FamilySearch Scanning
(Scanstone)

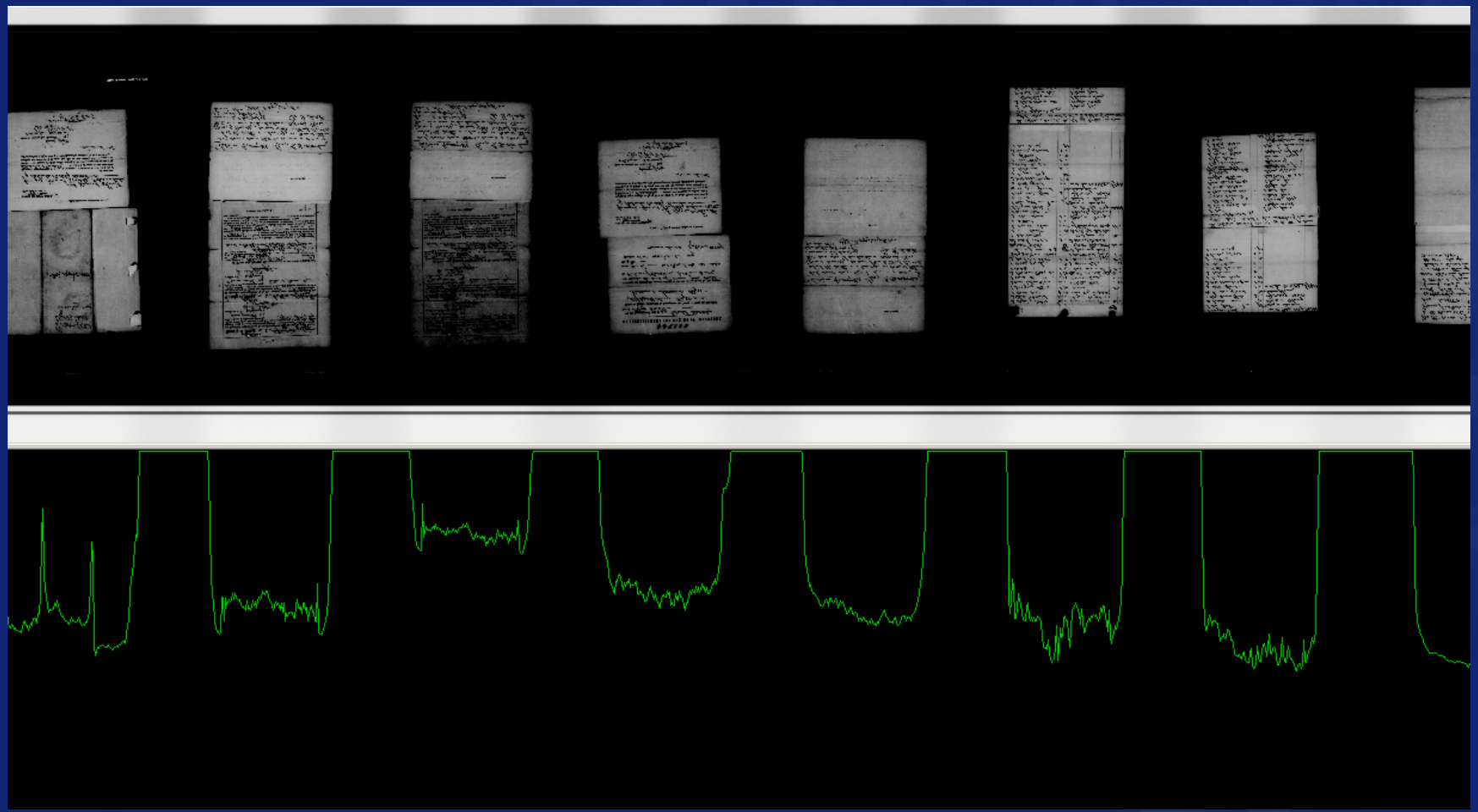
An Automated Exposure Method For Scanning Microfilm

Heath Nielson
nielsonhe@ldschurch.org

Digitizing Microfilm

- Scan as efficiently as possible
  - At 30 minutes a roll, it would require a single scanner 142 years of constant scanning to scan 2.5 million rolls
- Maintain consistent (good) quality
  - Adjust lamp level

Changing light level



Lamp level: 200

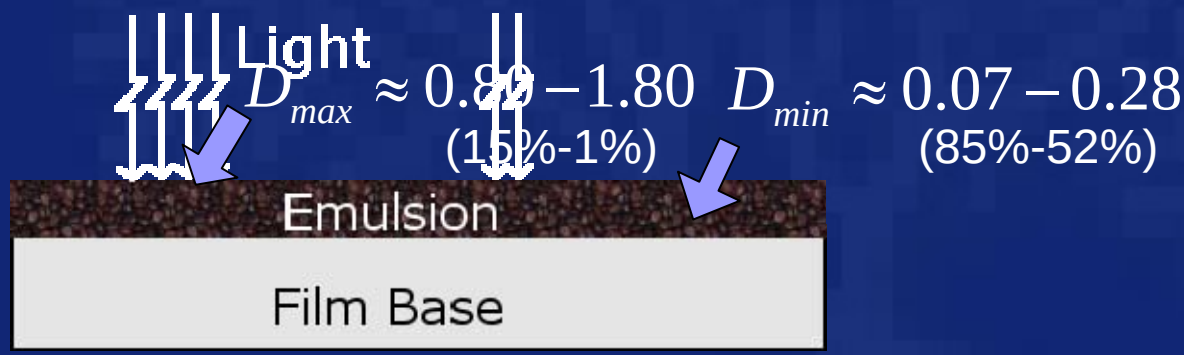
Goal

- To maintain efficiency and quality, we need to automatically identify the optimal lamp setting for a given film
- Optimal lamp level is dependent upon:
  - Film density
  - Focal length
  - Light output

Film Density

- Amount of incident light (I_i) transmitted through the film (I_t) is measured in terms of optical density D

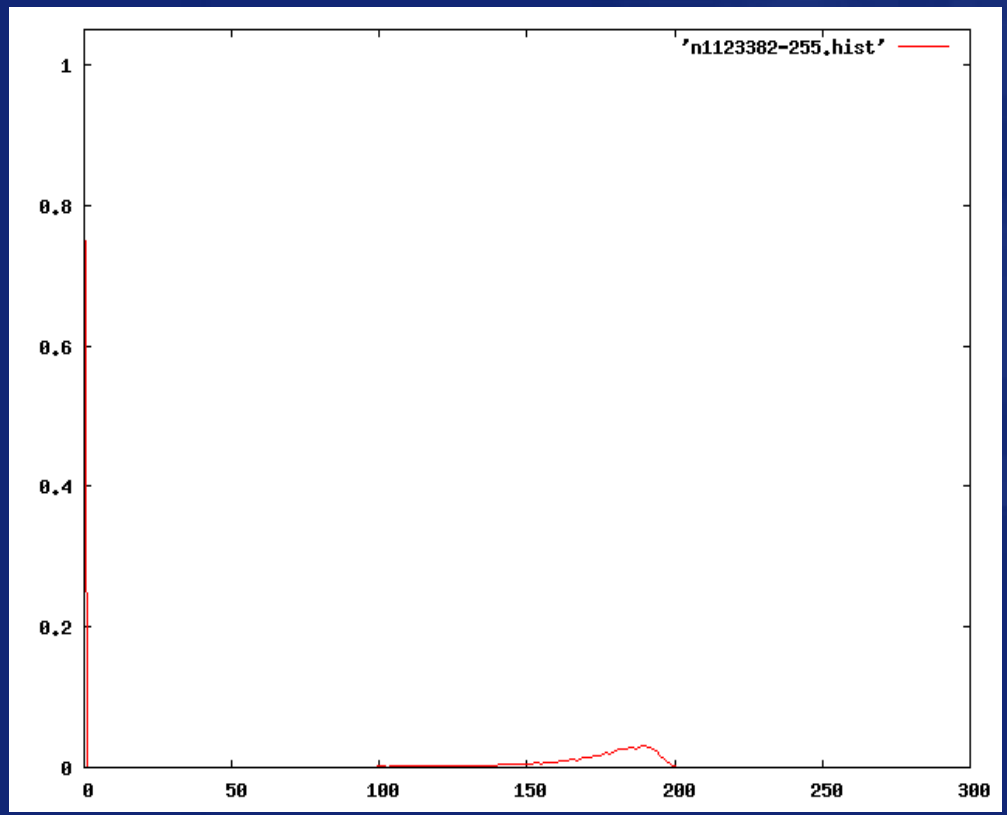
$$D = \log\left(\frac{I_i}{I_t}\right) \qquad 0 \le D \le \infty$$

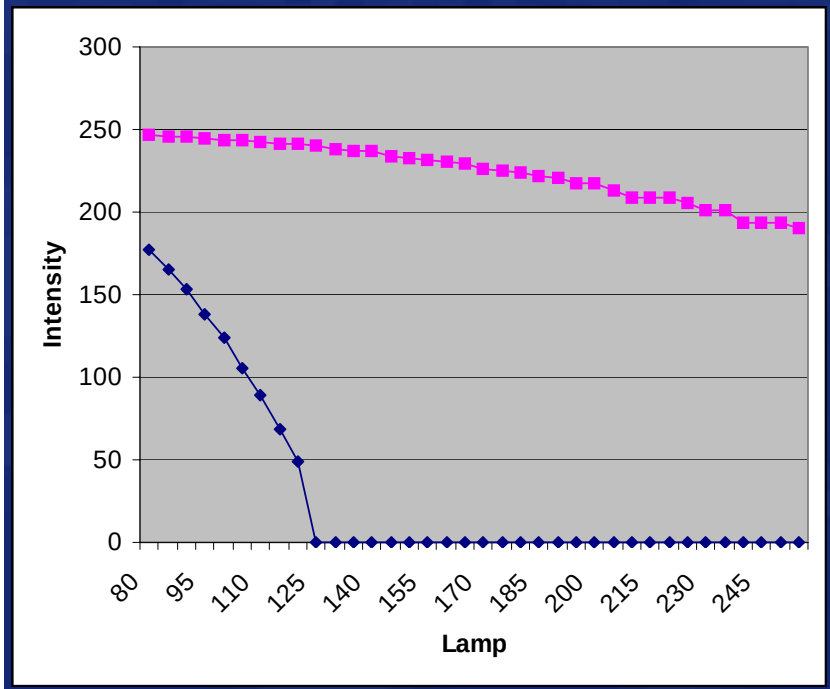


- Film Contrast is dependent upon range between D_{min} and D_{max}

Density vs. Lamp Level

Histogram View





$D_{min} \approx 0.16$

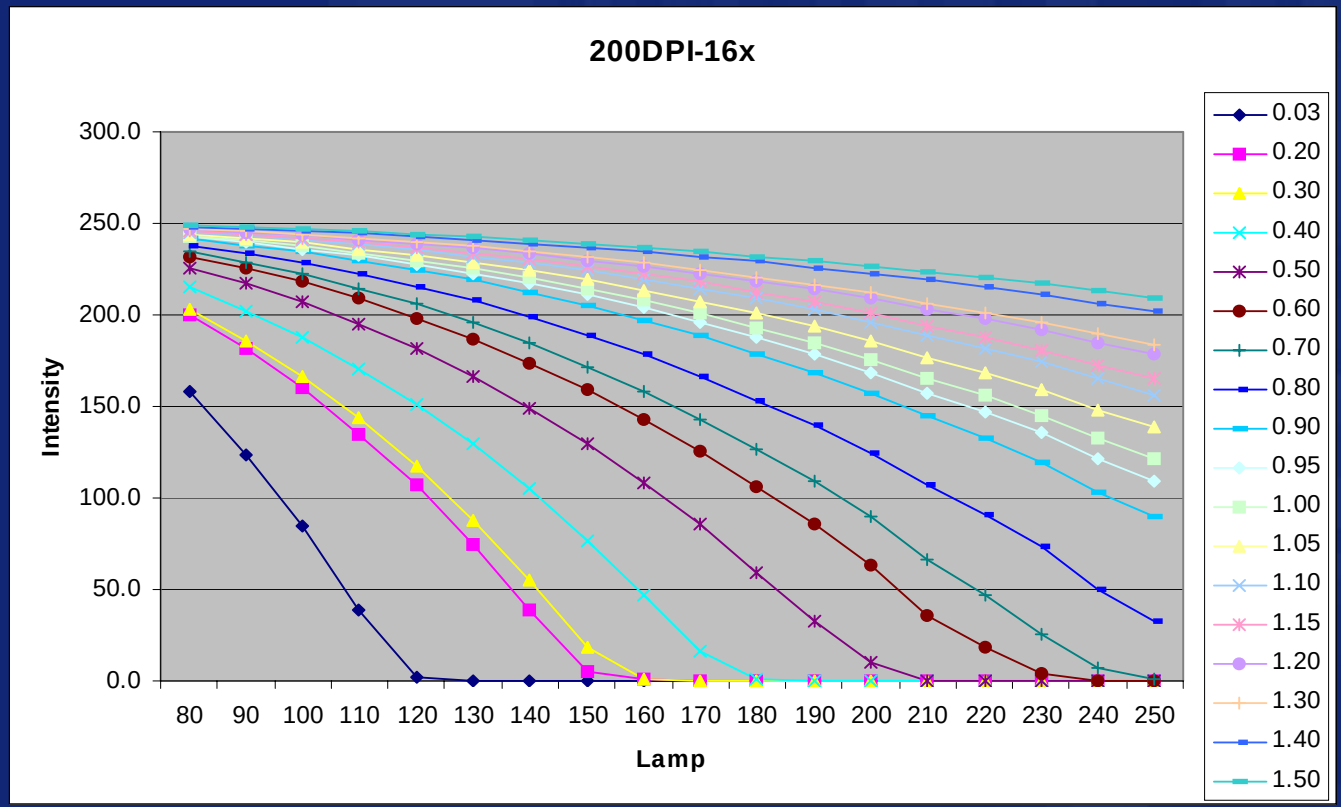
$D_{max} \approx 1.40$

Density Test

- Created test film comprised of 17 sections
  - The density of each section was constant
  - The sections represented a range of densities from 0.03 to 1.50
- Film was scanned multiple times across a range of different lamp levels
- The mean intensity for each density was calculated

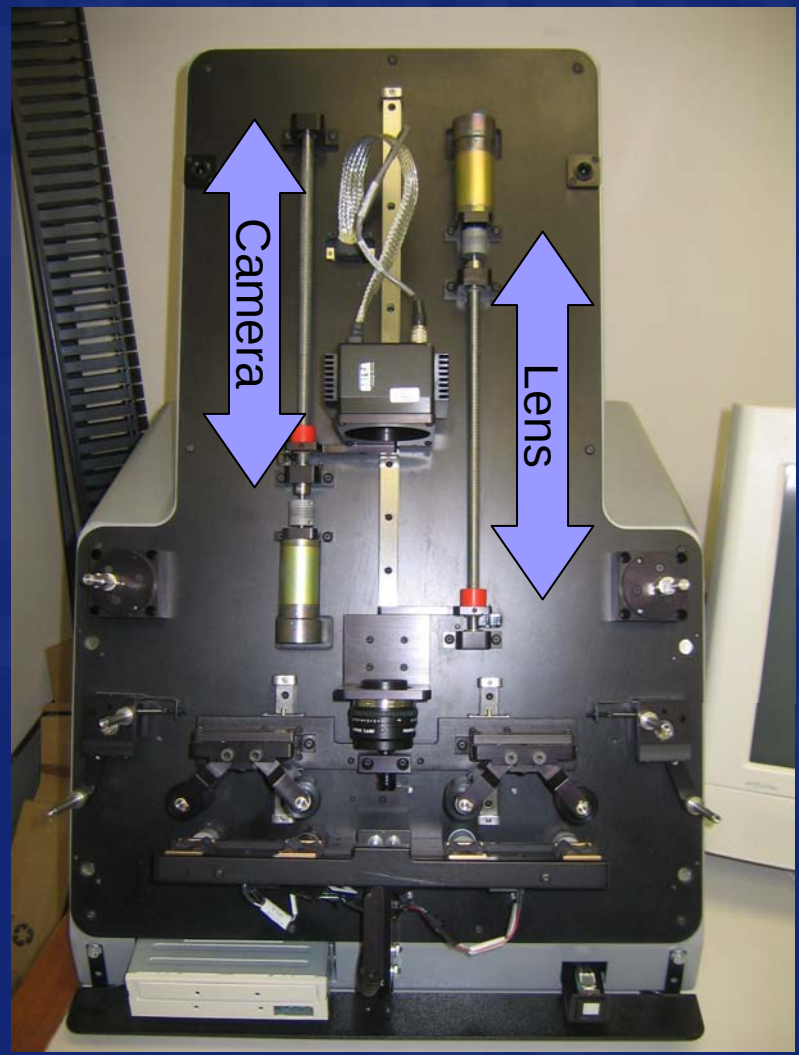
Density Roll

Result of Test Roll

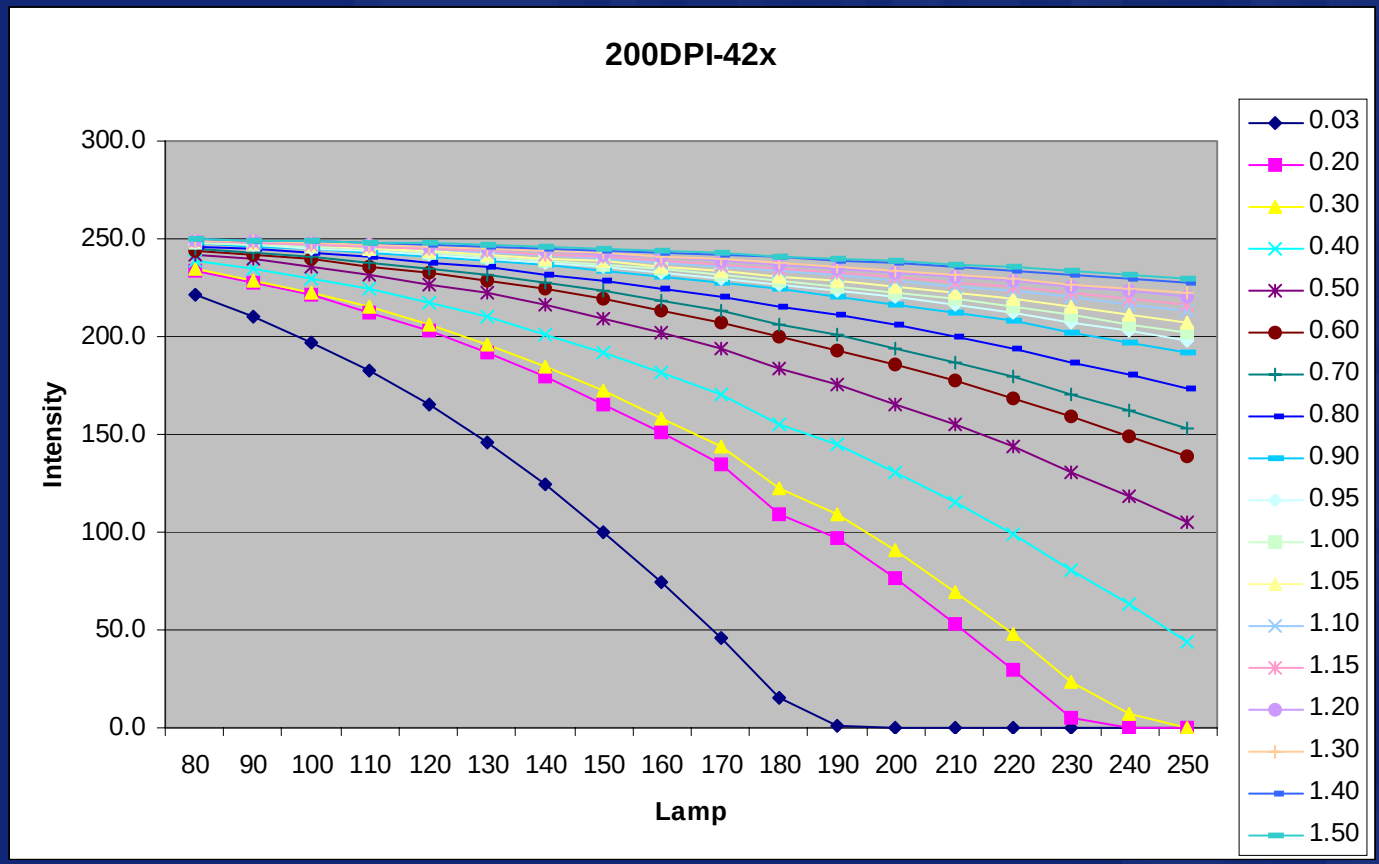


Focal length

- Camera and lens move
  - Affects the amount of light received by the camera
  - Position is determined by the selected DPI and reduction ratio of the film prior to scanning

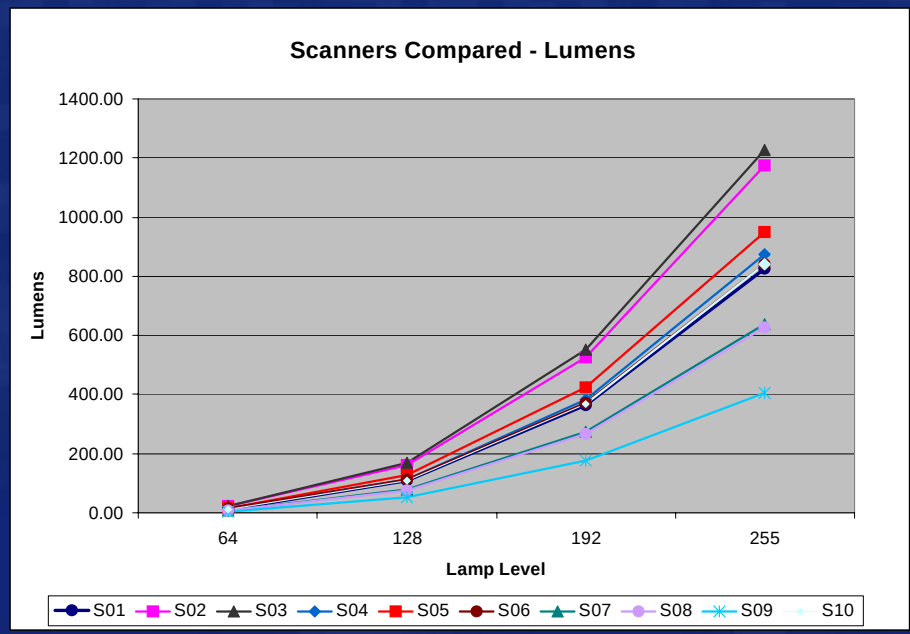


Focal Length vs. Lamp Level



Scanner Light Output

- Light output is inconsistent across scanners
- Mechanical differences
  - Presence of diffuser
  - Out of calibration
  - Age of lamp



Developing an Exposure Strategy

- Input
  - Unknown film density
  - Variable focal length
  - Inconsistent light output across scanners
- Output
  - Inter-frame background and handwriting with an intensity ~0
  - Frame background with an intensity ~200 or greater
- How should the lamp be adjusted?

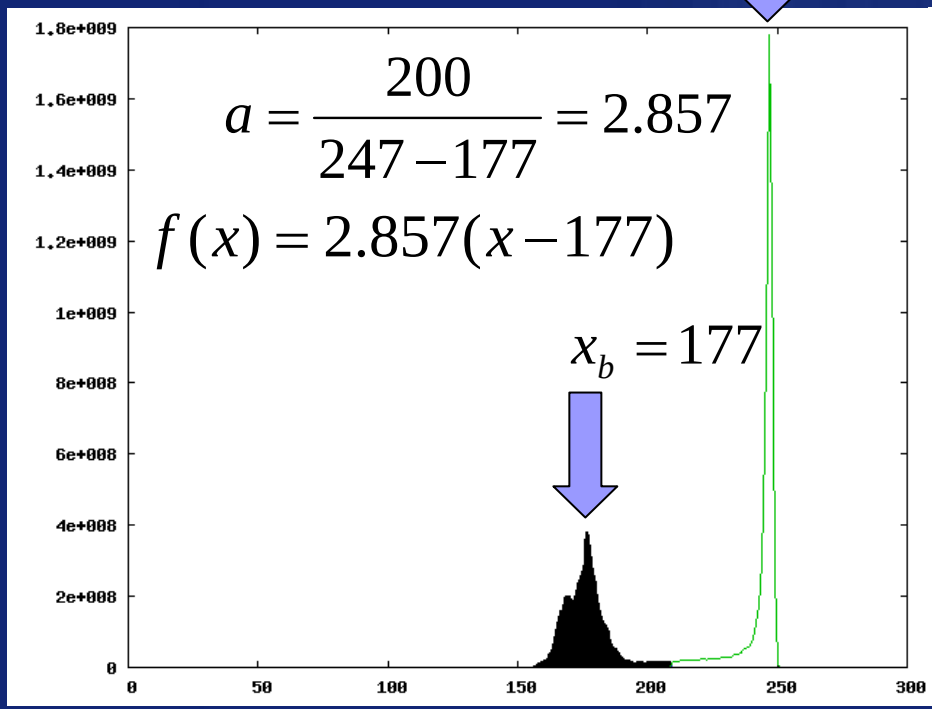
Histogram Stretching

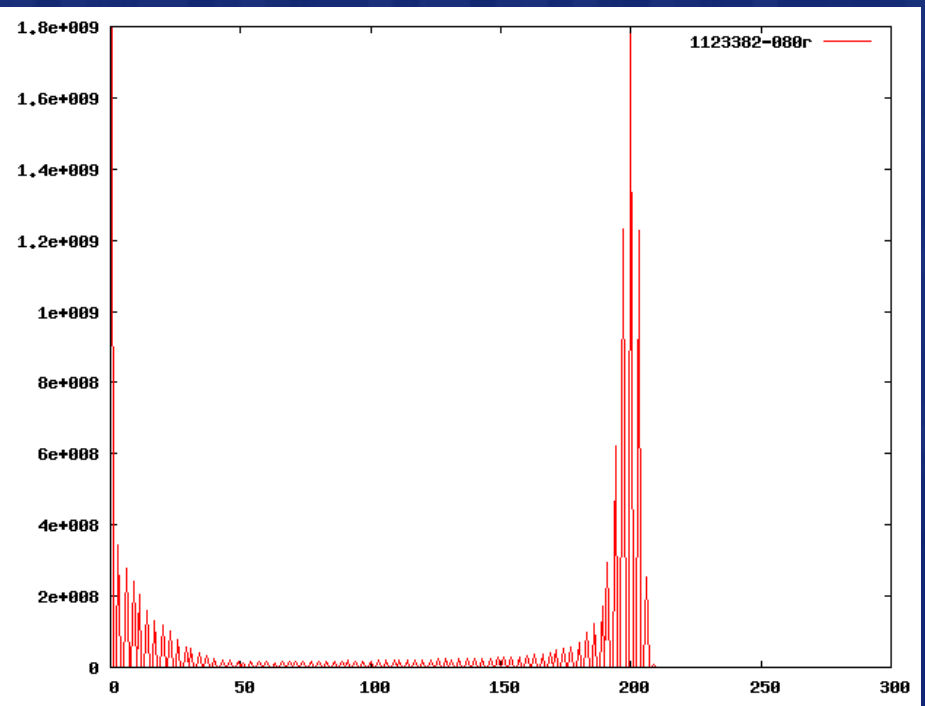
- Set the lamp to an “average” value for the selected focal length
- Scan the film
- Find the mode of the black and white distributions (x_b, x_w)
- Create a stretching function to remap the scanned pixel data

$$y = a(x - x_b) \qquad a = \frac{200}{x_w - x_b}$$

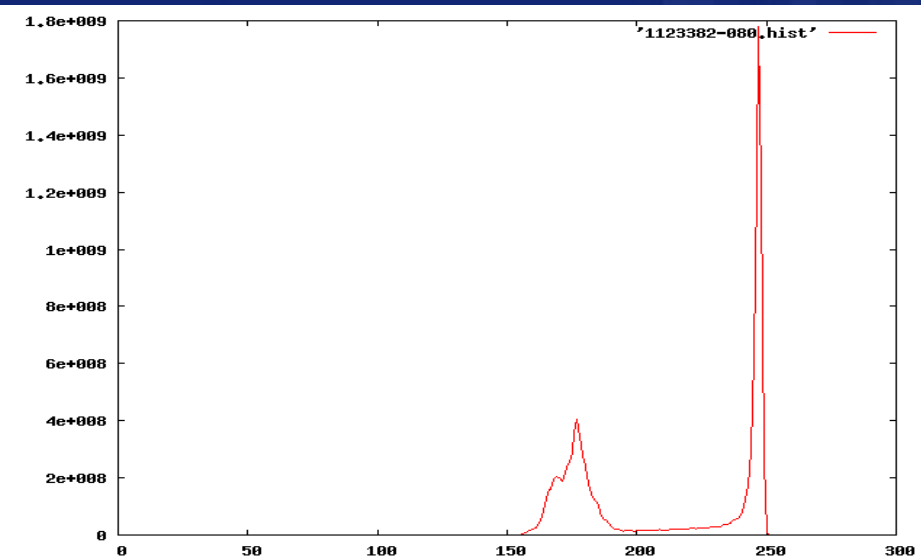
Example

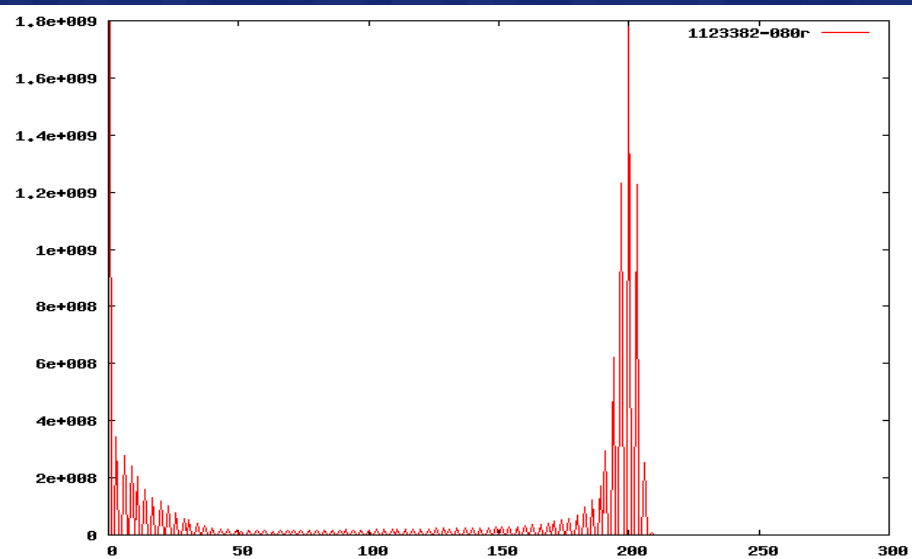
$x_w = 247$

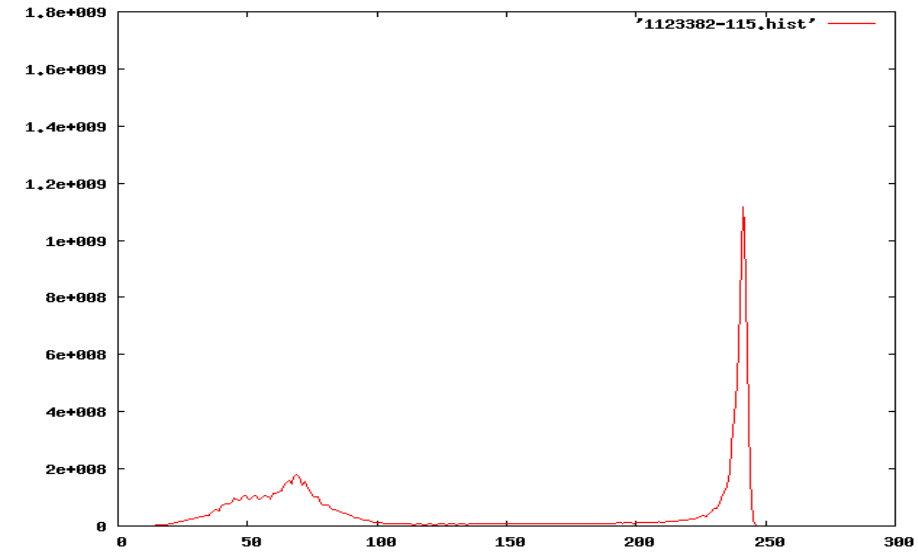


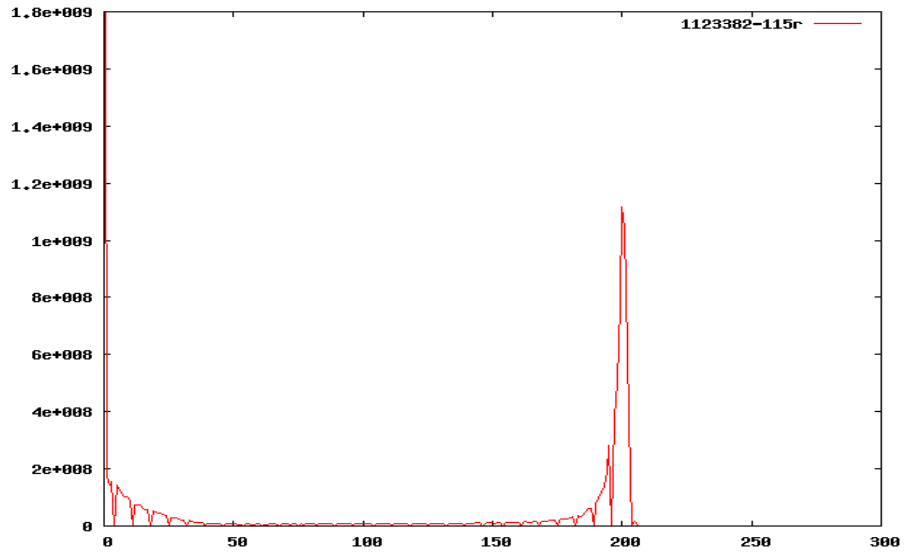


Invariant to minor lamp changes









Potential problems

- Still some dependence upon lamp level
  - Extreme values cause camera saturation
- Assumes a single black or white distribution
  - What value produces acceptable results for a multi-modal distribution?
- Fewer gray levels by stretching the image data post-scan

Manual Audit

- A random sample of 469 films were selected from the 2.5 million rolls
- Each film was scanned and audited for quality
- If quality was deemed unacceptable, the contrast could be overridden

Results

- 185 (39%) films were manually adjusted
  - As the test proceeded, the number of films adjusted decreased
  - In 155 films (84%), the contrast was increased
- Average number of gray levels
  - 187 before audit
  - 169 after audit

Future Directions

- Better metrics for determining a ‘good’ quality image
- Localized histogram stretching
  - Per region
  - Per frame
- 12-bit scanning
  - Decreases further, dependence upon lamp
  - Increases likelihood of capturing full dynamic range between D_{min} and D_{max}

Conclusion

- We have shown an exposure algorithm which produces consistent output in the presence of differing amounts of light and is automated, maintaining the efficiency necessary to scan large numbers of film