Binarization Driven Blind Deconvolution for Document Image Restoration

37th German Conference on Pattern Recognition

Thomas Köhler, Andreas Maier, Vincent Christlein
08.10.2015
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen Graduate School in Advanced Optical Technologies (SAOT)

SAOT erlangen graduate school in advanced optical technologies

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Introduction

Typical applications for automatic text image analysis

- Optical character recognition (OCR) and handwritten text recognition (HTR)
- Writer identification and verification
- Structural document segmentation

Two common subproblems in these fields

- **Text image binarization** on high quality images for feature extraction
- **Image deconvolution** for text image restoration using to enhance reliability of features

Our proposition: Text binarization and deconvolution should be coupled and solved together instead of solving both problems separately

Blind Deconvolution for Text Document Images

Natural scene statistics vs. text-specific approaches

- Blind deconvolution using natural scene statistics
  - Total variation based priors [1]
  - Hyper-Laplacian priors [2]

$\Rightarrow$ Reasonable for natural images but fails to model the characteristics of text

- Text-specific blind deconvolution approaches:
  - Modeling priors by text-specific properties (contrast, color-uniformity, ...) [3, 4]
  - Convolutional Neural Networks (CNN) learning [5]

$\Rightarrow$ Consider deconvolution and binarization as independent subproblems

[1] Chan, T. F., & Wong, C. K. (1998). Total variation (TV) blind deconvolution. IEEE Transactions on Image Processing 7(3)
[2] Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms. In Proc. CVPR 2009.
[3] Cho, H., Wang, J. & Lee, S. (2012).Text Image Deblurring Using Text-Specific Properties. Proc. ECCV 2012
[4] Pan, J., Hu, Z., Su, Z., & Yang, M.-H. (2014). Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. In Proc. CVPR 2014
[5] Hradiš, M., Kotera, J., Zemčík, P. & Šroubek, F. (2015). Convolutional Neural Networks for Direct Text Deblurring, Proc. BMVC 2015

Coupling Blind Deconvolution and Binarization

„Combined" methods

- Directly recover deblurred binarization from blurred intensity image (but without recovering of a deblurred intensity image) [1]
- Intensity based clustering for regularization of TV based blind deconvolution [2]

$\Rightarrow$ Simplified models employed for image binarization (e. g. intensity based)

Proposed method: Binarization driven blind deconvolution

- Binarization and blind deconvolution in one common framework
- Feature-based binarization used as a prior for blind deconvolution
- Blind deconvolution used to refine the binarization

[1] Zhang, J. (2012). An Alternating Minimization Algorithm for Binary Image Restoration. IEEE Transactions on Image Processing 21(2)
[2] Lelandais B. & Duconge, F. (2015). Deconvolution regularized using fuzzy c-means algorithm for biomedical image deblurring and segmentation. Proc. ISBI 2015

Image Deconvolution Model

Image Deconvolution Model

Modeling of the image formation process

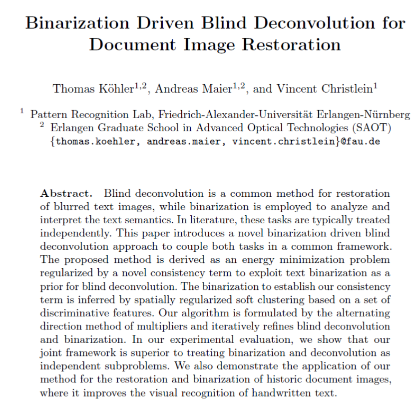
- Linear and space invariant convolution model:

$$\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \epsilon \tag{1}$$

  Formation of blurred image $\boldsymbol{y}$ from original image $\boldsymbol{x}$ with blur kernel $\boldsymbol{h}$ and additive noise ϵ

- We assume that $\boldsymbol{h}$ is unknown $\rightarrow$ blind deconvolution

- In binarization driven blind deconvolution:
  For each image $\boldsymbol{x}$ (and $\boldsymbol{y}$) there exists a corresponding binarization probability map $\boldsymbol{s}$
  - $s_i = 0$: i-th pixel belongs to a character
  - $s_i = 1$: i-th pixel belongs to a background



Binarization Driven Blind Deconvolution for Document Image Restoration

Thomas Köhler[1,2], Andreas Maier[1,2], and Vincent Christlein[1]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT)
{thomas.koehler, andreas.maier, vincent.christlein}@fau.de

Abstract. Blind deconvolution is a common method for restoration of blurred text images, while binarization is employed to analyze and interpret the text semantics. In literature, these tasks are typically treated independently. This paper introduces a novel binarization driven blind deconvolution approach to couple both tasks in a common framework. The proposed method is derived as an energy minimization problem regularized by a novel consistency term to exploit text binarization as a prior for blind deconvolution. The binarization to establish our consistency term is inferred by spatially regularized soft clustering based on a set of discriminative features. Our algorithm is formulated by the alternating direction method of multipliers and iteratively refines blind deconvolution and binarization. In our experimental evaluation, we show that our joint framework is superior to treating binarization and deconvolution as independent subproblems. We also demonstrate the application of our method for the restoration and binarization of historic document images, where it improves the visual recognition of handwritten text.

Original image $\boldsymbol{x}$

Binarization Driven Blind Deconvolution for Document Image Restoration

Blurred image $\boldsymbol{y}$

Energy Minimization Formulation

Blind deconvolution as joint energy minimization problem

- Common formulation of blind deconvolution as joint energy function [1]:

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{h}) = \mathcal{D}(\boldsymbol{x}, \boldsymbol{h}) + \lambda_x \mathcal{R}(\boldsymbol{x}) + \lambda_h \mathcal{H}(\boldsymbol{h}) \tag{2}$$

  $\mathcal{D}(\boldsymbol{x}, \boldsymbol{h})$ is the data fidelity term for the image $\boldsymbol{x}$ and the blur kernel $\boldsymbol{h}$,
  $\mathcal{R}(\boldsymbol{x})$ and $\mathcal{H}(\boldsymbol{h})$ are regularization terms with weights $\lambda_x \geq 0$, $\lambda_h \geq 0$

- Proposed formulation: Augment joint energy function with binarization driven consistency term

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{s}) = \mathcal{D}(\boldsymbol{x}, \boldsymbol{h}) + \lambda_x \mathcal{R}(\boldsymbol{x}) + \lambda_h \mathcal{H}(\boldsymbol{h}) + \lambda_c \mathcal{C}(\boldsymbol{x}, \boldsymbol{s}) \tag{3}$$

  $\mathcal{C}(\boldsymbol{x}, \boldsymbol{s})$ couples the image $\boldsymbol{x}$ with its binarization $\boldsymbol{s}$ with the weight $\lambda_c \geq 0$
  $\rightarrow$ Consistency term as additional prior for blind deconvolution

[1] Kotera, J., Šroubek, F. & Milanfar, Peyman. (2013). Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors. Proc. Computer Analysis of Images and Patterns

Definition of the Energy Terms

Deconvolution data fidelity and regularization terms

- Deconvolution data fidelity term $\mathcal{D}(\boldsymbol{x}, \boldsymbol{h})$ assuming additive white Gaussian noise:

$$\begin{aligned} \mathcal{D}(\boldsymbol{x}, \boldsymbol{h}) &= ||\boldsymbol{x} * \boldsymbol{h} - \boldsymbol{y}||_2^2 \\ &\equiv ||\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}||_2^2 \end{aligned} \tag{4}$$

- Image regularization term $\mathcal{R}(\boldsymbol{x})$ formulated as a Hyper-Laplacian prior:

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=1}^{n} \left([\nabla_h \boldsymbol{x}]_i^2 + [\nabla_v \boldsymbol{x}_i^2] \right)^{\frac{p}{2}} \quad \text{where } 0 < p \leq 1 \tag{5}$$

- Blur kernel regularization term to enforce non-negativity:

$$\mathcal{H}(\boldsymbol{h}) = \sum_{i=1}^{m} \mathcal{H}(h_i) \qquad \text{where } \mathcal{H}(h) = \begin{cases} h & h \geq 0 \\ \infty & h < 0 \end{cases} \tag{6}$$

Definition of the Energy Terms

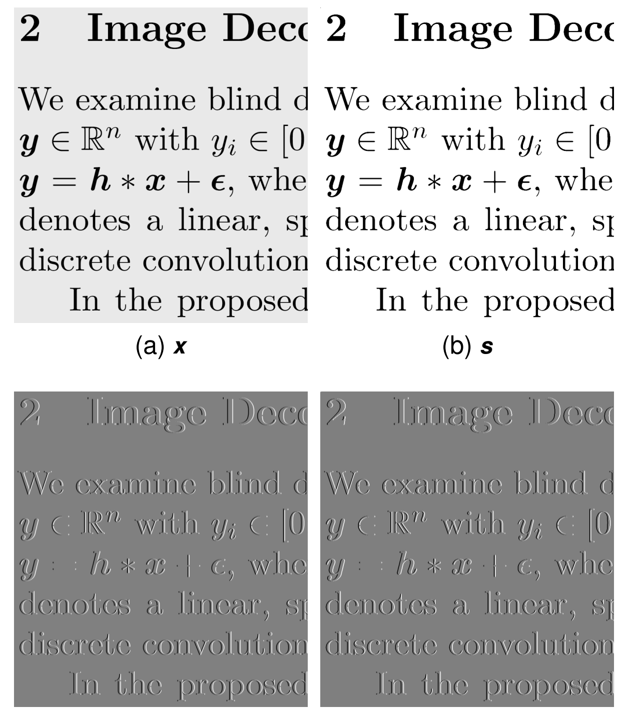
Binarization driven consistency term

- We exploit the fact that discontinuities in $\boldsymbol{x}$ and $\boldsymbol{s}$ should be aligned
- Formulation in the gradient domain:

$$\mathcal{C}(\boldsymbol{x}, \boldsymbol{s}) = ||\nabla_h \boldsymbol{x} - \nabla_h \boldsymbol{s}||_2^2 + ||\nabla_v \boldsymbol{x} - \nabla_v \boldsymbol{s}||_2^2 \quad (7)$$

- Not true in general for natural images but reasonable assumption for document images:
  - Exactly fulfilled in the background and inside characters
  - Gradients equal up to scale on boundaries

Text image and binarization with gradients:



(a) $\boldsymbol{x}$ (b) $\boldsymbol{s}$

(c) $\nabla_h \boldsymbol{x}$ (d) $\nabla_h \boldsymbol{s}$

Binarization Driven Blind Deconvolution Algorithm

erlangen graduate school in advanced optical technologies

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

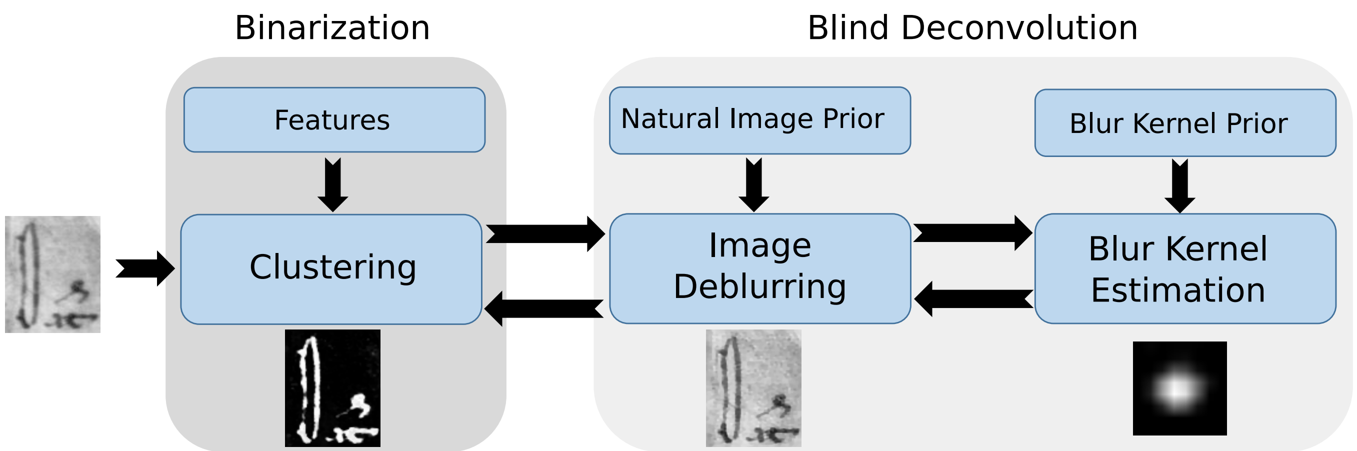
Energy Minimization Problem

Blind deconvolution via joint energy minimization

- Iteratively estimate the image $\boldsymbol{x}$ and the blur kernel $\boldsymbol{h}$:

$$\left(\boldsymbol{x}^{(t)}, \boldsymbol{h}^{(t)}\right) = \underset{\boldsymbol{x},\boldsymbol{h}}{\arg\min}\, \mathcal{E}\left(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{s}^{(t)}\right) \tag{8}$$

- Since the binarization $\boldsymbol{s}$ is unknown, we gradually refine $\boldsymbol{s}$ over the iterations
- Alternating minimization scheme to estimate all coupled variables



Estimation of the Binarization

Subproblem to update the binarization using spatially regularized soft clustering

$$(\boldsymbol{\mu}^{(t)}, \boldsymbol{s}^{(t)}) = \underset{\boldsymbol{\mu}, \boldsymbol{s}}{\arg\min} \left\{ \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{c} s_{ij}^{q} \left|\left| \boldsymbol{f}_{\boldsymbol{x},i} - \boldsymbol{\mu}_j \right|\right|_2^2}_{\text{Cluster data fidelity}} + \underbrace{\alpha \cdot \mathcal{R}_{\text{homogeneity}}(\boldsymbol{s})}_{\text{Cluster homogeneity}} \right\} \tag{9}$$

- Soft clustering procedure: alternating minimization for $\boldsymbol{\mu}$ and $\boldsymbol{s}$ [1]
  - Pixels in image $\boldsymbol{x}$ represented by feature set $\boldsymbol{f}_{\boldsymbol{x},1}, \ldots, \boldsymbol{f}_{\boldsymbol{x},n}$ with $\boldsymbol{f}_{\boldsymbol{x},i} \in \mathbb{R}^d$ and clusters represented by their centers $\boldsymbol{\mu}_j \in \mathbb{R}^d$
  - Membership degree of pixel i to cluster j modeled by s_{ij} (the binarization probability map) with weighting parameter $q > 1$

- In our approach: $c = 2$ clusters (background/characters) and scale space analysis using median filtering for feature extraction over $d = 3$ levels

[1] Yang, Y. & Huang, S. (2012). Image Segmentation by Fuzzy C-Means Clustering Algorithm with a Novel Penalty Term. Computing and Informatics 26(1).

Estimation of the Deblurred Image

Subproblem to update the deblurred image in the intensity domain

$$\boldsymbol{x}^{(t)} = \arg\min_{\boldsymbol{x}} \left\{ \mathcal{D}\left(\boldsymbol{x}, \boldsymbol{h}^{(t-1)}\right) + \lambda_x \mathcal{R}(\boldsymbol{x}) + \lambda_c \mathcal{C}\left(\boldsymbol{x}, \boldsymbol{s}^{(t)}\right) \right\} \tag{10}$$

- Alternating direction method of multipliers (ADMM) for efficient solution:

$$\begin{aligned} \arg\min_{\boldsymbol{x}, \boldsymbol{v}_h, \boldsymbol{v}_v} \Big\{ & ||\boldsymbol{H}^{(t-1)}\boldsymbol{x} - \boldsymbol{y}||_2^2 \underbrace{+ \lambda_v \left(||\boldsymbol{v}_h - \nabla_h \boldsymbol{x} - \boldsymbol{b}_h||_2^2 + ||\boldsymbol{v}_v - \nabla_v \boldsymbol{x} - \boldsymbol{b}_v||_2^2 \right)}_{\text{penalty terms with Lagrangian multiplier } \lambda_v} \\ & + \lambda_x \sum_{i=1}^{n} \left([\boldsymbol{v}_h]_i^2 + [\boldsymbol{v}_v]_i^2 \right)^{\frac{p}{2}} + \lambda_c \left(||\boldsymbol{v}_h - \nabla_h \boldsymbol{s}^{(t)}||_2^2 + ||\boldsymbol{v}_v - \nabla_v \boldsymbol{s}^{(t)}||_2^2 \right) \Big\} \end{aligned} \tag{11}$$

- Alternating minimization for $\boldsymbol{x}$ (in the Fourier domain) and auxiliary variables $\boldsymbol{v}_h$ and $\boldsymbol{v}_v$ (using soft thresholding and look-up tables)
- Bregman variables $\boldsymbol{b}_h$ and $\boldsymbol{b}_v$ updated per iteration [1]

[1] Goldstein, T., & Osher, S. (2009). The Split Bregman Method for L1-Regularized Problems. SIAM Journal on Imaging Sciences, 2(2)

Estimation of the Blur Kernel

Subproblem to update the blur kernel in the gradient domain

$$\boldsymbol{h}^{(t)} = \arg\min_{\boldsymbol{h}} \left\{ \mathcal{D}\left(\nabla \boldsymbol{x}^{(t)}, \boldsymbol{h}\right) + \lambda_h \mathcal{H}(\boldsymbol{h}) \right\} \quad (12)$$

- ADMM iterations for efficient solution:

$$\arg\min_{\boldsymbol{h},\boldsymbol{g}} \left\{ ||\nabla \boldsymbol{X}^{(t)} \boldsymbol{h} - \nabla \boldsymbol{y}||_2^2 + \lambda_h \mathcal{H}(\boldsymbol{g}) + \underbrace{\lambda_g ||\boldsymbol{h} - \boldsymbol{g} - \boldsymbol{b}_g||_2^2}_{\text{penalty term with Lagrangian multiplier } \lambda_g} \right\} \quad (13)$$

- State-of-the-art optimization scheme for kernel estimation adopted from Kotera et al. [1]

[1] Kotera, J., Šroubek, F. & Milanfar, Peyman. (2013). Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors. Proc. Computer Analysis of Images and Patterns

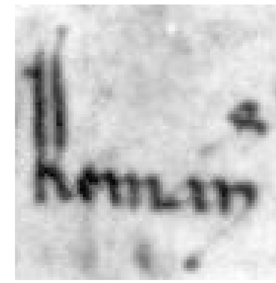
Overall Numerical Optimization

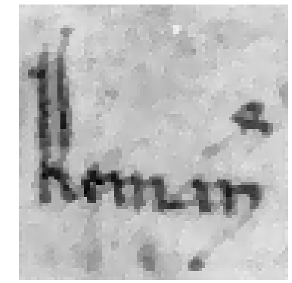
Coarse-to-fine optimization for $\boldsymbol{x}$, $\boldsymbol{h}$ and $\boldsymbol{s}$

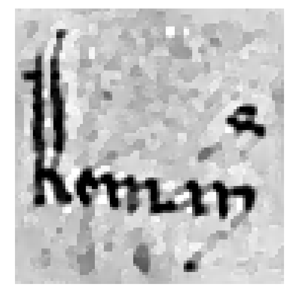
1. Estimate only $\boldsymbol{x}$ and $\boldsymbol{h}$ without binarization consistency ($K \geq 1$ levels)
2. Estimate $\boldsymbol{x}$, $\boldsymbol{h}$ and $\boldsymbol{s}$ with binarization consistency (on finest level)
3. Refinement of deblurred image (ringing and noise removal) by means of guided filtering using binarization as guidance image:

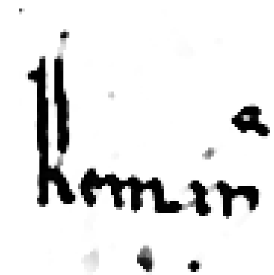
$$\boldsymbol{x} = \frac{1}{2}\Big(\mathrm{GF}(\tilde{\boldsymbol{x}}, \boldsymbol{s}) + \mathrm{GF}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}})\Big) \qquad (14)$$

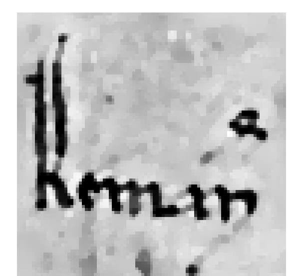
GF($\boldsymbol{p}$, $\boldsymbol{q}$): guided filter with input image $\boldsymbol{p}$ and guidance image $\boldsymbol{q}$


(a) Original


(b) Stage 1


(c) Stage 2


(d) Stage 2 (bin.)


(e) Stage 3

Experiments and Results

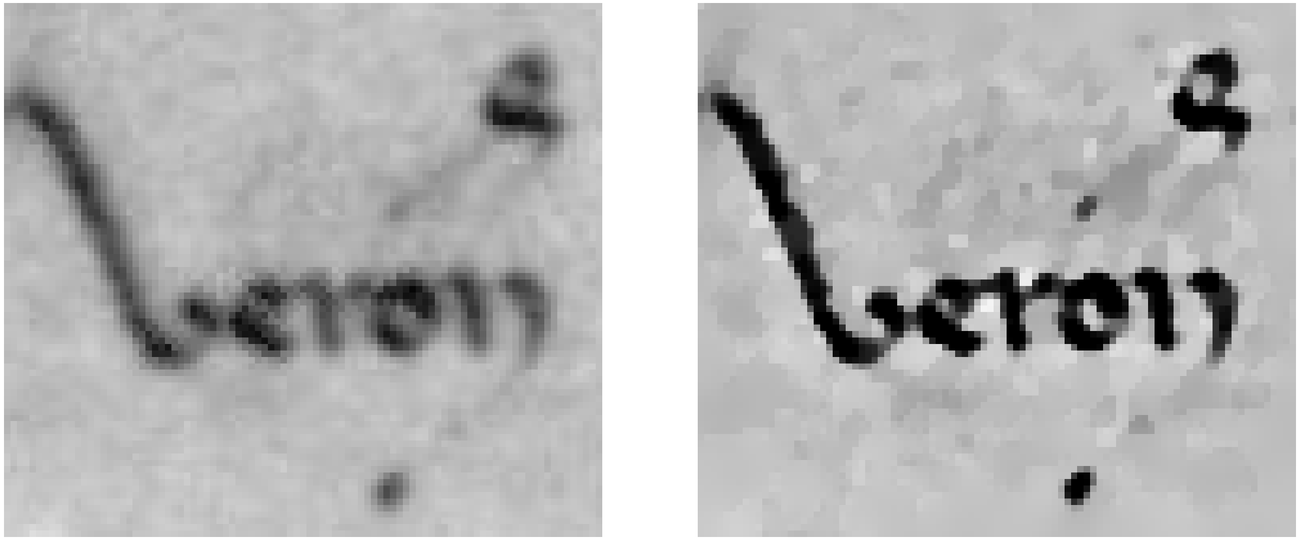
erlangen
graduate school in
advanced optical technologies

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Experiments for Image Blind Deconvolution

Comparison to different state-of-the-art blind deconvolution algorithms

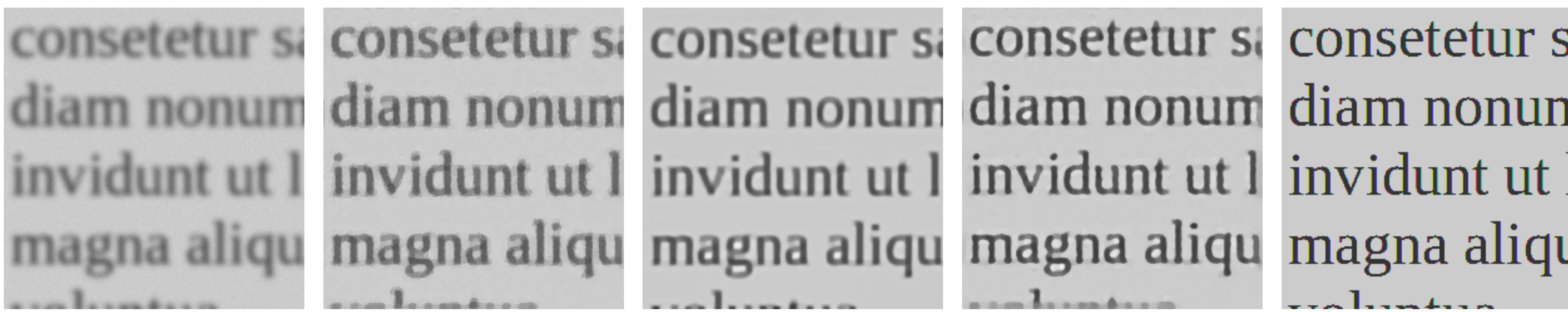
- Natural scene statistics approach of Kotera et al.
- Text-specific approach proposed by Pan et al.



Results on Simulated Images

Simulated images disturbed by out-of-focus blur and Gaussian noise

- Example for fixed noise standard deviation ($\sigma_n = 0.01$):



(a) Original (b) Kotera et al. (c) Pan et al. (d) Proposed (e) Ground truth

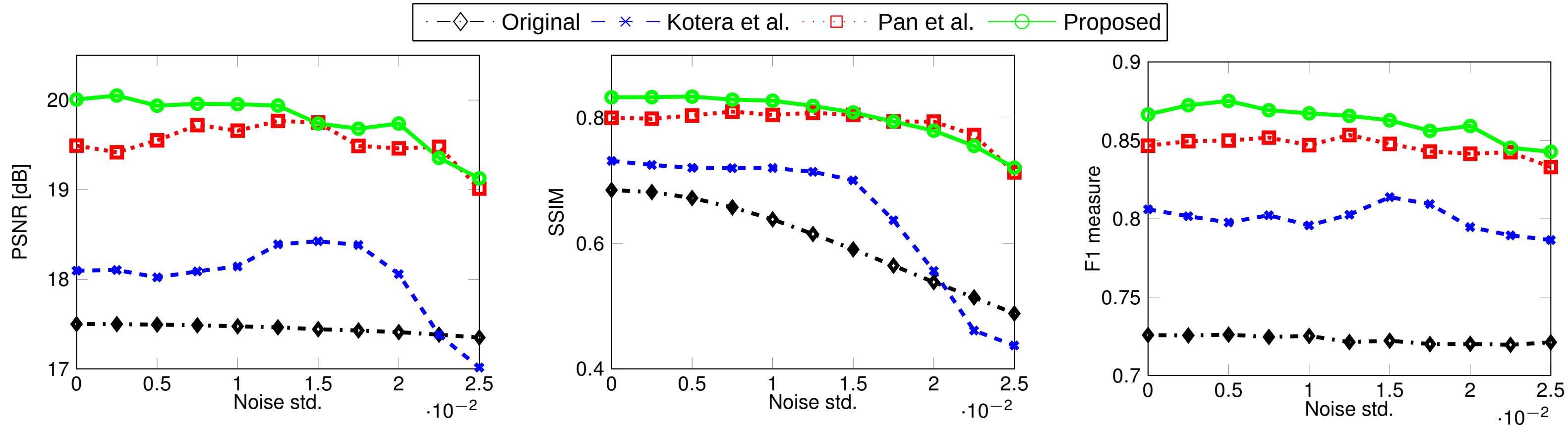
- Performance measures for fixed noise level averaged over 18 test images:

	PSNR (in dB)	SSIM	F1 measure*
Original	17.57 ± 0.42	0.64 ± 0.02	0.72 ± 0.01
Kotera et al.	18.30 ± 0.54	0.73 ± 0.03	0.79 ± 0.02
Pan et al.	19.79 ± 0.49	0.81 ± 0.02	0.85 ± 0.02
Proposed	**20.08 ± 0.61**	**0.83 ± 0.02**	**0.87 ± 0.02**

*) Binarizations for Kotera et al. and Pan et al. obtained in two-stage approach: blind deconvolution followed by thresholding using Otsu's method

Noise Robustness Results

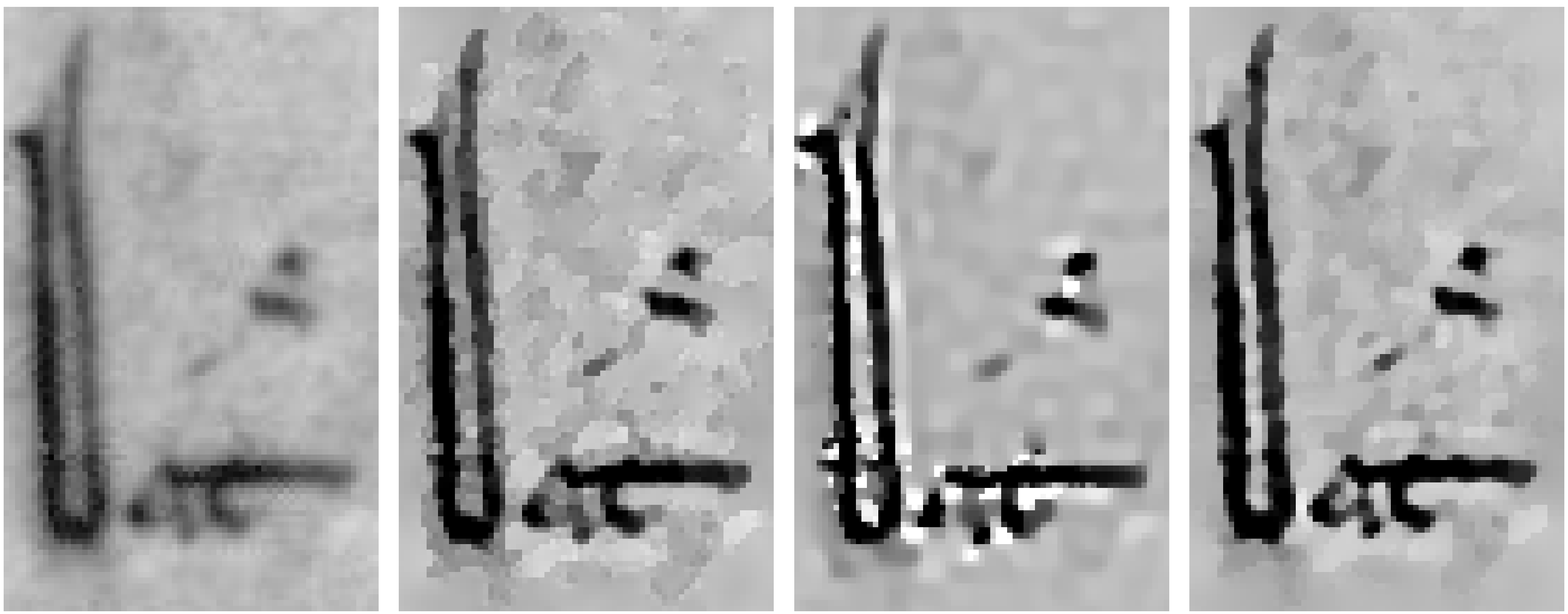
Simulated images with varying amount of Gaussian noise



- For small noise levels: binarization driven blind deconvolution substantially outperformed the state-of-the-art
- For moderate/large noise levels: competitive to the method of Pan et al. but improved robustness compared the method of Kotera et al.
- Our binarization outperformed two-stage blind deconvolution and binarization

Results on Real Document Images

Application: restoration of scanned handwritten historical documents [2]



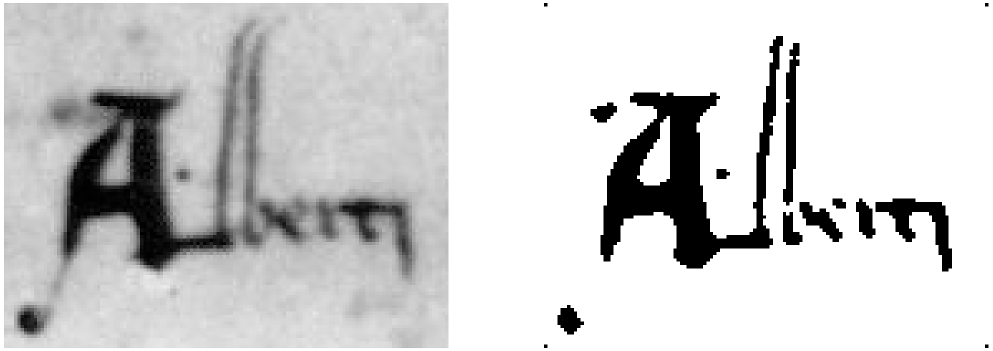
(a) Input | (b) Kotera et al. | (c) Pan et al. | (d) Proposed

[2]Image source: Göttingen Academy of Sciences & Humanities

Experiments for Image Binarization

Evaluation of the performance of our binarization method

- Comparison to established image binarization algorithms:
  - Global thresholding using adaptive threshold selection by Otsu's method
  - Local methods of Sauvola & Pietikäinen [1], Su et al. [2] and Bradley & Roth [3]
- Binarization directly on the original, blurred images and on deblurred images



[1] Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. Pattern Recognition, 33(2)

[2] Su, B., Lu, S., & Tan, C. L. (2010). Binarization of historical document images using the local maximum and minimum. In Proc. 8th IAPR International Workshop on Document Analysis Systems

[3] Bradley, D., & Roth, G. (2007). Adaptive Thresholding using the Integral Image. Journal of Graphics, GPU, and Game Tools, 12(2)

Results on Simulated & Real Document Images

F1 measures for binarization on simulated and handwritten document images

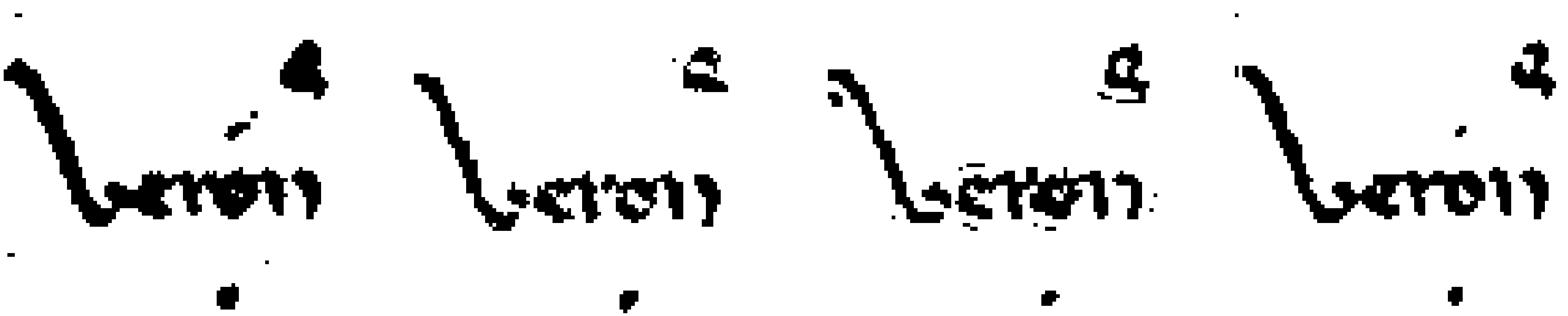
	Artificial	**Handwritten**	**Merged**
Otsu	0.72 ± 0.01	0.75 ± 0.15	0.73 ± 0.11
Sauvola	0.79 ± 0.04	$\mathbf{0.78} \pm \mathbf{0.07}$	0.79 ± 0.06
Bradley	$\mathbf{0.85} \pm \mathbf{0.01}$	$\mathbf{0.78} \pm \mathbf{0.08}$	$\mathbf{0.82} \pm \mathbf{0.07}$
Su	0.80 ± 0.01	$\mathbf{0.79} \pm \mathbf{0.09}$	0.80 ± 0.07
Kotera + Bradley	0.82 ± 0.01	0.71 ± 0.10	0.76 ± 0.09
Pan + Bradley	0.85 ± 0.02	0.73 ± 0.09	0.79 ± 0.09
Proposed	$\mathbf{0.87} \pm \mathbf{0.02}$	0.76 ± 0.11	$\mathbf{0.81} \pm \mathbf{0.10}$

- Comparison to global/local thresholding techniques:
  - On simulated data: best F1 measure by our method
  - On real data: better F1 measures by local thresholding techniques
- Our binarization outperformed two-stage deconvolution and binarization (Kotera + Bradley, Pan + Bradley)

Results on Simulated & Real Document Images

Comparison of text binarization on example document image

(a) Reference (b) Otsu (c) Sauvola (d) Su



(e) Bradley (f) Kotera + Bradley (g) Pan + Bradley (h) Proposed

Summary and Conclusion

erlangen
graduate school in
advanced optical technologies

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Conclusion

Novel binarization driven blind deconvolution for text images

- Couples blind deconvolution and binarization in a common framework
- Outperforms state-of-the-art blind deconvolution based on natural image statistics and text-specific properties
- Text binarization as a by-product that is competitive to state-of-the-art local binarization techniques
- Outperforms two-stage approach using blind deconvolution followed by image binarization

Future Work

Applications and extensions of the proposed method

- Applications:
  Binarization driven deconvolution as preprocessing for text image analysis (HTR, OCR, . . .)
- Augment feature-based clustering:
  Comprehensive set of text-specific features for text image binarization
- Enhance blur kernel estimation:
  Text binarization as guidance for kernel estimation
- Investigation of binarization consistency terms:
  Priors proposed for multi-channel image reconstruction [1]

[1] Köhler, T., Jordan, J., Maier, A., & Hornegger, J. (2015). A Unified Bayesian Approach to Multi-Frame Super-Resolution and Single-Image Upsampling in Multi Sensor Imaging. Proc. BMVC 2015.

Thank you very much for your attention!

Thank you very much for your attention!

Thank you very much for your attention!