

Learning Document Image Binarization from Data

Yue Wu, Stephen Rawls, Wael AbdAlmageed and Premkumar Natarajan

Abstract

In this paper we present a fully trainable binarization solution for degraded document images. Unlike previous attempts that often used simple features with a series of pre- and post-processing, our solution encodes all heuristics about whether or not a pixel is foreground text into a high-dimensional feature vector and learns a more complicated decision function. In particular, we prepare features of three types: 1) existing features for binarization such as *intensity* [1], *contrast* [2], [3], and *Laplacian* [4], [5]; 2) reformulated features from existing binarization decision functions such those in [6] and [7]; and 3) our newly developed features, namely the Logarithm Intensity Percentile (LIP) and the Relative Darkness Index (RDI). Our initial experimental results show that using only selected samples (about 1.5% of all available training data), we can achieve a binarization performance comparable to those fine-tuned (typically by hand), state-of-the-art methods. Additionally, the trained document binarization classifier shows good generalization capabilities on out-of-domain data.

I. INTRODUCTION

As one of the most fundamental preprocessing methods in various document analysis work [8], [9], [10], [11], [12], [13], [3], [14], [5], [15], document binarization aims to convert a color or grayscale document image into a monotonic image, where all text pixels of interest are marked in black with a white background. Mathematically, given a document image $\mathbf{D} = \{D_{i,j}\}_{i \in [1,W], j \in [1,H]}$ of size $W \times H$, image binarization assigns each pixel $D_{i,j}$ a binary class label $B_{i,j}$ according to a decision function $f_{\text{binarize}}(\cdot)$ in a meaningful way, namely

$$B_{i,j} = \begin{cases} \text{foreground class 1,} & \text{if } f_{\text{binarize}}(D_{i,j}) < 0 \\ \text{background class 0,} & \text{else} \end{cases} \quad (1)$$

Authors are all associated with Information Sciences Institute, University of South California, Marina Del Ray, California 90230. Email: {yue_wu, srawls, wamageed, pnataraj}@isi.edu.

A successful document binarization process discards irrelevant and noisy information while preserving meaningful information in the binary image $\mathbf{B}=\{B_{i,j}\}$. This process reduces the space to represent a document image, and largely simplifies the complexity of advanced document analysis tasks [4].

Although human do not often face many difficulties in identifying texts even on some low-quality document images, the document image binarization problem is indeed subjective and ill-posed [4], and it involves many different challenges and combinations of challenges. For example, several of the well-known ones are 1) how to handle document degradations like ink blob, fade text etc.; 2) how to deal with uneven lighting; and 3) how to differentiate bleed-through text from normal text. In such difficult scenarios, human actually uses high-level knowledge that might not be easily captured by low-level features—such as a script character set and background texture analysis—to help decide which pixel is foreground text.

Classic solutions more or less seek heuristic thresholds in simple feature spaces. This can be further grouped into the so-called *global thresholding* and *local thresholding* methods [14] according to whether this threshold is location independent or not. For example, Otsu's method [1] binarizes a pixel $D_{i,j}$ by comparing its *pixel intensity* $I_{i,j}$ to an optimal global threshold G_{th} derived from *intensity histogram* [1] as shown in (2)

$$f_{\text{Otsu}}(i, j) = I_{i,j} - G_{\text{th}} \quad (2)$$

In contrast Niblack's method [6] uses the decision function (3)

$$f_{\text{Niblack}}(i, j) = I_{i,j} - \mu_{i,j}^R + k_{\text{Niblack}}\sigma_{i,j}^R \quad (3)$$

where k_{Niblack} is a parameter below 0, and $\mu_{i,j}^R$ and $\sigma_{i,j}^R$ denote the mean and standard deviation of pixel intensities within a region R of size $w \times h$. Although heuristic solutions are very efficient—may only requiring a constant number of operations per pixel, and work fairly well on many well-conditioned document images, it is clear that simple features and decision functions are insufficient for handling difficult cases.

To achieve robust document binarization, many efforts are being made in the areas of 1) image normalization/adaptation, 2) discriminative feature space, and 3) more complicated decision functions. For example, Lu *et al.*[16] proposes a local thresholding approach that mainly relies on background estimation and stroke estimation. Su *et al.*[2], [3] finds that Otsu's thresholding helps attain more discriminative power in a local contrast feature space. Sauvola *et al.*[17], [7] adds the parameter S_{Sauvola} to allow a non-linear decision plane (4).

$$f_{\text{Sauvola}}(i, j) = I_{i,j} - \mu_{i,j}^R - k_{\text{Sauvola}}\mu_{i,j}^R(\sigma_{i,j}^R / S_{\text{Sauvola}} - 1) \quad (4)$$

Although many of these attempts work well when method assumptions are satisfied and method parameters are appropriate, adapting a heuristic binarization method to a new domain is often not easy. Indeed, Lazzara et al. [18] show that the original Sauovla method might fail even for well-scanned document images because of text fonts of different sizes.

Unsupervised learning recently dominates document binarization area. In [19], a document image is first clustered into three classes, namely foreground, background and uncertain, and pixels in the uncertain class will be further classified into either the foreground or background class according to their distances from these two classes. In [4], [5], an image is first transformed into a Laplacian feature space, and a global energy function is constructed to ensure that resulting binary labels are optimal in the sense of a predefined Markov random field. In [20], an unsupervised ensemble of expert frameworks is used to combine multiple binarization candidates. Although these methods do not require a training stage, some rely on theoretical models or heuristic rules whose assumptions may not be necessarily satisfied, some require expensive iterative tuning and optimizations, and thus no surprise to see they are not reliable for certain types of degradations [21].

Although image binarization is clearly a classification problem, supervised learning-based binarization solutions are still rare in the community. In this letter we discuss our initial attempts to solve the the document image binarization problem using supervised learning. The remainder of our paper is organized as follows: Section II overviews our solution and discusses all used features. Section III provides implementation details related to training and testing. Section IV shows our experimental results, and Section V concludes this paper.

II. FEATURE ENGINEERING

Our goal is to develop a generic solution without preset parameters and pre- or post-processing. Specifically, we are interested in learning a decision function $f_{\text{ours}}(\cdot)$ that maps a nd feature vector $\vec{X}_{i,j}$ extracted around a pixel $D_{i,j}$ to a binary space $\{0, 1\}$ in a meaningful way, i.e.

$$B_{i,j} = f_{\text{ours}}(\vec{X}_{i,j}) \quad (5)$$

Detailed feature engineering discussions are given below.

1) Existing Features: Since a number of simple tasks can be accomplished just by applying Otsu's method. We thus include a pixel intensity $I_{i,j}$ and its deviation from the Otsu's threshold as features below

$$X_{i,j}^{\text{Localint.}} = I_{i,j} \quad (6)$$

$$X_{i,j}^{\text{Otsu diff.}} = I_{i,j} - G_{\text{th}} \quad (7)$$

In addition, we also use local statistics of Eqs. (8) and (9), but with respect to different scales, i.e.,

$$X_{i,j}^{\text{Local avg.}|R} = \mu_{i,j}^R = \sum_{p=-w/2}^{w/2} \sum_{q=-h/2}^{h/2} I_{i+p,j+q} / wh \quad (8)$$

$$X_{i,j}^{\text{Local std.}|R} = \sigma_{i,j}^R = \sqrt{\sum_{p=-w/2}^{w/2} \sum_{q=-h/2}^{h/2} I_{i+p,j+q}^2 / wh - \mu_{i,j}^R} \quad (9)$$

where we make the size $w=h=ks$ of local window R be associated with scales $k \in [1, 2, 4, 8]$, and estimate stroke width s using Su's method [3]. Inspired by the success of the Su [2], [3] and Howe methods [4], [5], we include their contrast and Laplacian features shown in (10) and (11).

$$X_{i,j}^{\text{Su}|R} = \frac{\arg \max_{p,q \in R} \{I_{i+p,j+q}\} - \arg \min_{p,q \in R} \{I_{i+p,j+q}\}}{\arg \max_{p,q \in R} \{I_{i+p,j+q}\} + \arg \min_{p,q \in R} \{I_{i+p,j+q}\} + \epsilon_{\text{su}}} \quad (10)$$

$$X_{i,j}^{\text{Howe}|R} = \nabla^2 \mu_{i,j}^R \quad (11)$$

2) *Exponential Truncated Niblack Index*: To include Niblack's decision function in our considerations, we first rearrange terms in (3) according to $f_{\text{Niblack}}=0$, as shown below

$$k_{\text{Niblack}}(i, j|R) = (I_{i,j} - \mu_{i,j}^R) / \sigma_{i,j}^R \quad (12)$$

and then compute a so-called Exponential Truncated Niblack Index (ETNI) feature as follows.

$$X_{i,j}^{\text{ETNI}|R} = \begin{cases} \exp\{k_{\text{Niblack}}(i, j|R)\}, & \text{if } I_{i,j} \leq \mu_{i,j}^R \\ 1, & \text{otherwise} \end{cases} \quad (13)$$

Fig. 1 compares an image in the original form and its corresponding ETRI feature space.

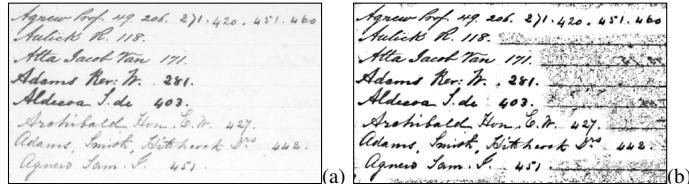


Fig. 1: ETRI features for image DIBCO2010_HW04. (a) Original image; (b) ETNI feature for R of size 64×64 .

3) *Logistic Truncated Sauvola Index*: Similarly, we rearrange terms in Sauvola's decision function (4) according to $f_{\text{Sauvola}} = 0$ for its key parameter k_{Sauvola} as follows,

$$k_{\text{Sauvola}}(i, j|R) \propto \frac{I_{i,j}/\mu_{i,j}^R - 1}{\sigma_{i,j}^R - S_{\text{Sauvola}}} \quad (14)$$

Since $k_{\text{Sauvola}}(i, j|R)$ could be $(-\infty, \infty)$, we normalize this index by using the logistic function shown in Eq. (15), and call it the Logistic Truncated Sauvola Index (LTSI),

$$X_{i,j}^{\text{LTSI}|R} = \begin{cases} 0 & , \text{if } \sigma_{i,j}^R > S_{\text{Sauvola}} \\ (1 + \exp\{-k_{\text{Sauvola}}(i, j|R)\})^{-1}, & \text{otherwise} \end{cases} \quad (15)$$

where the range of $X_{i,j}^{\text{TSI}|R}$ is $[0, 1]$, and the condition $\sigma_{i,j}^R < S_{\text{Sauvola}}$ ensures the sign consistency of $k_{\text{Sauvola}}(i, j|R)$. LTSI thus reflects the Sauvola decision surface. A sample result of the LTSI feature is given in Fig. 2.

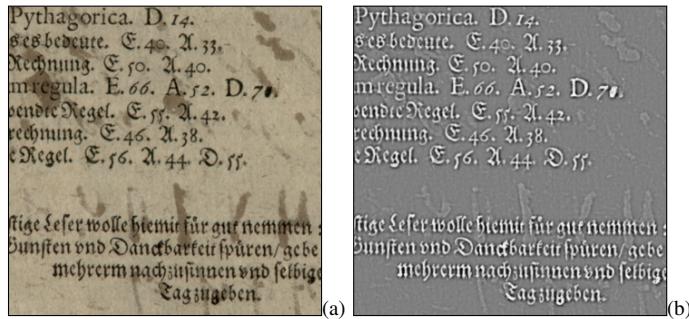


Fig. 2: LTSI features for image DIBCO2011_PR05. (a) Original image; (b) LTSI feature for R of size 8×8 .

4) Logarithm Intensity Percentile Features: Intuitively, the darkness of a pixel is related to whether it is a text pixel. Given a region S , the percentile of the pixel's intensity can be computed as

$$\text{perc}(i, j|S) = \sum_{p,q \in S} \frac{\mathbf{1}_{[0,\infty)}(I_{i,j} - I_{i+p,j+q})}{\|S\|} \quad (16)$$

where $\mathbf{1}_{[0,\infty)}(t)$ denotes the indicator function whose value is 1 when $t \in [0, \infty)$ and 0 otherwise, and $\|\cdot\|$ denotes the cardinality function. It is clear that this percentile is a type of rank feature, and thus is invariant to any monotonic transform on the original intensity space. To give a higher resolution for lower percentiles, we use the logarithm version of (16) as shown in (17), and call it Logarithm Intensity Percentile (LIP) feature. Here Th_{perc} is a threshold ($=.01$ in this paper).

$$X_{i,j}^{\text{LIP}|S} = \begin{cases} 1.0, & \text{if } \text{perc}(i, j|S) \leq \text{Th}_{\text{perc}} \\ \log_{\text{Th}_{\text{perc}}}(\text{perc}(i, j|S)), & \text{otherwise} \end{cases} \quad (17)$$

With regard to S , we make parallelogram S cover multiple rows, columns, diagonals, and inverse diagonals. The number of rows, columns, diagonals and inverse diagonals in S is made to be k times the estimated stroke width s . Finally, we also compute the LIP features with respect to the entire image and the maximum percentile among all previously extracted LIP features. Fig. 3 shows the original document with its corresponding features in the LIP spaces. As one can see, the LIP space indeed provides more discriminative powers.

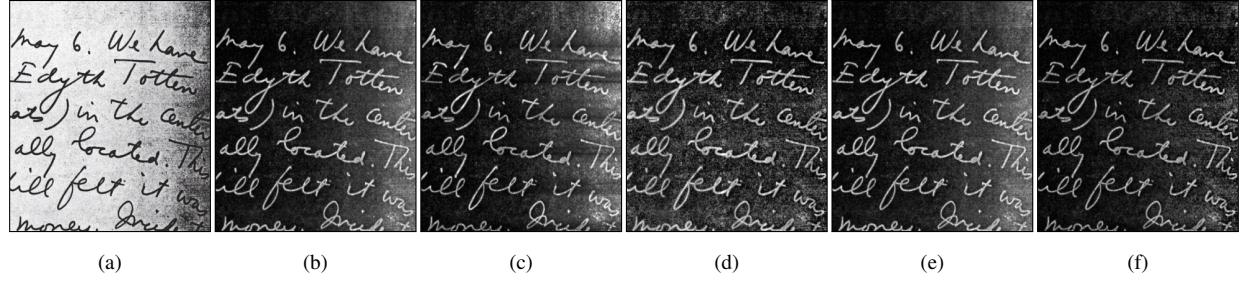


Fig. 3: LIP features for image DIBCO2011_HW1. (a) original image; (b) global LIP; (c)-(e) LIP along row, column, and diagonal; and (f) max LIP of all directions.

5) *Relative Darkness Index Features*: Inspired by the great success of local ternary patterns(LTP) [22] in face recognition, we borrow their essences here. LTP relies on the comparison of a center pixel's intensity with each pixel in a set of neighbors $\{N_1, \dots, N_k\}$ that are on a radius r circle, and the l th code in a length- k code string is defined as

$$\text{ltp}(P_{i,j}, l) = \begin{cases} +1, & \text{if } I_{i+r_l, j+c_l} \geq I_{i,j} + \text{tol} \\ -1, & \text{if } I_{i+r_l, j+c_l} \leq I_{i,j} - \text{tol} \\ 0, & \text{if } |I_{i+r_l, j+c_l} - I_{i,j}| < \text{tol} \end{cases} \quad (18)$$

where r_l and c_l denote the relative coordinates of a neighbor N_l w.r.t. a center pixel, and tol is a preset tolerance. However, the number of possible LTP codes is often huge to effectively encode. Though one may reduce this number by considering all shift-equivalent codes as one, or separating a ternary code into two binary codes, we find that the simple frequency count of each code in a code string has already revealed many intrinsic properties of pixels, and we call them the Relative Darkness Index (RDI) features. Precisely, given the code \mathcal{C} and neighbors on a radius r circle, the RDI feature can be defined as below

$$X_{i,j}^{\text{RDI}|\mathcal{C}, r} = \sum_{l=1}^k \frac{\mathbf{1}_0(\text{ltp}(P_{i,j}, l) - \mathcal{C})}{k}. \quad (19)$$

As one can see from Fig.4(c-e), most of the nearly homogeneous background parts are of high code 0 indices; pixels close to strong edges are dominated by code+1 indices, and foreground text pixels have high response on code-1 indices. To further enhance RDI's discriminative power, we compute the ratios of one code to the sum of itself and another code as well (see Fig.4(f-h)).

6) *Other Features*: Besides of features discussed above, we extract features from the global image statistics, including the mean and standard deviation of the entire image intensities, the mean and standard deviation of the percentile image, the 32 bins of normalized histogram (sum to 1) for image intensities, and the 32 bins of a normalized logarithmed histogram.

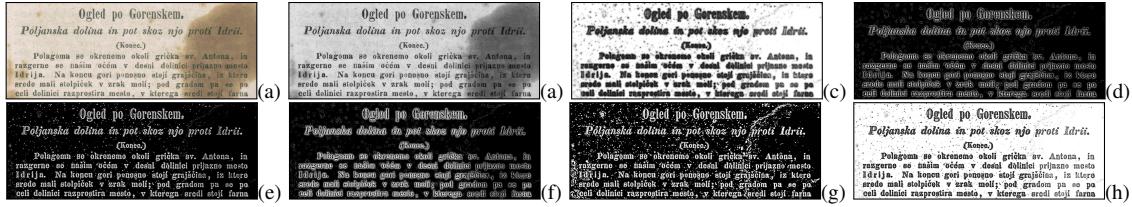


Fig. 4: RDI features for image DIBCO2013_PR05 (darker pixels indicate a value close to 0). (a) original color image; (b) original image; (c)-(e) RDI feature $X^{\text{RDI}|\mathcal{C},8}$ for $\mathcal{C} \in \{0, -1, +1\}$, respectively; (f) $\frac{X^{\text{RDI}|\mathcal{C}=+1,8}}{X^{\text{RDI}|\mathcal{C}\in\{0,+1\},8}}$; (g) $\frac{X^{\text{RDI}|\mathcal{C}=-1,8}}{X^{\text{RDI}|\mathcal{C}\in\{-1,+1\},8}}$; and (h) $\frac{X^{\text{RDI}|\mathcal{C}\in\{-1,0,+1\},8}}{X^{\text{RDI}|\mathcal{C}\in\{-1,0\},8}}$.

III. TRAINING AND TESTING SETTINGS

In experiments, we use the widely accepted Document Image Binarization Contest (DIBCO) from 2009 to 2014 [8], [9], [10], [11], [12], [13] as our training and testing data; it totals 76 images. We adopt the leave-one-out strategy where we first pick a DIBCO image set of a particular year as our testing set, and use the rest as our training set.

1) *Feature Summary*: We summarize all used features with dimensions and corresponding normalization considerations in Table I. Here, the stroke width s can be estimated via various methods; we use Su’s method [3]. ‘Scale’ indicates the side of local square region R .

TABLE I: Used Features

Type	Scale	Dimension	Normalization
Local int.	N/a	1	divide by 255
Otsu diff.	N/a	1	divide by 255
Local avg./std.	$1s, 2s, 4s, 8s$	4/4	divide by 255
Su/Howe	$1, 1s, 2s, 4s$	4/4	MinMax
ETRI/LTSI	$1s, 2s, 4s, 8s$	4/4	N/a
LIP	$1, 1s, 2s, 4s, 8s$	$1+4 \times 4+1$	N/a
RDI	$1, 1s, 2s, 4s, 8s$	5×6	N/a
Global int. avg./std.	N/a	1/1	divide by 255
Global perc. avg./std.	N/a	1/1	N/a
Global int./perc. logist.	N/a	32/32	N/a
Total		142	

2) *Sampling Strategy*: Selecting training samples is essential in task. First, one may not be handle a big training set of this task. These 76 images totally contain more than 80 million pixels. Assuming each feature is store in float32 format, we need $80 \times 142 \times 4\text{MB}$ ($\gg 256\text{GB}$) memory for just training features, while this requirement clearly beyond the capacities of most computers nowadays. Second, one may notice the imbalanced training data. We know both the background nontext class and foreground text class in the binarization problem actually cover different subclasses [19], [23], while we also know nearly homogeneous background and foreground dominate our training data.

To solve both problems, we first artificially classify all pixels in an image into 16 subclasses, each is represented as a 4-bit string, where each bit indicates whether or not this pixel should be treated as a pixel in Otsu’s foreground, in Niblack’s foreground, within s pixels away from reference image edges, and in a reference annotated foreground. We draw the same number of random samples for each subclass. Fig. 5 illustrates the samples we extracted that balanced both foreground and background subclasses.

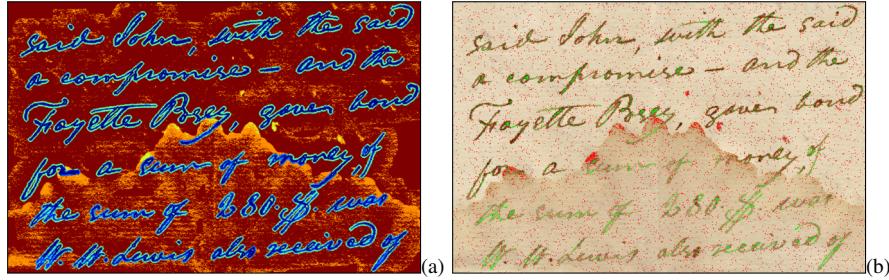


Fig. 5: Sampling strategy. (a) pixels with subclass labels for image DIBCO2012-HW02 (each color denotes a subclass); (b) samples extracted from DIBCO2012-HW02 image balanced subclasses (red/green dots indicate background/foreground.)

3) *Training and Testing Strategies*: In all of the following experiments, we perform a two-pass training. We first extract 9,600 samples (subclass balanced) from each training image and train a simple classifier, say Gaussian Naive Bayes. We use this classifier to decode all training images, and extract additional 9,600 erroneous samples (subclass balanced) from each image, and use all extracted samples to train an more complicated *sklearn* [24] *ExtraTrees* classifier [25]. Note in total we extract 19,200 samples per image, which only account for roughly about 1.5% of all samples. Classifier parameters are obtained from a 10-folded cross-validation using all samples. A final classifier is trained by using all extracted samples and validated parameters. Fig. 6 plots the feature importance of each feature type in terms of the overall contribution and the averaged dimensional contribution with respect to each feature type. As one can see, RDI, Global int. hist. and LIP are the three most useful feature categories in terms of overall contributions; and Su, LTSI and RDI are the three best features in terms of dimensional contributions.

In testing, we use the final classifier to predict the class label for all pixels in a testing image. Depending on the size of an image, the decoding time may vary between 5s to 30s.

IV. EXPERIMENTAL RESULTS

1) *Performance on DIBCO Datasets*: Table II lists performance of our proposed supervised binarization solution over the DIBCO 2012 [11], 2013 [12], and 2014 [13] datasets using standard metrics F1-score, peak signal-to-noise ratio (PSNR), and distance reciprocal distortion (DRD) (metric definitions can be found in [8], [9], [10], [11], [12], [13]). As we can see, our performance is comparable to the top five

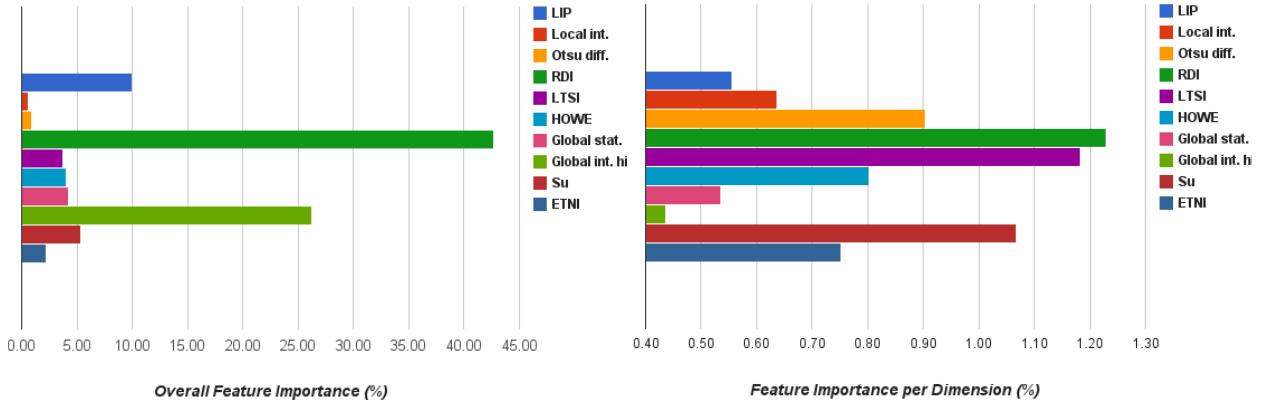


Fig. 6: Feature importance. Left: overall importance of each feature type; and Right: dimensional feature importance for each feature type.

methods. We also notice that our binarization classifier’s performance is very stable among all three datasets, especially since it always keeps a DRD score below 3 pixels. Sample decoding results are compared to the top two contest methods in Fig. 7. As one can see, our supervised solution successfully learnt knowledge to handle difficult cases: 1) faded text; and 2) text on a dirty background.

TABLE II: Performance Evaluations On DIBCO Datasets

	Method	Contest Rank	F1%	PSNR	DRD
DIBCO2012	[4]	1	89.47	21.80	3.400
	Lelore <i>et al.</i> ’s [11]	2	92.85	20.57	2.660
	[2]	3	91.54	20.14	3.048
	Nina’s [11]	4	90.38	19.30	3.348
	Yazid <i>et al.</i> ’s [11]	5	91.85	19.65	3.056
	Ours		92.01	19.92	2.601
DIBCO-2013	Su <i>et al.</i> ’s method [12]	1	92.12	20.68	3.100
	[5]	2	92.70	21.29	3.180
	[20]	3	91.81	20.68	4.020
	[26]	4	91.69	20.54	3.590
	[23]	5	90.92	19.32	3.910
	Ours		91.40	20.13	2.637
DIBCO-2014	Mesquita <i>et al.</i> ’s [13]	1	96.88	22.66	0.902
	[5]	2	96.63	22.40	1.001
	[27]	3	93.35	19.45	2.194
	Ziaratban <i>et al.</i> ’s [13]	4	89.24	18.94	4.502
	Mitianoudis <i>et al.</i> ’s [13]	5	89.77	18.49	4.502
	Ours		92.69	19.47	2.571

2) *Learning Curve*: As we mentioned previously, only about 1.5% of all available training samples are used in our experiments. We investigate the relationship between the amount of training samples and the binarization performance using the test set of DIBCO 2012 in Table III. As in many pattern recognition

DIBCO2012_H07	DIBCO2014_HW07
<p>the city seeking to <u>destroy</u> it without war seeking to dissolve the Union, and divide effects by re-negotiation. Both parties deprecated war; but one</p>	<p><i>Linaria Osiris quorundam</i>. 782. 773.774.775. <i>Linaria 1^o Panorum Cluy</i>. 790 <i>Linaria tertium genus.</i> <i>Linaria tertia Sync Cluy</i>. 795 <i>Trago.</i> <i>Linana Moravica quarta</i> <i>Clusij.</i> 797. <i>Linaria diversa de Colom</i> <i>Linaria cernua Apula</i> <i>fratilla Colone.</i> — (a)</p>
<p>the city seeking to <u>destroy</u> it without war seeking to dissolve the Union, and divide effects by re-negotiation. Both parties deprecated war; but one</p>	<p><i>Linaria Osiris quorundam</i>. 782. 773.774.775. <i>Linaria 1^o Panorum Cluy</i>. 790 <i>Linaria tertium genus.</i> <i>Linana tertia Sync Cluy</i>. 795 <i>Trago.</i> <i>Linana Moravica quarta</i> <i>Clusij.</i> 797. <i>Linaria diversa de Colom</i> <i>Linaria cernua Apula</i> <i>fratilla Colone.</i> — (b)</p>
<p>the city seeking to <u>destroy</u> it without war seeking to dissolve the Union, and divide effects by re-negotiation. Both parties deprecated war; but one</p>	<p><i>Linaria Osiris quorundam</i>. 782. 773.774.775. <i>Linaria 1^o Panorum Cluy</i>. 790 <i>Linaria tertium genus.</i> <i>Linana tertia Sync Cluy</i>. 795 <i>Trago.</i> <i>Linana Moravica quarta</i> <i>Clusij.</i> 797. <i>Linaria diversa de Colom</i> <i>Linaria cernua Apula</i> <i>fratilla Colone.</i> — (c)</p>
<p>the city seeking to <u>destroy</u> it without war seeking to dissolve the Union, and divide effects by re-negotiation. Both parties deprecated war; but one</p>	<p><i>Linaria Osiris quorundam</i>. 782. 773.774.775. <i>Linaria 1^o Panorum Cluy</i>. 790 <i>Linaria tertium genus.</i> <i>Linana tertia Sync Cluy</i>. 795 <i>Trago.</i> <i>Linana Moravica quarta</i> <i>Clusij.</i> 797. <i>Linaria diversa de Colom</i> <i>Linaria cernua Apula</i> <i>fratilla Colone.</i> — (d)</p>
<p>the city seeking to <u>destroy</u> it without war seeking to dissolve the Union, and divide effects by re-negotiation. Both parties deprecated war; but one</p>	<p><i>Linaria Osiris quorundam</i>. 782. 773.774.775. <i>Linaria 1^o Panorum Cluy</i>. 790 <i>Linaria tertium genus.</i> <i>Linana tertia Sync Cluy</i>. 795 <i>Trago.</i> <i>Linana Moravica quarta</i> <i>Clusij.</i> 797. <i>Linaria diversa de Colom</i> <i>Linaria cernua Apula</i> <i>fratilla Colone.</i> — (e)</p>

Fig. 7: Binarization results for image. (a) original images; (b) reference binarized images (highlighted red regions indicate disagreements); (c) results of contest rank 1; (d) results of contest rank 2; and (e) our results.

problems, the improvement of binarization performance gets smaller as the number of samples increases.

TABLE III: Performance v.s. Training Samples

#Samples	1,920	5,760	9,600	13,440	15,360	17,280	19,200
F1%	91.47	91.77	91.90	91.93	91.95	92.01	92.01
PSNR	19.64	19.81	19.86	19.86	19.88	19.93	19.92
DRD	2.797	2.689	2.637	2.634	2.618	2.599	2.601

3) *Document Binarization in the Wild* : Although images in DIBCO datasets have already covered a wide range of variations, there are clearly more variations and combinations of variations that are not included in DIBCO training data. We therefore test our learned classifier on out-of-domain document images, and we observe satisfactory results (see Fig. 8).

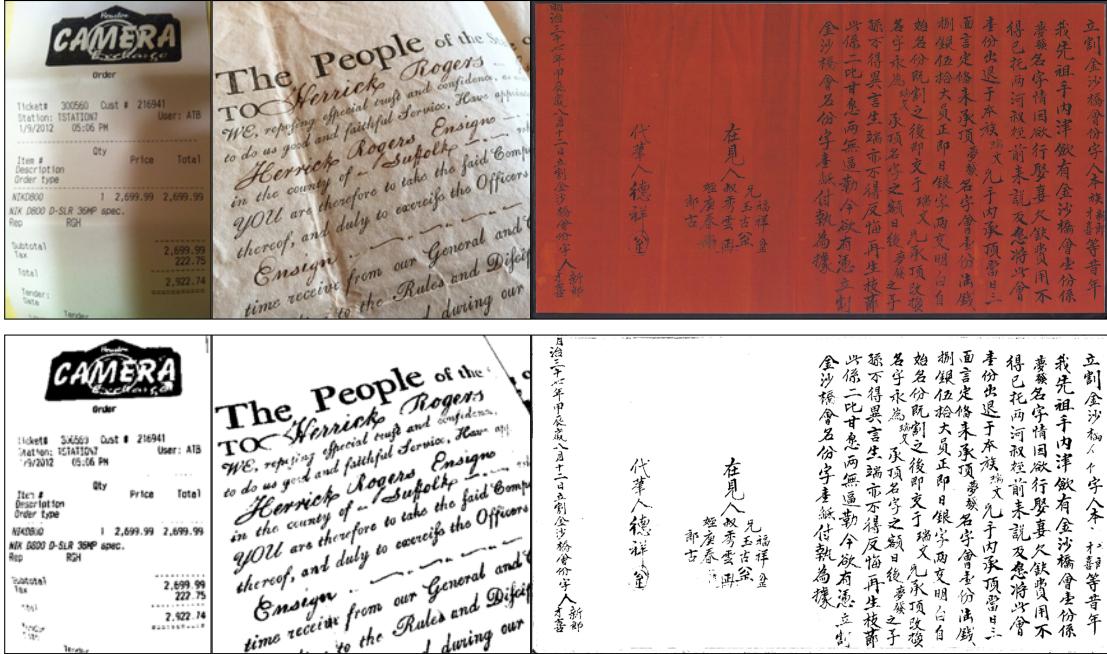


Fig. 8: Binarization results of out-of-domain data

V. CONCLUSION

In this paper we investigate the document binarization solution via supervised learning. Unlike previous efforts, this solution is parameter-free and fully trainable. Our experimental results showed that one can learn a reasonably well binarization decision function from a small set of carefully selected training data. Such a learned decision function not only works well for in-domain data, but can also apply to out-of-domain data. In future work, we will explore several interesting aspects such as discriminative features (e.g., image moments and connected component attributes) and classifier adaptation on the fly.

REFERENCES

- [1] N. Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [2] B. Su, S. Lu, and C. L. Tan, “Binarization of historical document images using the local maximum and minimum,” in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. ACM, 2010, pp. 159–166.
- [3] ———, “Robust document image binarization technique for degraded document images,” *Image Processing, IEEE Transactions on*, vol. 22, no. 4, pp. 1408–1417, 2013.
- [4] N. R. Howe, “A laplacian energy for document binarization,” in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011, pp. 6–10.
- [5] ———, “Document binarization with automatic parameter tuning,” *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 3, pp. 247–258, 2013.
- [6] W. Niblack, *An Introduction to Digital Image Processing*. Prentice-Hall, 1986.
- [7] J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern recognition*, vol. 33, no. 2, pp. 225–236, 2000.
- [8] B. Gatos, K. Ntirogiannis, and I. Pratikakis, “Icdar 2009 document image binarization contest (dibco 2009),” in *Document Analysis and Recognition (ICDAR), 2009 International Conference on*, vol. 9, 2009, pp. 1375–1382.
- [9] I. Pratikakis, B. Gatos, and K. Ntirogiannis, “H-dibco 2010-handwritten document image binarization competition,” in *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*. IEEE, 2010, pp. 727–732.

- [10] ——, “Icdar 2011 document image binarization contest (dibco 2011),” in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, 2011, pp. 1506–1510.
- [11] ——, “Icfhr 2012 competition on handwritten document image binarization (h-dibco 2012).” *ICFHR*, vol. 12, pp. 18–20, 2012.
- [12] ——, “Icdar 2013 document image binarization contest (dibco 2013),” in *Document Analysis and Recognition (ICDAR), 2013 International Conference on*. IEEE, 2013, pp. 1471–1476.
- [13] K. Ntirogiannis, B. Gatos, and I. Pratikakis, “Icfhr2014 competition on handwritten document image binarization (h-dibco 2014),” in *2014 14th International conference on frontiers in handwriting recognition*, 2014, pp. 809–813.
- [14] ——, “A combined approach for the binarization of handwritten document images,” *Pattern Recognition Letters*, vol. 35, pp. 3–15, 2014.
- [15] X. Peng, H. Cao, R. Prasad, and P. Natarajan, “Text extraction from video using conditional random fields,” in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, Sept 2011, pp. 1029–1033.
- [16] S. Lu, B. Su, and C. L. Tan, “Document image binarization using background estimation and stroke edges,” *International journal on document analysis and recognition*, pp. 1–12, 2010.
- [17] J. Sauvola, T. Seppanen, S. Haapakoski, and M. Pietikainen, “Adaptive document binarization,” in *Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on*, vol. 1. IEEE, 1997, pp. 147–152.
- [18] G. Lazzara and T. Géraud, “Efficient multiscale sauvolas binarization,” *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 17, no. 2, pp. 105–123, 2014.
- [19] B. Su, S. Lu, and C. L. Tan, “A learning framework for degraded document image binarization using markov random field,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 3200–3203.
- [20] R. F. Moghaddam, F. F. Moghaddam, and M. Cheriet, “Unsupervised ensemble of experts (eoe) framework for automatic binarization of document images,” in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013, pp. 703–707.
- [21] H. Ziae Nafchi, R. Farrahi Moghaddam, and M. Cheriet, “Phase-based binarization of ancient document images: Model and applications,” 2014.
- [22] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [23] M. A. Ramírez-Ortegón, E. Tapia, L. L. Ramírez-Ramírez, R. Rojas, and E. Cuevas, “Transition pixel: A concept for binarization based on edge detection and gray-intensity histograms,” *Pattern Recognition*, vol. 43, no. 4, pp. 1233–1243, 2010.
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [25] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [26] T. Lelore and F. Bouchara, “Fair: A fast algorithm for document image restoration,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 2039–2048, Aug 2013.
- [27] H. Z. Nafchi, R. F. Moghaddam, and M. Cheriet, “Historical document binarization based on phase information of images,” in *Computer Vision-ACCV 2012 Workshops*. Springer, 2013, pp. 1–12.