SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images - Single and Multiple Distortions

Nibal Nayef, Muhammad Muzzamil Luqman, Sophea Prum,
Sebastien Eskenazi, Joseph Chazalon and Jean-Marc Ogier
L3i Laboratory, University of La Rochelle, France
nibal.nayef@univ-lr.fr

***Abstract*—Smartphones are enabling new ways of capture, hence arises the need for seamless and reliable acquisition and digitization of documents. The quality assessment step is an important part of both the acquisition and the digitization processes. Assessing document quality could aid users during the capture process or help improve image enhancement methods after a document has been captured. Current state-of-the-art works lack databases in the field of document image quality assessment. In order to provide a baseline benchmark for quality assessment methods for mobile captured documents, we present in this paper a dataset for quality assessment that contains both singly- and multiply-distorted document images.**

The proposed dataset could be used for benchmarking quality assessment methods by the objective measure of OCR accuracy, and could be also used to benchmark quality enhancement methods. There are three types of documents in the dataset: modern documents, old administrative letters and receipts. The document images of the dataset are captured under varying capture conditions (light, different types of blur and perspective angles). This causes geometric and photometric distortions that hinder the OCR process. The ground truth of the dataset images consists of the text transcriptions of the documents, the OCR results of the captured documents and the values of the different capture parameters used for each image. We also present how the dataset could be used for evaluation in the field of no-reference quality assessment. The dataset is freely and publicly available for use by the research community at http://navidomass.univ-lr.fr/SmartDoc-QA.

Keywords*—*Quality Assessment Dataset, Document Image Quality, Capture-based Distortions, Predicting OCR Accuracy

I. Introduction and Related Work

Modern smartphones have had a revolutionary impact on the way people digitize paper documents. The goal of digitizing paper documents is not only to archive them for sharing but also to process them by automated document processing systems. The latter extract the content of the document images for recognizing it, indexing it, verifying it, matching it against a database etc. However, it is a known fact that the cameras of the smartphones are optimized for capturing natural scene images. Taking a simple photo of a paper document does not ensure that its content would be exploitable by automated document image processing systems. This could happen because of the light conditions, the resolution of the image, the camera noise, the perspective distortion, the physical distortions of the paper (folds etc.), the out-of-focus blur and/or the motion blur during capture. To ensure that a captured document image is exploitable by automated systems, it is important to automatically assess the quality of a captured document image in real-time. In many cases, it is not possible to re-capture a document, because the original paper document is not available anymore. Assessing the quality of a captured document image is also required as a step preceding quality enhancement methods.

The research community working on quality assessment of natural scene images have created standard datasets for evaluating and benchmarking their methods [1][2][3]. However, there is still a lack of datasets for evaluating quality assessment methods of document images. To the best of our knowledge, the only available dataset for quality assessment of smartphone captured document images was proposed in 2013 by Kumar et al. [4]. This dataset deals with only one type of documents, one type of capture distortions (out-of-focus blur) and the images are captured using one smartphone. The dataset of Kumar et al. [4] has 29 different documents used to capture 375 images with varying degrees of out-of-focus blur.

In the field of document image quality assessment [5], the dataset proposed in this paper makes the following contributions over state-of-the-art:

- Using 3 different real-world paper document types
- Considering multiple capture distortions (light conditions, motion blur, out-of-focus blur, perspective distortions)
- Considering both single and multiple distortions
- Using multiple smartphone cameras
- Building a reproducible and semi-automatic capture process which could be used to create future datasets

The quality of a document image depends on the degradation (or distortions) present in the image. Degradation in document images results from imaging conditions, imaging device, or from poor quality of paper, the printing process, ink blot or fading, document aging, extraneous marks, scanning noise, etc. Document image distortions can be categorized as follows:

- Scene-related: resulting from capture conditions such as light, motion blur, out-of-focus blur, resolution (from the aspect of the distance between the camera and the document), scene background, position of the camera with respect to the document etc.

978-1-4799-1805-8/15/$31.00 ©2015 IEEE

- Device-related: camera noise, resolution etc.
- Document-related: folding, age, stains, copying noise (from fax or printer), warping etc.

In our dataset we mainly consider scene-related distortions such as light conditions, blur and perspective distortions. However, we use a very simplified background in order to minimize the effect of background on the automatic processing of documents. As for device distortions, we use different smartphones whose specifications are given in the ground truth. The document-related distortions are partially addressed by using different types of paper documents.

Throughout the sections of this paper, we will explain the details of the creation process of the proposed dataset, and discuss how such a dataset could be used for benchmarking quality assessment methods.

II. The "SmartDoc-QA" Dataset

A. Documents

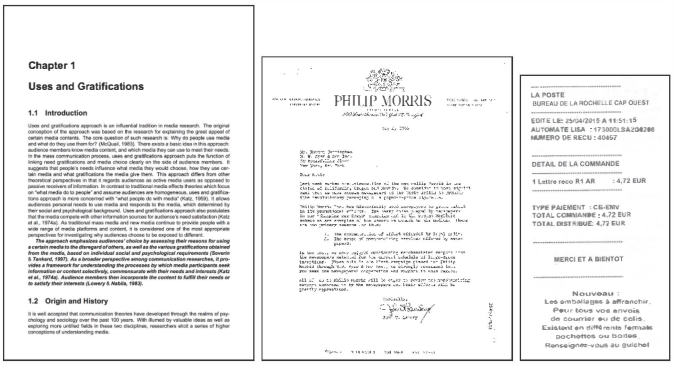
In order to cover different contexts of application and make our proposed dataset useful in real world commercial applications, we use the following three different categories of paper documents to create our dataset:

- Category 1: Contemporary documents (selected from SmartDoc competition dataset /challenge 2 [6]): This set contains 10 text-only documents. The document contents are generated from real text from wiki-books and cooking recipes from the Internet.
- Category 2: Old administrative documents (selected from the Tobacco dataset [7]): This set contains 10 documents. Those selected documents are relatively clean and readable, and they contain little salt and pepper noise. Some of the documents in this set contain small zones of image and handwriting.
- Category 3: Receipts: This set contains 10 real receipts from various shops. These receipts are relatively clean without folding traces. Contrary to the two other sets, this receipt set is in French language.

In total, there are 30 different documents used to capture 4260 images of our dataset. Figure II-A shows examples of those paper documents of the three document categories. It is worth mentioning that the documents have been chosen with simple layout in order to minimize the effect of the document layout complexity on OCR results. Hence, the difference between the OCR results of different documents could be majorly attributed to the quality of the captured document image. Therefore, only one-column text and clean documents are used. Some documents from the Tobacco dataset (category 2) contain small images of logos, which might be challenging for some OCR systems. However, based on our experiments, these images do not have a big impact on OCR results.

B. Logistics

The document images for the dataset are captured in a precisely controlled and repeatable environment. In order to simulate the possible light conditions that we face in real life scenarios, we performed image capture in a room with fully controlled light conditions. In order to precisely simulate the possible capture positions and the motion blur during the captures, we have employed a *Fanuc LR Mate 200iD* robotic arm. To diversify the cameras for the captures, we have employed two modern smartphones, *Samsung Galaxy S4* and *Nokia Lumia 920*, whose cameras are based on different sensor technologies and they capture images at different resolutions.



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

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

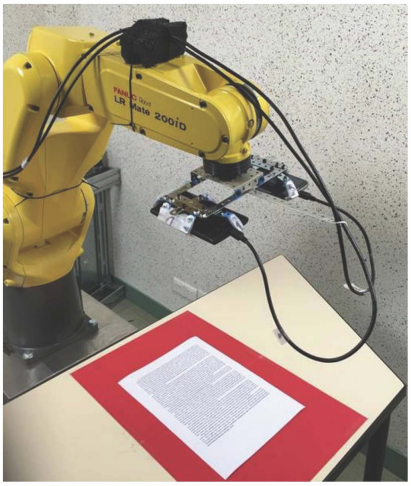


Fig. 2: The robotic arm holding smartphones over a document on a fixed simple background.

Fig.2 presents a photo of the robotic arm holding the smartphones over a document to be captured. The robotic arm and the smartphones are programmed to communicate to and to be controlled by – in real-time – a server program running on a computer. The precise position of the smartphone camera with respect to the center of the document is managed by the server program which in turn controls the robotic arm. Once a desired position and other capture conditions are achieved, the server program communicates with the smartphone API to trigger the capture. The only manual steps are changing the paper document, and changing light conditions. The overall sequence of both automatic and manual steps of the capture process is managed and synchronized by the server program.

C. Capture Protocol

The dataset has a total of 4260 document images captured from 30 different paper documents. 142 different images are captured per document (71 captures per phone), those captures are taken using representative values of different distortions (see capture parameters below). For each image, the information about the document and about capture conditions is stored as ground truth for evaluation purposes.

1) Fixed Capture Parameters:

- Background: one colored, clear contrast with the paper documents in order to minimize the effect of the page segmentation process
- Document: is completely inside the image with a fixed orientation
- Smartphone flash: always deactivated

Fig.2 shows the fixed capture settings related to the documents and the background.

2) Variable Capture Parameters:

- Smartphone camera: 2 smartphones
- Light: 5 light conditions
- Out-of-focus blur: 4 values
- Motion blur: 2 types
- Position of the smartphone with respect to the paper document (5 positions):
    - Perspective 1: Longitudinal incidence angle (mobile rotation around Y-axis).
    - Perspective 2: Lateral incidence angle (mobile rotation around X-axis).
    - Distance between the camera and the document: 1 value

In the following, we explain our protocol in capturing images of the documents using specific values of the variable capture parameters mentioned above. The captures are divided into two categories: singly- and and multiply-distorted document images. The multiply-distorted images are the ones encountered in real life. However, having singly-distorted images is very useful for developing and testing quality assessment methods. For example, testing the effect of certain distortions on image quality, or for developing methods which combine assessment metrics where each metric is capable of assessing only one type of distortions. For both distortion categories, every capture is taken with the following two smartphones:

- Samsung Galaxy S4 (camera: 13MP)
- Nokia Lumia 920 (camera: 8.7MP)

3) Single Distortions: We consider light, out-of-focus blur and motion blur distortions separately. All the captures for single distortions are taken at a position where the perpendicular distance between the camera and the center of the document is 35cm and both the longitudinal incidence angle and lateral incidence angle equal zero. We call this the parallel position where perspective distortion is minimal. Under these conditions, we execute the capture process as follows:

For light distortions, five images are captured for each document under each of the following light conditions:

- Light condition 1: Day light only (without any artificial lights)
- Light condition 2: Day light + ceiling neon light
- Light condition 3: Night + table lamp light
- Light condition 4: Table lamp light + an object casting a shadow on a large part of the document
- Light condition 5: Table lamp light + an object casting a grid shadow on the document

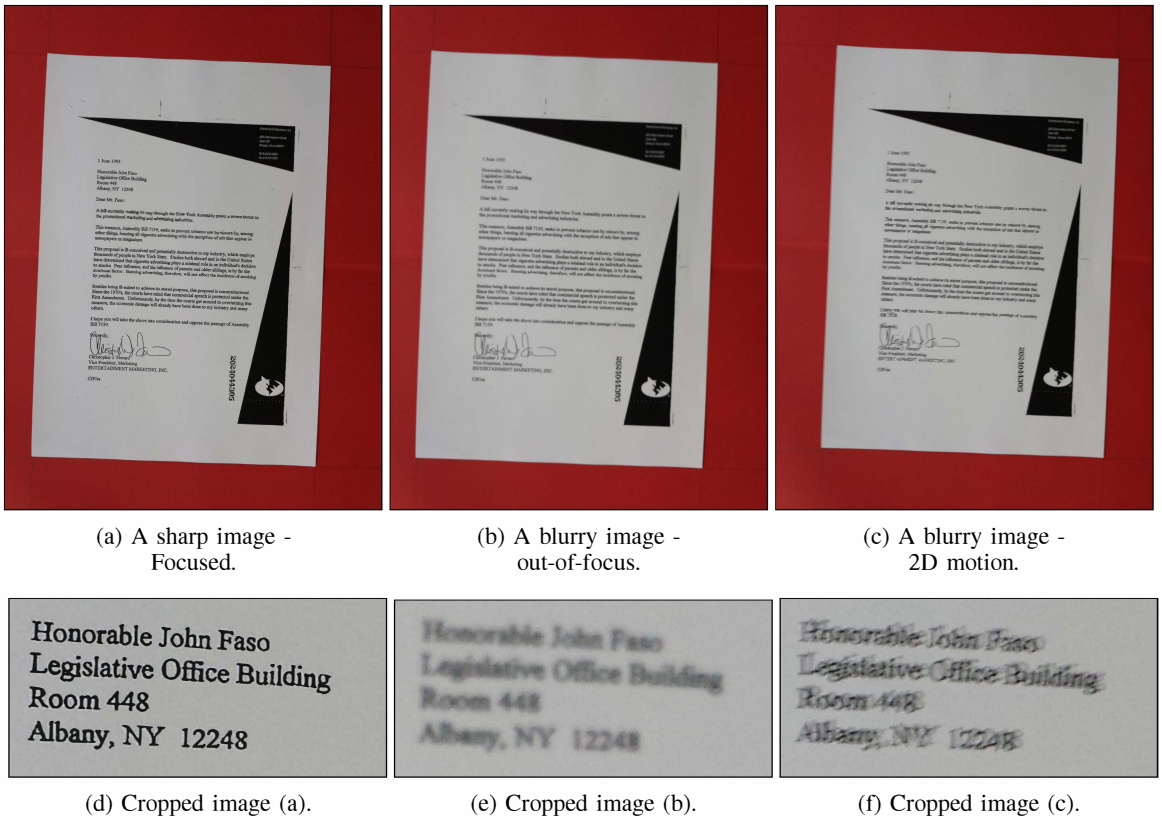
For out-of-focus blur, all the captures are also taken at the parallel position described above, with the light condition fixed to "day light + ceiling neon light". The smartphone camera is forced to focus on the document at a distance shorter than 35cm (22cm), then a capture is taken at the position with 35cm distance. This results in an image that has out-of-focus blur. This is repeated four times to capture four images at varying degrees of out-of-focus blur, where each time the focus point is 1cm closer to the document, while the capture is always taken at 35cm.

For motion blur, we use the same camera position and light condition as the out-of-focus blur captures. The motion blur is simulated as follows. While the robot arm is moving according to a specific speed and direction, the camera is triggered to take the capture. We use two motions, a horizontal one and a 2D one. The capture could happen at any instant during the motion, hence producing different degrees of severity of blur present in the images. Two captures are taken this way for each document.

Figure 3 shows three different captures of the same document taken at the second light condition and the parallel position. The first image shows a focused (sharp) capture, the second shows a blurry image due to out-of-focus blur and the third shows a blurry image due to motion blur. The complete process – with a total of 11 captures – is repeated using each smartphone camera, where we ensure that only one type of blur distortion is present in an image.

4) Multiple Distortions: Here we consider a combination of different capture conditions for each image. We have selected 3 light conditions, 5 camera positions and 3 blur values. Additionally, we capture a reference image that is not blurry (sharp / focused) at each combination of light condition and position. This creates 60 captures per document.

For the light conditions, we have chosen conditions 2, 3 and 4. The position of the camera with respect to the document is expressed as (perspective 1, perspective 2, distance), where those variables are as mentioned above in the variable capture parameters. The following five positions are considered: (0°, 0°, 35cm), (-10°, -5°, 35cm), (-10°, 5°, 35cm), (-5°, 10°, 35cm) and (5°, 10°, 35cm). For the three blur values, two of the out-of-focus blur values mentioned above are selected, and the 2D motion blur.



(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.

III. Image Quality Assessment using the "SmartDoc-QA" Dataset

Our proposed dataset is to be mainly used for benchmarking no-reference quality assessment (NR-QA) methods. Such methods aim at computing an image quality score that best correlates with either human perceived image quality or an objective quality measure, without any prior knowledge of reference images. In our dataset we focus on the objective quality measure of OCR. Hence, the dataset can be used to judge the ability of an NR-QA method to predict the OCR accuracy of a document. The reader is referred to the works in [4] and [5] for more details.

The advantage of predicting OCR accuracy of a given document image, or more generally, predicting the performance of a given task on a given input, permits us to adjust the final decision about how to handle a given image (recapture it, ask to improve it, reject it, ask a human to manually process it, adapt its automatic processing, etc.).

In the following subsections, we describe the ground truth of our dataset, and how a quality assessment method can be evaluated using state-of-the-art evaluation metrics.

A. Ground Truth

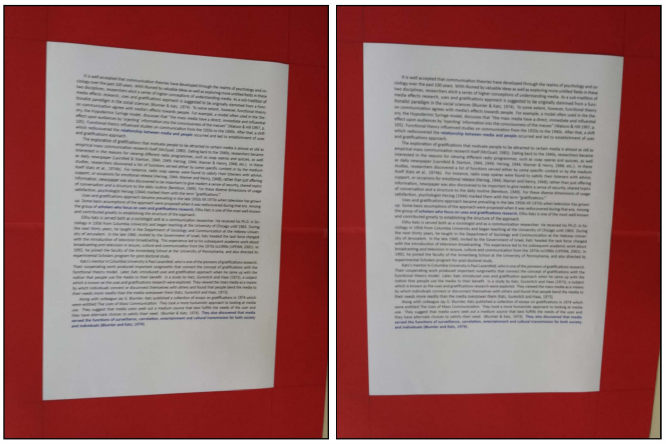
Each image of the dataset is provided with the following ground-truth information:

- The transcription of the text in a document
- The capture parameters (distortion types and values) and the ID of a captured document
- The results of two OCR systems: Abbyy Finereader Engine 11 and Tesseract
- The evaluation of the results of the OCR systems with ISRI-UNLV tools [8]
- A sharp "reference" image of the document at each combination of position and light

Tesseract was used with the default settings, those are: fully automatic page segmentation without orientation and script detection. Finereader was used with the TextUnicode default setting which permits a text-only output. We note however that Tesseract performs very poorly on the receipt document images, because it cannot segment them properly. Hence, those OCR results might not be useful for estimating the quality of the receipt images. As for computing the OCR accuracy results of applying the two OCR systems on our dataset, we used the accuracy program provided by ISRI-UNLV tools [8].

B. Evaluation Methodology and Metrics

For each input image of our proposed dataset, a quality assessment method outputs a quality score associated with that image. The output of a set of images is a list of quality scores. In order to test the ability of the QA method to predict OCR accuracies of the images, the predicted scores have to correlate well with the corresponding OCR accuracies of the images.



(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.

The majority of the works on document image quality assessment use two metrics: the Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank-order Correlation Coefficient (SROCC). They are used to compute the correlation between OCR accuracy values and the predicted quality scores of a quality assessment method. PLCC is used to evaluate prediction accuracy, while SROCC is used to assess prediction monotonicity. A good objective quality measure is expected to achieve high values in PLCC and SRCC. These metrics can be used on the proposed dataset with the two results of Tesseract and Abbyy Finereader Engine 11.

Some quality assessment methods are trained to identify the distortion type present in an image. Hence, the accuracy of the correct identification of distortion types can also be used as an evaluation metric.

The "reference" images – provided for each set of captures at each combination of position and light – could be used for two evaluation purposes: Firstly, for judging the performance of quality enhancement methods. Secondly, to evaluate full-reference quality assessment methods, such methods judge the quality of a distorted image with respect to a high quality reference image.

IV. Conclusions

In our pursuit to address the lack of datasets for mobile captured documents, we have created a new dataset, named *SmartDoc-QA*, that targets quality assessment tasks for the objective of later digitization and OCR processes. This dataset provides three innovative aspects. Firstly, it proposes both single and multiple capture distortions for subsets of the images, allowing to address a specific issue or real conditions. Secondly, it uses different real paper document types. Thirdly, it contains a complete ground-truth with type and amount of each distortion contained in the images, outputs from common OCR systems, OCR accuracy values and reference images.

We believe such dataset can be helpful to the community for investigating two topics in particular: assessing document image quality, and improving image quality, both in the perspective of OCR processing and document digitization. The precise quantification of the capture conditions for each image, along with OCR outputs, are valuable for training OCR quality predictors. Reference images and OCR ground-truth are, on the other hand, helpful for evaluating the quality of improved or restored images.

Our future work plans include the improvement of human motion reproduction and the extension of the dataset to more document types, more smartphones and more distortions.

Acknowledgments

This work is funded by the *Conseil General de la Charente Maritime (France)*, and is supported by the *European Commission* and the *Conseil Rgional de Poitou-Charentes (France)* under the FEDER program DATA-PC. We would also like to thank Ms. CELIA DUPART and Mr.CLEMENT BERARD for their technical support.

References

[1] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality

[2] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *ACSCC*, 2012, pp. 1693–1697.

[3] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *TIP*, vol. 15, no. 11, pp. 3440–3451, 2006.

[4] J. Kumar, P. Ye, and D. Doermann, "A dataset for quality assessment of camera captured document images," in *CBDAR*, 2014, pp. 113–125.

[5] P. Ye and D. Doermann, "Document image quality assessment: A brief survey," in *ICDAR*, 2013, pp. 723–727.

[6] J. Burie, J. Chazalon, M. Coustaty, S. Eskenazi, M. Luqman, M. Mehri, N. Nayef, J. Ogier, S. Prum, and M. Rusinol, "Icdar2015 competition on smartphone document capture and ocr (smartdoc)," in *ICDAR*, 2015, p. to appear.

[7] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J.Heard, "Building a test collection for complex document information processing," in *ACM SIGIR*, 2006, pp. 665–666.

[8] S. Rice, F. Jenkins, and T. Nartker, "The fifth annual test of OCR accuracy," Information Science Research Institute, Tech. Rep., 1996. [Online]. Available: http://www.stephenvrice.com/images/AT-1996.pdf