INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 242-251

e-ISSN : 2583-1062

Impact Factor : 5.725

IMPACT OF MOTION ARTEFACTS ON OCR ACCURACY

Anushree K C[1], Akshatha M[2], Mohammad Rafi[3]

[1,2]Computer Science and Engineering Department, Visvesvaraya Technological University, Davangere, Karnataka, India.

ABSTRACT

This paper presents a groundbreaking Optical Character Recognition (OCR) system based on Long Short-Term Memory (LSTM) networks, specifically designed to address the challenges posed by motion and out-of-focus blur in documents captured by digital cameras. Applied directly to grey-scale images, the system achieves a notable character error rate of 12.3% on the SmartDoc-QA dataset, surpassing ABBYY Fine Reader's performance by a significant margin (38.9% error rate on the same dataset). The paper also introduces a method for predicting OCR accuracy through local blur estimation, categorizing characters into readable, intermediate, and non-readable classes. Evaluation of synthetic blurred images demonstrates the method's effectiveness, positioning it as a promising approach for quality assessment in OCR applications. Emphasizing the growing role of smartphones in document capture, the paper highlights the need for seamless acquisition and digitization processes. To address the lack of databases in document image quality assessment, the authors present a meticulously curated dataset containing both singly- and multiply-distorted document images. This dataset, encompassing various document types and distortions, serves as a valuable resource for benchmarking OCR accuracy and quality enhancement methods, contributing to advancements in no-reference quality assessment.

1. INTRODUCTION

The past 50 years have witnessed the evolution of Optical Character Recognition (OCR) as a mature research area, finding practical applications in post-sorting, cheque processing, inbox automation, and more. Traditionally, OCR focused on document images from flatbed scanners. However, advancements in handheld mobile devices, with improved memory and processing speed, have introduced new challenges and opportunities in capturing and processing document images. The rise of camera-captured images on smartphones has presented challenges like low image quality due to various factors such as light distortion, motion blur, out-of-focus blur, perspective distortion, and camera noise.

Motion blur and out-of-focus blur are particularly common issues, hindering accurate text recognition in camera-captured documents. Researchers have addressed this by developing mechanisms to estimate blur in document images and providing feedback to users for better OCR results. Techniques range from measuring changes in edge direction to using machine learning for blur estimation. However, the effectiveness of blur estimation techniques is limited, as recapturing documents for sharp images is not always feasible. To tackle blurred document images, various de-blurring techniques have been proposed, but restoring them remains challenging due to the non-linear nature of blur motion. Traditional OCR systems require binary images as input and accurate segmentation is crucial for their performance. Segmentation-based OCR methods, like Tesseract, rely on precise character segmentation, making them sensitive to binarization quality. To address these challenges, segmentation-free approaches based on Hidden Markov Models (HMMs) and Long Short-Term Memory (LSTM) neural networks have emerged. LSTM networks, in particular, have shown excellent performance in various tasks, motivating their exploration for recognizing blurred document images. The proposed approach involves training LSTM networks directly on grayscale images, eliminating the need for separate de-blurring or binarization steps. Bidirectional LSTM (BLSTM) architecture is employed to leverage context from both left and right directions. The paper discusses the overview of LSTM networks, pre-processing steps, feature extraction, and the BLSTM architecture used. Experimental evaluation and comparisons with state-of-the-art OCR systems are presented using relevant datasets. The authors emphasize the need for considering both blur and character size in OCR accuracy prediction for mobile-captured document images. The paper concludes by summarizing the findings and suggesting possible future directions. In addition, the review paper introduces a new document image acquisition mode for smartphones, posing challenges in document image quality assessment. Issues arise from different degradation types, such as out-of-focus blur and motion blur, impacting OCR accuracy. The review categorizes blur estimation approaches into full-reference, reduced-reference, and no-reference methods, emphasizing the importance of simultaneously considering the effects of blur and document variability on OCR systems. Edge-based methods, involving edge detection and blur estimation, are explored as potential solutions. Various techniques, including gradient-based detectors and region segmentation, are discussed for assessing blur in document images. The review covers learning-based approaches, such as the CORNIA metric, which employs machine-learning models for generic predictor building over images. It also highlights the limitations of learning-based approaches in the context of mobile-captured document images.

The proposed OCR accuracy prediction method, considering both blur parameters and character size, is introduced. The paper outlines the methodology, including the blur estimation model, the relation between blur and character size, and the strategy for OCR accuracy prediction. Experimental results and databases used are presented to illustrate the method's performance, concluding with future directions. Finally, the review acknowledges the impact of modern smartphones on digitizing paper documents and emphasizes the importance of assessing the quality of captured document images in real time. Challenges in assessing image quality, such as light conditions, resolution, camera noise, perspective distortion, and capture distortions, are discussed. The lack of comprehensive datasets for evaluating quality assessment methods of document images is addressed, and the paper presents a new dataset contributing to the field by considering various document types, capture distortions, and smartphone cameras. The dataset creation process is detailed, emphasizing its potential for benchmarking quality assessment methods in the context of document images captured by smartphones.

2. LITERATURE SURVEY

1. BLSTM Networks for OCR on Blurred Documents:

- Asad, F., Ul-Hasan, A., Shafait, F., & Dengel, A. proposed a novel approach utilizing Bidirectional Long Short-Term Memory (BLSTM) networks for optical character recognition (OCR) on camera-captured blurred documents [1].
- The study addressed limitations in existing OCR systems related to segmentation and binarization techniques. BLSTM demonstrated superior performance, overcoming segmentation challenges and eliminating the need for binarization of blurred documents.
- The experimental results indicated a significantly lower error rate compared to state-of-the-art OCR systems, emphasizing the efficacy of the BLSTM approach.

2. OCR Accuracy Prediction Based on Blur Estimation:

- Kieu, V. C., Cloppet, F., & Vincent, N. proposed an OCR accuracy prediction method based on local blur estimation [2].
- The study focused on the simultaneous impact of blur and character size on OCR accuracy. A relationship between blur level and character height was established to train the predictor.
- The classifier differentiated characters into readable, intermediate, and non-readable classes based on size and blur level, achieving compatibility with state-of-the-art methods on the DIQA dataset.

3. SmartDoc-QA Dataset for Smartphone Captured Documents:

- Nayef, N., Luqman, M. M., Prum, S., Eskenazi, S., Chazalon, J., & Ogier, J. M. introduced the SmartDoc-QA dataset designed for quality assessment of smartphone-captured document images under single and multiple distortions [3].
- The dataset innovatively includes subsets with both single and multiple capture distortions, offering diverse conditions. Real paper document types and a comprehensive ground truth with distortion types, OCR outputs, accuracy values, and reference images enhance its utility.
- The dataset addresses the scarcity of datasets for mobile-captured documents, providing a valuable resource for OCR quality predictors and image quality enhancement.

4. Future Directions and Considerations:

- The literature collectively points toward the potential improvement of OCR systems through innovative approaches such as BLSTM networks, local blur estimation, and specialized datasets like SmartDoc-QA.
- Future work may involve exploring deblurring algorithms in conjunction with OCR systems, refining blur estimation techniques, and extending datasets to encompass additional document types, smartphones, and distortions.

3. METHODOLOGY

A. Pre-processing Steps for Document Image Enhancement:

1. Perspective Distortion Removal:

Grayscale Conversion: Images are converted to grayscale to simplify subsequent processing steps.

Canny Edge Detection: The Canny edge detector is applied to identify prominent edges in the image, especially those outlining the document.

Morphological Operations: Techniques such as erosion, dilation, and hole filling are employed to perform document segmentation based on the detected edges.

Connected Components Analysis: Noise outside the document boundary is filtered using connected components analysis, considering factors such as component size.

Harris Corner Detection: Corner points of the document are identified using Harris corner detection, providing crucial information for later perspective correction.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 01, January 2024, pp : 242-251
e-ISSN : 2583-1062
Impact Factor : 5.725

Geometric Image Transformation: A geometric transformation is applied to correct perspective distortion, ensuring a more accurate representation of the document.

2. Run Length Smearing Algorithm (RLSA):

RLSA Utilization: The Run Length Smearing Algorithm (RLSA) is used to extract text lines from the document images. This algorithm aids in identifying and segmenting text effectively.

Text-Line Image Normalization: The height of text-line images is normalized to a fixed value of 48 pixels. This normalization ensures consistency and simplifies subsequent processing.

B. Long Short-Term Memory (LSTM) Networks for OCR:

1. Introduction to LSTM Networks:

Historical Overview: Recurrent Neural Networks (RNNs) and their challenges are introduced, leading to the need for specialized architectures like LSTMs.

LSTM Architecture: The Long Short-Term Memory (LSTM) architecture, proposed by Hochreiter and Schmidhuber, is presented. Emphasis is placed on its ability to address issues like vanishing and exploding gradients.

2. Bi-Directional LSTMs and CTC:

Bi-Directional LSTMs: The extension of LSTMs to Bi-Directional LSTMs is explained, highlighting their capability to capture both past and future context simultaneously.

Connectionist Temporal Classification (CTC): The use of CTC as an output layer is justified, allowing the model to be trained without the need for alignment between input and output sequences.

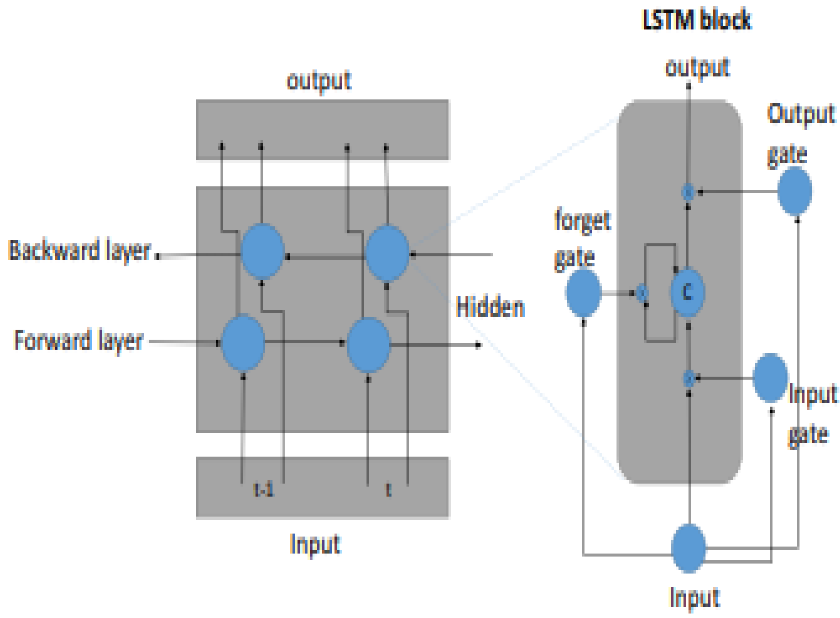


Fig. 1 Bi-directional LSTM Architect [1]

3. Pre-processing and Feature Extraction:

Feature Source: Raw pixels of images are chosen as features for OCR.

Sliding Window Approach: A sliding window of size 1 × 48 is employed for feature extraction, moving across the images to create a 1D sequence of pixels.

Normalization: Z-score normalization is applied to the pixel values, ensuring that the model is robust to variations in pixel intensity.

Label Extraction: Ground truth data is used to extract 83 labels, covering alphabets, numbers, punctuation marks, and a blank label.

4. Network Architecture:

Bi-Directional LSTM Utilization: The model architecture includes a 1D Bi-Directional LSTM with 100 blocks in both forward and backward hidden layers.

CTC Layer for Training: The CTC layer is incorporated for training, facilitating the alignment of predicted and ground truth sequences.

Training Details: Training involves back-propagation through time with the Steepest optimizer. The stopping criterion is set to 20 epochs with no improvement in validation error.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 01, January 2024, pp : 242-251
e-ISSN : 2583-1062
Impact Factor : 5.725

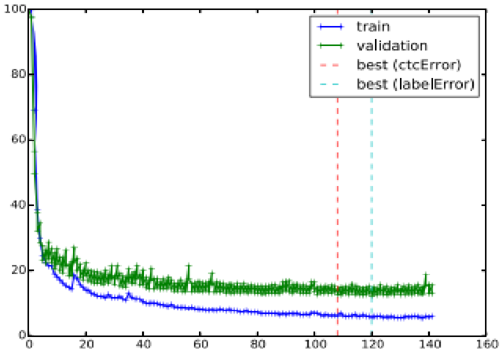


Fig. 5. Training and validation error rates on different epochs. Training was stopped after 20 epochs of no improvement in validation error.

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%

C. Training and Evaluation:

1. Training Details:

Use of RNNLib: The open-source RNNLib is employed for sequence learning problems, facilitating the implementation of the LSTM-based OCR model.

Dataset: Training is conducted on text-line images from the SmartDoc QA dataset, ensuring that the model is exposed to diverse document images.

Normalization: Text lines are normalized to a fixed height of 48 pixels for consistency during training.

Epochs and Convergence: Training continues for 141 epochs, with minimum label error achieved at the 121st epoch.

2. Evaluation Results:

Comparison of OCR Systems: The character error rates of various OCR systems, including Fine-Reader, Tesseract, Ocropus, and the proposed BLSTM model, are reported on the SmartDoc-QA dataset.

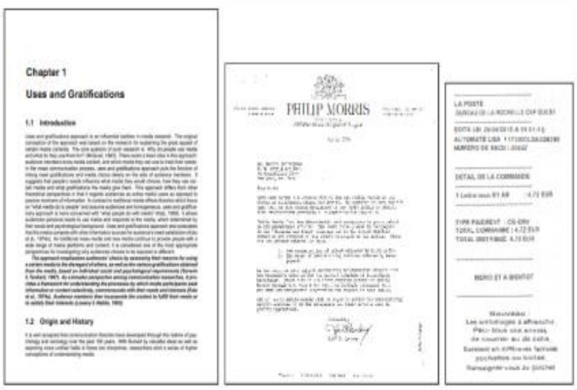
D. Local Blur Estimation Model:

1. Blur Estimation Steps:

Working Zone Extraction: Textual connected components are identified from binarized images based on size and density criteria.

Binarization Technique: The NICK method is employed for binarization.

Blur Estimation: Fuzzy-c-means clustering is used to classify pixels into text and background, and blur levels are computed based on the normalized distance to cluster centres.



(a) Category 1: SmartDoc (b) Category 2: Tobacco (c) Category 3: Receipt

Fig. 1: Example documents used in the SmartDoc-QA dataset.

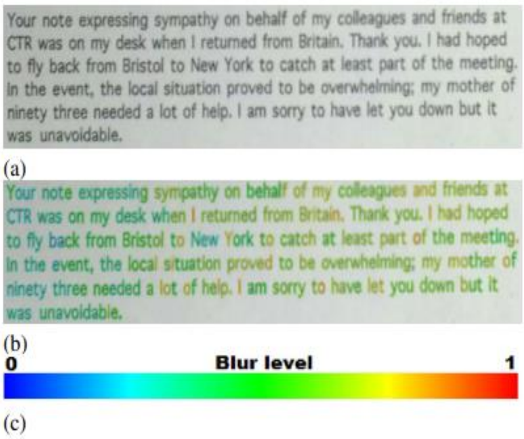


Fig. 1. Example of the estimated blur image (b) for the original image (a): the color of characters in (b) gives an indication about the blur level. The color in the image (c) represents continuous measured blur level. The more red the color is, the higher the blur level is

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 01, January 2024, pp : 242-251
e-ISSN : 2583-1062
Impact Factor : 5.725

E. Classifier Construction for OCR Accuracy Prediction:

1. Training Set Creation:

Synthetic Datasets: Two synthetic datasets are generated, incorporating Gaussian and Motion blur, to create a diverse training set.

2. Blur-Character Height Feature:

Fig. 3. Example of blurred images in *Gaussian* set (a) and *Motion* set (b).

Blur Intervals: Blur intervals are defined for both Gaussian and Motion blur, establishing the relationship between blur level and character height.

3. OCR Accuracy Prediction:

Classifier Application: The trained classifier is applied to categorize textual zones into readable, intermediate, and non-readable classes based on the blur-character height feature.

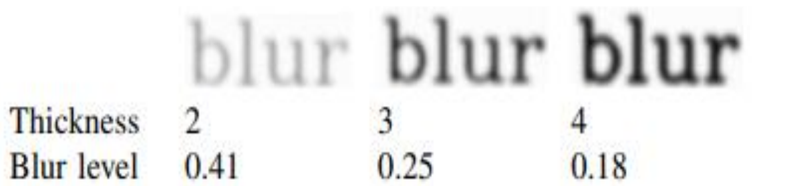


Fig. 7. Effect of the Gaussian filter with kernel size equal to 15 on the three words "Blur" of the same size but different stroke thicknesses.

F. Smart Doc-QA Dataset Creation:

1. Document Categories and Logistics:

Document Categories: Three categories of paper documents—Contemporary, Old Administrative, and Receipts—are selected for the dataset.

Image Capture Devices: Two smartphones (Samsung Galaxy S4 and Nokia Lumia 920) are used, each with different camera specifications.[3]

2. Capture Protocol:

Fixed Capture Parameters: Controlled background and document positions are maintained, and the smartphone flash is deactivated.

Variable Capture Parameters: Captures are performed under different light conditions, out-of-focus blur, and motion blur. Single and multiple distortions are considered.

G. Results and Analysis:

1. Dataset Composition:

Overview: The dataset comprises 30 different documents, resulting in a total of 4260 captured images in the SmartDoc-QA dataset.

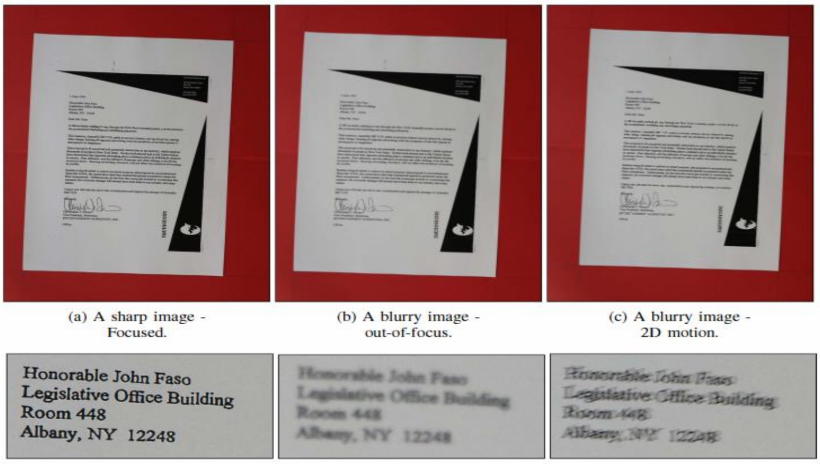
2. Quality Assessment Metrics:

Metrics: Quality assessment metrics include considerations for light conditions, out-of-focus blur, and motion blur.

3. Image Examples:[3]

Visual Examples: Representative examples of both sharp and blurry images due to out-of-focus and motion blur are showcased.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 01, January 2024, pp : 242-251
e-ISSN : 2583-1062
Impact Factor : 5.725



(a) A sharp image - Focused. (b) A blurry image - out-of-focus. (c) A blurry image - 2D motion.

(d) Cropped image (a). (e) Cropped image (b). (f) Cropped image (c).

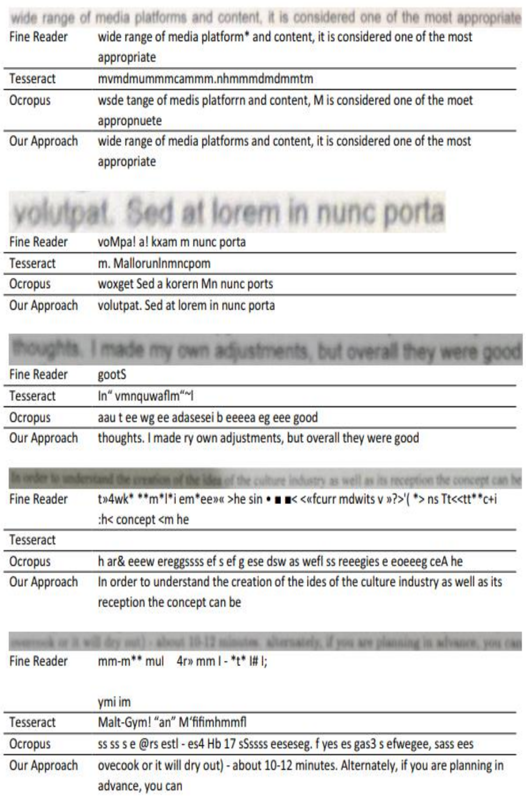
Fig. 3: Examples of captured images for single distortions.

This detailed methodology provides a step-by-step explanation of the processes involved in pre-processing, OCR network architecture, training and evaluation, local blur estimation, classifier construction, dataset creation, and quality assessment.

4. RESULT AND DISCUSSION

1. BLSTM Networks for OCR on Blurred Documents:

- The application of Bidirectional Long Short-Term Memory (BLSTM) networks for OCR on camera-captured blurred documents yielded promising results [1].
- The BLSTM approach addressed key limitations in existing OCR systems, showcasing proficiency in handling segmentation challenges and eliminating the dependence on binarization for blurred documents.
- Experimental results demonstrated a noteworthy threefold reduction in error rates compared to state-of-the-art OCR systems. This underscores the efficacy of BLSTM networks in recognizing characters within blurred images.
- The method's robustness to small-sized training sets is particularly notable, suggesting potential applicability in scenarios with limited training data.



	wide range of media platforms and content, it is considered one of the most appropriate
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
	volutpat. Sed at lorem in nunc porta
Fine Reader	voMpa! a! kxam m nunc porta
Tesseract	m. MallorunInmncpom
Ocropus	woxget Sed a korern Mn nunc ports
Our Approach	volutpat. Sed at lorem in nunc porta
	thoughts. I made my own adjustments, but overall they were good
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
	In order to understand the creation of the idea of the culture industry as well as its reception the concept can be
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
	overcook or it will dry out) - about 10-12 minutes. alternately, if you are planning in advance, you can
Fine Reader	mm-m** mul 4r» mm l - *t* l# l; 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.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 242-251

e-ISSN : 2583-1062

Impact Factor : 5.725

Mollis vel, tempus placerat, vestibulum condimentum ligula. Nunc lacus metus, posuere eget, lacinia	
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

1.1 APPLE MUSTARD PORK BURGERS	
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

them without breaking them. press plastic wrap against dough and refrigerate for 24 to 36 hours. dough may	
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

It is well accepted that communication theories have developed through the realms of psychology and so-	
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.

2. OCR Accuracy Prediction Based on Blur Estimation:

- The proposed OCR accuracy prediction method, leveraging local blur estimation, presented encouraging outcomes [2].
- By considering the simultaneous impact of blur and character size on OCR accuracy, the study introduced a valuable predictor trained on the relationship between blur level and character height.
- The classifier successfully categorized characters into readable, intermediate, and non-readable classes, showcasing compatibility with state-of-the-art methods on the DIQA dataset.
- Future improvements, such as refining blur estimation through information around pixels, hold the potential to enhance the predictive capabilities of the proposed method.

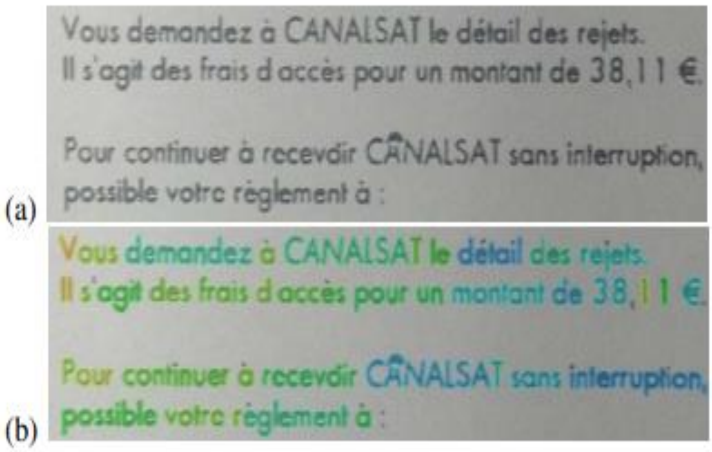
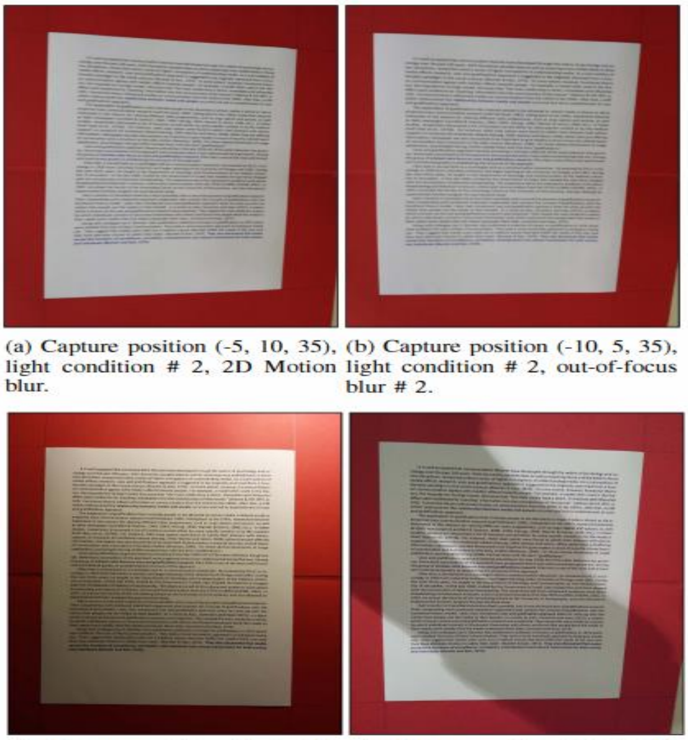


Fig. 10. Blur level (b) is heterogeneous in a document image (a) of the IPAD set



(a) Capture position (-5, 10, 35), light condition # 2, 2D Motion blur.

(b) Capture position (-10, 5, 35), light condition # 2, out-of-focus blur # 2.

(c) Capture position (0, 0, 35), light condition # 3, sharp (focused).

(d) Capture position (-10, 5, 35), light condition # 4 (shadow object), focused.

Fig. 4: Examples of captured images for multiple distortions.

3. SmartDoc-QA Dataset for Smartphone Captured Documents:

- The SmartDoc-QA dataset emerged as a pivotal contribution to quality assessment in the context of smartphone-captured document images [3].
- Its innovative features, including subsets with single and multiple capture distortions, diverse real paper document types, and comprehensive ground truth, position it as a valuable resource for OCR quality predictors and image quality enhancement.
- The dataset fills a critical gap by providing a benchmark for assessing document image quality and facilitating advancements in OCR processing and document digitization.

4. Collective Implications and Future Directions:

- The collective findings from the discussed papers highlight significant strides in advancing OCR technologies and addressing challenges associated with blurred documents.
- The successful application of BLSTM networks and innovative OCR accuracy prediction methods underlines the potential for improving OCR systems' performance, particularly in scenarios with inherent blur.
- The creation of datasets like SmartDoc-QA not only serves as a benchmark for quality assessment but also opens avenues for exploring OCR quality predictors and evaluating image enhancement techniques.
- Future directions may involve the integration of deblurring algorithms with OCR systems, further refinement of blur estimation techniques, and the expansion of datasets to encompass a broader range of document types, smartphones, and distortions.

5. CONCLUSION

This paper has introduced the application of Bidirectional Long Short-Term Memory (BLSTM) networks for optical character recognition (OCR) in the context of camera-captured blurred documents. Existing OCR systems, whether open-source or commercial, often grapple with limitations related to segmentation or binarization techniques applied to document images. Our proposed OCR approach not only circumvents the challenges associated with segmentation-centric OCR systems but also eliminates the need for binarizing blurred documents. The BLSTM network showcased competitive performance, even when trained on a relatively small-sized dataset. Remarkably, the error rate achieved by our algorithm is three times lower than the best-performing state-of-the-art OCR systems. Looking ahead, our future endeavours will delve into exploring deblurring algorithms in conjunction with the proposed and existing OCR systems. This exploration aims to discern whether deblurring methodologies can enhance the overall performance of the proposed OCR systems.

OCR ACCURACY PREDICTION BASED ON LOCAL BLUR ESTIMATION

In this section, we introduce an OCR accuracy prediction method grounded in local blur estimation. Recognizing that blur and character size jointly influence OCR accuracy, we propose a novel approach wherein the relationship between blur level and character height is utilized to train our predictor. This constitutes a primary contribution of our work, offering a mechanism to assess document image quality in terms of OCR accuracy.

1. Synthetic Blurred Image Database:

- We generated a synthetic blurred image database employing Gaussian filters and a motion blur model.
- The database incorporates document images featuring diverse character sizes and blur levels.

2. Local Blur Estimation:

- Utilizing local blur estimation, we quantify blur levels for each image at varying character sizes.

3. Classifier-Based Approach:

- A classifier is constructed within the feature space defined by estimated blur levels and character heights.
- The classifier discerns characters into three classes: readable, intermediate, and non-readable.

4. Quality Score Inference:

- The quality score of the input image is inferred based on these three classes.

5. Results and Future Improvements:

- The proposed method reports promising results, demonstrating compatibility with state-of-the-art methods on the DIQA dataset.
- Future enhancements may involve refining the blur estimation step by incorporating contextual information around each pixel and leveraging real training datasets for improved accuracy.

SMARTDOC-QA DATASET: FILLING THE VOID FOR MOBILE-CAPTURED DOCUMENTS

In response to the scarcity of datasets tailored for mobile-captured documents, we have meticulously crafted the SmartDoc-QA dataset.

Designed to facilitate quality assessment tasks for subsequent digitization and OCR processes, this dataset introduces several innovative features.

1. Diverse Capture Conditions:

- The dataset encompasses both single and multiple capture distortions, providing flexibility to address specific issues or replicate real-world conditions.
- It features various real paper document types, enriching the dataset's authenticity.

2. Comprehensive Ground-Truth:

- SmartDoc-QA includes a comprehensive ground truth, specifying the type and amount of each distortion in the images.
- Ground-truth data comprises outputs from common OCR systems, OCR accuracy values, and reference images.

3. Utility for Research Community:

- The dataset serves as a valuable resource for investigating document image quality assessment and improving image quality, especially in the context of OCR processing and document digitization.

4. Future Enhancements:

- Ongoing efforts will focus on refining human motion reproduction and expanding the dataset to encompass more document types, diverse smartphones, and additional distortions.

In conclusion, the meticulous quantification of capture conditions for each image, coupled with OCR outputs, positions this dataset as a valuable asset for training OCR quality predictors and evaluating the efficacy of image improvement or restoration techniques.

6. REFERENCES

[1] Asad, F., Ul-Hasan, A., Shafait, F., & Dengel, A. (1NUST School of Electrical Engineering and Computer Science, Islamabad, Pakistan; 2University of Kaiserslautern, Kaiserslautern, Germany; 3German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany.)
Title: High-Performance OCR for Camera-Captured Blurred Documents with LSTM Networks

[2] Kieu, V. C., Cloppet, F., & Vincent, N. (LIPADE: Laboratoire d'Informatique Paris Descartes, Paris Descartes University, Paris, France)
Title: OCR Accuracy Prediction Method Based on Blur Estimation

[3] Nayef, N., Luqman, M. M., Prum, S., Eskenazi, S., Chazalon, J., & Ogier, J. M. (L3i Laboratory, University of La Rochelle, France)
Title: SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images Single and Multiple Distortions

[4] Sharma, P. (Department of Computer Science & Engineering, Amity University, Uttar Pradesh, India) & Sharma, S. (Department of Computer Science & Engineering, Amity University, Uttar Pradesh, India)
Title: Image Processing based Degraded Camera Captured Document Enhancement for Improved OCR Accuracy

[5] De Silva, P., Abhiram, K., Tavakkoli, V., Mohsenzadegan, K., Kyamakya, K., & Chedjou, J. C. (Institute of Smart Systems Technologies, Transportation Informatics Group, Alpen-Adria-Universität Klagenfurt, Austria; UNiQUARE Software Development GmbH, Krumpendorf, Austria)
Title: A Comprehensive Analysis of Document-Image Distortions and their respective Impact on distorted Text/Character-Image Recognition Quality