retroReveal.org: Semi-automated Open-source Algorithms and Crowdsourcing Tools for the Discovery, Characterization and Recovery of Lost or Obscured Content

Hal M. Erickson; Div. Cardiothoracic Surgery; University of Utah Health Sciences Center; Salt Lake City, UT, U.S.A.

Joyce Ogburn; J. Willard Marriott Library; University of Utah; Salt Lake City, UT, U.S.A.

Abstract

This interactive session will mark the formal launch of retroReveal 1.0, a free service located at www.retroReveal.org. Supported and hosted by the University of Utah Marriott Library. The server provides semi-automated forensic-style enhancement of digital images of varying quality from cameras and scanners. Though originally targeted at archivists, curators, conservators and scholars, retroReveal algorithms have proven useful to others during the development phase, most notably archaeologists.

retroReveal accepts source images of modest-to-good quality and provides an accessible interface allowing non-experts to apply forensic-style algorithms and to easily review, flag and tag sets of processed images from their own or institutional uploads. retroReveal is not intended to replace advanced imaging technologies for the most difficult cases. It merely delivers free user-friendly tools for discovery and adequate identification/ characterization of relevant content in that vast majority of cases where sophisticated investigation is infeasible.

That conceded, recent problems solved include:

1. rendering for translation of a papyrus palimpsest so faint that the scholar hadn't even correctly identified the language; the palimpsest is Moslem "cover scroll" that reportedly predates the previous earliest known instance;

2. reading through strong redaction in a Mark Twain letter;

3. reading through endpapers down to a vellum letter used as a book's sewing support in the binding process ;

4. visualizing medieval Arabic content through gold leaf;

5. recovering Beethoven's water-damaged instructions to a publisher concerning musical details of a score;

6. imaging miscarriage-related erasures on a sketched self-portrait of Frida Kahlo with Diego Rivera;

7. revealing fine detail of Stradivarius wood and varnish in a nearly black section of a journal-published photomicrograph;

8. recovering content from massively-overexposed microfilms of objects that can no longer be accessed;

9. revealing from a poor-quality low-light aerial photo slide that a well-surveyed small and unimportant archaeological site was actually a major archaeoastronomical complex;

10. from low-quality source images, visualizing exposure-faded traces of messages written in wagon axle-grease on "message stones" along the Oregon Trail..

11. recovering a title from faint scratches on the skiver-covered front board of a Brigham Young diary covering one of the most contentious summers in the history of Utah settlement.

At present much of the image enhancement is database-managed script-driven processing using NIH ImageJ [1], though a switch to compiled binaries is pending. All scripts and processing source code will evolve as open-source, and all processed imagery and metadata are distributed under Creative Commons licensing. While processed images can be saved to the user's hard drive, very soon any processed result may be assigned a DOI. This persistent academically-citable locater of digital resources will allow others to invoke the original algorithm and regenerate/inspect the view that was seen by the citing, corresponding or collaborating author.

In the near future, the server will also crowdsource the flagging of lost/obscured content in the digitized collections of partner institutions, but these features are still under development and may not be available at the launch of retroReveal 1.0. When implemented, online tutorials will train volunteers to spot such features as palimpsests, redactions, authorial erasures, bleached or erased provenancial markings/marginalia, alterations, content beneath pasted-down binding materials, fugitive media, content on verso of mounted items, water-damaged media, obscuring mold or soiling/staining, and archaeological visualization both of land use in remote imaging and fugitive content on stone.

The nature of retroReveal workproduct is such that any substantive description must be highly dependent on pairs of high-resolution images. at least one of which is typically full-color. This dependence means that a useful communication (1) cannot be usefully rendered in grayscale printing, and (2) will be pushed beyond the page constraints of a Proceedings paper. The interested reader may download an IS&T-formatted white paper from the University of Utah institutional repository at http://uspace.utah.edu/ or from the retroReveal site itself at http://www.retroReveal.org/pubs/Archiving2012.pdf.

References

[1] Wayne S. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, 1997-2011). http://imagej.nih.gov/ij/.

Author Biography

Hal Erickson is a mathematical and chemical physicist by training, working as a bioinformaticist and imaging specialist in a research group focused on stem cell therapies in cardiovascular applications. For thirteen years he also taught the conservation materials science course sequence in the book-&-paper-focused Masters degree program offered by the Preservation and Conservation Studies program, University of Texas-Austin. He still regularly teaches invited seminar and courses in conservation science and imaging science.

Joyce L. Ogburn is the Dean of the J. Willard Marriott Library and University Librarian for the University of Utah. She oversees the University of Utah Press, is the current President of the Association of College and Research Libraries, and is a prolific author and speaker on library and scholarly communication issues.