https://doi.org/10.2352/EI.2023.35.8.IQSP-301

Image Demosaicing: subjective analysis and evaluation of image quality metrics

Tawsin Uddin Ahmed, Seyed Ali Amirshahi, and Marius Pedersen
Colourlab, Norwegian University of Science and Technology, Gjøvik, Norway
tawsin.uddin@gmail.com, s.ali.amirshahi@ntnu.no, marius.pedersen@ntnu.no

Abstract

Most cameras use a single-sensor arrangement with a color filter array. Color interpolation techniques performed during image demosaicing can be the reason behind visual artifacts generated in a captured image. While the severity of the artifacts depends on the demosaicing methods used, the artifacts themselves are mainly zipper artifacts (block artifacts across the edges) and false-color distortions. In this study and to evaluate the performance of demosaicing methods, a subjective pair-comparison method with 15 observers was performed on six different methods (namely Nearest Neighbours, Bilinear interpolation, Laplacian, Adaptive Laplacian, Smooth hue transition, and Gradient-Based image interpolation) and nine different scenes. The subjective scores and scene images are then collected as a dataset and used to evaluate a set of no-reference image quality metrics. Assessment of the performance of these image quality metrics in terms of correlation with the subjective scores shows that most of the evaluated no-reference metrics cannot predict perceived image quality.

Introduction

A single matrix Charge-Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) is employed in most cameras to measure color at each pixel. To capture colors, a Color Filter Array (CFA) covers the plane of a CCD or CMOS sensor that is integrated into modern digital cameras. The sensor's photodetectors assess the intensity of light, while CFAs separate the light wavelength into red, green, and blue color components. The demosaicing technique produces a full-color image from a raw sensor Bayer image captured using a single sensor array covered with a color filter array. When transforming from a Bayer image to an RGB image, demosaicing algorithms are designed to fill in the empty pixels. Demosaicing techniques employ different methods to estimate the missing data; nevertheless, because the data is approximated, artifacts may appear in the final image, raising the need for evaluating image quality.

In this study, we first aim to conduct a subjective analysis on simulated images from the ISETCam toolbox [1] to investigate the quality of the images generated using different demosaicing approaches. Then, using the subjective data collected, we aim to evaluate the performance of different no-reference image quality metrics.

The content of this article is organized in the following way: first we give a summary of related research conducted in image quality assessment of demosaiced images. Then we introduce the methodology, before we present the results, and in the final section we conclude and present possible future directions.

Related Works

Lu et al. [2] provide two additions to CFA demosaicing. That is, a more effective image demosaicing technique for producing images with improved quality and an advanced method to assess the performance of a demosaicing technique. The proposed demosaicing technique involves two phases: an approximation phase that uses spatial and spectral associations within surrounding pixels to approximate empty pixel information and a post-processing step that employs median filtering to decrease obvious demosaicing distortions. ΔE^*_{ab} and PSNR is used to evaluate fidelity, and a new measure is proposed to quantify zipper artifacts. They conclude that these measures are useful for evalutating demosaicing algorithms.

Lamb et al. [3] subjectively evaluated four different demosacing algorithms on 31 reference images. The algorithms used were Bilinear Interpolation, Freeman, Alternating Projection (AP), and High-quality linear interpolation. The analysis of the results indicates that blur and color halos are the most important quality aspects.

Sergej et al. [4] conducted subjective experiments in which observers manually marked visible artifacts on demosaiced images. These subjective markings were further compared to the results from image quality metrics, namely SSIM, HDR-VDP2, S-CIELAB, and MSE. HDR-VDP2 was best correlated with subjective markings.

Gasparini et al. [5] investigated how distortions introduced after demosaicing impact image quality and suggested a new no-reference metric to evaluate them. A subjective experiment found blur and the zipper pattern to be important, and this was incorporated into their no-reference metric. The metric was evaluated by 9 observers on 10 images reproduced by 9 algorithms. The metric was shown to produce a good correlation with subjective ratings.

Methodology

Simulation of a camera imaging pipeline

ISETCam [1] is a Matlab toolbox that experts can use to evaluate image quality and simulate imaging systems. We use this toolbox to simulate a camera imaging pipeline. Six common and exemplary demosaicing approaches are considered for this research, each with a different level of complexity and working mechanism. nearest-neighbour, bilinear interpolation, smooth hue transition, gradient-based color interpolation, Laplacian, and adaptive Laplacian are the image demosaicing techniques taken into account in this work.

The color information of the nearest neighbor is used to fill the blank pixels in nearest-neighbor interpolation. This type of interpolation produces unappealing blocky effects and is rarely utilized unless a fast execution is required. An empty pixel is filled via bilinear interpolation which is the average of its non/empty neighboring pixel [6]. There are obvious chroma artifacts in places containing detail in bilinear interpolation where color information has been interpolated and the detail is somewhat displaced across color channels. We can decrease chrominance artifacts by interpolating hue values instead of just the chrominance values if hues are characterized as the proportions of chrominance and luminance assuming that we have previously approximated the luminance. The green channel is used to estimate luminance, while the red and blue channels are used to estimate chrominance in the smooth hue transition method [7]. This is sufficient since 50 percent of the pixels collected are green as the human visual system is more sensitive to the green wavelength in the visual spectrum. The hue approximation uses the green channel, which is bilinearly interpolated by

$$G = \hat{G} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{1}$$

The relative intensity of red is determined with respect to the green channel using

$$R = G^{pt.wise} \times \left((\hat{R}^{pt.wise} \div G) * \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \right) \tag{2}$$

which is also used for blue channel estimation. In EQ. (2), $\hat{R}$, $\hat{B}$, $\hat{G}$ refers to the Red, Blue, and Green layes of the Bayer image and R, G, B correspond to the constructed RGB image channels after interpolation.

A common problem among the different techniques discussed so far is that they approximate color information over edges, resulting in less sharp borders in the output images and distortions across the edges. With the adaptive approximation of the green pixels based on local horizontal and vertical gradients, the gradient-based interpolation approach tries to prevent interpolating over edges. The horizontal and vertical gradient can be determined by calculating the second derivative in both directions. In doing so, the artifacts along the edges are minimized. Using

$$H_{x,y} = \left|\frac{S_{x,y-2} + S_{x,y+2}}{2} - S_{x,y}\right|, \tag{3}$$

and

$$V_{x,y} = \left|\frac{S_{x-2,y} + S_{x+2,y}}{2} - S_{x,y}\right|, \tag{4}$$

the second derivatives in the horizontal (H) and vertical (V) direction can be calculated. It should be noted that all S values for any conceivable x and y result from the same pixel. The green pixels can be approximated using

$$G_{x,y} = \begin{cases} \hat{G}_{x,y}, & \text{if } S_{x,y} \text{is green} \\ \frac{\hat{G}_{x,y-1} + \hat{G}_{x,y+1}}{2}, & \text{if } H_{x,y} < V_{x,y} \\ \frac{\hat{G}_{x-1,y} + \hat{G}_{x+1,y}}{2}, & \text{if } H_{x,y} > V_{x,y} \\ \frac{\hat{G}_{x,y-1} + \hat{G}_{x,y} + \hat{G}_{x-1,y} + \hat{G}_{x+1,y}}{4}, & \text{if } H_{x,y} = V_{x,y} \end{cases}. \tag{5}$$

After applying edge-aware interpolation on the green channel, gradient-based algorithm approximates only the red and green chrominance by eliminating the luminance component G. Then red and blue chrominance interpolation can be applied with no gradient consideration due to unnoticeable edge approximation. Finally, both the red and blue channels are reconstructed by adding the green channel.

Figure 1. *Reference images used for ISETCam demosaicing simulation.*

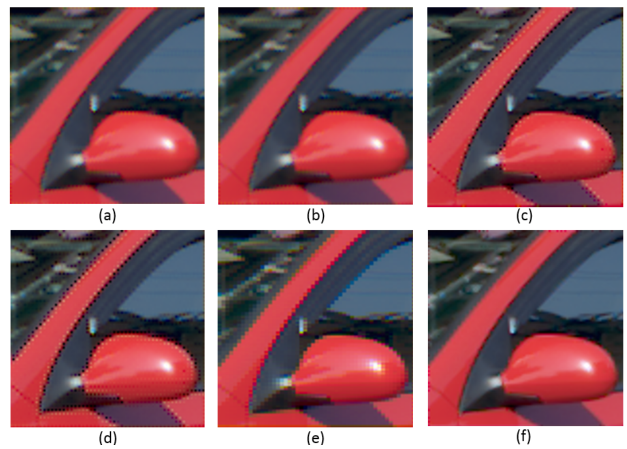

Figure 2. *Cropped versions of six demosaiced images simulated using six distinct image demosaicing algorithms (a) Bilinear Interpolation (BI) (b) Smooth Hue Transition (c) Adaptive Laplacian (d) Laplacian (e) Nearest Neighbours and (f) Gradient-Based color interpolation*

Dataset Collection and Preprocessing

For conducting our subjective experiment, the Colourlab Image Database: Image Quality (CID:IQ) [8] was used. CID:IQ contains 23 reference images spanning a wide variety of aspects, such as spatial information and colorfulness. However, the reference images for our work were selected with a couple of parameters in mind. First, the duration of the subjective experiment as observers would attend the experiment and it should not be tedious for them. Seven images from the original dataset wass selected for the final experiment (1). Second, generally, the artifacts of the demosaiced images are observed mostly on the edge of the images, and so the images taken into account for the experiment have a comparatively higher number of vertical or horizontal edges. In addition to the CID:IQ dataset, a single image was collected from the LIVE [9] dataset and another was captured by the authors (Fig. 3). However, portions of the images were cropped at the size of a 400x400 pixel window to allow observers to focus on a specific region with more edge information. Then six demosaicing algorithms were applied to those cropped images (Fig. 2).

After applying the demosaicing algorithms during imaging system simulation, two types of image artifacts were induced.

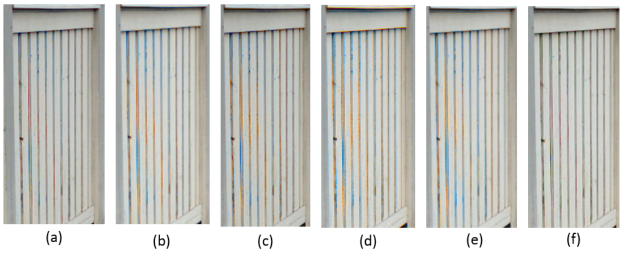

Figure 3. The false-color distortion is introduced with a variety of levels in the demosaiced images. The demosaicing methods are (a) Adaptive Laplacian (b) Bilinear Interpolation (BI) (c) Laplacian (d) Nearest Neighbours (e) Smooth Hue Transition (f) Gradient-Based color interpolation

The zipper artifacts (Fig. 2) where the zipper distortions along the edges are generated at different levels by the demosaicing algorithms. The false-color distortion where in the case of one of our reference images we can notice the false-color artifacts along the vertical edges with also level variation after applying the same color interpolation methods (Fig. 3).

Psychophysical Experiment

The subjective experiment is conducted in a controlled environment where all observers are provided with an identical environment or experimental setup during the experiment. A calibrated display with 80 cd/m^2 and D65 white point is used for the experiment. To control the viewing distance a chin rest is also placed at a distance of 60cm from the calibrated display (Fig. 4).

The subjective experiment is carried out in a paired comparison format [10]. In this pairwise comparison, two images, generated by two distinct demosaicing techniques, are placed side by side. The experiment consisted of 135 pair comparisons. A total of 15 observers participated in this experiment following the recommendation by CIE [11]. Observers were instructed to select the image which has the perceptually better quality. The sequence of pair comparison is random for all observers. A neutral gray background is placed behind each of the image pairs. The whole experiment is hosted on the QuickEval [12] platform. The ratings from the observers were converted into z-scores [13] (Fig. 5).

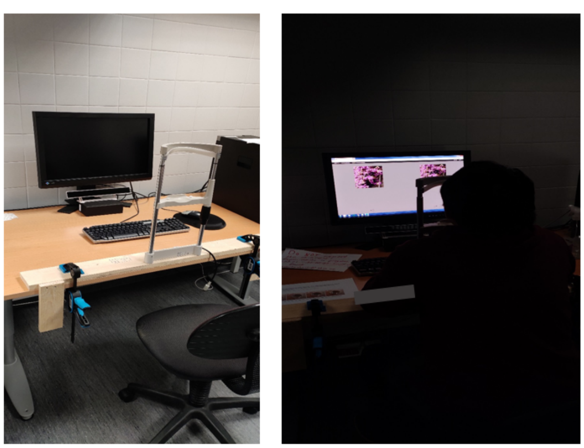

Figure 4. Subjective experiment set-up consisting of calibrated display and a chin wrist to hold observers' viewing direction stability

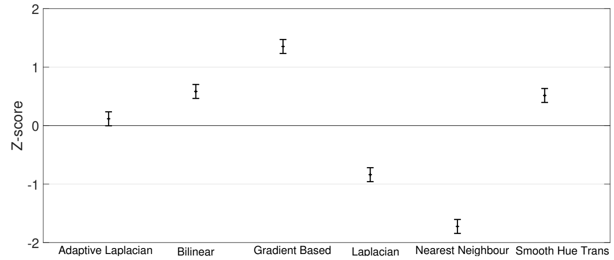

Figure 5. Z-scores from subjective experiment for all images.

Result and Discussion

Subjective results

From the z-scores calculated, we can see that the gradient based interpolation produces the most preferred images by the observers while the nearest neighbourhood interpolation produces the least preferred images. Investigation of the individual images show that there are differences between the images based on content. In general, observers have been able to differentiate between the interpolation methods.

Almost all the demosaiced algorithms produce chromatic distortion in the image that has some kind of repetitive patterns. This kind of artifact can be addressed as false color distortion. Bilinear, Laplacian, nearest-neighbor, and smooth hue transition introduce this pattern of distortion in the image, which is significantly reduced with adaptive Laplacian (Fig. 3). However, Gradient-based color interpolation is able to avoid this kind of chromatic artifact.

Image Quality Metrics

We have selected 35 state of the art no-reference IQMs. These are ARISMC [14], ARISML [14], BIQAA [15], BIQI [16], BIQME [17], BLIINDS2 [18], BlurMetric [19], BQMS [20], BRISQUE [21], CPBDM [22], ContrastNoReference [23], EBCM [24], ENIQA [25], FRIQUEE [26], GIF [27], HOSA [28], IEDD [29], ILNIQUE [30], JNBM [31], JNDDCT [32], JPEG2000 [33], JPEGF [34], JPEGQS [35], JPEGS [35], LPCGray [36], NFERM [37], NIQMC [17], NJQA [38], NR-JPEG2000 [39], PSI [40], QCCE [41], SF [42], SISBLIM [43], SPARISH [44], and niqe [45]. Linear Pearson (Fig. 6) and Spearman (Fig. 7) correlation was calculated for each of the IQMs and the subjective scores collected for the dataset.

From the results (Fig. 6) we can see that JNBM has the highest Pearson correlation coefficient and that many IQMs in general perform poorly. This indicates that the dataset is difficult for most IQMs. The analysis of the Spearman correlation (Fig. 7) shows a similar trend, with JNBM having a slightly lower coefficient. Further analysis indicated that JNBM predicts images with the lowest subjective scores more correctly, which gives a correlation between a cluster of images with the lowest subjective scores and the rest, but that the rank order within these is lower. Overall, the dataset is a challenging task for the IQMs.

We also calculated the linear Pearson correlation per image, where the correlation between the IQMs and the subjective scores for each of the six demoasaiced versions of each image has been calculated (Fig. 8). We can see that JNBM provides higher correlation coefficients for each image. There are also IQMs that have higher correlation coefficients for some images, such as QCCE

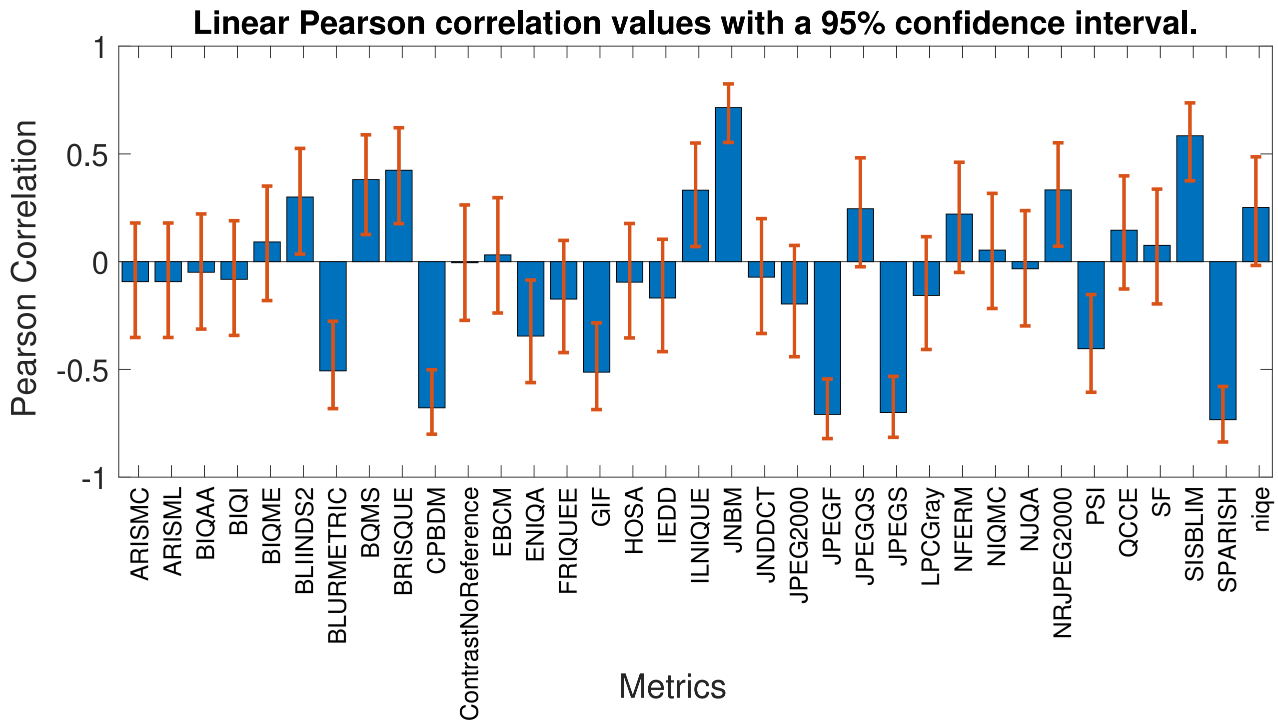

Figure 6. *Linear Pearson correlation between the subjective scores (z-scores) and IQM values. Higher value indicate better performance.*

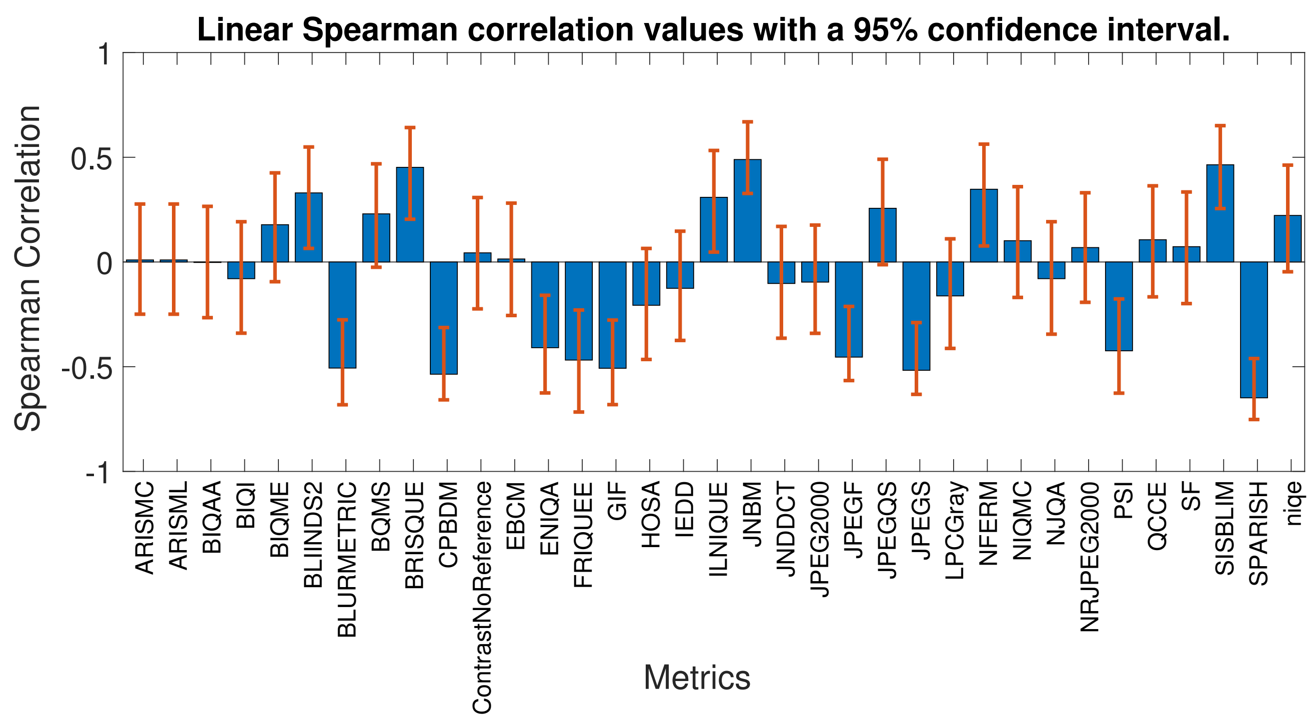

Figure 7. *Linear Spearman correlation between the subjective scores (z-scores) and IQM values. Higher value indicate better performance.*

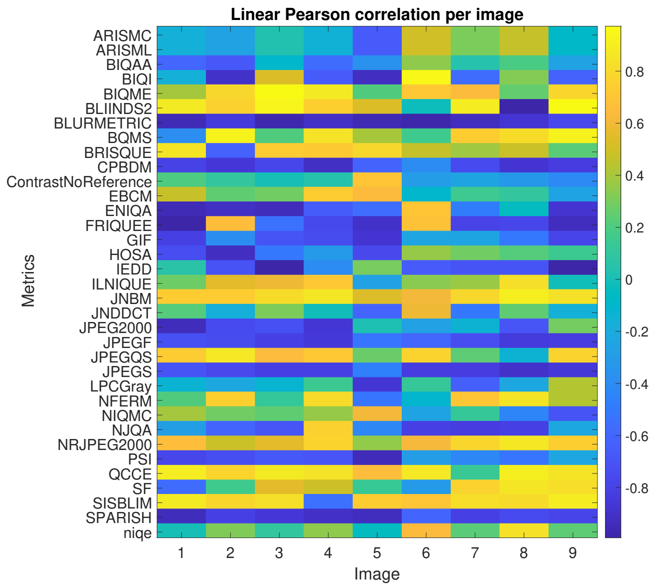

Figure 8. *Linear Pearson correlation for each image in the dataset for each IQM. For each value there are six datapoints. Higher value indicate better performance.*

and SISBLIM. QCCE has a lower overall Pearson correlation (Fig. 6) which indicates scale differences between images, which has been reported for other IQMs in the previous works [46].

Conclusion

The objective of this research includes formulating a psychometric experiment in a controlled environment for investigating several demosaicing algorithms. The images, used in the subjective experiment, are generated using a camera imaging pipeline with ISETCam. Demosaicing algorithms are incorporated while transforming sensor images to RGB images. In most cases, the gradient-based demosaicing technique provides visually more pleasant images while nearest neighbour interpolation produces the comparatively low-perceptual quality images. However, the content of an image also plays a significant role, which can be explored more in the future. The dataset containing images and subjective scores has also been used to evaluate no-reference image quality metrics to see if metrics can predict perceived image quality. The results indicate that many metrics are not capable of predicting perceived image quality. Some metrics have a higher correlation with perceived quality. The dataset can be downloaded from `www.colourlab.no`.

References

[1] Joyce E Farrell, Feng Xiao, Peter Bert Catrysse, and Brian A Wandell. A simulation tool for evaluating digital camera image quality. In *Image Quality and System Performance*, volume 5294, pages 124–131. International Society for Optics and Photonics, 2003.

[2] Wenmiao Lu and Yap-Peng Tan. Color filter array demosaicking: new method and performance measures. *IEEE transactions on image processing*, 12(10):1194–1210, 2003.

[3] Anupama B Lamb and Madhuri Khambete. Image quality assessment database for demosaicing artifacts. In *2018 International Conference on Communication and Signal Processing (ICCSP)*, pages 1100–1105. IEEE, 2018.

[4] Tomasz Sergej and Radosław Mantiuk. Perceptual evaluation of demosaicing artefacts. In *International Conference Image Analysis and Recognition*, pages 38–45. Springer, 2014.

[5] Francesca Gasparini, Mirko Guarnera, Fabrizio Marini, and Raimondo Schettini. No-reference metrics for demosaicing. In *Image Quality and System Performance VII*, volume 7529, page 752911. International Society for Optics and Photonics, 2010.

[6] Lanlan Chang and Yap-Peng Tan. Hybrid color filter array demosaicking for effective artifact suppression. *journal of Electronic Imaging*, 15(1):013003, 2006.

[7] Zhiqiang Li and Peter Z Revesz. Bilinear and smooth hue transition interpolation-based bayer filter designs for digital cameras. *International Journal of Circuits, Systems and Signal Processing*, 9:211–221, 2015.

[8] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. Cid: Iq–a new image quality database. In *International Conference on Image and Signal Processing*, pages 193–202. Springer, 2014.

[9] HR Sheikh. Live image quality assessment database release 2. *http://live. ece. utexas. edu/research/quality*, 2005.

[10] Maria Perez-Ortiz and Rafal K Mantiuk. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686*, 2017.

[11] Jan Morovic. Guidelines for the evaluation of gamut mapping algorithms. *PUBLICATIONS-COMMISSION INTERNATIONALE DE L ECLAIRAGE CIE*, 153:D8–6, 2003.

[12] Khai Van Ngo, Christopher André Dokkeberg, Ivar Farup, Marius Pedersen, et al. Quickeval: a web application for psychometric scaling experiments. In *Image Quality and System Performance XII*, volume 9396, page 93960O. International Society for Optics and Photonics, 2015.

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

[14] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang. No-reference image sharpness assessment in autoregressive parameter space. *IEEE Transactions on Image Processing*, 24(10):3218–3231, 2015.

[15] Salvador Gabarda and Gabriel Cristóbal. Blind image quality assessment through anisotropy. *JOSA A*, 24(12):B42–B51, 2007.

[16] AK Moorthy and AC Bovik. A modular framework for constructing blind universal quality indices. *IEEE Signal Processing Letters*, 17:7, 2009.

[17] Ke Gu, Weisi Lin, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Chang Wen Chen. No-reference quality metric of contrast-distorted images based on information maximization. *IEEE transactions on cybernetics*, 47(12):4559–4565, 2016.

[18] Michele A Saad, Alan C Bovik, and Christophe Charrier. Dct statistics model-based blind image quality assessment. In *2011 18th IEEE International Conference on Image Processing*, pages 3093–3096. IEEE, 2011.

[19] Frederique Crete, Thierry Dolmiere, Patricia Ladret, and

Marina Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In *Human vision and electronic imaging XII*, volume 6492, pages 196–206. SPIE, 2007.

[20] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang. Learning a blind quality evaluation engine of screen content images. *Neurocomputing*, 196:140–149, 2016.

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

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

[23] Yuming Fang, Kede Ma, Zhou Wang, Weisi Lin, Zhijun Fang, and Guangtao Zhai. No-reference quality assessment of contrast-distorted images based on natural scene statistics. *IEEE Signal Processing Letters*, 22(7):838–842, 2014.

[24] Azeddine Beghdadi and Alain Le Negrate. Contrast enhancement technique based on local detection of edges. *Computer vision, graphics, and image processing*, 46(2):162–174, 1989.

[25] Xiaoqiao Chen, Qingyi Zhang, Manhui Lin, Guangyi Yang, and Chu He. No-reference color image quality assessment: from entropy to perceptual quality. *EURASIP Journal on Image and Video Processing*, 2019(1):1–14, 2019.

[26] Lixiong Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 40:1–15, 2016.

[27] Ying Lingkai, Li Ziyin, and Zhang Congcong. No-reference sharpness assessment with fusion of gradient information and hvs filter. *Journal of image and graphics*, 20(11):1446–1452, 2015.

[28] Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25(9):4444–4457, 2016.

[29] Mykola Ponomarenko, Nikolay Gapon, Viacheslav Voronin, and Karen Egiazarian. Blind estimation of white gaussian noise variance in highly textured images. *Electronic Imaging*, 2018(13):382–1, 2018.

[30] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.

[31] 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.

[32] XH Zhang, WS Lin, and Ping Xue. Improved estimation for just-noticeable visual distortion. *Signal Processing*, 85(4):795–808, 2005.

[33] Hamid R Sheikh, Alan C Bovik, and Lawrence Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE Transactions on image processing*, 14(11):1918–1927, 2005.

[34] JPEG F NR. `https://github.com/scienstanford/iqmetrics/blob/master/noRef/JPEG_F_NR.m`.

[35] Zhou Wang, Hamid R Sheikh, and Alan C Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Proceedings. International conference on image processing*, volume 1, pages I–I. IEEE, 2002.

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

[37] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*, 17(1):50–63, 2014.

[38] S Alireza Golestaneh and Damon M Chandler. No-reference quality assessment of jpeg images via a quality relevance map. *IEEE signal processing letters*, 21(2):155–158, 2013.

[39] ZM Parvez Sazzad, Yoshikazu Kawayoke, and Yuukou Horita. No reference image quality assessment for jpeg2000 based on spatial features. *Signal Processing: Image Communication*, 23(4):257–268, 2008.

[40] 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.

[41] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Qcce: Quality constrained co-saliency estimation for common object detection. In *2015 Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2015.

[42] Shutao Li, James T Kwok, and Yaonan Wang. Combination of images with diverse focuses using the spatial frequency. *Information fusion*, 2(3):169–176, 2001.

[43] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Transactions on Broadcasting*, 60(3):555–567, 2014.

[44] Leida Li, Dong Wu, Jinjian Wu, Haoliang Li, Weisi Lin, and Alex C Kot. Image sharpness assessment by sparse representation. *IEEE Transactions on Multimedia*, 18(6):1085–1097, 2016.

[45] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

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