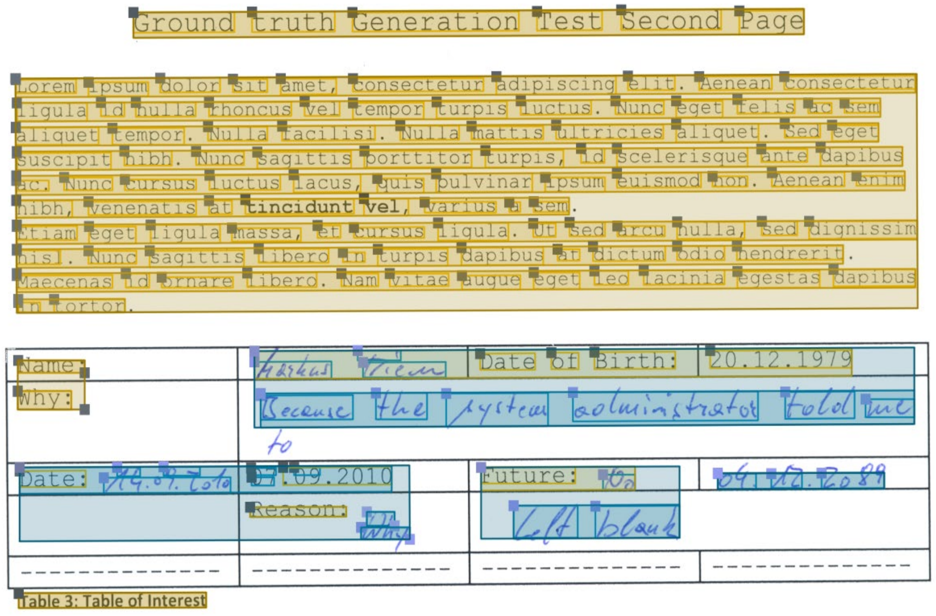


DIA: Preprocessing

2VU 183.628, 3ECTS066 Visual Computing, SS2025

Robert Sablatnig, Florian Kleber,
Rafael Sterzinger

Excercises

- Reminder
  - 19.04.2025 deadline exercise part 1
  - 28.04.2025 exercise presentations 1
  - Contact Group member
  - Hints & Remarks -> TUWEL

  Presentation:

  5 min.

  shortly present your results + main findings

  Report:

  3-5 pages

  present and discuss methodology based on results (advantages / disadvantages)

  results + conclusion

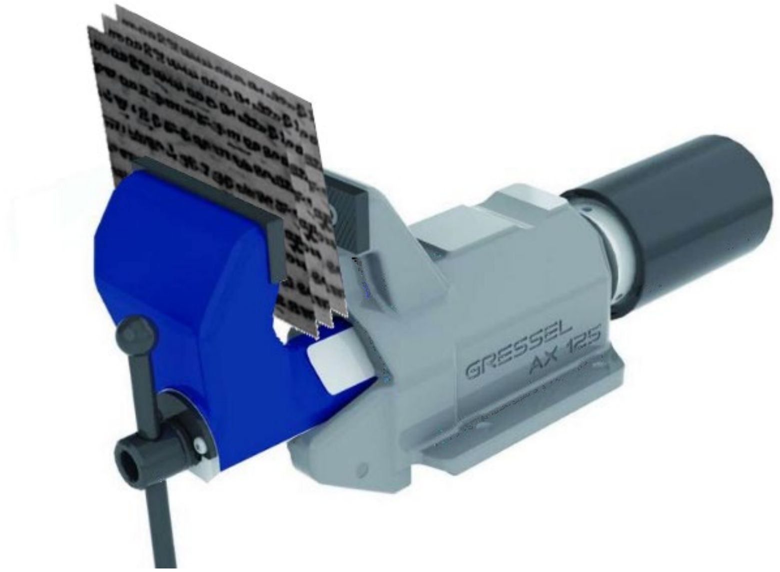
  General Questions:

  TUWEL forum

Statistical Analysis for MSI

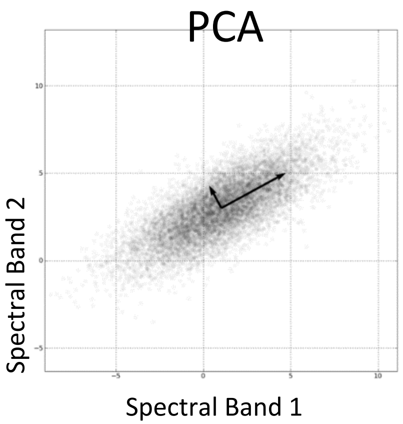
Statistical Analysis - Recap

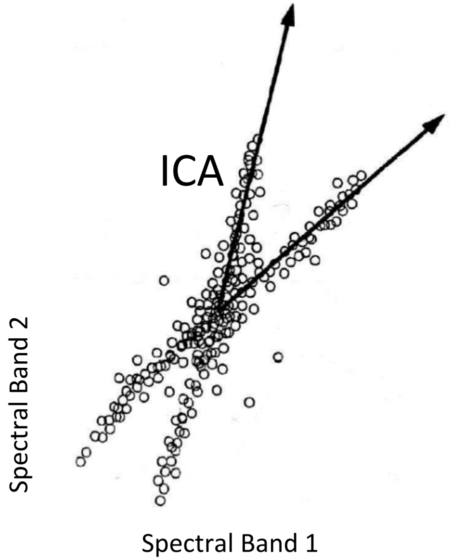
- MSI dataset:
  - Lots of layers
  - Hard to interpret/visualize
  - Much correlation & redundancy
- Goals:
  - 'Compress' the information contained in an MSI set
  - Isolate sources (e.g. support, inks, pigments)



Methods - overview

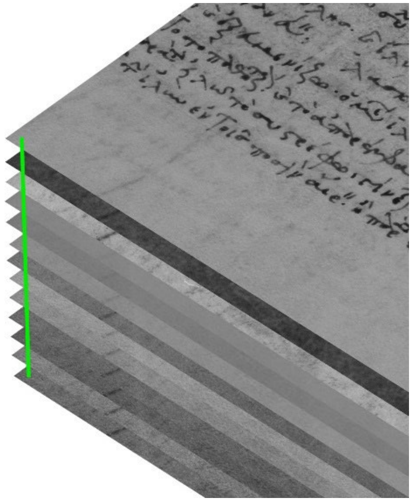
- Unsupervised methods
  - Find structure in data, without additional input
  - E.g.: Principal/Independent component analysis
- Supervised methods
  - Classical machine learning: provide algorithm with examples of classes (background, inks, pigments,…)
  - E.g. Linear Discriminant Analysis (LDA), (deep) neural networks





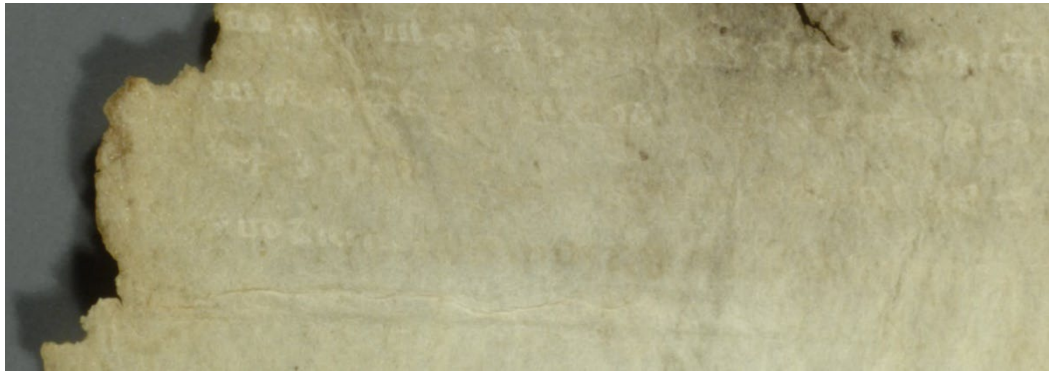
$$\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

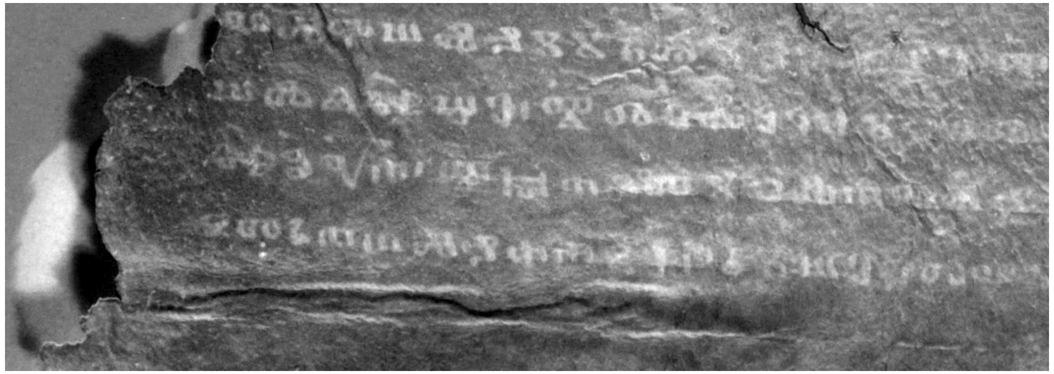
example for „background“



Unmixing

- Explicitly estimate
  - Sources, also called endmembers
  - Abundance fractions
- Prominent field in remote sensing
- Popular algorithm:
  - Vertex Component Analysis (VCA)
    - => Unsupervised





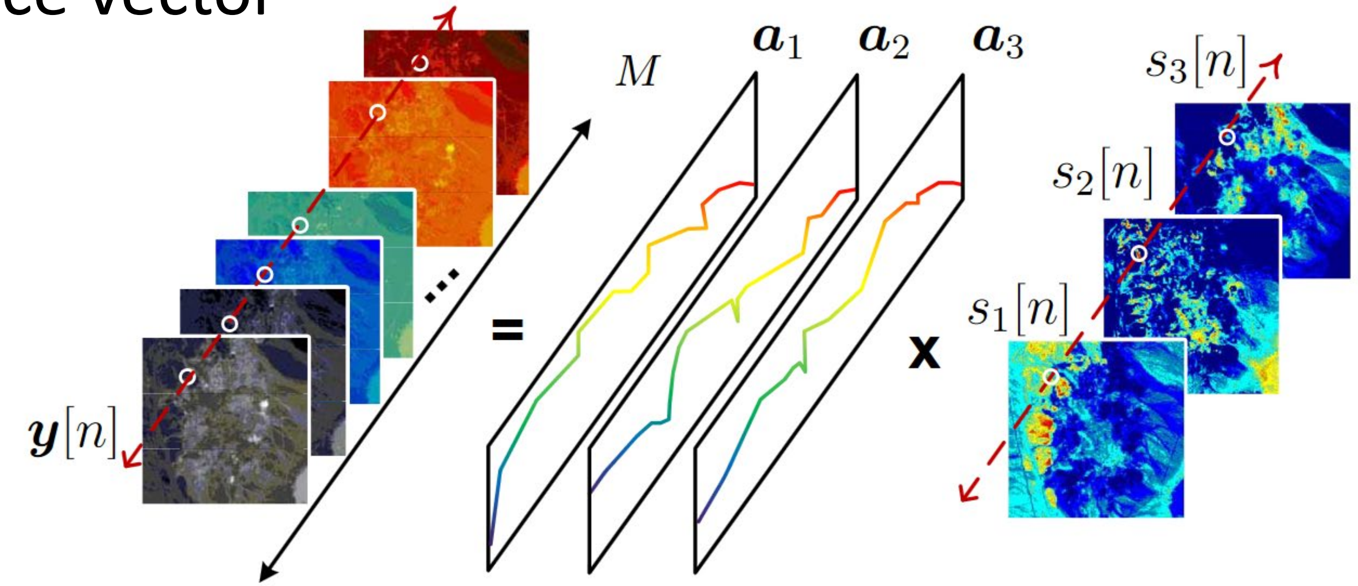
Linear Mixing Model – Example

- **Cocktail-Party Problem**:
- Two persons speaking with each other - $s_1(t)$, $s_2(t)$
- Two microphones - $x_1(t)$, $x_2(t)$
- Problem: separate the mixed signals

$$x_1(t) = a_{11}\, s_1 + a_{12}\, s_2$$
$$x_2(t) = a_{21}\, s_1 + a_{22}\, s_2$$

Linear Mixing Model

- $y(n) = A * s(n)$
  - where $y(n)$ is a spectral vector at pixel *n*
- $A = [a_1 \ldots a_N]$ is endmember signature
- $s[n]$ is abundance vector



Multispectral data Signature Abundance maps

Palimpsest Example

- Combine all three measurements the extract the three classes
- Determine the mean grayvalue of the 3 classes
  - Parchment
  - Overwriting
  - Underwriting
- Assumption: Gray values (Intensities) „add“ to produce the colors

	Parchment	Overwriting	Underwriting
Red	R1	R2	R3
Green	G1	G2	G3
Blue	B1	B2	B3

Endmembers

Signature

Palimpsest Example

- Assumption: Gray values (intensities) „add“ to produce the colors
- The product of the matrix and a vector of classes is the output color

	Parchment	Overwriting	Underwriting
Red	R_1	R_2	R_3
Green	G_1	G_2	G_3
Blue	B_1	B_2	B_3

Underwriting only

$$\begin{pmatrix} R_1 & R_2 & R_3 \\ G_1 & G_2 & G_3 \\ B_1 & B_2 & B_3 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} R_3 \\ G_3 \\ B_3 \end{pmatrix}$$

	Parchment	Overwriting	Underwriting
Red	R_1	R_2	R_3
Green	G_1	G_2	G_3
Blue	B_1	B_2	B_3

Pixel include equal amounts of Over- and Underwriting

$$\begin{pmatrix} R_1 & R_2 & R_3 \\ G_1 & G_2 & G_3 \\ B_1 & B_2 & B_3 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} R_{O+U} \\ G_{O+U} \\ B_{O+U} \end{pmatrix}$$

Palimpsest Example

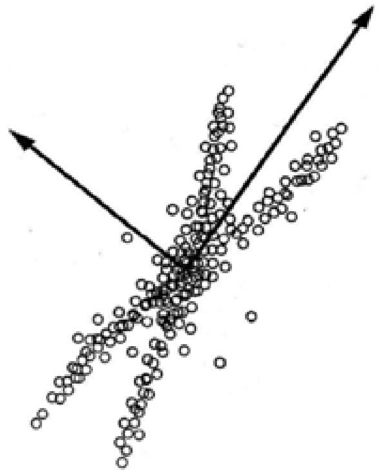
- Inverse Operator
  - Find the matrix that, when applied to (R,G,B) values of pixel, calculates the percentage of each „class“ of object (parchment, overwriting, underwriting)
- This is the „Inverse Matrix“

$$A = \begin{array}{c|ccc} & \text{Parchment} & \text{Overwriting} & \text{Underwriting} \\ \hline \begin{array}{c}\text{Red}\\ \text{Green}\\ \text{Blue}\end{array} & \multicolumn{3}{c}{\begin{pmatrix} R_1 & R_2 & R_3 \\ G_1 & G_2 & G_3 \\ B_1 & B_2 & B_3 \end{pmatrix}} \end{array}$$

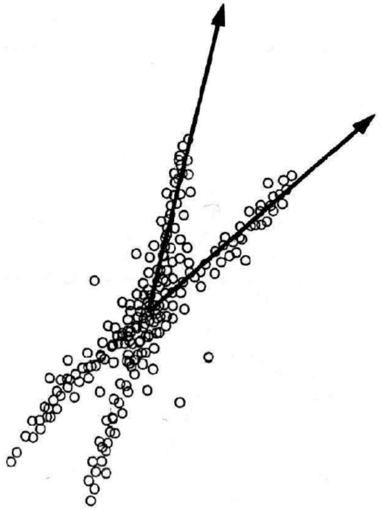
$$A^{-1} = \begin{array}{c|ccc} & \text{Red} & \text{Green} & \text{Blue} \\ \hline \begin{array}{c}\text{Parchment}\\ \text{Overwriting}\\ \text{Underwriting}\end{array} & \multicolumn{3}{c}{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}} \end{array} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} P \\ O \\ U \end{pmatrix}$$

Blind Source Separation Techniques:

- Principal Component Analysis (PCA):
  - Decomposes observations in k (k ≤ m) **uncorrelated** signals

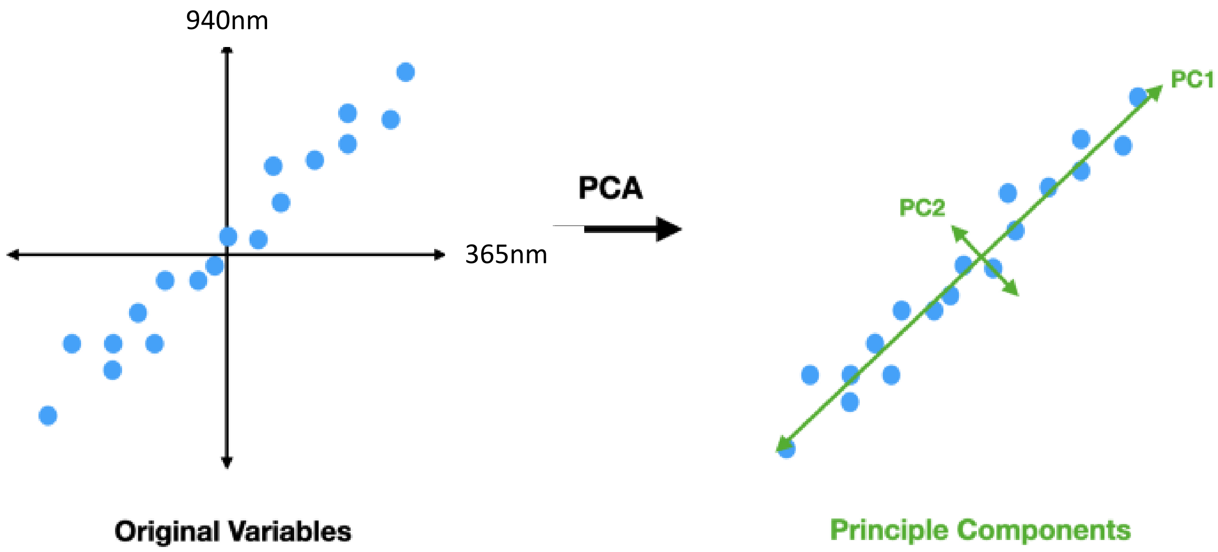


- Independent Component Analysis (ICA):
  - Decomposes observations k **independent** signals
  - Especially useful if there are more than 2 dominant sources (= different inks)



Principal Component Analysis (unsupervised)

- A general, widely used method for all kinds of data (not just MSI!)
- Decomposes observations to **uncorrelated** variables ("orthogonal")
- Principal components are **ordered** by the variance in the data that they explain
- → most information is "compressed" in the first *k* components



Simple example: input is a multispectral image with two layers (365nm and 940nm). Each point represents a pixel, its position is given by the intensity measured in either of the wavelengths.
After PCA, the points are transformed to a new coordinate system, where PC1 represents most of the variance in the data. PC2 is orthogonal to PC1 and contains the remaining variance

Illustrative Example of PCA

- **Two-Band Image**, e.g., image under red light and blue light
- Pixels from two object classes A, B: denoted in histograms by different symbols (circle ●, triangle ▲)
- Histogram: graph of pixel population vs. gray value
  - estimate of probability of each gray value

Image Histogram, Band 1

- Graph of probability of pixel gray values
- Pixels in Class “A” not distinguished from those in Class “B” by gray value

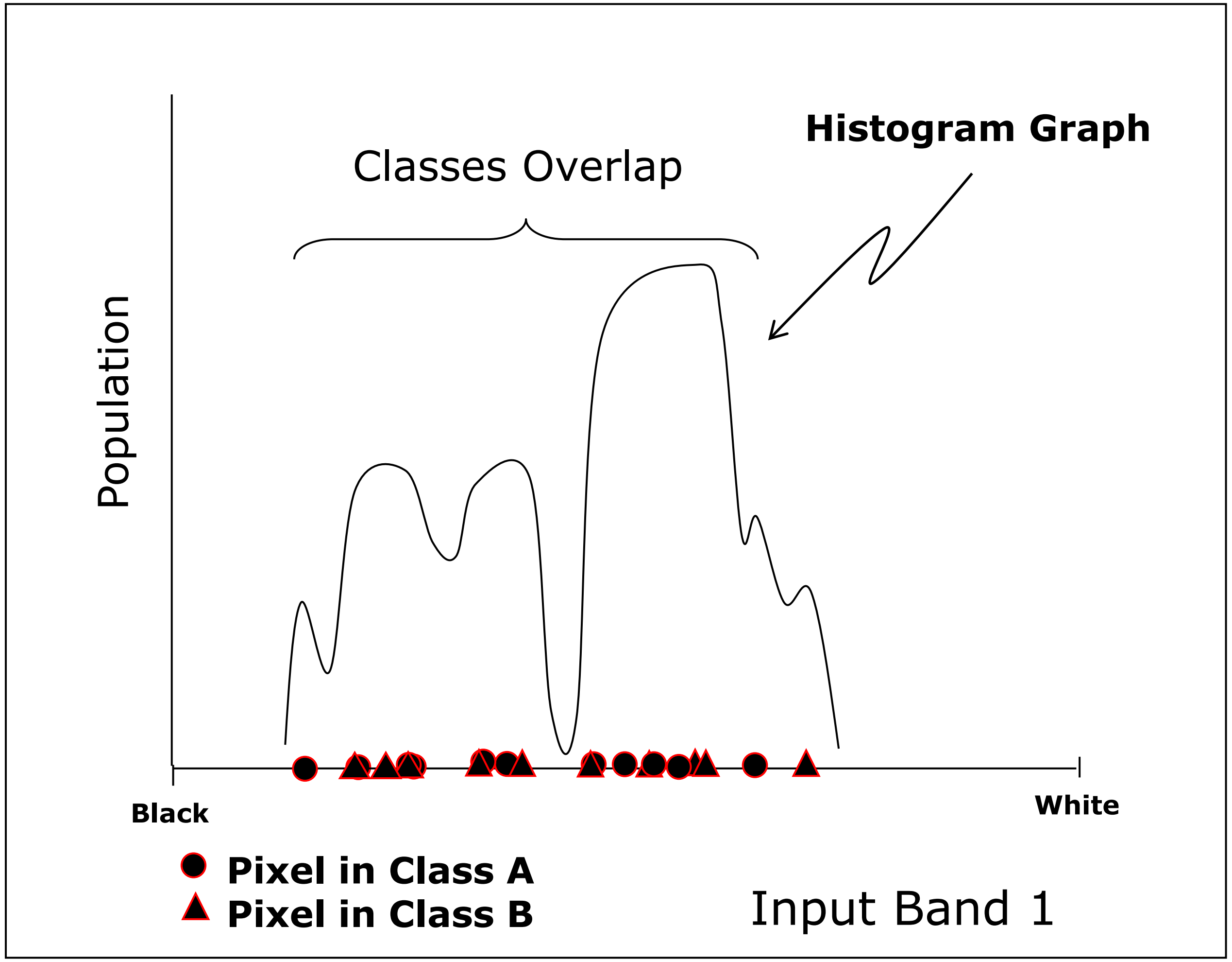
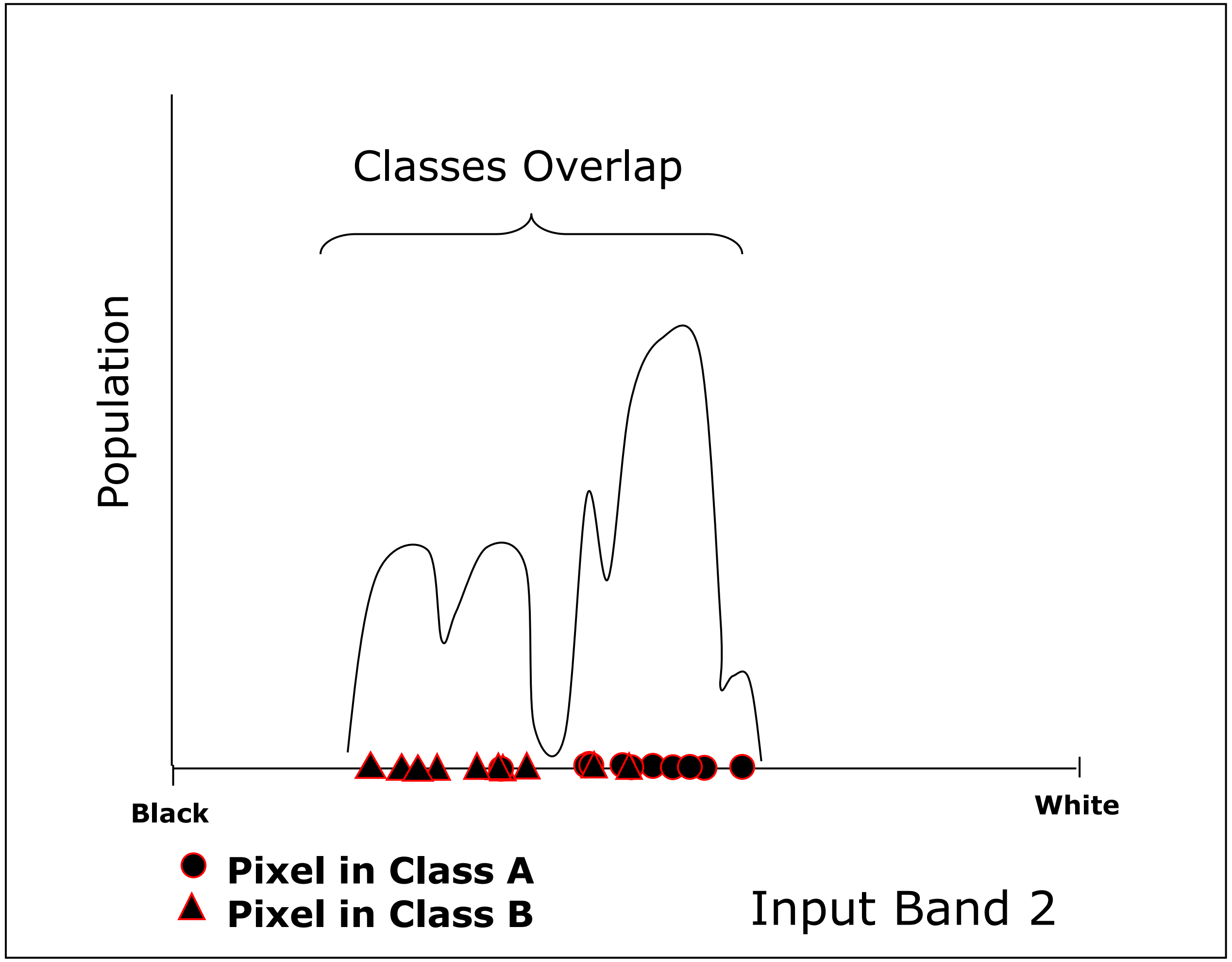


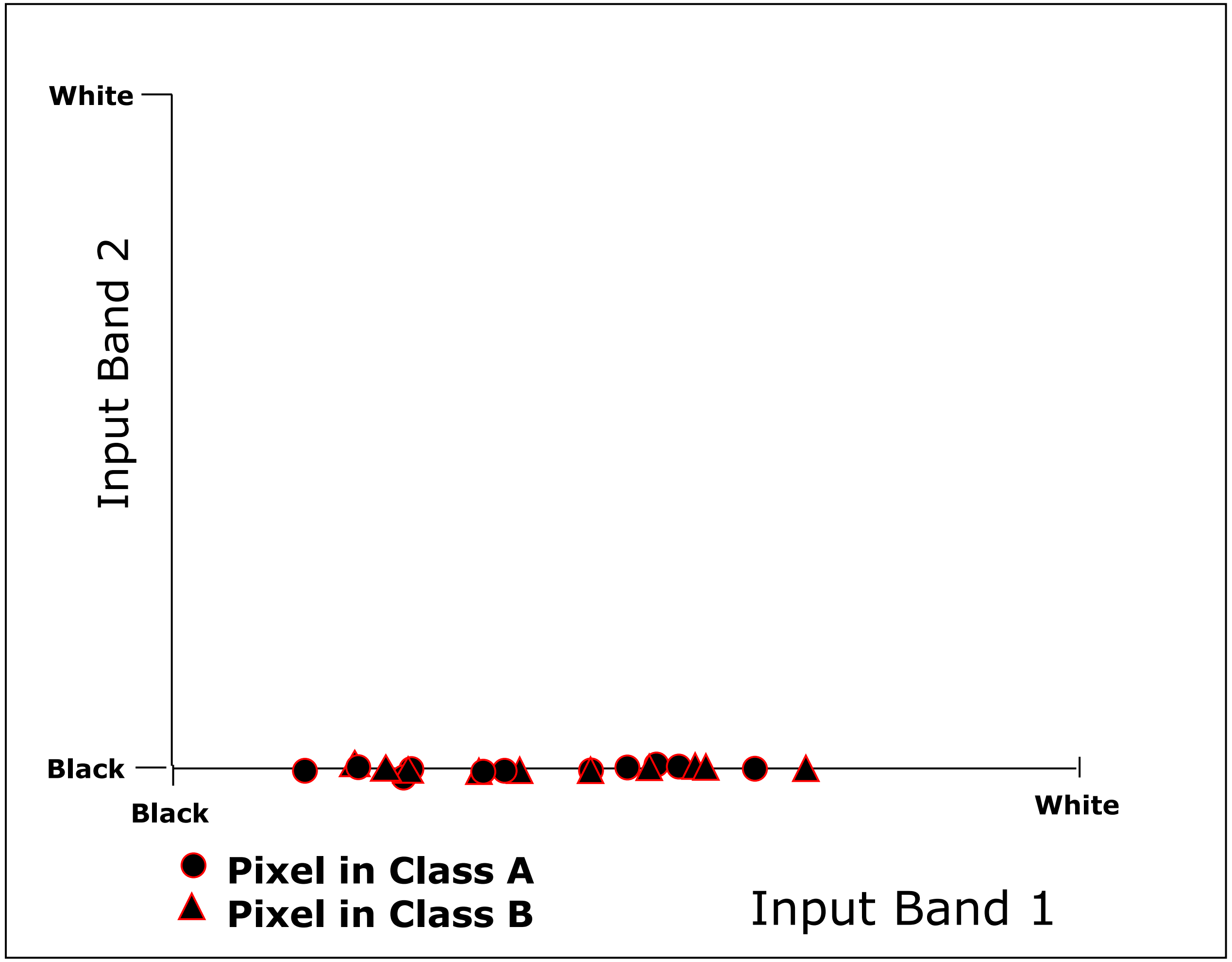
Image Histogram, Band 2

- Again, pixels in Class “A” are not distinguished from those in Class “B” by gray value



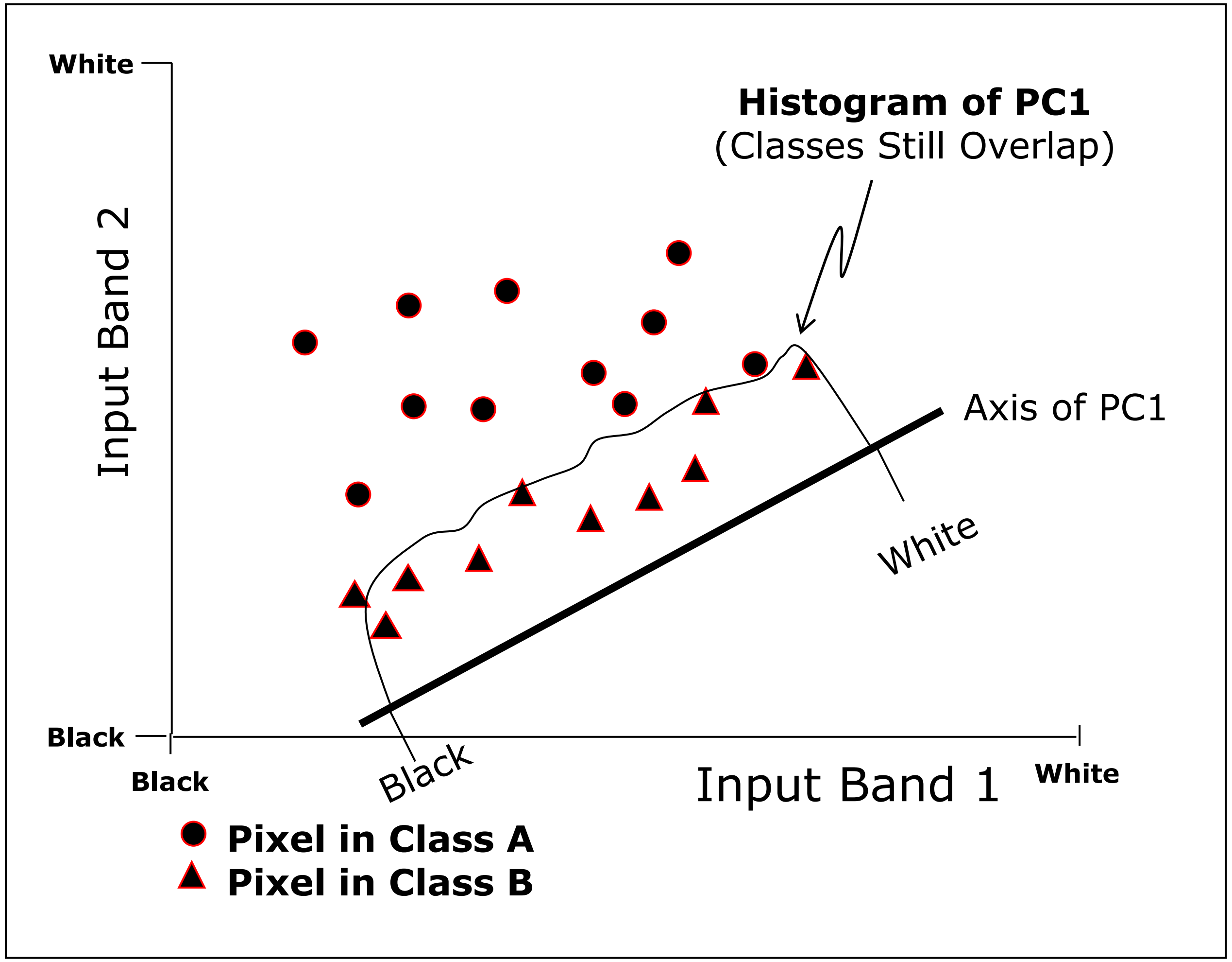
2-D Histogram of Bands 1,2

- "Simultaneous" probability of pixel gray values in two images
  - How many pixels have same pair of gray values in two images?



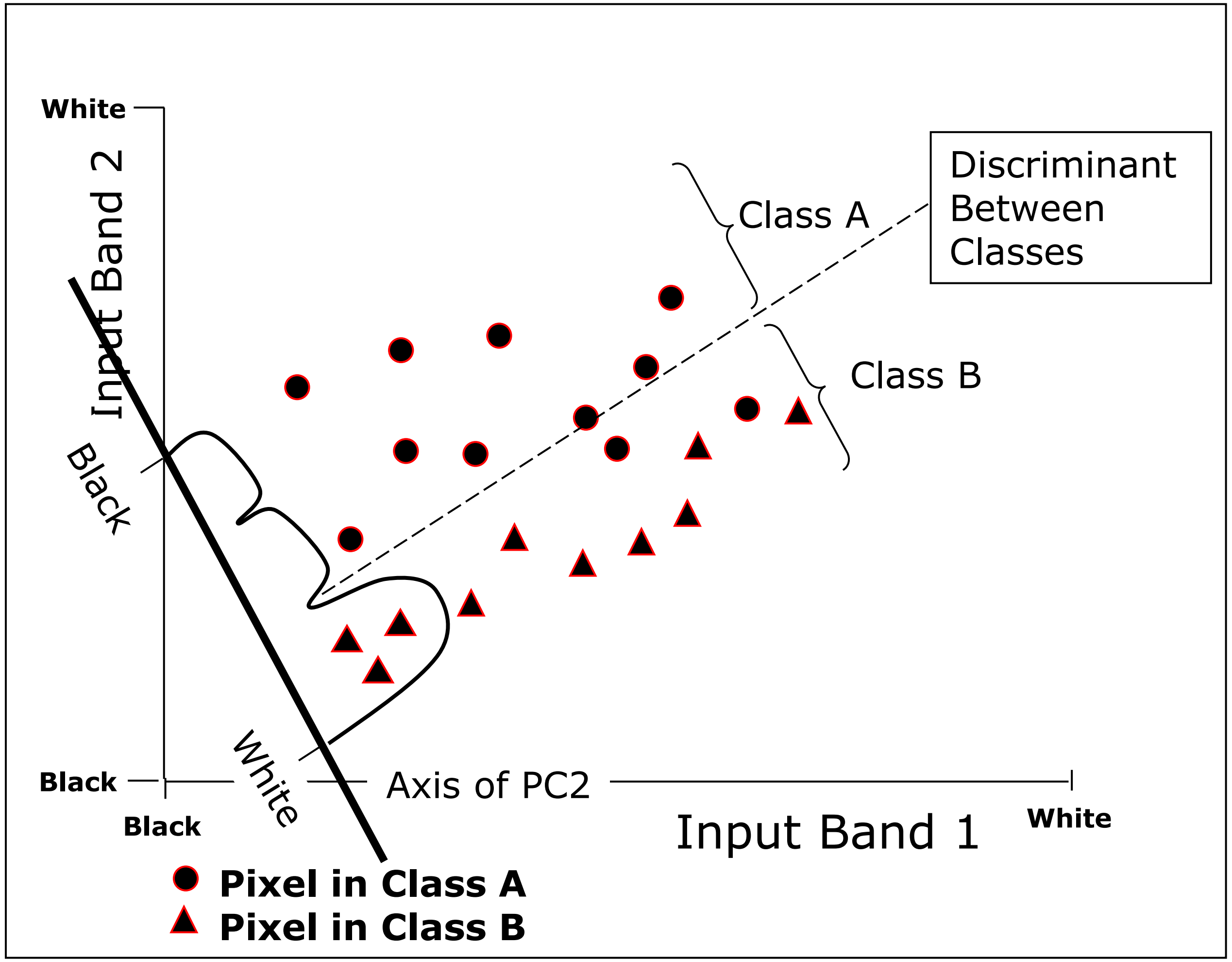
1st Principal Component

- Project pixels onto axis with largest variance
- Map ends of axis to “black” and “white”
- Forms new image as weighted sum of constituent images

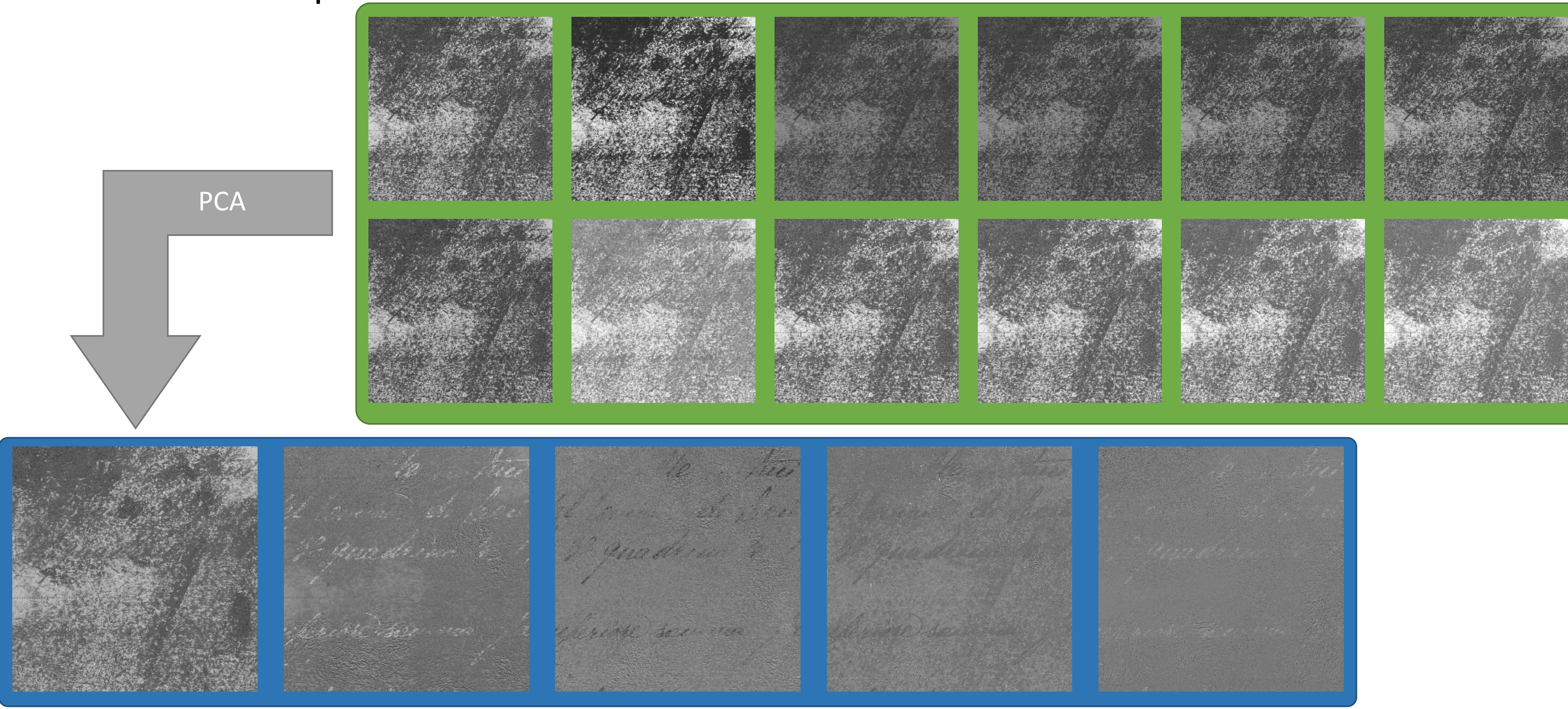


2nd Principal Component

- Project onto perpendicular axis with next largest variance
- Map ends of axis to “black” and “white”
- Forms new image as weighted sum of constituent images
  - enhanced contrast

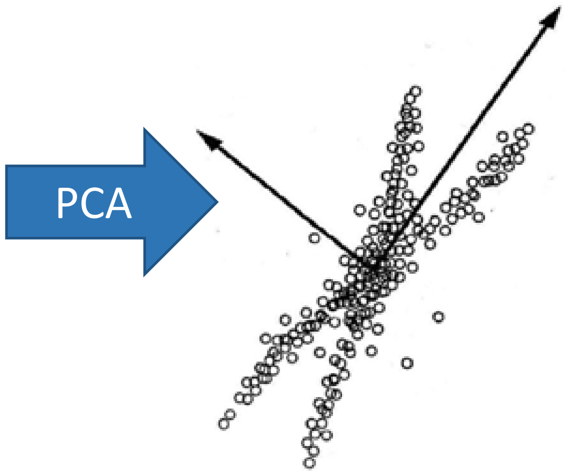


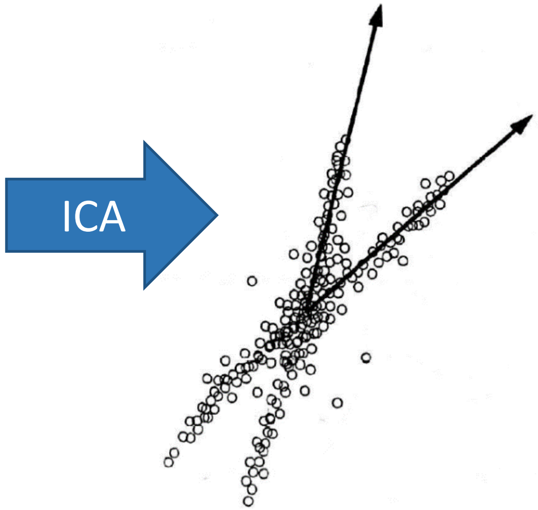
PCA - Example

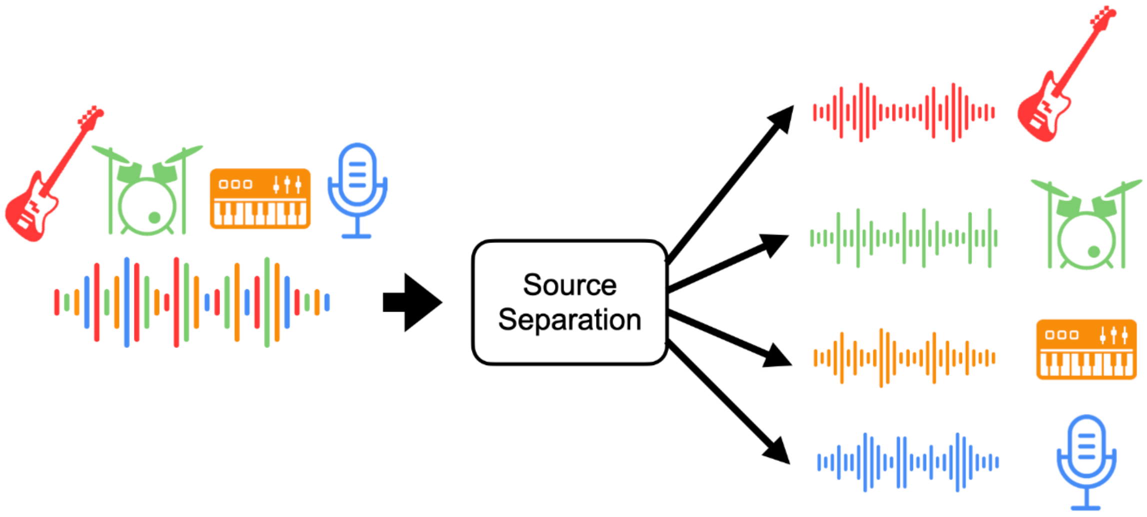


Independent Component Analysis (unsupervised)

- Related to PCA, but:
- New axes may be non-orthogonal
- Especially useful for separating multiple sources
- Components are **not** ordered
- Assumes independent, non-gaussian sources



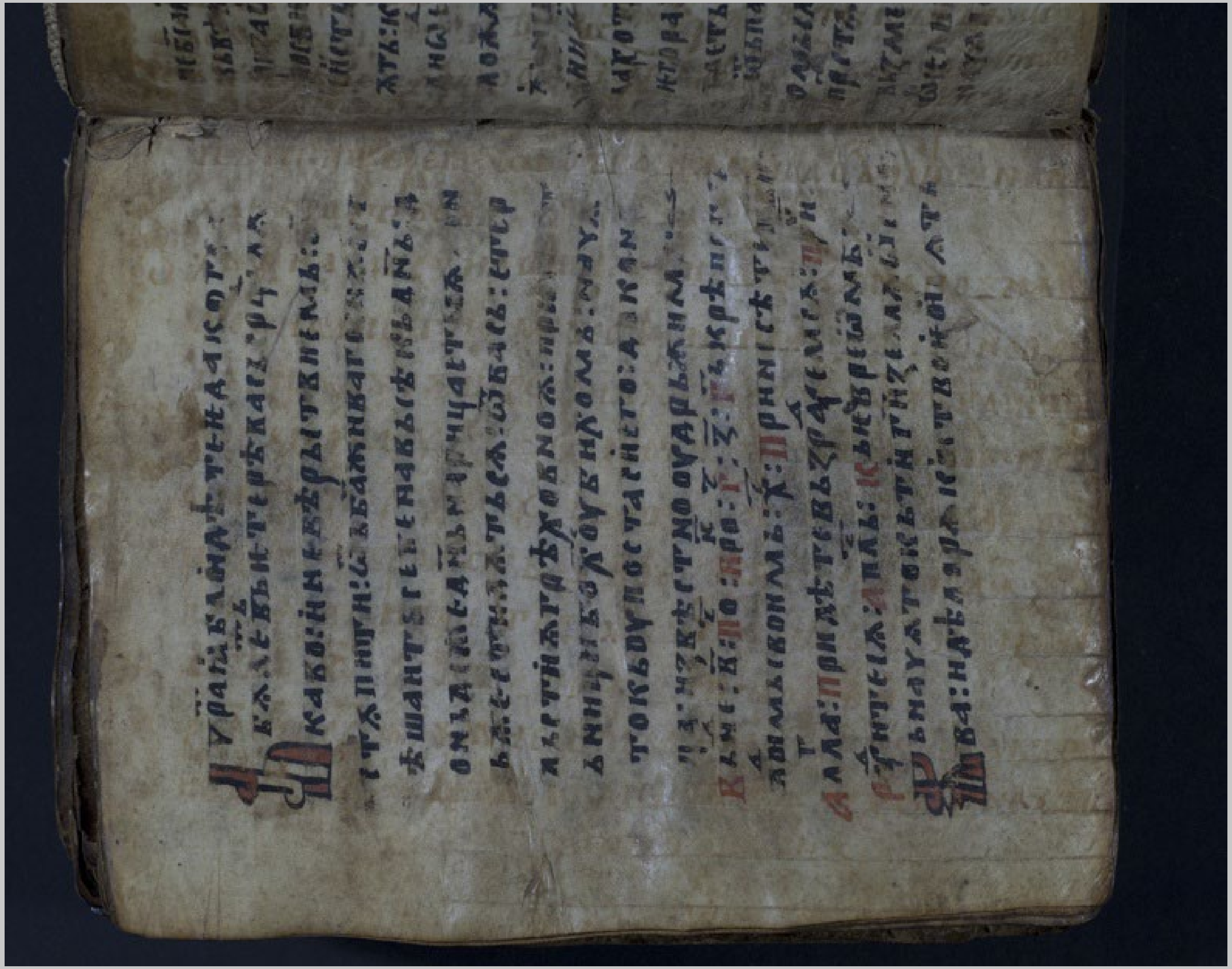




Computer Vision Lab

PCA vs. ICA comparison

- ICA assumes independence of sources
  => not always fulfilled!
- In such cases a dimension reduction with PCA is extremely helpful
- PCA transformation uses orthogonal basis, ICA not
- PCA is faster
- ICA performs often better if a huge number of dominant sources (e.g. inks) is present

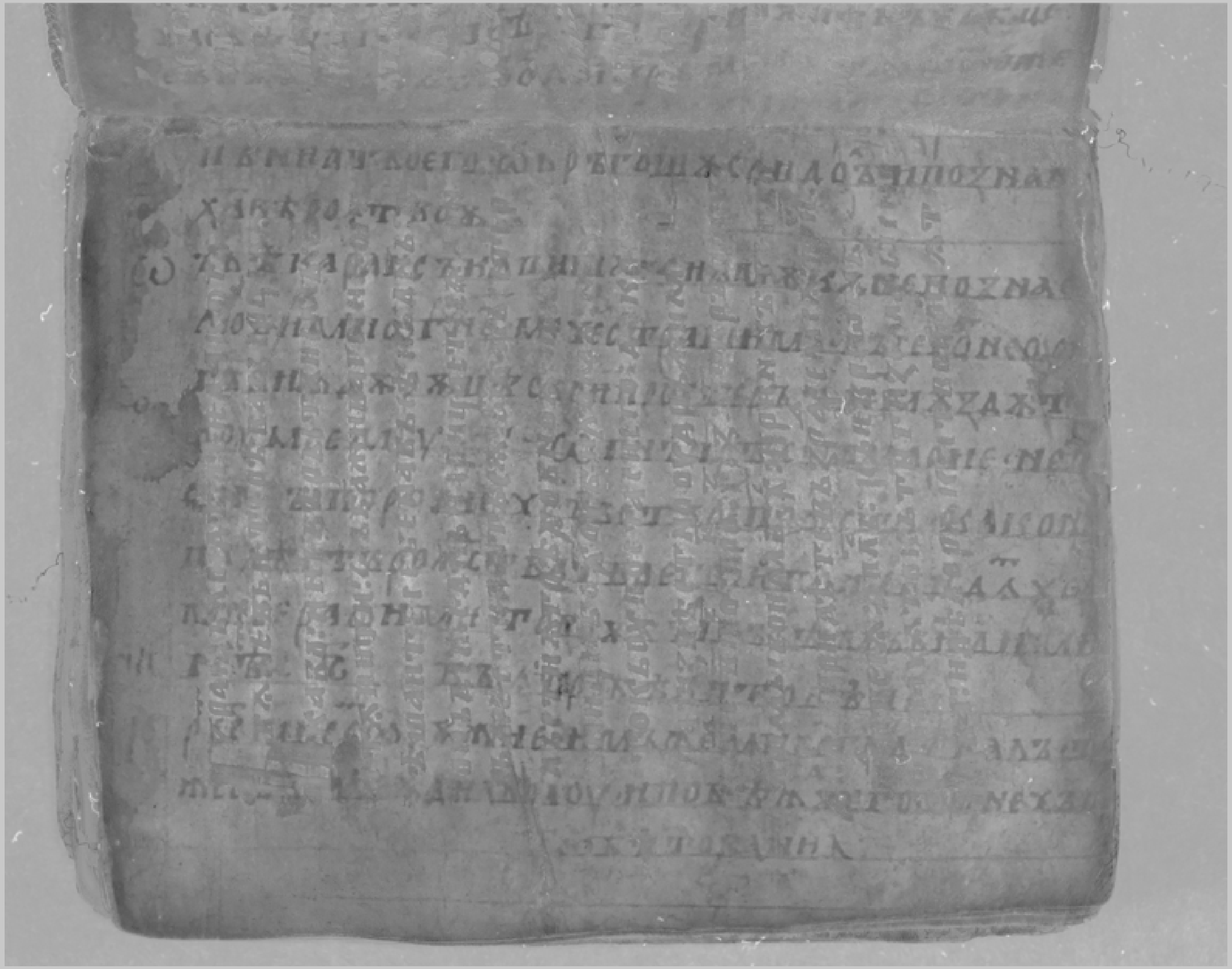


ICA Example: RGB image

ICA Example: UV Fluorescence

Independent Component #1

Independent Component #2



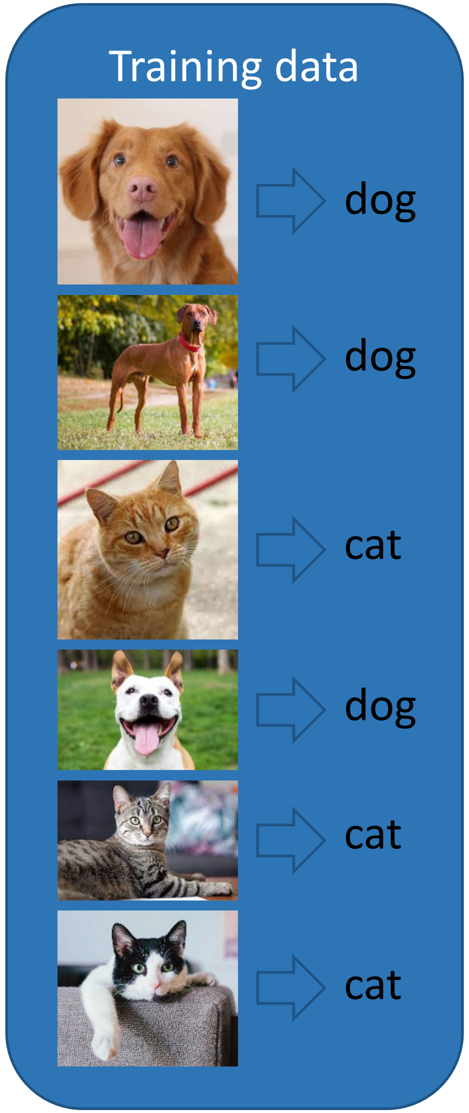
Independent Component #7

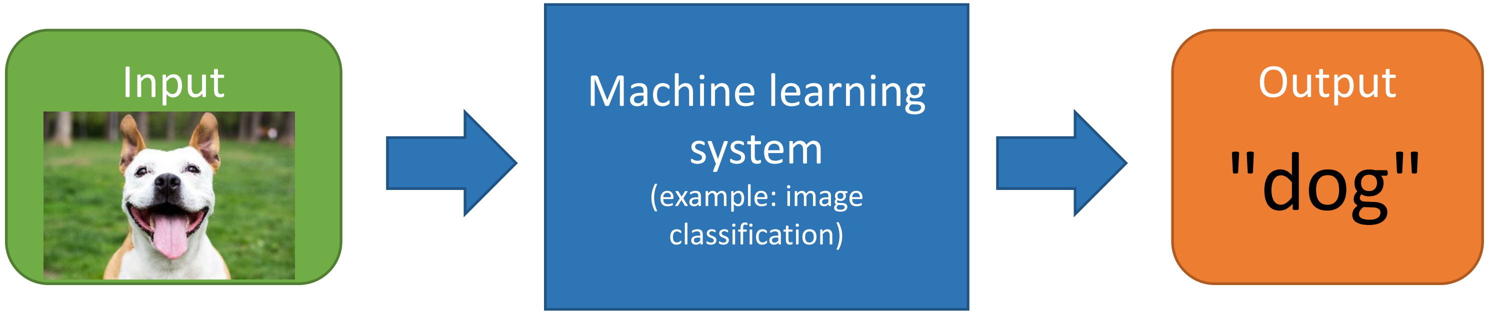
Principal Component #1

Principal Component #4

Supervised methods – classical machine learning

- Goal: create a system that estimates a target property from complex input data, e.g.:
  - size, color → fish species (machine learning example for beginners)
  - raster image → depicted object (image classification)
  - sound wave → spoken text (speech recognition)
- Train the system with pairs (input, expected output)
  - Define an error function (expected output vs. system's output)
  - Optimize the system parameters, such that the error becomes small for the training set



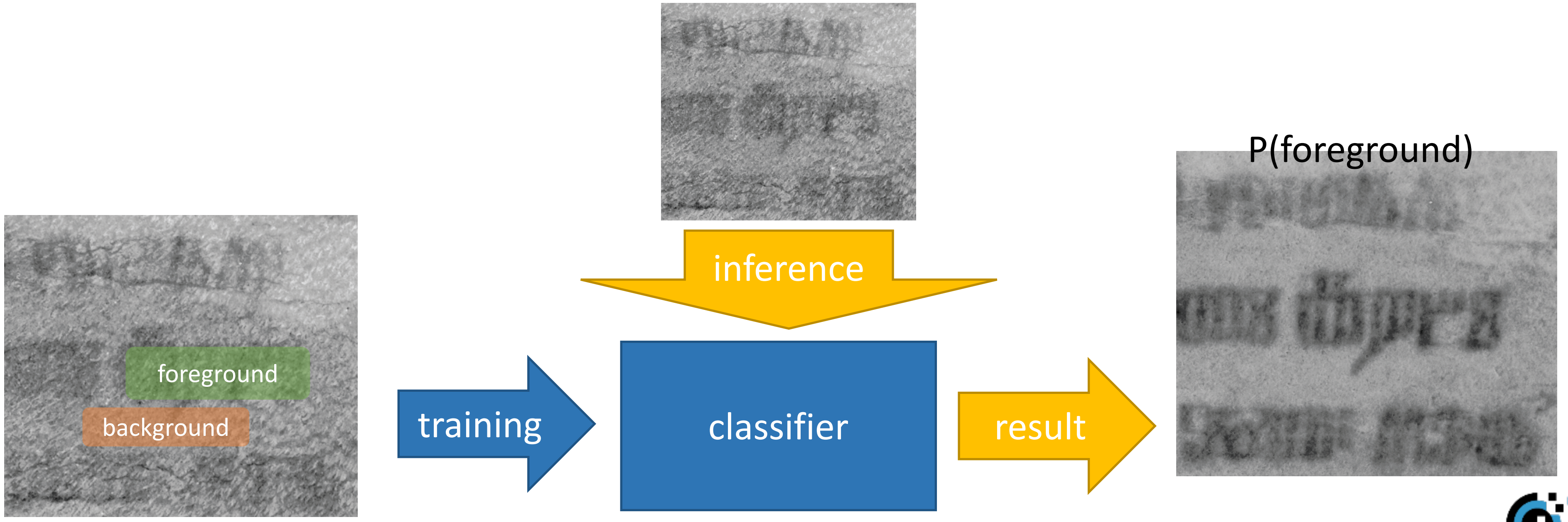


Applied to multispectral images…

- On pixel level
  - $\underbrace{(x_1, x_2, \dots, x_n)}_{\text{Spectral signature}} \Rightarrow \underbrace{(parchment|ink_1|ink_2)}_{\text{Material class}}$

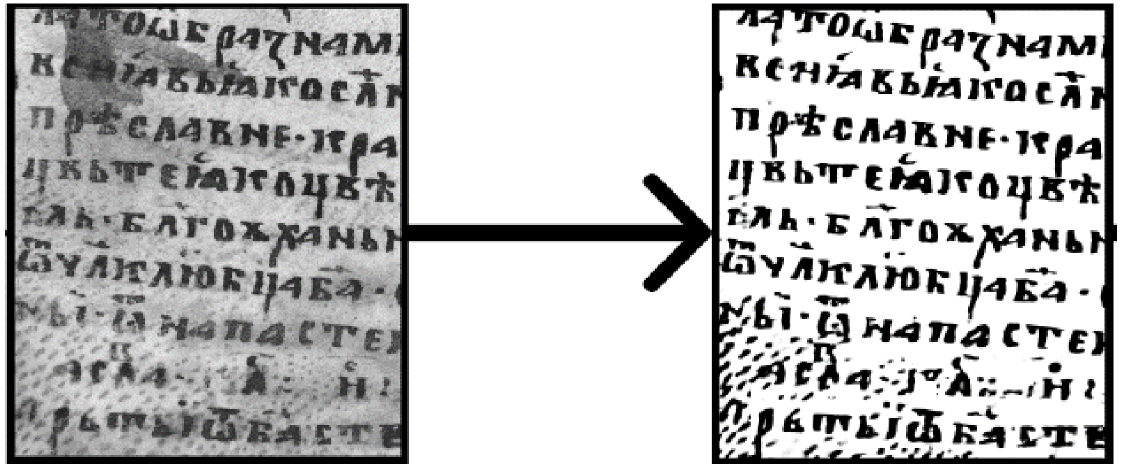
Simplest workflow:

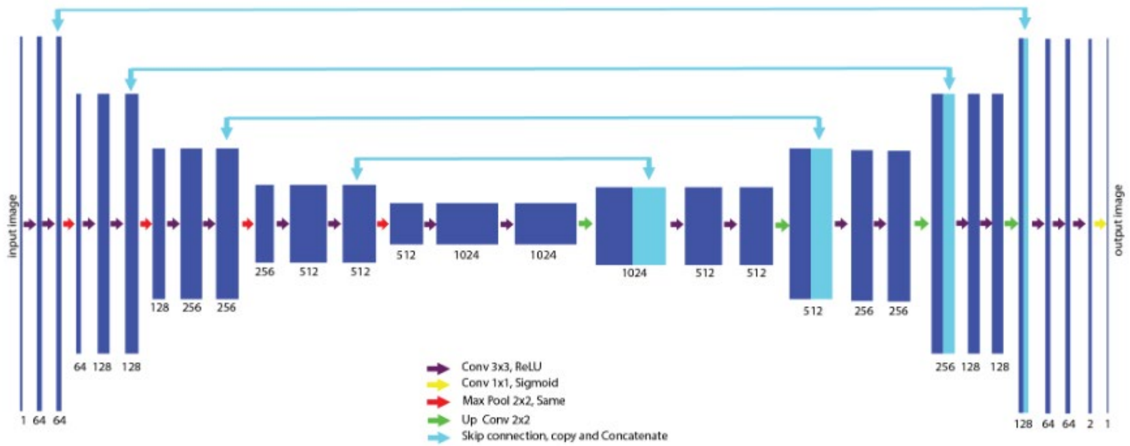
- label some pixels (as training data)
- Train classifier
- Estimate the classes of all remaining pixels



Applied to multispectral images…

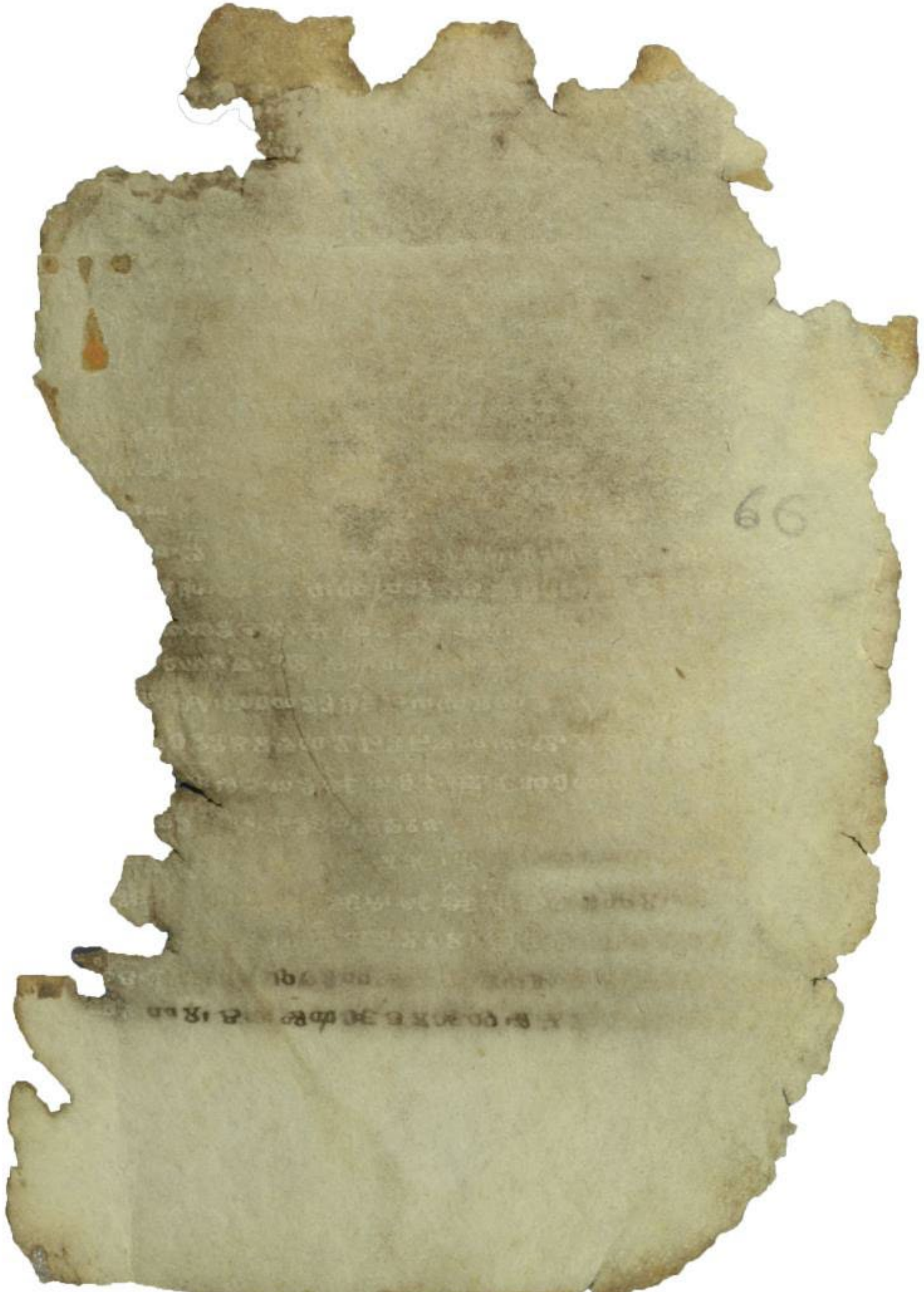
- On image level
  - Multispectral image ➔ processed image (e.g. binary image)
  - Convolutional Neural Networks – "Deep Learning"



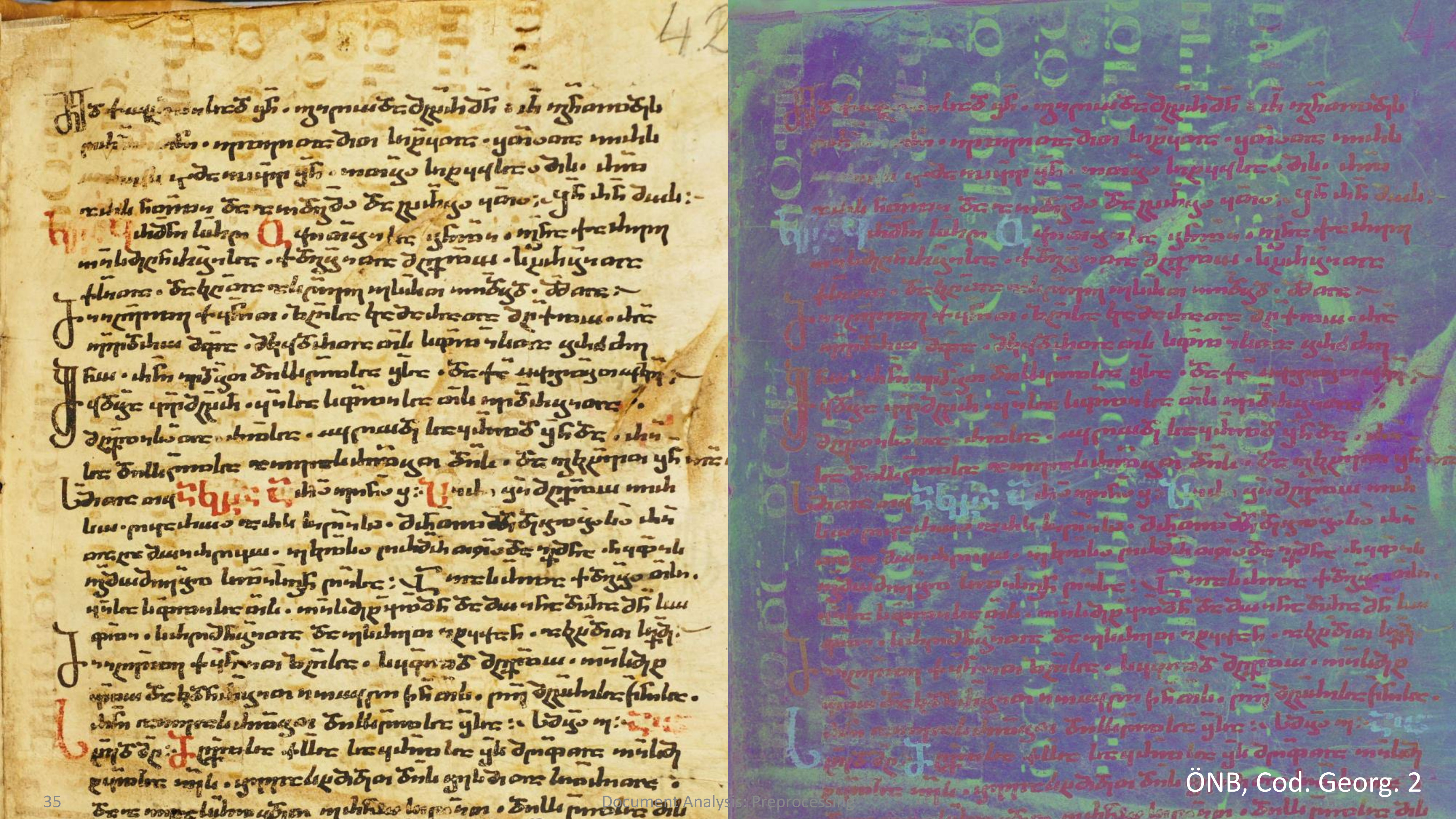


"DE-GAN": Document Enhancement Generative Adversarial Network

Examples



"Missale“
(Euchologium II) Sinaiticum

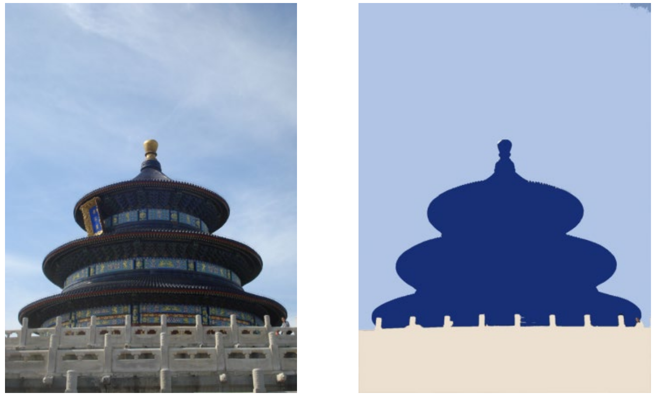


ÖNB, Cod. Georg. 2

Binarization

Image Segmentation

- The objective of image segmentation is to group image pixels according to pre-defined rules.
- Segmentation is often considered to be the first step in image analysis.
  - The purpose is to subdivide an image into meaningful non-overlapping regions, which would be used for further analysis.
  - It is hoped that the regions obtained correspond to the physical parts or objects of a scene (3-D) represented by the image (2-D).



Computer Vision Lab

Document Binarization

- Image Segmentation consists of 2 classes: foreground (written text) and background (paper)
- It converts a gray-scale document image into a binary document image
- Document image binarization is [was] usually performed in the preprocessing stage of different document image processing related applications:
  - OCR
  - Writer Identification
  - Layout Analysis, …

Solved Problem in DA?

Global Threshold:
200


Ground Truth Generation for Snippets and Segmentation

First author, second author and third author
University of Technology,
Somewhere on Earth
author@earth.com

Abstract - Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean consectetur ligula id nulla rhoncus vel tempor turpis luctus. Nunc eget felis ac sem aliquet tempor. Nulla facilisi. Nulla mattis ultricies aliquet. Sed eget suscipit nibh. Nunc sagittis porttitor turpis, id scelerisque ante dapibus ac. Nunc cursus luctus lacus, quis pulvinar ipsum euismod non. Aenean enim nibh, venenatis at tincidunt vel, varius a sem. Etiam eget ligula massa, et cursus ligula. Ut sed arcu nulla, sed dignissim nisl. Nunc sagittis libero in turpis dapibus at dictum odio hendrerit. Maecenas id ornare libero. Nam vitae augue eget leo lacinia egestas dapibus in tortor.

First Paragraph

Ut mattis, orci vitae ultrices venenatis, nisl quam viverra ipsum, sed rutrum nunc nibh non nunc. Nulla luctus mauris a turpis dictum rhoncus cursus leo vehicula. Integer elementum ipsum at tellus posuere aliquam. Ut sed vehicula neque. Maecenas malesuada mollis erat posuere varius. Curabitur leo massa, mattis nec lobortis pellentesque, feugiat vel libero. Vivamus id rhoncus orci. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Sed nisi orci, tempus quis interdum fringilla, ullamcorper vel neque. Sed sit amet nisi risus, sit amet dictum elit.

No.	Precision	Recall	Algorithm
1.	86%	15%	Name 1
2.	90%	36%	Name 2
3.	5%	100%	Test 3
4.	30%	2%	Ut mattis 4
5.	45%	17%	Sed sit 5

Table 1: Results of Lorem Ipsum

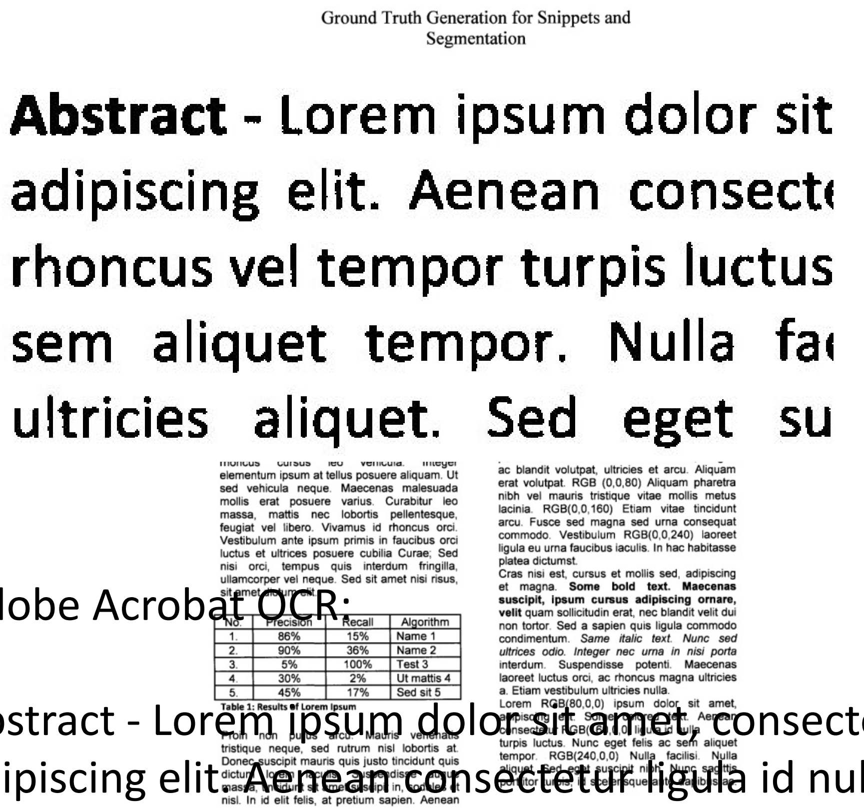
Proin non purus arcu. Mauris venenatis tristique neque, sed rutrum nisl lobortis at. Donec suscipit mauris quis justo tincidunt quis dictum lorem iaculis. Suspendisse augue massa, tincidunt sit amet suscipit in, sodales et nisl. In id elit felis, at pretium sapien. Aenean vitae mi est. Nullam massa dui, fringilla eu hendrerit ut, rutrum et arcu. Proin a quam augue, nec tristique enim. Quisque eleifend ligula id augue posuere ut cursus tortor sodales. Nullam eu ligula sem. Donec erat nulla, iaculis ullamcorper interdum ac, facilisis vel eros. In tincidunt iaculis mi non semper. Quisque convallis, mi id imperdiet tempus, elit augue porttitor sem, vel sollicitudin mauris odio eu ipsum. Praesent eget nulla velit. Suspendisse potenti.

Caption 1	Caption 2	Caption 3
Algorithm 1	95%	Good results
Algorithm 2	20%	Bad results

Table 2: test algorithm of the presented paper

Vestibulum et lectus tellus, a tempor metus. Ut ligula RGB (0,80, 0)neque, bibendum facilisis ultricies a, aliquam eget tellus. Suspendisse nec nunc nec RGB(0,160,0) magna ultrices vestibulum. Duis tempus, leo sed porta auctor, sem enim imperdiet est, RGB(0,240,0)vitae porta nisl nisi vel erat. In tellus lorem, feugiat ac blandit volutpat, ultricies et arcu. Aliquam erat volutpat. RGB (0,0,80) Aliquam pharetra nibh vel mauris tristique vitae mollis metus lacinia. RGB(0,0,160) Etiam vitae tincidunt arcu. Fusce sed magna sed urna consequat commodo. Vestibulum RGB(0,0,240) laoreet ligula eu urna faucibus iaculis. In hac habitasse platea dictumst.

Cras nisi est, cursus et mollis sed, adipiscing et magna. **Some bold text. Maecenas suscipit, ipsum cursus adipiscing ornare, velit** quam sollicitudin erat, nec blandit velit dui non tortor. Sed a sapien quis ligula commodo condimentum. *Same italic text. Nunc sed ultrices odio. Integer nec urna in nisi porta* interdum. Suspendisse potenti. Maecenas laoreet luctus orci, ac rhoncus magna ultricies a. Etiam vestibulum ultricies nulla. Lorem RGB(80,0,0) ipsum dolor sit amet, adipiscing elit. Some colored text. Aenean consectetur RGB(160,0,0) ligula id nulla turpis luctus. Nunc eget felis ac sem aliquet tempor. RGB(240,0,0) Nulla facilisi. Nulla aliquet. Sed eget suscipit nibh. Nunc sagittis porttitor turpis, id scelerisque ante dapibus ac.


Ground Truth Generation for Snippets and Segmentation

Abstract - Lorem ipsum dolor sit
adipiscing elit. Aenean consecte
rhoncus vel tempor turpis luctus
sem aliquet tempor. Nulla fa
ultricies aliquet. Sed eget su

Adobe Acrobat OCR:

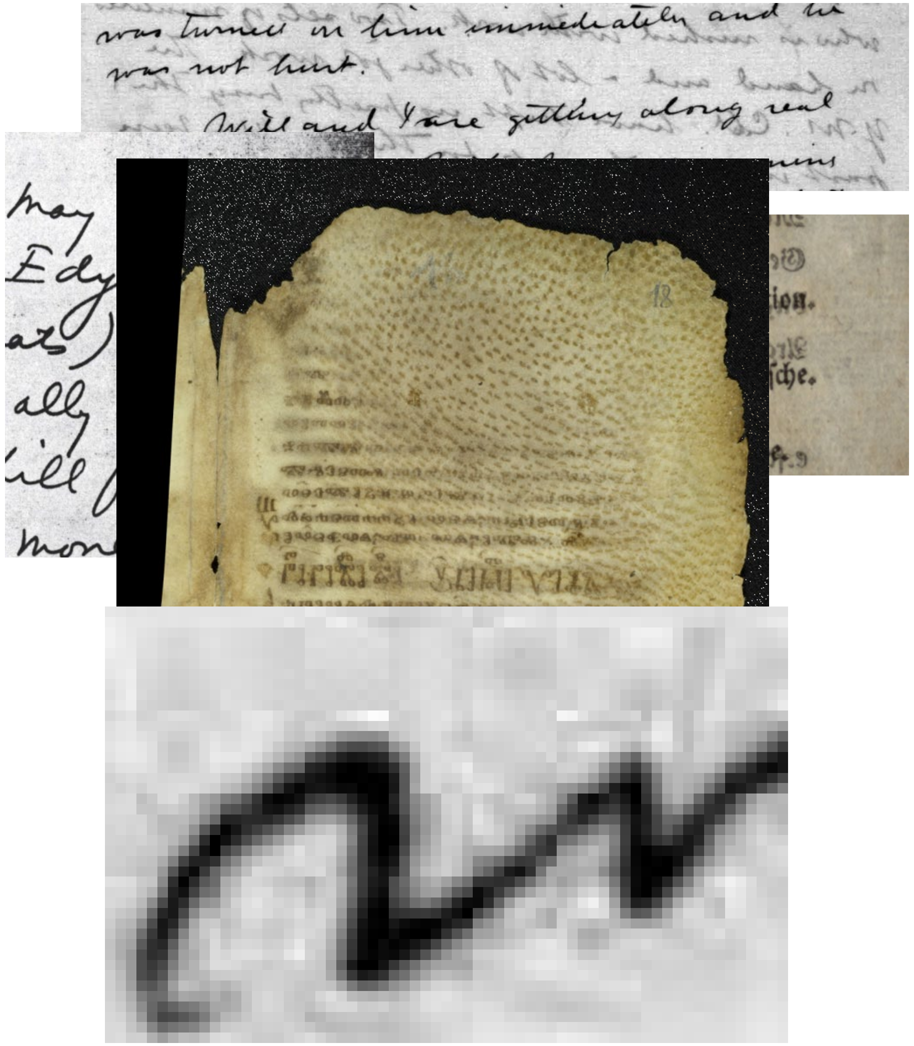
Abstract - Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean consectetur ligula id nulla rhoncus vel tempor turpis luctus.

Computer Vision Lab

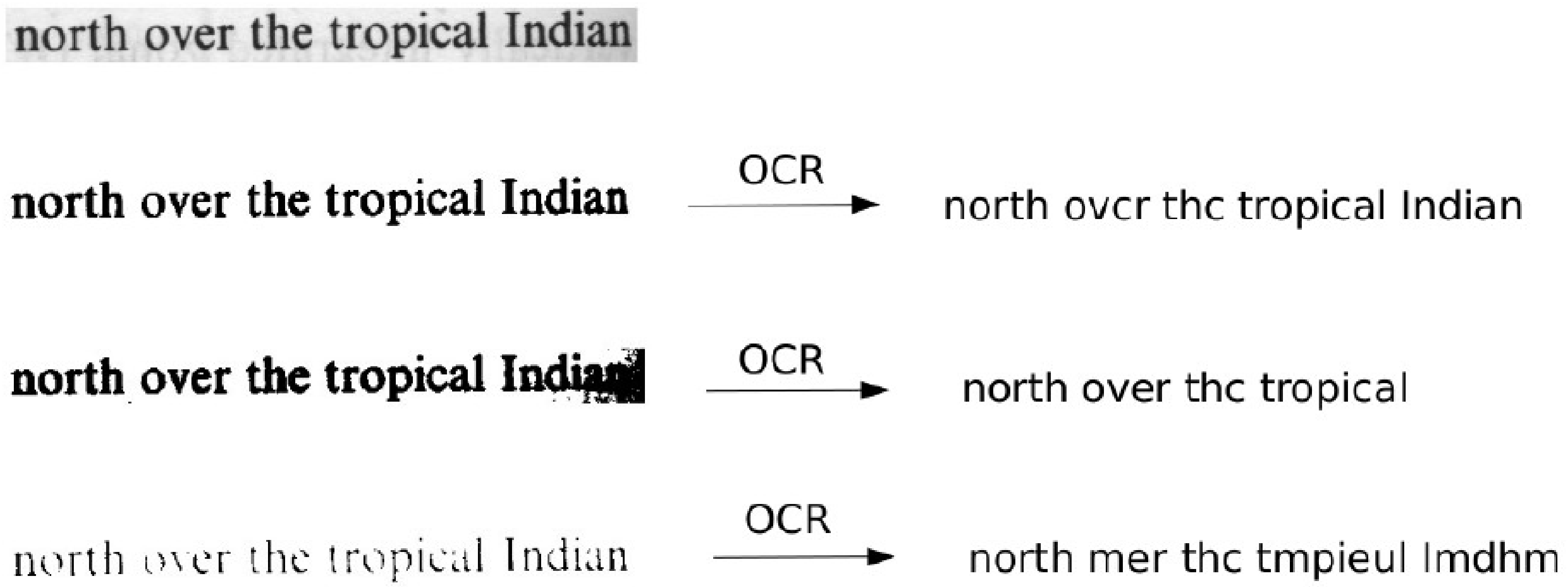
Binarization Problems

- Bleed Through Text
- Background Variation
- Low Contrast
- Distortion/Background Clutter
- Image Compression Artefacts
- Uneven Illumination

- Combinations of the listed Problems



Effect of Binarization on OCR

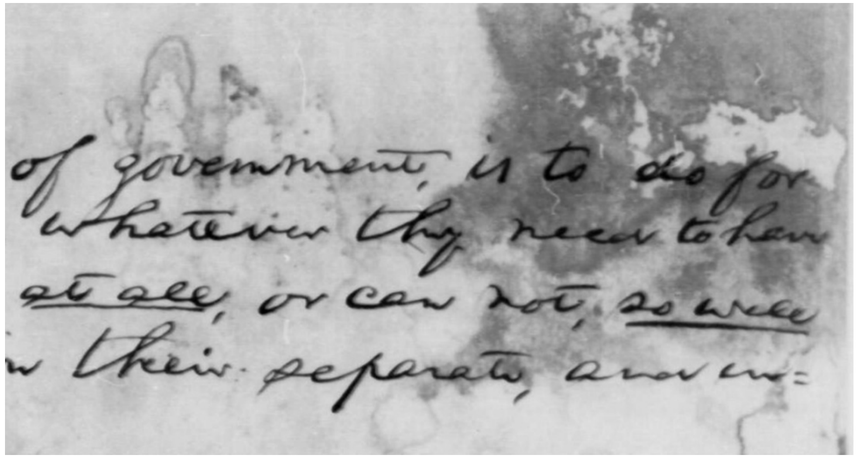


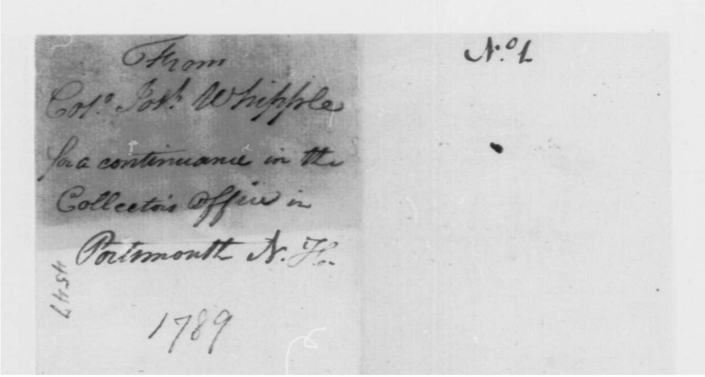
courtesy by Faisal Shafait

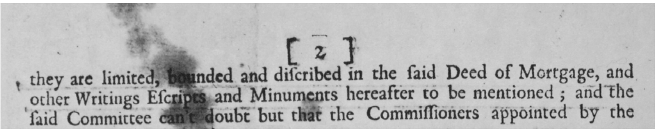
Binarization Contest

- DIBCO – Document Image Binarization Contest
- H-DIBCO – Handwritten Document Image Binarization Contest
- Contests in conjunction with ICDAR, DAS, ICFHR

- DIBCO 2009 was the 1st Int. Document Image Binarization contest

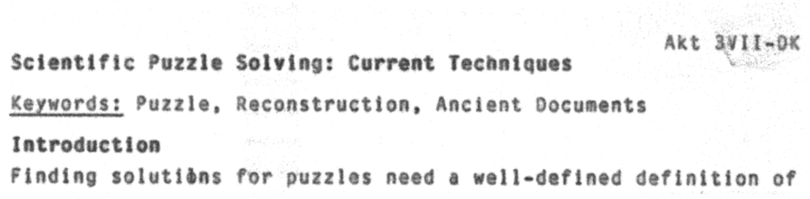

of government, is to do for
whatever they need to have
at all, or can not, so well
in their separate, and in-


From
Col. Jos. Whipple
For a continuance in the
Collector's Office in
Portsmouth N.H.
1789
No. 1


[2]
they are limited, bounded and diſcribed in the ſaid Deed of Mortgage, and other Writings Eſcripts and Minuments hereafter to be mentioned; and the ſaid Committee can't doubt but that the Commiſſioners appointed by the

Evaluation of Binarization Methods

- Human Evaluators
- Binary result is subjected to OCR – corresponding result is evaluated with respect to character or word accuracy

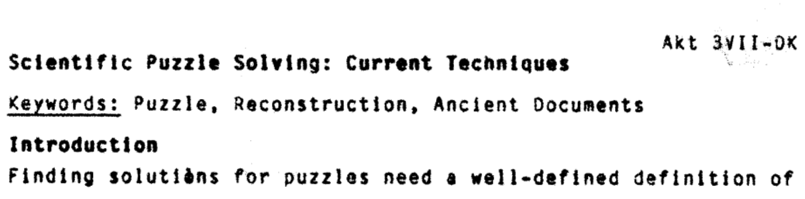

Akt 3VII-DK

Scientific Puzzle Solving: Current Techniques

Keywords: Puzzle, Reconstruction, Ancient Documents

Introduction

Finding solutiàns for puzzles need a well-defined definition of


Akt 3VII-DK

Scientific Puzzle Solving: Current Techniques

Keywords: Puzzle, Reconstruction, Ancient Documents

Introduction

Finding solutiàns for puzzles need a well-defined definition of

Adobe Acrobat OCR:

Akt l:~ u-01(

Scientific Puzzle SolyIng: Current Techniques
Keywords: Puzzle, Reconstruction. Ancient Documents
Introduction
Finding solutlins fur pUlzles need I well-defined definition of

Evaluation of Binarization Methods

- Binary Image is used as Ground Truth (GT) image - result is compared at pixel level
  - Noise is added to GT image to create a grey value image (synthetic images)
  - Gray scale image is binarized by a human

gray value image

GT image

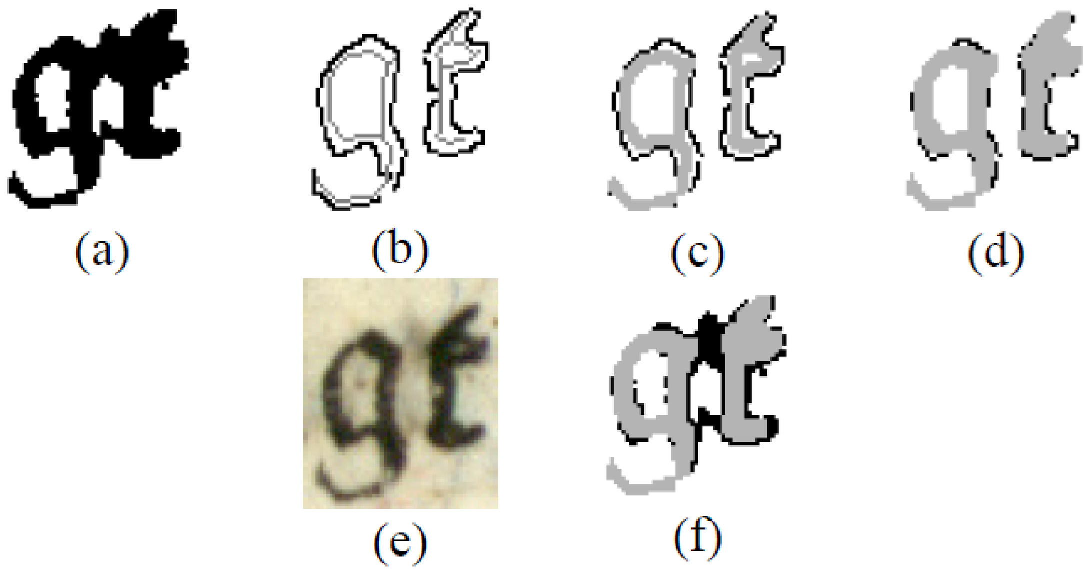
global threshold 90

- Combination of human oriented and evaluation and OCR results accuracy

See Nitrogiannis K., Gatos B. And Pratikakis I., *An Objective Evaluation for Document Image Binarization Techniques*, 9th IAPR Workshop on Doucment Analysis Systems, pp. 217-224, 2008.

DIBCO Evaluation

- DIBCO-GT: semi automatic procedure, which calculates skeleton of the image binarized by an adaptive method, followed by a dilation which is limited by the edges of strokes detected by Canny-Edge Detector



Courtesy of Ntirogiannis et al., 2008

(a) Binary image
(b) Corresponding edges + skeleton
(c) Dilated components (one dilation) – 15.29% of the edges are covered
(d) Dilated components (second dilatio) – 54.9% of the edges are covered
(e) Original image
(f) Estimated GT (gray), placed over binary image

DIBCO Evaluation Measures

- Binary image is compared with GT image pixelwise
- Evaluation measures
  - F-Measure, pseudo-F-Measure
  - PSNR (Peak Signal to Noise Ratio)
  - DRD (Distance Reciprocal Distortion Metric) (DIBCO 2011)

B. Gatos, K. Ntirogiannis and I. Pratikakis, ICDAR 2009 Document Image Binarization Contest (DIBCO 2009), 10th Int. Conf. On Document Analysis and Recognition, pp. 1375-1382, 2009.

I. Pratikakis, B. Gatos and K. Ntirogiannis, ICDAR 2011 Document Image Binarization Contest (DIBCO 2011), 11th Int. Conf. On Document Analysis and Recognition, pp. 1506-1510, 2011.

DIBCO Evaluation Measures

- F Measure

http://utopia.duth.gr/~ipratika/DIBCO2011/Evaluation.html

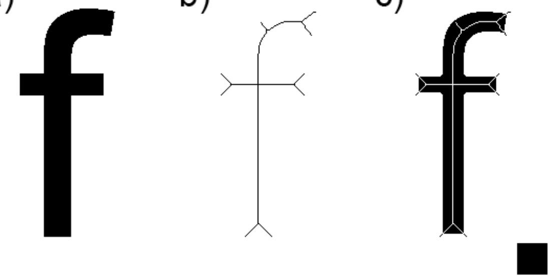
$$F - Measure = \frac{2 \cdot \mathrm{Re}\,call \cdot \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \qquad \mathrm{Pr}\,ecision = \frac{TP}{TP + FP}$$

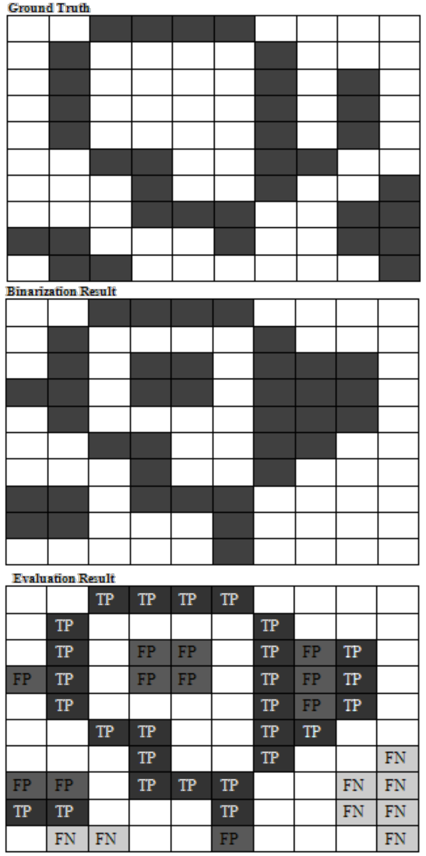
TP, FP, FN denote the True Positive, False Positive and False Negative Values. SG denote the Skeletonized Ground Truth, and B the binary image.

$$p - \mathrm{Re}\,call = \frac{\sum\sum SG(x,y) \cdot B(x,y)}{\sum\sum SG(x,y)} \qquad p - FM = \frac{2 \cdot p - \mathrm{Re}\,call \cdot \mathrm{Pr}\,ecision}{p - \mathrm{Re}\,call + \mathrm{Pr}\,ecision}$$

a) b) c)



a) GT image
b) Skeleton of GT
c) Binarized image: white Pixel of GT‘s skeleton have an impact on pseudo-Recall



TP = 27, FP = 11, FN = 8
RC = 0.77, PR=0.71, FM = 73.9%

Computer Vision Lab
WIEN

DIBCO Evaluation Measures

- PSNR

$$PSNR = 10 \cdot \log\left(\frac{C^2}{MSE}\right)$$

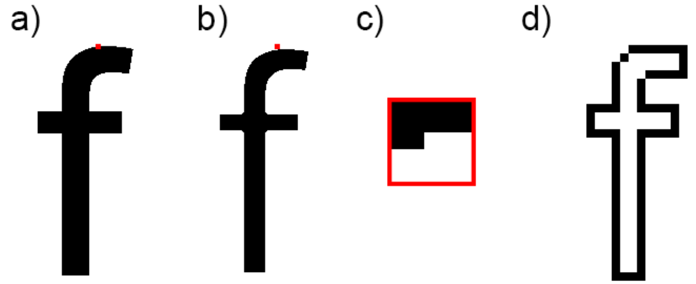
$$MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I'(x,y))^2}{M \cdot N}$$

PSNR is a similiarity measure.
The higher the value, the higher is the similiarity of 2 images.
Difference between Background and Foreground equals to C

- DRD (Distance Reciprocal Distortion Metric)

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN}$$

$$DRD_k = \sum_{i=-2}^{2}\sum_{j=-2}^{2} \left|GT_k(i,j) - B_k(x,y)\right| \cdot W_{Nm}(i,j)$$



a) GT image
b) Binarized image
c) 5x5 pixel window centered at first flipped pixel (white pixel have an impact on DRD)
d) NUBN

y document images.
el
n the 5x5 block of the
pixel at (x,y) in the
ocks in the GT image.

H. Lu, A.C. Kot and Y.Q. Shi, *Distance-Reciprocal Distortion Measure for Binary Document Images*, IEEE Signal Processing Letters, vol. 11, No. 2, pp- 228-231, 2004.

Evaluation Measures Example

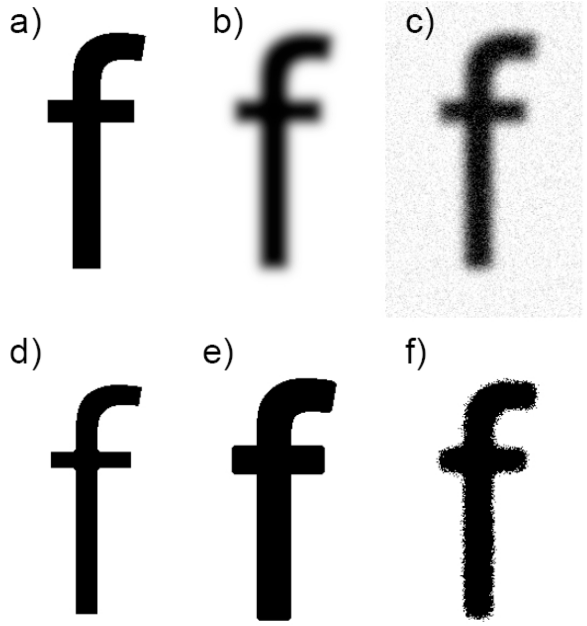


Figure 4.5: a) GT image b) GT image with Gaussian blur, σ=5 c) GT image, Gaussian noise added ($m = 0, \sigma = 0.01$) d) Eroded GT image (disc, $size = 4$), e) Dilated GT image (disc, $size = 4$), f) Otsu Threshold of c). See Table 4.2 for Evaluation Measures.

Input Image	"f" - Otsu (blur+noise) (Figure 4.5 f)	"f" eroded (Figure 4.5 d)	"f" dilated (Figure 4.5 e)	"f" border pixel flipped
Precision [%]	0.91	1	0.79	0.95
Recall [%]	0.97	0.74	1	1
p-Recall [%]	0.95	0.93	1	1
F-Measure [%]	94.51	85.48	88.42	97.88
p-F-Measure [%]	93.64	96.44	88.42	97.88
PSNR	17.99	14.48	14.35	22.18
NRM $\times 10^{-2}$	2.04	12.67	2.13	0.35
MPM $\times 10^{-3}$	0.38	0.63	1.30	0.10
DRD	7.58	19.30	20.07	2.45

Binarization Methods

- Global Binarization Methods
  - A single threshold is applied on every pixel in the image f(x,y): Thresholded image g(x,y) with threshold T is defined by

$$g(x,y) = \begin{cases} 1 & if \quad f(x,y) \geq T \\ 0 & otherwise \end{cases}$$

- Local (Adaptive) Binarization Methods
  - A (different) threshold for each pixel in the image is applied.
  - Local area information is exploited for the threshold of each pixel.

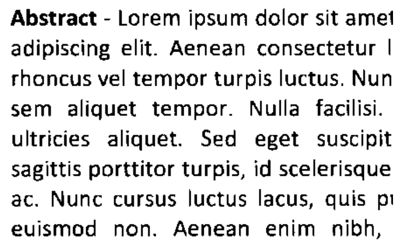
$$g(x,y) = \begin{cases} 1 & if \quad f(x,y) \geq t(x,y) \\ 0 & otherwise \end{cases}$$

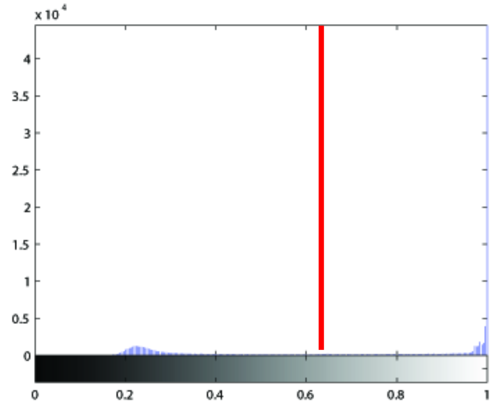
Binarization Methods

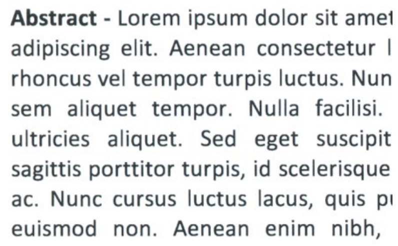
- Global methods are
  - Fast
  - Give good results if illumination is uniform and no noise is present
  - Fail if there are local changes in illumination
- Local methods are
  - Slow
  - Adapt to changes in illumination

Global Binarization Methods

- Global Methods are suitable for images with a bimodal gray value distribution.


Abstract - Lorem ipsum dolor sit amet adipiscing elit. Aenean consectetur l rhoncus vel tempor turpis luctus. Nun sem aliquet tempor. Nulla facilisi. ultricies aliquet. Sed eget suscipit sagittis porttitor turpis, id scelerisque ac. Nunc cursus luctus lacus, quis pı euismod non. Aenean enim nibh,

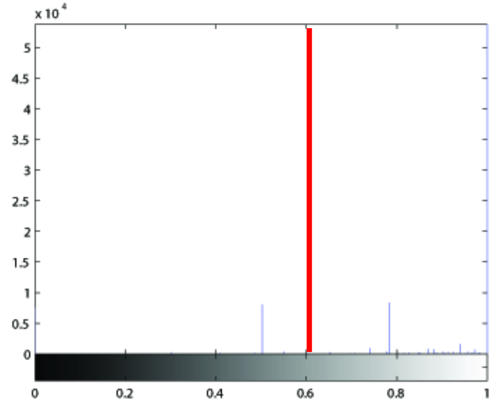



Abstract - Lorem ipsum dolor sit amet adipiscing elit. Aenean consectetur l rhoncus vel tempor turpis luctus. Nun sem aliquet tempor. Nulla facilisi. ultricies aliquet. Sed eget suscipit sagittis porttitor turpis, id scelerisque ac. Nunc cursus luctus lacus, quis pı euismod non. Aenean enim nibh,

This is a black text with RGB = (0,0,0)

This is a text with RGB = (128,128,128)

This is a text with RGB = (200,200,200)



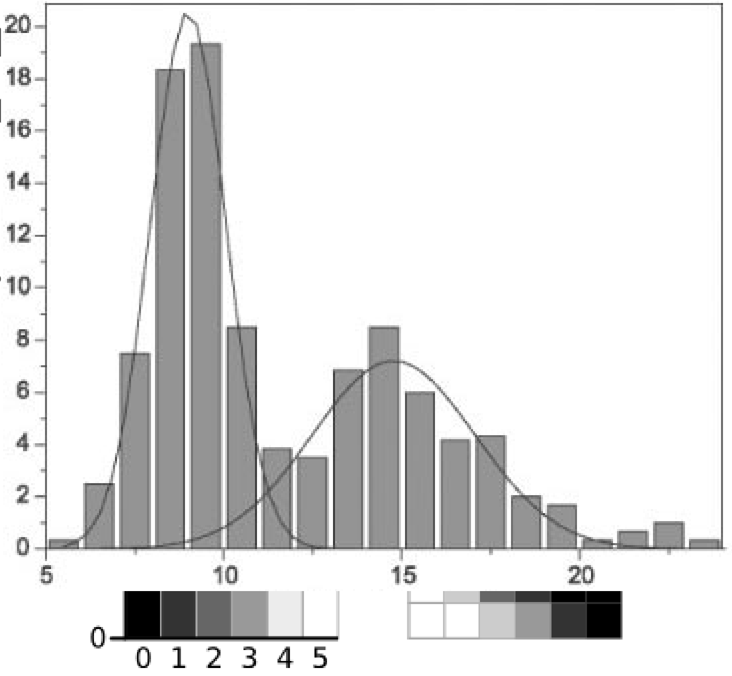
This is a black text with RGB = (0,0,0)

This is a text with RGB = (128,128,128)

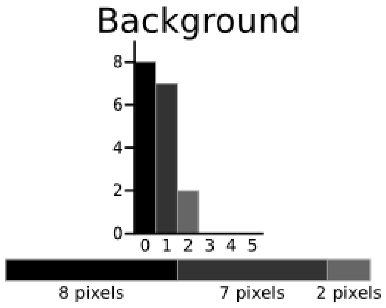
Global Binarization: Otsu

- Nobuyuki Otsu: *A threshold selection method from grey level histograms*. In: *IEEE Transactions on Systems, Man, and Cybernetics*. New York 9.1979, S.62–66. ISSN 1083-4419
- Otsu is a statistical method which assumes a bimodal histogram
  - Find a threshold that m n-class variance/maximizes the between-class varianc
  - Example:

    http://www.labbookpages. shold.html



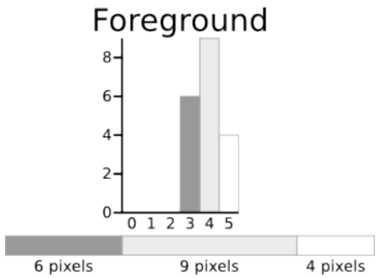
Otsu: Example I



$$\text{Weight} \quad W_b = \frac{8+7+2}{36} = 0.4722$$

$$\text{Mean} \quad \mu_b = \frac{(0\times 8)+(1\times 7)+(2\times 2)}{17} = 0.6471$$

$$\begin{aligned}\text{Variance} \quad \sigma_b^2 &= \frac{((0-0.6471)^2\times 8)+((1-0.6471)^2\times 7)+((2-0.6471)^2\times 2)}{17} \\ &= \frac{(0.4187\times 8)+(0.1246\times 7)+(1.8304\times 2)}{17} \\ &= 0.4637\end{aligned}$$



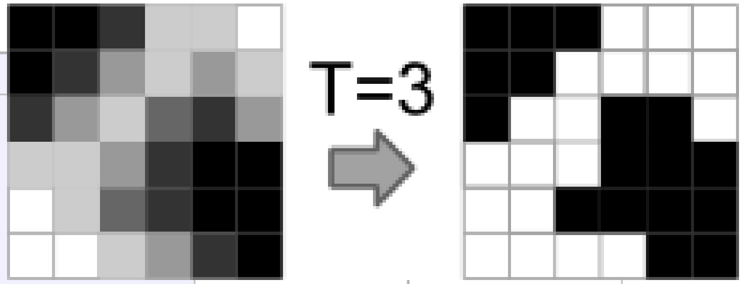
$$\text{Weight} \quad W_f = \frac{6+9+4}{36} = 0.5278$$

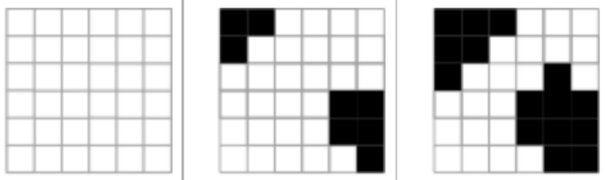
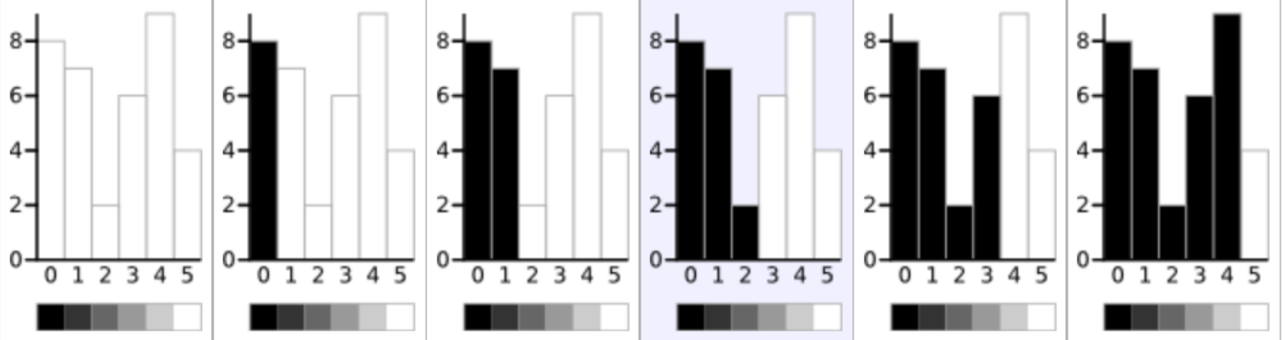
$$\text{Mean} \quad \mu_f = \frac{(3\times 6)+(4\times 9)+(5\times 4)}{19} = 3.8947$$

$$\begin{aligned}\text{Variance} \quad \sigma_f^2 &= \frac{((3-3.8947)^2\times 6)+((4-3.8947)^2\times 9)+((5-3.8947)^2\times 4)}{19} \\ &= \frac{(4.8033\times 6)+(0.0997\times 9)+(4.8864\times 4)}{19} \\ &= 0.5152\end{aligned}$$

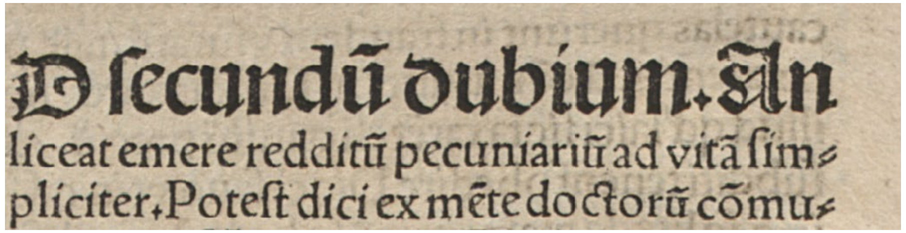
$$\begin{aligned}\text{Within Class Variance} \quad \sigma_W^2 = W_b\,\sigma_b^2 + W_f\,\sigma_f^2 &= 0.4722 * 0.4637 + 0.5278 * 0.5152 \\ &= 0.4909\end{aligned}$$

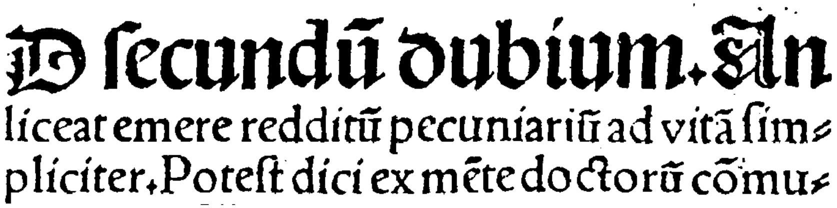
Otsu: Example II

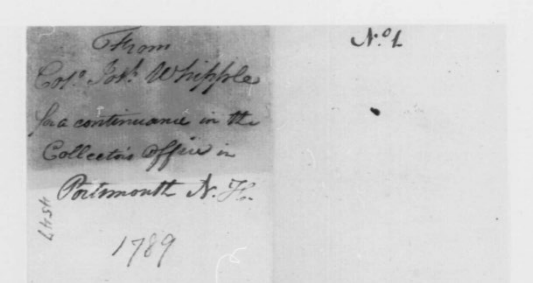


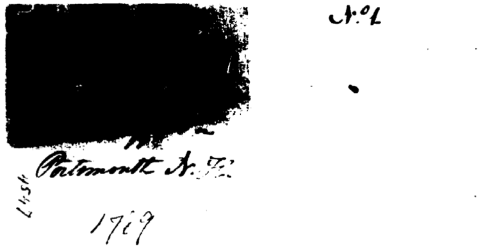
Threshold	T=0	T=1	T=2			
	 					
Weight, Background	$W_b = 0$	$W_b = 0.222$	$W_b = 0.4167$	$W_b = 0.4722$	$W_b = 0.6389$	$W_b = 0.8889$
Mean, Background	$M_b = 0$	$M_b = 0$	$M_b = 0.4667$	$M_b = 0.6471$	$M_b = 1.2609$	$M_b = 2.0313$
Variance, Background	$\sigma^2_b = 0$	$\sigma^2_b = 0$	$\sigma^2_b = 0.2489$	$\sigma^2_b = 0.4637$	$\sigma^2_b = 1.4102$	$\sigma^2_b = 2.5303$
Weight, Foreground	$W_f = 1$	$W_f = 0.7778$	$W_f = 0.5833$	$W_f = 0.5278$	$W_f = 0.3611$	$W_f = 0.1111$
Mean, Foreground	$M_f = 2.3611$	$M_f = 3.0357$	$M_f = 3.7143$	$M_f = 3.8947$	$M_f = 4.3077$	$M_f = 5.000$
Variance, Foreground	$\sigma^2_f = 3.1196$	$\sigma^2_f = 1.9639$	$\sigma^2_f = 0.7755$	$\sigma^2_f = 0.5152$	$\sigma^2_f = 0.2130$	$\sigma^2_f = 0$
Within Class Variance	$\sigma^2_W = 3.1196$	$\sigma^2_W = 1.5268$	$\sigma^2_W = 0.5561$	$\boldsymbol{\sigma^2_W = 0.4909}$	$\sigma^2_W = 0.9779$	$\sigma^2_W = 2.2491$

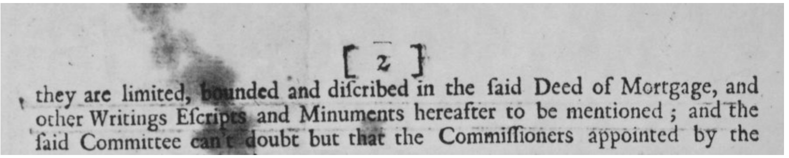
Binarization Results Otsu

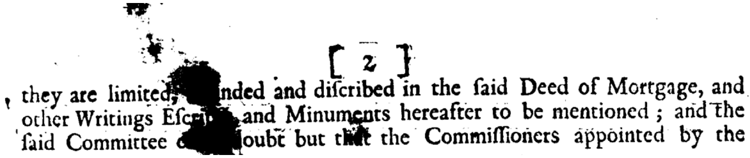

D secundũ dubium. An liceat emere redditũ pecuniariũ ad vitã sim- pliciter. Potest dici ex mẽte doctorũ cõmu-


D secundũ dubium. An liceat emere redditũ pecuniariũ ad vitã sim- pliciter. Potest dici ex mẽte doctorũ cõmu-


No 1
From Col. Jos. Whipple for a continuance in the Collector's Office in Portsmouth N.H.
1789




[2]
they are limited, bounded and discribed in the said Deed of Mortgage, and other Writings Escripts and Minuments hereafter to be mentioned; and the said Committee can't doubt but that the Commissioners appointed by the



Local Binarization Methods - Excerpt

- Niblack - define a threshold T based on the mean *m* and variance *s* within a local rectangular window by

$$T = m + k \cdot s$$

  - *Suggested windows size has to cover at least 1-2 characters according Gatos and Wolf et al.*
  - *Sensitive to noise*
- Sauvola and Pietkainen define a threshold T based on the mean *m*, *R* is the dynamic range of the standard deviation and *s* is the variance of the grayvalues. The parameter *k* is set to 0.5 by Sauvola and R is 128 (8 bit graylevel images).

$$T = m \cdot \left[1 + k \cdot \frac{s}{R - 1} \right]$$

  - *Less sensitive to background noise*
  - *Adaption of Niblacks algorithm*

Comparison

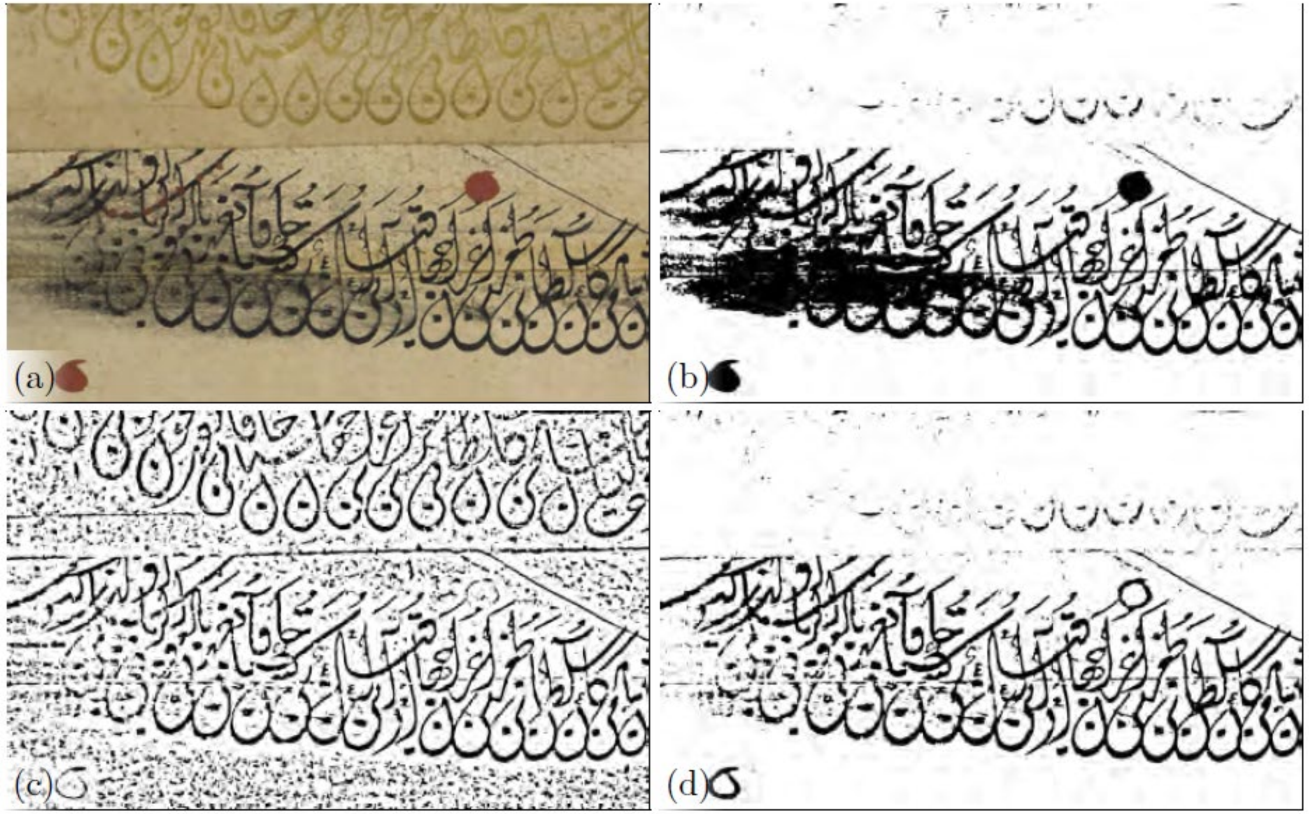
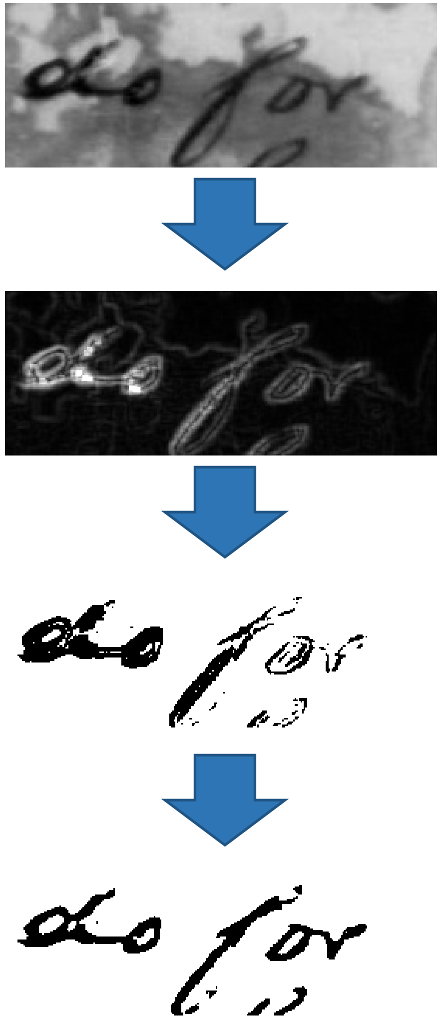


Figure 2.7: Binarization using global and local threshold techniques. (a) Original RGB image. (b) Result of Otsu [Ots79]. (c) Result of Niblack [Nib85]. (d) Result of Sauvola and Pietikäinen [SP00].

Local Binarization Methods – Su et al.

- Binarization of Historical Document Images Using the Local Maximum and Minimum, DAS 2010:
- Create Contrast Image
  - Makes use of the image contrast that is evaluated by using the local maximum and minimum within a local neighbourhood window.
- Detect High Contrast Points
  - Otsu‘s method is applied
- Generate final binary image
  - Each pixel is classified based on its intensity value compared with nearby high contrast points.



Su et al. – Contrast Image

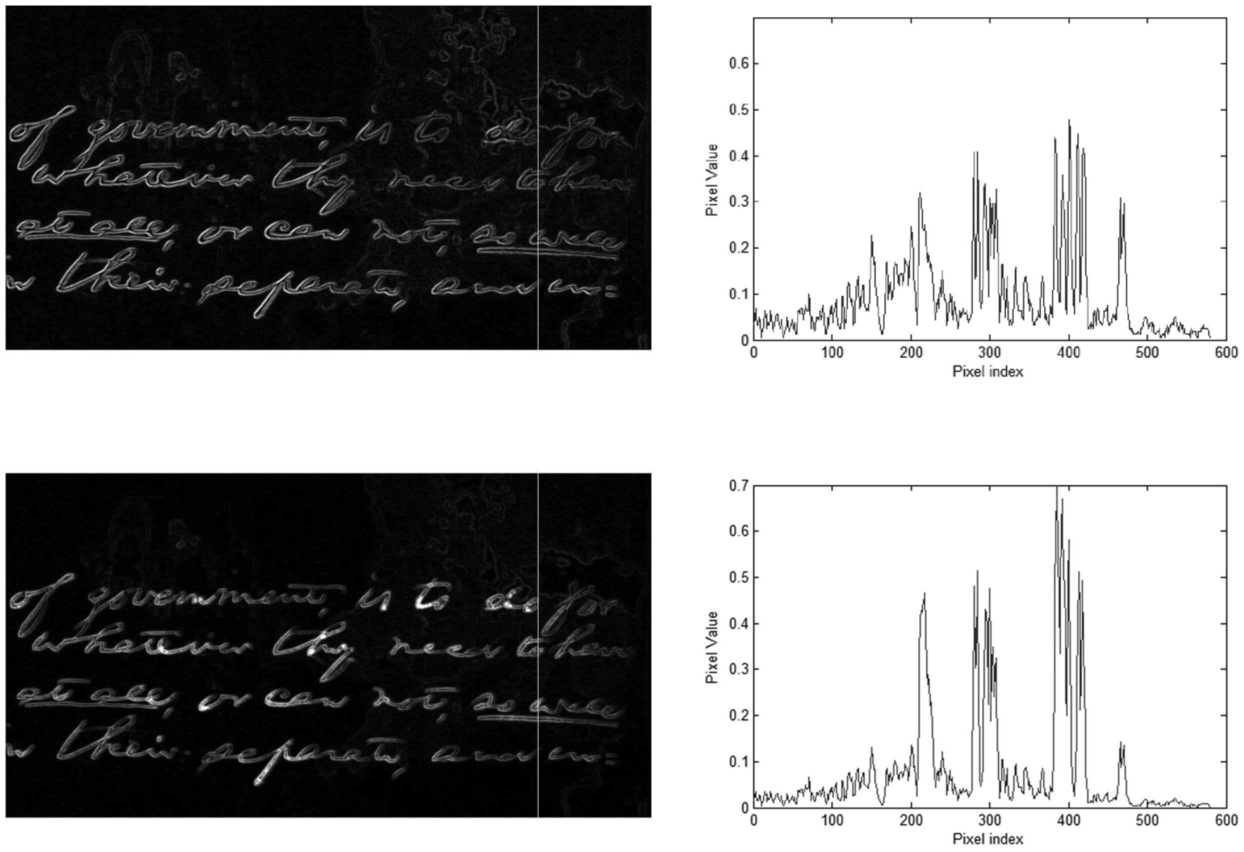
Local Contrast:

$$D(x,y)=\frac{f_{\max}(x,y)-f_{\min}(x,y)}{f_{\max}(x,y)+f_{\min}(x,y)}$$

Where f_{max} and f_{min} denote the maximum and minimum image intensities within a 3x3 local neighbourhood window.

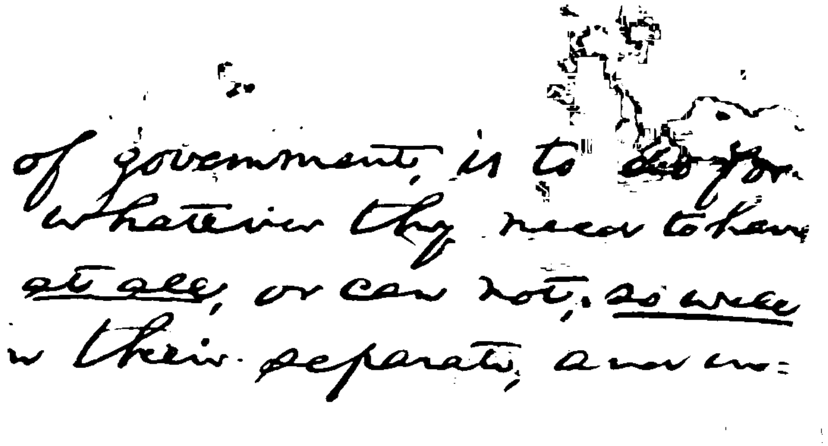
$$f_{\max}(x,y)+f_{\min}(x,y)$$

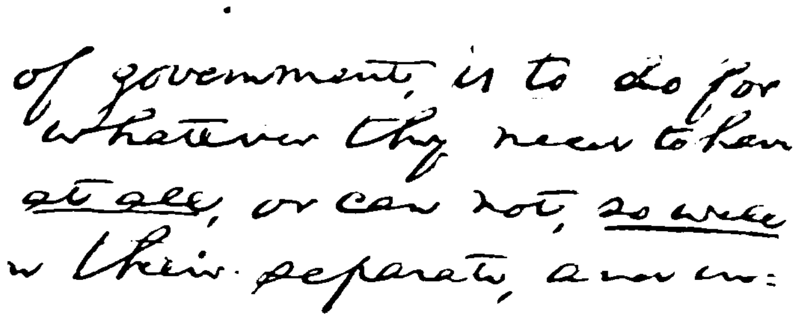
is a normalization factor that lowers the effect of the image contrast and brightness variation



Difference between image gradient and image contrast

Image Gradient vs. Image Contrast


of government, is to do for
whatever they need to have
at all, or can not, so well
in their separate, and in-


of government, is to do for
whatever they need to have
at all, or can not, so well
in their separate, and in-

Su et al. - High Contrast Point Detection

- High Contrast Pixel Detection
  - Apply Otsu to the Contrast Image
  - The text pixels should be close to the detected high contrast image pixels
  - The intensity of text pixels should be close to the intensity of nearby detected high contrast image pixels
  - The constructed contrast image has a clear bimodal pattern

Su et al. – Document Thresholding

- Document Thresholding

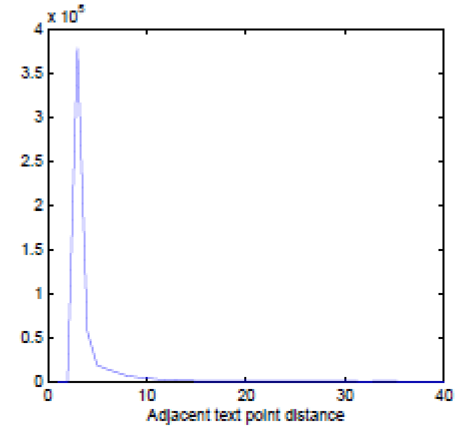
$$R(x,y) = \begin{cases} 1 & N_e \geq N_{\min} \quad and \quad I(x,y) \leq E_{mean} + E_{std}/2 \\ 0 & otherwise \end{cases}$$

$$E_{mean} = \frac{\sum_{neighbour} I(x,y) \cdot (1 - E(x,y))}{N_e} \qquad E_{std} = \sqrt{\frac{\sum_{neighbour} ((I(x,y) - E_{mean}) \cdot (1 - E(x,y)))^2}{N}}$$

E_{mean} and E_{std} are the mean and standard deviation of the image intensity of the detected high contrast image pixel
I is the input document image
E(x,y) refers to the binary high contrast pixel image. E(x,y) = 0 if I(x,y) is a high contrast pixel
N_e number of high contrast within local neighbourhood window

Su et al. Params

- Size of the neighbourhood window
- Minimum number of the high contrast image pixels N_{min}
- Both parameters are correlated to the width of the text strokes
- Size of the neighbourhood window should not be smaller than the text stroke width (text pixels within stroke will not be classified correctly otherwise because the local neighbourhood of high contrast pixels cannot be guaranteed)
- Minimum number should be around the size of the local neighbourhood window
- Estimate the text stroke width from the contrast image by building the adjacent high contrast points distance histogram.



Binarization Results - Comparison

Computer Vision Lab

Deep Learning based Binarization

- Hamza et al., 2005 (Shallow Network)
  - Use Self Organizing Maps (SOM) to cluster pixels based on (RGB) intensity values and the labeled output is used to train a MLP. The MLP classifies pixels in fore- and background.
- Afzal et al., 2015 (Deep Learning)
  - Use LSTM network which belongs to category of Recurrent Neural Networks (RNN)
- Westphal et al., 2018
  - Grid LSTM architecture (RNN architecture)
- Pastor-Pellicer et al., 2015
  - CNN architecture, with 2 conv. Layers and a MLP with 2 hidden layers and a output neuron

Deep Learning based Binarization

- Pastor-Pellicer et al., 2015

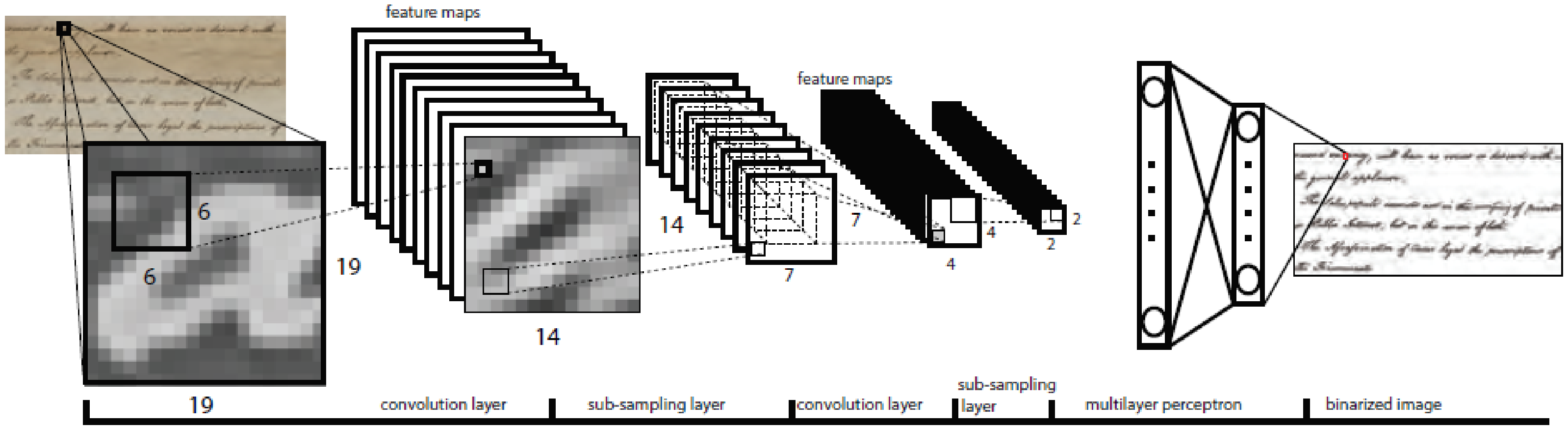


Figure 2.8: CNN architecture proposed Pastor-Pellicer et al. [PBZ+15]. Image taken from [PBZ+15].

Deep Learning based Binarization

- Tensmayer and Martinez, 2017
  - Suggest FCN for image binarization, Architecture has U-Shape
  - H-DIBCO performance improvement of 0.5% compared to 97.5% (p-FM)

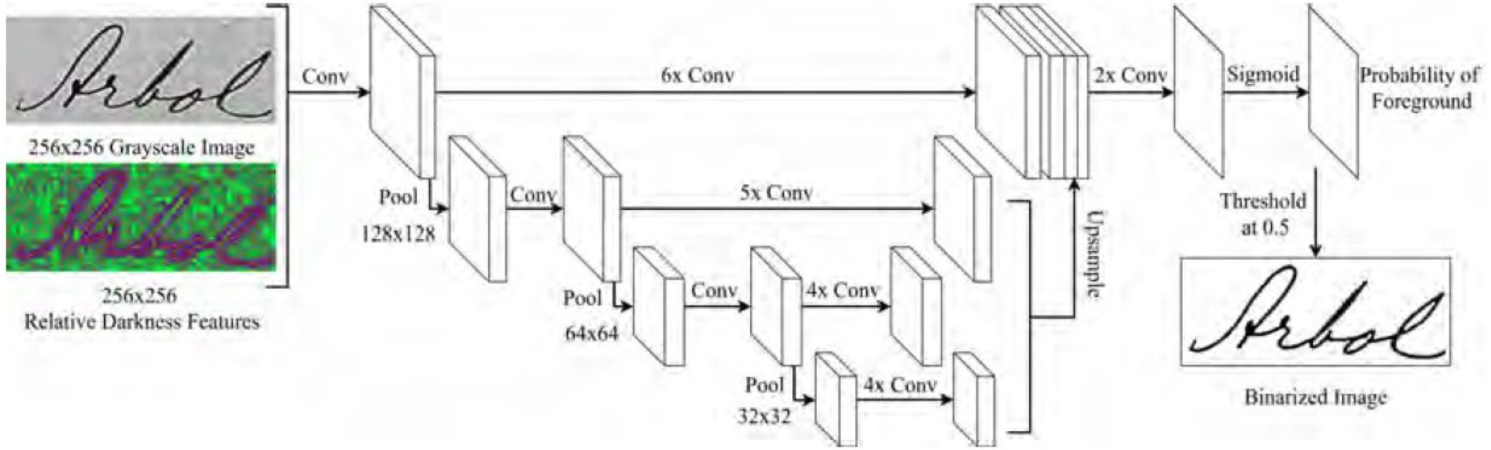


Figure 2.9: CNN architecture proposed by Tensmeyer and Martinez [TM17]. Image taken from [TM17].

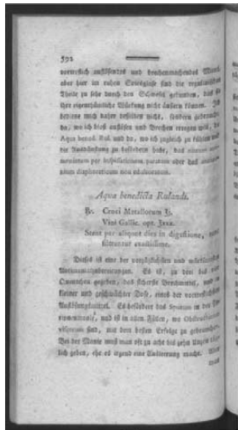
Deep Learning based Binarization

- Bezmaternykh et al., 2017
  - FCN, namely U-Net architecture
  - DIBCO 2017 winner
- DIBCO 2019: 4 out of 15 methods use a U-Net
  - Clustering approach was the winner although – this was attributed to the fact that a subset of the dataset has no similarities with previous datasets.
- Peng et al., 2017; Calvo-Zaragoza and Gallego, 2019
  - Auto-encoder method for binarization
- Tensmayer et al., 2019
  - Stress out the problem of limited training data
  - Propose a GAN to generate training data
  - Data is used to train a FCN
  - Outperforms in 6 out of 9 cases the winners

Dewarping

Document Dewarping

- Documents suffer from warping
  - Especially for camera-based acquisition
  - But also for scanners
- Two major solution categories:
  - 3D document shape reconstruction
  - 2D document image processing
    => rely only on 2D information



Scanner

Digital Camera

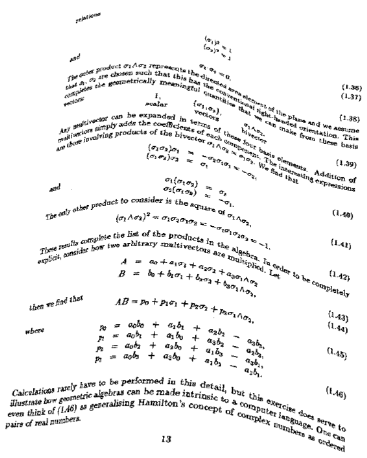
3D Shape Reconstruction

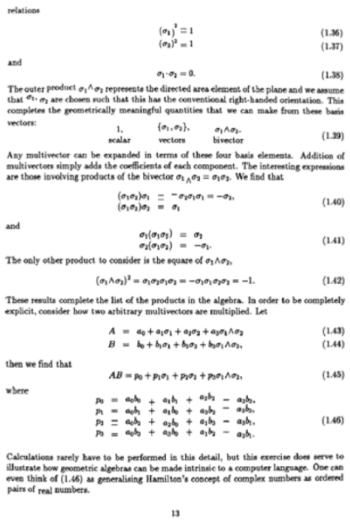
- With additional HW:
  - Laser scanners [Zhang08]
  - Stereo cameras [Ulges04]
  - Structured light [Brown04]
- Reconstruct 3D information from document:
  - Cylindrical model [Cao03]
    => image plane must be parallel to generatrix of page cylinder
  - Shape-from-shading [Tan06], [Zhang09]
    => requires knowledge of lighting

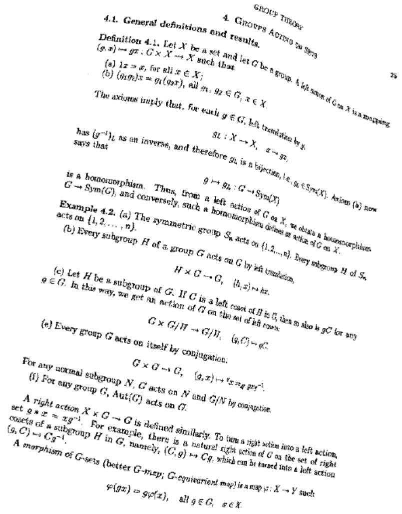
2D Document Image Processing

- Majority of these algorithms are based on detection of distorted text lines
  - Fit model to each text line (e.g. cubic B-splines, cubic splines)
- Some techniques emphasize on baseline finding
- Other techniques use
  - Document boundaries
  - Reference points

Dewarping Example



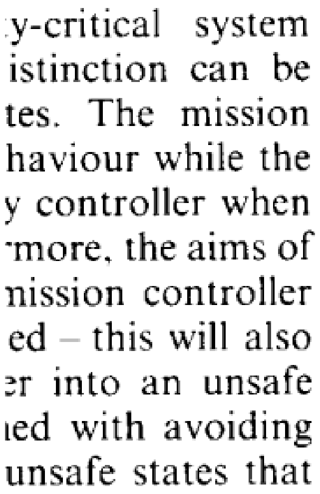


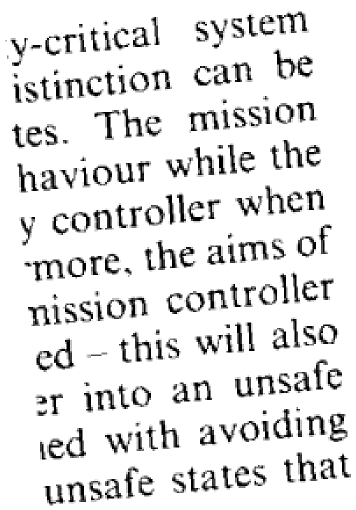


Skew Estimation

Skew Correction

- Pre-processing step of document layout analysis and OCR methods
- „*For Humans, rotated images are unpleasant for visualization and introduce extra difficulty in text reading*“ [Rafael Dueire Lins and Bruno Tenrio Avila, 2004]

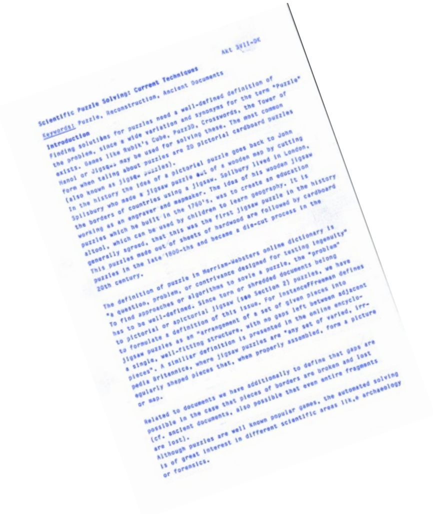

y-critical system
istinction can be
tes. The mission
haviour while the
y controller when
more, the aims of
nission controller
ed – this will also
er into an unsafe
ied with avoiding
unsafe states that

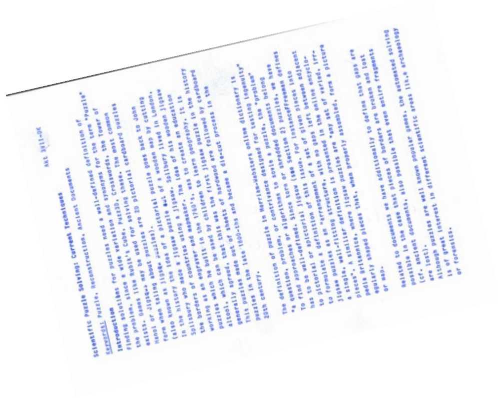

y-critical system
istinction can be
tes. The mission
haviour while the
y controller when
more, the aims of
nission controller
ed – this will also
er into an unsafe
ied with avoiding
unsafe states that

Skewed by 5°

Skew Correction

- Scanner – document not correctly aligned
- Scanner/copier – document feeder malfunction
- Document Image Acquisition with a camera system not perfectly aligned
- Document Images from Mobile Devices (e.g. smart phones)
  - Google Goggles, iBing Vision, ...





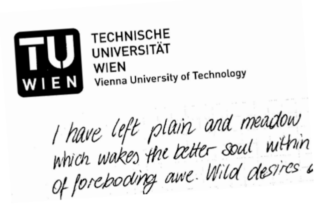
Current skew estimation algorithm have to be capable to deal with no restriction.

Skew Estimation Definition/Methods

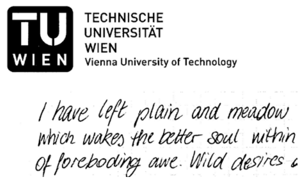
- Chen et al., 1995
  - *„The text skew angle of a document image is denoted by φ and is defined as its dominant (most frequently occurring) text baseline direction“*
  - Skew Estimation: *„Given a document image I with text lines at unknown skews φ1, φ2, ..., φn. Find φ‘, an estimate of the true text skew angle φ, to maximize the probability P(φ|I)“.*
- Methods
  - Up/Down Determination
  - Gradient based
  - Projection profile based
  - Analysis of the geometric distribution of CC (Connected Components)
  - Hough based Skew Estimation
  - Fourier based Skew Estimation
  - Distribution of local features

Skew Estimation problems

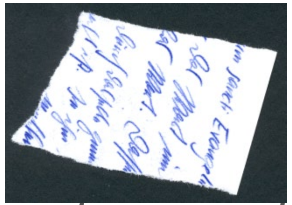
- Kavallieratou et al. 2002:
  - *Restriction of detectable angle range*
  - *Restriction on type or size of fonts*
  - *Dependence on page layout*
  - *A specific document resolution is required*
  - *Limitation to specific application*
  - *Large text areas are required*
  - *[...] Furthermore, the proposed algorithms can estimate the dominant skew angle and cannot deal with the cases of handwritten pages where the text lines may not be parallel to each other*

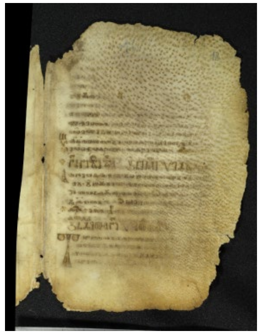


vs.



vs.



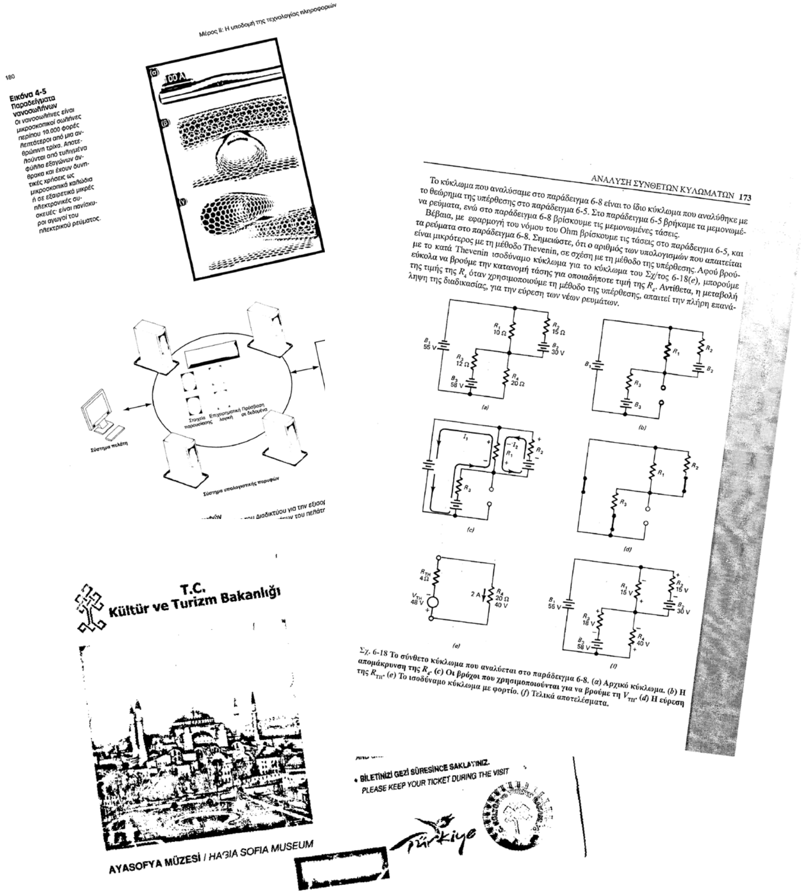


Document Image Skew Estimation Contest

- DISEC – Document Image Skew Estimation Contest
- Contest in conjunction with ICDAR 2013
- DISEC 2013 was the 1st Int. Document Image Skew Estimation Contest

DISEC 2013 Dataset

- Binary Images
- Rotated from -15° to +15°
- Images consist of pages of
  - Newspapers, scientific journals, travel guides, menus, comic books, literature books, …
  - Contain figures, tables, electrical circuits, architectural plans, …
- Different languages
  - English, Chinese, Greek, Japanese, Bulgarian, Russian, ancient Greek, …



DISEC 2013

Method	*AED*	*TOP80*	*CE*	*S*	*Overall Rank*
Ajou-SNU	2	2	2	6	2nd
Aria	8	12	12	32	11th
CMC-MSU	5	10	10	25	8th
CST-ECSU	10	11	11	32	11th
CVL-TUWIEN	4	5	6	15	5th
Gamera	5	4	3	12	4th
HIT-ICG-a	9	6	5	20	6th
HIT-ICG-b	10	9	9	28	9th
HP	12	8	8	28	9th
HS-Hannover	7	7	7	21	7th
LRDE-EPITA-a	1	1	1	3	1st
LRDE-EPITA-b	3	3	4	10	3rd

Method	Description
LRDE-EPITA-a	**EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France *(Jonathan Fabrizio):*** *This method uses the magnitude spectrum of a frequency Fourier transform to determine the orientation of the document image. The document image is preprocessed and all regions of the document are clustered using a KNN. At a next step, the Fourier transform is applied on the image to all clusters convex hull boundaries. In that way, in the frequency domain, the orientation is easier to be detected*
Anjou-SNU	**Ajou University, Suwon, Korea (*Hyung Il Koo*), *Seoul National University, Seoul, Korea (Nam Ik Cho):*** *This method estimates the skew by detecting straight lines in gray-scale and binary document images. therefore, it can take clues from text-lines, boundaries of figures, tables, vertical and horizontal separators as well as any combination of these entities. Specifically, a block-based edge detector that extracts several kinds of edges is developed. At a next step, straight lines are detected in edge maps and the skew angle is calculated by applying a maximum-likelihood estimation technique to the detected lines.*
LRDE-EPITA-b	**EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France *(E. Carlinet and J. Fabrizio):*** […] Based on the Line Segment Detector (LSD)

DISEC Evaluation Metrics

- E(j) is the distance of the skew estimation of the document image j and the GT

- Average Error Deviation (AED)

$$AED = \frac{\sum_{j=1}^{N} E(j)}{N}$$

- Average Error Deviation of the Top 80%

$$TOP80 = \frac{\sum_{j=1}^{M} sE(j)}{M}$$

sE(j) … $sorted\ Distances\ E(j)$ $in\ ascending\ order$

M … $0.8 * N$

DISEC Evaluation Metrics

- Percentage of Correct Estimations (CE)

$$CE = \frac{\sum_{j=1}^{N} K(j)}{N} \quad where \qquad K(j) = \begin{cases} 1 & if\ E(j) \leq 0.1 \\ 0 & otherwise \end{cases}$$

It is stated, that 0.1° was chosen due to the fact that a greater skew angle may be visible to a human observer

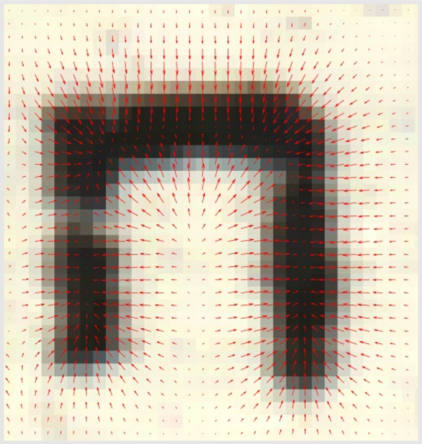
- For each criterion the ranking of each submitted method was calculated
- The final ranking is done by accumulating the ranking values for all criteria

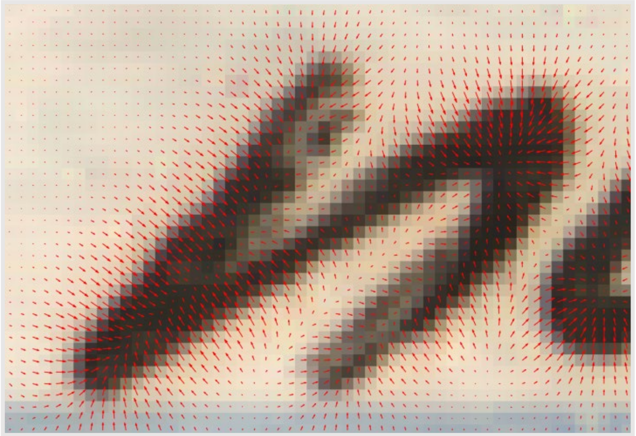
Up/Down Orientation

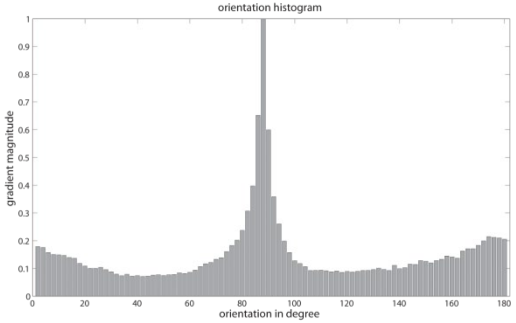
- Up/Down decision can be based on the statistical fact of the frequency of ascenders and descenders of roman letters
- Frequencies for English and German are shown in the following table
- See Caprari R., Algorithm for text age up/down orientation determination. Pattern Recogn. Lett., 21(4), pp. 311-317, 2000.

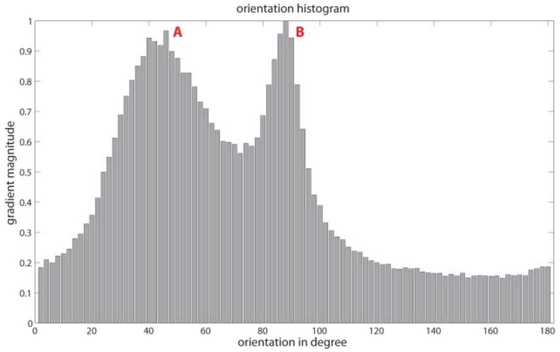
Stroke	Letters	Frequency English	Frequency German
Descender	j,p,q,y	4,15%	1,12%
Ascender	b,d,f,h,k,l,t	27,92%	24,19%
Neither	A,c,e,i,m,n,...	67,63%	74,96%

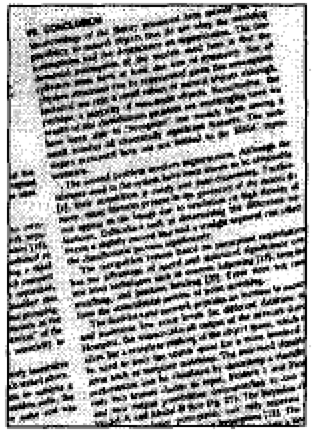
Gradient Based Methods

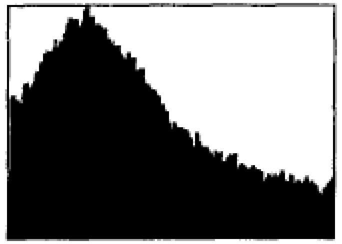








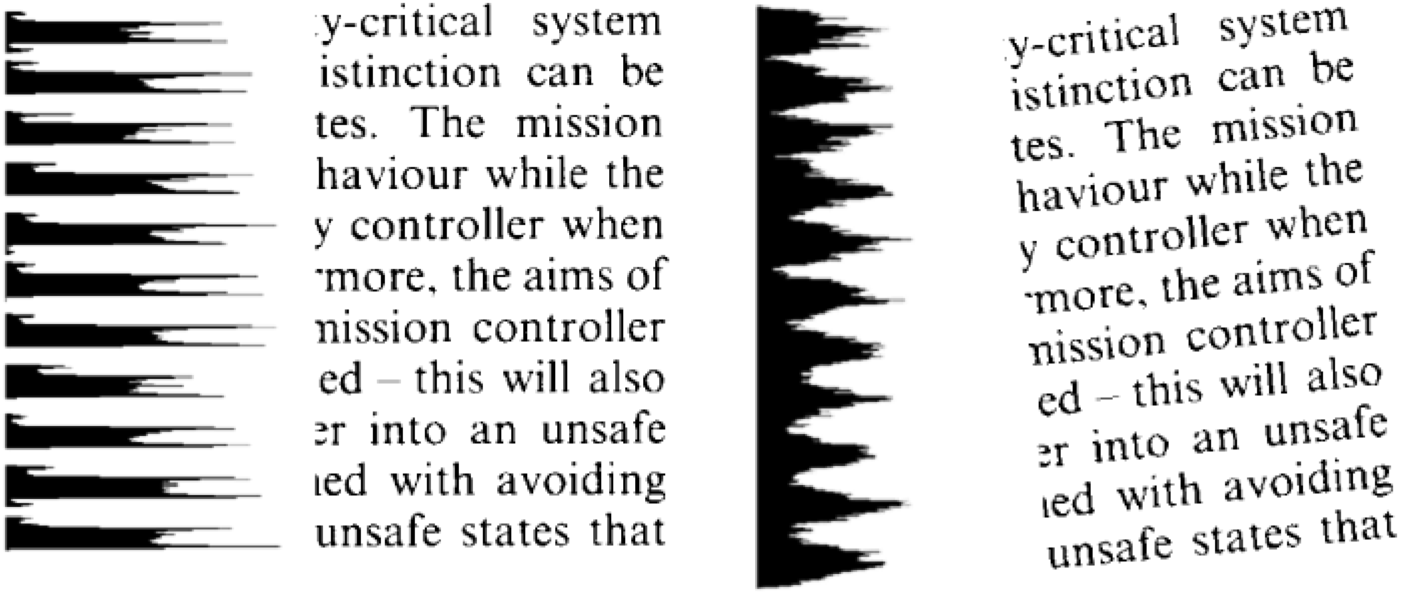




Sun C. And Si D., „*Skew and Slant Correction for Document Images Using Gradient Direction*“, 1997.

Projection Profile Based

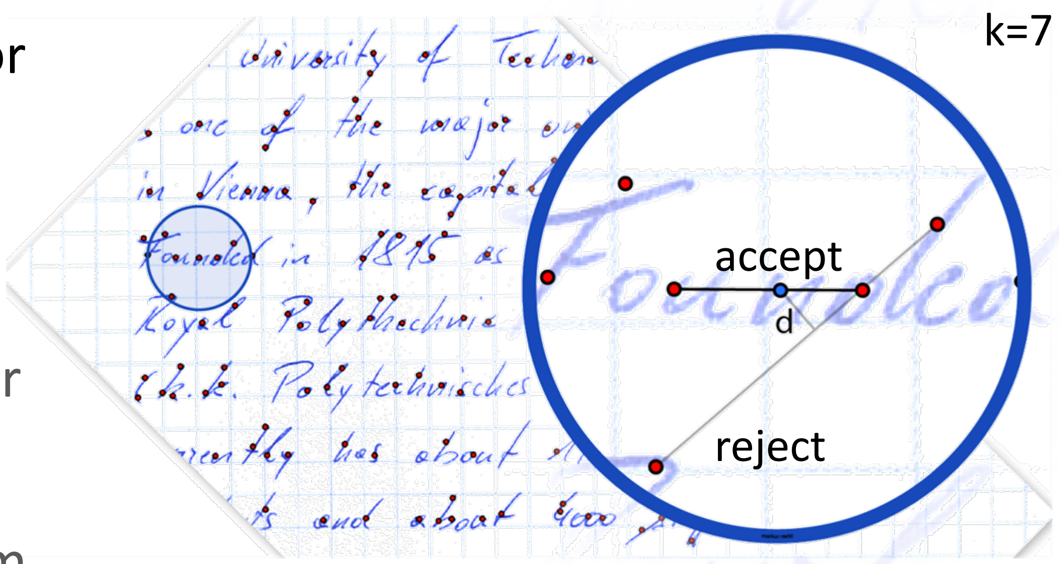
- Determine Horizontal Projection Profile
- Histogram has maximum amplitude and frequency when skew is 0° since the number of co-linear pixels is maximized



Focused Nearest Neighbour Clustering

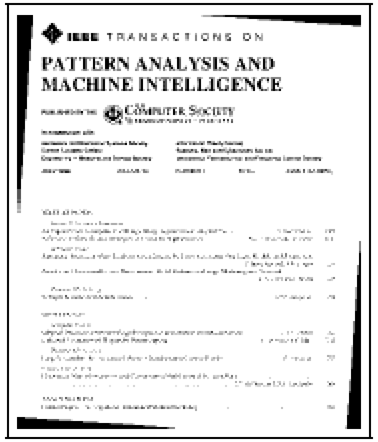
- Jiang et al., 1999
- Determine Feature Points
  - e.g. centroids of Connected Components (CC), Difference-of-Gaussians (DoG) interest points
- Allows determination of skew up to 180°

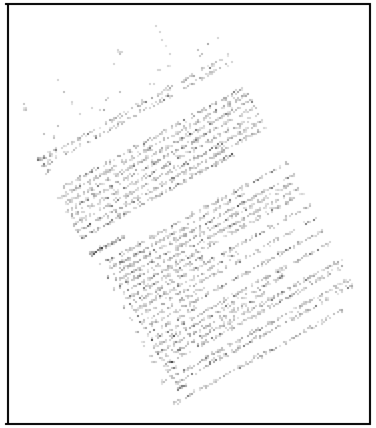
- Compute Local Skew angle for each feature point
  - Determine k nearest neighbours
  - Compute local skew line
  - Focused nearest-neighbour clustering (FNNC)
  - Each local skew angle is accumulated in a histogram

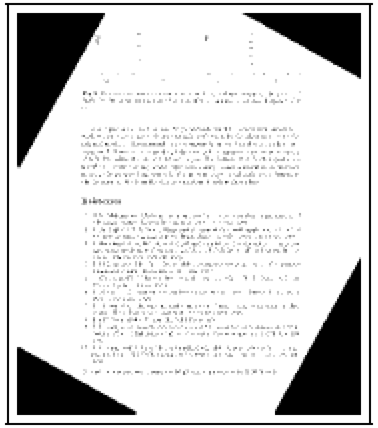


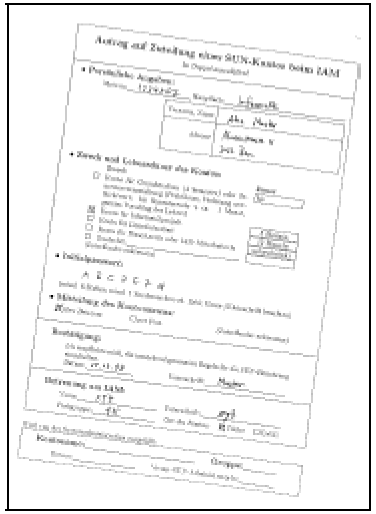
FNNC Results

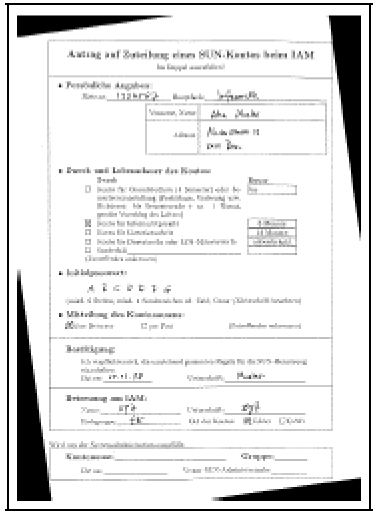

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE
COMPUTER SOCIETY

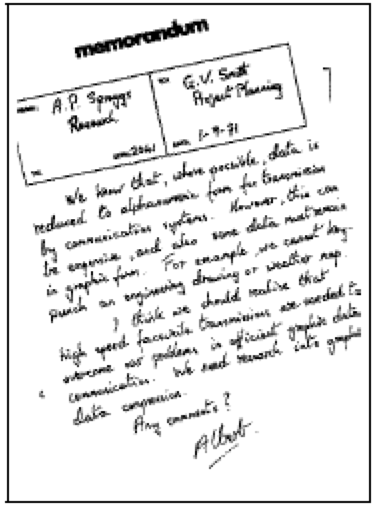

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE
COMPUTER SOCIETY

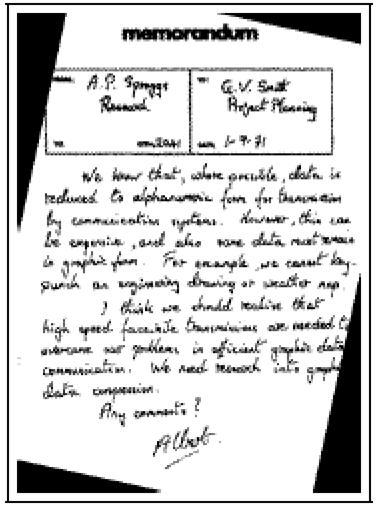





Antrag auf Zuteilung eines SUN-Kontos beim IAM


Antrag auf Zuteilung eines SUN-Kontos beim IAM

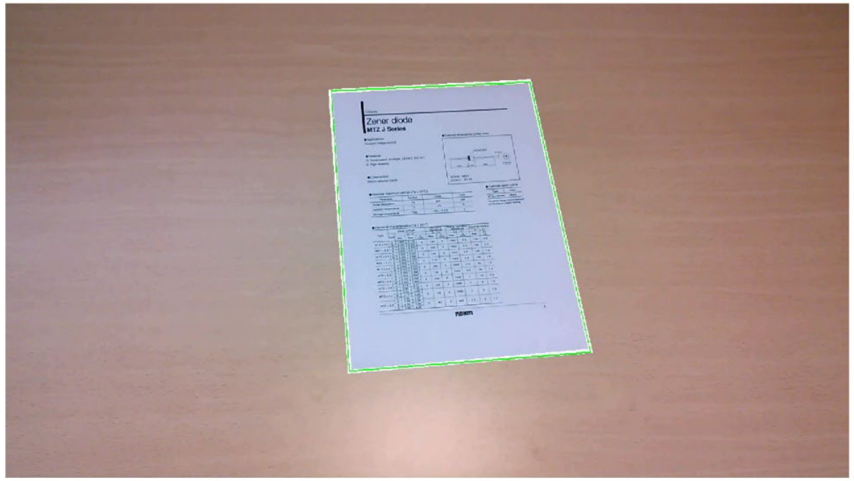

memorandum
A.P. Spriggs
Research
G.V. Smit
Project Planning
1-9-71
We know that, where possible, data is reduced to alphanumeric form for transmission by communication systems. However, this can be expensive, and also some data must remain in graphic form. For example, we cannot key-punch an engineering drawing or weather map.
I think we should realise that high speed facsimile transmissions are needed to overcome our problems in efficient graphic data communication. We need research into graphic data compression.
Any comments?
Albert

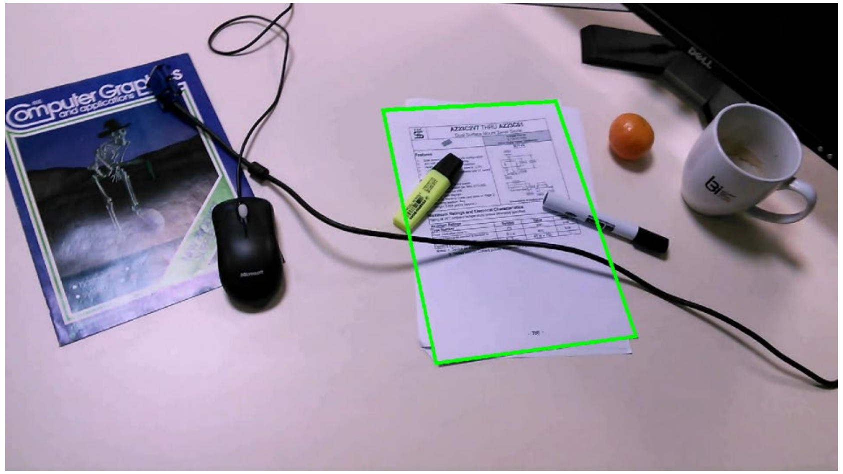

memorandum
A.P. Spriggs
Research
G.V. Smit
Project Planning
1-9-71
We know that, where possible, data is reduced to alphanumeric form for transmission by communication systems. However, this can be expensive, and also some data must remain in graphic form. For example, we cannot key-punch an engineering drawing or weather map.
I think we should realise that high speed facsimile transmissions are needed to overcome our problems in efficient graphic data communication. We need research into graphic data compression.
Any comments?
Albert

Page Segmentation

Page Segmentation

- Goal
  - Background removal (document segmentation)
  - Detect and segment the page outlines





Computer Vision Lab

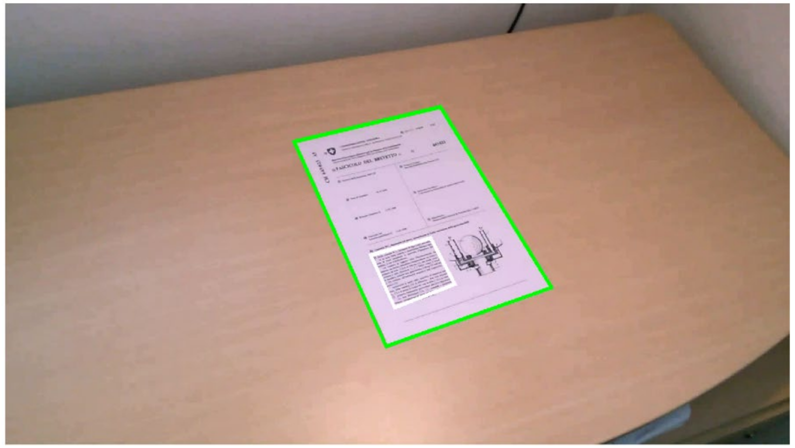
ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc)

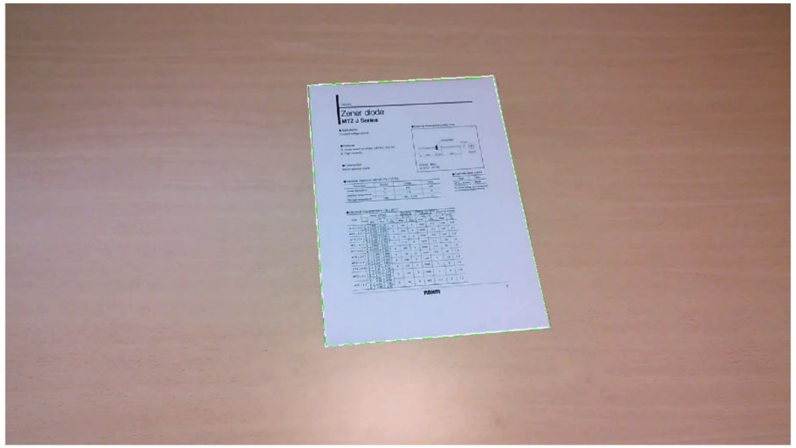
- Challenges
  - CHALLENGE-I: SMARTPHONE DOCUMENT CAPTURE (Page Detection)
  - CHALLENGE-2: SMARTPHONE OCR
- Evaluation for page segmentation
  - Jaccard Index (JI)

$$JI = \frac{area(GT \cap DP)}{area(GT \cup DP)}$$

  - where GT is the Ground Truth polygon and DP defines the Detected Polygon

Examples





SmartDoc @ICDAR2015

Winner of Method I

- LRDE (E. Carlinet and T. Geraud) [1]
  - Hierarchical representation of the image named Tree of Shapes
  - An energy on the tree is computed in order to select the shape that looks the most like a papersheet
- ISPL-CVML (S. Heo, H.I. Koo and N.I. Cho); 2nd place
  - Line Segment Detector is applied
  - Create document boundaries by selecting 2 horizontal and vertical segments that minimize a cost function exploiting color and egde features

- [1] Edwin Carlinet and Thierry Geraud. MToS: A tree of shapes for multivariate
- images. IEEE Transactions on Image Processing, 24(12):5330–5342,
- December 2015.

- SmartDoc @ICDAR2017 was cancelled

Next lecture: WI/WR 07.04.2025

The last Slide