Blind Document Image Quality Prediction Based on Modification of Quality Aware Clustering Method Integrating a Patch Selection Strategy

Alireza Alaei[1] Donatello Conte[2] Maxime Martineau[2]
Romain Raveaux[2]
1: School of Information and Communication Technology, Griffith Institute for Tourism, Griffith University, Australia
2: Laboratoire d'Informatique Fondamentale et Appliquées de Tours (EA 6300), 64 avenue Jean Portalis, Tours, France
emails: 1: alireza20alaei@gmail.com, 2: {name.surname}@univ-tours.fr

May 15, 2018

Abstract

The quality of document images has direct impacts on the performance of document image processing systems. Document Image Quality Assessment (DIQA) is, therefore, of fundamental importance to a numerous document processing applications. As manual quality assessment is almost impossible for a huge volume of document images generated in day-to-day life, it is critical to develop intelligent machine operated methods to estimate the quality of document images. In this paper, a blind document image quality assessment method is proposed to deal with the problem of DIQA in real scenarios, as reference images are not always available. To estimate the quality of a document image, the document is first sampled into a set of patches. The extracted patches are then filtered out based on their level of foreground information using a patch selection strategy. For every selected patch, a cluster assignment is then performed to obtain its quality from a quality aware bag of visual words constructed using k-means clustering. An average pooling is finally employed to estimate the quality of the input document image. To evaluate the proposed method, a dataset composed of document images and three scene image datasets were considered for experimentation. The results obtained from the proposed method demonstrate the effectiveness of the proposed DIQA method. These achievements in applied computational intelligence, expert and decision support systems make a good foundation for creating practical tools to automate document image forgery detection, and archiving process.

1 Introduction

Paper is still a medium frequently used to store and distribute information in day-to-day life Mesquita et al. (2015). Electronic management of paper-based systems is one of the most important infrastructures in modern administrations. However, implementing a scalable expert document processing system, dealing with a mass-volume of documents acquired from heterogeneous sources (scanner, smart phones, etc.), is not often an easy task. Image Quality Assessment (IQA) technology is a major element of such a system and has been applied in many information system rehabilitation projects for a long time due to its low setup cost and technical requirement Ye & Doermann (2013).

Over the last few years, IQA has been an active domain of research in the field of computer vision, image processing and expert systems. As a result, many IQA metrics for different applications have been proposed in the literature Ye & Doermann (2013); Chandler (2013); Verikas et al. (2011); Yang et al. (2011). Based on the availability of a reference image for estimating image quality, IQA methods in the literature are categorized into three groups: i) full reference (FR), ii) blind or no reference (NR), and iii) reduced reference (RR) IQA methods. In FR IQA methods, it is assumed that the reference image of a distorted image is available for assessing the quality of the distorted one. In NR IQA methods, however, reference images are not available for IQA.

Methods based on the RR approach utilise a feature set extracted from a reference image at a source node and compare it with the features extracted from the image received at a destination node to estimate the quality of the image at the destination (Chandler, 2013).

Image Quality Assessment is also of significant interest in the document digitisation industry. In an industrial document image-oriented application led by the ITESOFT Company, DIQA through a non-OCR technique was of the primary interest. The target images were documents captured by mobile devices that were needed to be sent to remote servers. Therefore, two use-cases were defined as follows.

The first use-case called "Remote storage for mass-volume document scanning services" is related to the document image storage workflow in a Wide Area Network (WAN). Mass-volume document scanning services (thin clients) access remote servers within the WAN and send images to be stored in a datacentre at the server-side. Furthermore, many corporations, such as insurance companies, have implemented their own websites, demanding clients to digitise their documents and then upload them to the system. However, because of limitations in bandwidth and many other technical server issues, the documents cannot be of a large size. Clients may need to compress their files before sending them to the servers. These images must have the minimal quality required to be reused by either humans or machines. In the case of human reusability, the stored documents need to be inspected either by expert technicians at the server-side to provide feedback to the clients about the quality of their documents, or by an automatic algorithmic DIQA method to assess the quality of clients' documents and provide them quality scores. Examining documents manually is very time consuming and inconvenient for both service providers and clients. The practical way is to send a quick and automatic machine generated feedback to the clients when uploading a low quality document. In this way, clients can rectify their documents and upload them again.

The second use-case called "Document image captured by mobile systems and stored in remote storage". In this case, end-users having mobile systems access remote servers within the WAN and send images to be stored at the server-side. The captured documents are checked to maintain the minimum quality of documents by end-users at the client side. Therefore, an automatic quality assessment method, as an important part of a fail-fast system, needs to be developed on the mobile system to immediately report any condition, which may cause failures.

In the field of document image analysis, quality assessment metrics proposed in the literature (Lu et al. (2004); Hale & Barney-Smith (2007); Kumar et al. (2012); Obafemi-Ajayi & Agam (2012); Ye & Doermann (2012); Ye et al. (2012); Nayef & Ogier (2015); Kang et al. (2014); Bhowmik et al. (2014); Alaei et al. (2015, 2016)) can be grouped into two main categories: i) OCR-based approaches (Hale & Barney-Smith (2007); Ye & Doermann (2012); Nayef & Ogier (2015); Kang et al. (2014); Bhowmik et al. (2014)), and ii) human perception-based approaches (Lu et al. (2004); Kumar et al. (2012); Obafemi-Ajayi & Agam (2012); Ye et al. (2012); Alaei et al. (2015, 2016)). In the OCR-based category, character-based features have commonly been used to characterize document images in order to assess their quality. Metrics obtained using character-based features and their correlations with the results of Optical Character Recognition (OCR) have mainly been considered as a means of Document Image Quality Assessment (DIQA) in most of the works in the literature Hale & Barney-Smith (2007); Ye & Doermann (2012, 2013); Nayef & Ogier (2015); Kang et al. (2014); Bhowmik et al. (2014). A compact review of the methods based on typical OCR results for DIQA is available in Ye & Doermann (2013).

However, in both the above mentioned use-cases the results of OCR may not be completely correlated to the quality of document images, as OCR may fail to provide appropriate results for those images despite their good quality. To get an idea about the result of OCR on a digital document, a good quality document image and its OCR results obtained from a recent commercial version of the ABBYY OCR product (ABBYY FineReader 12.04) are shown in Fig. 2(a) and Fig. 2(b).From Fig. 2(b), it is quite clear that the results of the OCR are far from perfect despite its good quality. It is also noted that the OCR cannot preserve the format of the document image at all. However, individuals assessed the image shown in Fig. 2(a) as a good quality image. Therefore in this case, we proposed to employ a document image quality assessment based on human perception to obtain quality of images.

An overview of the DIQA methods in the literature reveals that a few human perception-based DIQA methods have been developed in pastLu et al. (2004); Kumar et al. (2012); Obafemi-Ajayi & Agam (2012); Ye et al. (2012); Alaei et al. (2015, 2016). Subjective mean human opinion score (MHOS) has been used as the basis of image quality for performance evaluation in these method. In Lu et al. (2004), a DIQA method based on a distance-reciprocal distortion measure has been

proposed. The method works on binary document images and features are extracted in character level. It has been assumed that the distance between two pixels plays an important role in their mutual interference perceived by human beings. In Kumar et al. (2012), a DIQA system based on estimating sharpness of document images captured by smart-phones has been proposed. Sharpness estimation is performed in both the vertical and horizontal directions. The difference of differences in grey scale values of a median-filtered image has been used as an indication of edge sharpness. In Obafemi-Ajayi & Agam (2012), a system for DIQA in character level has been presented. Three groups of features based on morphological operations, spatial characteristics and noise properties have been extracted from different characters to estimate the quality of every character using a neural network. Recently, a method based on unsupervised feature learning using a visual codebook has been proposed for NR IQA in Ye et al. (2012). In the proposed method, a mapping model between feature vectors extracted from image patches and their corresponding quality scores has been learned during training Ye et al. (2012). It has been proven that learning-based methods Ye et al. (2012); Xue et al. (2013) provided better results compared to other methods designed based on natural scene statistics (NSS) Mittal et al. (2012) and distributions of normalized luminance coefficient features Mittal et al. (2013). The authors claimed that their method is a general-purpose approach; however, they have not tested their method using MHOS for document images. The Modified Gradient Magnitude Similarity Deviation method Alaei et al. (2015) and Texture Similarity Index Alaei et al. (2016) both are FR DIQA methods based on gradient magnitude and texture features that cannot be applied when no reference images are available.

As mentioned, the methods presented in Lu et al. (2004); Obafemi-Ajayi & Agam (2012), work only on binary document images having fully textual data. They have also focused on character level quality assessment, which is far beyond current real-world scenarios. The methods presented in Kumar et al. (2012); Ye et al. (2012) are the only methods that used perceptual quality concepts for image quality assessment. The method presented in Ye et al. (2012), however, needs human subjective image quality assessment provided by many individuals as ground truth for training the system. Providing human-based subjective ground truth for such training data is a challenging task and even impractical in real-world scenarios. The method proposed in Kumar et al. (2012) has also specifically been developed for sharpness distortion estimation that cannot be generalized. A brief comparison of the merits/demerits of the document image quality assessment methods reviewed in this research work is presented in Table 1.

Considering the amount of NR DIQA research work in the literature and in view of the fact that in most of document image processing applications original or reference images may not be available, designing expert blind quality assessment systems dedicated to document images is of high demand. As a learning-based approach is less sensitive to training data and distortion types and it has further provided better IQA results Ye et al. (2012); Xue et al. (2013) compared to other IQA techniques, a BDIQA method based on a bag-of-visual-words learning approach is proposed in this research work. Inspired by the methods presented in Ye et al. (2012); Xue et al. (2013), a modification of Quality Aware Clustering (QAC) Xue et al. (2013), which is hereafter called MQAC metric, is proposed in this research work. As foreground information is the most important part of document images, the modification is mostly performed based on the use and integration of foreground information in the MQAC. To do so, a document image foreground/background segmentation method followed by a patch selection strategy is proposed to first extract patches containing foreground information. For every selected patch, a cluster assignment is then performed to obtain its quality from a quality aware bag-of-visual-words constructed using k-means clustering. An average pooling is finally employed to estimate the quality of the input document image. The proposed MQAC is fast and does not need subjective image quality provided by human for training/learning. It is further capable of working on no-reference colour/grey document images containing different distortions.

Our contribution in this research work can be highlighted in three-fold. The first and core part of this research work is proposing an expert system based on a quality aware clustering method (bag-of-visual-words) and a cluster assignment technique to automatically assess quality of document images. The second one is integrating a segmentation technique to the proposed system to approximately extract foreground information from a document for the purpose of quality assessment. The third one is the use of foreground information by proposing a decision support strategy for patch selection that significantly improves the performance of the proposed document image quality estimation metric in terms of accuracy and speed and is suitable for real world applications.

Category	Method	Merit	Demerit
OCR-based methods	Hale & Barney-Smith (2007)	Proposing a linear model for document degradation generated by a scanner.	Uses only character level information, the model is mostly based on heuristics that cannot be generalised. The proposed model did not include any noise.
	Ye & Doermann (2012)	Using an unsupervised feature learning method, Using local information.	The system is an OCR dependent system, finding an optimal number of code book is crucial to achieve a high performance.
	Nayef & Ogier (2015)	Using a distortion-specific quality metric for each distortion.	It is quite difficult to determine the type and number of distortions in document images to subsequently design a specific metric for each of them.
	Kang et al. (2014)	Using a Convolutional Neural Network (CNN) for patch quality assessment.	The system is data-dependant, as training needs to be performed using the same type of data considered in testing.
	Bhowmik et al. (2014)	Using allographs and support vector regression for predicting OCR results.	Performance of the system highly depend on tuning different parameters.
Human perception-based methods	Lu et al. (2004)	Modelling human visual system using a distance-reciprocal distortion measure.	It is not suitable for colour and grey document images. Works mostly at character level.
	Kumar et al. (2012)	Using local information for sharpness detection.	Only one type of distortions has been modelled.
	Obafemi-Ajayi & Agam (2012)	Using a neural network multilayer perceptron (MLP) regression model to predict document image quality.	Character segmentation is a challenging and erroneous process. Character level information has been considered for quality assessment. Works only on binary document images.
	Ye et al. (2012)	Using local information, unsupervised learning, and sparse coding for quality assessment.	It needs human subjective image quality assessment provided by many individuals for training/learning the system.
	Alaei et al. (2015)	Using foreground information, local and gradient magnitude features.	The proposed method cannot be applied when there is no reference images.
	Alaei et al. (2016)	Texture features and local information.	The proposed method cannot be applied when there is no reference images.

Table 1: An overview of the DIQA methods in the literature

The rest of the paper is laid out as follows: Section 2 describes the proposed BDIQA method. Section 3 discusses the experiments, results and comparative analysis. Finally, Section 4 provides some conclusions and future work.

2 Proposed MQAC Metric

The block diagram of the proposed BDIQA method (MQAC) is demonstrated in Fig. 1. The proposed method is divided into two stages of learning and testing. In the learning stage, using a segmentation technique and a patch selection strategy, similarity measures between patches extracted from the original images and patches from the corresponding distorted ones are computed. Then, a codebook representing different quality levels is constructed as a representation of patch qualities. In the testing phase, given a test image, it is initially partitioned into a number of patches. The same segmentation and patch selection strategy, as in the training phase, is performed and the qualities of the selected patches are estimated by comparing them with the codebook. An average pooling process of the estimated patch qualities provides the final quality score for the test image. Details of each step is presented in the following subsections.

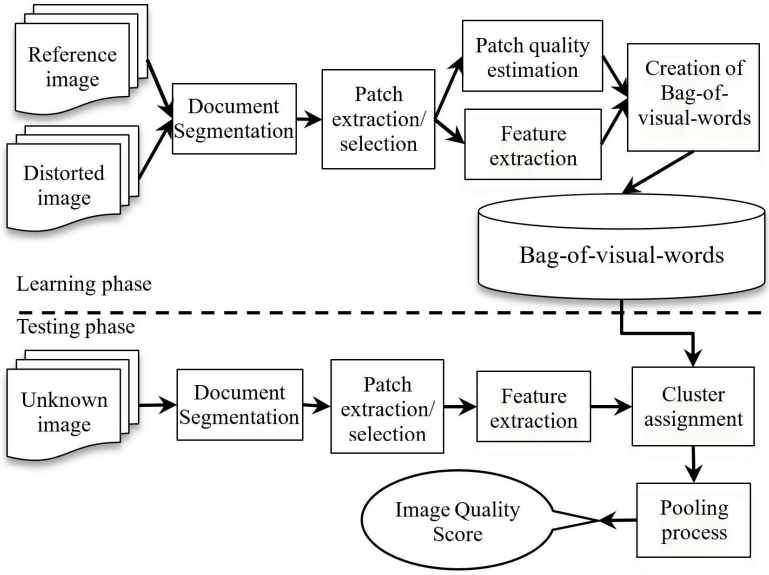


Figure 1: Block diagram of the proposed BDIQA method.

2.1 Applying piece-wise painting algorithm

Document images, unlike natural scene images, contain different types of information, such as textual information along with photo, graphical, signature, stamp and logo entities. Text, photo, graphics, logos, stamps, and tables, as foreground information, are generally used to convey a message or to serve as a proof and authentication document. On the other hand, the background does not hold consistent entities of the message and it is partially neglected by the Human Vision System (HVS).

Considering the fact that in document images the foreground carries more valuable information than the background, the quality of document images is highly dependent on the foreground information. That means, if the foreground information is distorted, the quality of a document image is low and vice-versa. An existing document image binarization technique may be used to extract foreground / background information in document images. However, an approximate foreground / background segmentation rather than an exact one is sufficient for the DIQA method proposed in this research work. Therefore, a state-of-the-art method for approximate foreground / background separation Alaei et al. (2011) is employed to approximately detect foreground information from document images for estimating their quality. The Piece-wise Painting Algorithm (PPA) is a powerful pre-processing technique evaluated on two different tasks, such as unconstrained handwritten text line segmentation Alaei et al. (2011) and logo detection Alaei & Delalandre (2014). The main

idea in this technique is to roughly segment a document image and provide an approximation of foreground and background information. In the PPA, a document image is initially divided into a number of vertical stripes of size s from the left to right direction. Intensity values in each row of a stripe are modified by the average intensity value of that row. Otsu's algorithm is then employed on each stripe separately to binarize the stripes. As a result, a painted image composed of a number of black and white portions is obtained. The black portions approximately represent foreground information irrespective of their content and white portions roughly signify the image background (Alaei et al., 2011). The resultant painted image obtained from the image shown in Fig. 2(a) is demonstrated in Fig. 2(c).



a) b) c) d)

Figure 2: (a) A sample of a distorted document image, (b) OCR result of (a), (c) the two-tone painted image obtained employing the PPA on the image shown in (a), (d) n patches sampled from the painted image shown in (c).

2.2 Patch extraction/selection

Since, the proposed method depends on mainly foreground image patches, the painted image obtained employing the PPA is sampled into n number of non-overlapping patches of size $p \times p$ say $\{P_1, P_2, \ldots, P_n\}$. The painted sampled image is shown in Fig. 2(d). To select a patch P_i, P_i needs to have a certain amount of foreground information. To formularize this statement, let SP be a set of selected patches, which is defined as:

$$SP = \left\{ P_i : T < \frac{|FG(P_i)|}{|P_i|} < 1, i : 1 \ldots m \right\} \tag{1}$$

where $|FG(P_i)|$ indicates the number of foreground (FG) pixels within the patch P_i and $|P_i|$ is the total number of pixels in the patch P_i. SP contains m patches where $m \leq n$. Fig. 2(d) shows the result of patch selection procedure, where a selected-patch is indicated by ✓and a discarded patch is shown by ×. The selected patches are further used for estimating the quality of the input image.

2.3 Patch quality estimation

As our goal is to use no human subjective quality scores in the learning phase, the most important part of the proposed BDIQA method is the assignment of a perceptual quality to each selected patch P_i. Basically, any existing FR IQA methods can be used for estimating the quality of a selected image patch providing it with a perceptual quality. From the literature, we noted that the FSIM method Zhang et al. (2011) has provided quite reasonable performance compared to the other FR IQA methods. Objective image quality obtained from the FSIMZhang et al. (2011) is highly correlated with the MHOS. Moreover, the dependency of training/learning based on the human opinion score is eliminated in our proposed method using the FSIM method Zhang et al.

(2011) to obtain the quality of patches. Therefore, the FSIM is used to estimate the qualities of distorted patches with respect to the corresponding patches in the reference image. Gradient magnitude and phase congruency are the features used in the FSIM for IQA Zhang et al. (2011). The Gradient Magnitude (GM) of a patch point $P_i(x, y)$ is computed using the following equations Zhang et al. (2011).

$$GM = \sqrt{G_h^2 + G_v^2} \tag{2}$$

$$G_h = \frac{1}{4} \times \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \times P_i(x, y) \tag{3}$$

$$G_v = \frac{1}{4} \times \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \times P_i(x, y) \tag{4}$$

The Phase Congruency at a patch point $P_i(x, y)$ is computed as follows.

$$PC(x, y) = \frac{\sum_o \sum_o W(x, y) \lfloor A_{so}(x, y) \Delta\Phi_{so}(x, y) - T \rfloor}{\sum_o \sum_o A_{so}(x, y) + \epsilon} \tag{5}$$

$$A_{so} = \sqrt{e_{so}(x, y)^2 - o_{so}(x, y)^2} \tag{6}$$

$$\Phi_{so} = \text{atan}(e_{so}(x, y)/o_{so}(x, y)) \tag{7}$$

where o and s indicate the index over orientations and scales respectively, the symbols $\lfloor \rfloor$ denote the enclosed value, which is equal to itself when it is positive, and zero otherwise. W is the sigmoid function used as a weighting parameter. A_{so} and Φ_{so} are amplitude of the response and the phase angle, respectively.

The patch quality score provided using the FSIM Zhang et al. (2011) is a real number ranging between 0 and 1. In FR IQA metrics, as the reference images are available, a simple average pooling of all estimated qualities for the patches provides an acceptable quality score of the image. The aim in this research work is, however, to construct a codebook for performing BDIQA. Since, the estimated image patch quality by the FSIM Zhang et al. (2011) can be different compared to the average of all estimated quality scores, direct use of the quality scores obtained from the FSIM Zhang et al. (2011) for constructing the codebook leads to a bias learning/training. Therefore, the estimated patch quality scores using the FSIM methodZhang et al. (2011) are normalized in order to be close to the overall perceptual quality of the image in which the patches are extracted. A percentile pooling technique Xue et al. (2013) is employed to normalize the patch quality score PS_i obtained using the FSIM method. To do so, PS_i is divided by a constant N to get PN_i which denotes the normalized quality score of the patch P_i. The constant N is computed as follows:

$$N = \frac{\sum_{i=1}^{m} PS_i}{10 \times \sum_{j=1}^{m/10} PS_j} \tag{8}$$

where m is the set of patches extracted from an image, j varies between 1 and 10% of the total number of patches extracted from the image ($m/10$), and PS_j are 10% of the patches having the lowest quality scores Xue et al. (2013). Since, in Wang & Shang (2006) it has been proven that the image patches with the worst estimated quality have a good linearity to human perception, 10% of the lowest quality patches have been used for normalization. The normalized values of PN_i are used for the creation of the bag-of-visual-words.

2.4 Feature extraction

One may expect that the features used for estimating the quality of an image patch in the previous section by employing the FSIM Zhang et al. (2011) method can be considered for the clustering and creation of bag-of-visual-words (BoVWs). These features are, however, only two features, phase congruency and gradient magnitude computed at the centre of the patch, are not sufficient for the construction of a fairly representative BoVWs from the extracted patches.

As pointed out in the literature Xue et al. (2013), to have a representative BoVWs, features representing the structural information of the input image are more suitable for the creation of visual codebooks. In this research work, therefore, structural features are extracted from all patches

P_i to construct a BoVWs, which also fairly represents the quality levels. The Difference of Gaussian (DoG) feature, whose proficiency in representing the structural information has been proven in the literature Young (1987), is used to extract a set of features for each patch P_i. In the DoG filter used for feature extraction, the support size of the first Gaussian becomes 1. The DoG feature for a point (x, y) of the patch P_i is computed as follows.

$$D(x,y) = (G_1(x,y) - G_\sigma(x,y)) \times P_i(x,y) \tag{9}$$

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{-(x^2+y^2)}{2\sigma^2}} \tag{10}$$

where three different scales ($\sigma = 0.5, 2, 4$) are considered as Gaussian filter parameters. The filtering outputs of patch P_i on the three scales are concatenated to extract the features for each patch P_i. As a result, a feature set composed of $432 (= 3 \times 12 \times 12)$ features for a patch of size 12×12 is computed.

2.5 Creation of bag-of-visual-words

Employing patch quality estimation, a set of normalized quality scores is obtained for all the selected patches $\{PN_1, PN_2, \ldots, PN_m\}$. Since normalized quality scores PN_i are real-values between 0 and 1, we uniformly quantize them into L levels, indicated by $q_l = l/L$, for all $l \in \{1, 2, \ldots, L\}$. Based on the normalized quality scores and the quality levels q_l, the selected patches are clustered into Lgroups of similar quality. As a result, L groups $(G_1, G_2, \ldots, G_L)$ are obtained employing a grouping process so that the quality of each group is also known Xue et al. (2013).

The k-means clustering algorithm using the Euclidean distance metric is then performed on each group G_l to create K clusters based on the DoG features extracted from the selected patches. As each cluster is represented by its centroid, in each group G_l a set of K centroids, which are called BoVWs, is obtained. As a result, L sets of BoVWs for L different quality levels are created. Those BoVWs are used to encode the quality of every newly selected patch to finally estimate overall quality of a new test image Xue et al. (2013).

2.6 Cluster assignment

As mentioned, each BoVW at each level l has an identified quality level q_l. To assign a quality score to a patch P_i, its nearest centroid is found at each level l. The Euclidian metric is employed to compute distances between the extracted DoG features from the patch P_i which is denoted as FP_i and the centroids of BoVWs at the level l, which is denoted as $CG_{l,k}$. The same process is applied to the other $L-1$ levels, that is, to find the nearest centroids for all the L levels as follows:

$$d_{i,l} = \min \|FP_i - CG_{l,k}\|, \forall k \in \{1, 2, \ldots, K\} \tag{11}$$

where i values vary from 1 to m, l values vary from 1 to L, which is the number of quality groups, and K is the number of clusters in each group.

For a given patch P_i, the distance vector elements $d_{i,l}$ obtained using the nearest neighbour rule are then sorted in ascending order. A smaller distance $d_{i,l}$ in the sorted distance list indicates the patch P_i should more likely have the same quality level as that of centroid $CG_{l,k}$. However, the quality levels in our codebook are of discrete values. To convert the discrete quality values obtained for each patch to a real value, a weighting average strategy is employed to determine the final quality score of the patch P_i:

$$PQ_i = \frac{\sum_{l=1}^{L} q_l e^{\frac{-d_{i,l}}{\tau}}}{\sum_{l=1}^{L} e^{\frac{-d_{i,l}}{\tau}}} \tag{12}$$

where τ is a control parameter. The patch quality PQ_i computed from the discrete quality levels q_l is a real value between 0 and 1 representing the quality of the selected patch P_i. This real value instead of discrete value q_l provides more accurate quality scores for the selected patches and finally image quality Xue et al. (2013).

2.7 Pooling process

Since, the proposed BDIQA method in this paper is based on the patch selection and patch quality score, a set of patch qualities (PQs) is obtained employing the cluster assignment procedures on the selected patches $(P_1, P_2, \ldots, P_m)$. To estimate the final image quality score (IQS) for the image based on the patch quality scores $(PQ_1, PQ_2, \ldots, PQ_m)$ estimated for the selected patches, a simple average pooling process is employed as follows:

$$IQS = \frac{1}{m}\sum_{i=1}^{m} PQ_i \tag{13}$$

where IQS is the estimated objective image quality score for the input document image ranging between 0 and 1. An IQS close to 0 indicates that the input image has a low quality image, whereas a value close to 1 specifies that the input image is of high quality.

3 Experimental Results and Discussion

3.1 Datasets and metrics of evaluation

Four different datasets were used to assess the performance of the proposed BDIQA (MQAC) method in this research work. To the best of our knowledge, no document-oriented dataset considering human opinion scores as ground truth for quality assessment is available in the literature. Therefore a document-oriented dataset called the ITESOFT dataset was created by the authors for experimentation. The ITESOFT is composed of 29 reference document images collected from real-world data using different capturing devices, such as mobile cameras and scanners. Some examples of documents available in the ITESOFT dataset are shown in Fig. 3. JPEG and JPEG2000 compression methods have been applied with different parameters to obtain 6 levels of distortion for each reference image. Accordingly, 348(= $29 \times 2 \times 6$) distorted images have been generated. MHOS has been provided for each document image based on the HOSs obtained from 23 individuals. To obtain HOSs, each individual was asked to evaluate the quality of a document image between 0 and 100 according to the following main question: "How would you rate the 'useability' of the document, if the document needs to be evaluated/used by a human operator?" Some detailed sub-instructions were also provided to further help the individuals for a better evaluation. The sub-instructions were: (i) "Is the image still informative or too distorted to be recognized by a person?", (ii) "Can the textual part be read easily (without considerable effort) by an individual?", (iii) "Are logos or stamps identifiable by an individual?", (iv) "Is the document still meaningful?", and (v) "Can you evaluate the quality of the information conveyed by the document?"



Figure 3: Some exemples of document images from the ITESOFT dataset.

The other three datasets are LIVE Sheikh (2003), TID2008 Ponomarenko et al. (2009) and CSIQ Larson & Chandler (2010) datasets, which are well-known scene-image datasets used frequently in the literature for evaluating the performances of IQA methods. These datasets were used to demonstrate that the proposed DIQA method is also suitable for estimating the quality of natural scene images.

For training the proposed BDIQA method, 10 images from the Berkeley image database Martin et al. (2001) rather than the ITESOFT, LIVE, TID2008 and CSIQ datasets were randomly selected

and used as reference images in order to validate the generalization and the database-independency of the method. In fact a good characteristic of the BoVWs methods is that we can create a meaningful codebook from a totally different dataset with respect to the ones used for testing MacQueen et al. (1998). Four common type of distortions, such as JPEG compression, JPEG2000 compression, Gaussian noise, and Gaussian blur distortions, were used to create distorted training images. These distortions were chosen for creating the distorted training images, since, JPEG compression, JPEG2000 compression, and Gaussian blur frequently occur in document images as a result of image compression and image capturing. Moreover, Gaussian noise is a basic noise model representing the effect of many random processes that occur in nature. For each image, its distorted versions were created on three quality levels by controlling the noise standard deviation (for distortion of Gaussian noise), the support of blur kernel (for distortion of Gaussian blur), the quality level (for distortion of JPEG compression) and the compression ratio (for distortion of JPEG2000 compression). As a result, a collection of $120(= 3 \times 4 \times 10)$ distorted images and 10 reference images were produced to use for the creation of BoVWs Xue et al. (2013). The choice of three distortion levels should be decided in such a way that the quality scores of the extracted patch samples from the collected images is uniformly distributed at all the quality levels. During the training phase, we initially used all the patches extracted from the image without employing our selection scheme for creating the BoVWs, since we needed to have enough information to create the best quality-aware centroids or BoVWs image quality representative of the training data for the proposed BDIQA. Based on our experimentation, we further noted that using the selected patches could also provide quite promising BoVWs resulting in comparative results. It is worth noting that for the training dataset, no human subjective scores were provided.

Performance of the proposed MQAC metric were evaluated computing the Pearson linear Correlation Coefficient (PCC), Spearman Rank order Correlation coefficient (SRC), and Root Mean Square Error (RMSE) metrics using the results provided by the MQAC method and the MHOS quality measure, which is the ground truth provided by human for images of every dataset. These three metrics have commonly been used for evaluation of various algorithms in the literature Ye et al. (2012); Mittal et al. (2012, 2013). To have better performance and efficiency in any IQA method, PCC and SRC must be close to 1, whereas RMSE must be close to 0. We further measured and reported the time complexity of the proposed MQAC method.

3.2 Parameters of implementation

There are five parameters in the proposed method that may affect the final DIQA results. The first two parameters (s, T) belong to the segmentation and patch selection strategy and the rest of the parameters (K, L, p) are related to the QAC method. To provide a clear view on the impacts of different parameters on the performance of the proposed method, we considered the ITESOFT dataset for the experimentation. The values of the parameters, which provide the best results, were further considered to compute the results on the other three datasets to show the generalization of our proposed metric.

3.2.1 Segmentation and patch selection parameters

In the implementation of the segmentation method Alaei et al. (2011) used in this work, s is the only parameter to be considered. Since the width of the stripes should be small enough to provide appropriate segmentation, some values with respect to the image width (2.5% and 5% of image width) were considered to analyse the impact of the parameter (s) on the results. From the results shown in Table 2, it is clear that the best results were obtained when s was set to 2.5% of the image width. To demonstrate the impact of the parameter T in the patch selection strategy, a range of values between 0.0 and 0.80 were considered for T. By setting $T = 0$, patch selection strategy discards the patches that do not have any foreground pixels. Experimental results are shown in Table 3. From Table 3 it is noted that the best results were obtained when T was set to 0 $(T = 0)$. Hence s and T were, respectively, set to 2.5% of the image width and 0 in our implementation to obtain the results on the other three datasets.

3.2.2 Parameters related to the QAC method

In the implementation of the QAC, 10 quality levels $(L = 10)$ were initially considered for sampling patch qualities. As a result, the worst quality level was between 0 and 0.1 $(q_1 = 0.1)$ and the

	p × p	
Result	2.5% of image width	5% of image width
RMSE	**0.109**	0.113
PCC	**0.891**	0.882
SRC	**0.875**	0.868

Table 2: The results obtained employing the proposed MQAC metric on the ITESOFT dataset based on different stripe-widths

Result	**All the patches**	T=0	T=0.25	T=0.50	T=0.65	T=0.80
RMSE	**0.138**	**0.109***	**0.110***	**0.110***	**0.111***	**0.114***
PCC	**0.817**	**0.891***	**0.888***	**0.888***	**0.886***	**0.880***
SRC	**0.804**	**0.875***	**0.875***	**0.874***	**0.871***	**0.864***

Table 3: The results obtained employing the proposed MQAC method on the ITESOFT dataset concerning different values of T. $*H0$ can be statiscally accepted. The value is equivalent to the mean ($\alpha = 0.05$).

best quality level was between 0.9 and 1 ($q_1 0 = 1$). To study the impact of different groups or quality levels (L), we further chose $L = 20$ to create the BoVWs and performed new experiments. The results obtained from the proposed method, when different values were considered for L, are presented in Table 4. The results shown in Table 4 point to that better performance is achieved when $L = 10$.

	Number of groups	
Result	L=10	L=20
RMSE	**0.109**	0.110
PCC	**0.891**	0.888
SRC	**0.875**	0.874

Table 4: The results obtained employing the MQAC metric on the ITESOFT dataset regarding different number of groups.

To observe the behaviour of the proposed method in relation to the number of clusters (K), values of 30 and 50 were considered for experimentation. From the experiments we noted that there was no important change in the results considering $K = 30$ or $K = 50$. However, the method was computationally less expensive when a smaller number was chosen for K in our implementation. Hence in our implementation L and K were, respectively, set to 10 and 30.

The last parameter in the proposed MCAQ method is the patch-size ($p \times p$) used to sample the image into a number of patches. In our implementation, a number of values from 7 to 12 were considered for p. To have a clear idea about the impact of different patch-sizes in the creation of BoVWs and the estimation of image quality, the results obtained considering different patch-sizes are shown in Table 5. The best result was obtained when p was set to 12 or a patch of size 12×12 was considered for image sampling.

	p × p					
Result	7×7	8×8	9×9	10×10	11×11	12×12
RMSE	0.135	0.121*	0.137*	0.137*	0.151*	**0.109***
PCC	0.828*	0.864*	0.821*	0.820*	0.777	**0.891**
SRC	0.814*	0.844*	0.814*	0.819*	0.780*	**0.875***

Table 5: The results obtained employing the proposed MQAC method considering different patch-size on the ITESOFT dataset. * indicates that the value is equivalent to the mean ($\alpha = 0.05$).

3.3 Experimental results

From the results (Table 2, 3, 4 and 5) obtained from the proposed MQAC on the ITESOFT dataset, it can be observed that most of the results do not significantly change when using different parameters for experimentation. To draw such a conclusion, statistical z-tests were carried out. We considered a null hypothesis of independence H0 between a given PCC score and the mean value over a set of PCC scores obtained with different parameters while assuming an admissible standard deviation of 0.005 (a gap of 0.5% to the mean). We computed, by means of a two-sided statistical hypothesis test, the probability (p-value) of getting a value of the statistic as extreme, or more extreme than observed by chance alone, if $H0$ is true. If a p-value is greater than 1.95 (1.95 can be found using a table of values from the Normal distribution for a risk of 5%), we can say that the hypothesis $H0$ of independence can be rejected with a risk of 5%. In such a case, we can conclude that the mean of the PCC score set is significantly different from a given PCC score. Each parameter value that fails at the test does deviate oddly from the mean of the set. For clarity reason, we chose the PCC metric but the same scheme holds for the other metrics. This observation reveals that the method is quite robust to the variation in parameters' values. From the result shown in Table 3 we can explicitly conclude that the patch selection strategy always produces better results compared to the results of the system when all the patches were used for predicting document image quality. It is also worth mentioning that considering T being bigger than 0 could provide reasonable results compared to the best results obtained with $T = 0$, while reducing drastically the computation time. Furthermore, the results reported in Table 5 demonstrate that comparative DIQA results can be achieved using other patch-sizes, such as 8×8, for image sampling. Particularly when the image sizes in the dataset are of small size (200×300), choosing a patch-size of 8×8 is recommended.

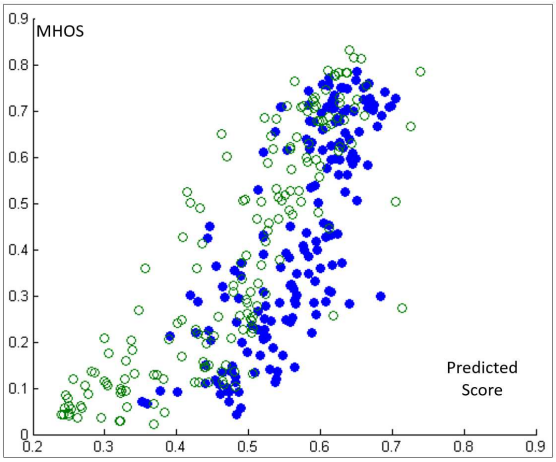


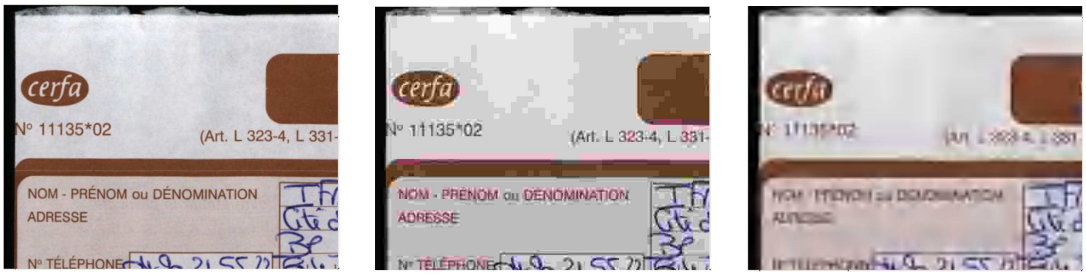
Figure 4: Scatter plot of the predicted quality scores employing our proposed method against subjective MHOS scores on the ITESOFT dataset. Blue and Green indicate JPEG and JPEG2000 distortions respectively.

As the ITESOFT dataset is composed of images with JPEG and JPEG2K distortions, the results of the proposed MQAC metric on each distortion are, furthermore, provided in Table 6. From Table 6 we noted that the proposed MQAC method performs better on the images including JPEG2K distortions, than the images including JPEG distortions. A pictorial demonstration of the results obtained from the proposed MQAC method on the ITESOFT dataset is shown in Fig. 4. From Fig. 4 it is clear that some points are diverging from the regression line. However, outliers are not the majority since both samples (objective and subjective results) fairly correlate with each other ($PCC > 0.85$).

To have a qualitative idea of the effectiveness of the proposed method, some parts of 3 document images from the ITESOFT dataset with different qualities, along with the estimated quality values based on the proposed MQAC metric, are shown in Fig. 5.

	Distortion		
Result	**JPEG**	**JPEG2K**	**All images**
RMSE	0.106	0.104	0.109
PCC	0.881	0.911	0.891
SRC	0.845	0.900	0.875

Table 6: The results obtained employing the proposed MQAC method on the ITESOFT dataset concerning different artifacts when $T = 0$ and the patch size is 12×12.



a) IQS = 0.75 b) IQS = 0.34 c) IQS = 0.29

Figure 5: Some images with their estimated qualities (a) an original image, (b) a JPEG distorted version, and (c) a JPEG2000 distorted version.

3.4 Comparative analysis and discussion

Since, the proposed MQAC method is a blind document image quality assessment metric, the results obtained from the proposed MQAC are compared to two state-of-the-art blind IQA methods. Four different datasets were used to have a variety of results in relation to document and scene images. The comparisons of results are tabulated in Table 7. The bold values indicate the best performances in each dataset. From Table 7, it is evident that in terms of all the three criteria (RMSE, PCC and SRC), the proposed MQAC method outperformed the QAC Xue et al. (2013) method on all four datasets. The improvement of performance (an average of 10% increase in the evaluation metrics) in the ITESOFT dataset is significant, as indicated in Table 7. This is because the proposed MQAC method mainly works on foreground information, which is well adapted to document images. Furthermore, the proposed MQAC method provided better results on the TID2008 Ponomarenko et al. (2009) and CSIQ Larson & Chandler (2010) datasets compared to the NIQE method Mittal et al. (2013). However, the results on the LIVE Sheikh (2003) dataset indicate that the best accuracy was obtained by the NIQE Mittal et al. (2013) method. The reason is that the NIQE Mittal et al. (2013) was adapted on the LIVE dataset. On the ITESOFT dataset, the results obtained from the proposed method are better than the results obtained from the QAC method Xue et al. (2013). Competitive results were obtained from the proposed MQAC method compared to the NIQE method Mittal et al. (2013). However, the NIQE method Mittal et al. (2013) is quite expensive in terms of computation time as it took on average more than 8.2 seconds to predict image quality, whereas computation time of our proposed method was less than 3 seconds on average for each image. This makes our proposed method useful for practical use, especially in the context of document image processing where the sizes of images are quite large.

It is also worth noting that the proposed method is independent of the training data, as the method uses a clustering technique to create the BoVWs from a totally separate dataset. Moreover, as the proposed method uses mostly foreground information to assess image quality, it is computationally fast compared to the state-of-the-art methods suitable for large size images. However, in the proposed method, the concept of BoVWs is used in the training phase that generally results in missing the spatial relationships (context information) among the patches. In addition, an equal importance is given to the selected foreground patches in the proposed DIQA that may cause erroneous results. Furthermore, the performance of the proposed DIQA method may drop when k-means clustering does not converge.

To overcome the problems of the BoVW, the encoding of first and second order statistics called Vector of Locally Aggregated Descriptors (VLAD) Jégou et al. (2010) and Fisher Vector (FV)) can be employed that may increase the performance while decreasing the codebook size Seeland et al.

Dataset	Method	RMSE	PCC	SRC
LIVE	**NIQE**	**10.91**	**0.915**	**0.916**
	QAC	12.80	0.881	0.879
	Proposed MQAC	11.69	0.902	0.901
TID2008	**NIQE**	0.963	0.794	0.783
	QAC	0.804	0.861	0.838
	Proposed MQAC	**0.790**	**0.867**	**0.849**
CSIQ	**NIQE**	0.138	0.873	0.869
	QAC	0.128	0.891	0.846
	Proposed MQAC	**0.128**	**0.892**	**0.849**
ITESOFT	**NIQE**	**0.108**	**0.894**	**0.878**
	QAC	0.138	0.817	0.804
	Proposed MQAC	0.109	0.891	0.875

Table 7: Comparisons of the Results Obtained from our Proposed MQAC, the QAC Xue et al. (2013) and NIQE Mittal et al. (2013) Methods.

(2017). A document specific full reference IQA may also be employed to compute more accurate patches qualities, resulting in a more sophisticated BoVWs. Moreover, second order statistics combined with sparse coding and a pooling strategy may further outperform the Fisher Vectors method Koniusz et al. (2017). To tackle the k-means clustering limitations, a hierarchical k-means, approximate k-means, or accelerated k-means may be used for creating BoVWs in the proposed method.

Since time complexity is an important issue in real-world applications, a time complexity analysis of our proposed method compared to the NIQE Mittal et al. (2013) and QAC Xue et al. (2013) methods is further provided. An experimental study was performed on a Desktop PC having 4GB RAM and Intel Core 2 DUO CPU@3GHz using the ITESOFT dataset for experimentation. Average computation time obtained, employing different methods, are shown in Table 8. As it is evident from Table 8, our proposed MQAC method performs faster than the NIQE Mittal et al. (2013) and QAC Xue et al. (2013) methods. It takes less than 1/2 of the time taken based on the QAC Xue et al. (2013) and approximately 1/3 of the time taken using the NIQE Mittal et al. (2013) to compute image quality for an image from the ITESOFT dataset. The results and computation time for the QAC Xue et al. (2013) and NIEQ Mittal et al. (2013) were computed using the Matlab implementation of the methods available on the authors' webpage.

Method	Time (sec.)	RMSE	D_0
NIQE Xue et al. (2013)	8.21	**0.108**	8.21
QAC Mittal et al. (2013)	6.76	0.138	6.77
Proposed MQAC	**2.96**	0.109	**2.97**

Table 8: The average Computation time and trade-off(D_0) between times and accuracies obtained employing different NR DIQA methods on the ITESOFT dataset.

As there is always a trade-off between accuracy and computation time, we propose a global index (D_0) that shows which method has the best compromise between time and accuracy. The proposed index (D_0) computes Euclidean distances between the origin and the positions where the time and RMSE are specified for all the methods considered for experimentation. The smallest distance belongs to the proposed MQAC method, which indicates the best performance concerning time and accuracy. Furthermore, in Fig. 6 the results of the proposed MQAC method and the other ones considered on the ITESOFT dataset are shown in terms of RMSE index (x-axis) and time (y-axis). The closer the point is to the plot's origin, the better is the method. From Fig. 6 it is noted that the best performance is obtained from the proposed MQAC method. Such a result makes our proposal suitable for expert information systems dealing with high resolution and large size documents.

4 Conclusions and Future Work

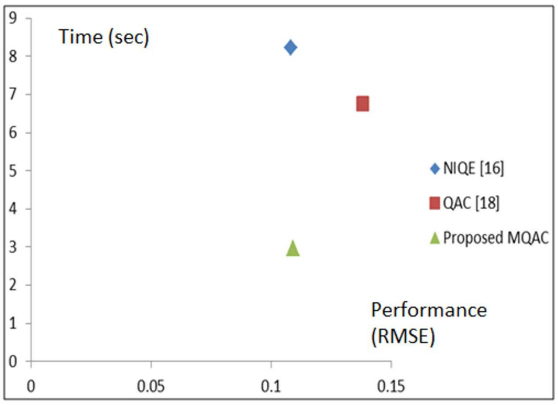


Figure 6: Trade-off between time and accuracy concerning RMSE of the considered NR DIQA methods on the ITESOFT dataset.

In this paper, an expert system (MQAC) is proposed to automatically estimate the quality of document images. The proposed system is independent to the training data, as a set of quality aware BoVWs by employing the k-means clustering technique is created from a completely different data. An automatic foreground/background segmentation technique followed by a patch selection strategy is further proposed to use local foreground information. The experimental results prove that the proposed segmentation and patch selection strategy can improve document image quality assessment results. Computation time of the method is also significantly lower than the state-of-the-art methods, as a small set of selected patches from each document image are used to predict document image quality. The results obtained from the proposed method reveal its suitability for expert systems embedded in mobile applications and applications dealing with large size images.

In future, the use of different features for characterizing the selected patches as well as incorporating spatial relation between the patches can be considered as a direction for further research. The effects of applying different clustering, and sparse encoding methods for learning BoVWs and also impact of using different distance measures for cluster assignment can also be investigated in future to further improve the performance of the proposed system.

Acknowledgments

Authors would like to thank V.P. D'Andecy and L. Guillemot of ITESOFT Company, and Prof. M. Blumenstein from the University of Technology Sydney for their useful suggestions and help.

References

Alaei, A., Conte, D., Blumenstein, M., & Raveaux, R. (2016). Document image quality assessment based on texture similarity index. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on* (pp. 132–137). IEEE.

Alaei, A., Conte, D., & Raveaux, R. (2015). Document image quality assessment based on improved gradient magnitude similarity deviation. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on* (pp. 176–180). IEEE.

Alaei, A., & Delalandre, M. (2014). A complete logo detection/recognition system for document images. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on* (pp. 324–328). IEEE.

Alaei, A., Nagabhushan, P., & Pal, U. (2011). Piece-wise painting technique for line segmentation of unconstrained handwritten text: a specific study with persian text documents. *Pattern Analysis and Applications*, *14*, 381–394.

Bhowmik, T. K., Paquet, T., & Ragot, N. (2014). OCR performance prediction using a bag of allographs and support vector regression. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on* (pp. 202–206). IEEE.

Chandler, D. M. (2013). Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, *2013*.

Mesquita, R. G., Silva, R. M., Mello, C. A., & Miranda, P. B. (2015). Parameter tuning for document image binarization using a racing algorithm. *Expert Systems with Applications*, *42*, 2593–2603.

Verikas, A., Lundström, J., Bacauskiene, M., & Gelzinis, A. (2011). Advances in computational intelligence-based print quality assessment and control in offset colour printing. *Expert Systems with Applications*, *38*, 13441–13447.

Yang, M.-D., Su, T.-C., Pan, N.-F., & Yang, Y.-F. (2011). Systematic image quality assessment for sewer inspection. *Expert Systems with Applications*, *38*, 1766–1776.

Hale, C., & Barney-Smith, E. (2007). Human image preference and document degradation models. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on* (pp. 257–261). IEEE volume 1.

Jégou, H., Douze, M., Schmid, C., & Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 3304–3311). IEEE.

Kang, L., Ye, P., Li, Y., & Doermann, D. (2014). A deep learning approach to document image quality assessment. In *Image Processing (ICIP), 2014 IEEE International Conference on* (pp. 2570–2574). IEEE.

Koniusz, P., Yan, F., Gosselin, P.-H., & Mikolajczyk, K. (2017). Higher-order occurrence pooling for bags-of-words: Visual concept detection. *IEEE transactions on pattern analysis and machine intelligence*, *39*, 313–326.

Kumar, J., Chen, F., & Doermann, D. (2012). Sharpness estimation for document and scene images. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 3292–3295). IEEE.

Larson, E. C., & Chandler, D. (2010). Categorical image quality (csiq) database. *Online, http://vision. okstate. edu/csiq*, .

Lu, H., Kot, A. C., & Shi, Y. Q. (2004). Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters*, *11*, 228–231.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *CAM Journal*, *10*, 31–36.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (pp. 416–423). IEEE volume 2.

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

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*, 209–212.

Nayef, N., & Ogier, J.-M. (2015). Metric-based no-reference quality assessment of heterogeneous document images. In *SPIE/IS&T Electronic Imaging* (pp. 94020L–94020L). International Society for Optics and Photonics.

Obafemi-Ajayi, T., & Agam, G. (2012). Character-based automated human perception quality assessment in document images. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *42*, 584–595.

Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., & Battisti, F. (2009). Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, *10*, 30–45.

Seeland, M., Rzanny, M., Alaqraa, N., Wäldchen, J., & Mäder, P. (2017). Plant species classification using flower images—a comparative study of local feature representations. *PloS one*, *12*, e0170629.

Sheikh, H. R. (2003). Image and video quality assessment research at live. *http://live. ece. utexas. edu/research/quality/*, .

Wang, Z., & Shang, X. (2006). Spatial pooling strategies for perceptual image quality assessment. In *Image Processing, 2006 IEEE International Conference on* (pp. 2945–2948). IEEE.

Xue, W., Zhang, L., & Mou, X. (2013). Learning without human scores for blind image quality assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 995–1002).

Ye, P., & Doermann, D. (2012). Learning features for predicting ocr accuracy. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 3204–3207). IEEE.

Ye, P., & Doermann, D. (2013). Document image quality assessment: A brief survey. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on* (pp. 723–727). IEEE.

Ye, P., Kumar, J., Kang, L., & Doermann, D. (2012). Unsupervised feature learning framework for no-reference image quality assessment. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1098–1105). IEEE.

Young, R. A. (1987). The gaussian derivative model for spatial vision: I. retinal mechanisms. *Spatial vision*, *2*, 273–293.

Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, *20*, 2378–2386.