High Performance OCR for Camera-Captured Blurred Documents with LSTM Networks

Fallak Asad[1], Adnan Ul-Hasan[2,3], Faisal Shafait[1] and Andreas Dengel[2,3]

[1]NUST School of Electrical Engineering and Computer Science, Islamabad, Pakistan
[2]University of Kaiserslautern, Kaiserslautern, Germany
[3]German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany.
fallak.asad@seecs.edu.pk, faisal.shafait@seecs.nust.edu.pk, adnan@cs.uni-kl.de, Andreas.Dengel@dfki.de

***Abstract*—Documents are routinely captured by digital cameras in today's age owing to the availability of high quality cameras in smart phones. However, recognition of camera-captured documents is substantially more challenging as compared to traditional flat bed scanned documents due to the distortions introduced by the cameras. One of the major performance-limiting artifacts is the motion and out-of-focus blur that is often induced in the document during the capturing process. Existing approaches try to detect presence of blur in the document to inform the user for re-capturing the image. This paper reports, for the first time, an Optical Character Recognition (OCR) system that can directly recognize blurred documents on which the state-of-the-art OCR systems are unable to provide usable results. Our presented system is based on the Long Short-Term Memory (LSTM) networks and has shown promising character recognition results on both the motion-blurred and out-of-focus blurred images. One important feature of this work is that the LSTM networks have been applied directly to the gray-scale document images to avoid error-prone binarization of blurred documents. Experiments are conducted on publicly available SmartDoc-QA dataset that contains a wide variety of image blur degradations. Our presented system achieves 12.3% character error rate on the test documents, which is an over three-fold reduction in the error rate (38.9%) of the best-performing contemporary OCR system (ABBYY Fine Reader) on the same data.**

I. Introduction

Optical Character Recognition (OCR) has grown into a mature research area over the last 50 years [1] and many practical systems, e.g. post sorting and routing, cheques processing and verification, inbox automation, etc. are in place [2]. However until the last decade, research in the field of OCR was limited to the document images obtained from the flatbed scanners. Recently, with the improvement in internal memory and processing speed of hand-held mobile devices such as Smart Phones, new ways of capturing and processing document images are emerging; hence arises the need to find reliable ways for text or information extraction from these camera-captured images. However, where advancement in the technology has given birth to new ways for the document capturing, it has also created the new challenges in the field of OCR. The camera-captured documents are subjected to low image quality that may occur due to light distortion, motion blur, out-of-focus blur, perspective distortion or camera noise, which makes the OCR from such images a difficult task [3].

Among the various image degradations, the out-of-focus blur and the motion blur are the most common one. Motion blur is caused by rapid movement of the camera or by relative speed between the object and the camera, while the out-of-focus blur is caused when light is not converged in the image. Motion blur and out-of-focus blur prevent accurate recognition of the text contents in camera-captured documents [4]. A few samples of text-line images extracted from SmartDoc-QA dataset [5] are shown in Fig. 1.

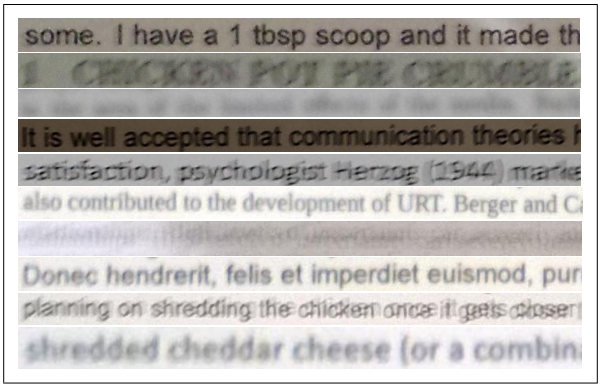


Fig. 1. Sample images of blurred text-lines from SmartDoc-QA dataset [5]

To tackle the problem of blur, the research has been focused on the mechanisms for estimating the blur in the document images to predict the OCR accuracy thus supplying the feedback to the user to re-capture the images to achieve better OCR results. Kumar et al. [6] proposed the use of $\vec{\Delta}$DOM (change in slope) to model the changes in the direction of edges to estimate the sharpness in images, whereas Rusiñol et al. [7] used different focus measure operators and combined their results to estimate the OCR accuracy in the image. Kai-Chieh et al. [8] used machine learning approaches for blur estimation achieving promising results. Later, Ye et al. [9] proposed a supervised feature learning approach for automatically learning features relevant to document image blur. However, usability of the blur estimation techniques is limited by the fact that re-capturing a document till one acquires a sharp image is not only time consuming, but not always feasible (for instance in the case of incidental text [10]).

To avoid this problem, many image de-blurring techniques have been developed; nevertheless restoring the blurred images is challenging because of the non-linear blur motion, which

978-1-5090-1792-8/16 $31.00 © 2016 IEEE
DOI 10.1109/DAS.2016.69

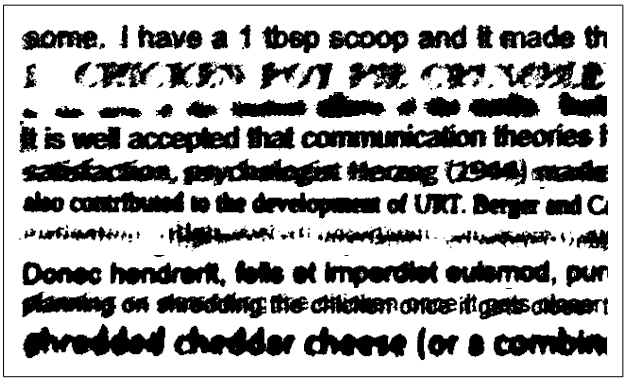


Fig. 2. Sample images from Fig. 1 thresholded using Sauvola binarization

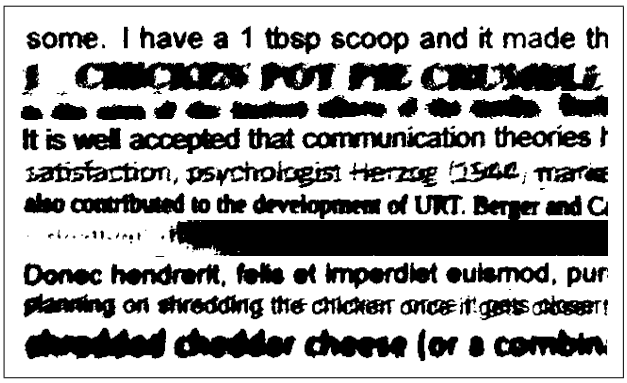


Fig. 3. Sample images from Fig. 1 thresholded using Otsu binarization

makes this problem strictly under-constrained. Most of the de-blurring techniques proposed in the literature [11], [12], [13] work well on the natural images but fail on images containing text due to the ringing artifacts that are produced as the result of de-blurring. Lu [4] proposed a de-blurring procedure for document images to increase the OCR accuracy but the results are not satisfactory because the image gets degraded due to ringing effect. Thus, restoring blurred document images remains a challenging task.

The primary challenge in directly recognizing blurred documents originates from the binarization step as existing OCR systems take binary images as input. Results of applying widely used Sauvola [14], [15] and Otsu [16] binarization algorithms on sample blurred text-lines from Fig. 1 are shown in Figs. 2 and 3. OCR systems can be broadly categorized into segmentation-based and segmentation-free approaches. Tesseract [17] is an example of segmentation-based OCR system, which works by segmenting the lines into character hypotheses and applying a shape based geometric classifier on the outline of each character hypothesis to recognize it. These segmentation-based approaches require accurate segmentation of characters, since the inaccurate segmentation would lead to an inaccurate OCR output. The segmentation process is highly sensitive to the quality of the binarization step. Hence segmentation errors originating from binarization failures become the limiting factor of these OCR methods. In [18] another segmentation based OCR system is proposed for recognition of low-resolution characters. A generative learning method is employed for training on isolated low-resolution characters. During recognition, character segmentation is done on the basis of features extracted from the characters and inter-character spaces.

To overcome the problems of accurate segmentation, many segmentation-free approaches based on Hidden Markov Model (HMMs) [19], [20] and scanning neural networks [21] have been developed for OCR. In recent years, the Long Short-Term Memory (LSTM) neural networks [22], which are a specialization of Recurrent Neural Networks (RNN), have been used for many tasks related to classification and prediction of time series data. These networks have outperformed alternative RNNs, HMMs, their hybrid and other sequence learning methods in many applications. Some benchmark records obtained with the help of LSTMs are optical character recognition of Latin [23] and Asian scripts [24], text-to-speech recognition [25] [26], and handwriting recognition [27]. This excellent performance of LSTM networks motivated us to investigate the application of LSTM network for recognition of blurred document images. To eliminate the need of image de-blurring or binarization for blurred documents, we have employed training of LSTM networks directly on the gray scale images. The bidirectional LSTM (BLSTM) architecture has been used to benefit from the context of each character from both left and right directions.

The rest of this paper is organized as follows: Section II presents a short overview of LSTM networks along with the pre-processing, feature extraction, and LSTM architecture used in this work. Section III describes the design of experimental evaluation, database used and compares the results obtained from the proposed system with other state-of-the-art commercial and open-source OCR systems. Section IV concludes the paper with a brief summary of the work and pointers to possible future directions.

II. Methodology

To reduce the impact of different image degradations, we proceeds the experiment with a few pre-processing steps: (i) To remove the perspective distortion, the image is first converted to a gray-scale image. Then we applied the canny edge detector [28] to extract the boundaries of the document. In order to segment the document from the background, we applied morphological operations (erosion, dilation and hole filling) on the image. To remove the noise outside the boundary of the document we find the connected components and filter out the components that have an area less than a predefined threshold. Then, we apply Harris corner detection [29] to detect the corners and filter the detected corners in order to extract the four corner points of the document. Finally, we apply geometric image transformation to warp the perspective using the four corner points. (ii) Run Length Smearing Algorithm (RLSA) [30] is then used for extracting text lines from the document images. (iii) The text-line images extracted from RLSA are normalized to a fixed height of 48 pixels.

A. Long Short-Term Memory (LSTM) Networks

RNNs were first proposed in 1980s and became quite popular due their power of representing temporal data but went out of flavor after a while due to problems in training these networks (vanishing and exploding gradient problem, i.e., the

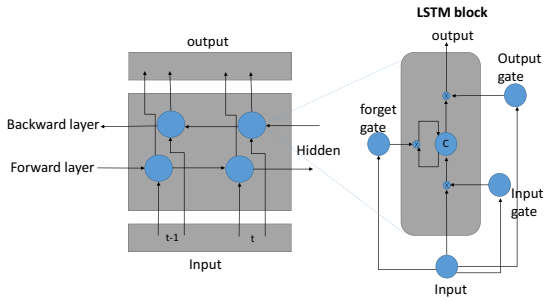


Fig. 4. Bi-directional LSTM architecture

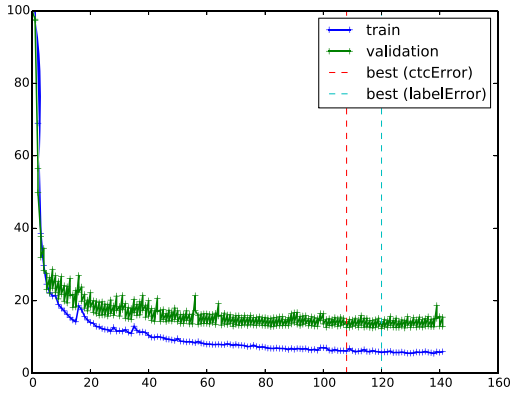


Fig. 5. Training and validation error rates on different epochs. Training was stopped after 20 epochs of no improvement in validation error.

TABLE I. CHARACTER ERROR RATE OF VARIOUS OCR SYSTEMS ON SMARTDOC-QA DATASET. ALL RESULTS ARE IN %

OCR System	Full-Image		Text-line Image	
	Default	Sauvola	Default	Sauvola
Fine-Reader	38.9	50.1	53.3	55.5
Tesseract	71.4	65.9	66.3	70.4
Ocropus	-	-	40.7	56.3
BLSTM Text-lines on Gray scale Images				12.3%

gradient either vanishes or explodes when back propagated through time) [31]. In 1997, Hochreiter and Schmidhuber introduced a new RNN architecture called as the Long Short-Term Memory (LSTM) which could propagate back the gradients in long time-steps without attenuating the gradient using a gating approach [22]. The network not only learns to predict the labels in the given data but also learns the gating i.e. when to write data to an LSTM cell and when to read data from the LSTM cell. In 2000, Gers et al. [32] suggested an improvement in LSTM cell by adding a forget gate which allowed the network to forget information over time. Graves introduced Bi-Directional LSTMs [25] in which he used two hidden layers; one moving from left to right (forward layer) and the other one moving from right to left (backward layer). Both the forward as well as the backward context could be learnt at the same time. Graves demonstrated the use of LSTMs on transcription tasks by aligning the output of the network with transcripts using a forward-backward algorithm referred to as CTC (Connectionist Temporal Classification) [33]. The network when employed with CTC as an output layer outperformed both an HMM recognizer and an HMM-RNN hybrid with the same RNN architecture [27].

B. Pre-processing and Feature Extraction

For the reported experiments, the raw pixels of the images are used as the features. The window of size 1×48 traverses through the images. The resulting one-dimensional (1D) sequence of pixels is normalized using Z-score, which transformed the data into new data with a mean of 0 and standard deviation of 1. This normalized data is then fed to the LSTM network for training. A total of 83 labels were extracted from the ground truth data including the alphabets in upper and lowercase, numbers, punctuation marks and a blank label. All foreign symbols were either removed from the ground truth or replaced with their English equivalent (e.g $\acute{E}$ would be replaced by the letter E).

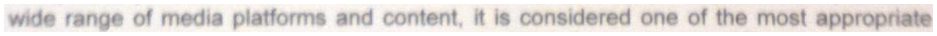
C. Network Architecture

Bidirectional LSTM is used in our work for the OCR of blurred documents. The architecture of BLSTM is shown in Fig. 4. For training, 1D BLSTM with 100 blocks in both the forward and backward hidden layers as well as the CTC (Connectionist Temporal Classification) layer have been employed. Training is carried out with the back-propagation algorithm through time and Steepest optimizer have been used with learning rate of 10^{-4} and with a momentum value of 0.9. Text line images and their corresponding ground-truth were extracted from the data-set provided in the SmartDoc QA data-set. These text lines images are then normalized to the fixed height of 48 and fed to the network for training. Training and validation errors were recorded. Training stops when the validation error shows no improvement in successive 20 epochs. For the experiment described in this paper, the open source library, RNNLib is used. RNNLib is a recurrent neural network library for sequence learning problems. This library provides the implementation of single and multi-dimensional recurrent neural networks, along with the CTC layer for aligning the output of the network with transcription using a forward-backward algorithm [33]. The training of the network proceeds with feeding the text-lines to the network and carrying out the forward propagation through the LSTM network and then performing forward-backward alignment of output labels with its ground truth. Lastly, backward propagation is carried out to adjust the weights of the network. The network took 141 epochs to converge and reported with minimum label error at 121 epoch on the validation set.

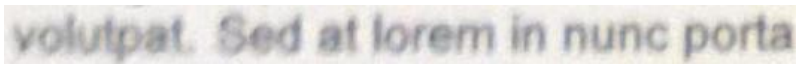
III. EXPERIMENTAL EVALUATION

A. Database

In order to evaluate the performance of the LSTM network, a subset of the SmartDoc-QA dataset was used. The SmartDoc-QA dataset contains three different categories of paper documents: category 1 contains contemporary documents (office documents, letters, etc.), category 2 contains old administrative documents, while category 3 contains a set of receipts. Receipts and old administrative documents have non-typical page layouts. As the primary focus of this paper is on the recognition of blurred text rather than page segmentation /



Fine Reader	wide range of media platform* and content, it is considered one of the most appropriate
Tesseract	mvmdmummmcammm.nhmmmdmdmmtm
Ocropus	wsde tange of medis platforrn and content, M is considered one of the moet appropnuete
Our Approach	wide range of media platforms and content, it is considered one of the most appropriate



Fine Reader	voMpa! a! kxam m nunc porta
Tesseract	m. Mallorunlnmncpom
Ocropus	woxget Sed a korern Mn nunc ports
Our Approach	volutpat. Sed at lorem in nunc porta

Fine Reader	gootS
Tesseract	In“ vmnquwaflm“~l
Ocropus	aau t ee wg ee adasesei b eeeea eg eee good
Our Approach	thoughts. I made ry own adjustments, but overall they were good

Fine Reader	t»4wk* **m*l*i em*ee»« >he sin • ■ ■< <«fcurr mdwits v »?>'(*> ns Tt<<tt**c+i :h< concept <m he
Tesseract	
Ocropus	h ar& eeew ereggssss ef s ef g ese dsw as wefl ss reeegies e eoeeeg ceA he
Our Approach	In order to understand the creation of the ides of the culture industry as well as its reception the concept can be

Fine Reader	mm-m** mul 4r» mm I - *t* l# I; ymi im
Tesseract	Malt-Gym! “an” M‘fifimhmmfl
Ocropus	ss ss s e @rs estl - es4 Hb 17 sSssss eeseseg. f yes es gas3 s efwegee, sass ees
Our Approach	ovecook or it will dry out) - about 10-12 minutes. Alternately, if you are planning in advance, you can

Fig. 6. Resultant output of different OCR systems on out-of-focus images. Notice that when the amount of blur increases, for instance in the bottom three cases, existing OCR systems produce unusable output. Our BLSTM network trained on gray scale images is able to extract most of the characters correctly even in those scenarios.

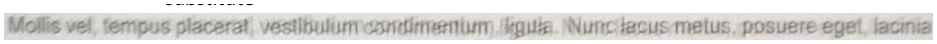
layout analysis [34], only category 1 documents have been considered for the experiments. For computational reasons, we sub-sampled random text-lines from the category 1 documents while ignoring images in the sharp / well-focused sub-category. A total of 9,557 images were used for the evaluation out of which 5,966 images were used for training, while 1,817 and 1,774 images were used for validation and testing respectively.

B. Evaluation Protocol

To evaluate the accuracy of BLSTM networks for recognition of blurred documents, we have used normalized edit distance for error calculation while ignoring the spaces. The overall accuracy is calculated using the ratio of insertions, deletions and substitution operations divided by the total number of characters in the transcription.

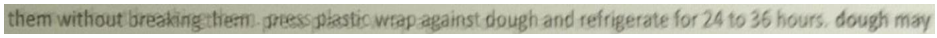
C. Comparative Analysis and Discussion

To compare the results with other OCR systems, we have used the same test set. For the comparison three state-of-the-art OCR system have been considered: Tesseract [17], FineReader [35] and OCRopus [36]. These OCR system have been evaluated on the full document image as well as on the extracted text lines images. Since these OCR systems are adaptive, they generally yield better results when presented with the full page image instead of individual text-lines. OCRopus uses the percentile filtering method [37] while Tesseract use Otsu

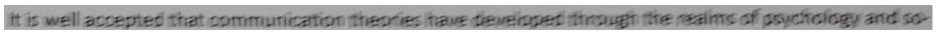


Fine Reader	Mdltsv# oos ^'WifOm^finurr kpu;t fturrEtecus metus posuereecjet taarua
Tesseract	“Us, mm; Wmmwm mum Wmmusmetm. mqu, m
Ocropus	Wogs wet temgus plicerat, wesiibukin 3Gnndimsrnntuum kguka 8unc iapus rmetus. 9osuee eget. tCaY
Our Approach	Mollis vel , tempus placerat , vestbulum condimenum , , iqulla . Niunlacus metus , posuereeget , lacinia

Fine Reader	1.1 APPLE MUSTARD PORK BURGEES
Tesseract	1.1 APPLE MUSTARD PORK BURGER”;
Ocropus	1.1 APPLE MUSTARD PORK EURrIEPS
Our Approach	1.1 APPLE MUSTARD PORK RURGEPS



Fine Reader	them without breaking hem 3tv&t DztstK wras agams* dough and refrigerate for 24 to hours dough m^ry
Tesseract	119th :efngeme f0! u to 36 hour; dough man]
Ocropus	WEIGHT VOLUME iNGREO)ES' YL>uuG HSTHUCTONS 8 1/2 oz cake fkour sMt ftours, baksg sodet malesuada
Our Approach	them without breaking thher. preess plasstic wrap against dough and refrigerate for 24 to 36 hours . dough may



Fine Reader	r v i xxx xx~'"x\<'jtrxx vxxx '*x~ i&**£ 'Uz*&crj&&t *1 "jxy. :ti* "c gf C'CJfj} *t\ti
Tesseract	n is net! Written «emafmam»
Ocropus	tt wwell azttepter ttset cprmunuCatror tsepsez rsaswe tesegser trrruuug tse searmrec sf 3erdcMogy annd L5-
Our Approach	Ht is well accepted that communication theores hhave deveoped thrrough te realms of psychology and so-

Fig. 7. Output of different OCR systems on a few sample text-lines images containing motion blur. In contrast to out-of-focus blur, the motion blurred text shows highly asymmetric degradation. Even though the BLSTM network was trained on mixed text-lines containing both types of blur, it is able to handle both cases with a high accuracy.

thresholding [16] for binarizing the images. We have additionally binarized the images with Sauvola binarization technique. Experiments are performed using different parameter values and window sizes of Sauvola to find the most suitable values. For full images, a threshold value of 0.1 and a window size of 100×100 yields the best results, whereas for text-line images, a threshold value of 0.24 and a window size of 30×30 are found suitable. Moreover, before feeding the images to the Fine Reader, “straighten text lines”, “correct image resolution” and “remove motion blur” options have been enabled. The results of these experiment are shown in the Table I. Out of these three existing systems, Fine Reader achieves the lowest error of 38.9% on the text lines images. However the BLSTM when directly trained on the gray-scale images out performed all these OCR systems. The training of the BLSTM took approximately 212 hours on a 2.4 GHz Intel Xeon machine with 56 GB of RAM. Fig. 6 and 7 show the output of the different OCR systems on out-of-focus blurred and motion blurred images respectively and demonstrate the effectiveness of the BLSTM network for recognition of blurred text. Note that the characteristics of both types of blur are unique. Out-of-focus blur exhibits a more Guassian-like distortion, whereas motion blur is asymmetric causing letters to cast shadows. Even though we trained a single BLSTM network on a mixture of text-lines that contained both of these distortions, the network was able to process both kinds of distortions accurately.

IV. Conclusion

This paper presented the use of BLSTM networks for optical character recognition of camera captured blurred documents. The existing open-source and commercial OCR systems are either limited by the factor of segmentation or by the techniques they employ for binarizing document images. Our OCR approach has not only overcome the problem of segmentation based OCR systems, but also eliminates the need of binarization of blurred documents. The BLSTM produced competitive results even when trained on a small-sized training set. The error rate achieved by our algorithm is three times lower than the best error rate obtained by state-of-the-art OCR systems. In future, we plan to explore deblurring algorithms in combination with proposed and existing OCR systems to

investigate weather debluring would improve the performance of the proposed systems.

REFERENCES

[1] S. Mori, C. Suen, and K. Yamamoto, "Historical review of OCR research and development," *Proceedings of the IEEE*, vol. 80, no. 7, pp. 1029–1058, 1992.

[2] G. Nagy, "Twenty years of document image analysis in PAMI," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 38–62, 2000.

[3] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 7, no. 2-3, pp. 84–104, 2005.

[4] Q. Xing Yu, L. Zhang, and C. L. Tan, "Motion Deblurring for Optical Character Recognition," in *Document Analysis and Recognition, Eighth International Conference*, pp. 389–393, 2005.

[5] N. Nayef, M. M. Luqman, S. Prum, S. Skenazi, J. Chazalon, and J.-M. Ogier, "SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images - Single and Multiple Distortions," in *CBDAR*, (Nancy, France), aug 2015.

[6] J. Kumar, F. Chen, and D. Doermann, "Sharpness Estimation for Document and Scene Images," in *21st International Conference on Pattern Recognition (ICPR 2012)*, (Tsukuba, Japan), nov 2012.

[7] M. Rusiñol, J. Chazalon, and J.-M. Ogier, "Combining Focus Measure Operators to Predict OCR Accuracy in Mobile-Captured Document Images," in *Document Analysis Systems (DAS), 11th IAPR International Workshop on*, (Tours, France), april 2014.

[8] Y. Kai-Chieh, C. G. Clark, and D. Pankaj, "Motion blur detecting by support vector machine," in *Proc. SPIE 5916, Mathematical Methods in Pattern and Image Analysis*, aug 2005.

[9] P. Ye, J. Kumar, L. Kang, and D. S. Doermann, "Real-time no-reference image quality assessment based on filter learning," in *IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA*, pp. 987–994, Jun 2013.

[10] T. Kimura, R. Huang, S. Uchida, M. Iwamura, S. Omachi, and K. Kise, "The reading-life log – technologies to recognize texts that we read," *12th International Conference on Document Analysis and Recognition*, pp. 91–95, Aug 2013.

[11] C. Taeg Sang, S. Paris, B. Horn, and W. Freeman, "Blur kernel estimation using the radon transform," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*, (Providence, RI,), june 2011.

[12] F. Rob, S. Barun, H. Aaron, T. R. Sam, and T. F. William, "Removing Camera Shake from a Single Photograph ," in *ACM Trans. Graph*, vol. 25, pp. 787–794.

[13] Z. Haichao, Y. Jianchao, Z. Yanning, and T. Huang, "Sparse Representation Based Blind Image Deblurring," in *Multimedia and Expo (ICME), IEEE International Conference*, (Barcelona, Spain), july 2011.

[14] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," in *Pattern Recognition*, vol. 33, p. 225–236, February 2000.

[15] F. Shafait, D. Keysers, and T. M. Breuel, "Efficient implementation of local adaptive thresholding techniques using integral images," in *Proc. SPIE Document Recognition and Retrieval XV*, (San Jose, CA, USA), pp. 101–106, Jan. 2008.

[16] N. Otsu, "A threshold selection method from gray-level histograms," in *IEEE Trans. Sys., Man., Cyber*, vol. 9, p. 62–66, 1979.

[17] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. Ninth Int. Conference on Document Analysis and Recognition (ICDAR), tesseract-version ="3.4.0"*, pp. 629–633, 2007.

[18] H. Ishida and 石田皓之, "A generative learning method for low-resolution character recognition," 2009.

[19] S. F. Rashid, F. Shafait, and T. M. Breuel, "An evaluation of HMM-based techniques for the recognition of screen rendered text," in *International Conference on Document Analysis and Recognition, Beijing, China*, pp. 1260–1264, Sep 2011.

[20] S. España-Boquera, M. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez, "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 767–779, Aug 2010.

[21] S. F. Rashid, F. Shafait, and T. M. Breuel, "Scanning neural network for text line recognition," in *10th IAPR International Workshop on Document Analysis Systems, Gold Coast, Queenslands, Australia*, pp. 105–109, Mar 2012.

[22] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in *Neural Computation*, vol. 9, pp. 1735–1780, Nov 1997.

[23] T. M. Breuel, A. Ul-Hasan, M. I. A. A. Azawi, and F. Shafait, "High-performance OCR for printed English and Fraktur using LSTM networks," in *12th International Conference on Document Analysis and Recognition, Washington, DC, USA*, pp. 683–687, Aug 2013.

[24] A. Ul-Hasan, S. Bin Ahmed, F. Rashid, F. Shafait, and T. Breuel, "Offline Printed Urdu Nastaleeq Script Recognition with Bidirectional LSTM Networks," in *Document Analysis and Recognition (ICDAR), 12th International Conference*, (Washington, DC), aug 2013.

[25] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," in *Neural Networks*, pp. 5–6, 2005.

[26] S. Fernández, A. Graves, and J. Schmidhuber, "An Application of Recurrent Neural Networks to Discriminative Keyword Spotting," in *Artificial Neural Networks (ICANN) ,17th International Conference*, (Porto, Portugal), pp. 220–229, sep 2007.

[27] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A Novel Connectionist System for Unconstrained Handwriting Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 855–868, may 2008.

[28] J. Canny, "A computational approach to edge detection," in *Pattern Analysis and Machine Intelligence, IEEE Transactions, year = 1986, volume = PAMI-8, number = 6, month = Nov, pages = 679 - 698,*.

[29] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988.

[30] K. Wong, R. Casey, and F. Wahl, "Document analysis system. IBM," in *Journal of Research and Development*, vol. 26, pp. 647–656, nov 1982.

[31] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of The 30th International Conference on Machine Learning*, vol. 28, 2013.

[32] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," in *Neural Computation*, vol. 12, pp. 2451–2471, Oct. 2010.

[33] A. Graves, S. Fernandez, F. Gomes, and J. Schmidhuber, "Connectionist Temporal Classification: Labeling Unsegemented Sequence Data with Recurrent Nerual Networks," in *ICML*, (Pennsylvania, USA), p. 369–376, 2006.

[34] F. Shafait, D. Keysers, and T. M. Breuel, "Performance evaluation and benchmarking of six page segmentation algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 941–954, 2008.

[35] S. Gaurav, "Abbyy finereader 12 professional," 2015. available at https://www.abbyy.com/finereader/professional/.

[36] OCRopus, Jan. 2015. available at https://github.com/tmbdev/ocropy.

[37] M. Z. Afzal, M. Kramer, S. S. Bukhari, M. R. Yousefi, F. Shafait, and T. M. Breuel, "Robust Binarization of Stereo and Monocular Document Images Using Percentile Filter," in *Camera-Based Document Analysis and Recognition*, vol. 8357, pp. 139–149, Aug. 2013.