 doi:10.20944/preprints202107.0200.v1

Article

SEDIQA: Sound Emitting Document Image Quality Assessment in a Reading Aid for the Visually Impaired

Jane Courtney[1]

[1] School of Electrical & Electronic Engineering, Technological University Dublin, City Campus, Ireland

* Correspondence: jane.courtney@tudublin.ie

Abstract: For Visually impaired People (VIPs), the ability to convert text to sound can mean a new level of independence or the simple joy of a good book. With significant advances in Optical Character Recognition (OCR) in recent years, a number of reading aids are appearing on the market. These reading aids convert images captured by a camera to text which can then be read aloud. However, all of these reading aids suffer from a key issue – the user must be able to visually target the text and capture an image of sufficient quality for the OCR algorithm to function – no small task for VIPs. In this work, a Sound-Emitting Document Image Quality Assessment metric (SEDIQA) is proposed which allows the user to hear the quality of the text image and automatically captures the best image for OCR accuracy. This work also includes testing of OCR performance against image degradations, to identify the most significant contributors to accuracy reduction. The proposed No-Reference Image Quality Assessor (NR-IQA) is validated alongside established NR-IQAs and this work includes insights into the performance of these NR-IQAs on document images.

Keywords: image quality assessment; image quality metrics; NR-IQAs; D-IQA; OCR accuracy; OCR prediction; OCR improvements; visual aids; visually impaired; reading aids; document images; text-based images

1. Introduction

With advances in smartphone technology, particularly in camera quality, several visual aids for VIPs are emerging [1,2] with Microsoft's Seeing AI as the current market front-runner. These assistive technologies range from navigation aids [3] to object detectors [4] and readers [5]. However, this last task has embedded in it a reliance on OCR accuracy and therefore on image quality. This means that the user's performance (hand motion, visual acuity, etc.) will affect the performance of the reader. Since these readers are both hand-held and designed for people with visual impairments, this is a fundamental issue that needs to be addressed.

To solve this issue, automatic processing can be done to improve OCR performance [6,7], but even the best performing pre-processors cannot achieve high OCR accuracy out of a low-quality image. Therefore, it is necessary to also assess the image quality before attempting OCR, and so a robust Image Quality Assessment (IQA) metric is needed.

For this application, in the absence of a reference image, No-Reference Image Quality Assessment (NR-IQA) – otherwise known as "blind" IQA – is required. Most established NR-IQAs concentrate on perceptual image quality [8] but it has been found that these are not suitable for the application of document images, as the degradations that affect text-based content, and subsequently, OCR accuracy, can be quite different, and what is considered "high quality" in a scene image does not correlate with document image quality [9]. In fact, in previous work by the author [10], an unexpected reverse relationship was discovered between established NR-IQAs such as BRISQUE [11], NIQE [12] and PIQE [13] and document image quality, with the metrics reporting lower quality results for ideal images than for their scanned counterparts.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

A number of blind Document Image Quality Assessors (D-IQAs) have been developed to date. Some concentrate on specific degradations such as compression [14] or blur [15,16] while others concentrate on perceptual quality [17–19]. More recently, methods have tended towards learning [20–23]. However, trained networks can be slow and are affected by the size and diversity of the training dataset. Some promising work has emerged in the area of screen content quality [24–26] and, in this application area, a relationship between entropy and text-based image quality in [27]. Other direct quality measures concentrate on the gradient of the image [28,29]. The design of SEDIQA builds on these findings to create a robust but directly measurable image quality metric.

As well as the metric, this work includes a document image enhancement technique with an emphasis on OCR accuracy improvements. Document image enhancement is still an open field of research, and the SmartDoc competition [30,31] continues to encourage development in the area and to allow evaluation of document image enhancers and improvements on OCR accuracy. Some contributions have been made [16,32] using the associated dataset, which is also used here for comparison.

This paper takes a systematic approach is taken to the investigation of OCR accuracy by first testing the relationship between accuracy and image degradations to determine which degradations should be the focus of the quality metric and the image enhancer. Performance of SEDIQA's *Q*-metric was evaluated by comparing it against image degradations and OCR accuracy as well as evaluating its performance alongside established NR-IQAs. The document image enhancer was evaluated by investigating improvements in OCR accuracy.

This full SEDIQA system is a Visual Reading Aid design that automatically captures, assesses and converts the camera-captured document images to audio outputs and ensures the best possible OCR accuracy for any given capture scenario.

The major contributions in this work include:

1) Testing of OCR Accuracy vs Image Degradations
2) Identification of the degradations that contribute most significantly to OCR accuracy reduction
3) A new, robust and directly measurable NR-IQA metric for document images. This is validated by testing on both synthetic and real images, against image degradations and OCR accuracy and alongside established NR-IQAs
4) Insights into the performance of established NR-IQAs on document images
5) Improvements in OCR accuracy in the full SEDIQA design
6) The full SEDIQA Design as a Visual Reading Aid

2. Materials and Methods

This system was designed in Python 3.7 with OpenCV 4.3. Testing was done on a PC using Tesseract [33] for OCR and MATLAB 2020b for BRISQUE, NIQE and PIQE image quality tests. Tests were performed on synthetically degraded images, live captured images and the SmartDoc dataset [34] as an established benchmark. The SmartDoc dataset provides an excellent testbed for this design as it contains images of the same documents captured under different capture conditions and with different capture parameters – very much representing a realistic page capture scenario. The dataset contains images with both single and multiple distortions, which include variations in lighting, focus, distance and motion blur. The full dataset was tested but graphical results are presented for an individual document from this dataset (*D1*) for clarity.

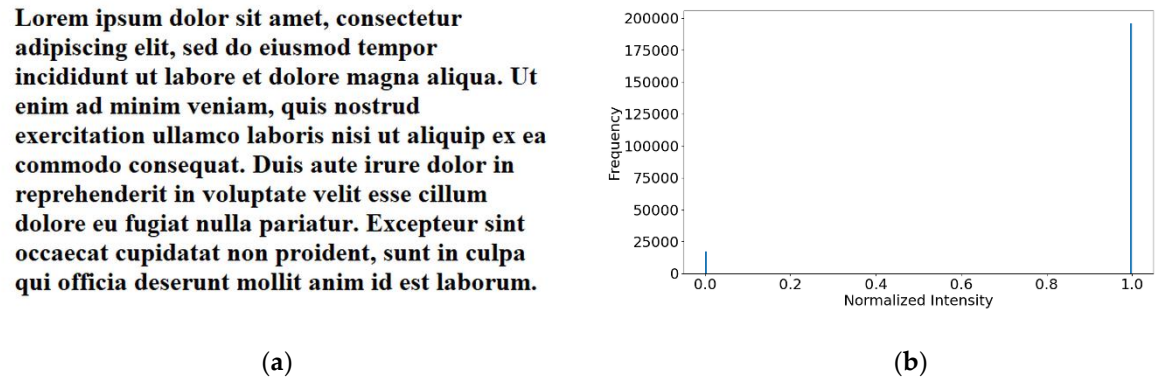
2.1 OCR Accuracy vs Quality

OCR performance has been found to deteriorate significantly under real image degradations [35,36]. Some of these degradations are due to camera parameters and are constant for a given capture setup, others will vary with user performance (how accurately the user targets the document with the camera) and external conditions such as lighting. Compression and resolution belong to the former type and are constant, as the same smartphone camera is used to capture raw image data. However, noise, blur, contrast and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

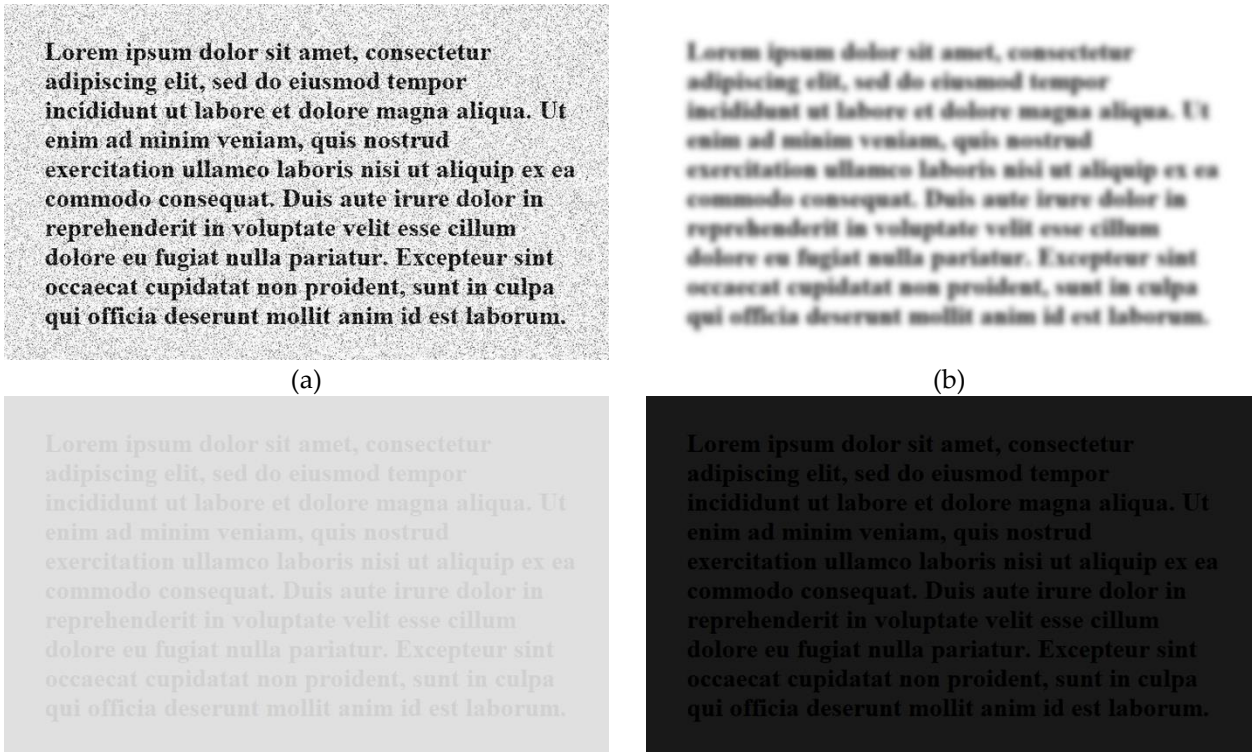
brightness will change for each capture scenario. While contrast has been the emphasis of many document image binarization techniques [37], it has been shown to have relatively little effect on OCR performance, while noise and blur have been found to be the degradations which contribute most significantly to accuracy reduction.

To confirm this, testing is done on synthetic images with noise, blur and contrast introduced separately. These images are created initially in imaging software, with clear black text on a homogenous white background and represent ideal text images. Degradations are then introduced incrementally to study the effects of each type of degradation. A synthetic image and its associated histogram are shown in **Figure 1** and samples of degraded versions of this are shown in **Figure 2**.



(**a**) (**b**)

Figure 1. A sample ideal text image alongside its corresponding histogram: (a) ideal text image; (b) intensity histogram.



(a) (b)

(c) (d)

Figure 2. Degraded counterparts of the synthetic text image in **Figure 1**: (a) noisy (Gaussian noise added); (b) blurred Gaussian blurring); (c) reduced contrast; (d) reduced brightness.

The set of synthetically degraded images was tested for OCR accuracy using the standard percentage Characters Correct as an accuracy measure. From the synthetic image tests, it is possible to ascertain which degradations have the most significant effect on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

OCR accuracy. These results were used in the development of the Q-metric and the Page Extractor, which is used to extract and enhance the document image.

2.2 SEDIQA Quality Measure

As can be seen in its image histogram (see **Figure 1**), an ideal text image is characterized by high median brightness (for dark text on a bright background), high contrast and low entropy. Entropy is visible as the spread of image values in the histogram, contrast approximates the width of the histogram and median brightness is the most common intensity value, in this case, white. The simple median intensity suffices for the majority of images and on the test dataset but will not work for bright text on dark backgrounds. To address this, a simple dominant intensity test can be used to determine whether to invert the image.

It has been shown that entropy, while considered proportional to quality in natural scene images, has a negative correlation with quality in text-based images, with higher entropy denoting lower quality [10,27]. As entropy captures both noise and blur deteriorations, it is potentially a good measure for predicting OCR performance:

$$E(I) = \sum^{N} p_I \log_2 p_I \tag{1}$$

However, since entropy does not vary with brightness or contrast, brightness and contrast approximation measures are also needed. The median intensity, $\tilde{I}$, of the image is a good approximation of brightness and should be high in a good quality text-based image with a bright background. The standard deviation, σ, of the image approximates contrast, as it is proportional to the width of the histogram:

$$\sigma(I) = \sqrt{\frac{\sum(I - \mu_I)^2}{N}} \tag{2}$$

However, in real images, degradations are rarely constant throughout the image, so the entropy and standard deviation of the whole image are not useful. To overcome this issue, two versions of the image are acquired: the Entropy Image (EI) and the Gradient Image (GI).

The entropy image gives local values of entropy for each location in the image. This local entropy should be high around text-content regions but low in the homogenous background. This means the median of the entropy image in a high-quality text image should be low while its standard deviation should be high. The gradient image captures the local contrast between the text and the background so in this image, again, the median should be low while the standard deviation should be high. In fact, the median of the gradient image is zero, so this term is omitted for computational efficiency.

$$Q = \begin{cases} \dfrac{\tilde{I} + \sigma(EI) + \sigma(GI)}{\widetilde{EI}} & if\ \widetilde{EI} > 0 \\ \tilde{I} + \sigma(EI) + \sigma(GI) & if\ \widetilde{EI} = 0 \end{cases} \tag{3}$$

where $\tilde{I} =$ intensity median, $\widetilde{EI} =$ entropy image median, $\sigma(EI) =$ entropy image standard deviation and $\sigma(GI) =$ gradient image standard deviation.

2.3 Validation of SEDIQA's Q metric

The Q values of the synthetically degraded images were measured to confirm the relationship between Q and image degradations noise, blur, contrast and brightness, as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

well as Q's relationship with OCR accuracy. The Q values and OCR accuracies of the SmartDoc dataset were also measured to confirm the relationship between Q and OCR accuracy under real-world conditions. To compare Q with other measures, this same testing approach on both synthetic and real images was used with a set of well-established NR-IQAs: BRISQUE, NIQE and PIQE.

2.4. SEDIQA Design

The full SEDIQA design consists of three main stages: page extraction, quality measurement and audio output. As well as emitting a tone during capture to guide the user, the system automatically retains the highest quality image, allowing for automatic best-image selection. This image is then passed through an OCR algorithm, followed by text-to-speech software.

Until the page extraction stage is successful, the output will be a repeated 'chirp' and no image is retained. Once the page extractor successfully finds text, this is cropped, perspective warped and cleaned to create a page image, which is both saved as an initial best image and passed to the quality measurement stage. The quality of the image is measured, saved as an initial Q_{MAX}, and converted to a tone where the frequency of the tone is proportional to the quality.

While SEDIQA is active, a new frame is grabbed from the camera and this process is repeated with an image added to the Best Image Array whenever its Q value exceeds Q_{MAX}, at which point Q_{MAX} is updated to the minimum Q in the array.

The full workflow for the system can be seen in **Figure 2**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

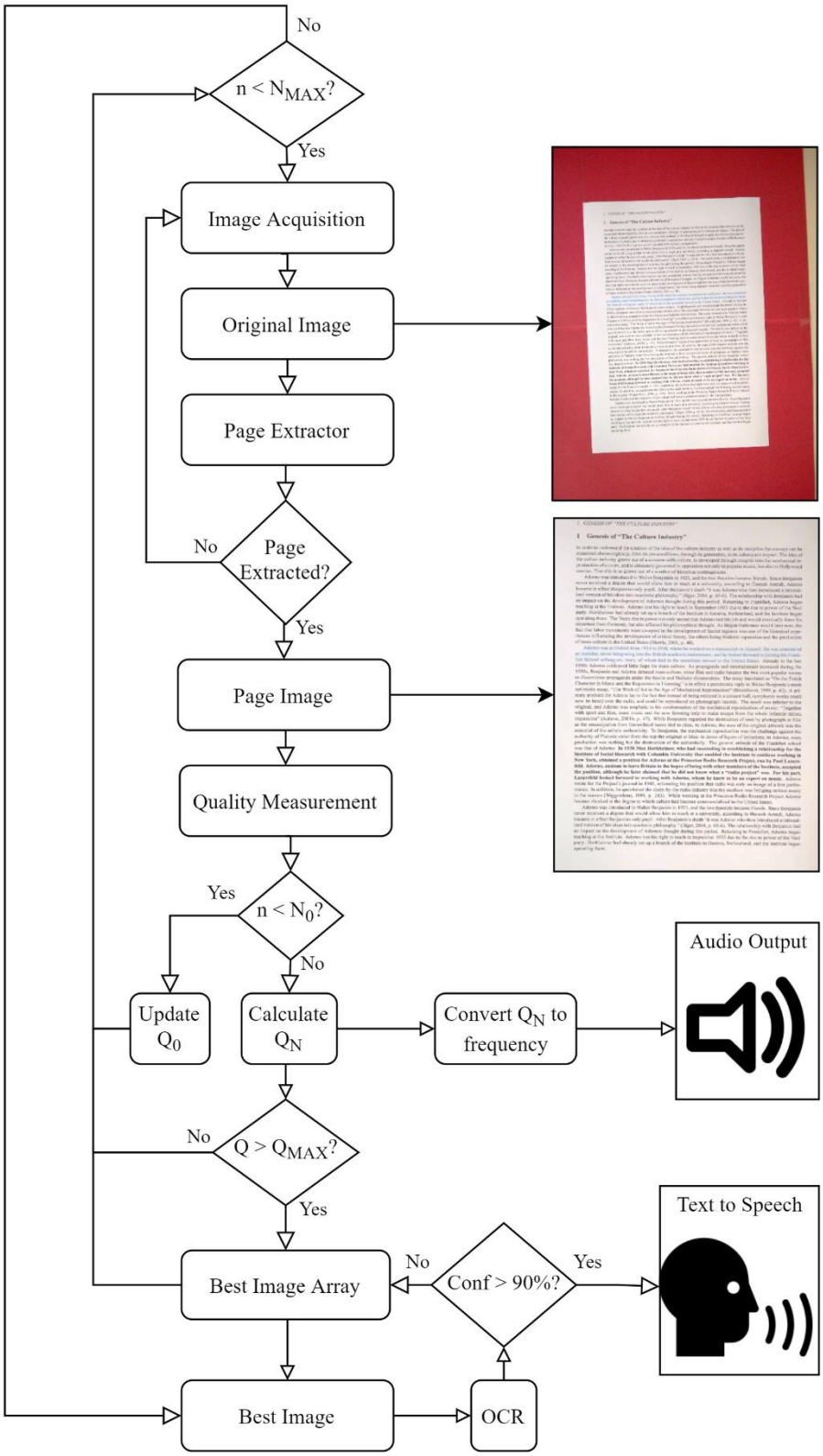


Figure 3. The workflow for the complete SEDIQA system; n = current image index, N_{MAX} = maximum no. of images to be captured, Q_0 = baseline Q, Q_{MAX} = lowest Q in the Best Image Array, Q_N = normalized Q, *Conf* = OCR confidence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

2.1.1 Page Extractor

A simple page extraction method is required to ensure that text has been found before a quality measure is taken. This method must be fast enough to ensure it can be implemented in near real-time but robust enough to operate under a variety of image degradations.

In this design, the image is first binarized using adaptive thresholding [38] followed by a dilation [39] to accentuate the boundary of the page. The largest contour in this image is taken as the page boundary and a polygon approximation is performed to establish the corners. While more sophisticated text detection methods, such as Google Vision or EAST [40], could replace this method, the design here was found to suffice for a typical page capture scenario, which is less challenging than the more general 'Text in the Wild' scenario, and has been shown to work on the SmartDoc dataset and in live testing.

The extracted page boundary is then used in a perspective transform [41] to warp and crop the image to only the page region. This is then resized using a cubic interpolation and cleaned by contrast enhancing, sharpening and denoising.

The full workflow for the system can be seen in **Figure 3**.

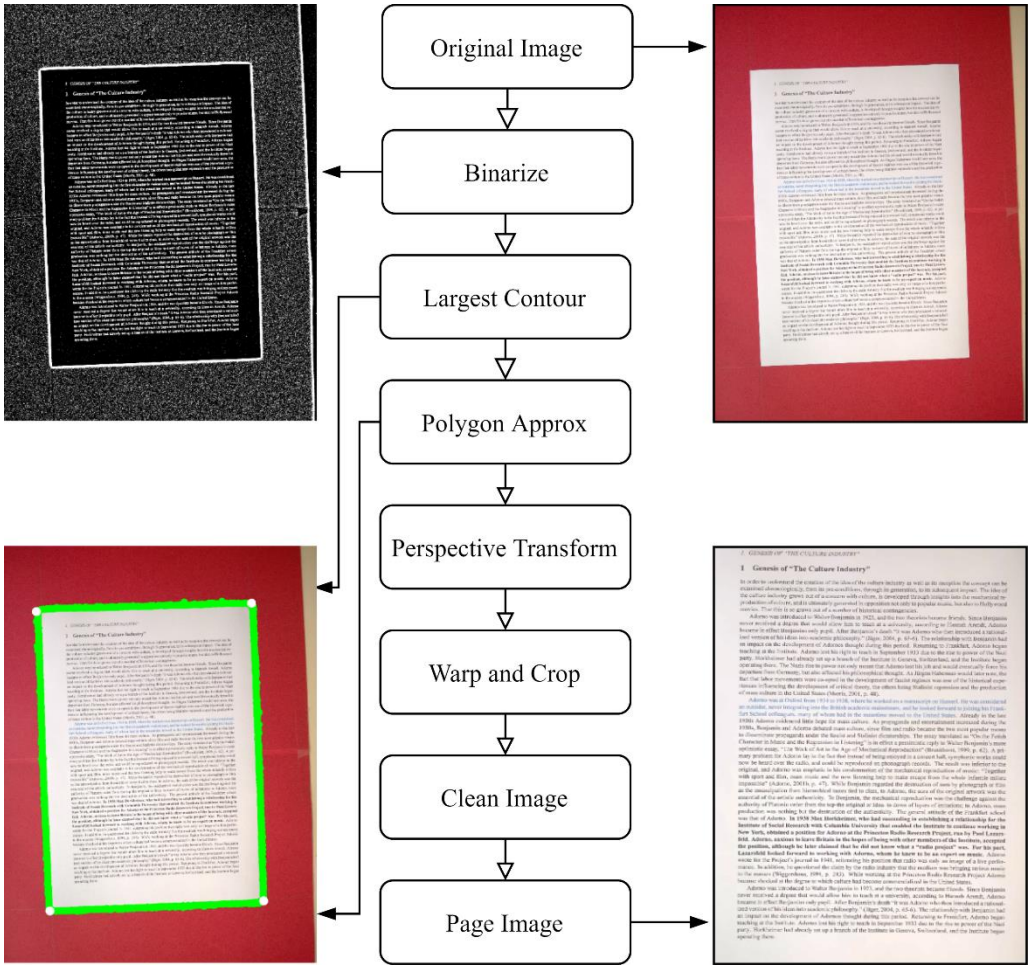


Figure 4. The workflow for the Page Extractor.

2.1.2. Quality Measurement

In the quality measurement stage of SEDIQA, Q is measured once text has been found. For the first N_0 images, the mean Q is calculated to act as a baseline, Q_0. This is used for an approximate normalization of Q to ensure that the frequency falls into the audible range. It is also used as an initial maximum, Q_{MAX}.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

$$Q_N = \frac{Q}{Q_0} \quad (1)$$

If $Q_N < Q_{MAX}$, no change occurs in Q_{MAX} and no image is captured, however Q_N is still calculated and passed to the audio output stage. If $Q_N > Q_{MAX}$, the image is retained as a member of the Best Image Array and Q_{MAX} is updated to the minimum Q in the array. The process is continued until the array a sufficient number of images have been tested and the array is full.

2.1.3. Audio Output

When the text is initially found by the Page Extractor, a neutral tone of 300 Hz is emitted. Once the baseline, Q_0, is established (after the first N_0 runs), Q_N is calculated. Decreases in Q_N are translated to lower frequency tones while increases are higher, tending towards the preferred 400-800 Hz range [42]. Q_N is found to vary by about $\pm 50\%$ over a full range of OCR accuracies from 0 to 99%, so to stay in the audible range and close to the preferred frequency range, Q_N is scaled by 400 to convert it to Hz. The frequency of the tone in Hz is then given by:

$$f = 400 Q_N \quad (6)$$

The sound only acts as a guide for the user as the best images will be retained automatically. At the end of the capture process, the best image is passed to a Tesseract OCR algorithm and converted to text. If the confidence value returned by this algorithm is too weak, the "best image" can be rejected and the next best image used. When the confidence value is sufficiently high, the text extracted by this process is then passed to text-to-speech software to be read aloud.

2.5. Accuracy Improvements

To investigate the effect of the SEDIQA system on OCR accuracy, each document from the SmartDoc dataset was passed through the SEDIQA system and the OCR accuracy was measured on both the original images and the extracted page images. The full dataset was tested but for clarity, results are presented for a sample document (*D1*). This document was selected for presentation as it was found that the range of images of this document in the dataset, captured under different conditions, led to a full range of initial accuracies ranging from 0 to 99%. Although the full SEDIQA system only retains the best of these, the full range is presented to demonstrate the extent of the accuracy improvements introduced by the system.

3. Results

Before the full SEDIQA system was tested, the relationship between OCR accuracy and image degradations was investigated using synthetically created and degraded images of text (see **Figure 1** and **Figure 2**). The *Q*-metric was validated on these synthetic images and its performance with respect to image degradations was compared with established NR-IQAs as well as with the OCR accuracy. Note that in these images, *Q* does not need to be normalized for audibility, so the original *Q* value is used.

The OCR accuracy, *Q*-metric and established NR-IQAs were also tested on real camera-captured images from the SmartDoc dataset. The full SEDIQA system was tested on this dataset and in live capture to confirm the relationship between OCR accuracy and the *Q*-metric and to test the system's accuracy improvements.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

3.1 Synthetically Degraded Images

Using synthetically created text images, such as the one shown in **Figure 1**, the relationship between OCR accuracy and different forms of image degradation can be established. The level of degradation is increased from zero (original, ideal image) to a maximum, and the OCR accuracy is tested for each level. Degradation is continued until OCR accuracy collapses, or a maximum degradation level is reached.

In previous work [35,36], it has been shown that noise and blur have significant effects on OCR accuracy while contrast and brightness have almost no effect. Although brightness and contrast show no effect on the OCR Accuracy in the ideal image case, variations in lighting throughout a real image can affect OCR performance and so these degradations are not ignored in developing the *Q*-metric.

The synthetic images were measured with the *Q*-metric and as a further validation, tested with well-established NR-IQAs: BRISQUE, NIQE and PIQE. Note that for these NR-IQAs, a lower value denotes higher quality. Results are presented here for each degradation type: noise, blur, contrast and brightness.

3.1.1. Noise

To test the metric's response to noise, Gaussian noise is incrementally added to the ideal image. The noise level is set by the sigma value, σ, in the probability density function:

$$G = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(I-\mu_I)^2}{2\sigma^2}} \tag{7}$$

The noise levels range from a value of 0.0125 for sigma at level 1 to 0.2375 at level 19, where OCR accuracy collapse occurs. Results are show in **Figure 5**. Although *Q*'s response to noise is a tad overdramatic, this ensures that images with high levels of noise would be rejected by the system and only the least noisy images would be retained as Best Image Array candidates. Of the other NR-IQAs, only PIQE shows the correct response to noise (decreasing quality with noise level). NIQE shows quality increasing with noise while BRISQUE shows almost no response (though it incorrectly shows lowest quality for the ideal image).

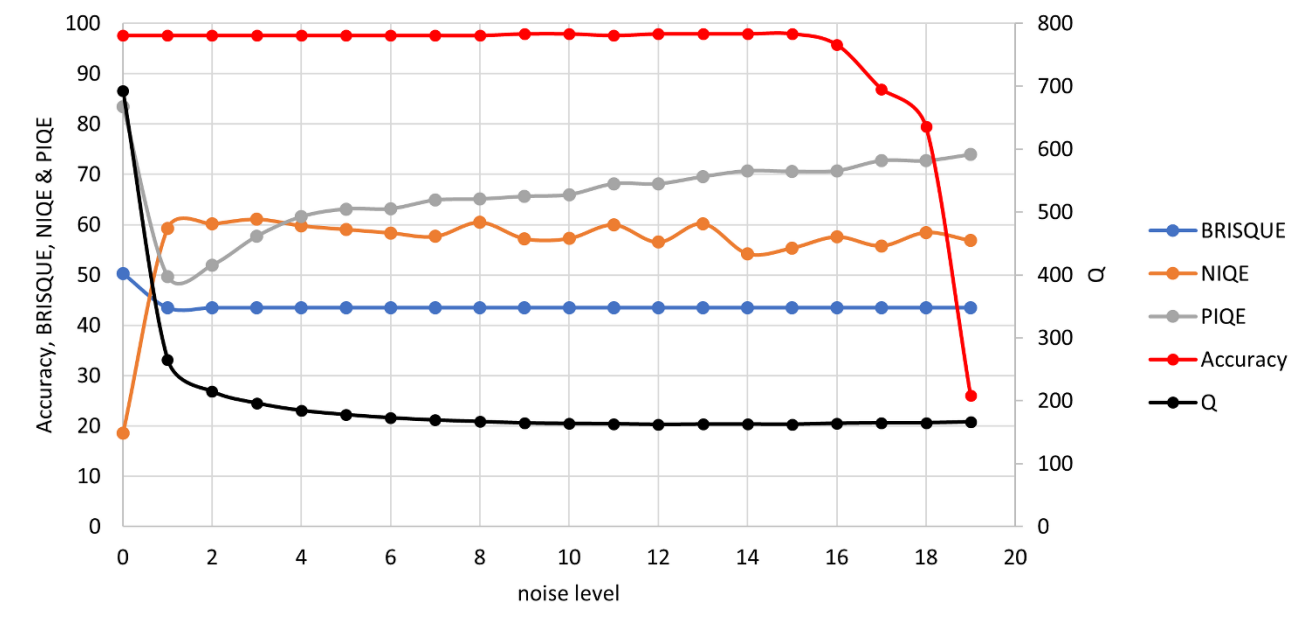


Figure 5. OCR Accuracy and Quality Metrics vs Noise Level. Note: high BRISQUE, NIQE and PIQE values denote low quality while high *Q* values denote high quality. A separate axis is used for *Q* here as the unnormalized *Q* value tends to be considerably higher than the other NR-IQAs, which tend to stay in the range 0 to 100.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

3.1.2. Blur

For blur, Gaussian blurring is again used. This time, the blur level is set by the size of the kernel. As the blurring kernel must center on a pixel, it can only have odd values, so the number of data points is limited. The blur levels range from a kernel size of 3x3 (level 3) to 19x19 (level 19), though OCR accuracy collapse occurs around level 15.

Results are show in **Figure 6**. Although Q erroneously shows an increase in quality at low blur, this is due to the median entropy in the ideal image being zero – a situation that does not arise in real images. In fact, the Q-metric tends towards infinity for an ideal image. BRISQUE and PIQE show the correct response to blur (decreasing quality with blur level) while NIQE shows quality increasing with blur.

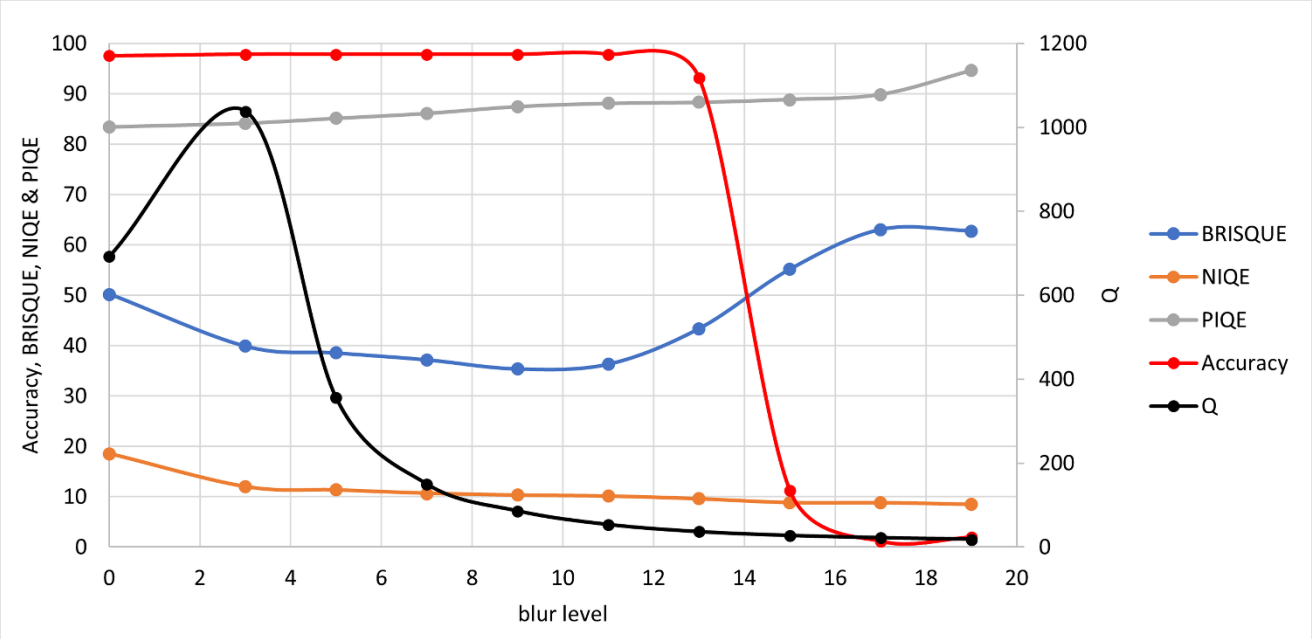


Figure 6. OCR Accuracy and Quality Metrics vs Blur Level. Note: high BRISQUE, NIQE and PIQE values denote low quality while high Q values denote high quality. Again, a separate axis is used for Q here as the unnormalized Q value tends to be considerably higher than the other NR-IQAs, which tend to stay in the range 0 to 100.

3.1.3. Contrast and Brightness

For contrast and brightness, the alpha and beta (gain and bias) parameters are used respectively:

$$D = \alpha I + \beta \tag{7}$$

where $D =$ degraded image, $I =$ original image, $\alpha =$ gain and $\beta =$ bias.

For contrast, the gain is decreased until a difference of just one intensity level (in the range 0 to 255) between text and background is observed. The contrast levels range from an α of 0.2 for level 1 to 0.02 for level 19. For brightness, the bias is decreased until the intensity level is just 1 in this same range 0 to 255. The β range is from 0.95 for level 1 to 0.05 for level 19.

Results are shown in **Figure 7** and **Figure 8**. The OCR Accuracy is not affected by either of these degradations in the synthetic image case, but it was found later in real image testing that those images with lighting issues tended to perform poorly in OCR, and so these degradations were included in testing here. This is most likely due to non-uniform lighting across the document in real capture. Q's response to both is linear, ensuring that images with lighting issues would be rejected by the system. Of the other NR-IQAs, only PIQE shows the correct response (decreasing quality with lighting issues). NIQE shows quality increasing with noise while BRISQUE shows almost no response.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

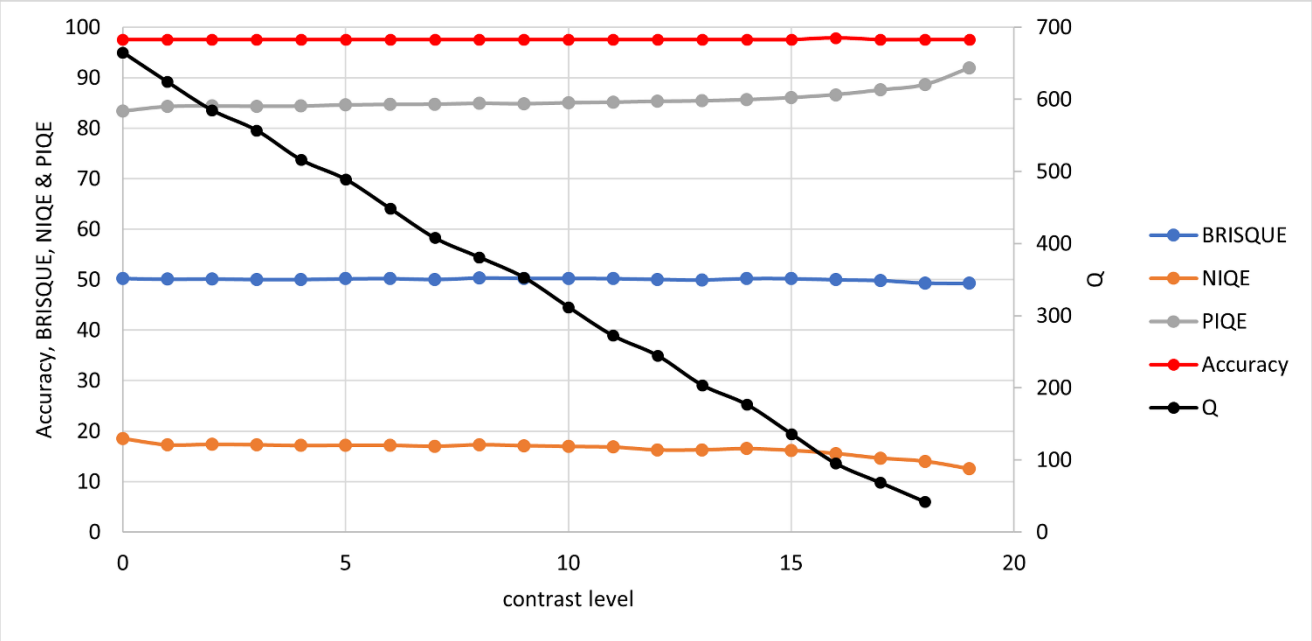


Figure 7. OCR Accuracy and Quality Metrics vs Contrast Level. Note: high BRISQUE, NIQE and PIQE values denote low quality while high *Q* values denote high quality. Again, a separate axis is used for *Q* here as the unnormalized *Q* value tends to be considerably higher than the other NR-IQAs, which tend to stay in the range 0 to 100.

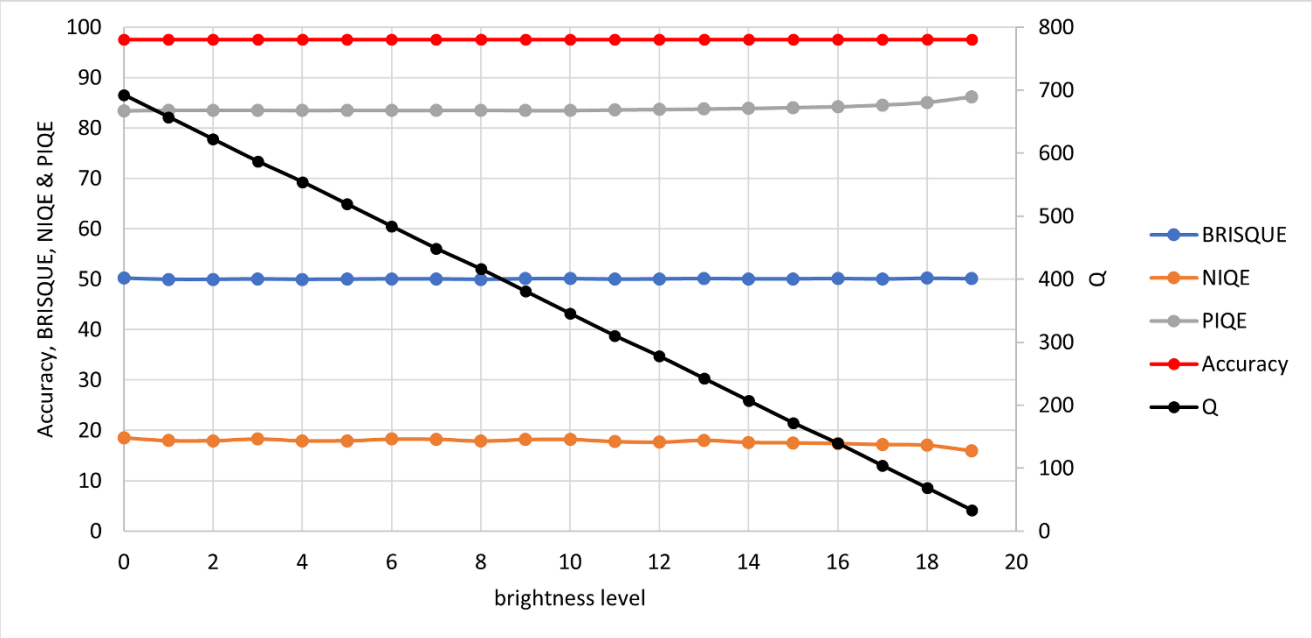


Figure 8. OCR Accuracy and Quality Metrics vs Brightness Level. Note: high BRISQUE, NIQE and PIQE values denote low quality while high *Q* values denote high quality. Again, a separate axis is used for *Q* here as the unnormalized *Q* value tends to be considerably higher than the other NR-IQAs, which tend to stay in the range 0 to 100.

These tests show SEDIQA's *Q*-metric responding strongly to both noise and blur – the most significant factors in reducing OCR accuracy – whereas its response to contrast and brightness, the changes are linear drop-offs. This means that the system will reject noisy, blurry and poorly lit images.

On an interesting side note, despite the fact that these established NR-IQAs continue to be used on text-based images, e.g. [43–45], it has been shown here that only PIQE responds correctly to these four common image degradations. This will be further

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

investigated on real camera-captured images in the next section. For useful reference, a summary of these findings is presented in **Table 1**.

Table 1. NR-IQAs and their ability to respond to image degradations in text-based images.

NR-IQA	Noise	Blur	Contrast	Brightness
BRISQUE	×	✓	×	×
PIQE	✓	✓	✓	✓
NIQE	×	×	×	×
SEDIQA	✓	✓	✓	✓

3.2 Real Camera-Captured Images

While the synthetic images allow individual degradations to be separated and examined, real camera-captured images are subject to random combinations of these degradations along with other imperfections.

To test SEDIQA's performance in real camera-captured images, the SmartDoc dataset was used as it contains images of the same documents captured under different capture conditions. It also allows comparison with other metrics and previous work, e.g. [16,32], as it is a well-established benchmark.

To demonstrate SEDIQA's performance, each version of a document was passed through the SEDIQA system. Although the system only retains a best image array based on the *Q*-metric, the full set of results is presented here for completeness. This test was repeated for each document set as well as on live captured images.

Only one document set is presented in **Figure 9** for clarity but a similar correlation between SEDIQA and OCR Accuracy is found throughout the dataset and across a variety of live capture scenarios.

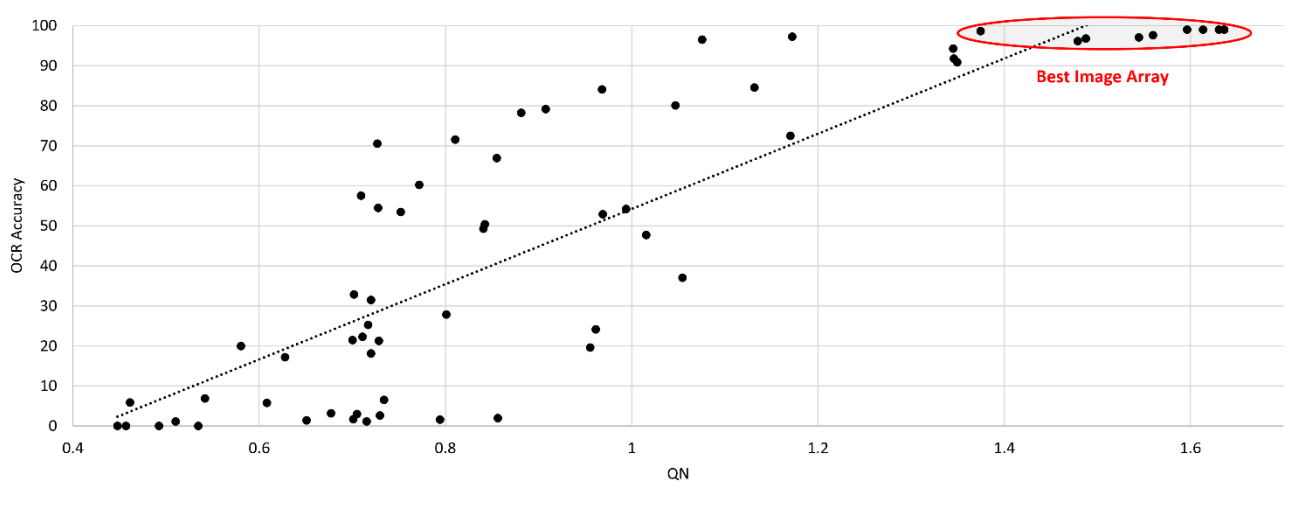


Figure 9. OCR Accuracy vs Q_N for document *D1* of the SmartDoc dataset. The system only retains those images in the Best Image Array for conversion to text but the full results across all versions of the document are presented here.

For comparison, the images cleaned by the SEDIQA system were also tested using the established NR-IQAs. Results are presented in **Figure 10**. For ease of comparison, all metrics were normalized to the range 0 to 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

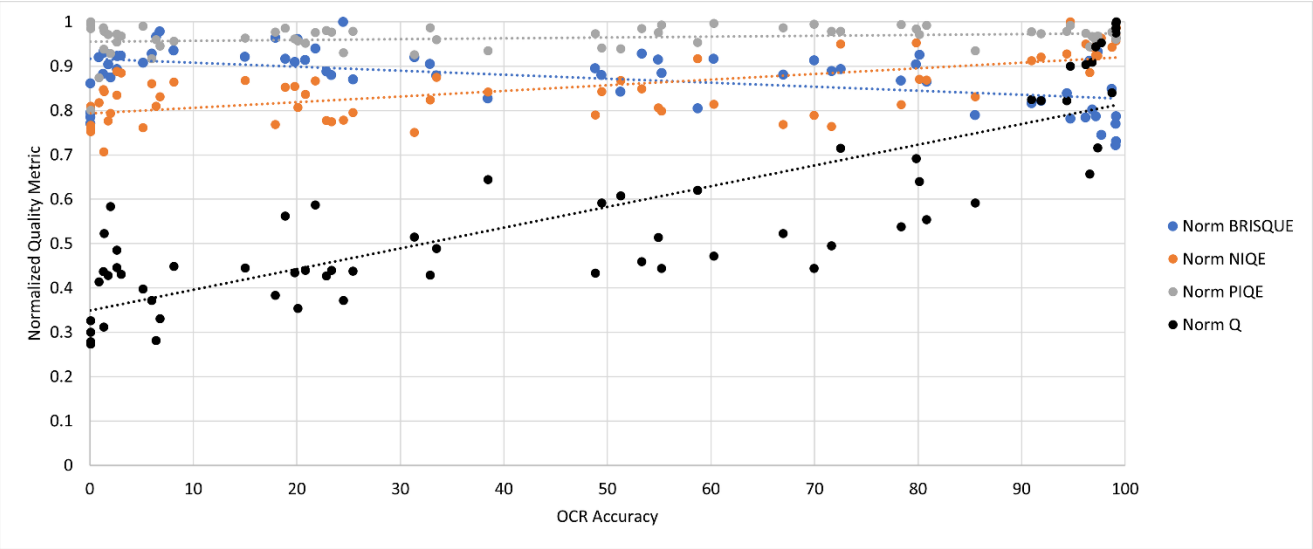


Figure 10. Normalized NR-IQAs vs OCR Accuracy for document *D1* of the SmartDoc dataset.

Despite its under-performance in the image degradation tests, BRISQUE showed some correlation with OCR Accuracy, while PIQE and NIQE showed weak and positive correlations (again, for these metrics, negative correlation is correct). The comparative correlation results are shown in **Table 2**, with SEDIQA's *Q*-metric showing the strongest correlation with OCR accuracy.

Table 2. Correlation of NR-IQAs with OCR Accuracy. Red denotes incorrect correlation.

NR-IQA	Correlation
BRISQUE	-0.5291
NIQE	0.6783
PIQE	0.2297
SEDIQA	**0.8463**

As the Best Image Array are the only candidates for text to speech conversion in the full SEDIQA system, their results are also presented in **Table 3**. Not only does the highest *Q* value correspond with the best performing image but all best array images give high accuracy.

Table 3. Normalized Q and OCR Accuracy results for the Best Image Array of document *D1*.

Image Rank	Accuracy	Norm Q
1	**99.11**	**1**
2	99.10	0.9967
3	99.08	0.9862
4	99.07	0.9757
5	97.68	0.9532
6	97.14	0.9440
7	96.80	0.9094
8	96.16	0.9039
9	94.69	0.8998
10	98.70	0.8402

Again, for comparison, the same best image selection method was performed using the other NR-IQAs, and the OCR Accuracies of their top-ranking images were tested. Results are shown in **Table 4** and confirm that, while BRISQUE shows some potential as an OCR Accuracy predictor, SEDIQA remains more robust and reliable.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0200.v1

Table 3. Accuracies for best images of document *D1* as determined by each NR-IQA. Red denotes unusably low accuracies.

	Accuracies			
Image Rank	**SEDIQA**	**BRISQUE**	**NIQE**	**PIQE**
1	**99.11**	99.08	1.32	0.07
2	99.10	**99.11**	31.34	0.86
3	99.08	97.68	0.07	5.98
4	99.07	99.07	5.13	31.34
5	97.68	0.07	0.07	2.00
6	97.14	94.69	**71.63**	24.47
7	96.80	96.16	0.07	38.44
8	96.16	0.07	17.91	**85.50**
9	94.69	97.14	66.96	1.32
10	98.70	99.10	23.31	51.25

3.3. Accuracy Improvements

As a final test of the full SEDIQA system, the OCR Accuracy of the original images was tested and compared to the accuracy of the cleaned images. The full set of results for document *D1* of the dataset are shown in **Figure 11**.

Although the system occasionally shows a decrease in accuracy, this is generally in images that are either of unusably poor quality or that do not pass the Page Extraction stage, where the image would be automatically rejected, and so these images are not passed to the Best Image Array.

The average increase in accuracy was 41% in the Best Image Array and 22% across the whole dataset, with a maximum increase for one image of 68% from just 25% accuracy in the original version to 94% in SEDIQA's cleaned version, converting it from an unusable image to a candidate for the Best Image Array.

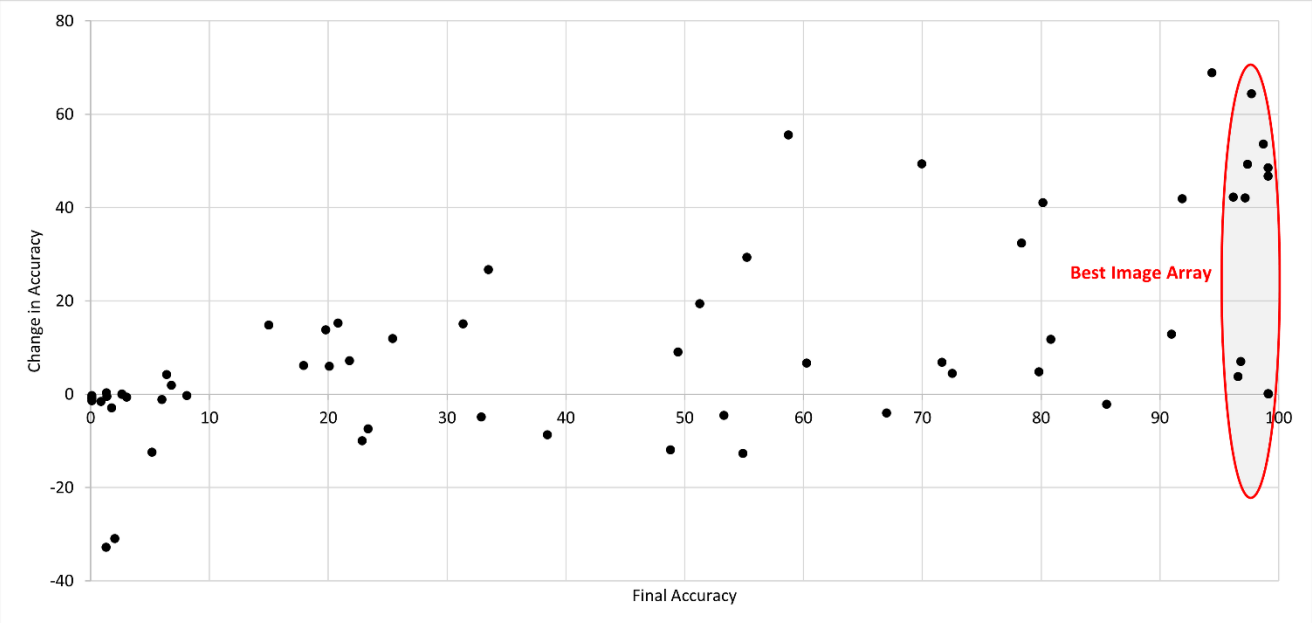


Figure 11. SEDIQA's effect on OCR Accuracy for document *D1* of the SmartDoc dataset. Again, the system only retains those images in the Best Image Array for conversion to text but the full results across all versions of the document are presented here.

 doi:10.20944/preprints202107.0200.v1

4. Discussion

Within this audio-based reading aid design lies some significant contributions to the field of document image quality assessment. First is a simple, robust and directly measurable NR-IQA for documents which picks up on four major sources of OCR accuracy reduction: noise, blur, contrast and brightness. This *Q*-metric has been validated by comparing to OCR accuracy, image degradations and three other established NR-IQAs: BRISQUE, PIQE and NIQE. This Q-metric shows the strongest correlation with OCR accuracy and correctly selects the highest performing image as the best image from a large dataset of images of different qualities.

As part of the validation process, some interesting discoveries were also made about the established NR-IQAs. First, NIQE was shown to neither respond to typical image degradations nor correlate with OCR accuracy. This is not necessarily surprising as it is intended for natural images, yet it continues to be used in text-based images intended for OCR [44,46]. PIQE showed good responses to image degradations but in real document images had a reverse correlation with OCR accuracy, selecting some of the worst performing images. BRISQUE only responded to blur out of the degradations tested but did show some promise in real images where it showed some correlation with OCR accuracy. However, it was not robust and gave high quality scores to images which completely fail at conversion to text. Still, the combination of these findings about BRISQUE would suggest that blur may be the biggest contributor to OCR accuracy reduction.

As well as the robust metric, SEDIQA includes a text detection, extraction and cleaning process that leads to significant improvements in accuracy in even some of the poorest performing images. Across the entire SmartDoc dataset, the rejection rate at the page extraction stage was approximately 10% with almost 90% of images successfully cropped, warped and cleaned.

There are some minor limitations to be addressed in future work. This system has not yet been ported to a smartphone app and as such, has not yet been tested for speed, user experience (UX) or compatibility. However, these initial tests and the simplicity of the metric suggest that the design has significant potential.

SEDIQA could also be applied to camera focusing systems, text-in-the-wild applications (such as sign reading in autonomous cars or navigation aids), and any other text-based applications, particularly those involving OCR. As a Reading Aid, SEDIQA offers much needed audio guidance, to aid in document capture. Although other reading aids, such as Microsoft Seeing AI, offer some audio guidance in the text location stage, these do not assess, or feedback to the user, the quality of the image, and as such, frequently lead to unsatisfactory results. As can be seen here, successful text detection does not necessarily mean successful OCR. With the addition of SEDIQA's robust *Q*-metric to the reading aid design, the user can sure of the best possible outcome from any given capture scenario.

Funding: This research received no external funding.

Acknowledgments: I would like to thank Claire Chambers, Susan McKeever and Thomas Lee for their valuable insights.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dockery, D.; Krzystolik, M. The Evaluation of Mobile Applications as Low Vision Aids: The Patient Perspective. *Invest. Ophthalmol. Vis. Sci.* **2020**, *61*, 935–935.
2. Akkara, J.D.; Kuriakose, A. Commentary: An App a Day Keeps the Eye Doctor Busy. *Indian J. Ophthalmol.* **2021**, *69*, 553.
3. El-taher, F.E.; Taha, A.; Courtney, J.; Mckeever, S. A Systematic Review of Urban Navigation Systems for Visually Impaired People. *Sensors* **2021**, *21*, 3103, doi:10.3390/s21093103.

4. Hisham, Z.A.N.; Faudzi, M.A.; Ghapar, A.A.; Rahim, F.A. A Systematic Literature Review of the Mobile Application for Object Recognition for Visually Impaired People. In Proceedings of the 2020 8th International Conference on Information Technology and Multimedia (ICIMU); IEEE, 2020; pp. 316–322.
5. Jiang, H.; Gonnot, T.; Yi, W.-J.; Saniie, J. Computer Vision and Text Recognition for Assisting Visually Impaired People Using Android Smartphone. In Proceedings of the 2017 IEEE International Conference on Electro Information Technology (EIT); IEEE, 2017; pp. 350–353.
6. Peng, X.; Wang, C. Building Super-Resolution Image Generator for OCR Accuracy Improvement. In Proceedings of the Document Analysis Systems; Bai, X., Karatzas, D., Lopresti, D., Eds.; Springer International Publishing: Cham, 2020; pp. 145–160.
7. Brzeski, A.; Grinholc, K.; Nowodworski, K.; Przybyłek, A. Evaluating Performance and Accuracy Improvements for Attention-OCR. In Proceedings of the Computer Information Systems and Industrial Management; Saeed, K., Chaki, R., Janev, V., Eds.; Springer International Publishing: Cham, 2019; pp. 3–11.
8. Zhai, G.; Min, X. Perceptual Image Quality Assessment: A Survey. *Sci. China Inf. Sci.* **2020**, *63*, 211301.
9. Ye, P.; Doermann, D. Document Image Quality Assessment: A Brief Survey. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition; August 2013; pp. 723–727.
10. Courtney, J. CleanPage: Fast and Clean Document and Whiteboard Capture. *J. Imaging* **2020**, *6*, 102.
11. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708, doi:10.1109/TIP.2012.2214050.
12. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212, doi:10.1109/LSP.2012.2227726.
13. Chan, R.W.; Goldsmith, P.B. A Psychovisually-Based Image Quality Evaluator for JPEG Images. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics; October 2000; Vol. 2, pp. 1541–1546 vol.2.
14. Alaei, A. A New Document Image Quality Assessment Method Based on Hast Derivations. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2019; pp. 1244–1249.
15. Kumar, J.; Chen, F.; Doermann, D. Sharpness Estimation for Document and Scene Images. In Proceedings of the Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012); IEEE, 2012; pp. 3292–3295.
16. Asad, F.; Ul-Hasan, A.; Shafait, F.; Dengel, A. High Performance OCR for Camera-Captured Blurred Documents with LSTM Networks. In Proceedings of the 2016 12th IAPR Workshop on Document Analysis Systems (DAS); April 2016; pp. 7–12.
17. Yang, H.; Fang, Y.; Lin, W. Perceptual Quality Assessment of Screen Content Images. *IEEE Trans. Image Process.* **2015**, *24*, 4408–4421, doi:10.1109/TIP.2015.2465145.
18. Shahkolaei, A.; Nafchi, H.Z.; Al-Maadeed, S.; Cheriet, M. Subjective and Objective Quality Assessment of Degraded Document Images. *J. Cult. Herit.* **2018**, *30*, 199–209, doi:10.1016/j.culher.2017.10.001.
19. Shahkolaei, A.; Beghdadi, A.; Cheriet, M. Blind Quality Assessment Metric and Degradation Classification for Degraded Document Images. *Signal Process. Image Commun.* **2019**, *76*, 11–21, doi:10.1016/j.image.2019.04.009.
20. Peng, X.; Wang, C. Camera Captured DIQA with Linearity and Monotonicity Constraints. In Proceedings of the Document Analysis Systems; Bai, X., Karatzas, D., Lopresti, D., Eds.; Springer International Publishing: Cham, 2020; pp. 168–181.
21. Gu, K.; Zhai, G.; Lin, W.; Yang, X.; Zhang, W. Learning a Blind Quality Evaluation Engine of Screen Content Images. *Neurocomputing* **2016**, *196*, 140–149, doi:10.1016/j.neucom.2015.11.101.
22. Li, H.; Qiu, J.; Zhu, F. TextNet for Text-Related Image Quality Assessment. In Proceedings of the Artificial Neural Networks and Machine Learning – ICANN 2018; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Springer International Publishing: Cham, 2018; pp. 275–285.
23. Lu, T.; Dooms, A. A Deep Transfer Learning Approach to Document Image Quality Assessment. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2019; pp. 1372–1377.

24. Qian, J.; Tang, L.; Jakhetiya, V.; Xia, Z.; Gu, K.; Lu, H. Towards Efficient Blind Quality Evaluation of Screen Content Images Based on Edge-Preserving Filter. *Electron. Lett.* **2017**, *53*, 592–594, doi:https://doi.org/10.1049/el.2017.0325.
25. Yang, J.; Zhao, Y.; Liu, J.; Jiang, B.; Meng, Q.; Lu, W.; Gao, X. No Reference Quality Assessment for Screen Content Images Using Stacked Autoencoders in Pictorial and Textual Regions. *IEEE Trans. Cybern.* **2020**, 1–13, doi:10.1109/TCYB.2020.3024627.
26. Shao, F.; Gao, Y.; Li, F.; Jiang, G. Toward a Blind Quality Predictor for Screen Content Images. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *48*, 1521–1530, doi:10.1109/TSMC.2017.2676180.
27. Zheng, L.; Shen, L.; Chen, J.; An, P.; Luo, J. No-Reference Quality Assessment for Screen Content Images Based on Hybrid Region Features Fusion. *IEEE Trans. Multimed.* **2019**, *21*, 2057–2070, doi:10.1109/TMM.2019.2894939.
28. Alaei, A.; Conte, D.; Raveaux, R. Document Image Quality Assessment Based on Improved Gradient Magnitude Similarity Deviation. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2015; pp. 176–180.
29. Li, H.; Zhu, F.; Qiu, J. CG-DIQA: No-Reference Document Image Quality Assessment Based on Character Gradient. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR); IEEE, 2018; pp. 3622–3626.
30. Burie, J.-C.; Chazalon, J.; Coustaty, M.; Eskenazi, S.; Luqman, M.M.; Mehri, M.; Nayef, N.; Ogier, J.-M.; Prum, S.; Rusiñol, M. ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc). In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2015; pp. 1161–1165.
31. Chazalon, J.; Gomez-Krämer, P.; Burie, J.-C.; Coustaty, M.; Eskenazi, S.; Luqman, M.; Nayef, N.; Rusiñol, M.; Sidère, N.; Ogier, J.-M. SmartDoc 2017 Video Capture: Mobile Document Acquisition in Video Mode. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR); November 2017; Vol. 04, pp. 11–16.
32. Javed, K.; Shafait, F. Real-Time Document Localization in Natural Images by Recursive Application of a Cnn. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2017; Vol. 1, pp. 105–110.
33. Smith, R. An Overview of the Tesseract OCR Engine. In Proceedings of the Ninth international conference on document analysis and recognition (ICDAR 2007); IEEE, 2007; Vol. 2, pp. 629–633.
34. Nayef, N.; Luqman, M.M.; Prum, S.; Eskenazi, S.; Chazalon, J.; Ogier, J.-M. SmartDoc-QA: A Dataset for Quality Assessment of Smartphone Captured Document Images-Single and Multiple Distortions. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2015; pp. 1231–1235.
35. Lundqvist, F.; Wallberg, O. *Natural Image Distortions and Optical Character Recognition Accuracy*; 2016;
36. Kolli, A. A Comprehensive Study of the Influence of Distortions on the Performance of Convolutional Neural Networks Based Recognition of MNIST Digit Images. PhD Thesis, Alpen-Adria-Universität Klagenfurt, 2019.
37. Mustafa, W.A.; Abdul Kader, M.M.M. Binarization of Document Images: A Comprehensive Review. *J. Phys. Conf. Ser.* **2018**, *1019*, 012023, doi:10.1088/1742-6596/1019/1/012023.
38. Wellner, P. Interacting with Paper on the DigitalDesk. *Commun. ACM* **1993**, *36*, 87–96.
39. Haralick, R.M.; Sternberg, S.R.; Zhuang, X. Image Analysis Using Mathematical Morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, 532–550.
40. Zhou, X.; Yao, C.; Wen, H.; Wang, Y.; Zhou, S.; He, W.; Liang, J. East: An Efficient and Accurate Scene Text Detector. In Proceedings of the Proceedings of the IEEE conference on Computer Vision and Pattern Recognition; 2017; pp. 5551–5560.
41. Barnard, S.T. Interpreting Perspective Images. *Artif. Intell.* **1983**, *21*, 435–462.
42. Vitz, P.C. Preference for Tones as a Function of Frequency (Hertz) and Intensity (Decibels). *Percept. Psychophys.* **1972**, *11*, 84–88, doi:10.3758/BF03212689.
43. Khare, V.; Shivakumara, P.; Raveendran, P.; Blumenstein, M. A Blind Deconvolution Model for Scene Text Detection and Recognition in Video. *Pattern Recognit.* **2016**, *54*, 128–148, doi:10.1016/j.patcog.2016.01.008.
44. Xue, M.; Shivakumara, P.; Zhang, C.; Xiao, Y.; Lu, T.; Pal, U.; Lopresti, D.; Yang, Z. Arbitrarily-Oriented Text Detection in Low Light Natural Scene Images. *IEEE Trans. Multimed.* **2020**.

45. Thanh, D.N.H.; Prasath, V.S. Adaptive Texts Deconvolution Method for Real Natural Images. In Proceedings of the 2019 25th Asia-Pacific Conference on Communications (APCC); IEEE, 2019; pp. 110–115.
46. Nakao, R.; Iwana, B.K.; Uchida, S. Selective Super-Resolution for Scene Text Images. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR); IEEE, 2019; pp. 401–406.