Blind Quality Assessment Metric and Degradation Classification for Degraded Document Images

Atena Shahkolaei[†], Azeddine Beghdadi[†]
and Mohamed Cheriet[†]

[†]*Synchromedia Laboratory for Multimedia Communication in Telepresence, École de technologie supérieure, Montreal (QC), Canada H3C 1K3*
[‡]*Department of Computer Science & Engineering, Paris 13 University, France*
atena.shahkolaei.1@ens.etsmtl.ca, beghdadi@univ-paris13.fr, mohamed.cheriet@etsmtl.ca

Abstract

Epochal documents suffer from several types of noises that accumulate and evolve over time. This significantly affects their quality and makes their storage and the interpretation of their visual content problematic. Digital preservation seems the most viable and the most promising. Moreover, measuring the amount of degradation and quality assessment of degraded documents is highly desirable for applications such as selecting the proper algorithms for enhancement and analysis of document images, filtering the damaged images, tuning the processing algorithms parameters, document repairing, psychological study, etc. The first contribution of this work is the proposition of an efficient Multi-distortion Document Quality Measure (MDQM) for quality assessment of physically degraded document images. The proposed MDQM metric is based on three sets of spatial and frequency image features. These features are extracted from two layers of text and non-text and mapped to the mean opinion scores (MOS) using the regression function. The second contribution of this work is to estimate the probability of four common document image distortion types, namely, paper translucency, stain, readers annotations and worn holes in the degraded images. In our experiment, the correlations of seven no-reference image quality assessment (NR-IQA) metrics with the MOS values are evaluated on two available datasets. It is shown that the performance of MDQM metric is significantly

better than the state-of-the-art NR-IQA metrics. Moreover, the experimental results demonstrate that MDQM metric not only leads to high efficacy for classification of the various degradations but also maintains a remarkable run-time efficiency. **It is worth to mention that the proposed method has been conducted for Arabic documents.**

Keywords: No-reference image quality assessment, degraded document images, physical noises, local phase, degradation classification, support vector machine.

---

1. Introduction

Historical document images represent an important part of the cultural heritage of countries and civilizations. Therefore, preserving and protecting these cultural heritages is of great importance and responsibility of governments and international organizations like UNESCO. In recent decades, digitizing historical documents and manuscripts to preserve and make them accessible via electronic media has received a considerable amount of attention [1]. Although the issue of digitizing these documents is mostly solved, the problem of analyzing them is still an ongoing challenge.

The image quality of these documents can be assessed subjectively and objectively. One factor that may affect the readability and interpretation of such media is image quality. Although subjective image quality assessment (IQA) assessment is the most reliable method, it demands human participants which makes the assessment time-consuming, tedious and expensive. The average result of a set of standard subjective tests is called MOS values. Therefore, objective assessment is the primary choice for IQA applications. Objective IQA automates the estimation of image quality by substituting the human perception process with some quality metrics. Objective assessment methods can be classified into three main categories according to the availability of the reference image. These categories are (1) full reference (FR), (2) reduced-reference (RR) and (3) no reference (NR) [2]. For FR metrics both original and distorted im-

ages are available [3–5]. RR metrics use the partial information about both the reference and degraded images [6–8]. Finally, for the NR methods, the evaluation of quality is based on some features and properties of the degraded image without referring to the original one. However, very often a priori knowledge of the distortions is used in the design of NR-IQA [9–13].

In recent years, several objective NR-IQA metrics have been proposed in the literature for different applications. In the following, we provide a brief review of NR-IQA metrics.

Moorthy et al. [9] proposed an interesting NR-IQA metric called Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) for NR-IQA. In their framework, first, scene statistics of the natural images are extracted to classify the images into different distortion types. Then, the extracted statistics are used to assess distortion-specific quality. The final quality score is found by a combination of the results from the aforementioned two steps. Another NR-IQA metric, BLind Image Integrity Notator using DCT coefficients statistics (BLIINDS-II) [10] has been proposed. It is based on fitting the DCT coefficients by a Generalized Gaussian Distribution (GGD) model derived from Natural Scene Statistics (NSS) study. Then, it transforms the model parameters into features. The popular BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) [13] metric takes advantage of the natural images statistics in the spatial domain. This metric is based on the statistics of fitting the Mean Subtracted Contrast Normalized (MSCN) coefficients by symmetric and asymmetric GGDs in two scales. The Curvelet Quality Assessment (CurveletQA) [14] metric exploits the statistical correlations between the curvelet energy distributions and distortions of the image using a log-pdf curvelet model. The High Order image Statistics (HOS-DIQA) [15] metric is proposed for Document Image Quality Assessment (DIQA) applications. It extracts a codebook from local image patches by using the k-means clustering. Then, the differences between the statistics of local features and these codewords are calculated as the quality metric. The Blind Quality Measure for SCIs (BQMS) [16] is the first proposed metric for quality assessment of the screen content images that ex-

ploit perceptual features according to the free energy measure and structural degradation information. In this metric, an immense number of training data was utilized in order to avoid the overfitting in comparison with other IQA metrics. NR Free Energy-Based Robust Metric (NFERM) is an NR-IQA metric which was proposed based on the free energy principle and important HVS inspired features followed by an SVM-based regression module [12]. Indeed, the extracted features can be classified into three groups in the NFERM metric: the features inspired by the free energy principle and the structural degradation model, HVS-inspired features and the possible losses of naturalness in the distorted image.

It is worth noticing that, the majority of the objective image quality assessment metrics work on the entire image information for evaluating the quality of images [9, 10, 12, 14, 16–18]. Therefore, the use of these metrics in the case of document images does not seem adequate. Indeed, the observer's visual attention is not taken into account in the sense that the effect of visual attraction by the text is not exploited. In recent years, some objective IQA measures based on different layers of images have been proposed for predicting the quality of images. Shahkolaei et al. [19] proposed the VDQAM metric for quality assessment of degraded document images. This measure is based on the statistical analysis of a set of selected features extracted from four layers of the image using a log-Gabor filters based image segmentation method. In the same vein, another approach based on a background/foreground image segmentation and a patch selection scheme has been proposed recently for the design of a new IQA [20].

In order to maintain, control and enhance the quality of degraded documents and also decrease the negative effect of distortions on diverse processing and analysis systems, it is necessary to estimate the probability of different distortion types in ancient document images.

To classify the degradation types in natural images, Moorthy and Bovik [11] introduced a two-step framework for blind image quality evaluation based on the natural scene statistics (NSS). This metric estimates the probability of dif-

ferent degradation types in natural images such as white noise, Gaussian blur, JPEG, etc. In [21], a degradation classification method for natural images was proposed that is based on the recognition accuracy of degradation type and overall image quality assessment. A Bayesian approach was utilized in order to predict the type of distortion in images using image quality metrics in [22]. In [23], a global FR image quality measure (IQM) was proposed based on linear discriminant analysis (LDA) classifier and neural approach. In this method, an artificial neural network (ANN) was used to improve the performances of the existing IQMs by selecting and combining the best IQMs for each considered degradation. Recently, Xiongkuo Min et al. [24] proposed a blind pseudo reference image (BPRI) metric in order to estimate some distortions such as blockiness, sharpness, and noisiness. This method is training-free except for degradation classification step.

Indeed, MOS values and OCR accuracy are two common criteria used for performance evaluation of document image quality. Furthermore, the richness and representativity of pictures in each category, are of great importance in the case of document images. Recently, two degraded document datasets with the human opinion scores were introduced in the literature [19, 25]. In these datasets, the pair comparison rating was used instead of the previously used OCR performance [26]. With the availability of such dataset, more research on degradation classification and modeling can be conducted in addition to the document image quality assessment. In what follows, we mention why human judgments are used in the introduced dataset instead of OCR accuracy. Firstly, the OCR engines are not perfect, especially for some of the languages, old writing styles and fonts. The second reason is that higher OCR accuracy does not necessarily mean that a document image is of high quality, rather it may mean that the text region is not degraded [19]. Thirdly, OCR methods work sufficiently well only for modern Latin documents. For ancient documents, especially with severe degradation, such objective evaluation is not recommended, because OCR methods may detect some degradations as the text or foreground.

In this paper, we propose a blind image quality assessment metric (MDQM)

for quality assessment of old manuscripts. The proposed metric assumes that the sensitivity of the human visual system varies depending on the observed region in the document image (background and foreground). The proposed IQA measure not only uses the existing statistical features, but it introduces a new feature, namely the local phase that follows the Gaussian distribution. **MDQM metric is evaluated on the MHDID dataset which contains Arabic scripts.** To the best of our knowledge, for the first time, a degradation classification algorithm is also proposed to estimate the probability of four common types of degradation in ancient document images.

In this work, we demonstrate that the proposed method could be used in an automatic degradation modeling scheme that allows detecting the different types of degradations that may affect the quality of ancient document images. As mentioned before, historical document images suffer many types of degradation. Some of these distortions destroy a large part of the document images after years, such as worn holes. This a priori knowledge on the type and nature of distortions is of great help in designing efficient image quality enhancement methods. Also, the MHDID dataset [25] and the proposed objective IQA metric can be useful in a set of possible applications, such as model selection, educational purposes, psychological study, parameter optimization and tuning to name a few [19].

This paper is organized as follows. Section 2 provides a detailed description of the proposed IQA measure. In section 2.5, we outline the degradation classification model. Section 3 provides experimental results followed by a conclusion in section 4.

2. Proposed Document Image Quality Measure (DIQM)

The proposed method is based on some observations and specificities of the document images related to the way the human perceives structured images. In the literature, many researchers have studied the statistics of natural images and their correlation with the Human Visual System (HVS) [13, 17, 18, 27]. One major difference between a document image and a natural image is the existence

of foreground and background that can be easily distinguished by the human visual system. Indeed, the document image is considered as a two-phase media where the object/target are identified and recognized instantaneously. We can mention the various types of noises that exist in the natural and historical document images as the second difference. Previous study [19] shows that document IQA on a properly segmented document image is a good strategy to design better performing metrics. A novel NR-IQA metric for quality assessment of old document images based on this approach is proposed.

Fig. 1 illustrates the framework of the proposed blind image quality measure (MDQM). The features used in MDQM metric can be grouped into three major classes: *i*) the first corresponds to the mean of the local phase of the three color image channels, *ii*) the second group of features is composed of the MSCN coefficients and MSCN coefficients of gradient information from two layers of foreground and background, and *iii*) statistical features such as mean and standard deviation are extracted from the MSCN coefficients and gradient of MSCN coefficients in the third group. Support vector regression (SVR) is utilized to map the extracted features to MOS values. In the following, the main steps of the proposed method are described and discussed.

2.1. Text/non-text image segmentation

It is worth to mention that the human visual system responses to the degradations appearing in the foreground and background of a given image are quite different. In other words, due to various types of noise and other distortions in text and non-text regions, the sensitively of human vision to different parts of document images is not the same. Therefore, it is of great importance to segment document images into different layers based on HVS sensitivity, because human perceptual observation is the ultimate judge of the image quality evaluation process.

Recently, a new blind IQA metric dedicated to document images and based on image segmentation has been proposed in [19]. The observed degraded document image is segmented into four layers namely, text, degradations far from

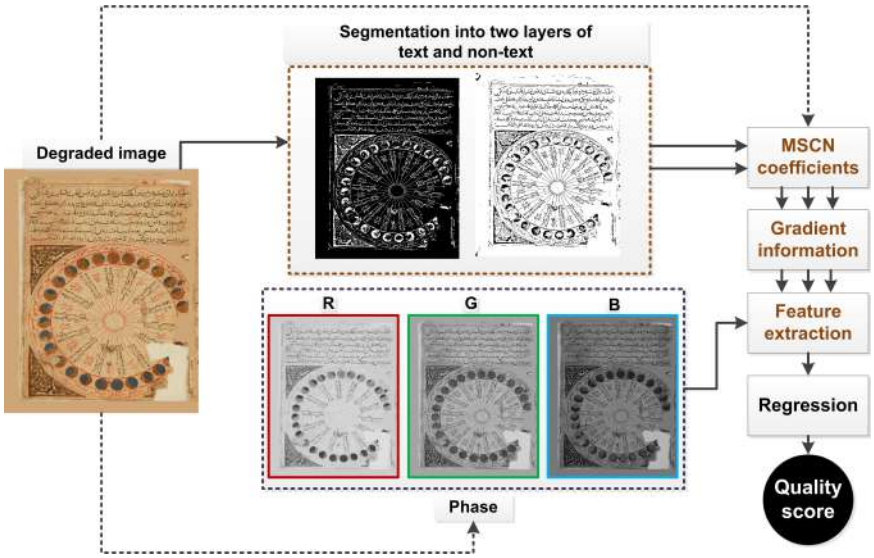


Figure 1: The framework of the proposed blind MDQM metric.

the text, degradations close to the text, and non-degraded pixels. Segmenting the image into four layers by using this approach is not always effective, because there is no guarantee that each layer is non-empty. Therefore, the feature analysis of these layers based on their MSCN coefficients histograms may not be accurate. To overcome this limitation, the segmentation of the degraded document image is restricted to two layers, namely text and non-text zones. The segmentation method is described below.

Simple global thresholding and other spatial domain filters cannot be used for the segmentation of degraded documents into different layers, due to the local variations in degraded document images [19]. The local adaptive methods can be used for segmentation of degraded document images. Log-Gabor filters have been reported to show good performance in many applications ranging from image quality assessment to image segmentation [17, 19, 28–30].

Different pairs of the symmetric and anti-symmetric functions, such as the Gabor, Log-Gabor, Gaussian derivatives, the difference of Gaussians, and Cauchy functions, which are known as quadrature pair filters were used in the literature

[31, 32]. In this work, we exploit the joint spatial and frequency localization of log−Gabor filter for segmenting the historical document images into two layers, namely foreground and background. Following the formalism and notation used in [31], let us consider the quadratic pairs defined by $M^e_{\rho r}$ and $M^o_{\rho r}$ which correspond to the even-symmetric and odd-symmetric wavelets at a scale ρ and orientation r. By considering $I(\mathbf{x})$ as a two-dimensional signal, the response associated with each quadratic pair of filters at each image point $\mathbf{x}$ forms a response vector obtained by convolving the wavelet filters with $I(\mathbf{x})$ [19] follows:

$$[e_{\rho r}(x), o_{\rho r}(x)] = [I(\mathbf{x}) * M^e_{\rho r}, I(\mathbf{x}) * M^o_{\rho r}] \tag{1}$$

where * denotes convolution, and $e_{\rho r}(x)$ and $o_{\rho r}(x)$ are the real and imaginary responses in the complex-valued frequency domain. The signal energy is defined as the summation of the responses across the five scales and directions as follows:

$$\mathrm{E(x)} = \sum_{\rho r} e_{\rho r}(x) \tag{2}$$

The degraded document image is segmented into two layers, namely text and non-text pixels, using a thresholding process defined as follows [19]:

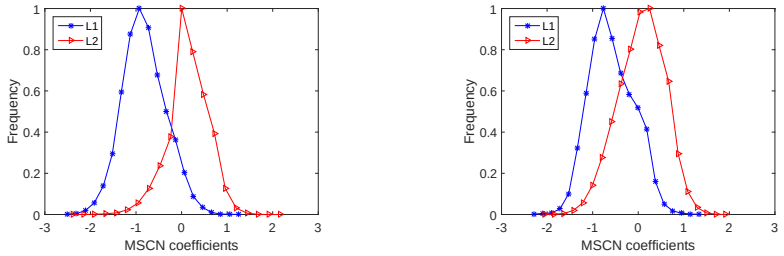
$$\begin{cases} L_1 : & \mathrm{E(x)} \leq \text{mean negative values} \\ L_2 : & \mathrm{E(x)} > \text{mean negative values} \end{cases} \tag{3}$$

where, layer L_1 approximates the text regions, while the layer L_2 corresponds to the non-text pixels.

Two typical degraded document images are shown in Fig. 2. It is clear that the proposed segmentation method has remarkable efficacy in discriminating between text and non-text pixels. From Fig. 2, it can also be observed that the histograms of MSCN coefficients for both layers of text and non-text of the degraded images exhibit a well shaped Gaussian distribution.

After segmenting document images into two layers, the statistics of the MSCN coefficients and their gradient information are computed for each layer

(a) Original image (b) Segmented image (c) Original image (d) Segmented image



(e) (f)

Figure 2: Illustrations of the proposed document image segmentation method and the associated MCSN coefficients histograms of text and non-text classes. Foreground and background pixels of two degraded images are shown in (b) and (d) images with white and black color, respectively. In (e) and (f) MCSN histograms where the blue plot is related to the text regions (L_1), while the red plot is related to the non-text regions (L_2).

---

and the entire document image. In the next subsection, more details about these statistics are given.

2.2. Document statistical analysis based on the MSCN coefficients

In order to remove the dependency between different image features, MSCN coefficients are used as a normalization technique in the proposed approach. The main behavior of the MSCN coefficients is their tendency toward a Gaussian distribution. MSCN coefficients are commonly used for NR-IQA of natural images [13, 17]. More details on the relevancy of the MSCN coefficients in the design of the proposed IQA metric are given in the following.

In [33], MSCN coefficients are computed by applying a local nonlinear op-

eration to the luminance channels, to remove local mean displacements of the luminance and normalize the local variances of luminance. Given an image $I(i,j)$, MSCN coefficient can be computed as follows:

$$\text{MSCN}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{4}$$

where $i \in 1, 2, ..., M$, $j \in 1, 2, ..., N$ are spatial indices of pixel; M, N are the height and width of the image respectively; $C = 1$ is a constant that prevents the denominator to be zero [17]. The local mean $\mu(i,j)$ and standard deviation $\sigma(i,j)$ are estimated as follows:

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j) \tag{5}$$

and

$$\sigma^2(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} \big(I_{k,l}(i,j) - \mu(i,j)\big)^2 \tag{6}$$

where, $\omega = \{\omega_{k,l} | k = -K, ..., K, l = -L, ..., L\}$ is a two-dimensional symmetric Gaussian weighting window.

It is worth to mention that the image gradient contains relevant information about edges and variations of local contrast that were utilized extensively in many FR and NR image quality assessment metrics [3, 34–36]. In this work, we use the gradient magnitude.

The estimated gradient magnitude M of the image $I(i,j)$ is given by:

$$M(i,j) = \sqrt{(I * H_x)^2(i,j) + (I * H_y)^2(i,j)} \tag{7}$$

where $*$ denotes the convolution operator. H_x and H_y are the convolution masks associated with the vertical and horizontal Sobel operator.

In this study, the statistics of MSCN coefficients of gradient magnitude are computed from each layer and the entire degraded image separately. It should be mentioned that the gradient magnitude features reflect the maximum intensity variation regardless of orientations and represent the strength of local

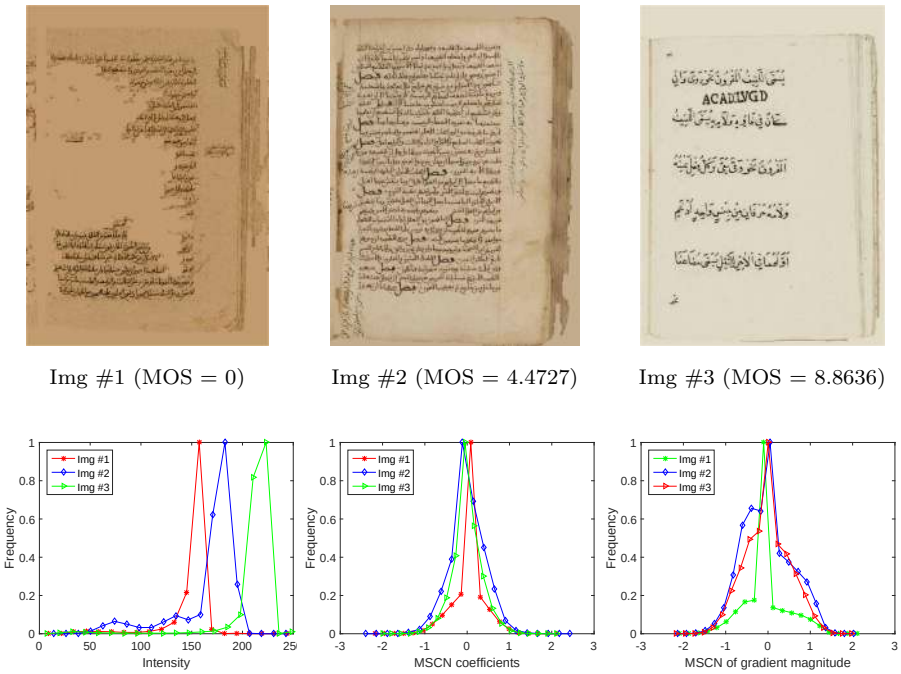


Figure 3: Histograms of intensity, MSCN and MSCN of gradient magnitude for three degraded document images with different MOS values, selected from the MHDID dataset.

luminance change [37]. In this work, we extract some statistical features from the MSCN of gradient magnitude to normalize the gradient magnitude to obtain a stable statistical image representation. Fig. 3 shows the histograms of the intensity, MSCN distributions and MSCN of gradient magnitude for three historical document images with different MOS values from the MHDID dataset. The Gaussian behavior of the MSCN coefficients and its gradient, unlike what is shown in the intensity histogram, can be observed from the plots. Besides, it can be seen that although the histogram of MSCN coefficients and its gradient follow the Gaussian distribution, the histogram of MSCN tends more to a bell curve in comparison with the histogram of MSCN of gradient magnitude. The gaussianity of the MSCN coefficients is expected since from equation 4. MSCN is the standard score of pixel intensity (plus the constant C in the denominator) in the local window of the image. The standard score is used in order to measure

how many standard deviations below or above the population mean a raw score is. Thus the standardized scores follow a Gaussian distribution (bell curve), as shown in Fig. 3.

2.3. Local phase information of RGB channels

It has been observed that the human visual system reacts more strongly to points where highly ordered phase information [38]. Since the human visual cortex is sensitive to phase congruent structures [39], the phase congruency (PC) value at a location can reflect how likely it is a relevant perceptual salient feature [2]. It is also worth to mention that phase carries more visual information than does magnitude [40]. Taking these observations into account, PC was used as a feature in several IQA metrics [2, 38, 41–43].

Our proposed IQA metric utilizes features derived from a locally weighted mean phase angle (LWMPA) [44], which is robust to noise, through the three color channels (R, G, B). In the following, the basic notions of LWMPA and its use for IQA are briefly described.

By using equation (1), we can derive the values of $e_{\rho r}(x)$ and $o_{\rho r}(x)$ which lead to find the local phase $\omega_{\rho r}(x)$ at given scale and orientation of the wavelet. This value is computed by the following formula:

$$\omega_{\rho r}(x) = arctan2(o_{\rho r}(x), e_{\rho r}(x)) \tag{8}$$

$$arctan2(i, j) = 2arctan\frac{i}{\sqrt{i^2 + j^2} + j} \tag{9}$$

Finally, the LWMPA $ph(x)$ is computed by summing all the response vectors in all the possible orientations and scales:

$$ph(x) = arctan2\left[\sum_{\rho,r} e_{\rho r}(x), \sum_{\rho,r} o_{\rho r}(x)\right] \tag{10}$$

LWMPA takes it values within the range $[-\pi/2, \pi/2]$.

It should be mentioned that for feature extraction, the RGB image components are quite efficient in comparison with the transformed color spaces. The

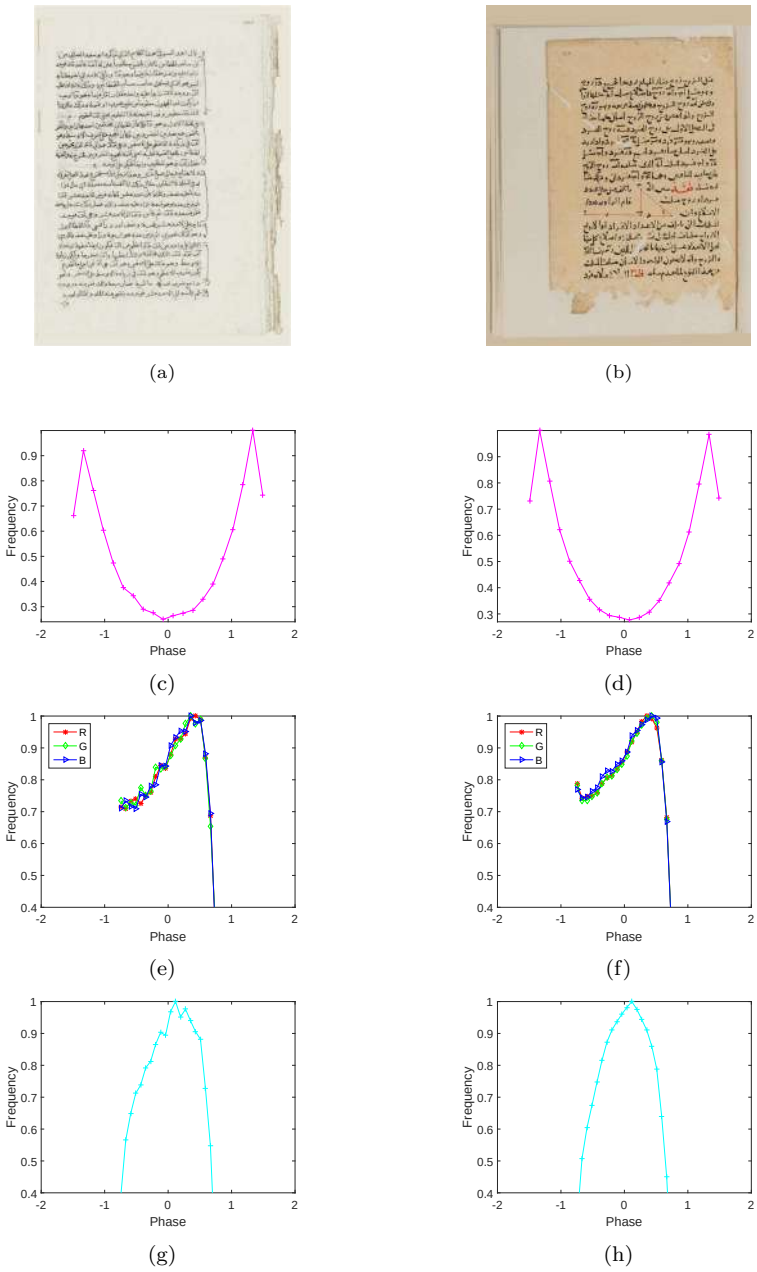


Figure 4: Histograms of the local phase of two degraded document images ((a)-(b)). The histograms in the second row indicate the distribution of local phase which is computed directly from gray scale version of images ((c)-(d)). The histograms in the third row correspond to the distribution of local phase on the RGB channels separately ((e)-(f)). The histograms of the mean of local phase of R, G and B channels are shown in (g) and (h).

interesting point of the proposed metric is that LWMPA is computed directly from each channel of ancient document images (R, G, B) separately. RGB image composites are quite efficient for feature extraction [45, 46]. Then, the final LWMPA (ph_{RGB}) is calculated from the mean of three LWMPAs for each degraded image (x) and expressed as:

$$ph_{RGB}(x) = \frac{1}{3} \sum_{c=R,G,B} ph_c(x) \tag{11}$$

Fig. 4 illustrates the local phase distribution of two degraded document images in the three cases according to *i*) gray-scale image, *ii*) individual RGB channels and *iii*) the combination of RGB channels ($ph_{RGB}(x)$) as proposed in this work. From the plots of Fig. 4, it can be seen that the local phase distribution of *iii* fairly tends towards a Gaussian distribution, while *i* and *ii* histograms are non-Gaussian. It is worth to notice that to the best of our knowledge, the local phase is utilized as a relevant feature for image distortion analysis for the first time in this work.

2.4. Symmetric and asymmetric generalized Gaussian distribution

The symmetric and asymmetric behaviors of the MSCN, its gradient, and local phase suggest fitting the model into asymmetric and symmetric Gaussian distributions. Extracted features are parametrized using a generalized Gaussian distribution. The zero mean generalized Gaussian distribution (GGD) can be expressed as:

$$f(x;\varphi,\sigma^2) = \frac{\varphi}{2\beta\Gamma(1/\varphi)} e^{\left(-\frac{|x|}{\beta}\right)^{\varphi}} \tag{12}$$

where φ controls the shape of the distribution and σ^2 is the variance of the distribution. The parameter β and gamma function Γ are defined as:

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\varphi)}{\Gamma(3/\varphi)}} \tag{13}$$

$$\Gamma(\varphi) = \int_0^{\infty} t^{\varphi-1} e^{-t} dt \ , \qquad \varphi > 0 \tag{14}$$

When the shape of the histogram is not symmetric, the GGD cannot be used for fitting [47]. Therefore, in this situation, asymmetric generalized Gaussian distribution (AGGD) is adequate [13]. The AGGD with zero-mode is defined by:

$$f(x;\digamma) = \begin{cases} \frac{\vartheta}{(\beta_l+\beta_r)\Gamma(1/\vartheta)}\exp\Big(-(\frac{-x}{\beta_l})^{\vartheta}\Big), & x<0 \\ \frac{\vartheta}{(\beta_l+\beta_r)\Gamma(1/\vartheta)}\exp\Big(-(\frac{x}{\beta_r})^{\vartheta}\Big), & x\geq 0 \end{cases} \tag{15}$$

The parameters β_l and β_r are defined as follows:

$$\beta_l = \sigma_l\sqrt{\Gamma(\frac{1}{\vartheta})/\Gamma(\frac{3}{\vartheta})} \tag{16}$$

$$\beta_r = \sigma_r\sqrt{\Gamma(\frac{1}{\vartheta})/\Gamma(\frac{3}{\vartheta})} \tag{17}$$

$\digamma$ is the set of three statistical features which can be defined as following:

$$\digamma = (\sigma_l^2, \sigma_r^2, \vartheta) \tag{18}$$

where σ_l^2 and σ_r^2 are the variance of the left side and right side of the distribution, respectively. The parameter ϑ controls the shape of the distribution.

These classical statistical functions provide useful information about the statistical behavior of the distributions. Therefore, two parameters of GGD (φ, σ^2) and three parameters of AGGD ($\sigma_l^2, \sigma_r^2, \vartheta$) on two layers of the degraded image constitute the first part of the feature vector of MSCN coefficients and MSCN of gradient magnitude for the proposed metric. Additionally, two parameters of GGD (φ, σ^2) and three parameters of AGGD ($\sigma_l^2, \sigma_r^2, \vartheta$) on the entire degraded image constitute the second part of the feature vector of MSCN coefficients, MSCN of gradient magnitude and local phase for the MDQM metric.

All the above statistical functions are calculated from two image scales; the original image scale and its reduced resolution by the "bicubic" function. **It should be mentioned that we used half the original pixel resolution and the reduced resolution for the feature extraction is 0.5**. Indeed, the final feature vector is a concatenation of the features from each resolution,

62 features are extracted from the original image and 62 features are extracted from the reduced resolution **(that is124 features)**.

2.5. Degradation identification of degraded document images

Identification of the type of distortion and estimation of the amount of each degradation can be specified by using the degradation model. It can be also useful for the optimization of enhancement parameters. For instance, parameters of the selected binarization methods can be properly adjusted by the provided information about the possible degradation types. In this section, the estimation of the probability of each distortion is provided for degraded document images in MHDID dataset.

In order to estimate the probabilities, we need to feed the previously extracted features to Support vector machine (SVM) classifier. SVM classifier is more popular in comparison with other types of classification techniques especially in IQA because it performs well in high-dimensional spaces, avoid over-fitting and have good generalization capabilities [48]. **Therefore, a multi-class SVM classifier with one-against-all classification strategy is used to classify the degraded images into four different distortion categories: paper translucency, readers' annotations, stain and worn holes in this work.**

For the classification stage, the images of each distortion category are subdivided into 80% for training and 20% for testing sets. The degradation types are denoted as κ_i, $i = \{1, ..., 4\}$, where each number corresponds to a type of distortion. A one-vs-one multi-class SVM classification is used in this work. In our work, a degree two polynomial is selected empirically as the kernel function to estimate the probability of each distortion type. The greatest value of the probabilities indicates the dominant distortion type of a historical document image.

LIBSVM toolbox [49] is used for implementing SVM classifier. Classification probabilities for four degraded document images on the newly proposed quality metric for degraded documents (VDQAM) [19] and our proposed metric

Figure 5: The performance of degradation classification for four ancient document images by two metrics, VDQAM and MDQM. The first raw (e(i)-e(iv)) and second raw (f(i)-f(iv)) bars show the MDQM and VDQAM operation for the probability estimation of different distortions types, respectively. The X-axis of bars demonstrates different distortion types in the degraded images of MHDID dataset: paper transluency (PT), readers' annotations (RA), stains (S) and worn holes (WH).

(MDQM) are shown in Fig. 5. The histograms in green color demonstrate the ability of MDQM metric for probability estimation of each distortion type in MHDID dataset, while red histograms show this probability on VDQAM metric. This figure clearly illustrates the superiority of the MDQM over VDQAM for degradation classification and probability estimation. For instance, MDQM truly estimates a higher probability of paper translucency for the image Fig. 5(a) which contains paper translucency. However, class probability estimations of VDQAM are not accurate for the same image. Moreover, some other

degradations such as readers' annotations and stain are remarkably detected by VDQAM metric in Fig. 5 f(i), while the amount of these degradations is not significant in the corresponding image (Fig. 5(a)).

3. Experimental results and discussion

In this section, the performance of the proposed metric is analyzed regarding its ability to predict subjective ratings of image quality on two datasets. To map our feature vectors to MOS values, we fed them into a support vector regression (SVR) process [50].

The relationship between objective and subjective quality scores is nonlinear in general. Therefore, a regression is utilized to remove this non-linearity [51]. The reported PCC values in this paper were computed after mapping the quality scores to MOS using the following logistic function:

$$f(x) = \beta_1 \Big(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \Big) + \beta_4 x + \beta_5 \tag{19}$$

where β_1, β_2, β_3, β_4 and β_5 are fitting parameters computed by minimizing the mean square error between quality predictions x and subjective scores MOS. The scatter plot of real MOS values and predicted MOS values of the proposed IQA model are shown in Fig. 6. Generally speaking, the points near the fitted curve demonstrate the better quality assessment, while the rest of them show the inaccurate quality assessment. It can be seen from Fig. 6 that the five degraded document images which their real MOS values have the maximum difference with their predicted MOS values are indicated in the case of 80% train and 20% test.

3.1. Performance comparison of the proposed metric

The proposed blind metric (MDQM) is compared with six state-of-the-art metrics on two datasets. Visual document image quality assessment (VDIQA) dataset is the first dataset of degraded document images along with human

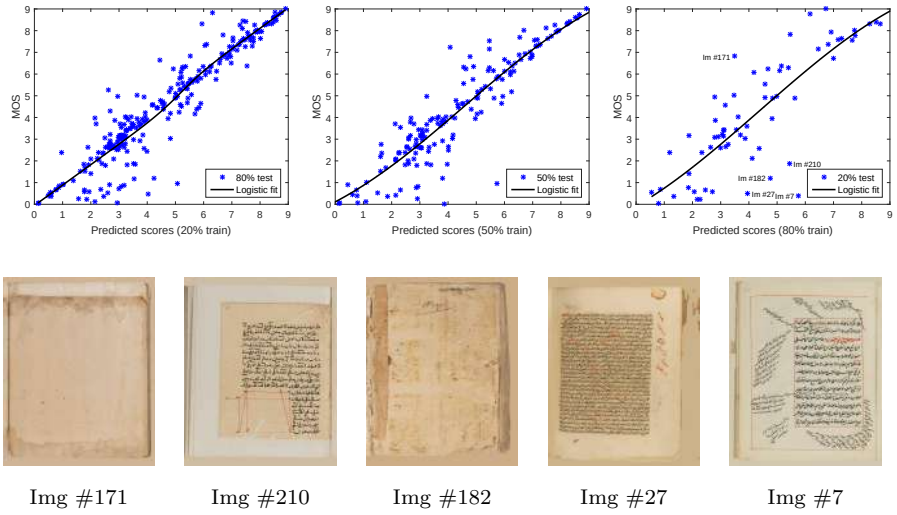


Figure 6: Scatter plots of quality scores against the subjective MOS on the MHDID dataset for the proposed model MDQM. These scatters are shown for the test set corresponding to the median PCC performance of the 1000 times train-test for three cases: 20% train 80% test, 50% train 50% test, and 80% train 20% test. The second **row** shows five degraded document images which their real MOS values have the maximum difference with their predicted MOS values in the case of 80% train and 20% test.

---

opinion scores for each document image [19]. This dataset contains 177 historical document images with various degradations. In order to judge the quality of degraded document images, 27 students were employed. The document images are taken from the library of Qatar University. The images are a subset of pages from several old Arabic books. It should be mentioned that the number **of** degraded documents in this dataset is not adequate for some future goals such as degradation classification and modeling. Also, no information about the type of degradations is provided in this dataset. Therefore, the MHDID dataset was introduced by Shahkolaei et al. [25] to overcome these limitations. The MHDID dataset consists of 335 degraded documents with four types of degradation categories, namely, paper translucency, stain, readers annotations, and worn holes. For the purpose of visual evaluation of the documents, 36 students were asked to judge the quality of degraded images. These degraded

document images were selected from the library of Qatar University. The language of these manuscripts is Arabic and these documents were collected from 130 different books edited during the 1st to the 14th Islamic centuries period [25]. **Pair Comparison Rating (PCR) method is used as a subjective rating method for evaluating the visual quality of degraded document images in VDIQA and MHDID datasets [52]**. In this method, two images are shown to subjects and the subjects are asked to select an image with higher quality in each trial. **More details about** both datasets were provided in [25].

The degree of concordance between the subjective and objective evaluations is analyzed in terms of three correlation measures: the Spearman Rank-order Correlation coefficient (SRC), the Pearson linear Correlation Coefficient (PCC), and the Kendall Rank Correlation coefficient (KRC). The PCC and SRC metrics measure prediction linearity and monotonicity, respectively. The KRC is used to evaluate the degree of similarity between quality scores and MOS. A good objective quality metric is expected to attain high values in SRC, PCC and KRC. We evaluated our proposed metric on the VDIQA and MHDID datasets. The results were reported based on the median value of 1000 times train-test for three cases of 20% train 80% test, 50% train 50% test and 80% train 20% test. The median SRC, PCC and KRC values across these 1000 train-test trials are tabulated in Tables 1-3. The top two metrics are highlighted in the tables. The correlations between objective image quality measures and MOS values show that the MDQM metric outperforms all the other NR-IQA metrics on both datasets. Also, it is clear from the Tables 1-3 **that the second best performing metric is the VDQAM metric.** The performance of the CurveletQA metric [14] is worst than other metrics on the VDIQA and MHDID datasets.

In order to show the accuracy and effectiveness of our proposed metric, VDIQA and MHDID datasets are utilized for testing and training. Table 4 indicates the performance of the proposed metric and six blind image quality assessment metrics in two cases: 1) when these metrics are trained on the MHDID dataset and tested on the VDIQA dataset. 2) when these metrics are

Table 1: Performance comparison of the proposed metric (MDQM) and six blind image quality assessment metrics on the VDIQA and MHDID datasets for 20% train and 80% test. Note the metrics that are specifically designed to assess the physical noises.

Database	Algorithms	Physical noise	20%-80%		
			SRC	PCC	KRC
VDIQA	DIIVINE [9]	No	0.6755	0.6925	0.4878
	BLIINDS-II [10]	No	0.5871	0.6062	0.4235
	BRISQUE [13]	No	0.5648	0.5926	0.4028
	CurveletQA [14]	No	0.5052	0.5511	0.3608
	BQMS [16]	No	0.6610	0.6823	0.4804
	VDQAM [19]	Yes	**0.7060**	**0.7204**	**0.5181**
	MDQM	Yes	**0.7224**	**0.7491**	**0.5323**
MHDID	DIVIINE [9]	No	0.6680	0.7004	0.4860
	BLIINDS-II [10]	No	0.6091	0.6357	0.4339
	BRISQUE [13]	No	0.6290	0.6680	0.4494
	CurveletQA [14]	No	0.5552	0.6161	0.3931
	BQMS [16]	No	0.6302	0.6692	0.4509
	VDQAM [19]	Yes	**0.6769**	**0.7102**	**0.4864**
	MDQM	Yes	**0.7286**	**0.7637**	**0.5349**

Table 2: Performance comparison of the proposed metric (MDQM) and six blind image quality assessment metrics on the VDIQA and MHDID datasets for 50% train and 50% test. Note the metrics that are specifically designed to assess the physical noises.

Database	Algorithms	Physical noise	50%-50%		
			SRC	PCC	KRC
VDIQA	DIIVINE [9]	No	0.7726	0.7878	0.5795
	BLIINDS-II [10]	No	0.6837	0.7008	0.5105
	BRISQUE [13]	No	0.7356	0.7539	0.5466
	CurveletQA [14]	No	0.6053	0.6419	0.4437
	BQMS [16]	No	0.7573	0.7815	0.5666
	VDQAM [19]	Yes	**0.7914**	**0.8095**	**0.6095**
	MDQM	Yes	**0.8300**	**0.8429**	**0.6432**
MHDID	DIVIINE [9]	No	0.7517	0.7838	0.5673
	BLIINDS-II [10]	No	0.7256	0.7490	0.5360
	BRISQUE [13]	No	0.7458	0.7783	0.5549
	CurveletQA [14]	No	0.6339	0.6930	0.4628
	BQMS [16]	No	0.7464	0.7779	0.5548
	VDQAM [19]	Yes	**0.7686**	**0.7878**	**0.5770**
	MDQM	Yes	**0.8098**	**0.8349**	**0.6219**

Table 3: Performance comparison of the proposed metric (MDQM) and six blind image quality assessment metrics on the VDIQA and MHDID datasets for 80% train and 20% test. Note the metrics that are specifically designed to assess the physical noises.

Database	Algorithms	Physical noise	80%-20%		
			SRC	PCC	KRC
VDIQA	DIIVINE [9]	No	0.7961	0.8292	0.6144
	BLIINDS-II [10]	No	0.7218	0.7590	0.5560
	BRISQUE [13]	No	0.7950	0.8283	0.6120
	CurveletQA [14]	No	0.6524	0.6985	0.4903
	BQMS [16]	No	0.7789	0.8213	0.5931
	VDQAM [19]	Yes	**0.8176**	**0.8481**	**0.6492**
	MDQM	Yes	**0.8447**	**0.8768**	**0.6785**
MHDID	DIVIINE [9]	No	0.7862	0.8191	0.6099
	BLIINDS-II [10]	No	0.7582	0.7891	0.5723
	BRISQUE [13]	No	0.7894	0.8245	0.6026
	CurveletQA [14]	No	0.6623	0.7250	0.4914
	BQMS [16]	No	0.7863	0.8266	0.5998
	VDQAM [19]	Yes	**0.7956**	**0.8294**	**0.6115**
	MDQM	Yes	**0.8395**	**0.8667**	**0.6566**

trained on the VDIQA dataset and tested on the MHDID dataset. For each performance measure, the best result is highlighted in bold. It can be observed in Table 4, the performance of MDQM metric is remarkably better than the other metrics.

3.2. Performance comparison of degradation classification model

As mentioned before, each degraded document image belongs to one of the four distortion categories of degradation in the MHDID dataset. For modeling the classification types, SVM classifier was used. Experimental results are done using 10-fold cross-validation setup where all samples were randomly divided into ten folds. **In each testing cycle for a class of degradation, one fold is used for testing and the rest of them are used for the training part. Indeed, this testing and training procedure is done for each class of degradation. Furthermore, the extracted features from each IQA metric are used as features for the classification.** The process is

Table 4: Performance comparison of MDQM metric and six NR-IQA metrics in two cases: 1) when these metrics are trained on the MHDID dataset and tested on the VDIQA dataset. 2) when these metrics are trained on the VDIQA dataset and tested on the MHDID dataset.

	Test on VDIQA			Test on MHDID		
Algorithms	SRC	PCC	KRC	SRC	PCC	KRC
DIIVINE [9]	0.7138	0.7330	0.5008	0.6989	0.7284	0.5061
BLIINDS-II [10]	0.6689	0.6901	0.4803	0.6310	0.6709	0.4762
BRISQUE [13]	0.7329	0.7510	0.5431	0.7098	0.7355	0.5329
CurveletQA [14]	0.6350	0.6609	0.5020	0.6136	0.6489	0.5049
BQMS [16]	0.7222	0.7467	0.5337	0.7022	0.7290	0.5218
VDQAM [19]	0.7584	0.7712	0.5690	0.7331	0.7656	0.5520
MDQM	**0.8023**	**0.8119**	**0.6001**	**0.7831**	**0.7948**	**0.6020**

repeated ten times and the final performance is calculated by the average result overall testing cycles.

F-measure values are calculated to evaluate the performance of the proposed metric for degradation detection. The comparison of F-measure values for seven NR-IQA algorithms is shown in Table 5. For each degradation type, the top two NR-IQA metrics are highlighted. As observed in Table 5, the proposed metric (MDQM) has the highest performance for detecting different physical noises in the MHDID dataset.

Table 5: The comparison of F-measure values for different document image degradation types in MHDID dataset for the seven NR-IQA metrics. We highlight the top two best performance metrics with boldface. Values are in percent.

	F-measure			
Algorithms	Paper translucecy	Readers' annotations	Stains	Worn holes
DIIVINE [9]	8.70	30.00	56.72	75.00
BLIINDS-II [10]	17.39	0	53.13	89.33
BRISQUE [13]	11.76	59.26	57.14	68.29
CurveletQA [14]	0	0	57.58	77.78
BQMS [16]	8.33	10.53	48.28	42.42
VDQAM [19]	**62.50**	**75.86**	**71.43**	**90.77**
MDQM	**78.57**	**82.35**	**82.93**	**96.77**

We can see that for almost all the metrics, the worn hole degradation has the highest value of F-measure. One reason is the fact that the worn holes distortion is one of the degradation prevailing in the degraded images of MHDID dataset. The other reason is that in the majority of images with worn holes degradation in the MHDID dataset, the amount of other degradations is negligible. Therefore, SVM classifier can detect this degradation category with more precision.

Table 6 lists the accuracy percentages for popular NR-IQA metrics on MHDID dataset. The results clearly indicate that the proposed metric has the best accuracy value for detecting degradation types in comparison with others. It means that the proposed metric has the highest ability in among others for putting each distorted document in the correct category of degradation. The accuracy of VDQAM and MDQM is more than the rest of the metrics because these metrics are proposed for quality assessment of degraded document images. The BQMS and VDQAM metrics have the worst and second best accuracy value in Table 6, respectively. It is clear from this table that the difference of accuracy value in the rest of the metrics is not considerable [9, 10, 13, 14], because these metrics were proposed for quality evaluation of natural images and the types of degradation are quite different.

Table 6: Accuracy comparison between six NR-IQA metrics and the proposed metric **on the MHDID dataset**. Note that top two metrics are highlighted. Values are in percent.

NR Indices	DIIVINE	BLIINDS-II	BRISQUE	CurveletQA	BQMS	VDQAM	MDQM
Accuracy	47.76	49.25	55.22	49.25	34.32	**76.11**	**85.07**

3.3. Computational complexity

Having illustrated that MDQM metric performs remarkably well in predicting the document image quality, now we demonstrate that this metric also has an acceptable complexity among other state-of-the-art metrics. Table 7 shows the number of extracted features and run time of seven NR-IQA metrics in milliseconds when were applied to the images of size 384×512, 1024×1280 and 2160×8840. It should be mentioned that all experiments in this section were

performed on a desktop computer **(Corei7, CPU E3-1230 v5 @ 3.40GHz, 32 GB of RAM, MATLAB 2015*b* and windows 7 Pro 64-bit)**.

It can be seen from Table 7 that MDQM metric is among the top four fastest indices, while the number of extracted features in the proposed metric is more than other metrics. Compared to the other competing indices, DIIVINE, BQMS, and BLIINDS-II, the proposed index MDQM is about 8 to 10 times, 20 to 22 times and 23 to 24 times faster, respectively. In Table 7, BRISQUE [13] and BLIINDS [10] metrics have the best and worst **run times** in among of other metrics, respectively. Another observation from Table 7 is that the ranking of indices might not be the same when they are tested on images of different size. For example, the proposed metric performs faster than CurveletQA metric in the images of size 1024×1280, but slower than the images of size 384×512 and 2160×8840.

Table 7: Comparison of run-time in terms of milliseconds and the number of features for seven NR-IQA metrics.

NR-IQA metrics	no. of features	384×512	1024×1280	2160×8840
1) BRISQUE [13]	36	66	263	4611
2) VDQAM [19]	40	170	1115	15525
3) MDQM	124	793	5557	84100
4) CurveletQA [14]	12	673	6250	47201
5) BQMS [16]	13	17545	115903	1689025
6) DIIVINE [9]	88	7693	46210	868093
7) BLIINDS-II [10]	24	19769	127875	2007044

In Table 8, we tabulate the percentage of time devoted to each of the steps in seven NR-IQA metrics. As it is clear in Table 8, the maximum time is dedicated to spatial correlation in DIIVINE, NSS feature extraction in BLIINDS-II, pair-wise products and AGGD in BRISQUE, NSS feature extraction in CurveletQA, features of structural degradation information in BQMS, Segmentation based on the Log-Gabor filter in VDQAM and feature extraction of Local phase in the proposed metric.

Fig. 7 shows the histogram of efficiency (seconds/image with the size of

Table 8: Informal complexity analysis of seven NR-IQA metrics. Tabulated values reflect the percentage of time devoted to each of the steps in these metrics for a degraded document image.

Metric	Step	Percentage of time
DIIVINE [9]	Steerable pyramid decomposition	0.08
	Divisive normalization	14.90
	Or. & scale selective statistics	0.20
	Orientation selective statistics	0.69
	Across scale correlations	2.56
	Spatial correlation	79.22
	Across orientation statistics	2.32
BLIINDS-II [10]	DCT coefficients	31.48
	NSS feature extraction	68.51
BRISQUE [13]	MSCN coefficients	28.36
	Fit GGD to MSCN	11.27
	Fit AGGD to pairwise products	60.35
CurveletQA [14]	CurveltQA transform	29.75
	NSS feature extraction in curvelet domain	70.24
BQMS [16]	Free energy entropy based on an autoregressive model	43.21
	Features of structural degradation information	56.78
VDQAM [16]	Segmentation based on the Log-Gabor filter	60.85
	MSCN coefficients	10.81
	GGD	13.37
	AGGD	14.95
MDQM	Segmented image into two layers	14.60
	Feature extraction of Local phase	63.58
	Feature extraction of MSCN coefficients	13.68
	MSCN of gradient magnitude	8.12

1024×1280) and accuracy for seven blind IQA metrics. The blue and red bars demonstrate the efficiency and accuracy, respectively. An efficient method should have a higher value of accuracy and lower value of efficiency. It is clear from Fig. 7 that the proposed method has the best accuracy and its efficiency is at an acceptable level in comparison with other NR-IQA metrics. It can be observed from Fig. 7 that the lowest efficiency and highest accuracy belong to the BRISQUE and MDQM metrics, respectively, while the run-time efficiency of the BLIINDS-II metric is the worst.

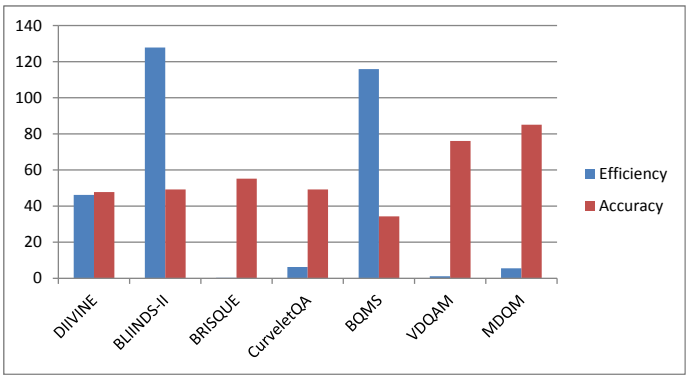


Figure 7: The histogram of efficiency (seconds/image) and accuracy for seven NR-IQA metrics.

4. Conclusion

We present a simple and effective blind image quality assessment metric (MDQM) to assess the quality of historical document images. In the proposed metric, each degraded document image is segmented into two layers (text and non-text) using the log-Gabor filters. Then, some statistical features such as MSCN coefficients, MSCN of gradient information and locally weighted mean phase angle are extracted from the two extracted layers and the entire degraded document image in two scales. Indeed, the spatial statistics of these layers are utilized for quality assessment. According to the performance analysis of the seven no-reference image quality assessment metrics, it is concluded that the MDQM metric significantly outperforms the other popular NR-IQA metrics on both VDIQA dataset and MHDID dataset. Indeed the proposed metric achieved the highest correlation with human judgments in comparison with other metrics. In this work, a

degradation classification model based on the proposed metric is also defined to estimate the probability of different types of degradation in ancient documents, namely, paper translucency, stain, readers annotations, and worn holes. SVM classifier was used for classifying the distorted images into four degradation categories. The experimental results demonstrate that the MDQM metric has the considerable ability for degradation classification of four common distortions in old manuscripts, while it has a moderate complexity. It is worth to note that this approach is language agnostic, although the results have been reported only on Arabic document images. In the future work, we will further consider other existing physical noises in old manuscripts by proposing a new dataset which has a huge number of degraded images. Moreover, proposing automatic degradation modeling will be considered as another target in our future work.

Acknowledgment

The authors thank the NSERC of Canada for their financial support under Grants RGPIN 138344-14.

References

[1] P. Ye, D. Doermann, Document image quality assessment: A brief survey, in: 12th International Conference on Document Analysis and Recognition, pp. 723–727.

[2] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A Feature Similarity Index for Image Quality Assessment, IEEE Transactions on Image Processing 20 (2011) 2378–2386.

[3] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, M. Cheriet, FSITM: A Feature Similarity Index For Tone-Mapped Images, IEEE Signal Processing Letters 22 (2015) 1026–1029.

[4] H. R. Sheikh, A. C. Bovik, Image information and visual quality, in: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pp. iii–709–12 vol.3.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (2004) 600–612.

[6] K. Gu, G. Zhai, X. Yang, W. Zhang, A new reduced-reference image quality assessment using structural degradation model, in: 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), pp. 1095–1098.

[7] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Transactions on Image Processing 21 (2012) 3378–3389.

[8] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE Journal of Selected Topics in Signal Processing 3 (2009) 202–211.

[9] A. K. Moorthy, A. C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, IEEE Transactions on Image Processing 20 (2011) 3350–3364.

[10] M. A. Saad, A. C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the dct domain, IEEE Transactions on Image Processing 21 (2012) 3339–3352.

[11] A. K. Moorthy, A. C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Processing Letters 17 (2010) 513–516.

[12] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, IEEE Transactions on Multimedia 17 (2015) 50–63.

[13] A. Mittal, A. K. Moorthy, A. C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Transactions on Image Processing 21 (2012) 4695–4708.

[14] L. Liu, H. Dong, H. Huang, A. C. Bovik, No-reference image quality assessment in curvelet domain, Signal Processing: Image Communication 29 (2014) 494–505.

[15] J. Xu, P. Ye, Q. Li, Y. Liu, D. Doermann, No-reference document image quality assessment based on high order image statistics, in: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3289–3293.

[16] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, Learning a blind quality evaluation engine of screen content images, Neurocomputing 196 (2016) 140–149.

[17] S. Abdalmajeed, J. Shuhong, No-reference image quality assessment algorithm based on Weibull statistics of log-derivatives of natural scenes, Electronics Letters 50 (2014) 595–596.

[18] D. Ghadiyaram, A. C. Bovik, Perceptual quality prediction on authentically distorted images using a bag of features approach, Journal of Vision, volume 17, Issue 1 (2017).

[19] A. Shahkolaei, H. Ziaei Nafchi , S. Almadad, M. Cheriet, Subjective and objective quality assessment of degraded document images, Cultural Heritage Journal, Vol 30, Pages 199-209 (2017).

[20] A. Alaei, R. Raveaux, D. Conte, Image quality assessment based on regions of interest, Signal, Image and Video Processing 11 (2017) 673–680.

[21] M. D. A. Chetouani, A. Beghdadi, A hybrid system for distortion classification and image quality evaluation, Image Communication Volume 27 (2012) Pages 948960.

[22] M. D. A. Chetouani, A. Beghdadi, Statistical modeling of image degradation based on quality metrics, in: ICPR.

[23] A. B. A. Chetouani, Image quality assessment based on distortion identification, in: SPIE Electronic Imaging Conferences, San Francisco, USA, Image Quality and System Performance VIII, Proceedings of SPIE, Volume: 7867,.

[24] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, C. W. Chen, Blind quality assessment based on pseudo reference image, IEEE Transactions on Multimedia (2017) 1–1.

[25] A. Shahkolaei, A. Beghdadi, S. Al-maadeed and M. Cheriet, A multi-distortion document image dataset towards no-reference quality assessment, in: 2nd IEEE Int. Workshop on Arabic and derived Script Analysis and Recognition (ASAR), 2018.

[26] W. Sun, F. Zhou, Q. Liao, Mdid: A multiply distorted image database for image quality assessment, Pattern Recognition 61 (2017) 153 – 168.

[27] W. Geisler, Visual perception and the statistical properties of natural scenes, 2008, volume 59 of *Annual Review of Psychology*, pp. 167–192.

[28] F. L. Bourgeois, E. Trinh, B. Allier, V. Eglin, H. Emptoz, Document images analysis solutions for digital libraries, in: First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings., pp. 2–24.

[29] B. Allier, H. Emptoz, Character prototyping in document images using gabor filters, in: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), pp. I–537–40, vol.1.

[30] A. K. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, in: 1990 IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings, pp. 14–19.

[31] G. Papari, N. Petkov, Edge and line oriented contour detection: State of the art, Image and Vision Computing 29 (2011) 79–103.

[32] D. Boukerroui, J. A. Noble, M. Brady, On the choice of band-pass quadrature filters, Journal of Mathematical Imaging and Vision 21 (2004) 53–80.

[33] Daniel L Ruderman, The statistics of natural images, Network: Computation in Neural Systems 5 (1994) 517–548.

[34] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, IEEE Transactions on Image Processing 21 (2012) 1500–1512.

[35] W. Xue, L. Zhang, X. Mou, A. C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Transactions on Image Processing 23 (2014) 684–695.

[36] D. Kundu, D. Ghadiyaram, A. C. Bovik, B. L. Evans, No-Reference Quality Assessment of Tone-Mapped HDR Pictures, IEEE Transactions on Image Processing 26 (2017) 2957–2971.

[37] W. Xue, X. Mou, L. Zhang, A. C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, IEEE Transactions on Image Processing 23 (2014) 4850–4862.

[38] Z. Liu, R. Laganire, Phase congruence measurement for image similarity assessment, Pattern Recognition Letters 28 (2007) 166–172.

[39] L. Henriksson, A. Hyvärinen, S. Vanni, Representation of cross-frequency spatial phase relationships in human visual cortex, Journal of Neuroscience 29 (2009) 14342–14351.

[40] A. V. Oppenheim, J. S. Lim, The importance of phase in signals, Proceedings of the IEEE 69 (1981) 529–541.

[41] A. Saha, Q. J. Wu, Perceptual image quality assessment using phase deviation sensitive energy features, Signal Processing 93 (2013) 3182–3191.

[42] A. S. Krzic, M. Donlic, M. Pejcinovic, D. Sersic, Image sharpness assessment based on local phase coherence and LAD criterion, in: 2016 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 1–4.

[43] Z. Wang, E. Simoncelli, Local phase coherence and the perception of blur, Neural information processing systems foundation, 2004, Vol. 16.

[44] P. Kovesi, Image features from phase congruency, Videre: J. Comput. Vis. Res. vol. 1 (1999) pp. 1–26.

[45] M. Tkalcic, J. F. Tasic, Colour spaces: perceptual, historical and applicational background, in: The IEEE Region 8 EUROCON 2003. Computer as a Tool., volume 1, pp. 304–308 vol.1.

[46] M. Tkalcic, J. F. Tasic, Colour spaces: perceptual, historical and applicational background, volume 1, IEEE, 2003.

[47] K. Sharifi, A. Leon-Garcia, Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video, IEEE Transactions on Circuits and Systems for Video Technology 5 (1995) 52–56.

[48] V. Vapnik, The Nature of Statistical Learning Theory, Berlin, Germany: Springer Verlag, 2000.

[49] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, ACM transactions on intelligent systems and technology (TIST) 2 (2011) 27.

[50] D. Basak, S. Pal, D. C. Patranabis, Support vector regression, Neural Information Processing-Letters and Reviews 11 (2007) 203–224.

[51] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Transactions on Image Processing 15 (2006) 3440–3451.

[52] L. Krasula, Quality Assessment Methodologies of Post-Processed Images, Ph.D. thesis, Universite de Nantes, 2017.