esaPROJECT: A PRELUDE TO DATA FUSION FOR X-RAY AND OTHER SPECTROSCOPIES

M. G. Dowsett[1] and A. Adriaens[2]

[1] EVA Surface Analysis, markdowsett.esa@gmail.com and Department of Physics, University of Warwick, Coventry, CV4 7AL, UK
[2] Department of Analytical Chemistry, Krijgslaan 281 S12, Ghent University, B9000, Belgium

Rationale

Like most synchrotron users we visit a variety of beam lines each with its own idiosyncratic data system and file format. These vary from the binary pseudo-TIFF structure found on CCD cameras to chaotic ad-hoc text formats like SPEC log files. Our data sets often contain 1000s of images, patterns or spectra as stacks or sequences, could be XRD, XAS or XEOL, and we naturally like to compare these with data from laboratory instruments. There was no single package which could:

- read all the raw formats
- extract patterns from diffractograms (especially when the detector is *not* on the beam axis)
- provide instantaneous comparisons (e.g. with references)
- manage different data types (diffraction, absorption, ... SIMS!)
- preprocess the data for other packages (e.g. IFEFFIT)
- extract trends
- remove background
- heal damaged data sets
- mathematically manipulate images, patterns and spectra
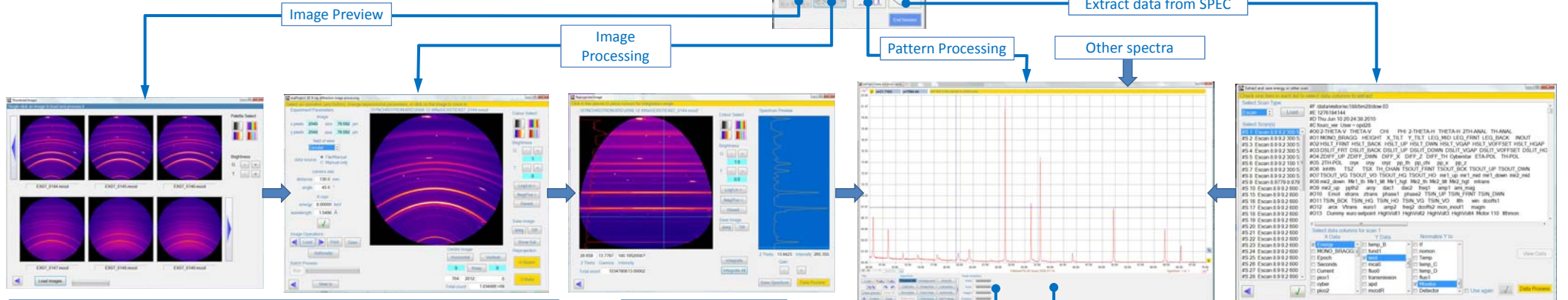- intensively batch process images, patterns and spectra

Moreover, the sheer tedium of subsequently using Origin, Gnu Plot or Sigmaplot to produce something for publication was overwhelming. We set out to fill this chasm with software of our own ... *esaProject**

Overview

Because of the mix of specialities of our coworkers, esaProject was designed for use by experts and novices alike. User input to the program is monitored by a contextually sensitive set of objects which ignore, inhibit, or warn of invalid text or operational demands.

For the most efficient combination of developability, stability and usability, esaProject is

- written in Microsoft Visual Studio languages
- runs under Windows XP onward
- takes advantage of the .NET environment

The software is structured in 7 main linked blocks each with its own GUI:

Image Preview

Image Processing

Extract data from SPEC

Pattern Processing

Other spectra

Pattern extraction

Patterns/spectra out

Stack/ sequence console

Publish

File Formats

esaProject can read anything where we can get or unpick the file format.
So far:

Images
- Mar CCD formats (pseudoTiff, big or little endian, FITS)
- Some Quantum formats
- esaProject binary format
- cbf (Pilatus compressed binary) and edf**

Patterns/Spectra
- Rapid II binary file sets (big and little endian)
- Generic one and two column text files (with or without text headers)
- esaProject text (human readable) format
- Brucker Diffrac raw files
- Panalytical *.xrdml files (subset – depends on experiment)
- SRS XANES and EXAFS fluorescence and transmission format (allows user to pick segments to include)
- Diamond EXAFS text file
- Diamond RGB text file

In most cases esaProject also automatically reads beam monitor data from separate files (e.g. SPEC files and Rapid 2 files) and configures it for point by point or whole data set normalization.

*The esaProject source code is ©2006-2015 Mark Dowsett, EVA Surface Analysis, UK.
**the edf format is an unstandardized mess and esaProject sometimes needs a tweak to read certain file sets

Pattern extraction

Diffraction images are transformed or "reprojected" into a space where the rings are straight lines. The transformation conserves intensity whilst incorporating an anti-aliasing strategy to reduce pixel sampling oscillations in the final spectrum. It employs an algorithm with a saturating series for the inverse cosine function so that it is immune to rounding errors which are problematic in that they result in an invalid computation

$$A = arcos(1 + \Delta) \text{ where } 0 < \Delta \ll 1$$

in some regions of the map.

Batch Processing

esaProject batch processes images heuristically:
Walk one through the processing steps (centring, reprojection, integration, saving images (for presentation) and patterns then Batch process as many as you like the same way.

Because pattern/spectrum operations are very fast and often require user judgement, esaProject batch processes spectra in parallel operations through a "Process All" feature

- TIFF
- wmf
- pdf (requires Acrobat)
- Printer

- Presentation
- Report
- Paper