urn:nbn:de:gbv:ilm1-2011imeko-082:8

Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

Mathematical Methods for Imaging Mass Spectrometry

Dennis Trede[1,2], *Jan Hendrik Kobarg*[2], *Klaus Steinhorst*[1] *and Theodore Alexandrov*[1,2,3]

[1]Steinbeis Innovation Center SCiLS (Scientific Computing in Life Sciences), Bremen, Germany
[2]Center for Industrial Mathematics, University of Bremen, Bremen, Germany
[3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, USA

Abstract – Imaging mass spectrometry (IMS) is an innovative and powerful measurement technology of analytical chemistry which, given a thin sample, is able to reveal its spatial chemical composition in the full molecular range. IMS produces a hyperspectral image, where for each pixel a high-dimensional mass spectrum is measured. A typical data set contains 10^8–10^9 values.

Analysis and interpretation of this huge amount of data is a mathematically, statistically and computationally challenging job. In this extended abstract we present some methods handling with processing IMS data sets.

Keywords: imaging mass spectrometry, hyperspectral imaging, mathematical image processing

1. INTRODUCTION

Mass spectrometry is a method of analytical chemistry to determine the elemental composition of a chemical sample. This task is accomplished through the experimental measurement of the masses of molecules of the sample to be examined.

Given a thin sample (usually a tissue slice), imaging mass spectrometry (IMS) measures high-dimensional mass spectra at its spatial points, providing a hyperspectral image with a mass spectrum measured at each pixel, see Figure 1. Each mass spectrum dimension represents the abundance of molecules with this molecular mass (so-called *m/z* value). Hence it is a natural viewpoint to represent an IMS data set as a hyperspectral image with thousands of channels, as done in other areas of science where multi-channel images are used, e.g. in astronomical hyperspectral imaging, in earth remote sensing, and in life sciences and bio-medicine (i.e. confocal Raman microscopy, near-infrared imaging), see Figure 2.

Currently, IMS is one of the few biochemical technologies able to establish the spatial biochemical composition of the sample in the full molecular range. Since 1970s, secondary ion mass spectrometry (SIMS) was the main IMS technique for surface analysis [1], although being unable to measure large molecules. With the advent of Matrix-assisted laser desorption ionization (MALDI) imaging mass spectrometry [2,3], the measurement of peptides and proteins became possible, which opened a door to the variety of biological and biomedical problems, in particular to detect and discover new biomarkers with a major focus in cancer research [4,5].

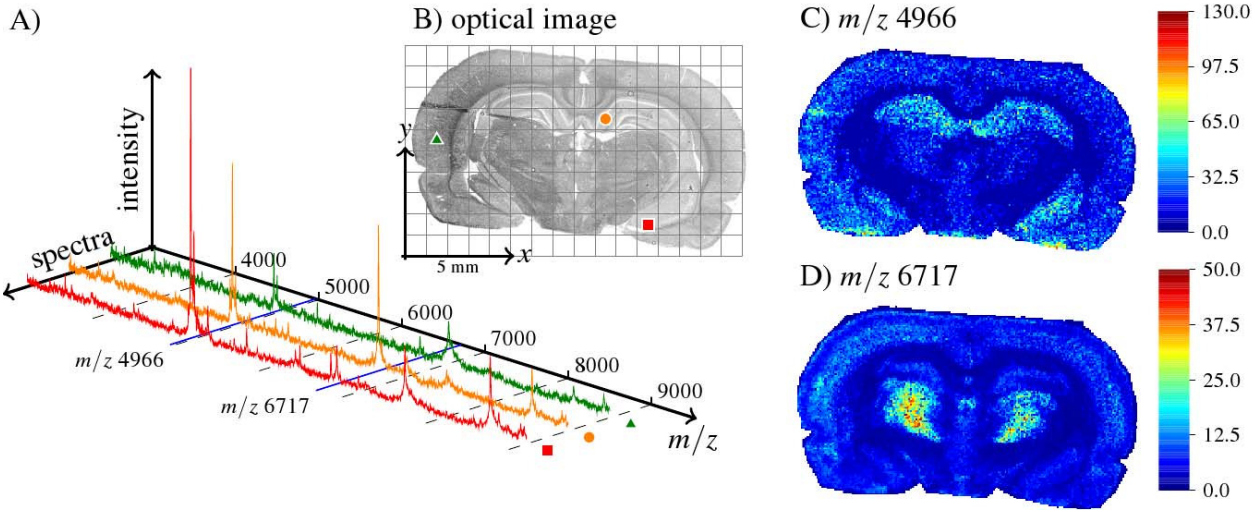

Fig. 1. An IMS dataset is a data cube. Spectra (A) are measured at spatial points of a sample (B) with spatial coordinates (x,y). Given a mass (channel), one obtains an intensity image; examples for the channels *m/z* 4966 and *m/z* 6717 are shown in (C–D).

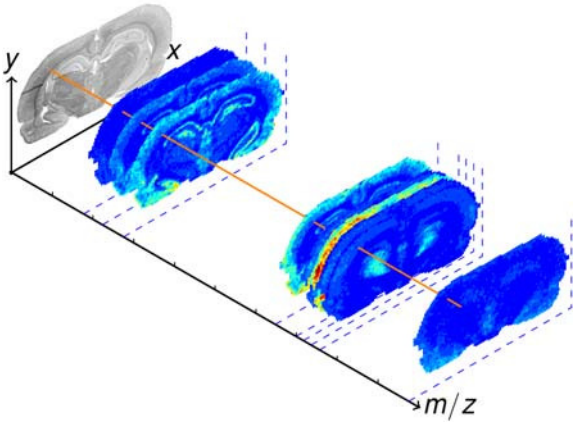

Fig. 2. Representation of an IMS data set as a hyperspectral image reduced to some relevant channels (*m/z* values).

Currently, the development of computational methods for IMS is lagging behind the technological progress [6]. The following computational problems are of interest in IMS data processing:

1. Preprocessing as e.g. normalization of data or noise reduction
2. Data compression in terms of peak picking or scale-space transformations as e.g. the discrete wavelet transform
3. Data representation using multivariate statistics as e.g. principal component analysis (PCA) and its variants
4. Spatial segmentation of an IMS dataset by means of spectra clustering
5. Supervised classification of IMS datasets based on training examples
6. Postprocessing as e.g. super-resolution and image registration

In the following, these problems are discussed in more detail.

2. PREPROCESSING

The IMS data set can be considered as a collection of spectra that have been measured independently. Hence normalization of spectra is an important task of image preprocessing. The most popular method is the so-called total ion count method, which normalizes every spectrum separately so that the intensity scales are identical. In [7], more advanced ways than spectrum-wise normalization are discussed, in particular normalization taking noise quantities of the spectra into account.

The presence of noise in IMS can be easily seen by visual inspection of *m/z*-images corresponding to some selected channels, see Figure 1. Since the noise in IMS is strong, another important preprocessing step is denoising. A fact to incorporate is that the noise variance changes both within an image and between different images. In [8], it has been shown that the noise variance at a spatial point linearly depends on the mean intensity around this point. This indicates that the noise is Poisson distributed. To treat that, in [8] a method for edge-preserving image denoising has been introduced that adjusts the level of denoising to the local noise level and to the local scale of the features to be resolved.

3. DATA COMPRESSION

The IMS data typically consists of thousands of different channels (10^3–10^4) which have to be evaluated statistically. To process the huge amount of data one could constrain the channels to the most relevant without losing significant information. In IMS, this process is called peak picking. Naturally, for processing huge IMS data sets we need an efficient peak picking method. At the same time, peak picking should be robust to strong noise, preventing the use of too simple local maxima or signal-to-noise ratio methods, which produce too many false positives. In [9], a peak picking method based on the orthogonal matching pursuit (OMP) is proposed and in [8] this method is applied to real-life mass spectrometry data. The main idea is to model each spectrum as a sum of Gaussian-shaped functions.

Alternative methods for reducing the amount of data for a later feature selection and classification are scale space methods as e.g. the discrete wavelet transform [10]. For the wavelet transform the idea is to use wavelet for which its scaling function closely matches the peak pattern of spectra, as e.g. the bi-orthogonal bior3.7 wavelet in [11].

4. DATA REPRESENTATION

Data mining of IMS data sets is currently a very time-consuming endeavor as it is mostly done manually and an IMS data set consists of thousands of channels for a single sample. Currently, complete mining of such data requires the user to click through each image and look for distributions that may correlate to the morphology of the sample analyzed. Unsupervised processing methods, which do not rely on labeling of data set elements, allow for automated extraction of data from a data set.

Principal component analysis (PCA) [12] and its variants [13,14] are typical unsupervised multivariate methods where data is statistically represented in fewer dimensions. The idea is to decompose the IMS data into its underlying trends and thus transform the data set into a small set of images showing main spatial features, as shown in Figure 3. The results of multivariate methods can be visualized as is or can be used in combination with other techniques (e.g. classification of PCA coordinates).

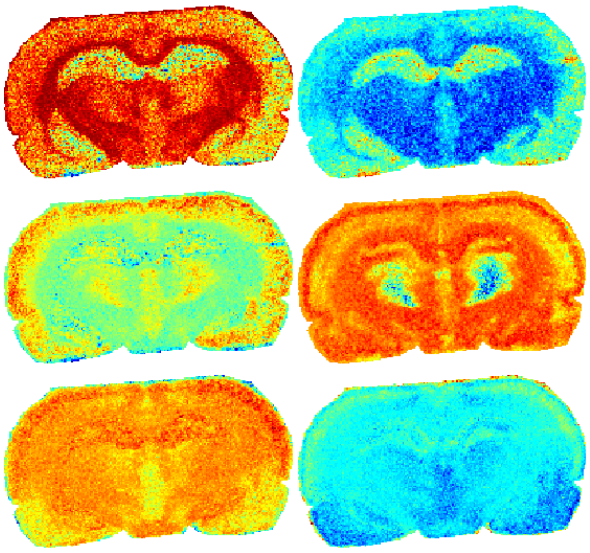

Fig. 3. The first six PCA abundance maps of an IMS data set.

5. SPATIAL SEGMENTATION

Another unsupervised method is spatial segmentation of a data set by clustering of spectra [15]. The results of clustering can be displayed as a spatial segmentation map (an integer-valued image, usually shown using pseudo-color), coloring identically points grouped into one cluster.

The main drawback of using straightforward clustering of mass spectra is that it is negatively affected by the pixel-to-pixel variability. Taking into account the spatial relations between spectra improves the segmentation maps considerably by suppressing the noise and pixel-to-pixel variability [8], see Figure 4. Most of the sophisticated clustering methods are computationally intensive due to slow high dimensional clustering. Use of simpler methods reduces the computation time but worsen the segmentation maps due to strong noise in data. In [16], an approach for segmentation of hyperspectral data has been proposed, that gets efficient due to a projection to fewer dimensions at the same time considering a spectrum together with its neighbors.

6. SUPERVISED CLASSIFICATION

Supervised methods are used for classification of spectra into several groups (i.e. 'abnormal' versus 'healthy' tissue) [11]. Typically, a region within a single sample is manually designated as having one histological state (i.e. tumor tissue) while the remainder of the sample is classified as another histological state (i.e. healthy tissue). Classification methods are e.g. being developed for biomarker detection since once the classification is performed and evaluated to be successful, one can find discriminative masses [5].

7. POSTPROCESSING

An important issue for any IMS technology is its relatively low spatial or lateral resolution (i.e. a large size of a pixel) as compared with microscopy. The state of the art resolution is around 20 microns for MALDI-imaging [17] versus maximum 0.25 microns for optical microscopy. So, when comparing an IMS data set or its segmentation map with a microscopy image, a significant difference in spatial resolution complicates the visual interpretation. In [18], a computational approach is proposed to improve the spatial resolution of a segmentation map of an IMS data set. In Figure 5, the a super-resolution image of a synthetic segmentation map is displayed.

Other imaging problems occur when extending the 2D IMS technique to three spatial dimensions with consecutive sections of tissue. Here one has to align a stack of hyperspectral images to each other. Methods for image registration of grey-scale images are available [19], but–to the best of our knowledge–not yet developed for 3D hyperspectral IMS data.

From a technical perspective, visualizing this 3D information is highly complex. From a medical perspective however, it still does not provide enough information for diagnosis. To draw conclusions from the data, it must first be correlated with 3D anatomical information (such as data obtained via magnetic resonance imaging). However, superimposing these two data sets originating from entirely different imaging modalities is complicated by the issue of image co-registration [20] and standard pipelines are not established, yet.

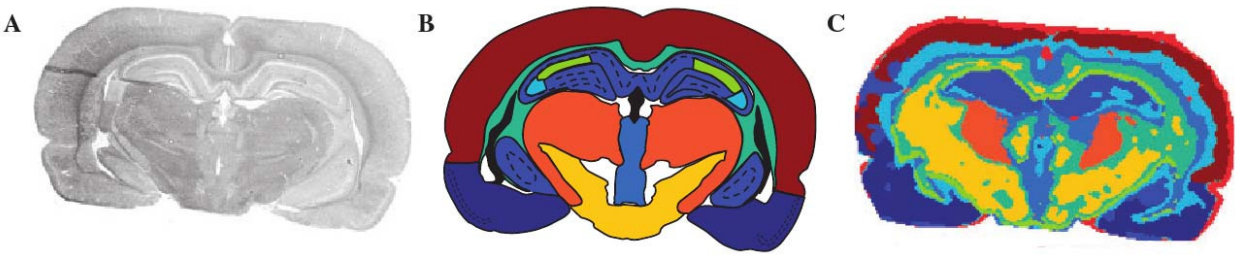

Fig. 4. Example of unsupervised spatial segmentation, from [16]. (A) Optical image. (B) Schematic representation based on the rat brain atlas. (C) Segmentation map with 10 clusters.

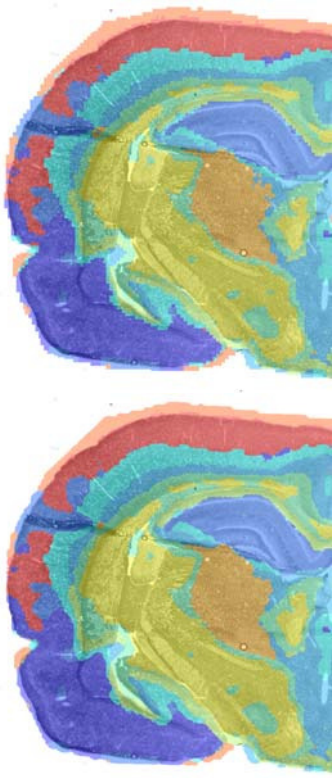

Fig. 5. Example of a super-resolution segmentation map co-registered with a microcopy image. Top: original segmentation map with 10 clusters. Bottom: its super-resolution version (magnification factor 10).

8. CONCLUSION

Data from imaging mass spectrometry can be represented as a hyperspectral image with thousands of channels. Since manual data mining of IMS data sets is very time-consuming, the development of computational methods is necessary. Mathematics offers a huge collection of methods from image processing, statistics and machine learning that can be used for simplifying and automating the analysis of IMS data.

Other areas of science where hyperspectral images incur use similar methods for related problems. Here, an interdisciplinary exchange of experiences can inspire each other and avoid gratuitous parallel developments.

ACKNOWLEDGEMENT

We gratefully thank Herbert Thiele and Michael Becker (both Bruker Daltonik) for providing the rat brain dataset and additionally Michael for drawing the schematic of the rat brain.

Klaus Steinhorst and Dennis Trede gratefully acknowledge the financial support of the Bremen Economic Development (WFB, project "3D MALDI-Imaging basic experiment", grant FUE0485B). Jan Hendrik Kobarg and Dennis Trede gratefully acknowledge the financial support of the European Union Seventh Framework Programme (project "UNLocX", grant 255931).

REFERENCES

[1] A. Benninghoven, E. Loebach, "Tandem mass spectrometer for secondary ion studies", *Review of Scientific Instruments*, **42**, 49–52, 1971.

[2] R.M. Caprioli, T.B. Farmer, J. Gile, "Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS", *Analytical Chemistry,* **69**, 4751–4760, 1997.

[3] M. Stoeckli, P. Chaurand, D.E. Hallahan, R.M. Caprioli, "Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues", *Nature Medicine*, **7**, 493–496, 2001.

[4] R.M.A. Heeren, D.F. Smith, J. Stauber, B. Kukrer-Kaletas, L. MacAleese, "Imaging mass spectrometry: hype or hope", *Journal of the American Society of Mass Spectrometry*, **20**(6), 1006–1014, 2009.

[5] S. Rauser, C. Marquardt, B. Balluff, S.-O. Deininger, C. Albers, E. Belau, R. Hartmer, D. Suckau, K. Specht, M.P. Ebert, M. Schmitt, M. Aubele, H. Höfler, A. Walch, "Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry", *Journal of Proteome Research*, **9**(4), 1854–1863, 2010.

[6] J.D. Watrous, T. Alexandrov, P.C. Dorrestein, "The evolving field of imaging mass spectrometry and its impact on future biological research", *Journal of Mass Spectrometry*, **46**(2), 209–222, 2011.

[7] S.-O. Deininger, D.S. Cornett, R. Paape, M. Becker, C. Pineau, S. Rauser, A. Walch, E. Wolski, "Normalization in MALDI-TOF imaging datasets of proteins: practical considerations", *Analytical Bioanalytical Chemistry*, **401**(1):167–181, 2011.

[8] T. Alexandrov, M. Becker, S.-O. Deininger, G. Ernst, L. Wehder, M. Grasmair, F. von Eggeling, H. Thiele, P. Maass, "Spatial Segmentation of Imaging Mass Spectrometry Data with Edge-Preserving Image Denoising and Clustering", *Journal of Proteome Research*, **9**(12), 6535–6546, 2010.

[9] L. Denis, D.A. Lorenz, D. Trede, "Greedy solution of ill-posed problems: error bounds and exact inversion", *Inverse Problems*, **25**(11), 115017, 2009.

[10] F.-M. Schleif, M. Lindemann, M. Diaz, P. Maass, J. Decker, T. Elssner, M. Kuhn, H. Thiele, "Support vector classification of proteomic profile spectra based on feature extraction with the bi-orthogonal discrete wavelet transform", *Computing and Visualization in Science*, **12**(4):189–199, 2009.

[11] T. Alexandrov, J. Decker, B. Mertens, A.M. Deelder, R.A.E.M. Tollenaar, P. Maass., H. Thiele, "Biomarker discovery in MALDI-TOF serum protein profiles using discrete wavelet transformation", *Bioinformatics*, **25**(5):643–649, 2009.

[12] R. Van de Plas, B. De Moor, E. Waelkens, "Imaging Mass Spectrometry Based Exploration of Biochemical Tissue Composition using Peak Intensity Weighted PCA", In: *Proc. of the Third IEEE/NIH BISTI Life Science Systems and Applications Workshop 2007*, 209–212, 2007.

[13] M. Hanselmann, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A. Heeren, F.A. Hamprecht, "Concise Representation of Mass Spectrometry Images by Probabilistic Latent Semantic Analysis", *Analytical Chemistry*, **80**(24):9649-9658, 2008.

[14] J.H. Kobarg, T. Alexandrov, "Algorithms to incorporate spatial information into clustering of hyper-spectral data", *unpublished*, 2011.
[15] S.-O. Deininger, M.P. Ebert, A. Fütterer, M. Gerhard, C. Röcken, "MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers", *Journal of Proteome Research,* **7**(12), 5230–5236, 2008.
[16] T. Alexandrov, J.H. Kobarg, "Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering", *Bioinformatics,* **27**(13), i230–i238, 2011.
[17] M. Lagarrigue, M. Becker, R. Lavigne, S.-O. Deininger, A. Walch, F. Aubry, D. Suckau, C. Pineau, "Revisiting rat spermatogenesis with MALDI imaging at 20 µm resolution", *Molecular and Cellular Proteomics,* mcp.M110.005991, 2010
[18] T. Alexandrov, S. Meding, D. Trede, J.H. Kobarg, B. Balluff, A. Walch, H. Thiele, P. Maass, "Super-resolution segmentation of imaging mass spectrometry data: solving the issue of low lateral resolution", *unpublished*, 2011.
[19] O. Schmitt, J. Modersitzki, S. Heldmann, S. Wirtz, B. Fischer, "Image registration of sectioned brains", *International Journal of Computer Vision*, **73**(1), 5–39, 2007.
[20] T.K. Sinha, S. Khatib-Shahidi, T.E. Yankeelov, K. Mapara, M. Ehtesham, D.S. Cornett, B.M.Dawant, R.M. Caprioli, J.C. Gore, "Integrating spatially resolved three-dimensional MALDI IMS with in vivo magnetic resonance imaging", *Nature Methods,* **5**, 57–59, 2008.

Corresponding Author: Dennis Trede is the author to whom correspondence shall be addressed. He is with Zentrum für Technomathematik, Universität Bremen, Bibliothekstrasse 1, D-28359 Bremen, Germany, and with Steinbeis Innovation Center SCiLS, Richard-Dehmel-Strasse 69, D-28359 Bremen, Germany, Tel:+49 421 218-63823, Fax:+49 421 218-98 63823, trede@scils.de.