorithm aimed at identifying chemical differentiations between molecule-specific images. With this method, data analysis time can be drastically reduced while adding a level of robustness and reproducibility to data analysis.

The goal of our approach is to identify and group spatially co-localized masses obtained by SIMS imaging. The algorithm automatically classifies mass images with similar spatial distributions into distinct groups using iterative clustering and image correlation as follows:

1. The first step of the iterative process forms clusters of pixels within the hyper-spectral stack.
2. Correlation values between mass images and each cluster are compared. Each mass image is assigned to a cluster group with the highest correlation value.
3. The algorithm cutoff point is determined by the rate at which the number of images per cluster changes (figure 1).
4. Resulting data are a montage for each cluster; a table containing masses, correlation values, and the assigned cluster group; and a 2D image containing cluster information.

The algorithm effectively partitions co-localized chemical signatures into distinct groups that are otherwise not obvious using manual visual analysis. Typical output resulting from this approach include: image overlays of distinct classes (figure 2) and image montages that are ranked in order of highest correlations with each cluster alongside individual masses corresponding to each molecule-specific image (figure 3). It is expected that this method can be used for analysis of other hyper-spectral imaging datasets.

References:

- [1] Funding for J.L. by National Research Council Research Associateship Programs fellowship.
- [2] Specimen preparation by K. Narayan, J. Fu, and S. Subramaniam (NCI/NIH).

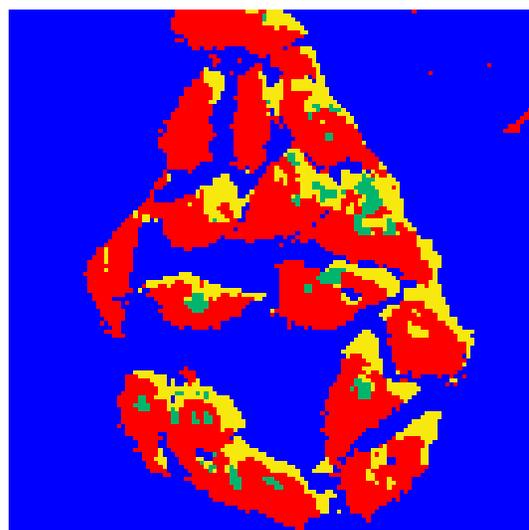
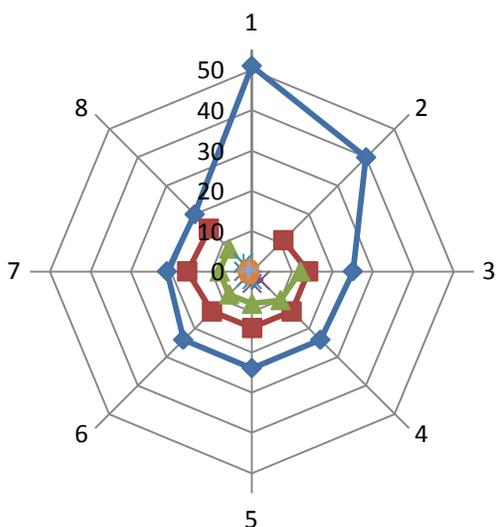


Figure 1. In the circularized graph, each spoke represents the iteration number. The number of clusters per iteration are shown in different colored lines. The radial axis shows the number of images per cluster. For example, in iteration 2, there are 2 clusters with approximately 40 and 10 images per cluster, respectively. The rate of change per cluster dictates the stopping point of the iterative process.

Figure 2. Cluster analysis of iteration number 4 shows three features (red, green, yellow) that are well-defined compared to each other and separated from the background (substrate; blue).

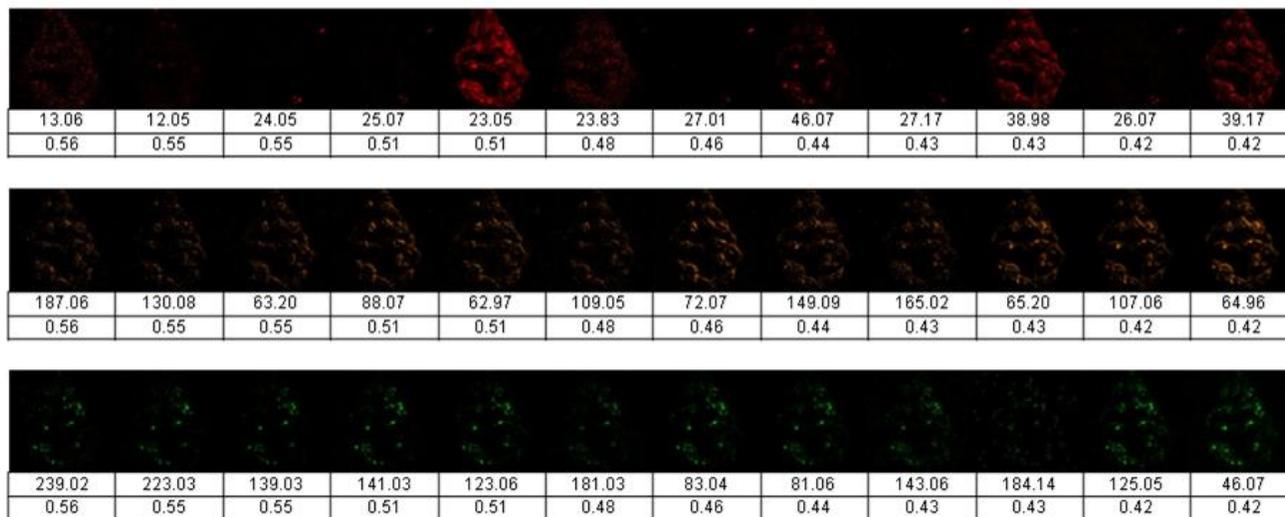


Figure 3. The top 10 images for clusters containing greater than three images are displayed in different colors, corresponding to clusters in figure 2. Below each row of images are the corresponding mass and correlation values.