2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

CG-DIQA: No-reference Document Image Quality Assessment Based on Character Gradient

Hongyu Li
AI Lab
ZhongAn Information
Technology Service Co., Ltd.
Shanghai, China
Email: lihongyu@zhongan.io

Fan Zhu
AI Lab
ZhongAn Information
Technology Service Co., Ltd.
Shanghai, China
Email: zhufan@zhongan.io

Junhua Qiu
AI Lab
ZhongAn Information
Technology Service Co., Ltd.
Shanghai, China
Email: qiujunhua@zhongan.io

***Abstract*—Document image quality assessment (DIQA) is an important and challenging problem in real applications. In order to predict the quality scores of document images, this paper proposes a novel no-reference DIQA method based on character gradient, where the OCR accuracy is used as a ground-truth quality metric. Character gradient is computed on character patches detected with the maximally stable extremal regions (MSER) based method. Character patches are essentially significant to character recognition and therefore suitable for use in estimating document image quality. Experiments on a benchmark dataset show that the proposed method outperforms the state-of-the-art methods in estimating the quality score of document images.**

I. Introduction

With the pervasive use of smartphones in our daily life, acquring document images with mobiles is becoming popular in digitization of business processes. The optical character recognition (OCR) performance of mobile captured document images is often decreased with the low quality due to artifacts introduced during image acquisition [1], which probably hinders the following business process severely. For example, during online insurance claims, if a document image of low quality, submitted for claims, is not detected as soon as possible to require a recapture, critical information may be lost in business processes once the document is unavailabe later. To avoid such information loss, therefore, automatic document image quality assessment (DIQA) is necessary and of great value in document processing and analysis tasks.

Methods for natural image quality assessment may not be suitable for document images because both the properties of document images and the objective of DIQA are totally different. To estimate the quality of document images, many no-reference (NR) assessment algorithms have been proposed, where the reference document image is not accessible in most practical cases. According to the difference of feature extraction, these NR DIQA methods can be categorized as two groups: learning-based assessment and metric-based assessment.

The learning-based DIQA methods take advantage of learning techniques, such as deep learning [2], to extract discriminant features for different types of document degradations. They perform well only on the dataset on which they were trained. However, it is unrealistic to collect sufficient document samples for training in real applications.

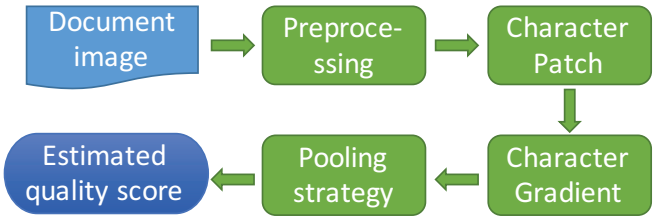


Fig. 1. Flowchart of the CG-DIQA method

The metric-based methods usually are based on hand-crafted features that are correlated with the OCR accuracy.Some degradation-specific quality metrics have been proposed to measure noise and character shape preservation [3]. Although much progress has been made in metric-based assessment, there still exists a clear problem. Features used in existing methods are generally extracted from square image patches, many of which do not have visual meaning involving character/text. Therefore, the resultant features, probably containing much noise, are not optimal for DIQA.

In addition, although the document image quality is affected by many degradations caused during image acquisition, blur is often considered as the most common issue in mobile captured document images, arising from defocus, camera motion, or hand shake [1], [4], [5]. Moreover, the blur degradation has a bad impact on the OCR accuracy, which suggests that detecting the blur degradation is more attractive and useful in practical applications.

The most striking feature distinguishing document images from other types of images is character/text. As a consequence, DIQA can be assumed as measuring the blur degradation of character/text. It is also observed that the gradient of the ideal character edge changes rapidly while the gradient of the degraded character edge has smooth change. Inspired by the assumption and based on the observation, we use the gradient of character edge to measure the blur degradation of document images. To find meaningful patches containing character/text in document images, we choose maximally stable extremal

978-1-5386-3787-6/18/$31.00 ©2018 IEEE

regions (MSER) [6] as character patches, which is often used for character/text detection in OCR.

In this paper, to ensure that the legibility of captured document images is sufficient for recognition, we propose a no-reference DIQA framework, *CG-DIQA*, based on character gradient. Without need of prior training, the proposed method first employs the MSER based algorithm to detect significant character patches and then uses the gradient of character edge to describe the quality model of a document image.

II. Approach

In the proposed CG-DIQA method, we first convert an input document image to a grayscale image followed by downsampling to a specific size, then detect character candidates as selected patches, and finally compute the standard deviation of character gradients as the estimated quality scores for the document image. The flowchart of the proposed method is demonstrated in Fig. 1. Different steps of the CG-DIQA method are described in detail in the consecutive subsections.

A. Preprocessing

To make quality assessment methods robust and efficient, preprocessing is generally required for DIQA. In the preprocessing step, a document image is initially converted into a grayscale image. Downsampling is also performed on the image in order to speed up the following processes if the resolution of document images is greater than 1000×1000. Smoothing is unnecessary in the proposed method, since most of image noise will be avoided after extracting character patches and smoothing may deteriorate the blur degradation of document images.

B. Character patch

Before measuring document quality, it is necessary to extract meaningful features for representing document images. Since the most significant feature of document images is character/text and DIQA is usually with respect to OCR performance, we replace a document image with the patches that contain characters during quality assessment. Using character patches can also make the proposed method more efficient, since it is easier for the method to handle patches rather than an entire image.

To extract character patches, the MSER based method [6] is first adpoted to detect character candidates, which perfroms well in scene text detection [7]. The main advantage of the MSER based method is that such algorithm is able to find most legible characters even when the document image is in low quality. To remove repeating character candidates, the pruning process is incorporated by minimizing regularized variations [7] in the MSER based method.

Since characters are often degraded to smaller broken strokes which cause extremely lower/higher width-height ratio, the width-height ratio r_c of characters is also used to remove those broken strokes and meaningless non-character regions obtained with the MSER based method. r_c is set between 0.25 and 4 in this paper. The eventual bounding boxes represent typical character patches in document images.

C. Character gradient

It is observed that the gradient of degraded character edges is with smooth change. Based on the observation, we use character gradient in character patches to measure the doument image degradation.

The image gradient can be calculated through convolving an image with a linear filter. We have studied several often-used filters, such as the classic Sobel, Scharr and Prewitt filters, and find that the Sobel filter performs best in predicting quality scores of document images. Thus, we choose the Sobel filter to calculate the character gradient in this paper. The Sobel filters on both directions are described as:

$$\mathrm{f}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathrm{f}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}. \tag{1}$$

Convolving f_x and f_y with a character patch (denoted by c) yields the horizontal and vertical gradient of the patch. The gradient magnitude of patch c at position (i, j), denoted by $\mathrm{m}_c(i, j)$, is computed as follows:

$$\mathrm{m}_c(i,j) = \sqrt{(\mathrm{c}\otimes\mathrm{f}_x)^2(i,j) + (\mathrm{c}\otimes\mathrm{f}_y)^2(i,j)}. \tag{2}$$

In future, if better ways of calculating the character gradient emerge, it is easy to incorporate such ways into the proposed DIQA framework.

D. Pooling strategy

Borrowing the idea in the literatures [8], [9] where gradient features are used for DIQA, we compute the overall quality score s of a document image as the standard deviation of character gradients via some pooling strategy. The widely used average pooling is adopted in this work to obtain the final quality score for a document image. That is, first take the average m_a of character gradients,

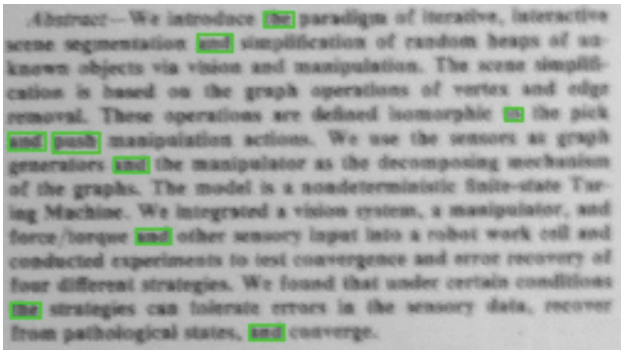
$$\mathrm{m}_a = \frac{1}{N}\sum_c\sum_{i,j}\mathrm{m}_c(i,j),$$

where N is the total amount of pixels in all character patches, and then compute the standard deviation of character gradients,

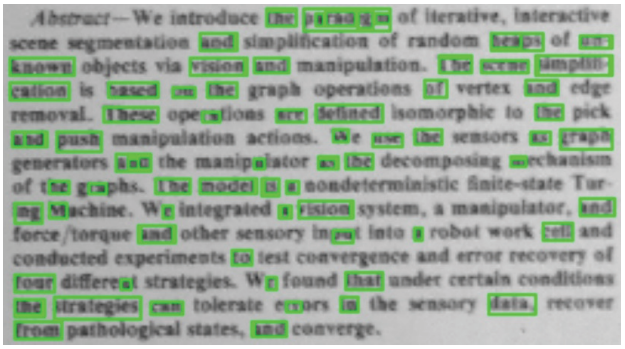
$$\mathrm{s} = \sqrt{\frac{1}{N}\sum_c\sum_{i,j}(\mathrm{m}_c(i,j) - \mathrm{m}_a)^2}. \tag{3}$$

s is eventually used to describe the overall quality prediction score for a document image.

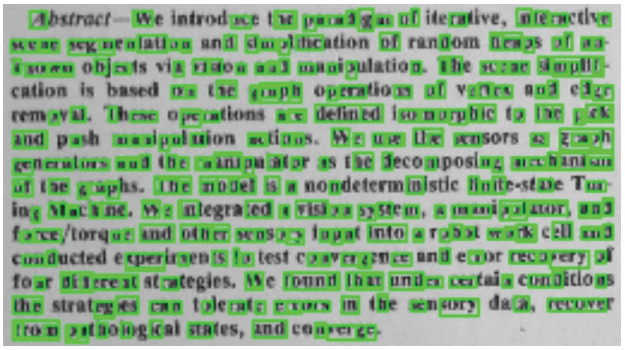
Since different character patches may contribute differently to the overall quality score in a document image, the final quality score can be computed through weighting character gradients. Weighted pooling may have better DIQA accuracy than average pooling, but weighted pooling will result in more computing overhead and can make the pooling process more complicated. Furthermore, quality scores predicted with weighted pooling is more nonlinear, which is not beneficial to the following business process.



(a) A low quality cropped document image



(b) A degraded cropped document image



(c) A high quality cropped document image

Fig. 2. Samples of cropped document images in the DIQA dataset. The extracted character patches are surrounded with green bounding boxes. These three images are respectively with the average OCR accuracies of 33.62%, 70.37% and 93.01%. Their quality prediction scores are 47.34, 61.45, and 72.21 respectively.

III. Experiments

A. Dataset and evaluation protocol

We present evaluation results on a public DIQA dataset [10] containing a total of 175 color images. These images with resolution 1840×3264 are captured from 25 documents containing machine-printed English characters using a smartphone. 6-8 photos were taken for each document to generate different levels of blur degradations. In Fig. 2, we show three samples of cropped document images with different degradation degrees from the DIQA dataset. Fig. 2(a) is a low quality image with severe blur that loses its readability. Fig. 2(b) is a document image with the light degradation but it is still recognizable with human perceptual systems. And Fig. 2(c) is a high quality image that can be easily read and recognized by OCR systems.

TABLE I
Results over the average OCR accuracy

	Median LCC	Median SROCC
Sparse [12]	0.935	0.928
Proposed	**0.9841**	**0.9429**

One traditional quality indicator for document images is the OCR accuracy [11]. Likewise, we define the OCR accuracy as the ground truth for each document image in our experiments. Three OCR engines: ABBYY Fine Reader, Tesseract, and Omnipage, were run on each image in the dataset to obtain the OCR results. The OCR accuracy ranging from 0 to 1 was obtained through evaluating the OCR results with the ISRI-OCR evaluation tool.

To evaluate the performance of the proposed method, the predicted quality scores need to be correlated with the ground-truth OCR accuracies. Thus, the Linear Correlation Coefficient (LCC) and the Spearman Rank Order Correlation Coefficient (SROCC) are used as performance indicators.

In our experiments, the LCC and SROCC are separately computed in a document-wise way. That is, for each document in the dataset, only its corresponding photos are taken into consideration while computing LCC or SROCC. Finally, we can get 25 LCCs and SROCCs involving the proposed method for this dataset. The medians of these 25 LCCs and SROCCs are used as the overall indicators for performance evaluation. Since three OCR engines have huge difference in accuracy, to avoid that the evaluation results are overwhelmingly dependent on a certain OCR engine, we claim to use the average OCR accuracy of three engines for the evaluation purpose.

B. Implementation and Results

To get the optimal parameter setting for the proposed method, we tested different sets of the stable area size s and the maximal variation v_{max} in the MSER based method for extracting character patches. Parameter s that is required to be neither too large nor small can play a role in eliminating the extracted false character regions. The maximal variation v_{max} is used to prune the area that has similar size to its children. Experimental results show that the proposed method generally performs best when v_{max} is set *0.2* and s is in a range *13* pixels to *0.001* of all pixels.

It is observed from our experiments that, once parameters are fixed in the MSER based method, the number of extracted character patches totally relies on the quality of document images. Although the extracted character patches are not too many for severely degraded document images, these patches, the most significant regions for character recognition, are enough to help estimate the quality score in a correct way.

It is also worth noting that it is not the amount of character patches that plays important effects on the character gradient, but the quality of such patches on behalf of the entire docu-

ment image. This strengthens that the character gradient can effectively reflect the document image quality.

For example, in Fig.2, there are only 48 character patches extracted in the severely degraded image with the average OCR accuray of 33.62% (Fig.2(a)), 309 patches in the slightly degraded image with the average accuray of 70.37% (Fig.2(b)), and 1615 patches in the high quality image with the average OCR accuray of 93.01% (Fig.2(c)). These three document images are respectively assessed to be with quality scores of 47.34, 61.45, and 72.21 using the proposed method.

We show our experimental results on the DIQA dataset in Table I. The median LCC and SROCC obtained with the proposed method are respectively 0.9841 and 0.9429. We compare the proposed method with the semi-supervised sparse representation based approach [12] that computes the correlation coefficients as well in view of the average OCR accuracy. As shown in Table I, the proposed method achieves the higher median LCC and SROCC than the sparse approach [12] that is based upon learning techniques.

TABLE II
RESULTS ON ALL DOCUMENT IMAGES

	LCC	SROCC
MetricNR[3]	0.8867	0.8207
Focus[4]	0.6467	N/A
Proposed	**0.9063**	**0.8565**

To avoid the bias towards the good results in terms of the document-wise evaluation protocol, we also directly compute one LCC (90.63%) and one SROCC (85.65%) over the average OCR accuracy for all of the 175 document images in this dataset. Table II shows that our method performs much better than the other two metric-based methods: MetricNR [3] and Focus [4].

C. Comparison

To compare with other state-of-the-art quality assessment approaches, we also compute the correlation values, LCC and SROCC, of the proposed method over three different OCR accuracies. In our experiments, seven general purpose DIQA approaches are selected for comparative anlaysis, including CORNIA [13], CNN [2], HOS [11], Focus [4], MetricNR [3], LocalBlur[5], and Moments[14]. Among them, the first three are based on learning techniques, while the others take advantage of hand-crafted features. Since most of these methods either only focus one OCR accuracry or have no result on all document images, and it is hard to re-implement them to get optimal experimental results, we have to choose different methods with available accuracies for comparion.

Table III illustrates the median LCCs and SROCCs of six DIQA algorithms in terms of the FineReader OCR accuracy. From the results, we can see that, the coefficient LCC of the proposed method is slightly lower than three methods: MetricNR [3], CORNIA [13], and HOS [11]. In addition, the proposed method is better than almost all other methods except the Focus method [4] in the SROCC coefficient. A nice, subtle highlight should be emphasized that the proposed method can perform well under both evaluation protocols, unlike other methods that can only work in a good way under a certain protocol. Since LCC can measure the degree of linear relationship between the predicted quality and OCR accuracy and SROCC can measure how well this relationship can be described using a monotonic function, experimental results demonstrate that the proposed CG-DIQA method is correlated with the FineReader OCR engine in a monotonic and linear way, which is more suitable for real applications.

TABLE III
COMPARISON OVER THE FINEREADER OCR ACCURACY

	Median LCC	Median SROCC
CORNIA[13]	0.9747	0.9286
CNN[2]	0.950	0.898
HOS[11]	0.960	0.909
Focus[4]	0.9378	**0.96429**
MetricNR[3]	**0.9750**	0.9107
Proposed	0.9523	0.9429

In Table IV, we show the median LCCs and SROCCs of four approaches on the Tesseract and Omnipage accuracies. It can be observed that the proposed method provids a lot better OCR prediction scores than the other three methods, even for different OCR engines.

TABLE IV
COMPARISON OVER THE TESSERACT AND OMNIPAGE OCR ACCURACIES

	Tesseract		Omnipage	
	Median LCC	Median SROCC	Median LCC	Median SROCC
LocalBlur[5]	N/A	0. 892	N/A	0. 725
Moments[14]	0.8197	0.8207	N/A	0.6648
Focus[4]	0.9197	N/A	0.8794	N/A
Proposed	**0. 9591**	**0. 9429**	**0. 9247**	**0. 8295**

IV. CONCLUSION AND FUTURE WORK

In this paper, we propose a character gradient based method for document image quality assessment. The method first extracts character patches via the MSER based method, then character gradients are computed for these patches, and finally the standard deviation of gradients are statistically obtained as the quality prediction score. In the method, it is assumed that character patches are more significant than widely-used square image patches in measuring the document image quality since they contain critical features for OCR. Our experimental results demonstrate that the proposed method can predict the quality score of document images very well in terms of both LCC and SROCC. The use of some new techniques to extract features of character patches and a weighting strategy based on patch size is our future research work.

REFERENCES

[1] P. Ye and D. Doermann, "Document image quality assessment: A brief survey," in *International Conference on Document Analysis and Recognition*, 2013, pp. 723–727.

[2] L. Kang, P. Ye, Y. Li, and D. Doermann, "A deep learning approach to document image quality assessment," in *IEEE International Conference on Image Processing*, 2014, pp. 2570–2574.

[3] N. Nayef, "Metric-based no-reference quality assessment of heterogeneous document images," in *SPIE Electronic Imaging*, 2015, pp. 94 020L–94 020L–12.
[4] M. Rusinol, J. Chazalon, and J. M. Ogier, "Combining focus measure operators to predict ocr accuracy in mobile-captured document images," in *Iapr International Workshop on Document Analysis Systems*, 2014, pp. 181–185.
[5] T. Chabards and B. Marcotegui, "Local blur estimation based on toggle mapping," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*, 2015, pp. 146–156.
[6] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image & Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
[7] A. Alaei, D. Conte, and R. Raveaux, "Document image quality assessment based on improved gradient magnitude similarity deviation," in *International Conference on Document Analysis and Recognition*, 2015, pp. 176–180.
[8] J. Kumar, F. Chen, and D. Doermann, "Sharpness estimation for document and scene images," in *International Conference on Pattern Recognition*, 2013, pp. 3292–3295.
[9] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, p. 39, 2016.
[10] J. Kumar, P. Ye, and D. Doermann, *A Dataset for Quality Assessment of Camera Captured Document Images*. Springer International Publishing, 2013.
[11] J. Xu, P. Ye, Q. Li, Y. Liu, and D. Doermann, "No-reference document image quality assessment based on high order image statistics," in *IEEE International Conference on Image Processing*, 2016, pp. 3289–3293.
[12] X. Peng, H. Cao, and P. Natarajan, "Document image quality assessment using discriminative sparse representation," in *Document Analysis Systems*, 2016, pp. 227–232.
[13] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 1098–1105.
[14] K. De and V. Masilamani, "Discrete orthogonal moments based framework for assessing blurriness of camera captured document images," in *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC–16')*, V. Vijayakumar and V. Neelanarayanan, Eds., 2016, pp. 227–236.