

Binarization Driven Blind Deconvolution for Document Image Restoration

Thomas Köhler^{1,2}, Andreas Maier^{1,2}, and Vincent Christlein¹

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.

1 Introduction

The automatic analysis of text images has become an essential tool within a wide range of applications in industry, forensics or historical research. Some of the most frequently required tasks for text image analysis include optical character recognition (OCR) or handwritten text recognition (HTR) [8], writer identification and verification [6] as well as structural document segmentation [9]. Hereby, most methods rely on the existence of accurate features extracted from document images, e. g., keypoints [6,9]. Another essential feature widely used for OCR and HTR is *binarization*, i. e., the segmentation of text images into character and background regions. The reliability of such features strongly depends on the quality of the underlying text images. To address this requirement, image enhancement and restoration techniques are commonly used for preprocessing prior to text analysis [8]. In this context, text image restoration by means of *deconvolution* is a technique to recover a sharp image from a blurred acquisition.

This work is partly supported by the German Federal Ministry of Education and Research (BMBF), grant-nr. 01UG1236a

Reasons for blurring can be motion blur, i. e., blur induced by moving the camera or a movement of the scene. Another reason that is relevant for document images acquired under a controlled environment, e. g., digital scanning, are limitations of optics and sensors. We consider the extraction of text features and image deconvolution as complementary problems that can be, however, strongly coupled in order to enhance both of them. In particular, this is the case for deconvolution and binarization. If a sharp text image obtained by image deconvolution is available, this serves as a reliable input for text binarization. Conversely, an accurate text binarization can be utilized as a strong prior for image deconvolution.

Text Image Binarization. Most methods for automatic binarization of text images can be categorized into two groups. The most basic *global* thresholding techniques estimate a single threshold, e. g., using Otsu’s method [18], to discriminate characters and background in two-tone images. This approach is computationally efficient but is sensitive to global illumination changes, which is a common issue in large document images. As a complementary approach, *local* thresholding techniques estimate a threshold per image patch or even per pixel to make binarization spatially adaptive [1,20].

Text Image Restoration. Document image restoration can be approached from a natural scene statistics or a text-specific point of view. Most general-purpose methods exploit natural scene statistics and make use of the fact that natural images are sparse in the gradient domain. This can be modeled by total variation [2,17] or heavy tailed priors [15]. A variety of algorithms has been proposed to solve image deconvolution as a *non-blind* problem [5] or as *blind* estimation [14] in which the blur characteristics are estimated simultaneously with the deblurred image. Despite their success, natural image statistics typically fail to model the characteristics of text images [4] since they provide a too weak prior. For this reason, various blind deconvolution approaches have been proposed which exploit the properties of text images. One class of methods utilizes the two-tone property of document images for blur kernel estimation and deblurring [16,3]. A different strategy has been proposed by Zhang [23] that directly restores a binarization from a blurred two-tone image. However, this method does not provide a deblurred intensity image. The use of more comprehensive text-specific properties in addition to the intensity has been examined in the work of Cho et al. [4]. These properties guide the image deconvolution and include contrast, color-uniformity and gradient statistics of characters and background, respectively. However, the success of this algorithm relies on the stroke width transform (SWT) [7] that is used to describe text-specific properties. For this reason, a complementary approach has been proposed by Pan et al. [19]. In their method, text deblurring is formulated via an L_0 norm regularized energy function using intensity and gradient information. Deblurring yields outstanding results in presence of severe motion blur for both pure text images as well as natural images containing text. However, it yields only a deblurred image with the associated blur kernel estimate without considering binarization.

Proposed Binarization Driven Blind Deconvolution. This work faces document image restoration from a different point of view. Similar to prior

work [3,16,23], we exploit properties of two-tone text images. In doing so, we consider blind deconvolution and binarization as coupled problems and aggregate them in a novel energy minimization framework. The proposed algorithm gradually refines text binarization that is exploited as a prior for blind deconvolution. The advantage of this novel strategy is twofold: 1) deconvolution is guided by binarization as a strong prior compared to priors derived from natural image statistics, and 2) text binarization benefits from deconvolution and is incrementally refined in our optimization. In detail, our contributions are:

- a novel energy minimization formulation for blind deconvolution that exploits text binarization as guidance,
- a binarization method using a soft clustering algorithm as inner optimization loop in the proposed framework,
- demonstration of the impact of our method in a comprehensive evaluation for document restoration on synthetic data as well as real historic text images.

2 Image Deconvolution Model

We examine blind deconvolution of single-channel images linearized to a vector $\mathbf{y} \in \mathbb{R}^n$ with $y_i \in [0; 1]$. Our method is derived from the image formation model $\mathbf{y} = \mathbf{h} * \mathbf{x} + \boldsymbol{\epsilon}$, where $\mathbf{x} \in \mathbb{R}^n$ denotes the unknown deblurred image, $\mathbf{h} \in \mathbb{R}^m$ denotes a linear, space invariant blur kernel in vector notation and $*$ is the discrete convolution operator. The signal $\boldsymbol{\epsilon} \in \mathbb{R}^n$ models additive noise.

In the proposed model, for \mathbf{x} there exists a corresponding binarization \mathbf{s} describing the partitioning of \mathbf{x} into characters and background, respectively. This binarization is encoded by a probability map $\mathbf{s} \in [0; 1]^n$, where s_i is the probability that the i -th pixel belongs to the background. The image \mathbf{x} and its binarization \mathbf{s} can be considered as coupled variables. If one knows an ideal image \mathbf{x} , \mathbf{s} could be determined accurately by means of image binarization. Conversely, if an ideal binarization \mathbf{s} would be known, blind deconvolution could be guided by \mathbf{s} . Hence, we formulate blind deconvolution as the joint energy function:

$$\mathcal{E}(\mathbf{x}, \mathbf{h}, \mathbf{s}) = \mathcal{D}(\mathbf{x}, \mathbf{h}) + \lambda_x \mathcal{R}(\mathbf{x}) + \lambda_h \mathcal{H}(\mathbf{h}) + \lambda_c \mathcal{C}(\mathbf{x}, \mathbf{s}), \quad (1)$$

where $\mathcal{D}(\mathbf{x}, \mathbf{h})$ and $\mathcal{R}(\mathbf{x})$ with weight $\lambda_x \geq 0$ denote the data fidelity and regularization term for image deconvolution, respectively. $\mathcal{H}(\mathbf{h})$ with weight $\lambda_h \geq 0$ denotes a regularizer for the blur kernel \mathbf{h} . $\mathcal{C}(\mathbf{x}, \mathbf{s})$ with weight $\lambda_c \geq 0$ describes a consistency term that couples the image \mathbf{x} with the associated binarization \mathbf{s} . We define the data fidelity term as:

$$\mathcal{D}(\mathbf{x}, \mathbf{h}) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2, \quad (2)$$

where $\mathbf{H} \in \mathbb{R}^{n \times n}$ denotes the blur kernel \mathbf{h} in matrix notation. Mathematically, $\mathcal{D}(\mathbf{x}, \mathbf{h})$ provides a maximum likelihood estimate under additive, zero-mean Gaussian noise. The regularization term for the deblurred image is given by a Hyper-Laplacian prior [14] derived from natural image statistics:

$$\mathcal{R}(\mathbf{x}) = \sum_{i=1}^n ([\nabla_h \mathbf{x}]_i^2 + [\nabla_v \mathbf{x}]_i^2)^{\frac{\beta}{2}}, \quad (3)$$

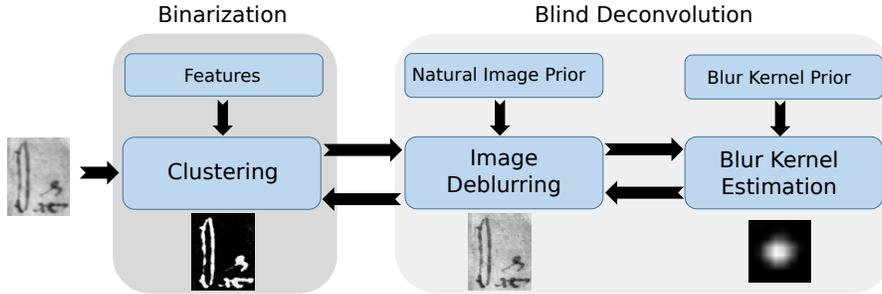


Fig. 1: Flowchart of the proposed binarization driven blind deconvolution method.

where $0 \leq p \leq 1$, and ∇_h and ∇_v denote the gradient of \mathbf{x} in horizontal and vertical direction (computed pixel-wise). This term exploits sparsity of \mathbf{x} in the gradient domain for the regularization of image deconvolution. For the blur kernel estimation, we enforce non-negativity of its elements h_i [14] according to:

$$\mathcal{H}(\mathbf{h}) = \sum_{i=1}^m \mathcal{H}(h_i) \quad \text{where } \mathcal{H}(h) = \begin{cases} h & h \geq 0 \\ \infty & h < 0 \end{cases}. \quad (4)$$

In order to guide the estimation of \mathbf{x} , we propose a new consistency term to couple deconvolution with the associated binarization \mathbf{s} . In doing so, we exploit the fact that discontinuities in \mathbf{x} and \mathbf{s} should be aligned. Although this is not completely true for natural images, it is a reasonable assumption for document images. For instance, a discontinuity in \mathbf{s} associated with the boundary of a character corresponds to a discontinuity of the intensities in \mathbf{x} and the gradients $\nabla \mathbf{s}$ and $\nabla \mathbf{x}$ are equal up to scale. In background regions, one can even assume equal gradients $\nabla \mathbf{s}$ and $\nabla \mathbf{x}$. We enforce this consistency in the gradient domain by the term:

$$\mathcal{C}(\mathbf{x}, \mathbf{s}) = \|\nabla_h \mathbf{x} - \nabla_h \mathbf{s}\|_2^2 + \|\nabla_v \mathbf{x} - \nabla_v \mathbf{s}\|_2^2, \quad (5)$$

where we construct \mathbf{s} such that its gradient $\nabla \mathbf{s}$ has a consistent direction with the gradient $\nabla \mathbf{x}$. For this purpose, a character at the i -th pixel in \mathbf{x} associated with a low intensity x_i corresponds to a low probability s_i in the binarization \mathbf{s} .

3 Binarization Driven Deconvolution Algorithm

The proposed method is based on the minimization of the energy function in Equation (1) and requires knowledge of the binarization \mathbf{s} to exploit the consistency $\mathcal{C}(\mathbf{x}, \mathbf{s})$. However, in practice \mathbf{s} is unknown and it would be error-prone to obtain it from the blurred image directly using standard image binarization techniques. For this reason, one strategy would be a formulation as joint energy minimization w. r. t. the deblurred image, the latent blur kernel and the binarization. Unfortunately, this approach is only computationally tractable with simplified models of image binarization. In the proposed method, we solve Equation (1) w. r. t. \mathbf{x}

and \mathbf{h} while the binarization \mathbf{s} is gradually refined over $t \geq 1$ iterations. The outline of our approach as depicted in Figure 1 is as follows: First, we obtain the binarization $\mathbf{s}^{(t)}$ using soft clustering in an inner optimization as proposed in Section 3.1. Then, we minimize Equation (1) w. r. t. \mathbf{x} and \mathbf{h} by exploiting the binarization $\mathbf{s}^{(t)}$ as shown in Section 3.2 and Section 3.3 according to:

$$(\mathbf{x}^{(t)}, \mathbf{h}^{(t)}) = \arg \min_{\mathbf{x}, \mathbf{h}} \mathcal{E}(\mathbf{x}, \mathbf{h}, \mathbf{s}^{(t)}). \quad (6)$$

These stages are solved alternately, where $(\mathbf{x}^{(t-1)}, \mathbf{h}^{(t-1)}, \mathbf{s}^{(t-1)})$ is propagated from iteration $t - 1$ to obtain refined estimates at iteration t . For an efficient implementation that avoids local minimums, iterations are performed in a coarse-to-fine scheme [19,14]. Starting at the coarsest level that is obtained by downsampling the input image and the support of the blur kernel, we gradually estimate the deblurred image and the blur kernel over different scales without using the consistency term in a first pass. Then, the blur kernel and the deblurred image with its binarization are refined on the finest scale using the full model with our consistency term in a second pass. Finally, we propose a further refinement of the deblurred image $\tilde{\mathbf{x}}$ by means of guided filtering [12] that exploits the binarization $\tilde{\mathbf{s}}$ to remove remaining deblurring artifacts, e. g. ringing [19]. We compute this refinement as the average image $\mathbf{x} = \frac{1}{2}(\text{GF}(\tilde{\mathbf{x}}, \tilde{\mathbf{s}}) + \text{GF}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}))$, where $\text{GF}(\mathbf{p}, \mathbf{q})$ denotes the guided filter with input image \mathbf{p} and guidance image \mathbf{q} .

3.1 Text-Specific Binarization

We formulate text binarization as a soft clustering problem using $c = 2$ clusters corresponding to characters and background. Each cluster is described in a feature space by the center $\boldsymbol{\mu}_j \in \mathbb{R}^d$ with $j \in \{1, 2\}$. To take the property of text into account that adjacent pixels should be assigned to similar clusters, we incorporate spatially regularized Fuzzy C-means clustering [22] to the proposed text binarization. Based on the features $\mathbf{f}_\mathbf{x}$ that are pixel-wise extracted as $\mathbf{f}_{\mathbf{x},i} \in \mathbb{R}^d$ at the i -th pixel in the image \mathbf{x} , the binarization is obtained by:

$$(\boldsymbol{\mu}^{(t)}, \mathbf{s}^{(t)}) = \arg \min_{\boldsymbol{\mu}, \mathbf{s}} \sum_{i=1}^n \sum_{j=1}^c s_{ij}^q \|\mathbf{f}_{\mathbf{x},i} - \boldsymbol{\mu}_j\|_2^2 + \alpha \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^n w_{ik} s_{ij}^q (1 - s_{kj})^q, \quad (7)$$

where $q > 1$ is a weighting parameter for the fuzzy cluster membership $s_{ij} \in [0; 1]$ of the i -th feature vector to the j -th cluster. The weights w_{ik} are set to $w_{ik} = 1$ if the i -th and the k -th pixel are adjacent in an 8-neighborhood and $w_{ik} = 0$, otherwise. The data fidelity term of soft clustering and the regularization term for adjacent pixels in Equation (7) are weighted to each other by $\alpha \geq 0$.

To define the feature set $\mathbf{f}_\mathbf{x}$, we perform a scale space analysis over $d \geq 1$ scales on the image $\mathbf{x}^{(t-1)}$. The features associated with the i -th pixel are assembled as $\mathbf{f}_{\mathbf{x},i} = (Q_i(\mathbf{x}^{(t-1)}, \omega_1) \ Q_i(\mathbf{x}^{(t-1)}, \omega_2) \ \dots \ Q_i(\mathbf{x}^{(t-1)}, \omega_d))^\top \in \mathbb{R}^d$, where $Q_i(\mathbf{x}^{(t-1)}, \omega_j)$ is the i -th pixel of a filtered version of $\mathbf{x}^{(t-1)}$ using the filter size ω_j . In this work, we implement the discrete filter Q_i via 2-D median filtering.

This provides edge preserving filtering and enables character analysis over various scales in soft clustering. Once the features \mathbf{f}_x are extracted, the cluster centers and the binarization map \mathbf{s} are computed as the zero-crossings of Equation (7). Soft clustering is obtained by alternating computation of:

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^n s_{ij}^q \mathbf{f}_{x,i}}{\sum_{i=1}^n s_{ij}^q} \quad (8)$$

$$s_{ij} = \left(\frac{\sum_{k=1}^c \frac{\|\mathbf{f}_{x,i} - \boldsymbol{\mu}_j\|_2^2 + \alpha \sum_{l=1}^n (1 - s_{li})^q w_{jk}}{\|\mathbf{f}_{x,i} - \boldsymbol{\mu}_k\|_2^2 + \alpha \sum_{l=1}^n (1 - s_{lk})^q w_{jk}}}{\sum_{k=1}^c \frac{\|\mathbf{f}_{x,i} - \boldsymbol{\mu}_k\|_2^2 + \alpha \sum_{l=1}^n (1 - s_{lk})^q w_{jk}}{\|\mathbf{f}_{x,i} - \boldsymbol{\mu}_k\|_2^2 + \alpha \sum_{l=1}^n (1 - s_{lk})^q w_{jk}}} \right)^{-\frac{1}{q-1}}, \quad (9)$$

until convergence of the clustering procedure using $\mathbf{s}^{(t-1)}$ obtained at the previous iteration as initialization. Finally, the refined binarization $\mathbf{s}^{(t)}$ is assembled from the cluster membership degrees s_{ij} associated with the background.

3.2 Estimation of the Deblurred Image

In order to estimate the deblurred image $\mathbf{x}^{(t)}$ at the current iteration, we solve Equation (6) w. r. t. \mathbf{x} while keeping the blur kernel fixed as $\mathbf{h}^{(t-1)}$. Using the consistency term $\mathcal{C}(\mathbf{x}, \mathbf{s}^{(t)})$ that couples the image \mathbf{x} with the binarization map $\mathbf{s}^{(t)}$, the deblurred image is obtained by the energy minimization problem:

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

This unconstrained problem is solved by the alternating direction method of multipliers (ADMM) using Split Bregman iterations [11]. For this purpose, we derive a constrained problem that is equivalent to Equation (10):

$$\begin{aligned} \arg \min_{\mathbf{x}, \mathbf{v}_h, \mathbf{v}_v} \left\{ \|\mathbf{H}^{(t-1)} \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_c (\|\mathbf{v}_h - \nabla_h \mathbf{s}^{(t)}\|_2^2 + \|\mathbf{v}_v - \nabla_v \mathbf{s}^{(t)}\|_2^2) \right. \\ \left. + \lambda_x \sum_{i=1}^n ([\mathbf{v}_h]_i^2 + [\mathbf{v}_v]_i^2)^{\frac{p}{2}} \right\} \quad \text{s. t. } \mathbf{v}_h = \nabla_h \mathbf{x}, \mathbf{v}_v = \nabla_v \mathbf{x}, \end{aligned} \quad (11)$$

where \mathbf{v}_h and \mathbf{v}_v are auxiliary variables and $\mathbf{H}^{(t-1)}$ is constructed from the kernel estimate $\mathbf{h}^{(t-1)}$. This provides a decoupling of the data fidelity term $\mathcal{D}(\mathbf{x}, \mathbf{h})$ from the regularizer $\mathcal{R}(\mathbf{x})$ and our consistency term $\mathcal{C}(\mathbf{x}, \mathbf{s})$. For numerical optimization, Equation (11) is re-formulated to the unconstrained minimization problem:

$$\begin{aligned} \arg \min_{\mathbf{x}, \mathbf{v}_h, \mathbf{v}_v} \left\{ \|\mathbf{H}^{(t-1)} \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_v (\|\mathbf{v}_h - \nabla_h \mathbf{x} - \mathbf{b}_h\|_2^2 + \|\mathbf{v}_v - \nabla_v \mathbf{x} - \mathbf{b}_v\|_2^2) \right. \\ \left. + \lambda_x \sum_{i=1}^n ([\mathbf{v}_h]_i^2 + [\mathbf{v}_v]_i^2)^{\frac{p}{2}} + \lambda_c (\|\mathbf{v}_h - \nabla_h \mathbf{s}^{(t)}\|_2^2 + \|\mathbf{v}_v - \nabla_v \mathbf{s}^{(t)}\|_2^2) \right\}, \end{aligned} \quad (12)$$

where λ_v is a Lagrangian multiplier that weights the quadratic penalty terms derived from the constraints in Equation (11), and \mathbf{b}_h and \mathbf{b}_v denote the Bregman variables. We solve this minimization problem by coordinate descent for \mathbf{x} , \mathbf{v}_h and \mathbf{v}_v , where \mathbf{b}_h and \mathbf{b}_v are chosen per Bregman iteration. Similar to [14], the optimization of Equation (12) is performed in the Fourier domain.

3.3 Estimation of the Blur Kernel

The estimation of the blur kernel \mathbf{h} can be done in a similar way. For blur estimation, we minimize Equation (6) w. r. t. \mathbf{h} and keep the deblurred image given by $\mathbf{x}^{(t)}$ fixed. Hence, we can omit the consistency term and optimize only the deconvolution term. Similar to prior work [14,19], the kernel is estimated in the gradient domain resulting in the energy minimization problem:

$$\mathbf{h}^{(t)} = \arg \min_{\mathbf{h}} \left\{ \mathcal{D}(\nabla \mathbf{x}^{(t)}, \mathbf{h}) + \lambda_h \mathcal{H}(\mathbf{h}) \right\}. \quad (13)$$

Then, we use ADMM and introduce the auxiliary variable \mathbf{g} to substitute the blur kernel \mathbf{h} in the regularizer $\mathcal{H}(\mathbf{h})$, the associated Bregman variable \mathbf{b}_g , and the Lagrange multiplier λ_g . This yields the unconstrained problem:

$$\arg \min_{\mathbf{h}, \mathbf{g}} \left\{ \|\nabla \mathbf{X}^{(t)} \mathbf{h} - \nabla \mathbf{y}\|_2^2 + \lambda_h \mathcal{H}(\mathbf{g}) + \lambda_g \|\mathbf{h} - \mathbf{g} - \mathbf{b}_g\|_2^2 \right\}, \quad (14)$$

where we reformulated $\mathbf{H}\mathbf{x}^{(t)}$ as $\mathbf{X}^{(t)}\mathbf{h}$. Minimization is performed by coordinate descent for \mathbf{h} and \mathbf{g} in the Fourier domain with a re-centering of the kernel after each iteration [14]. The non-negativity constraint according to Equation (4) is enforced at each iteration using thresholding of the kernel elements h_i .

4 Experiments and Results

We evaluated our algorithm on artificial and real document images. For a quantitative evaluation, we used an artificial dataset consisting of 18 images of sizes between 120×120 and 240×240 pixel that were simulated by blurring ground truth images with a 15×15 Gaussian kernel of standard deviation $\sigma_b = 2.5$, see Figure 2. Moreover, images were disturbed by adding zero-mean Gaussian noise with varying standard deviation σ_n . To demonstrate the performance of our approach in terms of binarization, we used 19 excerpts of real historical documents with manually generated ground truth text binarizations. Our method was compared with the general-purpose deconvolution approach of Kotera et al. [14] and the method of Pan et al. [19] that has been recently proposed for text images. The F1 measure was used to assess the reliability of text binarization. For artificial data with known ground truth grayscale images, we also measured the peak-signal-to-noise ratio (PSNR) as well as structural similarity (SSIM).

Deconvolution Results. For a quantitative evaluation, the Gaussian noise standard deviation to simulate artificial text images was first set to $\sigma_n = 0.01$. Then, we adjusted the regularization weights of all compared deconvolution methods on one single training image taken from our simulated data by optimizing the PSNR of the deblurred image. In addition to pure blind deconvolution, we applied blind deconvolution and binarization as a two-stage approach using the methods of Kotera et al. [14] and Pan et al. [19] followed by hard thresholding using Otsu’s method [18]. To assess binarization achieved by our method, we employ the

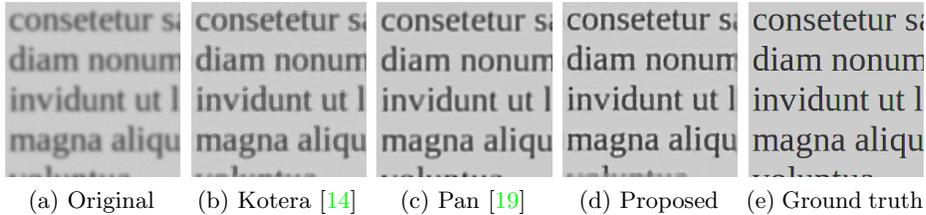
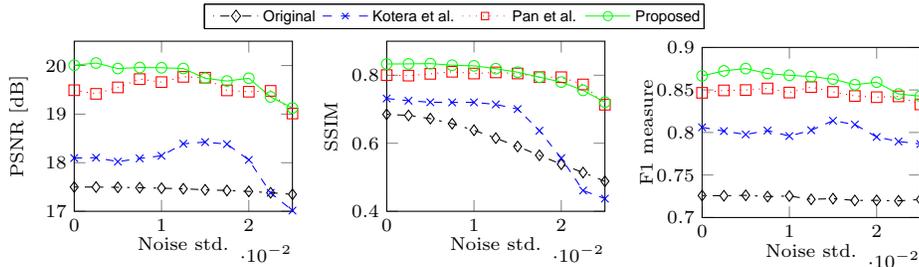


Fig. 2: Blind deconvolution results on simulated data with known ground truth.

Fig. 3: Influence of Gaussian noise with standard deviation σ_n to blind deconvolution and the binarization obtained from the deblurred images. We evaluated the median of all quality measures over 18 simulated images with ground truth data.

binarization provided in the final pass of our coarse-to-fine optimization. For this setup, Table 1 compares the blind deconvolution methods on the simulated dataset using the PSNR and SSIM measures on grayscale images and the F1 measure to assess the binarizations. In terms of all measurements, our method consistently outperformed the methods of Kotera et al. [14] and Pan et al. [19].

Another experiment was conducted to analyze the noise robustness of the different algorithms. Therefore, we varied the noise standard deviation of the simulated images and compared the different deconvolution methods. Figure 3 shows the median of different quality measures over 18 simulated images. Binarization driven blind deconvolution consistently outperformed the approach of Kotera et al. [14] and achieved higher robustness with respect to image noise. For small noise levels, it also achieved higher quality measures compared to blind deconvolution of Pan et al. [19] with competitive results in case of severe noise.

For qualitative results on real data, we tested our method with excerpts of document images, cf. Figure 4. As samples we used scans of historical handwritten documents. The images are of low quality and are affected by image noise. Compared to the results of Kotera et al. [14], binarization driven blind deconvolution was able to reconstruct sharper boundaries of characters with accurate denoising in the background.

Binarization Results. Although the main target of our method is image blind deconvolution, we show that it provides also binarization results comparable to

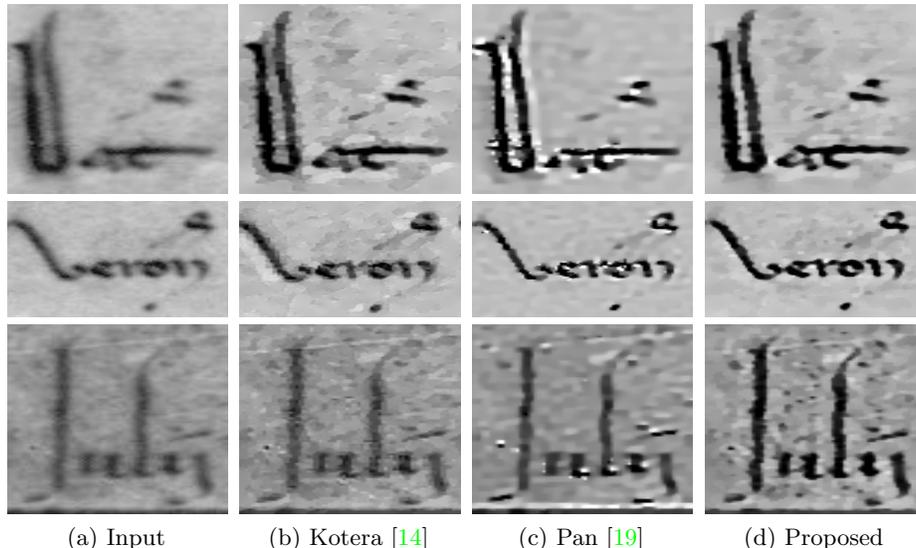


Fig. 4: Results of the different blind deconvolution methods applied on scanned handwritten documents (source: Göttingen Academy of Sciences & Humanities).

Table 1: Mean and standard deviation of all quality measures for blind deconvolution on simulated images (noise level $\sigma_n = 0.01$). Binarizations for [14] and [19] were obtained by applying Otsu’s method [18] on the deblurred images.

	PSNR (in dB)	SSIM	F1 measure
Original	17.57 ± 0.42	0.64 ± 0.02	0.72 ± 0.01
Kotera et al. [14]	18.30 ± 0.54	0.73 ± 0.03	0.79 ± 0.02
Pan et al. [19]	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

the state of the art. To investigate the performance of the binarization routine of our framework, we used the same artificial dataset as in the previous experiments as well as the handwritten dataset with manually generated ground truth binarization. Table 2 shows the results of our binarization in contrast to other local [1,20,21] and global binarization methods [18] applied to the original images. Moreover, we evaluated the deconvolution methods of Kotera et al. [14] and Pan et al. [19] by applying the method of Bradley and Roth [1] on the deblurred images. Binarization driven deconvolution outperformed all other binarization methods on the artificial dataset. On the historical dataset our method performed worse than local thresholding, but substantially outperformed global thresholding. The proposed method also achieved higher results than the other deconvolution methods. Overall the method of Bradley and Roth [1] achieved slightly better results. A visual comparison between all methods is shown in Figure 5.

Table 2: Mean and standard deviation of F1 measure on binarizations obtained from the artificial and handwritten documents as well as the merged dataset. The best and second best measures per dataset are highlighted.

	Artificial	Handwritten	Merged
Otsu [18]	0.72 ± 0.01	0.75 ± 0.15	0.73 ± 0.11
Sauvola and Pietikäinen [20]	0.79 ± 0.04	0.78 ± 0.07	0.79 ± 0.06
Bradley and Roth [1]	0.85 ± 0.01	0.78 ± 0.08	0.82 ± 0.07
Su [21]	0.80 ± 0.01	0.79 ± 0.09	0.80 ± 0.07
Kotera [14] + Bradley and Roth [1]	0.82 ± 0.01	0.71 ± 0.10	0.76 ± 0.09
Pan [19] + Bradley and Roth [1]	0.85 ± 0.02	0.73 ± 0.09	0.79 ± 0.09
Proposed	0.87 ± 0.02	0.76 ± 0.11	0.81 ± 0.10

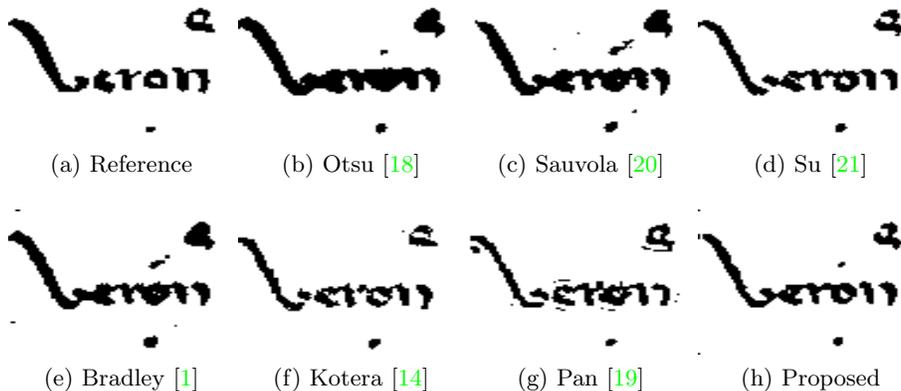


Fig. 5: Binarization results on a scanned handwritten document image.

5 Conclusions

In this work, we presented a new method for image blind deconvolution. Our algorithm explicitly incorporates text binarization as guidance. For this purpose, a novel consistency term serves as regularizer that couples deconvolution and binarization. Compared to existing blind deconvolution algorithms, our method provides more accurate deblurred intensity images as demonstrated for simulated and handwritten documents. In addition, our method provides a binarization as a by-product that is comparable to state-of-the-art text binarization techniques.

In our future work, we would like to incorporate our method as preprocessing step for successive processes like HTR or paleographic analysis. From an algorithmic point of view, our soft clustering procedure could benefit from the use of text-specific features like the stroke width transform [7]. By incorporating text-specific segmentation methods, we could extend our method to text images in the wild, i. e., non document texts. Finally, other domains like super-resolution [10,13] could benefit from our concept of a binarization driven deconvolution.

References

1. Bradley, D., Roth, G.: Adaptive Thresholding Using the Integral Image. *Journal of Graphics, GPU, and Game Tools* 12(2), 13–21 (2007) [2](#), [9](#), [10](#)
2. Chan, T.F., Wong, C.K.: Total Variation Blind Deconvolution. *IEEE Transactions on Image Processing* 7(3), 370–375 (Jan 1998) [2](#)
3. Chen, X., He, X., Yang, J., Wu, Q.: An Effective Document Image Deblurring Algorithm. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 369–376. IEEE (Jun 2011) [2](#), [3](#)
4. Cho, H., Wang, J., Lee, S.: Text Image Deblurring Using Text-Specific Properties. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *Computer Vision ECCV 2012, Lecture Notes in Computer Science*, vol. 7576, pp. 524–537. Springer Berlin Heidelberg (2012) [2](#)
5. Cho, S., Wang, J., Lee, S.: Handling Outliers in Non-Blind Image Deconvolution. In: *IEEE International Conference on Computer Vision (ICCV)*. pp. 495–502 (2011) [2](#)
6. Christlein, V., Bernecker, D., Hönig, F., Angelopoulou, E.: Writer Identification and Verification Using GMM Supervectors. In: *IEEE Winter Conference on Applications of Computer Vision*. pp. 998–1005 (2014) [1](#)
7. Epshtein, B., Ofek, E., Wexler, Y.: Detecting Text in Natural Scenes with Stroke Width Transform. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2963–2970. IEEE (Jun 2010) [2](#), [11](#)
8. Espana-Boquera, S., Castro-Bleda, M., Gorbe-Moya, J., Zamora-Martinez, F.: Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(4), 767–779 (April 2011) [1](#)
9. Garz, A., Diem, M., Sablatnig, R.: Detecting Text Areas and Decorative Elements in Ancient Manuscripts. In: 2010 International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 176–181 (Nov 2010) [1](#)
10. Ghesu, F.C., Köhler, T., Haase, S., Hornegger, J.: Guided Image Super-Resolution: A New Technique for Photogeometric Super-Resolution in Hybrid 3-D Range Imaging. In: Jiang, X., Hornegger, J., Koch, R. (eds.) *Pattern Recognition*. pp. 227–238 (2014) [11](#)
11. Goldstein, T., Osher, S.: The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences* 2(2), 323–343 (Jan 2009) [6](#)
12. He, K., Sun, J., Tang, X.: Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(6), 1397–409 (Jun 2013) [5](#)
13. Köhler, T., Jordan, J., Maier, A., Hornegger, J.: A Unified Bayesian Approach to Multi-Frame Super-Resolution and Single-Image Upsampling in Multi-Sensor Imaging. In: *Proceedings of the British Machine Vision Conference*. pp. 143.1–143.12. BMVA Press (2015) [11](#)
14. Kotera, J., Šroubek, F., Milanfar, P.: Blind Deconvolution Using Alternating Maximum a Posteriori Estimation with Heavy-Tailed Priors. In: Wilson, R., Hancock, E., Bors, A., Smith, W. (eds.) *Computer Analysis of Images and Patterns, Lecture Notes in Computer Science*, vol. 8048, pp. 59–66. Springer Berlin Heidelberg (2013) [2](#), [3](#), [4](#), [5](#), [7](#), [8](#), [9](#), [10](#)
15. Levin, A., Weiss, Y., Durand, F., Freeman, W.: Understanding and Evaluating Blind Deconvolution Algorithms. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1964–1971. IEEE (Jun 2009) [2](#)

16. Li, T.H., Lii, K.S.: A Joint Estimation Approach for Two-Tone Image Deblurring by Blind Deconvolution. *IEEE Transactions on Image Processing* 11(8), 847–58 (Jan 2002) [2](#), [3](#)
17. Oliveira, J.a.P., Bioucas-Dias, J.M., Figueiredo, M.A.: Adaptive Total Variation Image Deblurring: A Majorization-Minimization Approach. *Signal Processing* 89(9), 1683–1693 (Sep 2009) [2](#)
18. Otsu, N.: A Threshold Selection Method From Gray-Level Histograms. *IEEE Transactions on Systems, Man, And Cybernetics* 9(1), 62–66 (Jan 1979) [2](#), [8](#), [9](#), [10](#)
19. Pan, J., Hu, Z., Su, Z., Yang, M.H.: Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2901–2908 (June 2014) [2](#), [5](#), [7](#), [8](#), [9](#), [10](#)
20. Sauvola, J., Pietikäinen, M.: Adaptive Document Image Binarization. *Pattern Recognition* 33(2), 225–236 (2000) [2](#), [9](#), [10](#)
21. Su, B., Lu, S., Tan, C.L.: Binarization of Historical Document Images Using the Local Maximum and Minimum. In: 9th IAPR International Workshop on Document Analysis Systems. pp. 159–165. Boston, MA (Jun 2010) [9](#), [10](#)
22. Yang, Y., Huang, S.: Image Segmentation by Fuzzy C-Means Clustering Algorithm with a Novel Penalty Term. *Computing and Informatics* 26(1), 17–31 (2012) [5](#)
23. Zhang, J.: An Alternating Minimization Algorithm for Binary Image Restoration. *IEEE Transactions on Image Processing* 21(2), 883–888 (Feb 2012) [2](#), [3](#)