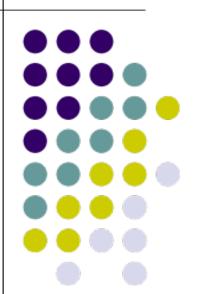
# **Image Capture 101**

Larissa Halishoff
Mike Conroy
UBC Digital Initiatives Unit
July 22, 2010



# contents



- Definition
- Types of Scanners
- The camera
- Atiz Bookdrive
- Workflow or how a digital object gets made
- What happens next?

# digitization



- The conversion of analog material to digital format
- Photographs, text, audio, video
- For print, called scanning or image capture
- Use of cameras or scanners

# cradle scanner (ATIZ)





- Bound books, journals, newspapers etc.
- Ideal for materials with fragile spines

# multi-page feeder (Fujitsu)





- Disbound items, unbound documents
- Ideal for multi-page, high-volume projects with straightforward text and pictures
- Can do 2-sided documents quickly

#### flatbed scanner





- Photographs, single pages, smaller size items
- Ideal for items that will lay flat

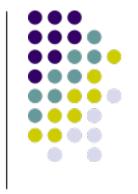
# large-format scanner (Vidar)





- Maps, newspapers, posters.
- Large-format, flat items that will not fit in conventional scanners.

# studio kit

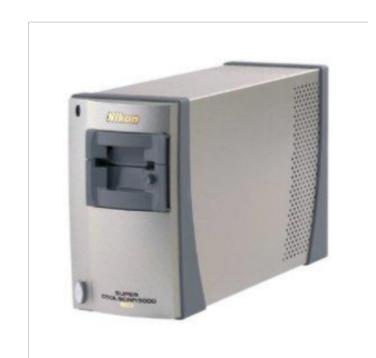




- Ideal for items that cannot be scanned through conventional scanners.
- Projects can use combination of ie. ATIZ and studio kit (ie. pullout map)

### 35mm slide scanner





Slide format only

# microfilm scanner





Microfilm reels only

#### the camera



- DSLR (Digital Single Lens Reflex)
- Detachable lens
- Large sensor
- Preview depth of field





### the camera



- Currently use Canon Rebel Xsi (450) /
   T1i (500)
  - Fixed lens (no zoom)

# the shutter

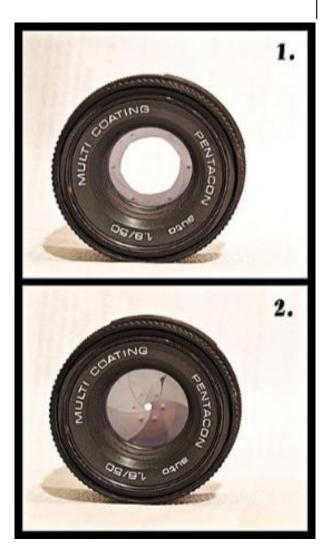


- Allows a pulse of light to hit the sensor by revealing the sensor for a fixed amount of time.
- Typically we use 1/50 of a second, which is a relatively slow shutter speed.

# the aperture



- The hole through which light travels from outside to the sensor.
- Measured in F stops
- Standard for our set up is F 5.0



#### how it all works



- Internal light meter determines how much light is required to produce an ideal exposure
- Interplay between shutter speed (how long light hits the sensor) and aperture (how much light gets through the hole)
- The smaller the aperture the greater the depth of field (focal range)

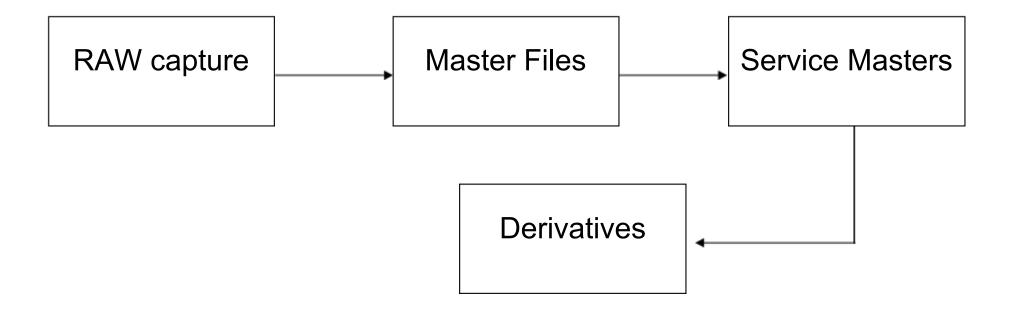
# **ATIZ Bookdrive Pro**



http://www.youtube.com/watch?v=fltTOvgPJAM

# production workflow





# RAW capture



#### Atiz image capture

- Canon RAW format (.CR2)
- High quality
- Camera's "native" format
- Not ideal for preservation
  - Non-standard
  - Proprietary
- Not usually kept

# master files



- Tiff format
- Derived from RAW files
  - Batch conversion using Bookdrive Editor pro, Irfanview, Photoshop....
  - Lossless
  - Format is controlled by Adobe but in public domain
  - Originally created as a standard for scanners
  - Considered appropriate for digital preservation
- NO OTHER IMAGE PROCESSING OCCURS AT THIS STAGE

#### service masters



- Derived from Master Files
  - created in image editing software such as Bookdrive Editor Pro.
  - Images are cropped, sharpened, colour corrected for optimal OCR
  - Used to generate derivative files
- Service Masters are also Tiff images

# derivatives



- resized, reformatted files created from the service masters
  - Optimized for use on the web
  - o Eg. Jpeg or PDF
- In the case of Atiz created ebooks:
  - PDF/A files are derived from the tiff service masters

# next steps



- Optical Character Recognition
  - Usually occurs in the derivative stage
- Metadata creation
- Upload
  - To digital asset management product
    - Eg CONTENTdm, Dspace (clRcle), ebrary etc.

# OCR: optical character recognition



- OCR software translates scanned images into machine-encoded text
- This allows you to search and find words or text in the image
- ABBYY FineReader 10

#### how OCR works



- 1. Analyzes the structure of the scanned image
- 2. Divides the image into blocks (text, tables, images etc.)
- 3. Blocks broken down to characters
- 4. OCR software processes many possible options to find most probable outcome
- 5. Presents best solution.

Replicates "natural" human recognition patterns

#### **OCR** software



#### Takes the scanned image ...

If the water supply is insufficient and the soil parsimonious, this prodigious consumption cannot be supplied, and dessication of organs takes place. This begins in the oldest leaves, and nearly always the little leaves at the base of the stem become soft, flabby, and withered. Analyses have been made which show that these leaves let escape some nitrogenized matter, phosphoric acid and potash, which they contained when living, green and turgescent. Thus the closing of one of these groups of little cell factories by the dessication of a leaf is a very important process to the plant, for less vegetable matter is elaborated than if it had continued its work. In dry years a shortening of the stems and a comparatively small amount of straw results.

If the water supply is insufficient and the soil parsimonious, this prodigious consumption cannot be supplied, and dessication of organs takes place. This begins in the oldest leaves, and nearly always the little leaves at the base of the stem become soft, flabby, and withered. Analyses have been made which show that these leaves let escape some of the nitrogenized matter, phosphoric acid and potash, which they contained when living, green and turgescent. Thus the closing of one of these groups of little cell factories by the dessication of a leaf is a very important process to the plant, for less vegetable matter is elaborated than if it had continued its work. In dry years a shortening of the stems and a comparatively small amount of straw results.

... and turns it to text.

### **OCR** software



what ABBYY can read

- 191 languages
- 36 languages with dictionary and spell check
- 6 alphabets
- Simple chemical formulas

# **OCR** software

what ABBYY	
can't read	

Inhalt

-

Daß der Arbeitgeber zu Unnecht einen Mehrwert vorenthält, folgt aus der Tatsache, daß jede Ware ihren Wert ausschließlich durch die auf ihre Serftellung verwandte Arbeit erhölt.

- 2. Die Bildung des Mehrwertes bedingt insolge der technischen Fortschritte einerseits eine stein sortschreitende Ausbentung der Arbeiter und eine immer größer werdende Jusammenziehung des durch die Produktion getvormenen Genenums in die Hände der Rapitalisten, anderseits eine Berdrüngung der keinern Kopitalisen und Betriebe durch die größern Alffunusation der Bermissen, Beresendung der Rassen, Konzentration der Bermissen.
- 3. Die burch die "Anarchie der Produktion" hervorgerusenen Krisen verfährten diese Emitvidlung: sie vermehren das Heer der Arbeitslosen und das Angebot von billigen Arbeitskräften und beschlennigen den Zusammendruch weniger leiftungsfähiger Navitaliften.
- \$7. Anrge Prafung ber Grundlagen bes Margismus . . . . 91-102
  - I. Die materialistische Geschichtsauffassung, bas ausschließlich bie wirtichaftlichen Berhöltnisse bie gestinge Ordnung der Renichen bestimmen, ist unhaltbar:
    - Neben ben Bebürfniffen bes "unmittelbaren Lebensunterhaltes" bestimmen babon unabhängige geistige Interessen bie Menschen; bieser "geistige Oberban" wirft gestaltenb und ändernd auf die wirtschaftlichen Berhältnisse.
    - 2. Die Menschieltsgeschichte hat durchtweg gang andere treibende Rächte gehabt als die Gorge für dem Lebensunterhalt. Bollends die bedeutsamste Tafjache, die Entstehung des Christentung, ift aus wirtschaftlichen Ursachen nicht zu ernflären.
  - II. Die Bilbung eines unverdienten Mehrtvertes in der fapitag liftischen Production liegt nicht vor.

#### Biertes Rapital: Die Musbreitung bes Margismus in Deutschland 103-134

- - I. Nach Laffalles Tobe läßt die Bewegung des Allgemeinen Deutsiden Arbeitervereins in ihrer Austraft nach:
    - 1. Die Nachfolger Laffalles in ber Leitung bes Bereins find unbedeutende Berfonlichteiten.
    - Durch Gründung von Gewerkschaften sucht v. Schweizer die Betvegung neu zu beleben, während die Fortschrittspartei (hirsch) durch Gründung von Gewerkvereinen wieder Einftuß zu gewinnen sucht.

# let's do it!

