The Common Fold: Utilizing the Four-Fold to Dewarp Printed Documents from a Single Image

Sagnik Das, Gaurav Mishra, Akshay Sudharshana, Roy Shilkrot
{sadas, roys} @cs.stonybrook.edu
ACM DocEng '17, Valletta, Malta

Stony Brook University

Introduction

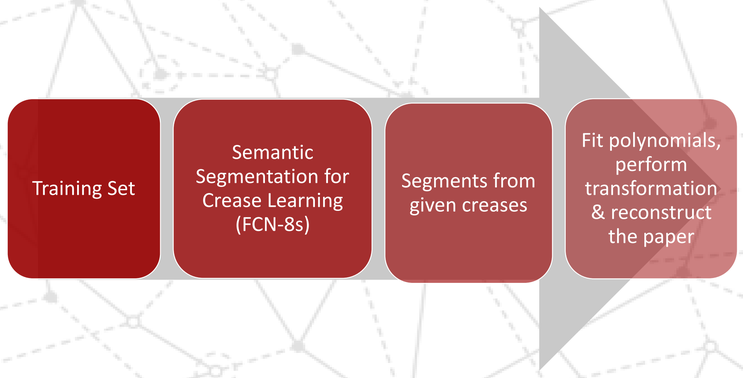
In many cases paper documents are scanned after being folded or warped, exhibiting non-uniform geometric deformation that renders digitization challenging and severely cripples OCR performance.

We propose **a deep learning based de-warping approach** utilizing the half-fold, using a convolutional neural network combined with piecewise polynomial regression.

Motivations

- A de-warping technique without calibrated hardware [1]
- No dependencies on content [2].
- Work with a single image [3].

Dewarping System Description



• Dataset Creation

Created a synthetic realistic dataset of more than 12,000 ground-truth images using Blender.

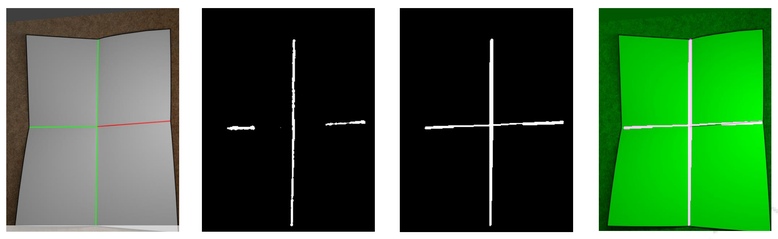
To resemble realistic images, we considered -

a. Random camera angles.
b. Random lightings.
c. Random geometric deformations.

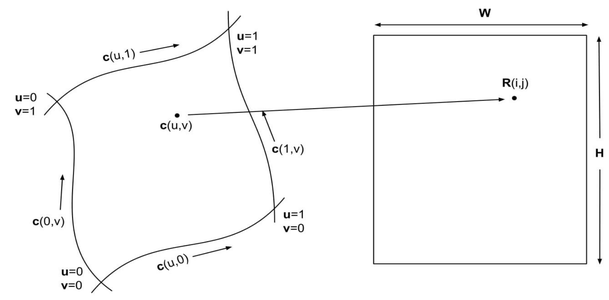
Training Images

• Semantic segmentation for learning creases

a. We learn the parameters of folding i.e. **the creases** by semantic segmentation on two classes namely, non-crease and crease.
b. Used pre-trained FCN-8s [4] trained on our own dataset.
c. As the preprocessing step we used Local Contrast Normalization
d. Probabilistic Hough transform and simple 2D line geometry is used as post-processing.



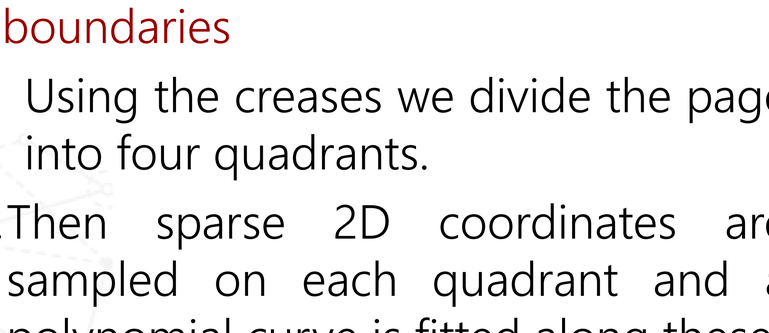
Ground-truth | Detected Crease | Post-processed | Infused



Representation of Coons Patch. $\boldsymbol{c}(u,v)$ represents an interior point, where $0 \le u,v \le 1$; R(i, j) is an interior pixel in the mapped quadrilateral

• Polynomial Fitting on segment boundaries

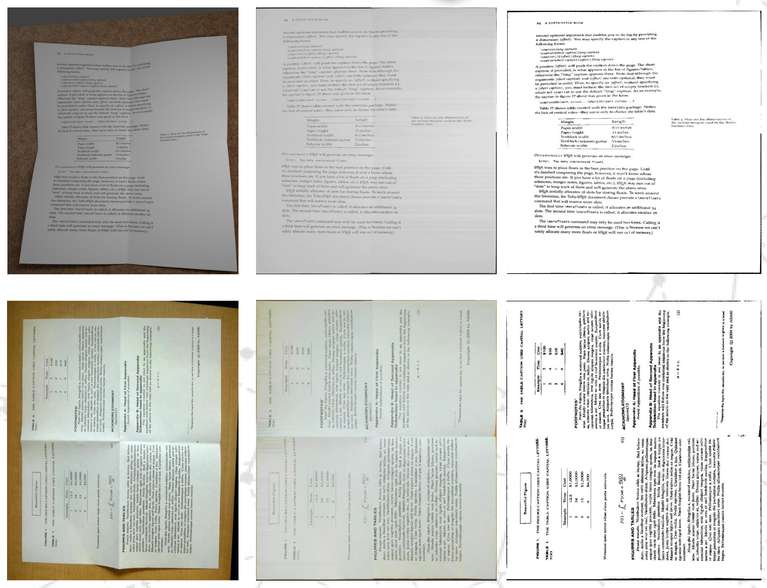
a. Using the creases we divide the page into four quadrants.
b. Then sparse 2D coordinates are sampled on each quadrant and a polynomial curve is fitted along these.
c. Given four polynomials and the four corner intersection points, we formulate a Coon's Patch .
d. Each pixel $c(u,v)$ is mapped to a pixel in R by computing the two parameters u and v s.t. $(u,v) = \left(\frac{i}{H}, \frac{j}{W}\right)$, where H and W is the height and width of R.



One segment | Fitted polynomials | Flattened segment

• Merging and Reconstruction

a. Find best matching seam between segments.
b. Blend the segments on the matching seam.
c. We do luminance correction [5] to tackle non-uniform luminance.



Original | Dewarped | Thresholded

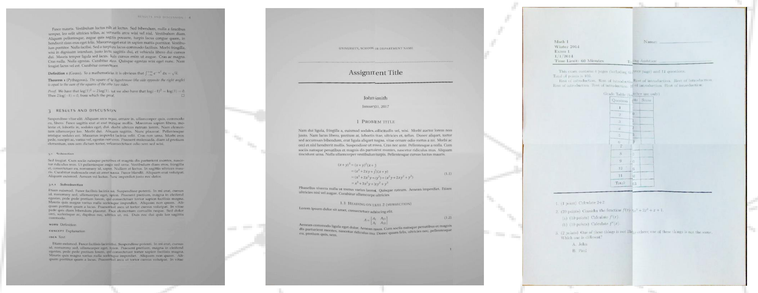
Results

- Performed OCR analysis using pytesseract (v0.1.7)
- 20 document images of 1200x1600 resolution from a synthetic test set with total 3260 words.
- Baseline OCR accuracy is obtained on original document images before folding.

Method	% Words Correct
Our Method	56.44 ± 14
Perspective Rectification	33.89 ± 20
No rectification	20.08 ± 9
Baseline	69.94 ± 15

Limitations

- Half-fold assumption
- Crease detection
- 2D boundary method
- Segment merging



Failed Examples

Future Work

- Evaluate applicability on a mobile device.
- Address the existing limitations.

References

1. Li Zhang, Yu Zhang, and Chew Tan. 2008. An improved physically-based method for geometric restoration of distorted document images. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 4(2008), 728–734.
2. Changsong Liu, Yu Zhang, Baokang Wang, and Xiaoqing Ding. 2015. Restoring camera-captured distorted document images. International Journal on Document Analysis and Recognition (IJDAR) 18,2(2015),111–124.
3. Shaodi You, Yasuyuki Matsushita, Sudipta Sinha, Yusuke Bou, and Katsushi Ikeuchi.2017. Multiview Rectification of Folded Documents. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).
4. Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.3431–3440.
5. Michael S Brown and Y-C Tsoi. 2006. Geometric and shading correction for images of printed materials using boundary. IEEE Transactions on Image Processing 15, 6 (2006), 1544–1554.