Document Image Quality Assessment based on Improved Gradient Magnitude Similarity Deviation

Alireza Alaei, Donatello Conte, Romain Raveaux
Laboratoire d'Informatique (LI EA6300), Université François-Rabelais de Tours, France
{alireza.alaei, donatello.conte, romain.raveaux}@univ-tours.fr

***Abstract*—Digitization of business processes and the use of mobile devices as portable scanner lead to a massive production of document images that is beyond manual handling. In such a scenario, automatic estimation of document image quality is a concern in order to adapt as early as possible document image analysis methods. In this paper, a method for full reference document image quality assessment (DIQA) using mainly foreground information is proposed. In the proposed method, a segmentation technique is employed on a reference document image to approximately separate foreground and background information. Foreground information of the document image are then considered in the form of foreground patches for computing image quality. For each foreground patch, corresponding gradient maps, obtained from the reference and distorted gradient magnitude maps, are used to compute a gradient magnitude similarity map of the patch. Gradient magnitude similarity deviation of the patch is then calculated by the means of standard deviation over all the values in the gradient magnitude similarity map obtained for the patch. An average pooling is finally performed on all the standard deviations obtained for all the foreground patches to obtain the final image quality metric of the distorted document image. To evaluate the proposed method, we used 3 different datasets. The first dataset was a dataset composed of 377 document images of which 29 were reference images and 348 were distorted images. The other datasets were LIVE and CSIQ datasets composed of scene images with MHOS as ground truth. The results obtained from the proposed system are encouraging.**

Keywords: Document image, Image quality assessment, Gradient magnitude, Foreground separation, Patch extraction.

I. INTRODUCTION

Evaluating the quality of output images generated by many image processing applications such as image acquisition, compression, restoration and transmission is a challenging task. Quality assessment can be performed subjectively or objectively. In subjective image quality assessment (IQA), the quality measure is based on human beings judgments [9]. With the development of the Internet, and the availability of image capturing devices such as digital cameras, mobile phone cameras and image scanners, the sizes of digital image collections especially digital born document images are increasing rapidly. In such a context, manual and subjective IQA is too tedious, expensive and time-consuming.

Considering current digital technologies and bulk of information in the form of digital scene images, many researchers from image and video processing field have widely worked on the topic of objective IQA and as results many image quality metrics have been introduced in the literature [1, 3, 4, 11-14]. Based on the availability of a reference/original image for measuring the quality of a degraded image, objective IQA metrics can be categorized into three main groups as: a) Full Reference (FR) [3, 4, 12, 14], b) Blind or No-Reference (NR) [1, 11, 13], and c) Reduced-Reference (RR) methods. Most existing approaches for IQA are known as full-reference, meaning that the reference/original image is available for assessing the quality of distorted images.

Concerning document images, volume of digital document images especially digital born documents are increasing day-by-day in the context of new technologies. These documents are mostly color images with different background, texture, and content. However, most of the methods for DIQA focused on two-tone document images and the results of OCRs have commonly been considered as the metric for DIQA [5, 8]. A review of those methods for DIQA has been provided in [9]. Few research works have also been carried out on the topic of objective DIQA in the literature [2, 6, 7, 13] using subjective mean human opinion score (MHOS) as the ground truth for image quality.

In [2], a FR DIQA method based on a simple distance-reciprocal distortion measure has been proposed for binary document images at character level. It has been assumed that distance between two pixels plays important role in their mutual interference perceived by human beings. In [6], a DIQA system based on estimating sharpness of document images (captured by smart-phones) has been proposed. Sharpness estimation is performed separately in the X and Y directions. The difference of differences in gray scale values of a median-filtered image has been used as an indication of edge sharpness. In [7], a system for typewritten DIQA at the character level has been provided. Three groups of features such as morphological-based features, noise removal-based features and spatial characteristics features have been extracted from different characters to train a neural network for estimating qualities of character images. Proposed document image quality measure has been shown to correlate well with human perception. Recently, a method based on unsupervised feature learning has been proposed for NR IQA in [13]. Authors claimed that their method is a general-purpose IQA approach; however, they have not tested their system on document images.

From the above-mentioned literature for DIQA concerning the use of MHOS for quality assessment, it can

978-1-4799-1805-8/15/$31.00 ©2015 IEEE

be noted that only one method so far has specifically been developed for color/gray DIQA [6], which estimates only one type of distortion (blur) in document images. The method presented in [13] seems to be working on document images, but it surely needs some adaptation to fairly work on such data. The rest of the techniques for DIQA work only on binary document images and results of different OCRs need to be considered as evaluation metric for the purpose. Since, most of the document images are generated in color/gray format using existing technologies, converting those document images into binary format and employing a quality image assessment technique on binary documents for obtaining their image qualities may not be feasible in many applications. OCRs results for quality assessment may not also completely fulfill users'/clients' needs and expectations, as users/companies may be interested to have complete outlook of document images without need of performing any OCR application. They may also need to apply some higher level document image processing techniques such as image acquisition, compression, transmission and spotting/retrieval process that mainly deal with color/gray document images.

In such scenarios, it is extremely necessary to develop computational approaches for automatically predicting perceptual document image quality, which also fairly correlates with human subjective evaluation. To fulfill this need, a FR DIQA method based on gradient magnitude feature obtained from foreground information is proposed in this research work. The use of gradient magnitude and foreground information are justified, as foreground generally carries out more important information in document images and it is usually more affected by any of distortion types. Human perception is also more focused on foreground distortion than background one for DIQA. The gradient magnitude characterizes edge distortion which has a big impact in document image quality. In our proposed method, initially, some preprocessing techniques such as color to gray conversion, filtering, and down-sampling are employed to enhance the image quality, and prepare both reference document image and its distorted version for further processing. To obtain the foreground information, a simple but effective foreground/background segmentation method (PPA) [10] is employed to approximately extract foreground in a reference document image. Then, gradient magnitudes of both reference and distorted document images are obtained. Instead of computing the document image quality using whole document image, a set of more informative patches obtained from the foreground information employing the PPA is considered for the purpose. Gradient magnitude similarities for all the extracted foreground patches are computed. The standard deviation of gradient magnitude similarity values in an extracted foreground patch is obtained as similarity deviation of that patch. An average pooling strategy on all the standard deviations obtained for the extracted patches is finally performed to obtain the final document image quality metric of the distorted document image. The contribution presented in this paper is three-fold: a) using a foreground/background separation method to show significance of using foreground information for DIQA, b) proposing a simple patch selection technique, which performs well in both document and scene images, and c) providing a new pooling strategy incorporating both standard deviation and mean statistics to obtain final metric for image quality.

The rest of the paper is laid out as follows: Section II describes our proposed FR DIQA method. Section III discusses the experiments, results and comparative analysis. Finally, Section IV concludes the paper.

II. PROPOSED FR DIQA METHOD

The proposed method is inspired by the gradient magnitude similarity deviation (GMSD) method [14]. The GMSD method is reported to be among the leading metrics in the literature in terms of accuracy and speed introduced for scene images. However, it does not perform well on document images. In this research work, the GMSD is entirely modified and adapted to document images. The modified GMSD, which is hereafter called as MGMSD performs well on document images. Block diagram of our proposed FR DIQA method is demonstrated in Fig. 1. Different steps of the proposed method are described in the consecutive subsections.

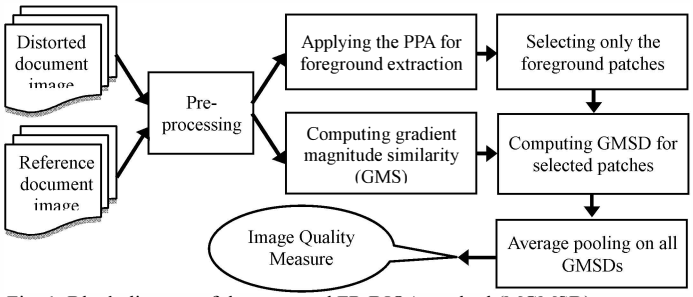


Fig. 1. Block diagram of the proposed FR DIQA method (MGMSD).

A. Pre-processing

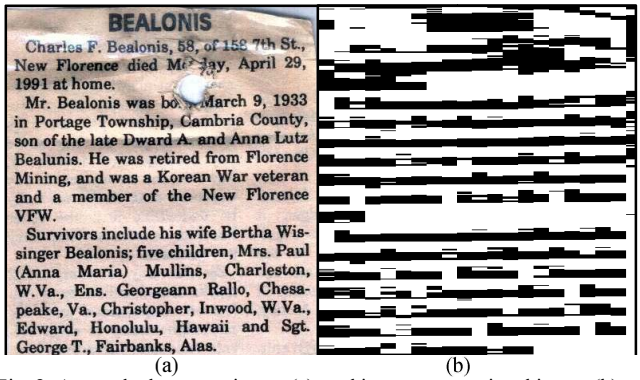
Since a FR DIQA method is proposed in this research work, both reference and distorted document images are necessary to be processed side by side. In pre-processing step, initially, both reference and distorted images are converted into gray-scale images. Then a 2×2 mean/average filter is applied on both images to smooth them [14]. Down-sampling is also performed on both document images in order to speed up the rest of the processes on document images, which are generally of big size. Furthermore, down-sampling has provided higher performance for a number of IQA models including the SSIM [3], FSIM [12] and GMSD [14].

B. Foreground extraction

The Piece-wise Painting Algorithm (PPA) is a novel idea recently introduced for text-line extraction in document images [10]. In this research work, this concept is used to represent a document image by black and white patches of different size. Black patches approximately represent foreground information (text, graphic, etc.) in document image irrespective of its content and shape, whereas white patches roughly signify document image background. Since in document images, foreground data carries more important

information compared to background information; in this research work, it is also assumed that foreground information plays an important role for estimating objective document image quality. Moreover, different artifacts such as those occur because of JPEG and JPEG2000 compression methods mainly affect edge structure of foreground information in document images [14]. Employing the *PAA* provides such information (foreground) to be used for DIQA in our proposed FR DIQA method.

In the PPA, initially, a document image is decomposed into vertical stripes from the left to the right direction. The width of stripes (*s*) is computed based on document image width. Subsequent to the division of the document image into several stripes, the gray value of each pixel in each row of a stripe is changed into the average gray value of all pixels present in that row of the stripe. The resultant gray-scale image is then converted into a two-tone image by applying the Otsu's method. A resultant two-tone image obtained employing the PPA on the image shown in Fig. 2(a) is demonstrated in Fig. 2(b). The black and white patches represent the foreground and background of document image, respectively [10].



(a) (b)

Fig. 2. A sample document image (a), and its two-tone painted image (b) obtained employing the PPA.

C. Gradient magnitude similarity (GMS)

Gradient magnitude in digital images is defined as the root mean square of image directional gradient along horizontal (*X*) and vertical (*Y*) directions. Gradient magnitude typically provides how an image is changing while employing some image processing techniques. Information extracted based on gradient features has been used for FR IQA in the literature [12, 14]. In most of the methods used gradient information for computing an image quality index, a similarity measure has been used to compute the image quality. One of the similarity functions frequently used in many research works is structural similarity (*SSIM*) presented in [3]. In *SSIM*, luminance, contrast, and structural similarities have been computed for every pixel and the product of these three similarities has been considered as a metric for image quality at pixel level. Average pooling has been used to obtain final quality of an image. Feature similarity (*FSIM*) index [12] and GMSD [14] are also two other FR IQA models which use a local similarity map obtained from gradient magnitude features and some other features for predicting image quality.

In this research work similar to the method presented in [14], a Prewitt filter with a *3×3* template gradient is employed to obtain document image gradient. To formulize the problem, let *R* and *D* be a reference document image and a distorted version of the reference image respectively. The Prewitt filters in *X* and *Y* directions are defined as F_X and F_Y. The gradient magnitudes for a pixel located at a position (i, j) in *R* and *D,* which are called as G_R and G_D, are computed employing the following formulas.

$$G_R(i,j) = \sqrt{(R \otimes F_X)^2(i,j) + (R \otimes F_Y)^2(i,j)} \quad (1)$$

$$G_D(i,j) = \sqrt{(D \otimes F_X)^2(i,j) + (D \otimes F_Y)^2(i,j)} \quad (2)$$

where the sign $\otimes$ is the convolution operation. Considering the gradient magnitudes G_R and G_D computed for reference and distorted images, a gradient magnitude similarity (*GMS*) map is computed as local quality map of the distorted image *D*. The *GMS* value for a pixel located at (i, j) is defined in the following.

$$GMS(i,j) = \frac{2 \times G_R(i,j) \times G_D(i,j) + c}{G_R^2(i,j) + G_D^2(i,j) + c} \quad (3)$$

To have a stable result for the *GMS*, a constant *c* is included to the *GMS* equation [12, 14]. The optimum value of *c* is defined in section III. The *GMS* can be computed for a patch as well; since, a patch is a small part of an image.

D. Patch selection and pooling strategy

As demonstrated in Fig. 1, a patch selection strategy is proposed in this research work. To do that, the resultant painted image obtained from the PPA is considered and a connected component labeling method is employed to obtain the foreground information as a set of black patches of different size. The foreground patches are selected and considered for computing document image quality index.

Average pooling has commonly been used to obtain final quality score for an image from its local quality map. More sophisticated pooling strategies such as weighted pooling and standard deviation have also been developed for computing final image quality of an image [12, 14].

In this research work, a combination of the average and standard deviation pooling is proposed to obtain a quality score for an image. The proposed pooling strategy takes the advantages of both average and standard deviation pooling strategies. The standard deviation pooling is employed at the patch level and the average pooling is applied at the image level. To formulize this definition, let $\{P_1, P_2, \ldots, P_k, \ldots, P_m\}$ be *m* patches obtained from the painted image. For each foreground patch say P_k, its corresponding patch in the *GMS* map say $PGMS_k$ is considered and the standard deviation (*STD*) of values in $PGMS_k$ called patch gradient magnitude similarity deviation (*PGMSD*) is computed as follows.

$$PGMSD_k = STD(PGMS_k) \quad (4)$$

Final image quality score (*MGMSD*) for a document image is obtained employing the average pooling as:

$$MGMSD = \frac{1}{m}\sum_{k=1}^{m} PGMSD_k \quad (5)$$

where *m* is the number of foreground patches contributed in the computation of final DIQA score for an image. In other

words, the *MGMSD* is simply figured out employing the average pooling on all the PGMSD values obtained from the selected foreground patches. It is worth mentioning that patches of different size have equal weight in the average pooling for computing final quality score. Higher value for the *MGMSD* means the image is more distorted and correspondingly the image quality is lower.

III. Experimentation Results and Discussion

A. Datasets and metrics of evaluation

To evaluate the performance of the proposed FR DIQA method, three different image datasets were used in this research work. The first one is ITESOFT dataset containing 29 reference document images collected from real world data using different capturing devices such as mobile camera, steel camera, and scanner. JPEG and JPEG2000 at 6 different levels have been applied on reference images to generate 348 (29×2×6) distorted images. MHOS has been provided for each document image based on the HOSs obtained from 23 individuals. The other two datasets are LIVE [15] and CSIQ [16] datasets, which are well-known scene-image datasets in the literature. Some statistics about these datasets are shown in Table I.

To demonstrate the performance of the proposed *MGMSD* method, Pearson linear Correlation Coefficient (PCC), Spearman Rank order Correlation coefficient (SRC), and Root Mean Square Error (RMSE) were computed based on the results provided by the *MGMSD* method and the MHOS provided for images of a dataset as ground truth. These three evaluation metrics have widely been used for evaluation of various algorithms in the literature. Values of PCC and SRC should be near to 1, whereas value of RMSE should be close to 0 for better performance and efficiency of any IQA method.

B. Implementation and parameter settings

There are two parameters in the proposed *MGMSD* technique that need to be tuned during the training phase. The first parameter is the width of stripes (s) in the PPA that should be small enough to take care of intensity variation in an image. Based on the experimentation results obtained from the proposed method using the stripe width as 2.5% and 5% of the image width, there was no significant difference between the final results. Hence, in the implementation, this parameter was set to 5% of the image width. The second parameter is the value of c in equation (3). We considered different values for c and computed the results, we noted that using the value c=170 and c=250 provided the best results on the ITESOFT dataset. Hence, c was set to 170 as in [14].

C. Results and discussion

Concerning three different datasets of document and scene images, results obtained from the proposed method (*MGMSD*) are tabulated in Tables II, III and IV. From Table II, it is noted that the RMSE obtained from the proposed FR DIQA method on a document-based dataset (ITESOFT) is quite good, since it is close to 0. PCC metric obtained for ITESOFT document images is 0.917, which means the results are highly correlated with the human opinion scores. To have a clear idea about the behavior of the proposed method on different type of distortions/artifacts, details of the results obtained for the JPEG and JPEG2000 distortions present in the ITESOFT dataset are shown in Table III. From Table III, it can be noticed that the proposed method performs better on the distorted document images generated using the JPEG2000 compression technique compared to the JPEG compression technique.

To have a visual perspective of the proposed method performance on ITESOFT dataset, a scatter plot of the predicted quality scores obtained from our proposed method against subjective MHOS scores of the ITESOFT dataset is provided in Fig. 3. From Fig. 3, it is evident that the proposed method produces less accurate quality prediction results when the distortion is severe or in other words when the subjective MHOS values are very small (less than 0.1). It is also clear that the distributions of residual values of predicted scores for JPEG and JPEG2000 distortions show almost the same behavior with respect to the MHOS scores.

TABLE I. SOME STATISTICS OF THE DATASETS USED FOR EXPERIMENTATION.

Database	No. of Reference Images	No. of Distorted Images	No. of Distortion Types	Type of Images	No. of Observers
ITESOFT	29	348	2	Document/Color	23
LIVE [15]	29	779	5	Scene/Color	161
CSIQ [16]	30	866	6	Scene/Color	35

TABLE II. THE RESULTS OBTAINED EMPLOYING THE PROPOSED FR DIQA METHOD ON THE ITESOFT DATASET.

Result	Stripe Size = 5% of image width
RMSE	0.096
PCC	0.917
SRC	0.916

TABLE III. THE RESULTS OBTAINED EMPLOYING THE PROPOSED METHOD ON THE ITESOFT DATASET CONCERNING DIFFERENT ARTIFACTS.

Result	JPEG	JPEG2000	All images
RMSE	0.097	0.090	0.096
PCC	0.902	0.935	0.917
SRC	0.895	0.934	0.916

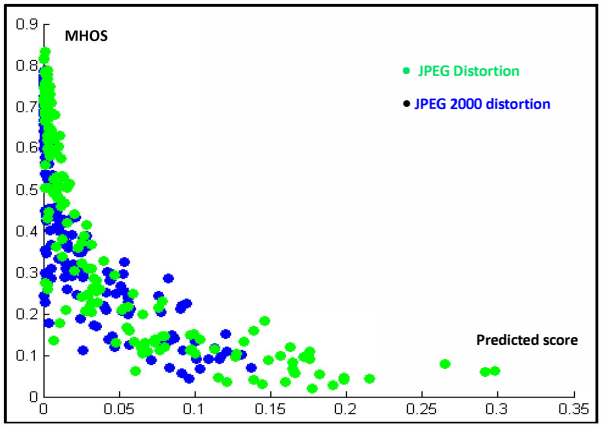


Fig. 3. Scatter plot of the predicted quality scores employing our proposed method against subjective MHOS scores on the ITESOFT dataset.

D. Comparative analysis

To the best of our knowledge, there is no public dataset composed of document images for DIQA to obtain/compare

the results of the proposed method in this paper. However, there are some publicly available scene-image datasets such as LIVE and CSIQ. In this research work, the results obtained employing the proposed DIQA method (*MGMSD*) on the ITESOFT dataset as a propertied document-based dataset and two public datasets of scene images are compared to the results achieved based on three state-of-the-art IQA methods named PSNR, SSIM [3] and GMSD [14]. The GMSD was reported to be one of the fastest and most accurate methods in the literature for IQA on scene images [14]. All the results were computed in terms of RMSE, PCC and SRC measures and they are tabulated in Table IV.

TABLE IV. COMPARISON OF THE RESULTS OBTAINED FROM OUR PROPOSED *MGMSD* METHOD, PSNR, SSIM [3] AND GMSD [14].

Dataset	Method	RMSE	PCC	SRC
CSIQ [16]	PSNR	0.173	0.751	0.806
	SSIM [3]	0.133	0.861	0.876
	GMSD [14]	0.079	0.954	0.957
	Proposed MGMSD	0.086	0.945	0.949
LIVE [15]	PSNR	13.36	0.872	0.876
	SSIM [3]	8.95	0.945	0.948
	GMSD [14]	7.62	0.960	0.960
	Proposed MGMSD	8.82	0.946	0.951
ITESOFT	PSNR	0.147	0.791	0.796
	SSIM [3]	0.129	0.843	0.837
	GMSD [14]	0.122	0.860	0.839
	Proposed MGMSD	**0.096**	**0.917**	**0.916**

From Table IV, it is noted that the proposed *MGMSD* method outperforms the GMSD [14], SSIM [3] and PSNR on ITESOFT dataset. The PCC and SRC obtained from the proposed *MGMSD* method using the ITESOFT dataset are 0.917 and 0.916 respectively. More than 8% improvement on both metrics is obtained by the proposed method compared to the SSIM [3]. The improvement is more than 5% compared to the GMSD [14]. Concerning the LIVE and CSIQ datasets, the results obtained from the proposed *MGMSD* method are better than the results obtained from the SSIM [3] and PSNR metrics, which are the metrics commonly used for natural images. The results obtained from the proposed *MGMSD* method on LIVE and CSIQ datasets are slightly lower compared to the results achieved from GMSD technique [14], but the results are pretty comparable. These results were expected; since in this research work, a segmentation technique mostly dedicated to document images was used for the foreground extraction. For such complex scene images, an advanced segmentation technique might be needed to obtain an accurate segmentation and consequently more accurate IQA results. It is also worth mentioning that the proposed *MGMSD* method uses only a portion of images (foreground information) for computing image quality index. Concerning this fact, the results obtained for the scene images are also quite encouraging. It further conveys that using foreground information (a part) of even a scene image can provide enough information for measuring the image quality.

Regarding execution time of the Matlab implementation of the proposed method on a Desktop PC having 4GB RAM and Intel Core 2 DUO CPU@3GHz, we noted that it took on an average 864 ms for each image of ITESOFT dataset to compute the quality score. This time was 426 ms, 400 ms, and 381 ms for the SSIM, GMSD and PSNR methods respectively. From the experiment results it is clear that the proposed MGMSD provided better accuracy compared to the GMSD, SSIM and PSNR. However, it is slower than the GMSD, SSIM and PSNR methods, as the proposed metric is utilizing the PPA and connected component labeling for foreground patch extraction.

IV. CONCLUSION AND FUTURE WORK

A FR DIQA method, which also performs well on scene images, is proposed in this research work. The method is based on the hypothesis that foreground information influences the human perception in measuring the image quality especially in document images. The experimental results obtained employing the proposed method on both document and scene image datasets prove the applicability and suitability of the hypothesis. The effect of employing some new feature extraction techniques for characterizing local patches to accurately predict image quality especially on document images and also proposing a weighting strategy based on patch-size are our future research work.

REFERENCES

[1] H. Tong, M. Li, H.-J. Zhang, C. Zhang "No-reference quality assessment for JPEG2000 compressed images," in Proc. of the ICIP, pp. 3539-3542, 2004.
[2] H. Lu, A. C. Kot, Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images," IEEE Signal Processing Letters, 11(2), pp. 228-231, 2004.
[3] Z. Wang, A. C. Bovik, B. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, 13(4), pp. 600-612, 2004.
[4] Z.Wang, A. C. Bovik, E. P. Simoncelli, "Structural approaches to image quality assessment," in Handbook of Image and Video Processing, pp. 961-974, Academic Press, 2nd Ed., 2005.
[5] C. Hale, E. H. Barney Smith, "Human image preference and document degradation models," in Proc. of the 9th ICDAR, pp. 257-261, 2007.
[6] J. Kumar, F. Chen, D. Doermann, "Sharpness estimation for document and scene images," in Proc. of the 21st ICPR, pp. 3292-3295, 2012.
[7] T. Obafemi-Ajayi, G. Agam, "Character-based automated human perception quality assessment in document images," IEEE Trans. On SMC-Part A: Systems and Humans, 42(3), pp. 584-595, 2012.
[8] P. Ye, D. Doermann, "Learning features for predicting OCR accuracy," in Proc. of the 21st ICPR, pp. 3204-3207, 2012.
[9] P. Ye, D. Doermann, "Document image quality assessment: a brief survey," in Proc. of the 12th ICDAR, pp. 723-727, 2013.
[10] A. Alaei, U. Pal, P. Nagabhushan, "A new scheme for unconstrained handwritten text-line segmentation", Pattern Recognition 44(4), pp. 917-928, 2011.
[11] A. Mittal, G. S. Muralidhar, J. Ghosh, A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," IEEE Signal Processing Letters, 19(2), pp. 75-78, 2012.
[12] L. Zhang, L. Zhang, X. Mou, D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. on Image Processing, 20(8), pp. 2378-2386, 2011.
[13] P. Ye, J. Kumar, L. Kang, D. S. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in Proc. of the CVPR, pp. 1098-1105, 2012.
[14] W. Xue, L. Zhang, X. Mou, A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Trans. on Image Processing, 23(2), pp. 684-695, 2014.
[15] H.R. Sheikh, K. Seshadrinathan, A.K. Moorthy, Z. Wang, A.C. Bovik, L.K. Cormack, "Image and video quality assessment research at LIVE," http://live.ece.utexas.edu/research/quality.
[16] E.C. Larson, D.M. Chandler, "Categorical Image Quality (CSIQ) Database," http://vision.okstate.edu/csiq.