Comparative Visualization for Comprehensive Two-Dimensional Gas Chromatography

Benjamin V. Hollingsworth[1] Stephen E. Reichenbach[2]

[1]GC Image, LLC, Lincoln NE

[2]Computer Science & Engineering Dept., University of Nebraska-Lincoln

GC Image

Software for comprehensive two-dimensional gas chromatography

University of Nebraska Lincoln

Goal: Comparative analyses with GCxGC

Approach: Extend image comparison methods for GCxGC

Processing: Registration and normalization

Comparison: Image-based methods and tools

Conclusions and future work

Comparative analyses with GCxGC

GCxGC is a powerful technology for chemical separations, providing significantly greater in separation capacity, higher-dimensional chemical ordering, and improved SNR.

Applications of GCxGC comparative analyses:

- Compare manufactured products with standards for quality control [Shellie et al., 2000].
- Monitor actual or potential pollution sites for environmental changes [Reddy, et al., 2002].
- Survey crime scenes for chemical "fingerprints" [Frysinger and Gaines, 2002].
- Assay classes of tissue samples for biomarker discovery [Welthagen et al., 2005].

Challenges for GCxGC comparisons

Challenges for comparative analyses with GCxGC:

- Data inconsistencies.
  - Sample amount variations.
  - Retention time variations.
  - Peak shape variations.
- Data and task complexities.
  - Multi-dimensional data.
  - Rich data (i.e., many chemical peaks).
  - Various dimensions of comparison, e.g., absolute differences, relative differences, etc.
- Software.
  - Software research and development.
  - Software training for analysts.

Comparison methods for digital images

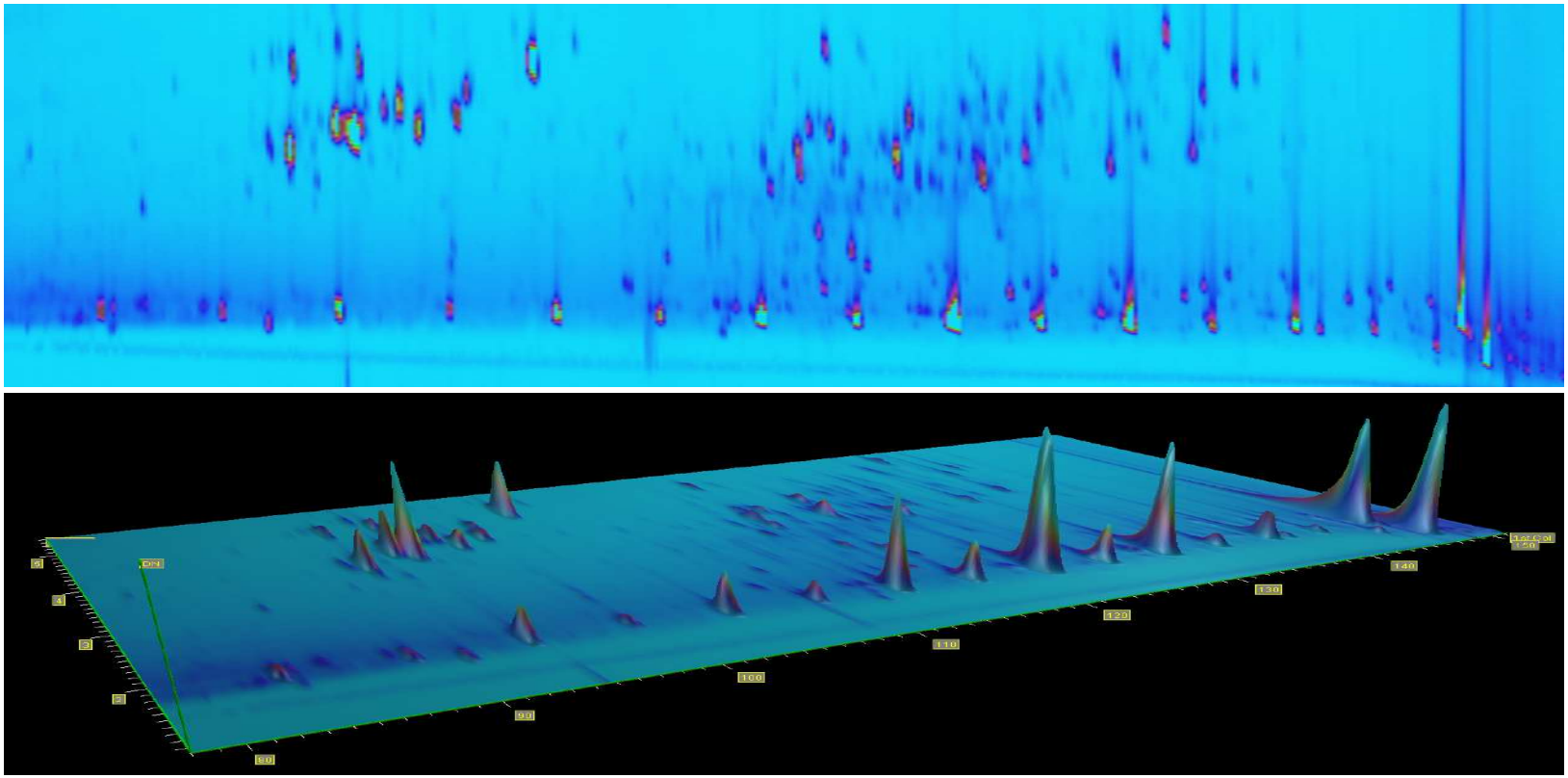
GCxGC data can be represented as a digital image with pseudo-coloring.

Traditional image comparison methods:

- Side-by-side comparisons.
- Flicker (cycle) between images.
- Difference image.
- Color addition (e.g., 2–3 datasets in color channels).

A GCxGC image can be interpreted as an elevation map, forming a 3D surface which can be projected to 2D.

GCxGC data viewed as 2D or 3D image



Reddy et al., Environ. Sci. & Tech., vol. 36, 2002.

Methods for comparing GCxGC images

Preprocess images to remove background and to detect, quantify, and identify peaks.

Register images using retention times of selected peaks to correct for retention time variations.

Normalize intensities using total responses (volumes) of selected peak(s) to correct for sample amount variations.

Use *colorized difference* to indicate both differences and underlying intensities.

Use local, *fuzzy difference* to correct for slight misregistration and peak shape variations.

Use tools for masking, tabular data, and 3D viewing to enhance analysis.

Registration

Registration aligns peak retention times between an *analyzed image* and a *reference image* for selected peaks (registration peak set) present in both images.

- Find affine transformation that minimizes mean-square error between transformed reference peak retention times and analyzed peak retention times.
- Eliminate (from registration peak set) 25% of peaks with largest misregistration after transformation.
- Recompute optimal affine transformation.
- Apply affine transformation to all pixels of reference image.

Example uses *n*-alkanes, solvent peak, and polar natural compound.

Normalization

Normalization scales for equivalent responses in selected peaks (quantitative peak set) present in both analyzed and reference images.

- Determine scale factor for summed responses (volumes) of analyzed peaks to reference peaks.
- Eliminate (from quantitative peak set) 25% of peaks with greatest difference after scaling reference peak responses.
- Recompute scale factor.
- Apply scale to all pixels of reference image.

Example uses *n*-alkanes.

Grayscale difference

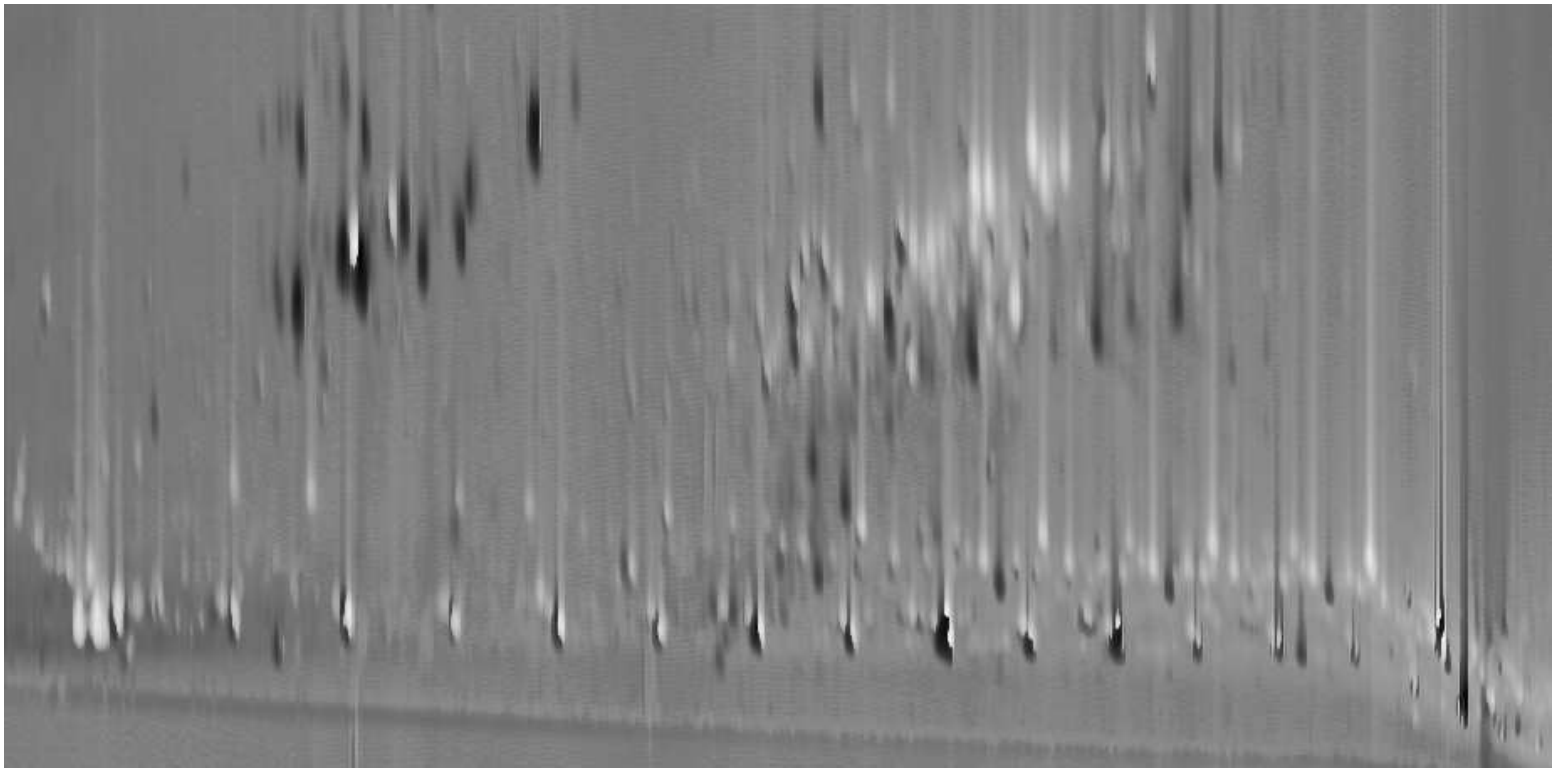
Pixel-by-pixel subtraction of the *analyzed image* minus the *reference image*.

Grayscale colorization, with positive values gray to white and negative values gray to black. Use same scale for both positive and negative values. Logarithmic scale can enhance visual differences.

Two problems:

- Grayscale shows only differences, not underlying peak intensities.
- Slight misregistration or peak shape differences cause adjacent dark and bright regions for the same peak

Grayscale difference



Grayscale difference.

Colorized difference

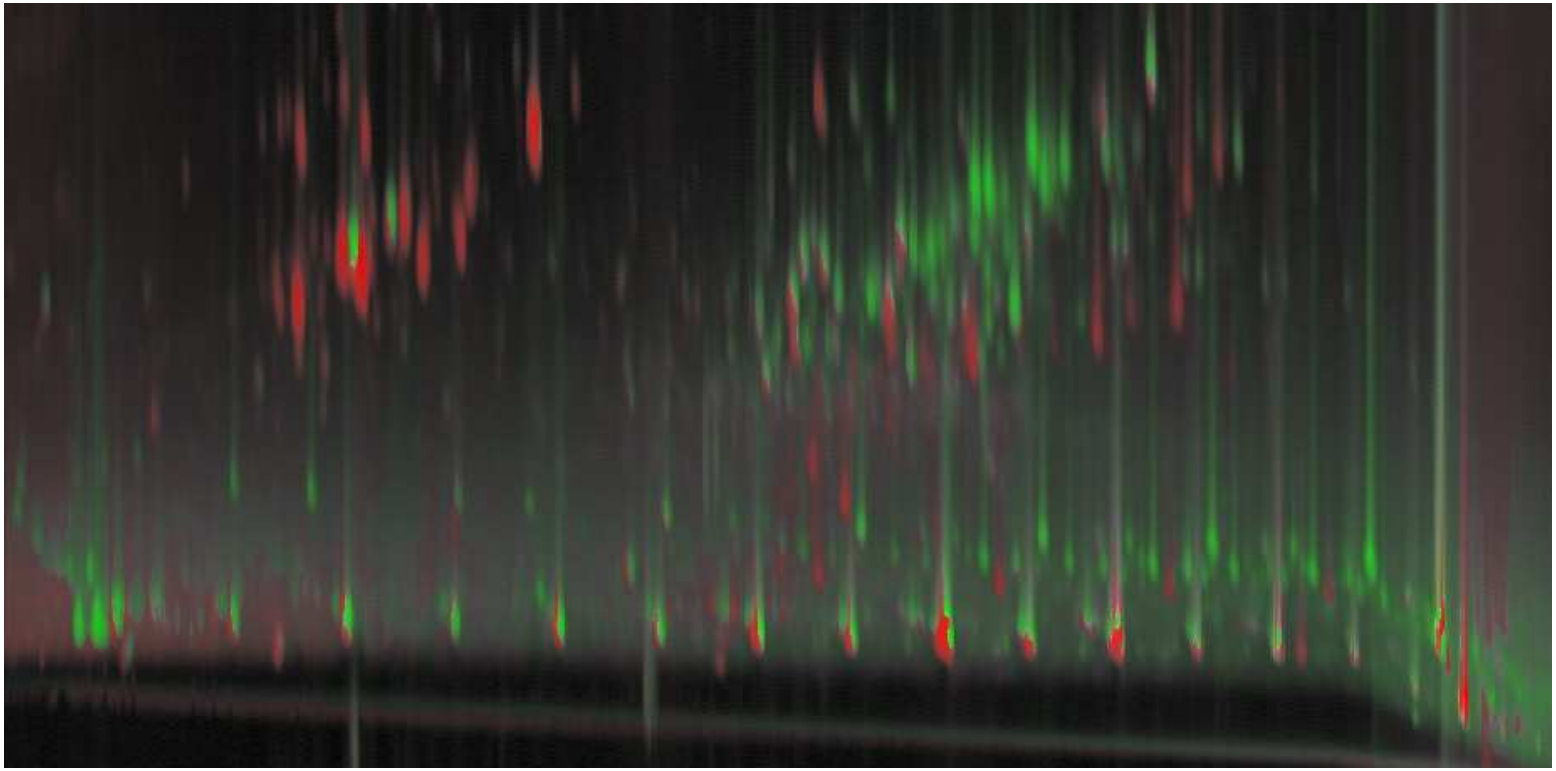
Set pixel pseudo-color:

- Hue to green for positive difference or red if negative difference
- Intensity to maximum of analyzed or reference image (scaled 0 to 1).
- Saturation to magnitude of difference (scaled 0 to 1).

Shows both difference and underlying intensity.

Adjacent red and green regions for the same peak.

Colorized difference



Colorized difference.

Grayscale fuzzy difference

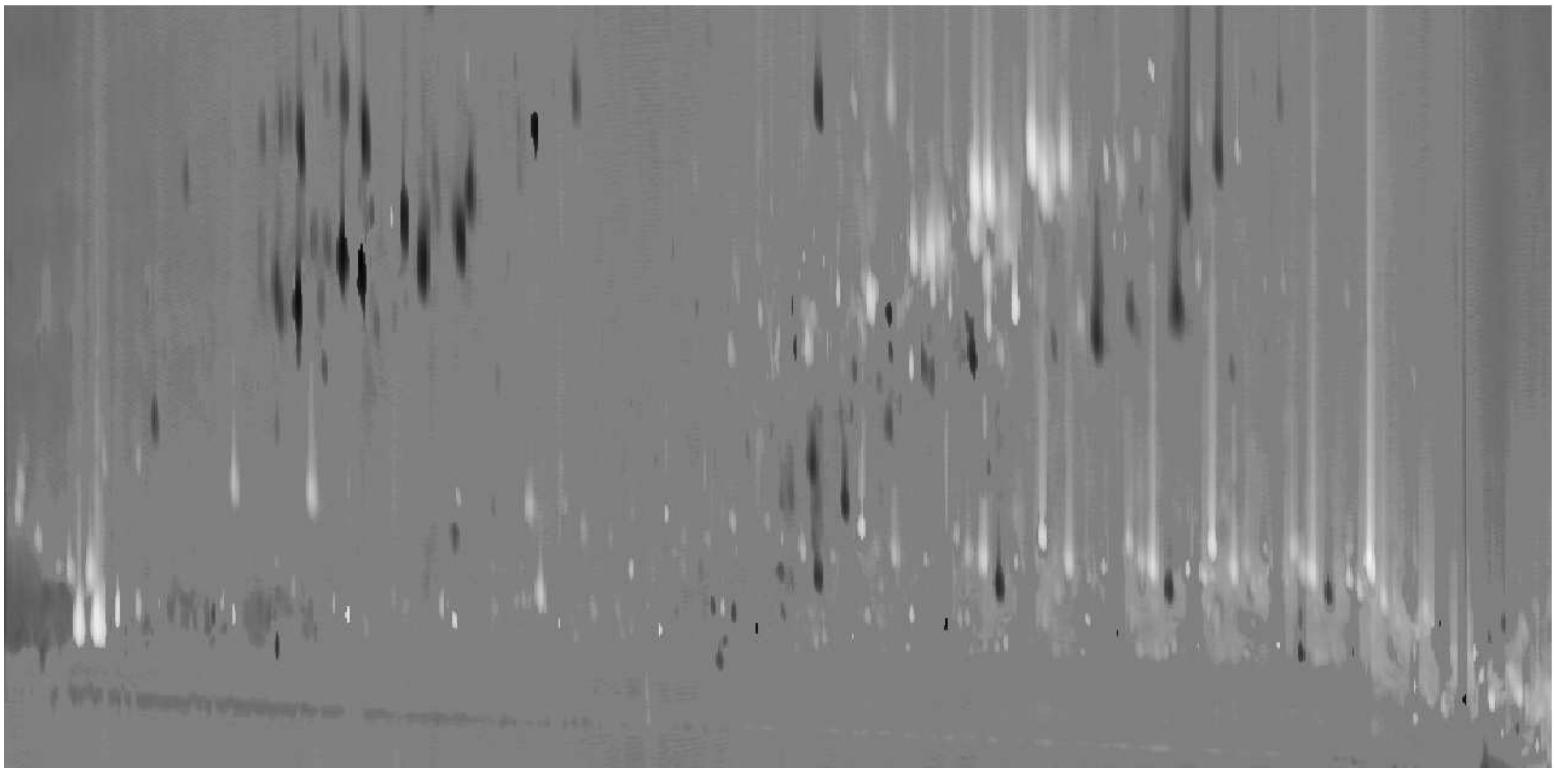
Reference image pixel values compared to values in neighborhood around corresponding pixel in analyzed image (and vice versa). Fuzzy difference is zero unless pixel is outside range of values in neighborhood.

Repeat for analyzed image pixel values compared to neighborhood in reference image. Then, take the value with the larger magnitude.

Differences for slight misregistration or peak shape are removed.

Grayscale shows only differences, not underlying peak intensities.

Grayscale fuzzy difference



Grayscale fuzzy difference.

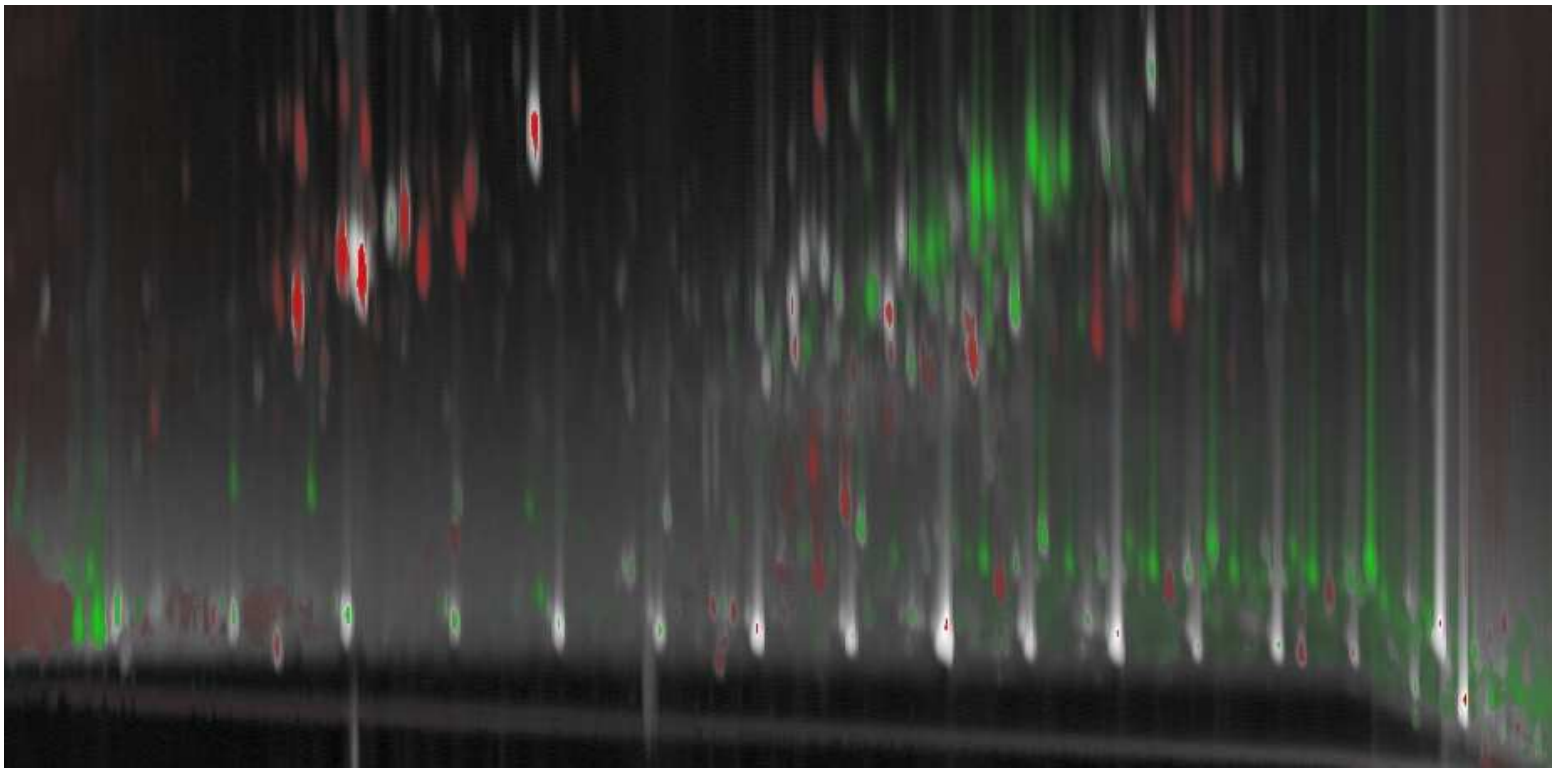
Colorized fuzzy difference

Use fuzzy difference for saturation.

Differences for slight misregistration or peak shape differences are removed.

Pseudo-color shows both differences and underlying peak intensities.

Colorized fuzzy difference



Colorized fuzzy difference.

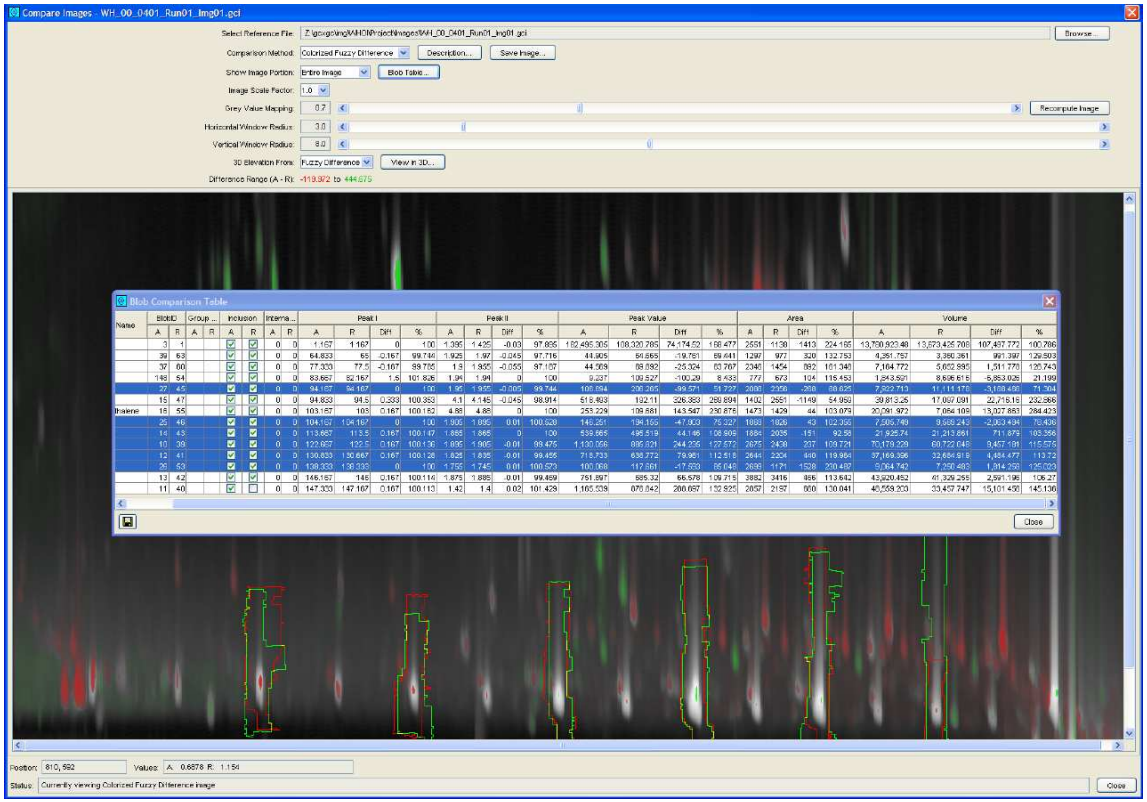
Tools for comparative analysis

Masking — excludes (or blocks) user-specified regions or peak sets from comparison.

Tabular data — presents quantitative data, differences, percents, etc. Allows interactive sorting and controls for graphical highlights in image.

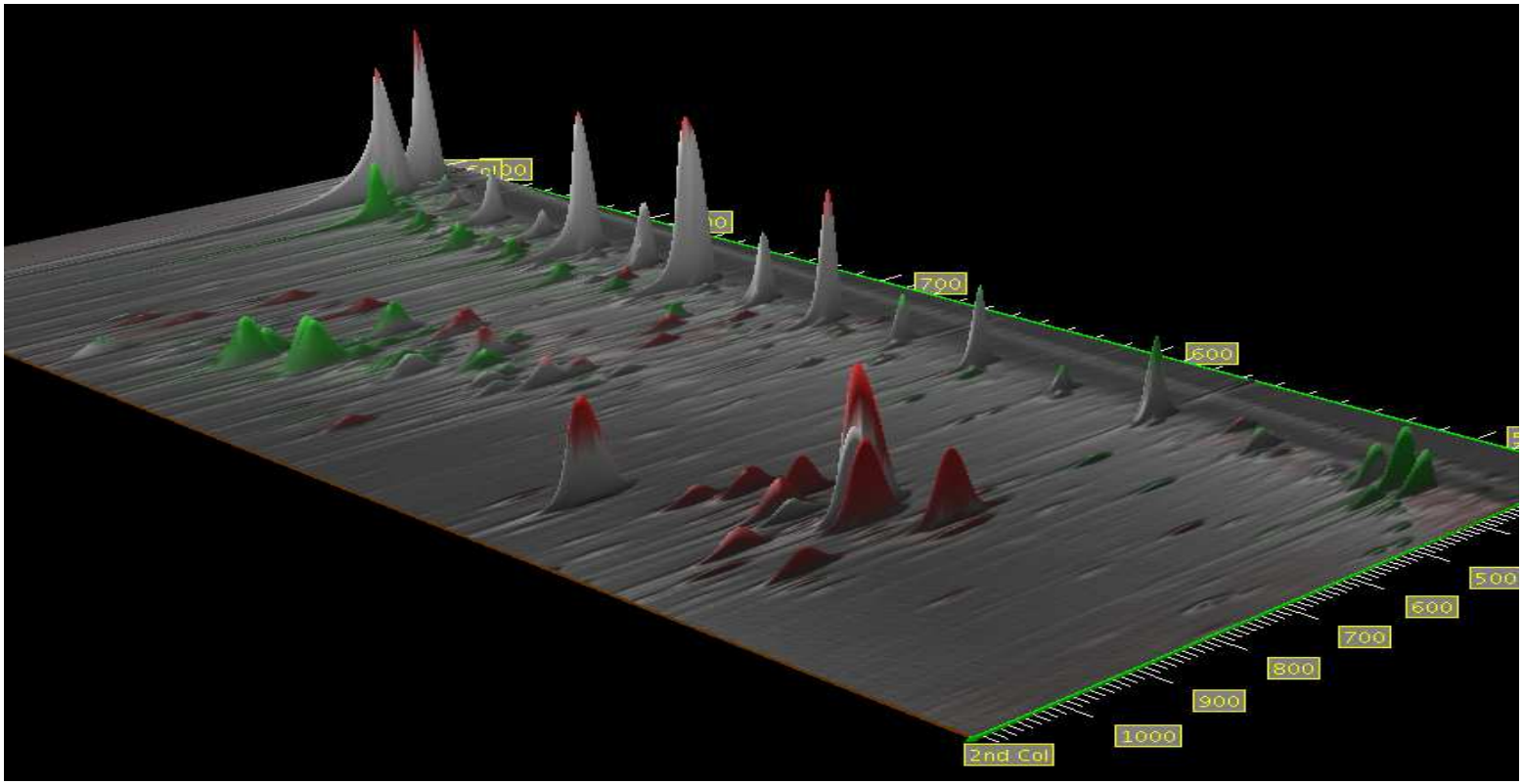
3D view — uses elevation as another dimension for visualization. Select any method for pseudo-colorization and then use value (analyzed, reference, or maximum) for elevation.

Tabular data and graphical highlights



Tabular data and graphical highlights.

3D colorized fuzzy difference



3D image with colorized fuzzy difference draped over maximum value elevation map.

Conclusions and future work

Image processing based on GCxGC metadata registers data to remove retention time variations and normalizes data to remove sample amount variations.

New colorized fuzzy difference successfully illustrates both differences and underlying intensities.

Masking, tabular, and 3D tools enhance comparative analyses.

Current work, supported by NSF and NIH, is investigating model-based analyses and comparisons.

Questions ?

technical: reich@cse.unl.edu

software: info@gcimage.com

www.gcimage.com

licensing: info@zoex.com

www.zoex.com