

Book Scanning – Technologies and Techniques

Mike Mansfield

Director of Content
Engineering

Ancestry.com /
Genealogy.com

Outline

- Project Analysis
- Scanning Parameters
- Book Scanners

Project Analysis Overview

- Scope
- Goals
- Project and Customer Requirements
- Content Evaluation

Project Analysis

- Assessment
- Selection
- Information Representation – Goals and Metrics
- Funding
- Planning and resource assignment
- Prepare the originals for digitization
- Scanning
- Quality Assurance
- Post-Processing – OCR, Compression, Format Conversion
- Return originals to the collection
- Host the data
- Archiving and preservation

Book Scanning Parameters Overview

- Resolution
- Bit Depth
- Dynamic Range
- Tonal Sensitivity
- Geometrical Corrections
 - De-Skew
 - Curve Correction
 - Text Crushing
- Masking and Cropping

Resolution

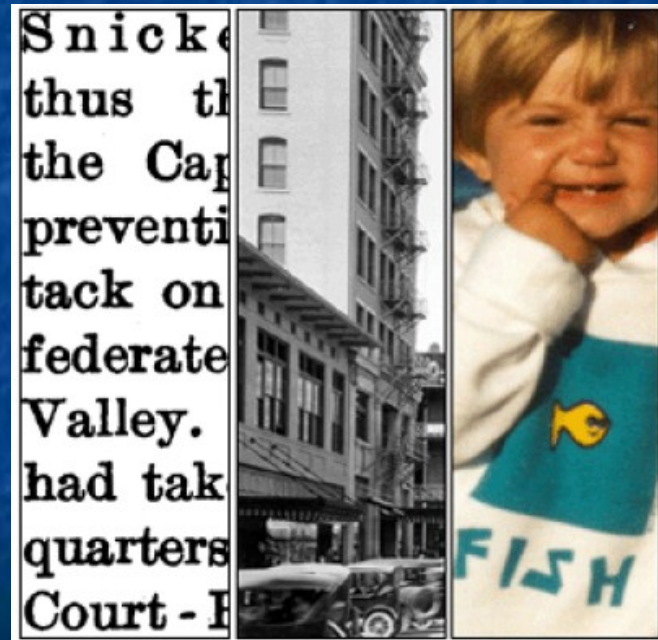
- Samples Per Inch (SPI), Dots Per Inch (DPI), Pixels Per Inch (PPI)
- Archival Quality
- Access Quality
- “Faithful” Representation of the page

Resolution and OCR

- Most OCR engines are optimized for 300 DPI images with typefaces in point sizes between 10 and 14.
- In cases where the font size of characters on an image are very small (point size of 6 or less), scanning images at 400 DPI can improve character recognition

Bit Depth

- Number of colors or “tones” a scanner can differentiate
- Bitonal
- Grayscale
- Color



Dynamic Range

- A scanner's dynamic range is a measure of how well the device can record changes in the brightness of the image it's scanning



Tonal Sensitivity

- The ability of a scanner to accurately represent similar, adjacent tonal values as distinct from each other

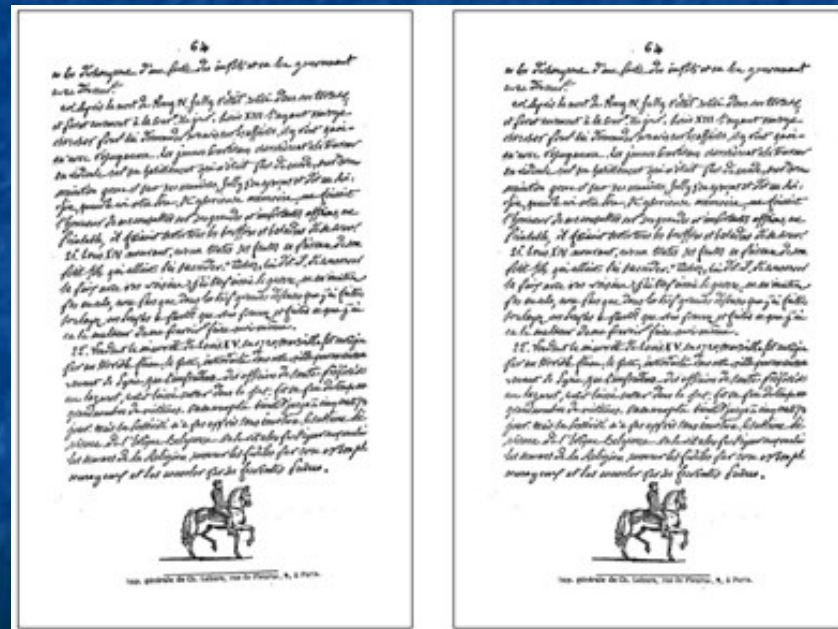


Geometrical Corrections

- Deskew
- Bookfold Corrections
 - Curve Correction
 - Text Crushing

Deskew

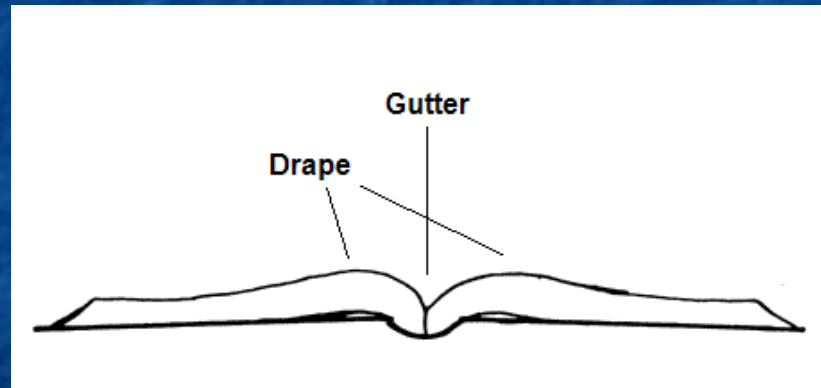
- Skew detection and correction

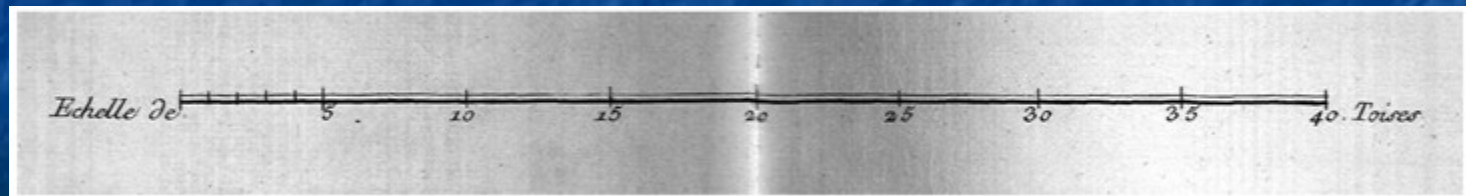
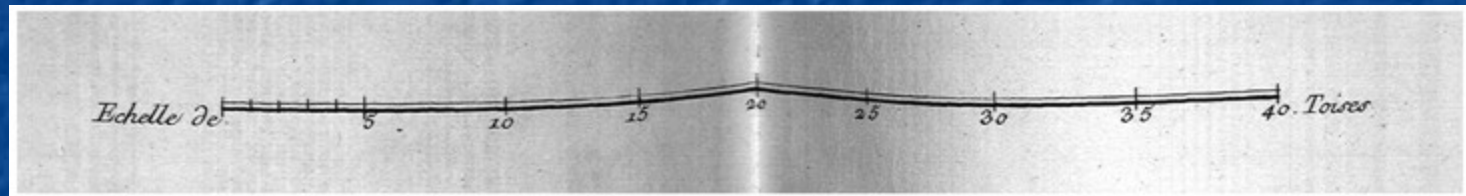
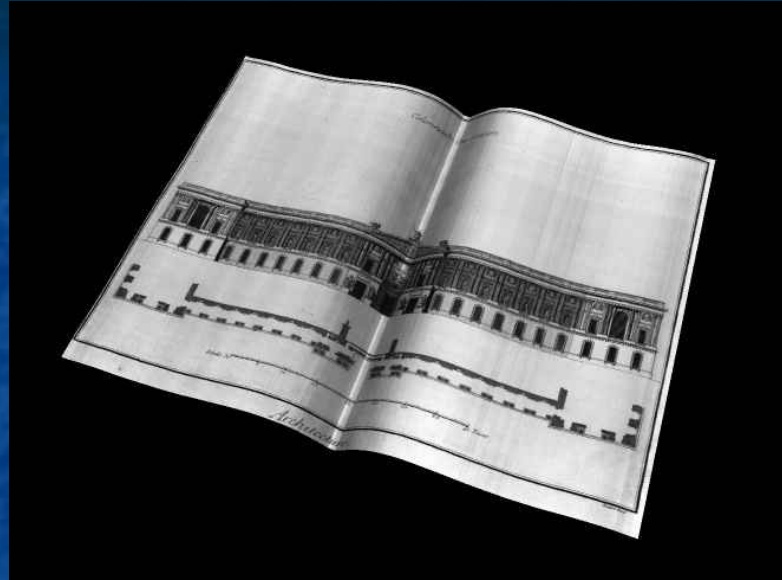


Bookfold Corrections

Curve Correction and Text Crushing

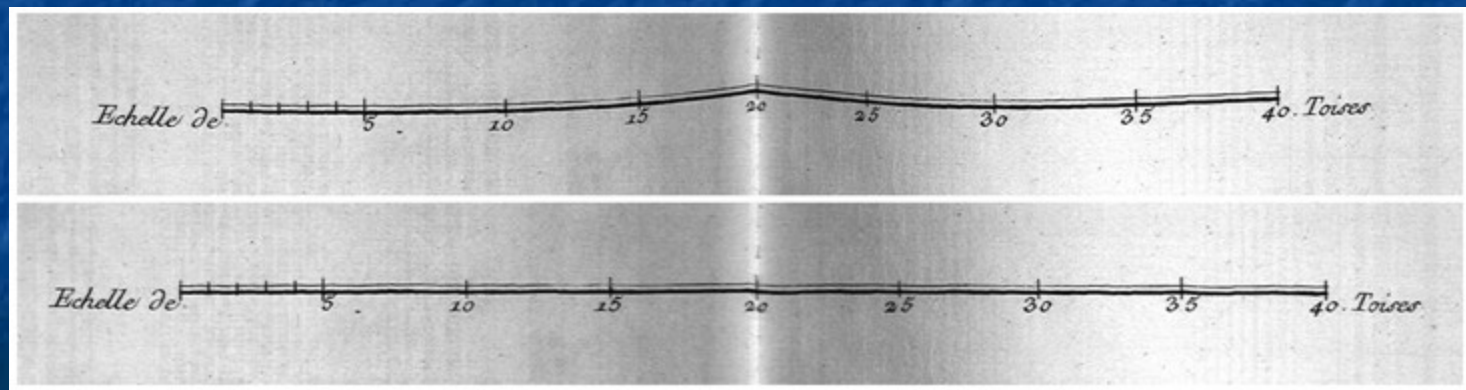
- Pages of bound books are three dimensional surfaces





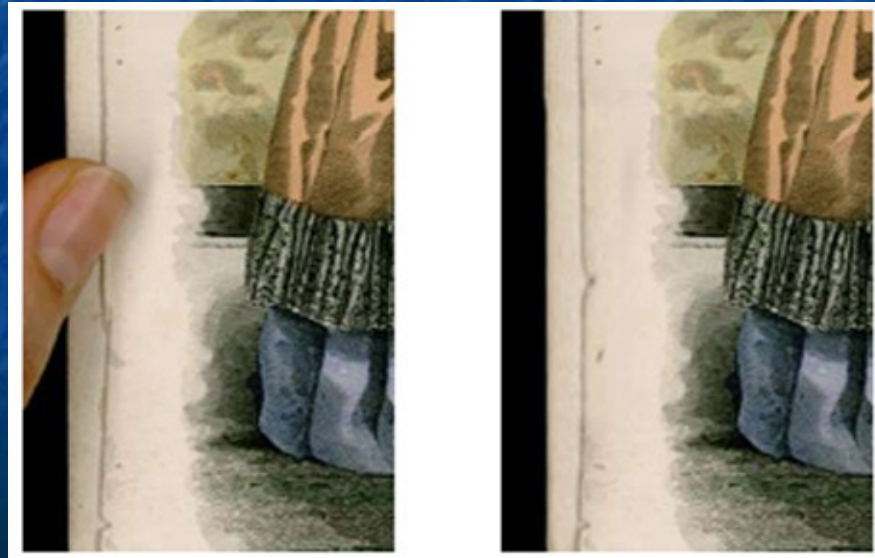
Curve Correction and Text Crushing Compensation

- Straighten curves and preserve uniform distances in the drape and gutters of scanned book pages



Finger Masking

- Methods to remove the images of the operator's fingers holding down the pages during scanning



Cropping and Page Splitting

- Detecting and cropping edges to remove portions of the image containing the book cover, end-papers, spine edges, and page fan-outs.
- Splitting double page images.

Not What We Want



What We Do Want

DATE.	NAMES.	RECORD.	VOL. PAGE.
1768. Aug. 10.	Hudson, Frederick, and Sarah Youngs, . . .	M. B.,	XIII. 173
1761. Mar. 31.	Hudson, William, and Elenor Murphy, . . .	"	IV. 125
1772. Jan. 13.	Hoeck, Elbertie, and Tobias Van Boeren, . . .	"	IX. 8
1778. Dec. 10.	Huestis, Edward, and Hanah McConnell, . . .	"	XXVI. 118
1757. April 22.	Huestis, Patty, and James Boyd, . . .	"	I. 511
1736. Nov. 22.	Huff, Catherine, and Isaac German, . . .	"	I. 8
1781. Oct. 21.	Huff, Gamaliel, and Jane Cutler, . . .	"	XXXIV. 2
1757. Feb. 22.	Huff, Hannah, and Richard Burd, . . .	"	I. 448
1762. Sept. 1.	Huff, Henry, and Susannah Hyatt, . . .	"	VI. 301
1736. July 26.	Huff, Lawrence, and Susannah German, . . .	"	I. 2
1782. April 29.	Huff, Uriah, and Deborah Townsend, . . .	"	XXXV. 144
1780. May 22.	Hugford, Mary, and Amos Dodge, . . .	"	HL 162
1760. Feb. 27.	Hugens, Mary, and Mathew Hopper, . . .	"	III. 55
1768. Nov. 7.	Hugeny, Margrietta, and Abraham G. Voorburgh, . . .	"	XIII. 229
1758. Oct. 14.	Huger, Catrina, and Johannes Maurritius Goetschius, . . .	"	II. 56
1781. Oct. 27.	Huggeford, Martha, and Elias Hardy, . . .	"	XXXIV. 9
1767. Oct. 28.	Huggeford, Sarah, and Elijah Drake, . . .	"	XII. 72
1781. Nov. 27.	Huggeford, Tiamur, and Henry Bogert, . . .	"	XXXIV. 61
1778. Aug. 11.	Huggins, Mary, and William Appleton, . . .	"	XXVI. 2
1774. Mar. 2.	Huggins, Sarah Child, and Edwin Gool, . . .	"	XXII. 138
1763. Aug. 5.	Hughes, Catharine, and James Simpson, . . .	"	VII. 262
1770. Aug. 25.	Hughes, Hannah, and Mary Denniston, . . .	"	XVI. 171
1761. July 6.	Hughes, James, and Phebe Lowther, . . .	"	IV. 284
1765. Aug. 5.	Hughes, John, and Mary Denniston, . . .	"	IX. 254
1756. Sept. 30.	Hughes, Jonathan, and Catharine Anderson, . . .	"	I. 203
1764. Dec. 6.	Hughes, Margaret, and Robert Johnson, . . .	"	VIII. 441
1763. June 6.	Hughes, Meredith, and Margaret Humphries, . . .	"	VII. 214
1762. Jan. 2.	Hughes, Nathaniel, and Eleanor Stuart, . . .	"	VI. 1
1761. June 5.	Hughes, Nathaniel, and Sarah Langford, . . .	"	IV. 196
1762. Oct. 18.	Hughes, William, and Mary Armstrong, . . .	"	VI. 371
1773. Sept. 21.	Hugner, Mary, and John Clute, . . .	"	XXI. 113
1781. Jan. 6.	Hugnie, Daniel, and Mary Garabants, . . .	"	XXXI. 14
1769. May 23.	Huick, Andries, and Lenah Shaver, . . .	"	II. 280
1770. Oct. 12.	Huick, Burger, and Elizabeth Goes, . . .	"	XVI. 209
1755. Sept. 12.	Huick, Direct, and Sarah Vanderusen, . . .	"	I. 174
1758. Dec. 4.	Huise, Anneke, and William Carr, . . .	"	II. 113
1764. Sept. 12.	Hulet, Rebecca, and John Mitchell, . . .	"	VIII. 304
1738. April 21.	Hulet, Benjamin, and Susannah Whitehead, . . .	"	I. 9
1769. Jan. 9.	Hewlett, Daniel, and Mary Mott, . . .	"	XXV. 7
1758. April 24.	Huett, Hannah, and Joshua Cornwall, . . .	"	I. 884
1737. Jan. 13.	Hulet, John, and Elizabeth Dosenbury, . . .	"	I. 8
1761. Oct. 17.	Hulet, William, and Abigail Still, . . .	"	V. 144
1780. Nov. 21.	Hulett, Elizabeth, and William Maffett, . . .	"	XXX. 119



(From by Atkinson, Troy, N. Y.)

William Barton

WILLIAM BARTON was born in Troy, N. Y., Feb. 7, 1805, and was next to the eldest in the family of three sons and three daughters of Robert and Mary (Carpenter) Barton, the former a native of Dutchess County, and the latter a native of Ulster Co., N. Y. His parents came to Troy in 1803. His father at that time was a farming mill maker; was afterwards in the grocery and provision business, and still later in the manufacture of soap and candles. He was an active business man, was connected with various enterprises in the early history of Troy, and was in politics one of the old Federal party. He died in Troy, in the year 1836, aged fifty-nine. His wife died in 1846, aged eighty-six. Both were members of the Society of Friends.

William Barton received a good education at the Friends' School, at Nine Partners, Dutchess Co., N. Y. For many years he was connected with his father in business in Troy. In 1832 (having previously studied engineering with Amos Eaton and others) he became assistant engineer in the construction of the macadamized road from Troy to Bennington. In 1833-34 he was assistant engineer on the Rensselaer and Saratoga Railroad, and in the construction of the bridge at Troy, over the Hudson River, for that road. From 1836 to 1839 he was engaged in the surveys for the Boston and Albany Railroad, and subsequently contracted to build sections of that road, and also to build the wharves and accommodations for the depot on the east side of the Hudson at Albany.

In May, 1839, Mr. Barton married Mary A., daughter of James and Rebecca Wells, of Springfield, Mass., and sister of David A. Wells, of Norwich, Conn. Returning

to Troy the same year, he was appointed city surveyor, and held the office almost continuously until 1859.—a term of nearly twenty years,—during which time, in 1858, he made the first complete map of Troy, embracing the whole city, and was also the engineer for the extension of the wharves along the city front.

In 1859-60 he was one of the organizers of the Arta Reid Steam Fire-Engine Company, and was its first president. In 1860-61 he organized the Troy and Lansingburgh Street Railroad Company, and the Troy and Colson Street Railroad Company, and was the engineer and superintendent until 1867. He has been one of the directors since the organization of both roads, and is now one of the executive committee. From 1870 to 1875 he was engaged in the coal business, at the corner of Jacob and River Streets.

Mr. Barton has taken an active part in the politics of his city and county. Was formerly a member of the Whig party; is now a Republican. For two years, 1847-48, he represented the Seventh Ward in the common council of Troy. He has ever been interested in all enterprises tending to the prosperity of the city where he has spent most of his life as an active business man.

Raised under Quaker influences, he in early life held that belief, but about 1836 he became a Unitarian, and after returning to Troy he became one of the founders of the Unitarian Church in that city.

Mr. Barton is well known in Troy as an enterprising, upright, and judicious business man. His correct habits, integrity, and honesty of purpose in all his relations of life have won the confidence of all who know him.

Bond of Cornelis Vos to Johanna de Hulter

[317] Appeared before me, Johannes La Montagne, in the service of the General Chartered West India Company commissary at Fort Orange and the village of Beverwyck, in presence of Abram Staets and Adriaen Gerritsen, magistrates of said jurisdiction, Cornelis Vosch, who acknowledges that he was well and truly indebted to Madam Johanna de Hulter in the sum of one hundred and ninety Carolus guilders in good, whole beavers, which he promises to pay on the first of August A°. 1659; as security⁷⁰ and special pledge for which sum he, Cornelis Vosch, offers the last payment on his new, sold house, which payment is due on the date above named; binding furthermore his person and estate, real and personal, under submission to all courts and judges. Done in Fort Orange, the 18th of July A°. 1658.

Abram Staets

Adriaen Gerritsen

CORNELIS VOS

Acknowledged before me,

LA MONTAGNE, Commissary at Fort Orange

Satisfaction of above bond

Appeared before me, Johannes La Montagne, commissary of Fort Orange and the village of Beverwyck, Madam Johanna Ebbings⁷¹ who declared that the above bond was fully satisfied and paid. Done in Fort Orange, the 8th of August A°. 1659.

JOHANNA EBBINCK

Acknowledged before me,

LA MONTAGNE, Commissary at Fort Orange

Bond and mortgage of Teunis Slingerlant to Johanna de Hulter

[318] Appeared before me, J. La Montagne, in the service of the General Chartered West India Company commissary at Fort Orange and the village of Beverwyck, in presence of the Honorable Jan Tomassen and François Boon, magistrates of said jurisdiction, Teunis Slingerlant, who acknowledged that he was well and truly indebted to Madam Johanna de Hulter in the sum of thirteen

⁷⁰ *Hypothec*, usually translated "mortgage."

⁷¹ Johanna Ebbings was the daughter of Johannes de Laet. Her first husband was Johannes de Hulter, with whom she came to New Netherlands in 1653.

Book Scanners Overview

- Document Scanners
- Planetary Book Scanners
- Flying Linear Arrays
- Digital Photography
- Robotic Page Turners

Document Scanners

- Cut the spine off of the book and scan the loose pages in a document scanner
 - The book is rendered almost useless for additional use
 - Rebinding is expensive and slow
 - Makes most sense when a sacrificial copy of the book exists.

Document Scanners

- Extremely Fast
- Feature Rich
- Relatively Inexpensive
- Large range of options and price points
- Some limited applications in the Family History and Genealogy domain

Document Scanners

- Major office equipment manufactures

- Canon
- Fujitsu
- Kodak
- Panasonic
- Ricoh



Document Scanners

- Resolution: 100-600 DPI
- Bit Depths: Bitonal, Grayscale, Color
- Simplex / Duplex
- 2 x 3 inch to 12 x 30 inch documents
- Rate: Few hundred pages per day to tens of thousands of pages per day
- Deskewing, cropping, dithering, dynamic thresholding, binarization, etc...

Planetary Book Scanners

- Specialized devices designed to do primarily one thing – scan bound books
- CCD Array, integrated lighting, specialized scan beds/book cradles, and book specific image processing options

Dissection of a Minolta PS 7000

- 7,500 Pixel Reduction type line CCD
- Halogen Lamp Lighting
- Up to A2 Size
- 200/300/400/600 DPI
- Bitonal or 8-bit Grayscale
- 4.5 Seconds per scan on an A4 page at 400 DPI



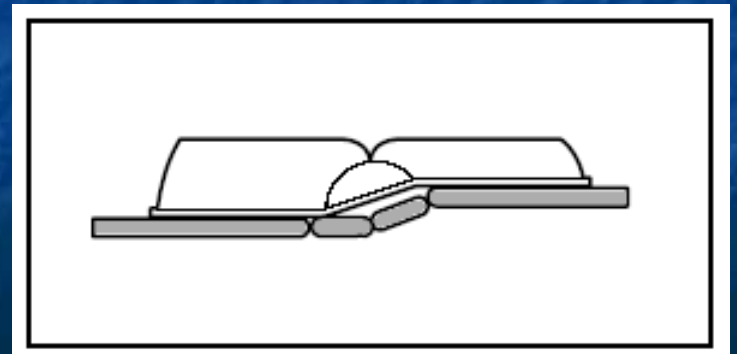
Dissection of a Minolta PS 7000

- Image Processing
 - Curvature Correction
 - Text Crushing Correction
 - Centering
 - Finger Masking
 - Spread/Single/Book Split
 - Linearization



Dissection of a Minolta PS 7000

- Articulating Book Cradle



Dissection of a Minolta PS 7000

- Scan buttons on the scan bed



Minolta



Bookeye



Zeutschel



Planetary Book Scanners

- Resolutions from 300 DPI to 600 DPI
- Bit-Depths: Bitonal, Grayscale, Full Color
- Rich feature set well suited to large production projects
 - Book cradles, glass plates to reduce page curvature, specialized image processing, human-factors, etc.
- Support for most book sizes from small books to large quarto volumes and smaller atlases
- Proven technology, few moving parts, highly reliable
- 1 page scan in 5-10 seconds
- 1,500 – 3,000 pages per 8 hour shift

Flying Linear Arrays

- Integrated flying linescan CCDs and lighting systems with specialized scan tables and book cradles



i2S – DigiBook, Zeutschel



Flying Linear Arrays

- Resolutions up to 800 DPI
- Bit-Depths: Bitonal, Grayscale, Full Color
- Very high quality scans
- Feature rich systems with book specific support
 - Book cradles, glass plates, human-operation factors
- Support for very large volumes, atlases, and maps to a meter square
- 1 page scan in 2-6 seconds
- 2,500 – 4,500 pages per 8 hour shift

Digital Photography

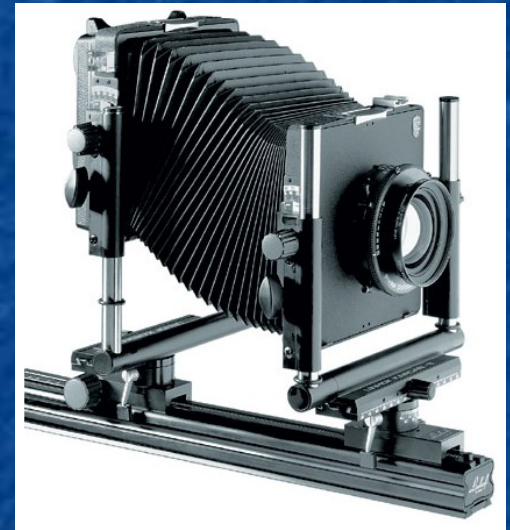
- Large Photographic Formats
- Scanbacks
- Professional Photographic Optics and Cameras
- Studio Lighting Systems
- Color Management Software
- Custom camera positioning and book holders

Large Format Photography

- Better Tonality
- Higher color depth
- Sharper
- Grain-free
- More control on the final geometry and perspective of the photographed pages

Scanbacks

- Large Format Cameras
- Trilinear array, 1-pass scan
- 6000 x 7250 pixels = 43 megapixels
- 30+ seconds for a single scan



Digitizing the Gutenberg Bible



Digitizing the Gutenberg Bible



Anagramm Picture Gate 8000 Scanback

- Trilinear CCD Array, 1-pass scan
- Large Format Camera 9 x 12 cm
- 8000 x 9700 Pixel Optical Resolution = 77.6 megapixels
- 48 Bit Color Depth
- 444 MB File in 48 Bit Color
- 40 Seconds for a full scan
- Fiber-optic connection

Digital Photography

- Super high quality images
- Custom lighting and positioning
- Slow
 - Scanning page images is slow
 - Positing each page is slow
- Skilled and experienced photographers
- Few applications in the Family History and Genealogy domain

Robotic Page Turners

- Kirtas APT BookScan 1200
- i2S DigiBook – Digitizing Line



Conclusion

- Analyze your project's requirements and scope
- Understand the content and determine the scanning metrics
- Match the scanning technology to the content and project goals

Questions

MMansfield@Myfamilyinc.com