Evaluation of Image Quality Metrics for Sharpness Enhancement

Yao Cheng*†, Marius Pedersen†, and Guangxue Chen*
*State Key Laboratory of Pulp and Paper Engineering,
South China University of Technology, Guangzhou, China
†The Norwegian Color and Visual Computing Laboratory, Department of Computer Science,
NTNU - Norwegian University of Science and Technology, Gjøvik, Norway
Email: marius.pedersen@ntnu.no

***Abstract*—Image quality assessment has become a meaningful research field due to the explosive growth of image processing technologies in imaging industries. It is becoming more usual to quantify the quality of an image using image quality metrics, rather than carrying out time-consuming psychometric experiments. However, there is little research on the performance of image quality metrics on quality enhanced images. In this paper, we focus on images that have been enhanced by sharpening. A psychometric experiment was designed with observers giving scores to different images enhanced by sharpening on a display in a controlled dark environment. The results showed that full reference image quality metrics performed well when sharpening did not improve the visual image quality, while in images where sharpening increased the visual quality the performance was lower. No reference image quality metrics show better predictions than full reference image quality metrics in most cases.**

I. Introduction

Recently, our daily life depicts a situation that we are surrounded by a tremendous amount of images. Images become a vital information carrier compared with graphics and words [1], [2]. Image quality (IQ) is used to describe how good the image is, which can be understood as the amount of distortion referred to original image (deemed as the highest quality image). However, even without comparison to the original image, human observers still can distinguish objects, background, foreground, contour, texture and so on in the image, and then give a perceptual quality score to it.

Image enhancement has been widely applied in image processing to improve the appearance of images. The principal objective of image enhancement is modifying image attributes to get a more pleasing output in a given case [3]. However, the effects of image enhancement have not been studied in detail so far. This is a challenge in objective IQ assessment, as discussed in [4], [5]. As shown by Zhang et al. [6], observers in general prefer a sharpened image to the original image since the average level of preferred sharpness is consistently higher than the detection threshold across image contents and subjects. Traditionally, the way to improve sharpness is with respect to the edges of images [7], [8], [9].

There are generally two methods of evaluating IQ: objective and subjective. Objective methods are using IQ metrics to evaluate the quality of images, while subjective methods are based on human observers giving scores to or rank the images according to a specific guideline. Since humans are the final receiver of images, the correlation between objective results and psychometric results has been used as a performance evaluation of the objective methods.

In this paper, we designed a psychometric experiment where observers are giving scores to images with different sharpness levels on a display in a controlled environment. The goal is to compare the results of the human observers with the results of state-of-the-art IQ metrics. This is done by calculating the correlation between the psychometric scores and the scores from IQ metrics.

This paper is organized as follows; Section II summarizes the basic method of image enhancement on sharpness and a selection of existing IQ metrics. Section III introduces the experimental setup including the selection of test images, viewing condition and experimental method. Section IV provides the subjective and objective results and an analysis and discussion of the results. Lastly, in Section V, conclusions and future works are presented.

II. Background

A. Sharpness enhancement

Image enhancement aims to improve the perceptual IQ or to get a better output for future image processing, such as image analysis, image detection, image segmentation and image recognition. In a given situation, image enhancement can reach its objective by modifying IQ attributes. There are many IQ attributes used to describe the quality of an image [10], [11], such as sharpness, contrast, color, lightness, and artifacts. Sharpness is an important attribute, which usually relates to the definition of edges and visibility of details [10].

The basic method of sharpening images is by using Unsharp Masking (USM) [6]. As the name implies, USM enhances edges through subtracting an unsharp version of image (since edges can be treated as the high frequency signal, so the unsharp version of image can be found by applying a low-pass filter) by the original image, then adds this part back to original image. USM has many advantages such as it is a linear space-invariant filter, which can be easily implemented as a spatial-domain convolution. It is computationally inexpensive and robust. However, it may also have some drawbacks. It can result in overshoot and undershoot to the edges, which can produce halo artifacts. Since it cannot recognize noise, which may amplify the background noise in smooth regions. Last, it cannot sharpen all edges since it uses a fixed sharpening strength.

B. Image quality metrics

IQ metrics have three main categories depending on accessibility of the original image [12]. Metrics that are using the original image in addition to the distorted image are commonly referred to as full reference (FR) IQ metrics. Reduced reference (RR) metrics uses only partial information about the images. The last category is no reference (NR) IQ metrics, which uses only the distorted image to determine IQ. Many of these IQ metrics are based on the human visual system (HVS), and they have the ultimate goal of predicting perceived IQ.

1) Full reference metrics: More and more IQ metrics have been applied for IQ (surveys can be found in [13], [14], [15]), and FR metrics become increasingly mature. FR metrics can be divided into two groups: Pixel-based and HVS-based. For the former, the earliest IQ metrics are the Mean squared error (MSE) and Peak Signal to Noise Ratio (PSNR), which are computing the distance between corresponding pixels in the reference and distorted images. In these cases, the assessment is usually not correlated with perceptual IQ. For the other group, there are two kinds of framework [16]. First is the bottom-up framework, which needs to simulate the processes of the HVS. For example, S-CIELAB is a spatial extension to the CIELAB color metric, which applied a spatial filtering operation to simulate the spatial blurring of the HVS, at the same time consistent with the basic CIELAB calculation for large uniform areas [17]. Adaptive Bilateral Filter (ABF) is used for color image difference evaluation, which avoided the undesirable loss of edge information introduced by filtering using contrast sensitivity functions [18]. Spatial Hue Angle MEtric (SHAME) took into account the HVS by incorporating information about region of interest [19]. The Total Variation of Difference (TVD) metric [20] removes information imperceptible to the observer, and then calculates the difference between the original and reproduction. The other framework is the top-down framework, which models the overall function of the HVS given a special condition. For example, Structure SIMilarity (SSIM) index [21] is based on the degree of structure similarity in the reference and distorted images. Visual Information Fidelity (VIF) [22], [23] depicts the connection between image information and IQ depending on natural scenes statistical (NSS). Visual Signal-to-Noise Ratio (VSNR) [24] employs a wavelet-based model to determine distortions compared to the threshold of visual detection. Feature Similarity (FSIM) Index [25] uses phase congruence and gradient magnitude as features to characterize the image local quality. Gradient Magnitude Similarity Deviation (GMSD) [16] computes a local quality map by comparing the gradient magnitude maps of the reference and distorted image, and uses standard deviation to obtain the final IQ score. Amirshahi et al. [26] proposed an IQ metric based on features extracted from Convolutional Neural Networks (CNNs), which produced good results on different databases. Zhao et al. [27] evaluated IQ metrics for perceived sharpness of projection displays, where the images were blurred, and found that SSIM, FSIM and VIF produced good results.

2) No reference metrics: Recent NR metrics are based on the assumption that the distortion types are known (such as blocking artifacts [28], blur and noise [29], [30], [31], JPEG [32] or JPEG2000 compression [33], [34], and others [35], [36]). There is an important notion proposed by Ferzli and Karam: Just-Noticeable Blur (JNB) [37], [38], which takes into account the response of the HVS to sharpness at different contrast levels. The derived HVS-based sharpness perception model is used to predict the relative perceived sharpness in images with different content [39]. Later more and more research is based on the JNB [40], [41], [42], [43], [44]. The other way to predict a certain distortion is transform-based, such as Discrete cosine transform (DCT) [45] and Discrete wavelet transform (DWT) [46], [47]. Local Phase Coherence - Sharpness Index (LPC-SI) proposed by Hassen et al. [48], [49], which identifies sharpness as strong local phase coherence (LPC) near distinctive image features evaluated in the complex wavelet transform domain.

However, human observers do not know exactly what the distortions are in the images. Therefore, more and more NR metrics based on statistic characteristics are proposed. Moorthy and Bovik [50] proposed a two-step framework for NR IQ assessment based on natural scene statistics (NSS): the blind image quality index (BIQI) [51]. The first stage is a classification stage, which is based on a description of distorted image statistics to classify an image into a particular distortion category. In their demonstration, this set consists of JPEG, JPEG2000 (JP2K), white noise (WN), Gaussian Blur (Blur) and Fast fading (FF) and it can be extended to any number of distortions. The second stage evaluates the IQ along the amount or probability of each of these distortions, so the quality score is expressed as a probability-weighted summation. Mittal et al. [52], [53] designed a NSS based Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE), which extracts the point wise statistics of local normalized coefficients of luminance signals in the spatial domain, as well as pairwise products of adjacent normalized luminance coefficients which provide distortion orientation information. These coefficients can be used as statistical features that correlate well with human judgments of IQ. Saad et al. [54] introduced the BLIINDS index (BLind Image Integrity Notator using DCT Statistics), which is based on predicting IQ through observing the statistics of local DCT coefficients of a number of features (contrast, structure, sharpness and orientation anisotropies). Mittal et al. derived the Natural Image Quality Evaluator (NIQE) [55], [56], which is based on the construction of a quality aware collection of statistical features based on a simple and successful space domain NSS model. Besides, there are also some machine learning methods. Li et al. [57] developed a general regression neural network (GRNN), which is trained by related perceptual features (as phase congruency, entropy and image gradient), to estimate IQ by approximating the functional relationship between these features and subjective scores.

C. Performance measures for image quality metrics

In order to assess the performance of IQ metrics, it is common to calculate the correlation between the observer results and the IQ metrics. There are commonly two different correlation coefficients used for this:

1) Pearson's correlation coefficient: Pearson's correlation coefficient r assumes linear relationship between two random samples X and Y:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{1}$$

where, $x_1, ..., x_n$ belong to sample X and x_i represents one of them. $y_1, ..., y_n$ belongs to sample Y and y_i represents one of them. n is the number in the samples. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and similar for $\bar{y}$. The range of value r is $[-1, 1]$. If the value is higher than 0, then X and Y have positive association; if the value is lower than 0, then X and Y have negative association; while value equals to 0, X and Y have no association.

2) Spearman's rank correlation coefficient: Spearman's rank correlation coefficient ρ assumes monotonic relationship between samples X and Y:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{2}$$

where, $x_1, ..., x_n$ belongs to sample X and x_i represents one of them. $y_1, ..., y_n$ belong to sample Y and y_i represents one of them. n is the number in the samples. $d_i = x_i - y_i$, is the difference between ranks. Spearman's ρ is a non-parametric coefficient. When X and Y have strictly monotone increasing relationship, the value of ρ is 1; When X and Y has strictly monotone decreasing relationship, the value of ρ is -1; while ρ equals to 0, Y tends to be flat when X is increasing.

III. Experimental Setup

In our experiment, six test images (Fig. 1) from the Colourlab Image Database: Image Quality [58] were selected. These images are selected since they contain different characteristics; such as fine details, sharp edges, lines, texture and so on. We generate five different levels of sharpness by using the Matlab function *imsharpen* altering the *radius* parameter as 0 (the original image), 1, 5, 20, and 50.

22 naive human observers (10 men and 12 women, age 20-27), following the recommendation of minimum 15 observers by CIE [59] and ITU [60], were invited to give perceptual ratings to the different levels sharpness of images using a five force-choice based category judgment. The five categories, bad, poor, fair, good, and excellent, were represented by numbers from 1 to 5. A higher value means the higher quality. The raw data from the experiment was processed into Z-scores [61] using the Colour Engineering Toolbox [62].

We have followed the CIE guidelines [59] with regards to viewing conditions on display. The chromaticity of the white displayed on colour monitor has been set to CIE standard illuminant D65 and the luminance level of the white displayed on the monitor has been set to 80 cd/m^2. The two steps were calibrated by using Eye-one device before the experiment. The experiment was conducted in a dark environment. The viewing distance was approximately 54 cm and the images were shown in real size on the display, calibrated to sRGB.

Before the experiment, the visual acuity of the observers was evaluated. In order to show those different sharpness levels of each image randomly to get more accurate judgment from human observers, we designed a Matlab GUI to present the images to the observers. Once the observer gave score to one image, he/she continued to the next image. The observers were not informed about the changes done to the images.

Based on the IQ metrics described in Section II, we choose two FR metrics [63] (SSIM [21], VIF [22]) and four NR metrics (JNBM [39], LPC-SI [48], BRISQUE [52], [53], NIQE [55], [56]) to predict IQ. These metrics can be considered to be state of the art, and has shown to perform well in existing evaluation studies.

IV. Results and Analysis

First we introduce the results from the psychometric experiment, then the evaluation results of the IQ metrics.

A. Subjective Results

The Z-scores from the psychometric experiment are shown in Fig. 2. For the turtle image, there is a tendency that the IQ can be improved by a certain amount sharpening. For flowers and buildings, the IQ tends to decrease as the sharpness level increases. For mountain, sunflower and leaves, the IQ tends to increase when the sharpness level is increasing. This is an indication that the preferred amount of sharpening is dependent on the image. For all images the results indicate that some amount of sharpening is preferred among the observers.

B. Objective Results

We will assess the performance of the metrics by investigation of their correlation with the percept (in this case the Z-scores). As we can see in Fig. 3, when we use FR metrics to predict the IQ, it turns out that only those images which were distorted (i.e. having a lower quality) after sharpening have a high Pearson correlation coefficient. SSIM is good at predicting the quality of the images flowers, buildings and turtle. The VIF metric is good at predicting quality of the images flowers and mountain. FR metrics using both reference images and tested images as input, and they assume that the original image is pristine. Therefore, when the quality is increased by sharpening, FR metrics cannot predict perceived IQ. The exception is that the VIF metric is giving a high correlation for the mountain image, despite the fact that the observers generally consider the original image (without sharpening) having a lower quality compared to the sharpened versions. The results for Spearman coefficients are very similar to those of Pearson.

As for NR metrics when it comes to Pearson correlation, JNBM provides equivalent results to BRISQUE and LPC-SI for the sunflower image. LPC-SI metric is suitable for mountain, sunflower and leaves. BRISQUE gives good results for flower, sunflower and turtle. NIQE has a high correlation metric for leaves. For Spearman the results are similar, but we can notice that that BRISQUE has a lower rank correlation coefficient in the flowers image than for Pearson, the same can be seen for LPC-SI for the leaves image.

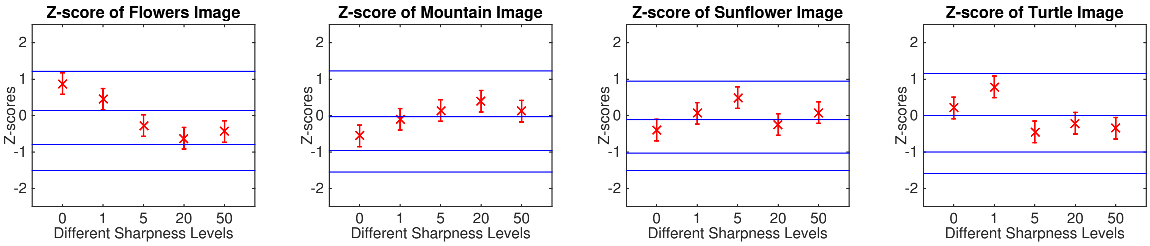
Overall it is interesting to notice that none of the metrics perform well for in all six test images, but there is always one IQ metric that produces acceptable results. This might indicate that the selection of the IQ metric to be applied could be linked to the content of the image.

V. Conclusion and future work

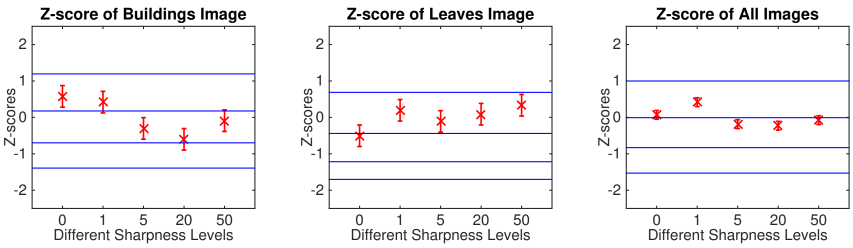
In this work, we designed a psychometric experiment to study the performance of existing image quality metrics for enhanced images by using human observer evaluations as references. The subjective results indicate that the preferred

(a) (b) (c) (d) (e) (f)

Fig. 1. Six selected test images from the Colourlab Image Database: Image quality [58] are used for the psychometric experiment. Each image has a resolution 800 × 800 pixels. (a) Flowers, (b) Mountain, (c) Sunflower, (d) Turtle, (e) Buildings, (f) Leaves. They were sharpened in Matlab with *imsharpen* function with the *radius* parameter varying from 1, 5, 20, to 50. This results in each image having five level of sharpness including the original image.



(a) (b) (c) (d)



(e) (f) (g)

Fig. 2. The Z-scores of perceptual ratings collected from 22 human observers based on 5 sharpening levels of 6 test images. All the figures' horizontal axis represent the increasing sharpness, where 0 refers to the original image, and 1, 5, 20, 50 represent the image sharpened from original image with corresponding radius. The vertical axis represents the Z-score value, which has the range [-2.5, 2.5]. The cross markers are mean Z-scores for each images, and the red lines through markers represent 95% confidence interval. The blue horizontal lines separate vertical area into five categories, i.e., excellent to bad from top to bottom. (g) shows Z-scores of all test images.

amount of sharpness can be linked with content. Evaluation of full reference metrics showed that they performed well when sharpening did not improve the visual quality of the images, while in images where sharpening increase the visual quality the performance was lower. In most cases no reference metrics showed better predictions than full reference metrics.

There are many future works to be included. For example, additional test images and image quality metrics should be added. Besides, choosing other image attributes to extend the range of the evaluations.

VI. Acknowledgements

The project is supported by Science and Technology Planning Project of Guangdong Province, China (No.2013B090600060). This research has been funded by the Research Council of Norway through project no. 221073 HyPerCept Colour and quality in higher dimensions.

References

[1] Vassilis Athitsos, Michael Swain, and Charles Frankel. Distinguishing photographs and graphics on the world wide web. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 10–17, 1997.

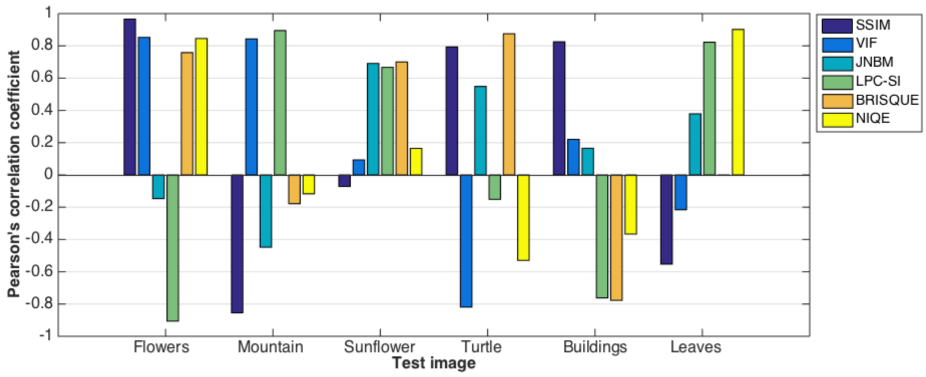
[2] Nahum Gershon, Stephen G Eick, and Stuart Card. Information visualization. *interactions*, 5(2):9–15, 1998.

[3] SS Bedi and Rati Khandelwal. Various image enhancement techniques-a critical review. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(3), 2013.

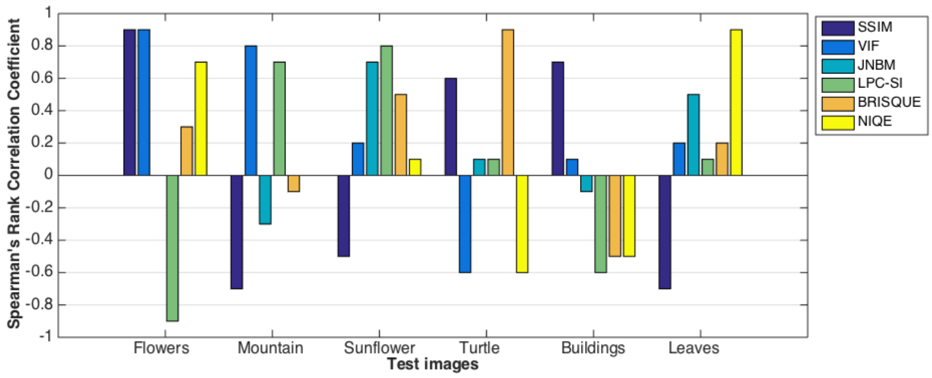
[4] Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, 2013:53 pages, ID 905685, 2013.

[5] Damon M Chandler, Md M Alam, and Thien D Phan. Seven challenges for image quality research. In B. E. Rogowitz, T. N. Pappas, and H. de Ridder, editors, *Human Vision and Electronic Imaging XIX*, pages 901402–901402, San Francisco, CA, Feb 2014.

[6] Buyue Zhang, Jan P Allebach, and Zygmunt Pizlo. An investigation of perceived sharpness and sharpness metrics. In Rene Rasmussen



(a)



(b)

Fig. 3. (a) Pearson's correlation coefficients for different test images, (b) Spearman's rank correlation coefficients for different images.

and Yoichi Miyake, editors, *Image Quality and System Performance II*, pages 98–110, San Jose, CA, USA, Jan 2005.

[7] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.

[8] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217, 1980.

[9] J.F. Canny. Finding edges and lines in images. Technical report, No. AI-TR-720. MIT Artifical Intelligence Laboratory, 1983.

[10] Marius Pedersen, Nicolas Bonnier, Jon Yngve Hardeberg, and Fritz Albregtsen. Attributes of image quality for color prints. *Journal of Electronic Imaging*, 19(1):011016–1–011016–13, Jan 2010.

[11] Marius Pedersen, Nicolas Bonnier, Jon Y Hardeberg, and Fritz Albregtsen. Image quality metrics for the evaluation of print quality. In *Image quality and system performance VIII*, pages 786702–786702, Jan. 2011.

[12] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[13] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *Image Processing, IEEE Transactions on*, 15(11):3440–3451, 2006.

[14] Lin Zhang, Lei Zhang, Xuanqin Mou, and Dejing Zhang. A comprehensive evaluation of full reference image quality assessment algorithms. In *International Conference on Image Processing*, pages 1477–1480. IEEE, 2012.

[15] Marius Pedersen and Jon Yngve Hardeberg. Full-reference image quality metrics: Classification and evaluation. *Foundations and Trends® in Computer Graphics and Vision*, 7(1):1–80, Jan 2012.

[16] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *Image Processing, IEEE Transactions on*, 23(2):684–695, 2014.

[17] X. Zhang and B.A. Wandell. A spatial extension of cielab for digital color-image reproduction. *Journal of the Society for Information Display*, 5(1):61–63, 1997.

[18] Zhaohui Wang and Jon Yngve Hardeberg. Development of an adaptive bilateral filter for evaluating color image difference. *Journal of Electronic Imaging*, 21(2):023021–1, 2012.

[19] Marius Pedersen and Jon Yngve Hardeberg. A new spatial filtering based image difference metric based on hue angle weighting. *Journal of Imaging Science and Technology*, 56(5):50501–1–50501–12(12), September 2012.

[20] Marius Pedersen and Ivar Farup. Improving the robustness to image scale of the total variation of difference metric. In *Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on*, pages 116–121. IEEE, 2016.

[21] Zhou Wang, Alan Bovik, Hamid Rahim Sheikh, and Eero Simoncelli.

Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[22] Hamid Rahim Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[23] Hamid Rahim Sheikh, Alan C Bovik, and Gustavo De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.

[24] Damon M Chandler and Sheila S Hemami. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *Image Processing, IEEE Transactions on*, 16(9):2284–2298, 2007.

[25] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: a feature similarity index for image quality assessment. *Image Processing, IEEE Transactions on*, 20(8):2378–2386, 2011.

[26] Seyed Ali Amirshahi, Marius Pedersen, and Stella X Yu. Image quality assessment by comparing cnn features between images. *Journal of Imaging Science and Technology*, 60(6):60410–1, 2016.

[27] Ping Zhao, Yao Cheng, and Marius Pedersen. Objective assessment of perceived sharpness of projection displays with a calibrated camera. In *2015 Colour and Visual Computing Symposium (CVCS)*, pages 1–6, Aug 2015.

[28] Zhou Wang, Alan C Bovik, and BL Evan. Blind measurement of blocking artifacts in images. In *International Conference on Image Processing*, volume 3, pages 981–984. IEEE, 2000.

[29] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. A no-reference perceptual blur metric. In *International Conference on Image Processing.*, volume 3, pages III–57–III–60, Rochester, NY, Sep 2002. IEEE.

[30] Xiang Zhu and Peyman Milanfar. A no-reference sharpness metric sensitive to blur and noise. In *International Workshop on Quality of Multimedia Experience*, pages 64–69, San Diego, CA, Jul 2009. IEEE.

[31] Min Goo Choi, Jung Hoon Jung, and Jae Wook Jeon. No-reference image quality assessment using blur and noise. *International Journal of Computer Science and Engineering*, 3(2):76–80, 2009.

[32] Zhou Wang, Hamid R Sheikh, and Alan C Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *International Conference on Image Processing*, pages I–477–I–480. IEEE, 2002.

[33] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. *Signal processing: Image communication*, 19(2):163–172, 2004.

[34] Hamid Rahim Sheikh, Alan Conrad Bovik, and Lawrence Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *Image Processing, IEEE Transactions on*, 14(11):1918–1927, 2005.

[35] Zhou Wang and Alan C Bovik. A universal image quality index. *Signal Processing Letters, IEEE*, 9(3):81–84, 2002.

[36] Ji Shen, Qin Li, and Gordon Erlebacher. Hybrid no-reference natural image quality assessment of noisy, blurry, jpeg2000, and jpeg images. *Image Processing, IEEE Transactions on*, 20(8):2089–2098, 2011.

[37] Rony Ferzli and Lina J Karam. Human visual system based no-reference objective image sharpness metric. In *International Conference on Image Processing*, pages 2949–2952. IEEE, 2006.

[38] Rony Ferzli and Lina J Karam. A no-reference objective image sharpness metric based on just-noticeable blur and probability summation. In *International Conference on Image Processing*, pages 445–448, 2007.

[39] Rony Ferzli and Lina J Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Transactions on Image Processing*, 18(4):717–728, 2009.

[40] Nabil G Sadaka, Lina J Karam, Rony Ferzli, and Glen P Abousleman. A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling. In *IEEE International Conference on Image Processing*, pages 369–372, 2008.

[41] Niranjan D Narvekar and Lina J Karam. A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In *International Workshop on Quality of Multimedia Experience*, pages 87–91, San Diego, CA, USA, Jul 2009. IEEE.

[42] Niranjan D Narvekar and Lina J Karam. A no-reference image blur metric based on the cumulative probability of blur detection (cpbd). *IEEE Transactions on Image Processing*, 20(9):2678–2683, 2011.

[43] Sheng-hua Zhong, Yan Liu, Yang Liu, and Fu-lai Chung. A semantic no-reference image sharpness metric based on top-down and bottom-up saliency map modeling. In *International Conference on Image Processing*, pages 1553–1556, Hong Kong, Sep 2010. IEEE.

[44] Christoph Feichtenhofer, Hannes Fassold, and Peter Schallauer. A perceptual image sharpness metric based on local edge gradient analysis. *Signal Processing Letters, IEEE*, 20(4):379–382, 2013.

[45] Shizhong Liu and Alan C Bovik. Efficient dct-domain blind measurement and reduction of blocking artifacts. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1139–1149, 2002.

[46] Hanghang Tong, Mingjing Li, Hongjiang Zhang, and Changshui Zhang. Blur detection for digital images using wavelet transform. In *IEEE International Conference on Multimedia and Expo*, volume 1, pages 17–20. IEEE, 2004.

[47] Rony Ferzli and Lina J Karam. No-reference objective wavelet based noise immune image sharpness metric. In *International Conference on Image Processing*, pages I–405. IEEE, 2005.

[48] Rania Hassen, Zhou Wang, Magdy M Salama, et al. Image sharpness assessment based on local phase coherence. *Image Processing, IEEE Transactions on*, 22(7):2798–2810, 2013.

[49] Rania Hassen, Zhou Wang, and Magdy Salama. No-reference image sharpness assessment based on local phase coherence measurement. In *International Conference on Acoustics Speech and Signal Processing*, pages 2434–2437, Dallas, TX, USA, Mar 2010. IEEE.

[50] Anush K Moorthy and Alan C Bovik. A two-step framework for constructing blind image quality indices. *Signal Processing Letters, IEEE*, 17(5):513–516, 2010.

[51] Anush K Moorthy and Alan C Bovik. BIQI software release 2009. http://live.ece.utexas.edu/research/quality/. Visited:2015-11-25.

[52] Anish Mittal, Anush K Moorthy, and Alan C Bovik. Blind/referenceless image spatial quality evaluator. In *Asilomar Conference on Signals, Systems and Computers*, pages 723–727. IEEE, Nov 2011.

[53] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *Image Processing, IEEE Transactions on*, 21(12):4695–4708, 2012.

[54] Michele Saad, Alan C Bovik, Christophe Charrier, et al. A dct statistics-based blind image quality index. *Signal Processing Letters, IEEE*, 17(6):583–586, 2010.

[55] Anish Mittal, Ravi Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *Signal Processing Letters, IEEE*, 20(3):209–212, 2013.

[56] Anish Mittal, Ravi Soundararajan, and Alan C Bovik. NIQE software release 2012. http://live.ece.utexas.edu/research/quality/. Visited:2015-11-25.

[57] Chaofeng Li, Alan Conrad Bovik, and Xiaojun Wu. Blind image quality assessment using a general regression neural network. *Neural Networks, IEEE Transactions on*, 22(5):793–799, 2011.

[58] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. CID: IQ–a new image quality database. In *Image and Signal Processing*, pages 193–202. Springer, 2014.

[59] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report ISBN: 3-901-906-26-6, CIE TC8-03, 156:2004.

[60] International Telecommunication Union. Recommendation itu-r bt.500-11: Methodology for the subjective assessment of the quality of television pictures. Technical report, International Telecommunication Union/ITU Radiocommunication Sector, 2009.

[61] Peter G Engeldrum. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek press, 2000.

[62] Phil Green and Lindsay MacDonald. *Colour engineering: achieving device independent colour*, volume 30. John Wiley & Sons, 2011.

[63] Marius Pedersen. Evaluation of 60 full-reference image quality metrics on the CID:IQ. In *Int. Conference on Image Processing*, pages 1588–1592. IEEE, 2015.