



Digital Microfilm Frame Detection

Christopher Nelson

Heath Nielson & Shane Hathaway

The Church of Jesus Christ of Latter Day
Saints

Microfilm Frame Detection



Scanning microfilm is much like taking pictures:

1. Scan a small strip of microfilm
2. Finish the scan in a place that looks like background
3. Look for a document in that strip and save it
4. Repeat



What if the entire microfilm roll was scanned into one extremely large image?



How would frame detection work?

Where are the Documents?

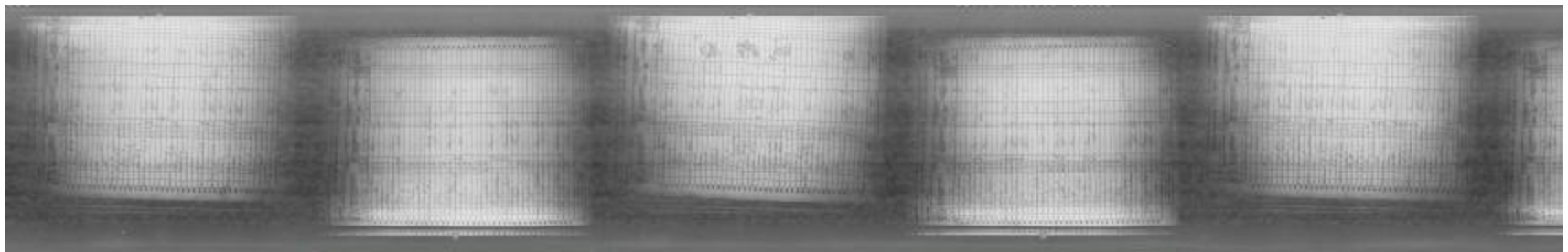


Why Find Documents?

- Saving document images off the film
- Indexing microfilm by document number / location
- Cataloging microfilm contents

Challenges

- Documents do not have consistent size
- Cluttered film / overlapping documents
- Poor microfilm quality / noise
- And much more...



Digital Microfilm Frame Detection



- 1) Generate a Ribbon Profile
- 2) Set the Threshold
 - a.* Generate the “Average Minimum Profile” using a Sliding Window
 - b.* Adjust Threshold to Allow for Gradual Changes
- 3) Mark the Document Segments
- 4) Detect Horizontal Frame Edges
 - a.* Generate Horizontal Profiles
 - b.* Set Thresholds using Histograms
 - c.* Select the Best Results

Ribbon File Format



- Uncompressed 8-Bit Grayscale Image File
- Millions of Pixels Long
- Average File Size: 20 – 30 Gigabytes
- Encoded as a Eight Level “Hierarchal Pyramid”



Frame Detection
Runs on the 5th Level

Generating the Ribbon Profile



Each pixel has an intensity value which ranges from 0 (pure black) to 255 (pure white)

Profile: sum of these values for each column



Documents = High Profile Values

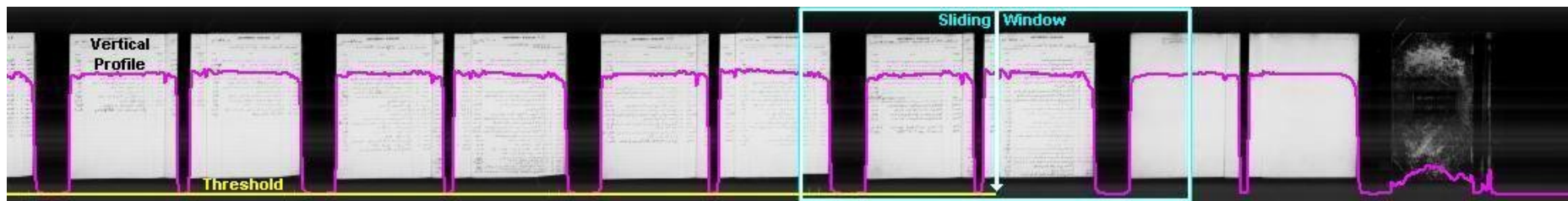
Background = Low Profile Values

Setting the Threshold



Threshold: dividing line between document and background profile values

- 1) Generate the “Average Minimum Profile” using a Sliding Window



- 2) Adjust Threshold to Allow for Gradual Changes



Marking Document Segments



Left and right document edges are found where threshold and profile values match

Ribbon segments containing documents occur where the profile lies above the threshold



Detecting Horizontal Frame Edges



1) Generate Two Ribbon Profiles

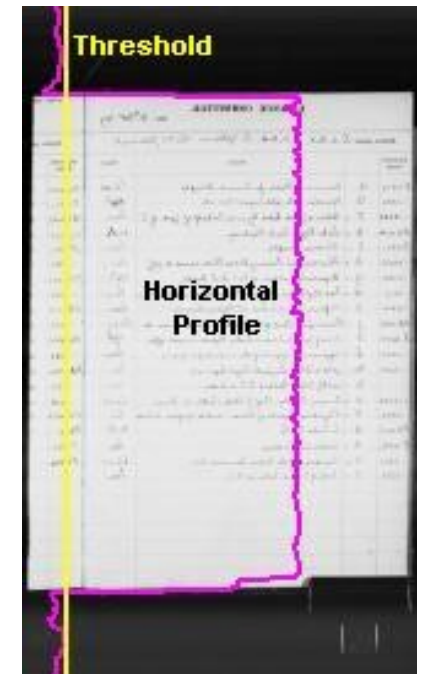
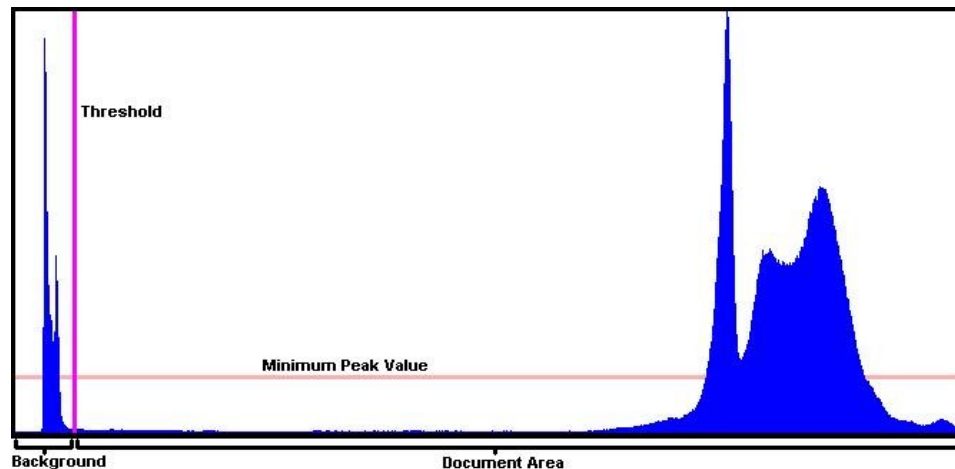
- Horizontal Pixel Intensity – sum of pixels in each row
- Horizontal Pixel Variance – variance for each row of pixels

2) Set Threshold using Histograms

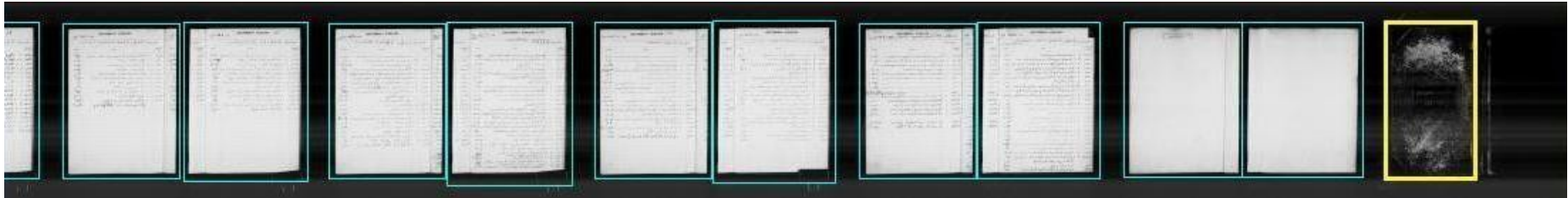
- Compute a “minimum peak value”
- Find the minima after first group of peaks

3) Select the Best Results

- Choose the one which creates the largest frame



Frame Detection Demonstration



- 1) Generate a Ribbon Profile**
- 2) Set the Threshold**
 - a.* Generate the “Average Minimum Profile” using a Sliding Window
 - b.* Adjust Threshold to Allow for Gradual Changes
- 3) Mark the Document Segments**
- 4) Detect Horizontal Frame Edges**
 - a.* Generate Horizontal Profiles
 - b.* Set Thresholds using Histograms
 - c.* Select the Best Results

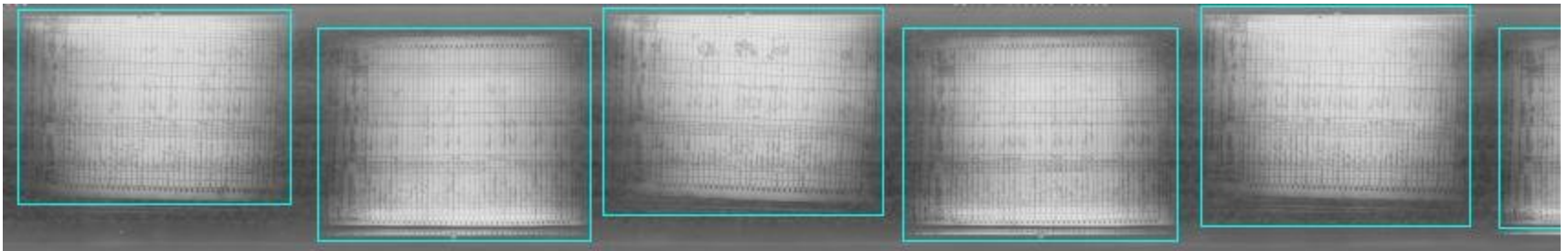
How Well Does this Work?



Accuracy Based on Microfilm Quality

- 91 Good Films: 99.86%
- 17 Fair Films: 99.47%
- 12 Poor Films: 94.36%

For Example...



We've Got Frames, Now What?



Improving Frame Detection

- Detecting Reverse Polarity Frames
- Finding Rotation / Mirroring Problems
- Separating Overlapping Frames

Uses for “Framed” Document Images

- Automatically Identifying the Contents of Frame
- Cataloging / Indexing Microfilm Ribbons
- Saving Document Images for Later Use
- Measure Microfilm, Frame, or Document Quality

Questions

