



**HAL**  
open science

## ”Quality” vs. ”Readability” in Document Images: Statistical Analysis of Human Perception

Alireza Alaei, Romain Raveaux, Donatello Conte, Bela Stantic

► **To cite this version:**

Alireza Alaei, Romain Raveaux, Donatello Conte, Bela Stantic. ”Quality” vs. ”Readability” in Document Images: Statistical Analysis of Human Perception. 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Apr 2018, Vienna, Austria. 10.1109/DAS.2018.42 . hal-01879782

**HAL Id: hal-01879782**

**<https://hal.science/hal-01879782>**

Submitted on 24 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# “Quality” vs. “Readability” in Document Images: Statistical Analysis of Human Perception

Alireza Alaei<sup>1</sup>, Romain Raveaux<sup>2</sup>, Donatello Conte<sup>2</sup>, Bela Stantic<sup>1,3</sup>

<sup>1</sup>School of Information and Communication Technology, Griffith University, Australia

<sup>2</sup>Laboratoire d'Informatique (LI EA6300), Université François-Rabelais de Tours, France

<sup>3</sup>Institute for Integrated and Intelligent Systems, Griffith University, Australia

{a.alaei, b.stantic}@griffith.edu.au, {romain.raveaux, donatello.conte}@univ-tours.fr

**Abstract**—Based on the hypothesis that a good / poor quality document image is most probably a readable / unreadable document, document image quality and readability have interchangeably been used in the literature. These two terms, however, have different meanings implying two different perspectives of looking at document images by human being. In document images, the level of quality and the degree of readability may have a relation / correlation considering human perception. However, to the best of our knowledge there is no specific study to characterise this relation and also validate the abovementioned hypothesis. In this work, at first, we created a dataset composed of mostly camera-based document images with various distortion levels. Each document image has then been assessed with regard to two different measures, the level of quality and the degree of readability, by different individuals. A detailed Normalised Cross Correlation analysis along with different statistical analysis based on Shapiro-Wilks and Wilcoxon tests has further been provided to demonstrate how document image quality and readability are linked. Our findings indicate that the quality and readability were somewhat different in terms of the population distributions. However, the correlation between quality and readability was 0.99, which implies document quality and readability are highly correlated based on human perception.

**Keywords:** Document image analysis; Quality and readability; Human visual system; Shapiro-Wilks test; Wilcoxon test.

## I. INTRODUCTION

Image quality is a set of perceptions, which represent the overall degree of excellence of an image observed by individuals. Image quality assessment has widely been used in the literature of natural image analysis to measure the perceived image degradation, especially compared to that of a perfect image. Assessing the quality of an image by individuals based on human perception is a subjective process. However, subjective image quality assessment by human is a time consuming process and impractical in real-time applications. On the other hand, many objective image quality assessment (IQA) metrics have been introduced in the literature [1-4] to work as good as human being handling a huge size of digital scene images generated in day-to-day life.

Document image quality assessment (DIQA) is, however, different from the natural image quality assessment as document images contain textual information in addition to graphical information present in natural scene images. Technicians, officers or bankers should be able to

recognize text or courtesy values, to verify signatures, to recognize a person on a photo or to identify stamps or logos, etc. in document images. Therefore, researchers have further used the term “document image readability” in the field of document image analysis to assess document images quality. This indicates that the readability of document images is important in document image analysis domain and can be assessed by human being or machine [22].

In relation to the human readability, document images are presented to individuals and asked them to look at the documents, read and assess their readabilities, whereas as in relation to the machine readability, document images are generally subjected to an optical character recognition (OCR) process and the accuracy obtained from OCR is considered as degree of machine readability of the documents [22]. In both human and machine readability, poor quality document images might result in low document image readability (documents are unreadable) [17-23].

In the field of document image analysis, many objective document image quality assessment metrics have been proposed in the literature [5-16] that can be grouped into: i) OCR-based [5-9], and ii) human perception-based approaches [11-16]. In the OCR-based (machine operator) methods, the results of an OCR have commonly been used to quantify the quality of a document image [5-9].

In the OCR-based category, character-based features have commonly been used to estimate the quality of document images. The metrics obtained as the quality of the document images were then associated with the results obtained from OCR to compute correlation values in order to evaluate the performances of the proposed methods in the literature [5-9, 23]. High correlation values indicate that the performances of the proposed methods / metrics are commendable.

On the other hand, the methods presented in [11-16] used subjective mean human opinion scores (MHOS) as the basis of document image quality for performance evaluation. Different features characterizing textural, visual and structural appearance of document images have been employed by different methods in the literature to estimate quality of the document images [11-16].

In addition to the methods in the literature that provide only a metric as the quality of a document image, there are also many studies which propose employing different pre-processing and filtering techniques in order to enhance the quality of document images (especially historical and degraded ones), results in improving the readability of those

documents [17-24]. Different pre-processing and filtering operations [17, 18, 19], binarization methods [20], and quality assessment followed by filtering operations [21, 22, 23] have been used to improve document image quality and readability. Eye tracking technology has also been used to aggregate reading behaviour of many readers in order to provide objective quality feedback [24]. In most of the studies in the literature [17-20, 22] human visual inspection has been used to determine the level of enhancement in relation to document image readability.

As different levels of image quality may result in different degrees of image readability, image readability and image quality have been used in the literature interchangeably [17-23]. However, we noted that a document image, which can be read by individuals, may not always be readable by OCR. In Figure 1(a) a French document image with a poor quality, but still readable by individuals is demonstrated, whereas erroneous result obtained from OCR is shown in Figure 1(b). This indicates the gap between current OCR technology (machine reading) and human reading capability, as humans are capable of utilizing more contextual information in document images compared to current OCR technology. The gap between OCR technology and human reading capability is even more when handwritten documents are considered for evaluation. Humans are also more tolerant in relation to noise and document image degradation [22]. Indeed, the level of document image quality and the degree of readability based on human perception may have a certain relation / correlation. However to the best of our knowledge, there is no specific study to demonstrate and thoroughly analyse the dependency of document images “quality” and “readability” based on human perception. In addition, there are many document image analysis applications such as digital libraries keeping a mass of documents in their original shape and providing a search facility using word spotting techniques that do not use / need OCR technology in their entire processes. In such a scenario, a human perception based document image readability would be the best solution to keep / transmit the document images with a certain quality for any further uses and processes.

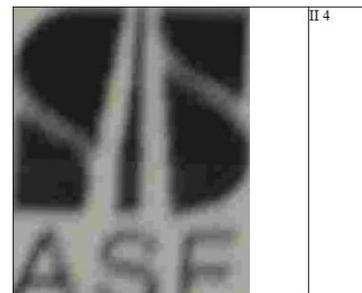
To address the above-mentioned issue, in this research work a dataset composed of document images with various quality levels extracted from different sources has initially been created. Two different human perception metrics as the level of quality and the degree of readability have been provided by individuals. A detailed Normalised Cross Correlation (NCC) analysis along with different statistical analysis based on Shapiro-Wilks and Wilcoxon tests has further been provided to demonstrate how document image quality and readability are related. The reliability of the hypothesis “a good / poor quality document image is most probably a readable / unreadable document” has finally been examined. Our findings indicate that document quality and readability are somewhat different in terms of the population distributions. However, correlation between quality and readability was high, which implies quality and readability are highly correlated based on human perception.

The rest of the paper is laid out as follows: In Section II, a brief description of the dataset as well as its quality and readability ground truths are provided. Section III describes the choice of statistical tests used in our research work.

Section IV discusses the statistical results and findings. Finally, Section V provides some conclusions and future work.



(a)



(b)

Figure 1. a) A document image of poor quality but readable by individuals, and b) Erroneous OCR result obtained for the image shown in Figure 1(a).

## II. DATASET DESCRIPTION

To the best of our knowledge, no document-oriented dataset considering human opinion scores as ground truth for both quality and readability assessment is publicly available in the literature. Therefore, a document dataset called ITESOFT dataset was created in this research work to study the different aspects of document image quality and readability in relation to the human perception. The ITESOFT dataset is composed of 29 document images (mostly printed bills and invoices) collected from real-world data using different capturing devices, such as mobile cameras and scanners. Generally, four common type of distortions, such as JPEG compression, JPEG2000 compression, Gaussian noise, and Gaussian blur distortions,

were used in the literature to create distorted images. Since in our investigation the main concern was the distortions occurred when applying different compression techniques to reduce the document image sizes for image transmission, we considered JPEG and JPEG2000 (JPEG2K) compression techniques to create degraded document images. Six different levels of compression for each (JPEG and JPEG2K) were applied on the reference images to generate 348 ( $29 \times 2 \times 6$ ) distorted images. As a result, 377 ( $348 + 29$ ) document images in total were gathered in the ITESOFT dataset.

A Mean Human Opinion Score (MHOS) has been computed for each document image based on the Human Opinion Scores (HOSs) obtained from 23 observers in relation to the quality of the document. Readability of each document image has also been provided using HOSs collected from the same 23 observers with respect to the readability of the document. Observers were undergraduates, PhD candidates and faculties from Computer Science Laboratory of University of Tours, France. Observers had normal or corrected to normal vision. They were naive to the purpose of the experiment and had not previously seen the documents. The images were shown to the observers in a random order on an 18 inches monitor. Observers had a distance of approximately 70cm from the monitor. The observers could not change the size of the document images shown in the monitor. In fact, a simple software was designed to demonstrate the document images on the monitor and collect the quality and readability scores from the observers. Each individual could complete the evaluation in one or more session using the same device (monitor) to demonstrate all the document images in an entire evaluation process. In our experiments for obtaining the HOSs, each individual was asked to evaluate the quality and readability of a document image between 0 and 100 according to the following main question: "How would you rate the overall degree of 'excellence' and 'useability' of the document, if the document needs to be 'read' and 'reused' by human operators?"

Some detailed sub-instructions for quality ( $Q_i$ ) and Readability ( $R_j$ ) were also provided to further help the individuals for a better evaluation. The sub-instructions were:

- "Q<sub>1</sub>: Is the image still informative or too distorted to be recognized by a person?",
- "Q<sub>2</sub>: Can you evaluate the quality of the information conveyed by the document?",
- "R<sub>1</sub>: Can the textual part be read easily (without considerable effort) by an individual?",
- "R<sub>2</sub>: Are logos or stamps identifiable by an individual?",
- "R<sub>3</sub>: Is the document still meaningful?".

A summary of the ITESOFT dataset including the size of the samples according to distortions and document types are shown in Table I. As shown in Table I, we have chosen two specific types of documents called Bill and Invoice captured images to provide a more detailed analysis of data based on the type of document and capturing device.

TABLE I. DATASET AND SAMPLE SIZES

Document \ Sample	Artifact			
	Entire dataset	Original images	JPEG	JPEG2K
Readability (R)	377	29	174	174
Quality (Q)	377	29	174	174

### III. SELECTING STATISTICAL TESTS

In statistics, a paired test is a type of location test used when comparing two sets of measurements to assess whether their populations are different or not. Moreover, a paired test is generally used when comparing two related samples, matched samples, or repeated measurements on a single population. In addition, in the case of dependent samples a paired test is more powerful than a 2-sample test to compare the independence of the samples. A paired test is not also subjected to the variation because the paired observations are dependent on the sense that they come from the same population observed through two viewpoints of readability and quality which do not necessarily have the same variance. As in this research work, our data are paired and come from the same population with different point of views, we used different paired tests with various parameters for experimentation and statistical analysis. Various statistical tests have been designed [26, 27, 30] and employed in different contexts [25, 28]. In this work, we have used different statistical tests including normalised cross correlation (NCC) [30], paired Wilcoxon test [26] and Shapiro-Wilk test [27]. The NCC demonstrates the degree of correlation between quality and readability, whereas Wilcoxon [26] and Shapiro-Wilk tests provide us with the idea of the distribution of data in the quality and readability dataset. These statistical tests are well suited to find out in a scientific manner whether or not quality and readability are equivalent. The Kolmogorov-Smirnov and chi-square tests were not used in our experimentation, since paired data cannot be addressed with these tests. Furthermore, Kolmogorov-Smirnov and chi-square tests require the elements to be independent for comparison [29], whereas this was not the case in our data.

#### A. Population and samples

Population and samples are two main parameters in every statistical test. In our dataset, the population is the number of images considered in each statistical test. Since Readability ( $R_i$ ) and Quality ( $Q_i$ ) of each image have been evaluated by individuals,  $Q$  and  $R$  are considered as two samples in all experimental statistical analysis in this work.

#### B. What do we want to test?

A summary of different paired tests, their hypotheses and characteristics considered in this research work is demonstrated in Table II. Three different paired tests are employed to analyse different aspects of data ( $Q$  and  $R$ ). The correlation, as a single value, is one of the most common statistics used in the literature to quantify the degree of relationship between two sets of values [30]. Here we used NCC to examine to which extend document image quality is correlated with document image readability.

The Wilcoxon test is a non-parametric statistical test and distribution free test designed to test the median equality of two dependent samples [26]. The Wilcoxon test is solely based on the order in which the observation from the two samples fall. This test is also called a general normality test designed to detect all the departures from normality. This test assesses whether the population mean ranks of two samples differ. Therefore, the null hypothesis (H0) is defined as “the medians of the two samples ( $Q$  and  $R$ ), which are defined by  $\tilde{Q}$  and  $\tilde{R}$ , are identical”.

The Shapiro-Wilks test (SW) [27] is a parametric test which allow us to prove the distribution of the samples are of a Gaussian distribution. Since, this test is also a paired test, the difference between  $Q$  and  $R$  ( $Q - R$ ) is considered for the analysis. The null hypothesis (H0) of this test is defined as the population of the differences ( $Q - R$ ) should be normally distributed. Alternative hypothesis (H1) is therefore defined as the population are not normally distributed.

TABLE II. TESTS CHARACTERISTICS

Hypothesis Test	Null hypothesis (H0)	Alternative hypothesis (H1)	Paired	Parametric
Normalised Cross Correlation (NCC)	Correlation is equal to 0	Correlation is not equal to 0	Yes	No
Wilcoxon signed-rank test (W)	$\tilde{Q} = \tilde{R}$	$\tilde{Q} \neq \tilde{R}$	Yes	No
Shapiro-Wilks test (SW)	$Q - R$ population is normally distributed	$Q - R$ population is not normally distributed	Yes	Yes

### C. How to make a decision?

The p-value is defined as the probability of obtaining a result equal to or "more extreme" than what was actually observed, when the null hypothesis is true. It is the error of the second kind, to accept the null hypothesis while H1 is true (false positive). If the p-value is less than a chosen alpha level (usually 5%), then the null hypothesis (H0) is rejected. On the contrary, if the p-value is greater than the chosen alpha level, then the null hypothesis (H0) cannot be rejected. In any case, it never leads to accept H1.

## IV. CORRELATION AND TEST RESULTS

The correlation values obtained on different subsets of database are shown in Table III. From Table III, it can be noted that the H0 (correlation is equal to 0) with 95% confidence (p-value=5%) can be rejected in all the experiments. This means that there is a high correlation between readability and quality in document images based on human perception assessment. For example, the correlation between readability and quality was more than 0.99 with a high significance (p-value) of 2.2e-16 when considering all the document images scores for experiment. From the results shown in Table III, it can further be concluded that in the JPEG2K compressed document images readability and quality are more correlated than that of the JPEG compressed ones. In fact, this is due to the specific design of the JPEG2K compression algorithm that keeps intact critical details compared to the JPEG compression algorithm.

TABLE III. CORRELATIONS OBTAINED EMPLOYING THE NCC ON DIFFERENT SUNSETS

Document Test	Artifact		
	Entire dataset	JPEG	JPEG2K
Correlation	0.9905	0.9856	0.9954
$p - value$	2.2e-16	2.2e-16	2.2e-16
5% risk level	H0 can be rejected	H0 can be rejected	H0 can be rejected

To understand the details of the relationship between quality and readability measures, the values of these measures for the entire ITESOFT dataset, the JPEG and JPEG2K distortions are plotted in Figures 2, 3 and 4, respectively. From the scatter plots shown in Figure 2, 3 and 4, it can be noted that for low and high values of quality and readability, the correlations are more evident. This means that when an image is globally poor both the readability and quality scores are low, and equally when an image is globally in a good condition both measures are highly correlated. This consideration may guide to build an automatic system to infer the readability from the quality (and vice versa) for a certain type of images, specifically for visually very poor or very good document images. Figure 3 further demonstrates that in JPEG document images the deviation of the readability and quality with respect to the fitted line is more significant compared to the JPEG2K document readability and quality measures shown in Figure 4. This is in line with the results reported in Table III and proves that readability and quality are more associated in JPEG2K documents compared to the readability and quality in JPEG documents.

The results obtained by employing the Wilcoxon and Shapiro-Wilk tests on the paired data are shown in Tables IV, and V, respectively. From the results shown in Table IV, it is clear that the Null hypothesis (H0) at 5% risk level is rejected in all the experiments. This indicates that the quality and readability do not have exactly the same population distribution, as the medians of the two group of samples are not the same.

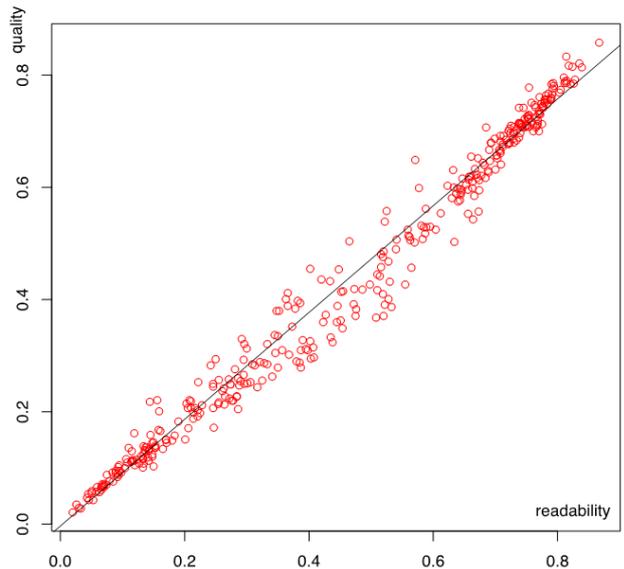


Figure 2. Scatter plot of the readability vs quality considering all document images in the dataset

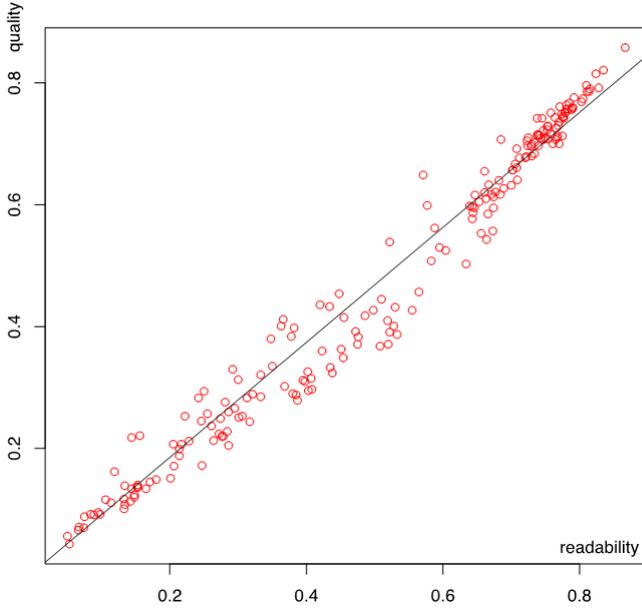


Figure 3. Scatter plot of the readability vs quality considering the document images with JPEG artifact

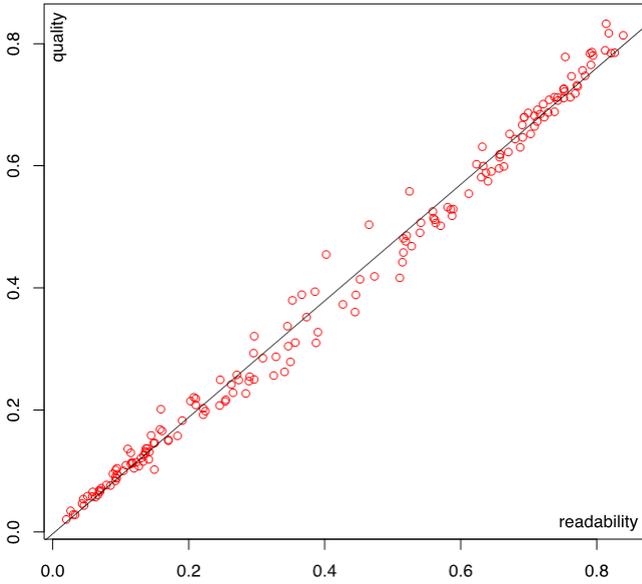


Figure 4. Scatter plot of the readability vs quality considering the document images with JPEG2K artifact

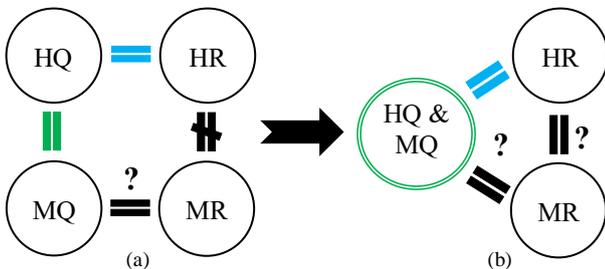


Figure 5. Relation between quality and readability in relation to machine technology and human perception: HQ = Human Quality; HR = Human Readability; MQ = Machine Quality; MR = Machine Readability

- ≡ : Two concepts are equal. ≡ : Machine performance is very close to human.
- ≡? : Relation between two concepts is not established.
- ≠ : There is a technological gap between human and machine readability.
- : Concepts are merged, as machines provide quality scores close to human scores.

From the results obtained based on the Shapiro-Wilk test (Table V), we can conclude that the Null hypothesis ( $H_0$ ) on the normality of the samples can be rejected in the JPEG subset and entire database, but not in the JPEG2K subset. These results reveal that again the distributions of quality scores and readability scores are not exactly the same Gaussian normal distribution. As a result, the statistical Wilcoxon and Shapiro-Wilk tests could prove that the readability and quality do not have exactly the same sample distributions. This means despite the existence of a high correlation between document image readability and quality, starting from one of the two measures it is difficult to infer directly the other.

Experimental analysis provided in this paper reveals that the document image quality and readability based on human perception, as two metrics have a high correlation, can be used interchangeably, keeping in mind that their distribution might be different (Figure 5(a)). Moreover, there are automatic document image quality assessment methods with high performance in the literature [14, 15, 16, 1] that can replace almost manually human inspection for quality assessment, so these two are merged in Figure 5(b). There is a technological gap between human readability and machine readability that needs more research work to implement an OCR can perform at human level. In addition, relation between these two concepts (human readability and machine readability) has not been established yet. Relation between human/machine quality assessment and machine readability also needs to be investigated.

TABLE IV. STATISTICS OBTAINED EMPLOYING THE WILCOXON TEST ON DIFFERENT DATA

Document	Artifact		
	Entire dataset	JPEG	JPEG2K
Parameter			
$p$ - value	2.2e-16	2.2e-16	2.2e-16
5% risk level	$H_0$ can be rejected	$H_0$ can be rejected	$H_0$ can be rejected

TABLE V. STATISTICS OBTAINED EMPLOYING THE SHAPIRO-WILK TEST ON DIFFERENT DATA

Document	Artifact		
	Entire dataset	JPEG	JPEG2K
Parameter			
$p$ - value	9.e-05	0.0083	0.4389
5% risk level	$H_0$ can be rejected	$H_0$ can be rejected	$H_0$ cannot be rejected

## V. CONCLUSION AND FUTURE WORK

In this paper, a dataset of document images along with two different scores for the readability and quality of each document image have been introduced. Different statistical tests, such as Normalised Cross Correlation (NCC), Shapiro-Wilks test and Wilcoxon test, have further been applied to demonstrate how document image quality and readability are related. From the results obtained from the test (NCC), it can be noted that the quality and readability scores collected based on human perception were highly correlated. The results obtained from the Shapiro-Wilks and

Wilcoxon tests considering 5% risk level reveal that readability and quality had non-identical sample distributions as their medians and Gaussian distributions were different. As a result, these findings clearly specify that starting from the quality or readability measure, one may not directly infer the other one. It is also worth mention that document image readability and quality are also application / domain dependent and the Quality / readability in handwritten documents may be different compared with printed documents.

In future, we plan to compare human readability and machine readability (OCR) in printed as well as in handwritten document images. Furthermore, finding the relation between machine quality score and machine readability would be also a great problem to solve. We further plan to investigate providing specific metrics to indicate the minimum qualities required in printed and handwritten document images to be comfortably read by human beings.

#### ACKNOWLEDGMENT

We would like to thank Dr Christophe Lenté and Dr Mohand Slimane from Université François-Rabelais de Tours for their helpful suggestions and discussions.

#### REFERENCES

- [1] K Ma, W. Liu, T. Liu, Z. Wang, D. Tao, "dipiQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs," *IEEE Transactions on Image Processing*, 26(8), pp. 3951-3964, 2017.
- [2] A. Mittal, G. S. Muralidhar, J. Ghosh, A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Processing Letters*, 19(2), pp. 75-78, 2012.
- [3] L. Zhang, L. Zhang, X. Mou, D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, 20(8), pp. 2378-2386, 2011.
- [4] Q. Li, Z. Wang, "Reduced-reference image quality assessment Using divisive normalization-based image representation," *IEEE Journal of Selected Topics in Signal Processing*, 3(2), pp. 202-211, 2009.
- [5] C. Hale, E. H. Barney Smith, "Human image preference and document degradation models," in *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR)*, pp. 257-261, 2007.
- [6] P. Ye, D. Doermann, "Learning features for predicting OCR accuracy," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, pp. 3204-3207, 2012.
- [7] N. Nayef, J.-M. Ogier, "Metric-based no-reference quality assessment of heterogeneous document images," in *Proceedings of the SPIE- Document Recognition and Retrieval (DRR) XXII*, 2015, doi:10.1117/12.2076150.
- [8] L. Kang, P. Ye, Y. Li, D. Doermann "A Deep Learning Approach to Document Image Quality Assessment," In *Proceedings of the International Conference on Image Processing (ICIP)*, pp. 2570-2574, 2014.
- [9] T. K. Bhowmik, T. Paquet, N. Ragot, "OCR Performance Prediction using a Bag of Allographs and Support Vector Regression," In *Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 202-206, 2014.
- [10] P. Ye, D. Doermann, "Document image quality assessment: a brief survey," in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 723-727, 2013.
- [11] H. Lu, A. C. Kot, Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images," *IEEE Signal Processing Letters*, 11(2), pp. 228-231, 2004.
- [12] J. Kumar, F. Chen, D. Doermann, "Sharpness estimation for document and scene images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, pp. 3292-3295, 2012.
- [13] T. Obafemi-Ajayi, G. Agam, "Character-based automated human perception quality assessment in document images," *IEEE Transactions On Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(3), pp. 584-595, 2012.
- [14] P. Ye, J. Kumar, L. Kang, D. S. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proceedings of the CVPR*, pp. 1098-1105, 2012.
- [15] A. Alaei, D. Conte, R. Raveaux, "Document Image Quality Assessment based on Improved Gradient Magnitude Similarity Deviation," in *Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 176-180, 2015.
- [16] A. Alaei, D. Conte, M. Blumenstein, R. Raveaux, "Document Image Quality Assessment Based on Texture Similarity Index," in *Proceedings of the 12th IAPR Workshop on Document Analysis Systems (DAS)*, pp. 132-137, 2016.
- [17] N. Nikolaos, V. Dimitrios, (2009), "Binarization of pre-filtered historical manuscripts images", *International Journal of Intelligent Computing and Cybernetics*, 2(1), pp. 148-174.
- [18] G. Agam, G. Bal, G. Frieder and O. Frieder, "Degraded document image enhancement," in *Proceedings of the SPIE 6500, Document Recognition and Retrieval XIV*, 65000C (2007). doi:10.1117/12.706484.
- [19] Z. Shi, V. Govindaraju, "Historical Document Image Enhancement Using Background Light Intensity Normalization," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, pp. 473-476, 2004.
- [20] E. Kavallieratou, E. Stamatatos, "Improving the Quality of Degraded Document Images," in *Proceedings of the Second International Conference on Document Image Analysis for Libraries*, pp. 340-349, 2006.
- [21] M. Cannon, J. Hochberg and P. Kelly, "Quality Assessment and Restoration of Typewritten Document Images," *International Journal of Document Analysis and Recognition*, 2, pp. 80-89, 1999.
- [22] V. Govindaraju, S. N. Srihari, "Image Quality and Readability," in *Proceedings of the International Conference on Image Processing (ICIP)*, Vol. 3, pp. 324-327 1995.
- [23] D. K. Reed, E. H. Barney Smith, "Correlating degradation models and image quality metrics," in *Proceedings of the SPIE-Document Recognition and Retrieval (DRR)*, 6518, 2008. DOI: 10.1117/12.766784
- [24] R. Biedert, A. Dengel, M. Elshamy, G. Buscher, "Towards Robust Gaze-Based Objective Quality Measures for Text," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 201-204, 2012.
- [25] J. Luengo, S. García, F. Herrera, "A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests," *Expert Systems with Applications*, 36 (4), pp. 7798-7808, 2009.
- [26] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*. 1 (6), pp. 80-83, 1945.
- [27] S. S. Shapiro, M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, 52(3-4), 1965.
- [28] J. Li, X. Tang, W. Zhao, J. Huang, "A new framework for identifying differentially expressed genes," *Pattern Recognition*, 40 (11), pp. 3249-3262, 2007.
- [29] A. J. Viollaz, "On the reliability of the chi-square test," *Metrika* 33(1), pp. 135-142. 1986. doi:10.1007/BF01894740.
- [30] E. W. Weisstein, "Cross-Correlation." From *MathWorld--A Wolfram Web Resource*. <http://mathworld.wolfram.com/Cross-Correlation.html>.