Discrete Wavelet Transform-based Multivariate Exploration of Tissue via Imaging Mass Spectrometry

Raf Van de Plas*
Katholieke Universiteit Leuven
Dept. of Electrical Eng. ESAT
SCD-SISTA (BIOI)
Kasteelpark Arenberg 10
B-3001 Leuven (Heverlee)
Belgium

Bart De Moor
Katholieke Universiteit Leuven
Dept. of Electrical Eng. ESAT
SCD-SISTA
Kasteelpark Arenberg 10
B-3001 Leuven (Heverlee)
Belgium

Etienne Waelkens
Katholieke Universiteit Leuven
Dept. of Mol. Cell Biology
Section Biochemistry
O & N, Herestraat 49 - bus 901
B-3000 Leuven
Belgium

This paper gives a short overview of work described in a technical report by the same authors [3], which is available upon request.

ABSTRACT

Mass spectral imaging (MSI) or imaging mass spectrometry is a developing technology that combines spatial information with traditional mass spectrometry. It enables researchers to study the spatial distribution of biomolecules such as proteins, peptides, and metabolites throughout organic tissue sections. MSI has particular merit in exploratory settings where there is no prior hypothesis of relevant target molecules. It is rapidly becoming a potent exploratory instrument for tissue biomarker studies.

MSI is a high-throughput technique that mines massive amounts of measurements from a single tissue section. As various parameters such as the covered tissue surface area, the spatial resolution, and the extent of the mass range grow, MSI data sets rapidly become very large, making analysis from a computational and memory standpoint increasingly difficult. In this paper we introduce the discrete wavelet transform (DWT) as a means of reducing the dimensionality of the data,while retaining a maximum amount of biochemical information. The DWT delivers a more compact description of each mass spectrum, expressed as wavelet coefficients. The efficacy of performing analyses directly in the DWT-reduced space is illustrated using unsupervised trend detection via principal component analysis (PCA) on the MSI measurement of a sagittal section of mouse brain.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Biology and genetics; G.1.2 [**Numerical Analysis**]: Approximation—*Wavelets and fractals*; I.5.2 [**Pattern Recognition**]: Design Methodology—*Pattern Analysis*; I.4.2 [**Image Processing and Computer Vision**]: Compression (Coding)—*Approximate methods*

*Corresponding author: raf.vandeplas@esat.kuleuven.be

SAC'08 March 16-20, 2008, Fortaleza, Ceará, Brazil

Keywords

bioinformatics, proteomics, imaging, mass spectrometry, discrete wavelet transform, principal component analysis

1. INTRODUCTION

MALDI-based imaging mass spectrometry (MSI) [2] preserves the link between a spatial tissue location and a biochemical characterisation of what was found there. It uses the molecular specificity and sensitivity of normal mass spectrometry to collect a spatial mapping of biomolecules (or rather their ions) from a tissue section. Meistermann *et al.* [1] give an example of its use in biomarker discovery. Fig. 1 shows an overview of an MSI experiment. A more thorough treatment is available from Stoeckli *et al.* [2] and Van de Plas *et al.* [4].

The result of an MSI experiment consists of a grid of measurement locations or 'pixels' covering the tissue section, with an individual mass spectrum connected to each pixel. The data structure can be considered as a three-mode array with two spatial modes (x and y) and one mass-over-charge mode (m/z). In this paper, principal component analysis (PCA) is used to decompose high-dimensional MSI data into a reduced set of uncorrelated biochemical tissue trends [4].

The size of an MSI data set is primarily influenced by two parameters: the number of measurement locations or pixels and the number of scanned m/z-bins. The pixel number increases as the tissue surface area that needs to be covered grows larger or the spatial resolution becomes higher. The number of m/z-bins is proportional to the extent of the mass range and the granularity of the mass resolution. As this technology develops, both parameters get pushed upward and data sets appear that stretch the computational and memory resources generally available or challenge the scalability of the algorithms that operate on them. Due to its multiway nature, MSI has, more so than standard mass spectrometry, a need for strong dimensionality reduction methods with minimal loss of biochemical information.

2. METHODS

In response to the issues raised in section 1, this paper introduces the discrete wavelet transform (DWT) as a method for reducing high-dimensional MSI data into a lesser-dimensional space with most relevant information left intact. The capability for operating analyses directly in the reduced space is demonstrated using trend detection via PCA [4] with satisfactory results. The general method is explained in [3] and Fig. 4, while section 3 demonstrates the procedure on a mouse brain case study.

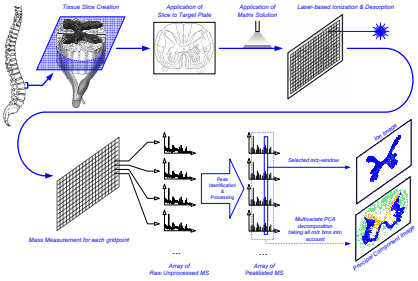

Figure 1: Overview of an MSI experiment on spinal cord. (wet-lab) A tissue section is cut using a microtome, mounted on a target plate, and covered with an appropriate chemical matrix to enable ionization. (mass spec) Individual mass spectra are collected from the tissue area of interest, while their spatial relationships are retained. (in silico) The data is collected into a three-mode array for analysis.

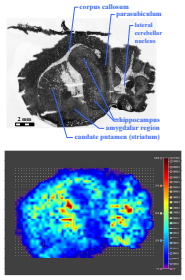

Figure 2: (top) Picture of the sagittal mouse brain section imaged in section 3. (bottom) Ion image showing the presence of m/z 14148 primarily in the corpus callosum and the lateral cerebellar nucleus.

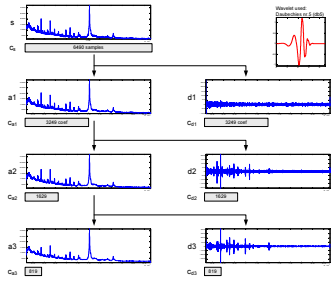

Figure 3: A multiple-level DWT performed on a mass spectrum from the case study of section 3. The chosen wavelet is db5 (top-right) and the decomposition tree has a depth of three levels. The profile of the original mass spectrum s is well preserved in approximations $a1$ to $a3$, while the required coefficients describing the waveform have been reduced from 6490 to 819.

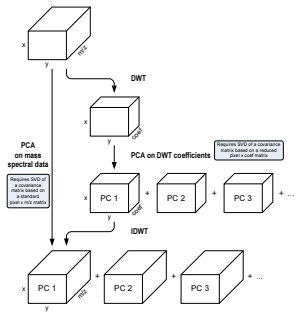

Figure 4: Overview of the DWT-based PCA decomposition of MSI data. DWT is used to reduce the feature space, and PCA is performed in the reduced space.

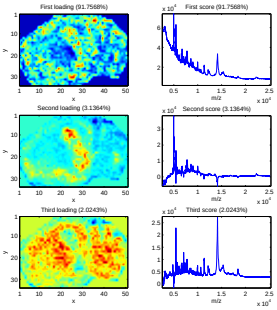

Figure 5: Non-DWT-based principal components. These are the three most important PCs found via direct application of PCA. They chemically delineate a number of known anatomical zones in the mouse brain (e.g. *corpus callosum*).

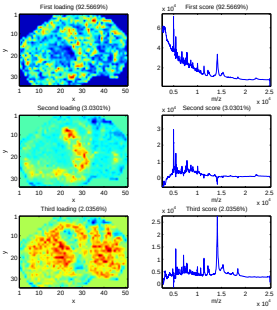

Figure 6: DWT-based principal components. These are the three primary PCs found via the method of Fig. 4. Although PCA was applied in the lower-dimensional coefficients-space, it yields almost identical results (see Fig. 5).

3. CASE STUDY

To establish an empirical test case, this section applies both the DWT-based method as well as direct PCA to the MSI measurement of a mouse brain section. The goal is an assessment on a real example of how well the information in the data set is retained, given a serious reduction in dimensionality. The example also demonstrates whether results from factor analysis methods such as PCA that are performed in the coefficients-space, approximate or retain relevance to results directly from the original m/z-space.

4. CONCLUSION

The excellent compression characteristics of the DWT provide us with a means of dimensionality reduction that retains the mass spectral information to a high degree while offering significant reductions in resource requirements. In this paper we have shown specifically that resource-hungry operations, such as biochemical trend detection via PCA, can be performed directly on the compacter description of the data with meaningful biochemical results that deviate little from non-DWT-reduced results. The approach has calculation time reducing, noise removing, and memory requirements dimishing aspects that can be useful in any MSI context. Additionally, it holds promise as a tool to make multivariate exploration feasible on very large MSI data sets, that might otherwise surpass available resources.

Acknowledgements

We kindly acknowledge Dagmar Niemeyer and Sören-Oliver Deininger from Bruker Daltonics in Bremen, Germany.

RVDP is a research assistant of the IWT at the K.U.Leuven, Belgium. BDM is a full professor at the K.U.Leuven, Belgium. EW is a full professor at the K.U.Leuven, Belgium. Additionally, RVDP, BDM, and EW are affiliated with the Interfaculty Centre for Proteomics and Metabolomics, ProMeta at the K.U.Leuven.

Research supported by Research Council KUL: GOA AMBioRICS, SymBioSys CoE EF/05/007, several PhD/postdoc & fellow grants; Flemish Government: - FWO: PhD/postdoc grants, projects G.0241.04, G.0499.04, G.0232.05, G.0318.05, G. 0553.06, G.0302.07, research communities (ICCoS, ANMMM, MLDM); - IWT: PhD Grants, GBOU-McKnow-E, GBOU-ANA, TAD-BioScope-IT, Silicos; SBO–BioFrame; Belgian Federal Science Policy Office: IUAP P6/25 & P6/28; EU-RTD: ERNSI; FP6-NoE; FP6-IP, FP6-MC-EST, FP6-STREP, ProMeta, BioMacS.

5. REFERENCES

[1] H. Meistermann *et al.*, Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat, *Mol Cell Proteomics*, 5:10, 2006, pp 1876–1886.

[2] M. Stoeckli *et al.*, Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues, *Nat Med*, 7:4, 2001, pp 493–496.

[3] R. Van de Plas *et al.*, "Discrete Wavelet Transform-based Multivariate Exploration of Tissue via Imaging Mass Spectrometry," *Internal Report, ESAT-SISTA, K.U.Leuven (Leuven, Belgium)*, 2007.

[4] R. Van de Plas *et al.*, "Prospective Exploration of Biochemical Tissue Composition via Imaging Mass Spectrometry Guided by Principal Component Analysis," *in Proceedings of the Pacific Symposium on Biocomputing 12*, Maui, HI, 2007, pp. 458-469.