


APPLICATION NOTE

Evaluation of Digital Image Recognition Methods for Mass Spectrometry Imaging Data Analysis

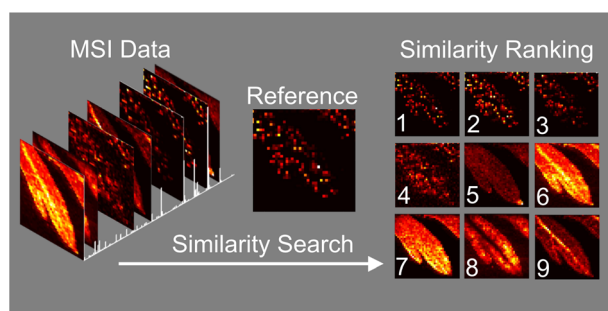
Måns Ekelöf,¹ Kenneth P. Garrard,¹ Rika Judd,² Elias P. Rosen,³ De-Yu Xie,²
Angela D. M. Kashuba,³ David C. Muddiman^{1,2,4} 

¹FTMS Laboratory for Human Health Research, Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA

²Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA

³Division of Pharmacotherapy and Experimental Therapeutics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

⁴Molecular Education, Technology, and Research Innovation Center (METRIC), North Carolina State University, Raleigh, NC 27695, USA



Abstract. Analyzing mass spectrometry imaging data can be laborious and time consuming, and as the size and complexity of datasets grow, so does the need for robust automated processing methods. We here present a method for comprehensive, semi-targeted discovery of molecular distributions of interest from mass spectrometry imaging data, using widely available image similarity scoring algorithms to rank images by spatial correlation. A fast and powerful batch search

method using a MATLAB implementation of structural similarity (SSIM) index scoring with a pre-selected reference distribution is demonstrated for two sample imaging datasets, a plant metabolite study using *Artemisia annua* leaf, and a drug distribution study using maraviroc-dosed macaque tissue.

Keywords: Mass spectrometry imaging, Image recognition, SSIM

Received: 3 July 2018/Revised: 24 August 2018/Accepted: 26 September 2018/Published Online: 15 October 2018

Introduction

Mass spectrometry imaging (MSI) datasets are highly complex, containing abundance and distribution information about thousands of chemical species. As sample probes and ionization techniques have evolved, the information density of untargeted (discovery) MSI data has increased to the point where comprehensive manual interpretation is not practical. Some degree of automation is often employed to extract features of interest in a semi-targeted fashion.

The desired outcome of discovery-type MSI experiments is typically the identification of molecular distributions

correlated to some other features such as a known region of the sample, or the distribution of some known compound such as a disease marker, isotopic label, or a drug. For this type of study, data interpretation comes down to finding images of a *particular appearance* from a limited search space. This is, in essence, an image recognition problem similar to that of facial recognition or compression quality evaluation in digital image processing [1].

The gold standard for calculating the perceived similarity of two given images is the structural similarity (SSIM) index [2, 3]. The SSIM algorithm arose from a need to automatically predict the perceived quality of digital images after compression or other processing. To calculate the SSIM index for a pair of aligned images, each image is subdivided into smaller sub-images, typically by generating a small window around each pixel. For each aligned pair of sub-images x and y , the arithmetic mean (μ_x, μ_y), standard deviation (σ_x, σ_y), and Pearson's

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13361-018-2073-0>) contains supplementary material, which is available to authorized users.

Correspondence to: David Muddiman; e-mail: dcmuddim@ncsu.edu

correlation coefficient ($\sigma_{xy}/\sigma_x\sigma_y$) are calculated. The mean intensity and standard deviations are converted into 0–1 scores which are multiplied together to generate the SSIM score as shown in Eq. (1). The final result can be shown either as a map of local similarities, or as a mean SSIM (MSSIM) score for the whole image as shown in Eq. (2).

$$\text{SSIM}(x_i, y_i) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} * \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} * \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \text{luminance} * \text{contrast} * \text{structure} \quad (1)$$

$$\text{MSSIM} = \frac{\sum_{i=1}^n \text{SSIM}(x_i, y_i)}{n-1} \quad (2)$$

Methods

MSiReader Implementation

To apply image recognition methods to real MSI data, the batch processing function of MSiReader [4, 5] was modified to enable correlation scoring for a range of MS images with a given reference image. The SSIM algorithm is included in the MATLAB Image Processing Toolbox (The MathWorks, Inc., Natick, MA, USA). The MATLAB implementation of SSIM calculates the index at each pixel by applying a circular gaussian weighting function of adjustable radius. The combined score at each pixel is then calculated as

$$\text{SSIM} = \text{luminance}(x, y)^\alpha \times \text{contrast}(x, y)^\beta \times \text{structure}(x, y)^\gamma \quad (3)$$

Where the weighting constants α , β , and γ can be set by the user. They default to 1. An example of SSIM output using 200×200 monochrome images is shown in Figure 1, illustrating both the individual components and final scoring (mean SSIM).

Evaluation of Imaging Datasets

In order to test the usefulness of image recognition for real problems, two imaging datasets were produced, selected to be representative of typical work done in our lab. Each image was acquired using IR-MALDESI ionization coupled to a Q Exactive Plus mass spectrometer operating at a nominal resolving power of 140,000 as previously described [6]. The raw data was converted to .imzml using msconvert [7] and imzmlconverter [8], and loaded into MSiReader for analysis. Normalization to maximum abundance per image was used to ensure matching based on relative rather than absolute ion abundance for the luminance score. All heatmaps were generated using the “hot” colormap preset in MSiReader.

All raw data used are provided in .mzml and .imzml format in the electronic supplement. The image recognition tools used are included in the current open-source and stand-alone versions of MSiReader (v. 1.01), available at <http://www.msireader.com>.

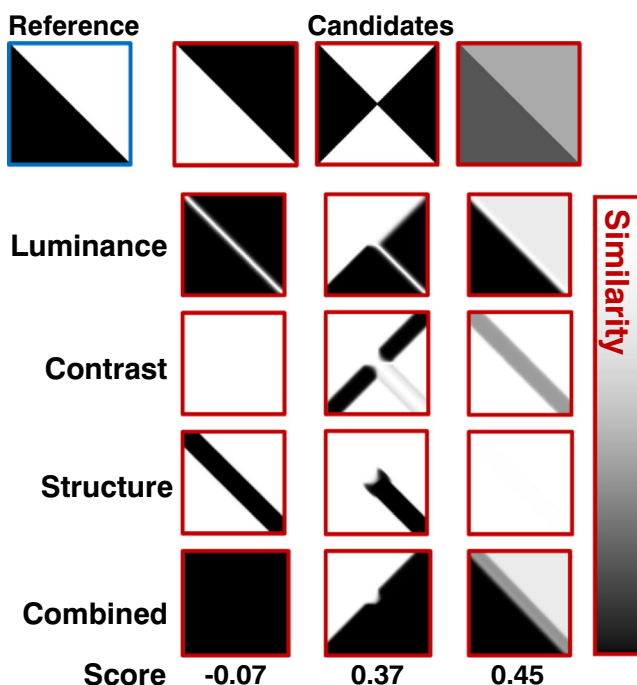


Figure 1. SSIM analysis of three example images for a given reference, with the separate components shown separately and combined into a numerical score (mean SSIM). The SSIM output shown was produced using the same MATLAB implementation as used for MSI evaluation, but with the Gaussian radius weighting set to a value of 6 for demonstration purposes

Imaging of Artemisia annua Leaf The sweet wormwood (*Artemisia annua*, Chinese: *Qinghao*), native to China, is notable as the primary natural source of artemisinin, a powerful antimalarial compound, the discovery of which was awarded the 2015 Nobel medicine prize [9]. Artemisinin and other related metabolites (e.g., its precursors and derivatives) are accumulated in glandular trichomes on the leaf surface, the size and density of which depend on spatial positions of leaves and plant ages [10, 11]. The unique chemical composition and localization of glandular trichomes on the leaf surface makes it suitable as a validation system for MSI data analysis.

Leaves on the 15–17th nodes of 2-month old *A. annua* plants, grown in the NC State phytotron, were collected and affixed to a glass microscopy slide using double-adhesive tape. A 2×2 mm region of interest was imaged in negative mode at a spatial resolution of $50 \mu\text{m}$ (40×40 scans), in the mass range of m/z 100–400. The molecular ion of intact artemisinin $[\text{M}-\text{H}]^-$ observed at m/z 281.1395 was selected as reference for image scoring. The MSiPeakfinder tool was used to pre-generate a list of 332 masses with a $2\times$ or higher abundance ratio in scans from leaf tissue compared to blank scans. This reduced dataset was used to evaluate the effect of the various SSIM parameters (α , β , γ , Gaussian radius).

Imaging of Drug-Dosed Macaque Lymph Node Combinations of antiretroviral (ARV) therapies have radically improved health outcomes for persons living with HIV. Interruption of

these regimens, however, leads to rapid viral rebound that may result from inadequate penetration of drug into tissues where virus primarily resides such as lymph nodes [12]. Tissue disposition of the viral entry-inhibitor maraviroc was investigated in the lymph node of a rhesus macaque, an animal model of infection, receiving 270 mg/kg maraviroc dosed twice daily. Since ARV tissue distribution can be highly heterogeneous [13], MSI analysis provides a useful tool in identifying ions accumulating in similar patterns to maraviroc that may participate in its trafficking and metabolism within the lymph node.

A 10- μm -thick section of dosed lymph node was imaged in positive mode at 100- μm spatial resolution (75×90 scans, or 7.5×9 mm), in the mass range of m/z 200–800. Comprehensive SSIM analysis was performed by binning the whole mass range into evenly spaced non-overlapping bins of 5-ppm width (277,259 bins), and subsequently comparing each bin against the reference distribution of maraviroc (m/z 514.3352) using default SSIM weightings. Duplicate hits resulting from the same peak being included in adjacent 5-ppm bins were removed, with only the highest ranked image at a given mass (10-ppm tolerance at m/z 550) kept for analysis.

Results and Discussion

Trichome-Bound Metabolites in A. annua

To find suitable constant parameters for the SSIM algorithm, SSIM evaluation for the *A. annua* sample was performed repeatedly, with the weighting parameters (α, β, γ) varied between 0.5 and 4 individually and pairwise. While changes to the weightings did affect the numerical SSIM score, the rank order was largely unchanged, and so the default weight of 1 to each parameter was used for all data here presented. Similarly, evaluating the SSIM scores with the Gaussian radius parameter varying between 1 and 5 showed only minor effects on the final ranking. It was observed that a small increase in the radius parameter led to significantly lower ranking of images with visible noise, caused either by low absolute ion abundance (shot noise) or significant chemical background noise. We found that a value of 2.25, raised from the default 1.5, resulted in somewhat improved contrast between visually identified “good hits” and “bad hits,” while still assigning high similarity scores to images with moderate levels of chemical noise. These parameters ($\alpha = \beta = \gamma = 1$; radius = 2.25) were used for all subsequent processing.

The pre-selected set of 332 tissue-correlated peaks was evaluated against the reference m/z 281.1395 (artemisinin), and sorted by similarity as shown in Figure 2. The final ranking correctly identified the reference mass itself as a perfect correlation match, with its own first isotope as a close second. Known artemisinin derivatives including deoxyartemisinin (m/z 265.144, rank 13) and dihydroartemisinin (m/z 283.155, rank 22) were identified as visually similar distributions despite large variance in actual ion abundance. The peak set contained 31 ion masses with the characteristic distribution pattern of artemisinin, which were correctly assigned ranks of 1–31.

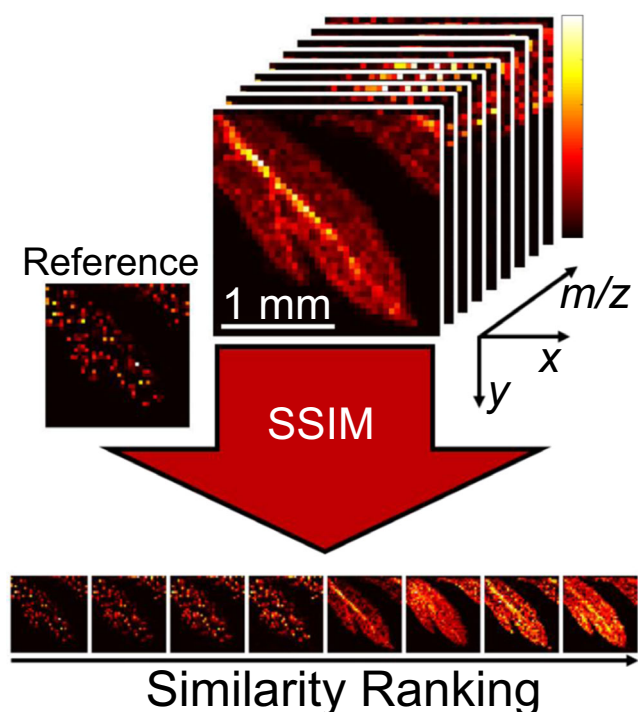


Figure 2. Illustration of the image recognition workflow applied to an MSI dataset acquired from IR-MALDESI imaging of *A. annua* leaf using a known metabolite (artemisinin, m/z 281.1395) as a reference. All images were internally normalized to a 0–1 abundance scale before processing and visualization. Some selected masses taken from the top 100 best matches are shown with similarity rank

Drug Distribution in ARV-Dosed Tissue

For the drug-dosed lymph node section, a comprehensive brute force search was performed, where the SSIM evaluation was performed separately for the mass image of each non-overlapping 5-ppm bin through the whole mass range of m/z 200–800. Performing the evaluation this way required a total of 20 h of computation time. This represents the most thorough search possible with the method, providing a “worst case” example of computation requirements.

The 500 best unique image matches were batch exported and inspected. The top 20 unique matches yielded images of very high visual similarity, with the top 10 almost indistinguishable from the reference. The exported images were all found to visually outline the tissue shape in part or whole, with lower ranked and partial images generally ranking lower. This is illustrated in Figure 3, showing a selection of images throughout the correlation range.

We have found SSIM to be a very robust and useful noise filter for images including some blank or off-tissue data. Caution must however be taken not to include so much blank data as to make the contrast between blank and sample dominate the correlation calculations. For images containing very large regions consisting exclusively of blank scans, or where matching to very localized distributions is desired, we recommend limiting the search to a pre-defined region of interest. Narrowing the search space this way has the additional effect of reducing

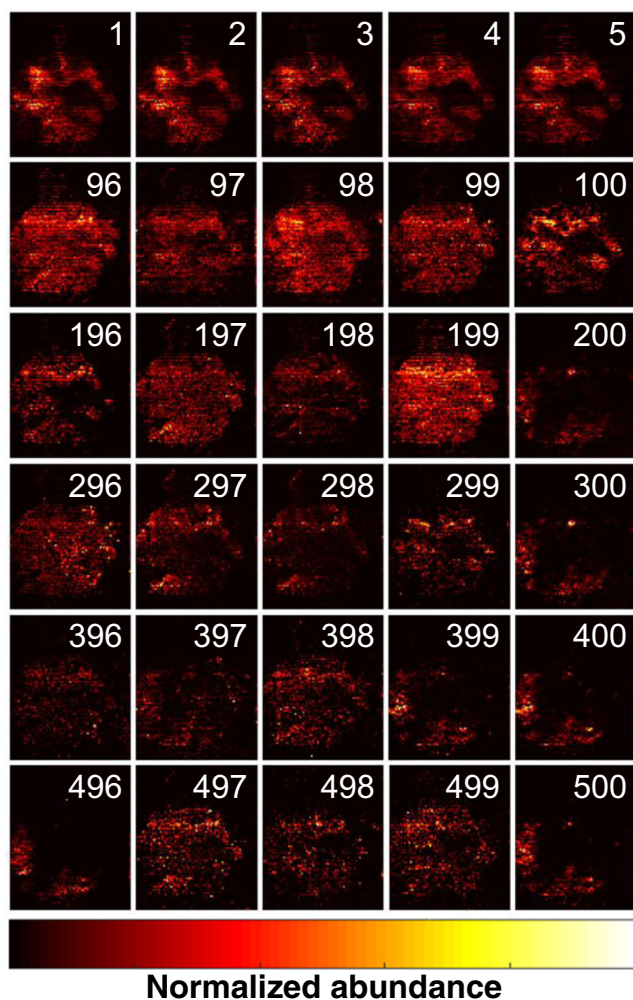


Figure 3. Selected output from a comprehensive SSIM search of drug-dosed lymph node tissue. The images are shown on the normalized 0–1 abundance scale used for SSIM comparisons, with similarity rank shown in white. The image ranked 1 is that of maraviroc (m/z 514.3352), used as a reference

processing times proportionately and can be applied for that purpose alone.

Conclusions

We have here described the implementation and use of an open-source tool using image similarity scoring to extract features of potential interest from high resolving power mass spectrometry imaging datasets. Using the SSIM method for image similarity scoring, the process of semi-targeted discovery can be performed in an automated fashion. Sorting or filtering data by structural similarity effectively reduces complex datasets down to a scale suitable for manual interpretation, and can be used as a reproducible pre-processing step for methods where computation time and memory requirements are limiting factors, e.g., principal component analysis.

All the algorithms used have been incorporated into the latest public release of MSiReader through the batch processing interface. The code is distributed under the BSD 3 license [4] and can be freely adapted to other platforms used for analyzing MSI data.

Acknowledgements

All mass spectrometry measurements were carried out in the Molecular Education, Technology, and Research Innovation Center (METRIC) at NC State University. The authors gratefully acknowledge the financial support received from the National Institutes of Health (R01AI111891, R01GM087964) and North Carolina State University.

Compliance with Ethical Standards

All animal experiments were performed in accordance with locally approved IACUC protocols.

References

1. Wang, Z., Bovik, A.C. Morgan & Claypool Publishers, San Rafael, Calif. (2006)
2. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *Ieee T Image Process.* **13**, 600–612 (2004)
3. Lin, W.S., Kuo, C.C.J.: Perceptual visual quality metrics: a survey. *J Vis Commun Image R.* **22**, 297–312 (2011)
4. Robichaud, G., Garrard, K.P., Barry, J.A., Muddiman, D.C.: MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. *J Am Soc Mass Spectr.* **24**, 718–721 (2013)
5. Bokhart, M.T., Nazari, M., Garrard, K.P., Muddiman, D.C.: MSiReader v1.0: evolving open-source mass spectrometry imaging software for targeted and untargeted analyses. *J. Am. Soc. Mass Spectrom.* (2017)
6. Robichaud, G., Barry, J.A., Muddiman, D.C.: IR-MALDESI mass spectrometry imaging of biological tissue sections using ice as a matrix. *J Am Soc Mass Spectr.* **25**, 319–328 (2014)
7. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., et al.: A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012)
8. Race, A.M., Styles, I.B., Bunch, J.: Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J. Proteome.* **75**, 5111–5112 (2012)
9. Tu, Y.Y.: Artemisinin—a gift from traditional Chinese medicine to the world (Nobel lecture). *Angew Chem Int Edit.* **55**, 10210–10226 (2016)
10. Alejos-Gonzalez, F., Qu, G., Zhou, L.L., Saravitz, C.H., Shurtleff, J.L., Xie, D.Y.: Characterization of development and artemisinin biosynthesis in self-pollinated *Artemisia annua* plants. *Planta.* **234**, 685–697 (2011)
11. Xie, D.Y., Ma, D.M., Judd, R., Jones, A.L.: Artemisinin biosynthesis in *Artemisia annua* and metabolic engineering: questions, challenges, and perspectives. *Phytochem. Rev.* **15**, 1093–1114 (2016)
12. Fletcher, C.V., Staskus, K., Wietgreffe, S.W., Rothenberger, M., Reilly, C., Chipman, J.G., et al.: Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2307–2312 (2014)
13. Thompson, C.G., Bokhart, M.T., Sykes, C., Adamson, L., Fedoriw, Y., Luciw, P.A., et al.: Mass spectrometry imaging reveals heterogeneous efavirenz distribution within putative HIV reservoirs. *Antimicrob. Agents Chemother.* **59**, 2944–2948 (2015)