

Visualization and Analysis of Large Three-Dimensional Hyperspectral Images

Stephen E. Reichenbach^{*a}, Xue Tian^a, Robert Lindquist^a,
Qingping Tao^b, Alex Henderson^c, and John C. Vickerman^c

^aComputer Science & Engineering Department,
University of Nebraska – Lincoln, Lincoln NE 68588-0115, USA;

^bGC Image, LLC, PO Box 57403, Lincoln NE 68505-7403, USA;

^cSurface Analysis Research Centre, Manchester Interdisciplinary Biocentre,
University of Manchester, Manchester M1 7DN, UK

ABSTRACT

New technologies for Secondary Ion Mass Spectrometry (SIMS) produce three-dimensional hyperspectral chemical images with high spatial resolution and fine mass-spectral precision. SIMS imaging of biological tissues and cells promises to provide an informational basis for important advances in a wide variety of applications, including cancer treatments. However, the volume and complexity of data pose significant challenges for interactive visualization and analysis. This paper describes new methods and tools for computer-based visualization and analysis of SIMS data, including a coding scheme for efficient storage and fast access, interactive interfaces for visualizing and operating on three-dimensional hyperspectral images, and spatio-spectral clustering and classification.

Keywords: hyperspectral image processing, three-dimensional image processing, secondary ion mass spectrometry (SIMS)

1. INTRODUCTION

This paper describes informatic technologies for visualizing and analyzing large, three-dimensional, hyperspectral datasets generated by a new generation of chemical imaging technologies, such as time-of-flight secondary ion mass spectrometry (ToF-SIMS). In ToF-SIMS,¹ a beam of primary ions is directed onto a target, eroding molecules and molecular fragments as neutral species and ions (i.e., secondary ions) from the target surface (as illustrated in Fig. 1). The secondary ions that are eroded from the target surface are electrostatically accelerated to a detector that measures their intensity as a function of flight time — data that can be converted to mass spectra. The primary-ion beam can be directed in a raster pattern to create a mass-spectral image and the raster scanning can be repeated to generate a three-dimensional mass-spectral image, as illustrated in Fig. 2.

The Ionoptika J105 3D Chemical Imager,² developed by Professor John C. Vickerman and his research group at the Manchester Interdisciplinary Biocentre (MIB) in conjunction with Ionoptika (Hampshire UK), combines several advances, including polyatomic primary-ion beams and an advanced buncher for secondary ions that facilitates a continuous-beam primary-ion probe. Polyatomic primary-ion beams (e.g., buckminsterfullerene, [C₆₀]⁺) provide greater secondary-ion yield, more uniformity in the secondary-ion yield, and less damage to the substrate of the target than traditional primary-ion beams.³ Greater yield improves the signal-to-noise ratio and sensitivity. Improved uniformity enhances the effective resolution and allows more accurate mapping of chemical constituents. With reduced sub-surface degradation as the surface is eroded, subsequent scans across the target yield more accurate depth profiling to improve three-dimensional chemical imaging. An innovative secondary-ion buncher shapes the electric field that propels the secondary ions for the time-of-flight mass spectrometer, thereby focusing the variable-sized, variable-positioned secondary ions. Time focusing obviates the need for pulsing the primary-ion beam to limit the time range of the secondary ions, which allows quasi-continuous operation of the primary-ion beam.

^{*}Corresponding author contacts: <mailto:reich@cse.unl.edu>, phone +1(402)472-5007

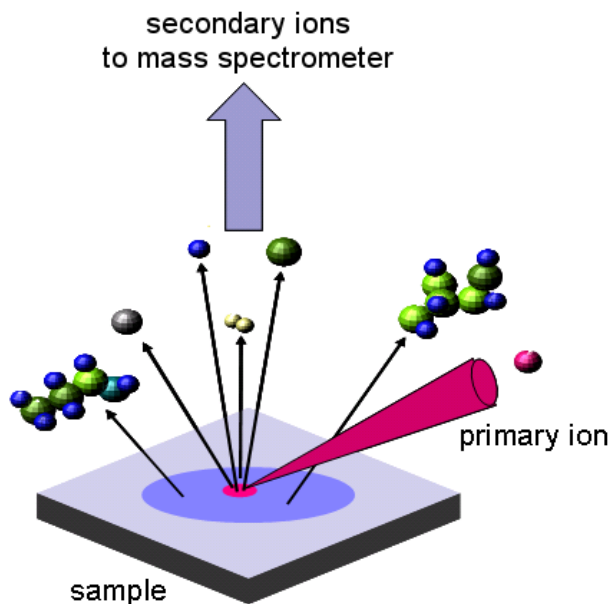


Figure 1. In ToF-SIMS, primary ions directed at the target erode secondary ions for analysis by mass spectrometry. The primary ion beam can be directed in a raster pattern to create a two-dimensional mass-spectral image. Repeating the raster pattern to drill through the target creates a three-dimensional mass-spectral image.

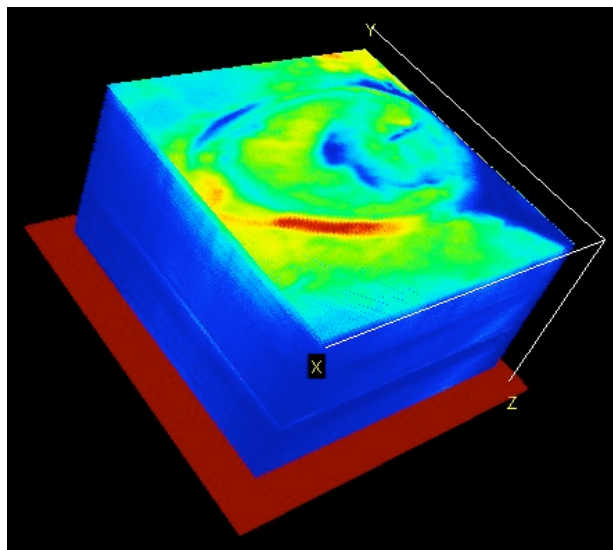


Figure 2. The three-dimensional mass-spectral data can be visualized as a three-dimensional data-cube projected onto two dimensions. Colorization of each data point can be based on the total intensity or ion count (TIC), selected intensity or ion count (SIC), or other spectral properties. (Data from Fletcher et al.²)

The continuous primary-ion beam provides faster analyses and increased spatial resolution. Then, the system uses a harmonic-field reflectron with the property that the time-of-flight in and out of the reflector depends on mass-to-charge only (and not on their variable energy). This creative design provides high-precision mass spectrometry even with continuous operation of the primary-ion beam. The system's high spatial resolution, fine mass precision, and high-sensitivity surface and depth-profile characterizations of the molecular chemistry of heterogeneous materials, including biological tissues and cells, promise to provide an informational basis for important advances in a wide variety of applications, including cancer treatments. However, the volume of data produced poses a significant challenge for interactive visualization and analysis.

The J105 ToF-SIMS instrument will produce three-dimensional datasets on the order of 512^3 spectra with tens of thousands of ToF channels. In the example datasets presented here, individual mass spectra are sampled in up to 85,000 ToF channels at a rate of 1 ns per 8-bit intensity (a raw data rate of 1 gigabyte per second) and accumulated in hardware with an Ortec Fastflight-2TM (Oak Ridge, TN). In the example datasets, 200 to 1000 raw spectra are accumulated per pixel, but the number may be larger or smaller depending on the application. If no more than 256 spectra are accumulated, each accumulated intensity can be represented with a 16-bit unsigned integer in the raw data file, reducing the data rate to 10 megabytes per second (MB/s). Still, a 128×128 , 128-layer image with 85,000 2-byte ToF channels requires 332 gigabytes (GB) without compression and even a 512×512 , one-layer image with 85,000 2-byte ToF channels requires 41GB. Datasets of tens to hundreds of gigabytes cannot be held in the fast memory of typical computer systems, which creates a bottleneck for interactive visualization and analysis with general-purpose imaging software. Real-time, interactive, three-dimensional visualization and analysis requires memory efficient and computationally efficient software.

This paper discusses three fundamental issues for visualization and analysis of large, three-dimensional, hyperspectral data:

- computationally efficient compression for real-time visualization,
- three-dimensional hyperspectral interactivity, and
- spatio-spectral clustering and classification.

2. COMPRESSION FOR REAL-TIME VISUALIZATION

This section describes a compression scheme for holding large three-dimensional ToF-SIMS data sets in memory for fast visualization. The access pattern is a critical consideration for data compression. The most important access mode for interactive SIMS visualization is retrieving spectra by pixel (i.e., spatial position). Analysts determine chemical compositions on the basis of mass spectral characteristics, so mass spectra viewing is fundamental. Common interactive operations are to view the mass spectrum at a point in the image space indicated by point-and-click and to view the mass spectrum summed over a spatial region indicated by drawing. An important operation is to generate a classification rule(s) based on mass spectra in two (or more) regions. Analysts also view a mass-spectral range (e.g., for a selected ion) across the image space, but such spectral-spatial viewing does not require immediate interactivity to the degree required in pixel-oriented spatial-spectral access. Moreover, operations to generate spatial maps from their spectra require pixel-by-pixel access to many or all ToF channels, e.g., to map regions that satisfy a classification rule.

Tretter, Memon, and Bouman⁴ cite two approaches among methods for lossless compression of hyperspectral images: predictive coding and reversible transforms, each followed by context modeling and coding. Both approaches can be applied either with respect to the spatial dimensions or to the spectral dimension (or both). Predictive coding has been the predominant approach for hyperspectral data. Lossless transform coding methods for hyperspectral data are newer and typically require greater computation than lossless predictive methods, but may achieve greater compression. Given the motivation of interactive visualization and analysis of SIMS data, low computational complexity is more important than optimal compression, so the more traditional approach of predictive coding may be better suited. Given the primary need for spatial access in SIMS analysis, each pixel spectrum should be compressed separately.

Reichenbach et al.⁵ recently described a method that codes individual spectra, consistent with the predominant access mode for SIMS analysis, based on statistical and structural characteristics of SIMS spectra. Unlike some hyperspectral data generated by remote sensing satellites, for which many hyperspectral compression methods have been developed, SIMS spectra have many zero values and the probability distribution of the intensity values is skewed significantly, decreasing rapidly with magnitude. Also, many of the non-zero values are in adjacent ToF channels, forming peaks in the mass spectra. These statistical characteristics can be exploited to give highly compressed data that can be accessed quickly.

The most notable characteristic of SIMS hyperspectral data is that many of the intensity values are zero. This characteristic of the data reflects the fact that some mass-to-charge ratios do not occur physically and the fact that the number of chemical constituents at a sample point of the target limits the number of secondary ions and therefore the number of peaks in each mass spectrum. Another important characteristic of the data is that the probability distribution of the intensity values decreases with intensity. Many of the non-zero values are equal to one (many of which may be noise but must be coded in a lossless method) and most non-zero values are less than 256. Because of the large number of zero values and the long-tailed skewed probability distributions, least-squares predictors, which are effective for remote sensing data, do not perform well for SIMS data. Another characteristic of the datasets is that ion peaks in the mass spectra may be wider than the ToF channels, so each mass spectral peak may cause several non-zero values in adjacent ToF channels.

The high probability of zero-values and the clustering of non-zero values suggest that run length encoding may be used to effectively code the long runs of zeros between non-zero values. Commonly used sparse array representations of mass spectra (i.e., recording the mass and value for each non-zero value) similarly take advantage of the large number of zeros to efficiently represent MS data. If, instead of the ToF channel index, the differential of indexes of non-zero-valued channels is used (i.e., the difference between the index of the next channel with a non-zero value and one more than the index of the current non-zero channel), the result is a run length code.

Based on the SIMS data characteristics, the compression method separately codes the ToF differentials and non-zero intensity values. Because many of the ToF differentials are zero, it is important to code them most efficiently. Because the compressed data will be decoded for visualization, the approach uses representations of integer byte-lengths which do not require computation for decoding — just byte copies. Accordingly, the method uses 2-bit length codes to record the number of bytes for each ToF differential and zero bytes are used if the differential is zero. The length codes (in binary) are: 00 if the differential is zero, with no separate representation

of the differential; 01 if the differential is in the range 1–255, with the differential coded in one byte; 10 if the differential is in the range 256–65535, with the differential coded in two bytes; and 11 if the differential is 65536 or larger, with the differential coded in four bytes. So, only two bits are required for the ToF differentials that are equal to zero, 10 bits are required for the ToF differentials in the range 1–255, etc. The differential codes can be retrieved quickly using byte copies.

The non-zero intensity values could be compressed by any method, but the scheme used for the ToF differentials can be used and is justified by the significant number of ones and small values. Also, the integer byte-length scheme allows quick retrieval of the intensity for a specific channel, decoding only the ToF differentials and the intensity byte-lengths to locate the byte(s) with the intensity value. So, the non-zero intensity values are reduced by one (so that the smallest value to be recorded, which is one, is mapped to zero) and then coded using the length-coding scheme described above. Decoding restores the non-zero values by adding one.

3. THREE-DIMENSIONAL HYPERSPECTRAL INTERACTIVITY

This section describes methods for interactive visualization of three-dimensional ToF-SIMS data. Interactivity with three-dimensional hyperspectral data requires several types of views.

Three-dimensional visualization is required to view the entire spatial data-cube. Display devices are two-dimensional, so three-dimensional data must be projected onto two dimensions for viewing, as shown in Fig. 2. The 4DTM ToF-SIMS software being developed at GC Image, LLC, supports perspective projection with mouse-controlled rotation to view the data from any three-dimensional angle and zoom to move the data-cube closer or more distant. Radio buttons provide six standard orthogonal views (one of the six data-cube faces shown fully in front) and eight isometric views (one of the eight data-cube corners with the three intersecting faces shown fully at equal angles). A control also is provided for the aspect ratio of the depth (or z) dimension.

However, there are important limitations of three-dimensional views. First, each data-point can be shown with only one color, whereas the mass spectrum at each data-point has many values. By default, the 4D ToF-SIMS colors each point by mapping the total ion or intensity count (TIC), i.e., the sum of intensities across the mass spectrum, through a color-map function (e.g., a conventional cold-to-hot color-map with shades of blue, green, yellow, and red indicating increasing value). By default, the mapping is linear, but the software also provides a color-temperature slider to implement logarithmic or exponential mapping. As described below, the software also allows user-specified mass spectral ranges, i.e., selected ion or intensity count (SIC), to be used instead of the TIC, which allows visualization of spectral channels indicative of chemically important ions. Recomputation of the value for each data-point (when the TIC or SIC is changed) is relatively fast if the compressed data can be held in memory.

A second problem is related to the two-dimensional nature of display devices and vision. With a projection from three dimensions to two, data on the inside of the data-cube can be obscured. Two-dimensional views, described next, allow slices through the data-cube to expose interior data, but various techniques can be used to help show “inside” data in the 3D projection. The 4D ToF-SIMS software allows a transparency threshold which makes data-points with values below the threshold transparent, allowing larger values inside the cube to be seen, as shown in Figure 3. A control for an upper threshold is being added, which will allow for isosurfaces, showing only data-points within an intensity range.

Two-dimensional visualization is convenient for viewing a spatial slice through the data-cube, which exposes interior data. The 4D ToF-SIMS software provides a slicing plane in the three-dimensional view that can be oriented perpendicular to any of the three spatial dimensions and positioned along that dimension with a slider. As the slicing plane is repositioned, the selected image slice is shown in a separate two-dimensional image viewer, as shown in Fig. 4. The slice image has the same source data (e.g., TIC or SIC) and colorization mapping as the three-dimensional view. Because the data is extracted from the array that is used for the three-dimensional view, the image view is shown in real-time as the slice plane is repositioned. This allows interactive two-dimensional animation of the three-dimensional data along any spatial dimension.

Displaying a two-dimensional slice as an image provides a convenient interface for precisely indicating data-points or drawing regions. The 4D ToF-SIMS software provides several drawing interfaces to interactively

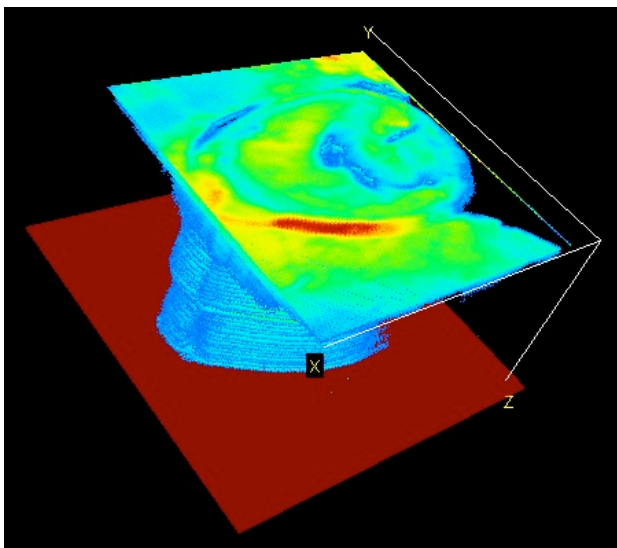


Figure 3. A threshold on the values shown allows viewing of larger valued data-points inside the data-cube.

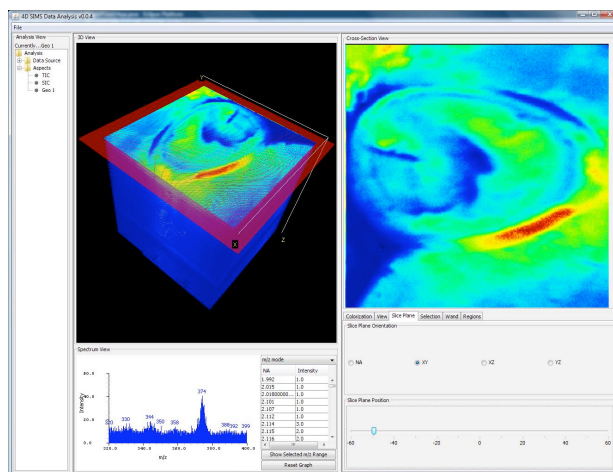


Figure 4. As the slice plane (shown in red in the three-dimensional view) is repositioned (with the slider in the lower-right panel), the two-dimensional image is updated.

delineate regions: rectangle (enclosed region, including single-point selection), polygon (enclosed region), free-hand (enclosed region), and scribble. The software allows users to build composite regions using discard (new), addition (union), subtraction, and replace (discard followed by addition). Composite regions are maintained as *geometric aspect functions* that can be saved, loaded, edited, and visualized with masks in the two-dimensional and three-dimensional views.

One-dimensional visualization is especially convenient for showing the mass spectrum of an indicated data-point or the summed mass spectrum for a selected region. In the 4D ToF-SIMS software, the spectrum viewer displays a spectrum in graphical and tabular formats, as illustrated at toward the bottom-left of the graphical user interface (GUI) in Fig. 4. The abscissa of the spectrum can be set to ToF, mass-to-charge (m/Q) or nominal mass-to-charge (rounded to whole numbers). ToF-SIMS systems generate hyperspectral data — intensity arrays with tens of thousands of values — so neither display screen-resolution nor visual acuity is sufficient to perceive all spectral intensities simultaneously. Therefore, the graphical view of the spectrum allows zooming to show sub-ranges of the spectrum and the tabular view supports scrolling. The tabular view can be sorted either by abscissa or intensity (the ordinal) with either increasing or decreasing values.

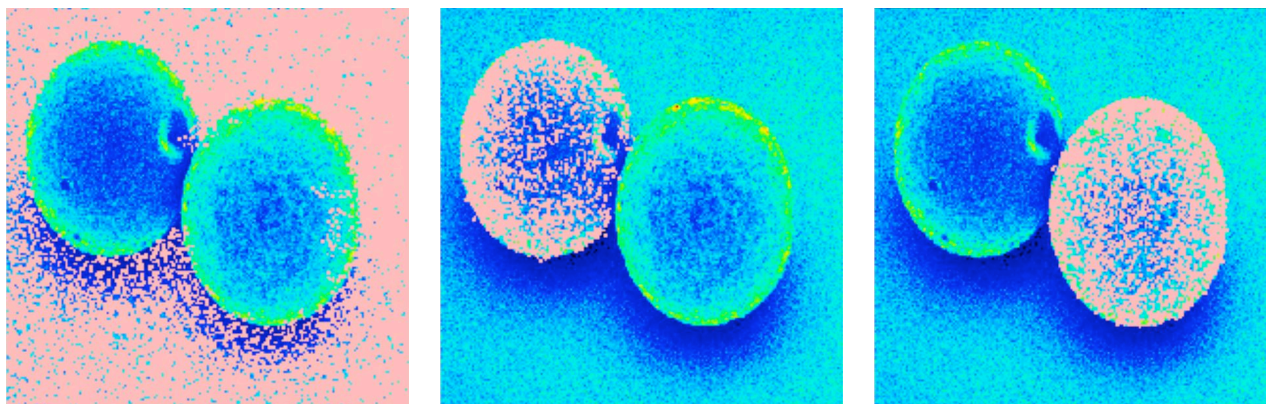
The graphical view allows interactive delineation of a SIC range by mouse click-and-drag. The SIC can then be visualized in the three-dimensional and two-dimensional spatial views with a button click. Generating new SIC visualizations requires summing intensities for the indicated range and may take a few seconds depending on the size of the data and SIC range. Specific SIC ranges (e.g., that are indicative of various chemically important ions) can saved, loaded, and edited, as well as visualized.

4. SPATIO-SPECTRAL CLUSTERING AND CLASSIFICATION

Analyses of ToF-SIMS images may entail both spatial and spectral features. For example, it may be necessary to delineate a spatial feature such as a cell or components within a cell. Similarly, it may be necessary to identify spectral features, such as a spectral signature indicating the presence of a drug.

The 4D ToF-SIMS software provides several interactive tools for spatial and structural analysis. The previous section described drawing tools that can be used to delineate spatial regions and examine the summed mass spectrum for the region. The Spectral Similarity and Magic Wand tools support semi-automated region selection.

The **Spectral Similarity** tool selects data points at which the similarity with a reference spectrum is greater than a specified threshold. First, the user selects a data point or region to provide a reference spectrum. Then, at



A. Seed in the background.

B. Seed in the left bead.

C. Seed in the right bead.

Figure 5. The Magic Wand can be used to select proximal and spectrally similar data points for an indicated seed point. (Data from Winograd and Braun.⁶)

every data point i , the similarity of spectrum s_i and the reference spectrum r is computed as the cosine between the vectors of the two spectra:

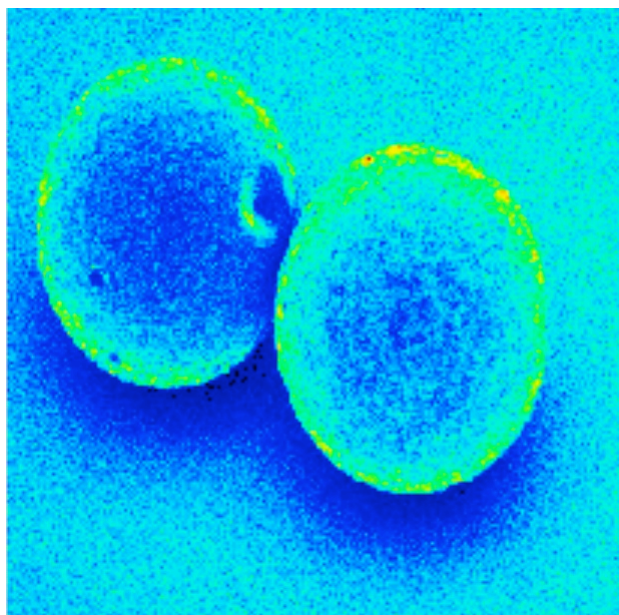
$$\text{spectralSimilarity}(r, s_i) = \frac{r \cdot s_i}{|r| |s_i|}, \quad (1)$$

which varies from 0 for orthogonal mass spectra to 1 for mass spectra that are identical after normalization. The Angular Threshold slider allows the user to interactively adjust the threshold for selection, i.e., if its similarity with the reference spectrum is greater than the threshold then the data point is included. Because the similarities for each pixel are computed only once, the thresholding can be performed in real-time as the slider is adjusted.

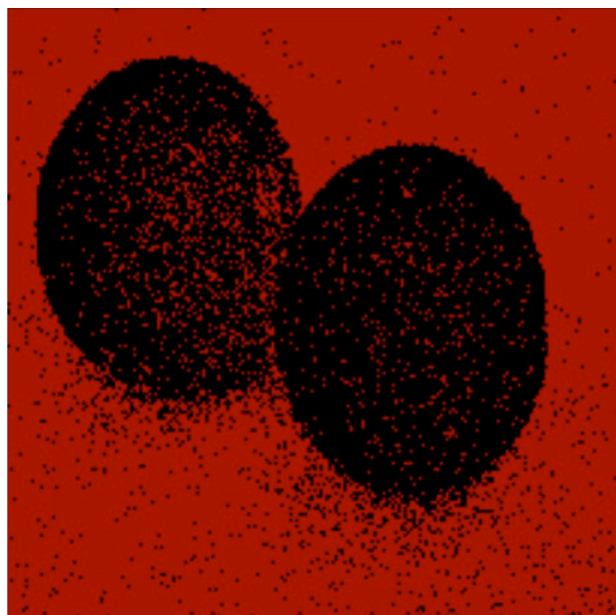
The **Magic Wand** tool selects data points based on both spectral similarity and spatial proximity. First, the user selects a data point to provide both a reference spectrum and a seed for the selected spatial region. Then, the software iterates a region-growing process. In each cycle of the iteration, data points within a specified spatial distance of any selected data point (initially just the seed) are tested for similarity with the reference spectrum. Every data point within the specified distance that meets the similarity criterion is added to the selected region. This iterative process repeats each time the region grows, then stops when no more data points nearby any selected data points are similar enough. Two sliders are provided to parameterize the Magic Wand: the jump parameter, which specifies the distance from a selected data point at which that the region-growing process can include additional data points, and the similarity threshold, which specifies the level of similarity required for new data points to be added to the selected region. Two types of similarity are allowed: TIC similarity and spectral similarity. Spectral similarity is given in Eq. (1). TIC similarity ranges linearly from 0 for the difference between the largest and smallest TIC values in the data to 1 for identical TIC values.

Fig. 5 illustrates use of the Magic Wand on ToF-SIMS data from two 200- μm polystyrene beads coated with different peptide mimics with molecular weights 226 and 547 on a silicon substrate.⁶ In each image, the region selected by the Magic Wand is shown as a mauve-colored overlay on the colorized TIC image. In Fig. 5A, the seed point is in the background and the Magic Wand successfully selects much of the background. In Figs. 5B and 5C, the seed point is respectively in the left and right bead and the Magic Wand selects much of the indicated bead.

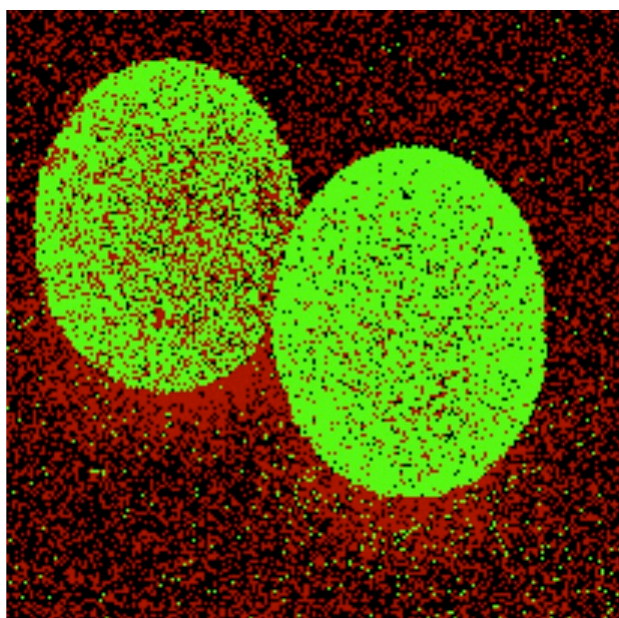
The **Spectral Clustering** tool provides a uniform interface for various clustering algorithms, including k -means,⁷ hierarchical clustering,⁸ and spectral clustering.⁹ Clustering algorithms group objects such that those in the same group are more similar with one another than with objects in other groups. In the 4D ToF-SIMS software, the normalized mass spectrum (each intensity in the spectrum divided by the largest intensity in the spectrum) is used for assessing the similarity of data points. First, the user selects a subset of data points to be clustered. Then, the user selects the clustering algorithm and provides its parameters, e.g., some algorithms



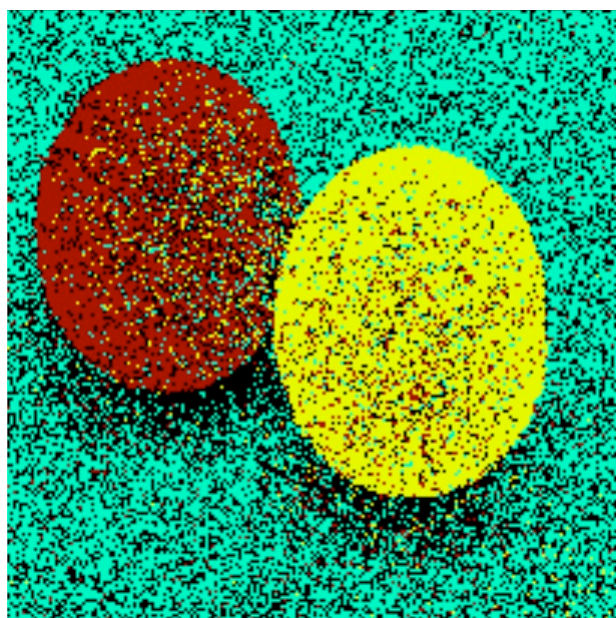
A. TIC.



B. Two clusters, k -means with $k=2$.



C. Three clusters, k -means with $k=3$.

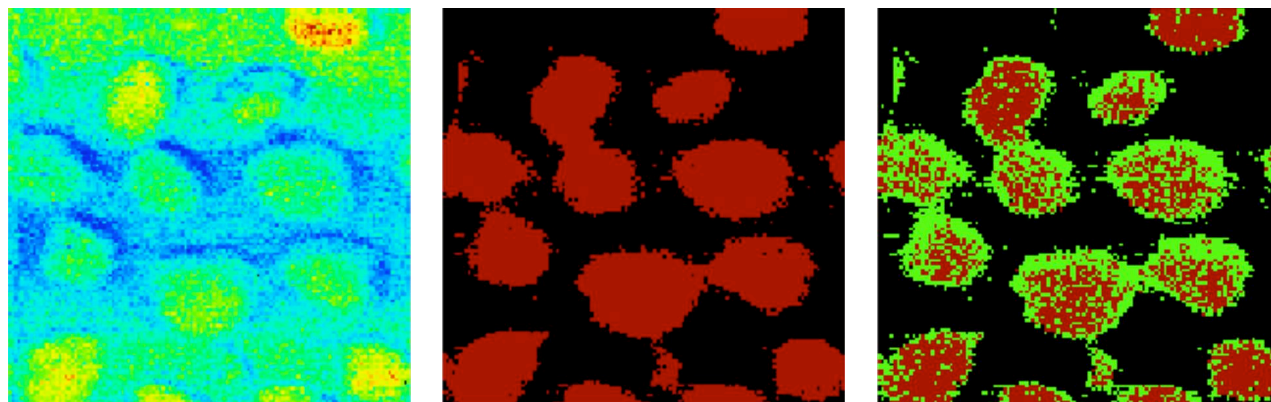


D. Four clusters, k -means with $k=4$.

Figure 6. Clustering of ToF-SIMS data of coated polystyrene beads.

require the user to specify the number of clusters. Then, the algorithm separates the data points into clusters based on their spectral similarity (as determined by the algorithm).

Fig. 6 illustrates clustering with the ToF-SIMS data from two polystyrene beads coated with different compounds. Fig. 6A shows the colorized TIC image. As can be seen in this image, the background regions below the beads and the centers of the beads have smaller TIC values, indicating less intense mass spectra. With only the TIC, the data points in the two beads cannot be distinguished from one another nor from the background. Fig. 6B shows k -means clustering with two clusters, i.e., $k=2$. This clustering is mostly successful in separating the two beads (colored black) from the background (colored red). Fig. 6C shows k -means clustering with



A. TIC.

B. Two clusters, k -means with $k=2$.

C. Three clusters, k -means with $k=3$.

Figure 7. Clustering of ToF-SIMS data of HeLa cells. (Data from S. Rabbani and J. Fletcher, Surface Analysis Research Centre, University of Manchester.)

three clusters ($k=3$). This clustering also separates the beads (green) from the background and divides the background into two clusters, apparently separating the more intense background (black) and the less intense background (red). With four clusters ($k=4$), in Figure 6D, the two beads are well separated (red and yellow) from each other and from the background (cyan and black).

Fig. 7 shows clustering with ToF-SIMS data from HeLa cells (an immortalized cervical cancer cell line widely used as a standard cell type for a wide range of biological research). The cells were cultured on poly(L-lysine) coated silicon shards for 24 hours at 37°C with 5% CO₂, washed briefly in 0.15 M ammonium formate solution to remove salt ions, then freeze-dried prior to ToF-SIMS analysis. Typical cell diameter is 20 μm. The data was acquired using 40 keV C₆₀⁺ primary ions scanned to 128×128 pixels over a field of view 88×108 μm² (a pixel dimension of approximately 0.7×0.8 μm²). The total ion dose used was 3 × 10¹⁴ ions cm². Fig. 7A shows a TIC image. Fig. 7B shows that k -means clustering with two clusters ($k=2$) separates the cells and background. In Fig. 7C, k -means clustering with three clusters ($k=3$) apparently separates the cell edges and interiors.

Spectral Classification is based on supervised training: two (or more) geometric aspect functions designate distinct classes of data points, then the classification algorithm assigns class membership to other data points based on mass spectral characteristics. Experiments with the bead data (Fig. 6) compared four classification algorithms: C4.5 decision trees,¹⁰ soft independent modeling of class analogy (SIMCA),¹¹ principal component analysis (PCA)¹² with discriminant function analysis (DFA),¹³ and the most similar neighbor with a probability-based spectrum similarity measure (MSN-PSSM).¹⁴ Two data sets were constructed, the first with 100 data points from each bead and the second with 50 data points from each bead. With these data sets, leave-one-out cross-validation, which is commonly used in chemometrics, was used for testing.

Overall classification accuracy and Fleiss' kappa statistic¹⁵ are used to quantitatively measure the performance of the different algorithms. Overall classification accuracy is defined as:

$$\text{Accuracy} = \frac{\# \text{ correctly classified samples}}{\# \text{ samples in dataset}}. \quad (2)$$

Fleiss' kappa statistic is a chance corrected measure of agreement between two sets of categorized data that assesses agreement compared to random-chance levels. In this study, kappa measures the agreement between samples' true labels and samples' classified labels for different algorithms. A kappa value of 1 means perfect agreement, 0 means agreement expected by chance, and -1 means perfect disagreement. Interpretation of the kappa values is based on Fleiss' cutoffs:¹⁶ kappa values exceeding 0.75 suggest agreement well above chance, values in the range of 0.40 to 0.75 indicate fair levels of agreement above chance, and values below 0.40 are indicative of poor agreement above chance levels. Table 1 shows the overall classification accuracy and Fleiss' kappa statistic for each classification algorithm. Decision trees and PCA with DFA performed best, followed by MSN-PSSM, then SIMCA.

Table 1. Classification results with four classification methods for two datasets.

Classifier	Dataset 1		Dataset 2	
	Accuracy (%)	Kappa	Accuracy (%)	Kappa
Decision trees	90.00	0.80	93.50	0.87
SIMCA	80.00	0.60	83.00	0.66
PCA with DFA	91.00	0.82	92.50	0.85
MSN-PSSM	89.00	0.78	90.00	0.80

5. CONCLUSION

This report summarizes new methods and tools for computer-based visualization and analysis of high-resolution, three-dimensional, hyperspectral ToF-SIMS data. The informatics suite includes a coding scheme for efficient storage and fast access, interactive interfaces for visualizing and operating on three-dimensional hyperspectral images, and spatio-spectral clustering and classification.

The goal of the work to-date is proof-of-concept and the development of a prototype foundation for future work. Future work will include continued evaluation of the coding effectiveness as the instrument evolves, interactive spatial operations for 3D drawing such as rotation and extrusion, algorithmic spatial operations such as dilation and erosion, improved clustering and supervised classification methods, and a general framework for spectral aspects functions such as PCA, ion ratios, etc.

ACKNOWLEDGMENTS

This work was supported by the USA National Science Foundation funding to S. E. Reichenbach (IIS-0431119) and Q. Tao (IIP-0741027) and by the UK Engineering and Physical Sciences Research Council's "Collaborating for Success through People" funding to John C. Vickerman (EP/FO12985). The authors gratefully acknowledge the support and data provided by John Fletcher, Sadia Rabbani, and others at the Surface Analysis Research Centre of the University of Manchester.

REFERENCES

- [1] Vickerman, J. C., "ToF-SIMS – an overview," in [*ToF-SIMS: Surface Analysis by Mass Spectrometry*], Vickerman, J. C. and Briggs, D., eds., 1–40, SurfaceSpectra, Manchester, UK (2001).
- [2] Fletcher, J. S., Rabbani, S., Henderson, A., Blenkinsopp, P., Thompson, S. P., Lockyer, N. P., and Vickerman, J. C., "A new dynamic in mass spectral imaging of single biological cells," *Analytical Chemistry* **80**(23), 9058–9064 (2008).
- [3] Fletcher, J. S., Lockyer, N. P., Vaidyanathan, S., and Vickerman, J. C., "TOF-SIMS 3D biomolecular imaging of xenopus laevis oocytes using buckminsterfullerene (C60) primary ions," *Analytical Chemistry* **79**(6), 2199–2206 (2007).
- [4] Tretter, D., Memon, N., and Bouman, C. A., "Multispectral image compression," in [*Handbook of Image and Video Processing*], Bovik, A., ed., 539–554, Academic Press, San Diego CA (2000).
- [5] Reichenbach, S. E., Henderson, A., Lindquist, R., and Tao, Q., "Efficient encoding and rapid decoding for interactive visualization of large three-dimensional hyperspectral chemical images," *Rapid Communications in Mass Spectrometry* **23**(9), 1229–1233 (2009).
- [6] Winograd, N. and Braun, R. M., "Imaging mass spectrometry and combinatorial chemistry," *Spectroscopy* **16**(9), 14–27 (2001).
- [7] MacQueen, J. B., "Some methods of classification and analysis of multivariate observations," in [*Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*], 281–297 (1967).
- [8] Johnson, S., "Hierarchical clustering schemes," *Psychometrika* **32**(3), 241–254 (1967).
- [9] Ng, A. Y., Jordan, M. I., and Weiss, Y., "On spectral clustering: Analysis and an algorithm," in [*Advances in Neural Information Processing Systems*], **14**, 849–856, MIT Press, Cambridge MA (2002).
- [10] Quinlan, J. R., [*C4.5: Programs for Machine Learning*], Morgan Kaufmann (1993).

- [11] Wold, S. and Sjoström, M., “SIMCA: A method for analyzing chemical data in terms of similarity and analogy,” in [*Chemometrics Theory and Application*], *American Chemical Society Symposium Series* **52**, 243–282 (1977).
- [12] Jolliffe, I. T., [*Principal Component Analysis*], Springer, New York NY, second ed. (2002).
- [13] Krzanowski, W. J., [*Principles of multivariate analysis: a user’s perspective*], Oxford University Press, New York NY, revised ed. (1988).
- [14] Tian, X., Reichenbach, S. E., Tao, Q., and Henderson, A., “Classification and cluster analysis of complex time-of-flight secondary ion mass spectrometry for biological samples,” in [*International Conference on Bioinformatics, Computational, Biology, Genomics and Chemoinformatics*], in press (2009).
- [15] Fleiss, J. L., “Measuring nominal scale agreement among many raters,” *Psychological Bulletin* **76**(5), 378–382 (1971).
- [16] Fleiss, J. L., [*Statistical Methods for Rates and Proportions*], Wiley, New York NY, second ed. (1981).