Improved Blur Detection of Historical Document Images Using a Neural Network

Family History Technology Workshop
Brigham Young University
February 27, 2018

Ben Baker
bakerb@familysearch.org

Presented by Mike Wynn

FamilySearch

Background

Built upon work presented at 2012
Family History Technology Workshop

"Blur Detection of Historical Document Images"

https://fhtw.byu.edu/static/conf/2012/baker-blur-fhtw2012.pdf

Was back on Imaging team for part of 2017 specifically to work on this project again, but now on Automated Content Extraction (ACE) team

FamilySearch

Review of Previous Work

- FamilySearch captures hundreds of millions of images annually using cameras at hundreds of sites throughout the world
- Images are audited by humans for image quality problems
- Majority of all image audit quality control failures are <u>still</u> due to blur and out of focus images (same was true back in 2011)

FamilySearch

Review of Previous Work

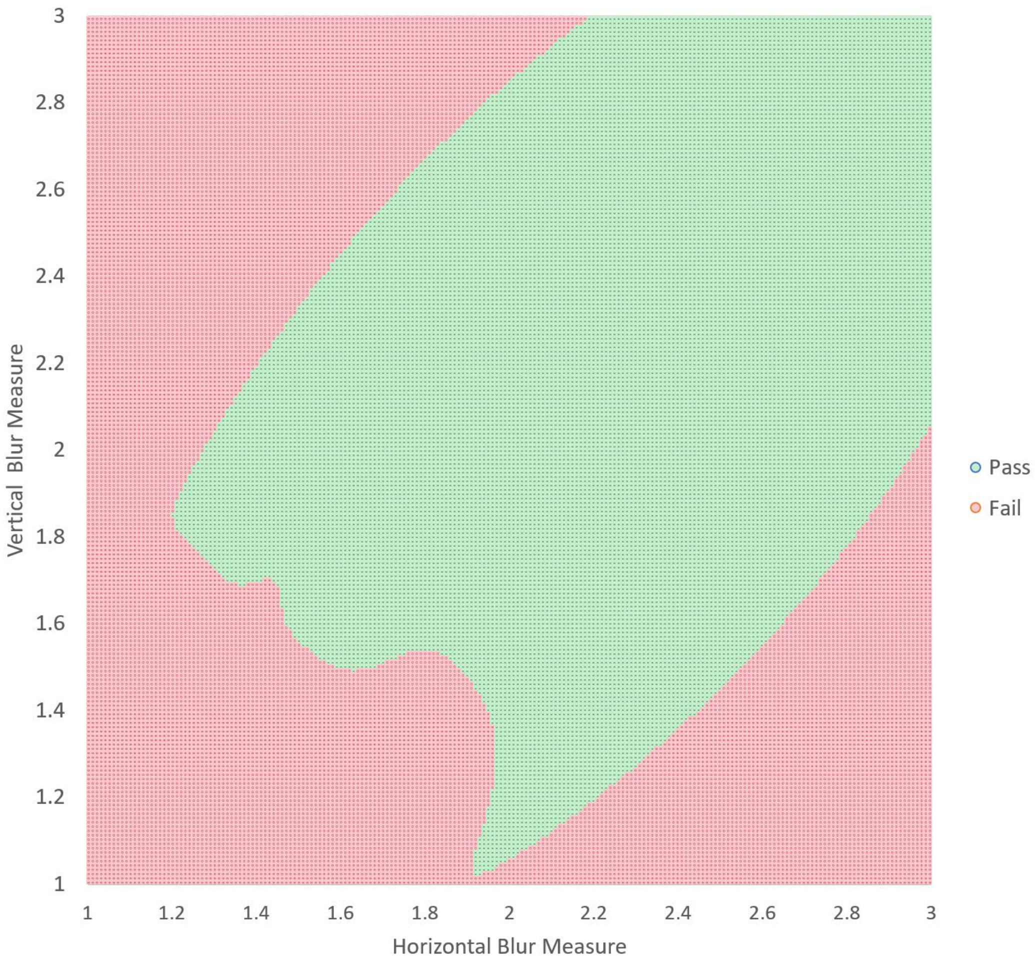
- Determine sharpness of edges of document text by measuring fit to logistic function
- Look at vertical and horizontal edges separately to detect directional motion blur
- Simple threshold could detect 81% of failed images and 84% of passing images correctly

Improvements in 2013

- Alan Cannaday improved results by:
  - Introducing M-shift when fitting edges to the logistic function
  - Training a Gaussian Mixture Model (GMM) on labeled images
- Accuracy improved as high as 89.2% for certain image sizes
- Still not good enough for camera operators, though, so feature was turned off

FamilySearch

GMM Visualization



GMM Implementation Results

Classification	Percent Correctly Identified
Blank	23.19%
In Focus	90.97%
Horizontal Blur	48.6%
Vertical Blur	50.25%
Out of Focus	76.86%

Overall Performance	
Weighted Precision-like score	87.44%
Weighted Recall-like score	57.79%
F-score	69.59%

- Precision-like score – weighted percent of correctly identified images
- Recall-like score – weighted percent of blurry and out of focus images that were correctly identified as “bad”
- Weights correspond to frequency of occurrence of each image class across all captured images

FamilySearch

Improved Methodology

- Expanded from two to five different classes
  - “Good”
    - Blank
    - In Focus
  - “Bad”
    - Horizontal Blur
    - Vertical Blur
    - Out of Focus
- Trained multi-layer perceptron classifier on 2,331 images mostly labeled by human auditors
- Used Weka to produce weights that were manually used in C++ imaging library code to determine classification
- Tested on 775 labeled images

FamilySearch

Neural Network Topology

- Fully connected neural network
- Edge sharpness metrics and image characteristics as features

  (120 features total – see paper for details)
- One hidden layer with 62 nodes
- Five output nodes

FamilySearch

Neural Network Implementation Results

Classification	Percent Correctly Identified
Blank	88.24%
In Focus	96.39%
Horizontal Blur	71.43%
Vertical Blur	73.58%
Out of Focus	55.56%

Overall Performance	
Weighted Precision-like score	95.86%
Weighted Recall-like score	67.31%
F-score	79.09%

FamilySearch

Comparing Approaches

- Neural network results are when prediction matches class exactly
- DCam camera capture software alerts operators when images need to be retaken
- Consider matching “good” and “bad” like the GMM classifier did
  - “Good” – One of Blank or In Focus for these two classes
  - “Bad” – One of Horizontal Blur, Vertical Blur or Out of Focus for these three classes

FamilySearch

Comparing Percent Correctly Identified as “Good”/“Bad”

Classification	GMM Implementation	Neural Network Implementation
Blank	23.19%	94.12%
In Focus	90.97%	97.53%
Horizontal Blur	48.6%	83.93%
Vertical Blur	50.25%	84.91%
Out of Focus	76.86%	79.63%

Comparing Precision and Recall

Overall Performance	GMM Implementation	Neural Network Implementation
Weighted Precision-like score	87.44%	97.30%
Weighted Recall-like score	57.79%	82.95%
F-score	69.59%	89.55%

- Precision-like score – weighted percent of correctly identified images
- Recall-like score – weighted percent of blurry and out of focus images that were correctly identified as “bad”
- Weights correspond to frequency of occurrence of each image class across all captured images

Future Plans

- More labeled images for training and testing
- Add one or more classes to deal with bleed out, rubber stamps or other problematic images
- Using a deeper neural network with a different activation function such as Relu instead of the sigmoid function
- Using a Convolutional Neural Network (CNN) based approach


Q&A