

Recovery of handwritten text from the diaries and papers of David Livingstone

Keith T. Knox
Air Force Research Laboratory
Maui, Hawaii

Roger L. Easton, Jr.
Rochester Institute of Technology
Rochester, NY

William Christens-Barry
Equipoise Imaging, LLC
Ellicott City, MD

Kenneth Boydston
Megavision, Inc.
Santa Barbara, CA

26 January 2011
Electronic Imaging Symposium
Computer Vision and Image Analysis of Art II

Manyema Field Diary, 1870-1871



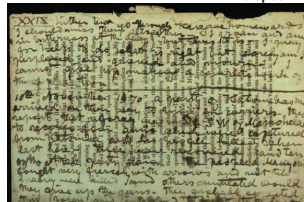
Historical Significance

- Tropical diseases
- African slave trade
- Just before Stanley

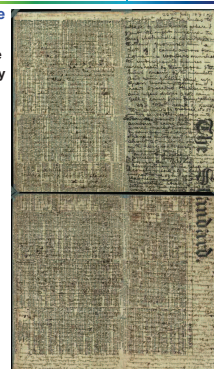
Ran out of paper, ink

- Newspapers
- Sermons
- Envelopes

Envelopes



Sermons



Newspapers

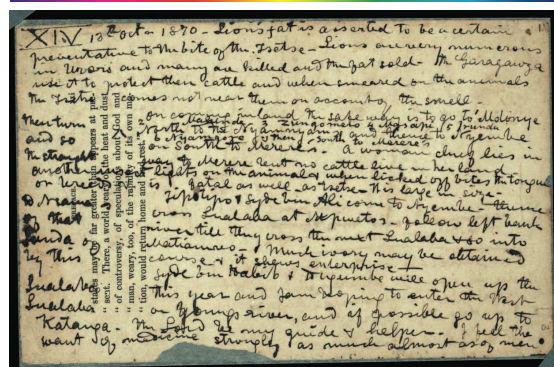
4 of 29

Outline

- **David Livingstone**
 - Famous missionary from Scotland
 - Writings influenced public opinion of African slave trade
- **Multi-spectral Imaging Team**
 - Recovered erased text from Archimedes Palimpsest
 - Imaged Livingstone's diaries at National Library of Scotland in June/July 2010
- **Image Processing Techniques**
 - Principal Components Analysis (separates writing)
 - Spectral Ratios (suppress printed text)
 - Pseudocolor (distinguish ink that bled through paper)

2 of 29

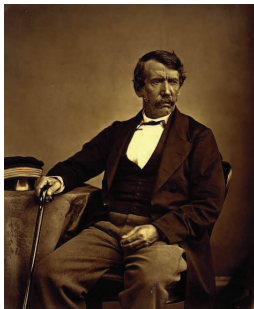
Many Leaves are Legible



5 of 29

David Livingstone 1813-1873

- **Missionary in Africa**
 - 1841 arrives in South Africa
 - 1847 sets up mission by Zambezi River
- **First Expedition**
 - 1853 expedition into interior
 - 1855 discovers Victoria Falls
 - 1856 publishes *Missionary Travels*
- **Second Expedition**
 - 1858 official expedition into interior
 - 1864 expedition recalled after wife dies
- **Third Expedition**
 - 1866 seeks source of Nile
 - 1871 encounters H. M. Stanley in Ujiji
 - 1873 dies of malaria in Zambia

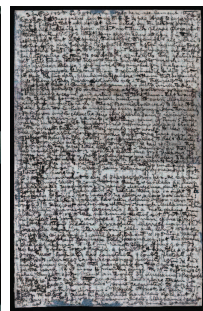


3 of 29

Many are Illegible



Fading ink -- interference from printed text



Bleed-through from other side

6 of 29

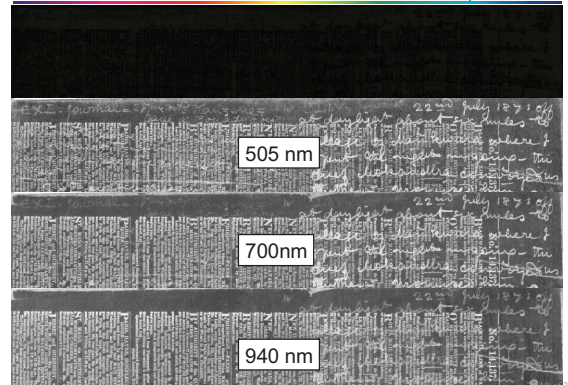
Multispectral Imaging Team



Doug Emery	Roger Easton Jr.	Bill Christens-Barry	Mike Toth	Keith Knox	Ken Boydston
Metadata	Image Processing	Lights	Management	Image Processing	Image Capture

7 of 29

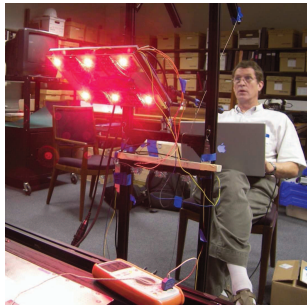
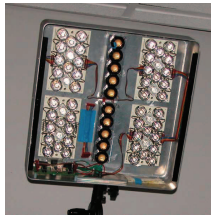
Spectral Dependence



8 of 29

LED Illumination Panel

Ultraviolet	450 nm	Infrared	735 nm
	465 nm		780 nm
	505 nm		870 nm
	535 nm		940 nm
	592 nm		
365 nm	638 nm		
	700 nm		

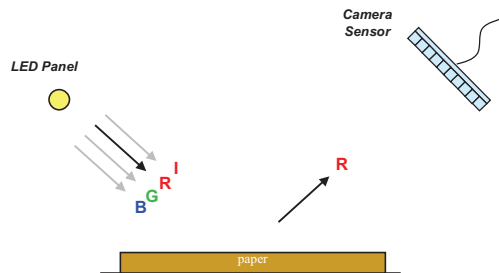


8 of 29

Principal Components Analysis to Separate Writings

11 of 29

LED Illumination



Images at every wavelength are in perfect registration

9 of 29

Principal Component Analysis

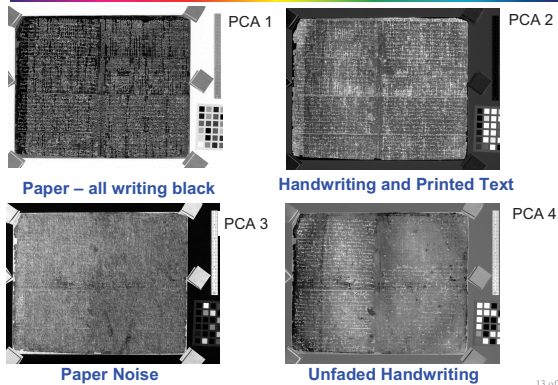
- Data Set**
 - Perfectly registered data cube
 - 12 wavelengths, 365 nm – 940 nm
- Principal Components Analysis**
 - Finds axes of maximum variation
 - Components ordered by significance
- Separating the Writings**
 - Three kinds of writing – printed text, handwriting and bleed-through
 - Sometimes more than one ink
 - Different spectral variation for each writing
- Application of PCA to data cube**
 - Goal is to isolate handwriting from printed text
 - PCA components seldom align perfectly with individual writings



Visible image of handwriting on newspaper

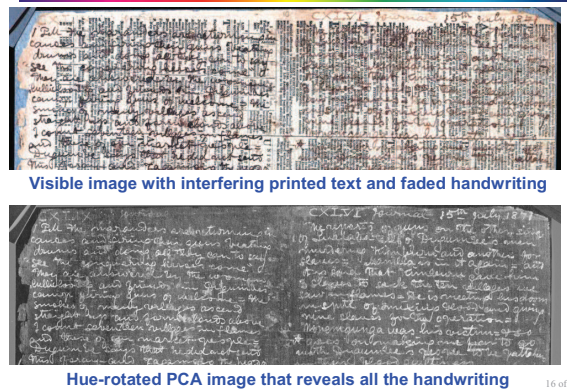
12 of 29

Principal Components



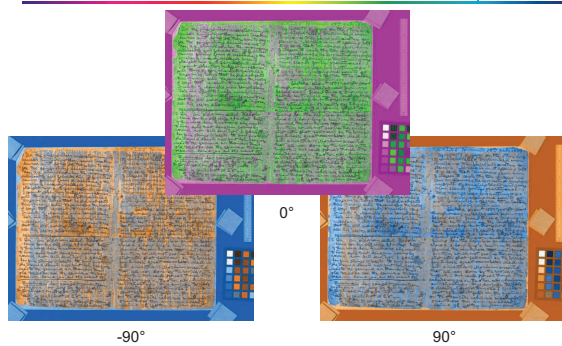
13 of 29

Handwriting Recovered



16 of 29

Hue Rotation



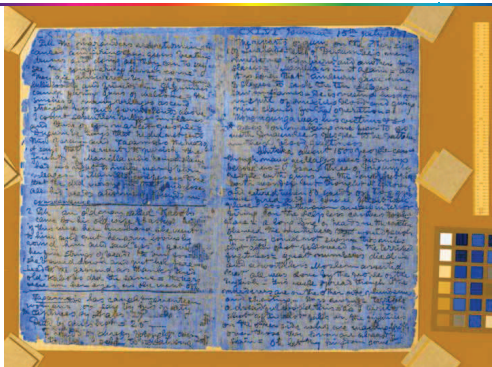
Insert PCA components 1 & 2 in a color image and rotate hue

14 of 29

Spectral Ratios Suppresses Printed Text

17 of 29

Hue Rotation



Choose hue angle that equates paper & text values in one color channel

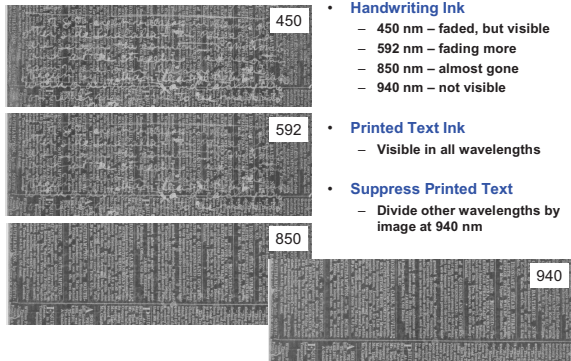
15 of 29

Printed Text Interferes with Handwriting



18 of 29

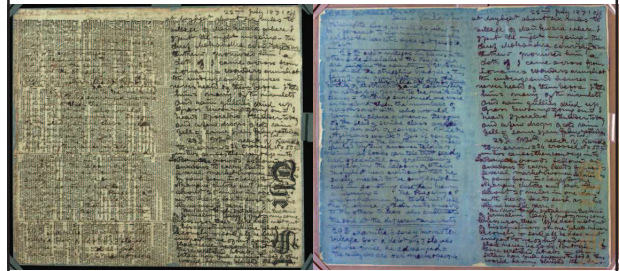
Spectral Dependence of Inks



- **Handwriting Ink**
 - 450 nm – faded, but visible
 - 592 nm – fading more
 - 850 nm – almost gone
 - 940 nm – not visible
- **Printed Text Ink**
 - Visible in all wavelengths
- **Suppress Printed Text**
 - Divide other wavelengths by image at 940 nm

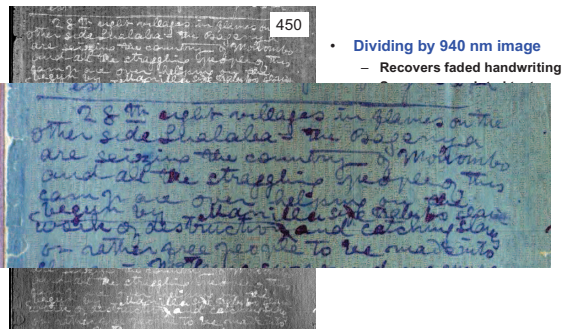
19 of 29

Full Page Result



22 of 29

Spectral Ratios



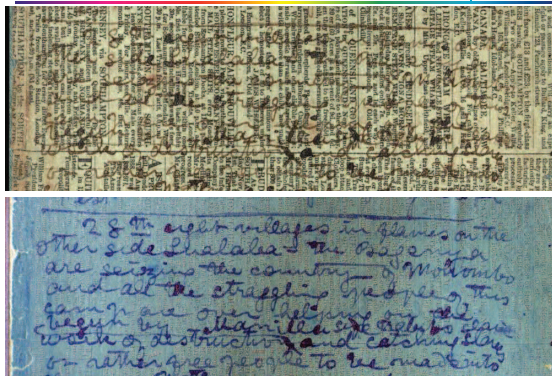
- **Dividing by 940 nm image**
 - Recovers faded handwriting

20 of 29

Pseudo-color to Suppress Bleed-through

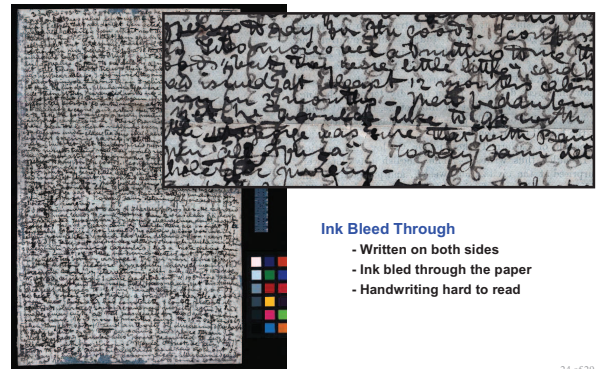
23 of 29

Color Image, Spectral Ratio Comparison



21 of 29

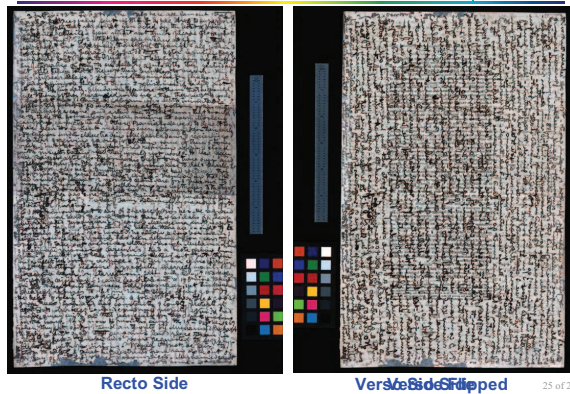
Ink Bleed Through



- Ink Bleed Through**
- Written on both sides
 - Ink bled through the paper
 - Handwriting hard to read

24 of 29

Scan Both Sides of Page



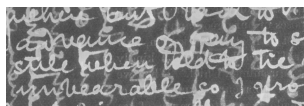
25 of 29

Summary

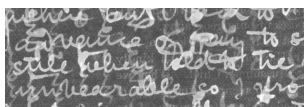
- **Principal Components Analysis**
 - Principal components not always aligned with writings
 - Hue rotation of color component image emphasizes writings
 - Method helps separate writings and recover handwriting
- **Spectral Ratios**
 - Handwriting ink is not visible in near infrared
 - Printed text is constant over the spectrum
 - Dividing by near IR image suppresses printed text
- **Pseudocolor**
 - Combine both sides of page in pseudocolor image
 - Uses color to distinguish writing from both sides

28 of 29

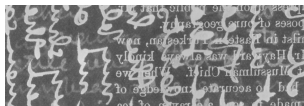
Pseudocolor Combination



505 nm Recto → R



780 nm Recto → G



505 nm Verso → B

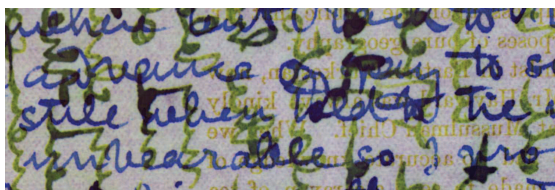
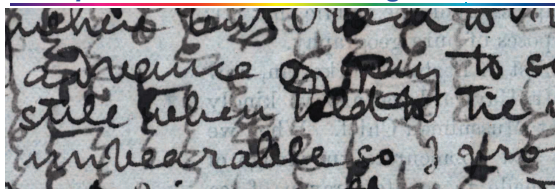
26 of 29

Acknowledgments

- **Dr. Adrian Wisnicki**
 - Research consultant with *Livingstone Online*
- **Grant from National Endowment for the Humanities**
 - **Volunteers and staff from**
 - National Library of Scotland
 - David Livingstone Centre
 - Edinburgh Napier University

29 of 29

Comparison with Visible Image



27 of 29