MUG: A Parameterless No-Reference JPEG Quality 
Evaluator Robust to Block Size and Misalignment

Hossein Ziaei Nafchi, Atena Shahkolaei, Rachid Hedjam, and Mohamed Cheriet, Senior Member, IEEE

Abstract—In this letter, a very simple no-reference image quality assessment (NR-IQA) model for JPEG compressed images is proposed. The proposed metric called median of unique gradients (MUG) is based on the very simple facts of unique gradient magnitudes of JPEG compressed images. MUG is a parameterless metric and does not need training. Unlike other NR-IQAs, MUG is independent to block size and cropping. A more stable index called MUG+ is also introduced. The experimental results on six benchmark datasets of natural images and a benchmark dataset of synthetic images show that MUG is comparable to the state-of-the-art indices in the literature. In addition, its performance remains unchanged for the case of the cropped images in which block boundaries are not known. The MATLAB source code of the proposed metrics is available at https://dl.dropboxusercontent.com/u/74505502/MUG.m and https://dl.dropboxusercontent.com/u/74505502/MUGplus.m.

Index Terms—Blockiness artifact, JPEG compression, JPEG quality assessment, MUG, No-reference quality assessment.

I. INTRODUCTION

JPEG lossy compression is one of the most common coding techniques to store images. It uses a block-based coding scheme in the frequency domain, e.g., discrete cosine transform (DCT), for compression. Since $B \times B$ blocks are coded independent of each other, blocking artifacts are visible in JPEG compressed images specially under low bit rate compression. Several no-reference image quality assessment (NR-IQAs) models have been proposed to objectively assess the quality of the JPEG compressed images [1]–[17]. NR-IQAs do not need any information of the reference image. NR-IQAs are of high interest because in most present and emerging practical real-world applications, the reference signals are not available [18]. In the following, we will have an overview on NR-IQAs for JPEG compressed images.

In [1] for each block, horizontal and vertical difference at block boundaries are used to measure horizontal and vertical blockiness, respectively. Tan and Ghanbari [2] proposed a blockiness metric via analysis of harmonics. They used both the amplitude and the phase information of harmonics to compute a quality score. Harmonic analysis was also used to model another blockiness metric in [3].

Wang et al. [4] modeled the blocky image as a nonblocky image interfered with a pure blocky signal. Energy of the blocky signal is then used to calculate a quality score. In the DCT domain, a metric was proposed in [5] that models the blocking artifacts by a two-dimensional step function. The quality score is calculated following the human vision measurement of block impairments. The metric proposed in [9] measures blockiness artifact in both the pixel and the DCT domains. In [14], zero values DCT coefficients within each block are counted and a relevance map is estimated that distinguishes between naturally uniform blocks and compressed uniform blocks. For this end, an analysis in both DFT and DCT domains is conducted.

Wang et al. [6] proposed an efficient metric that measures blockiness via horizontally and vertically computed features. These features are average differences across block boundaries, average absolute difference between in-block image samples, and zero crossing rate. Using a set of subjective scores, five parameters of this model are estimated via nonlinear regression analysis. In [7], the edge orientation changes of blocks were used to measure severity of blockiness artifacts. Perra et al. [8] analyzed the horizontal, vertical and intrablock sets of $8 \times 8$ blocks after applying the Sobel operator to the JPEG compressed images.

The difference of block boundaries plus luminance adaptation and texture masking were used in [10] to form a noticeable blockiness map. From which, the quality score is calculated by a Minkowski summation pooling. In [11], one-dimensional signal profile of gradient image is used to extract block sizes and then periodic peaks in DCT domain are analyzed to calculate a quality score. Chen et al. [12] proposed a very similar metric.

In [15], three features including the corners, block boundaries (horizontal, vertical, and intrablock), and color changes, together with the subjective scores are used to train a support vector regression model. Li et al. [16] measured the blocking artifacts through weighting a set of blockiness scores calculated by Tchebichef moments of different orders.

Lee and Park [13] proposed a blockiness metric that first identifies candidates of having blockiness artifacts. The degree of blockiness of these candidates is then used to compute a quality score. Recently, a blockiness metric is proposed that performs in three steps [17]. Block grids are extracted in the spatial domain and their strength and regularity is measured. Afterwards, a masking function is used that gives different weights to the smooth and textured regions.

The aforementioned indices have at least one of the following drawbacks. They might not be robust to block size and block misalignment (examples are [6]–[8], [10], [14], [16]). They are...
complex (examples are [5], [11], [14]–[17]), or have many parameters to set ([6], [11], [14]–[17]). Indices such as NJQA [14] and GridSAR [17] are too much slow. Some indices need training ([6], [15]). Also, the range of quality scores provided by some of the indices such as [6] is not well defined, or they show other numerical issues [17].

In this letter, we propose a quality assessment model for JPEG compressed images that overcomes all aforementioned drawbacks. The proposed index is very simple and efficient, it is parameterless, and robust to block size and misalignment. The proposed metric called MUG is based on two simple facts about blockiness artifact. As a result of more JPEG compression, the number of unique gradient magnitude values decreases, and the median value of unique gradient magnitude values increases. The proposed blockiness metric MUG uses these two simple facts to provide accurate quality predictions for JPEG compressed images. Unlike other metrics that presume position of blocks beforehand or localize the position of blocks, MUG is not a local model and hence does not need any information on the position of blocks.

II. PROPOSED METRIC (MUG)

The proposed index called MUG predicts the quality of JPEG compressed images as follows. Given the JPEG distorted image \( D \), the Scharr gradient operator is used to approximate horizontal \( G_x \) and vertical \( G_y \) gradients of \( D \): \( G_x = h_x \ast D \) and \( G_y = h_y \ast D \), where \( h_x \) and \( h_y \) are horizontal and vertical gradient operators, and \( \ast \) denotes the convolution. From which, the gradient magnitude is computed as \( G(x) = \sqrt{G_x^2(x) + G_y^2(x)} \).

It is worth to mention that within the context of the proposed metrics, the Scharr operator performs better than the Sobel and Prewitt operators. The proposed metric works directly on the gradient magnitude instead of directional gradients. Let us denote \( uG \) as the unique numerical values of \( G(x) \). We show in the following that two properties of \( uG \) can be used to predict quality of JPEG compressed images: 1) number of values in \( uG \), and 2) median of \( uG \) values.

A. Number of Unique Gradients (NUG)

The number of unique gradients (NUG), e.g., the number of values in vector \( uG \), indicates how many distinct edge strengths exist in JPEG compressed image \( D \). It is very likely that a JPEG compressed image with blocking artifacts has smaller values of NUG than its uncompressed version. To verify this statement, JPEG compressed images of TID2013 dataset were chosen. For each of the 25 distortion-free images in TID2013, there are five JPEG compressed images of different distortion levels. The values of NUG for each of the 25 sets are found inversely proportional to the amount of distortion:

\[
\text{Compression rate } \propto \frac{1}{\text{NUG}}. \tag{1}
\]

In other words, the Spearman rank-order correlation coefficient (SRCC) between NUG values and mean opinion score (MOS) values is equal to 1 for each of the 25 sets. This experiment shows that the aforementioned statement holds true. Fig. 1 shows a scatter plot of NUG scores against the subjective MOS on the LIVE dataset [19] (see Section III to see how this plot is drawn). This plot shows that there is a noticeable correlation between NUG scores and MOS on this dataset. Unfortunately, NUG does not take into account the content of original images. An image may originally have less edge strengths variation than another. Therefore, there are cases that NUG cannot fairly judge images having different contents. This issue is solved through including median of unique gradients (MUG) into the proposed model.

B. Median of Unique Gradients (MUG)

As mentioned above, the image content is a factor that needs to be taken into account. Let us repeat the same experiment on JPEG compressed images of the TID2013 dataset, but this time for the MUG. The experiments show that the same statement holds true, e.g., the values of MUG for each of the 25 sets are proportional\(^1\) to the amount of distortion:

\[
\text{Compression rate } \propto \text{MUG}. \tag{2}
\]

In fact, MUG determines how strong is the middle value of unique gradients, which helps in taking into account the content of images. However, the values of MUG are not always reliable because the image quality is not only related to the edge strengths. The distribution of the unique gradients \( uG \) is another factor that cannot be considered by direct median value. Therefore, a simple standard deviation normalization was applied on the \( uG \) values before median value being computed:

\[
uG' = \frac{uG}{\sqrt{\sigma(uG)}}. \tag{3}
\]

Unique gradients vector \( uG \) has different behavior for images having naturally uniform regions and block uniform regions. For images with mostly naturally uniform regions, the standard deviation in general decreases by more compression. In contrast, the standard deviation value in general increases by more

\(^1\)Except for one case where SRCC is equal to 0.6, not 1.
compression for images having less naturally uniform regions. Therefore, median of $uG'$ takes into account the content of images. The effect of standard deviation normalization is visually shown in the scatter plots of Fig. 2.

The proposed quality assessment model for JPEG compressed images (MUG) can be written by combining relations (1) and (2):

$$MUG = \frac{MUG}{\text{NUG}}$$

(4)

where, $MUG$ (in italic) is the median value of $uG'$. It can be seen that the proposed metric is parameterless. To the best of our knowledge, MUG is the only parameterless metric in the literature. MUG is therefore completely independent to the misalignment. This advantage is shown in the experimental results. Since the proposed metric is parameterless, it should be invariant to the block size as well. However, no dataset is available to experimentally verify this statement. It is worth to mention that when the input image is in color, MUG converts it to a luminance channel: $L = 0.06R + 0.63G + 0.27B$. According to [20], this conversion may be imperfect, but it is likely to offer accurate estimates of differential measurements. Therefore, image gradient computation from $L$ should yield more accurate results. Since MUG only uses the median value of unique gradients, it might not be very accurate for some cases. In the following, MUG is modified by adding a few more unique gradient values.

C. Stable MUG ($MUG^+$)

The distributions of unique gradient values can be very different for images having diverse edge information. This distribution might be skewed (usually right-skewed), bimodal, etc. The median value can suddenly be shifted to the left or right if some adjacent values of median are not considered. Therefore, the MUG index can become more stable by adding a few more values to the median. These values must be smaller than the median value because larger values than median have much more variations and might be unreliable. Suppose that $uG'$ values in vector $uG'$ are sorted from smallest to largest. In this case, MUG/2 is the index of the median value in $uG'$. One easy way to add a few extra values that mentioned above is to use corresponding values of these indices: $\text{NUG}/i, i \in \{2, 3, \ldots, M + 1\}$, where $i = 2$ is the index of median and $M$ is the total number of values used ($M = 19$ in this letter). In fact by adding these extra values, the proposed metric becomes numerically more stable. Moreover, there are cases that there are not $M$ unique values in vector $uG'$. This property often happens when the majority of the input image or the whole image is naturally uniform or textured. Suppose that there are $1 \leq N \leq M$ of such values available. The stable MUG (called $MUG^+$) takes into account this behavior by the following formulation:

$$MUG^+ = \frac{MUG}{M - N + 1}$$

(5)

where MUG = MUG for $N = M$.

Apart from the block misalignment problem, several JPEG quality assessment models such as [6] and [14] provide quite wrong predictions in special cases that image has large amount of naturally uniform regions and/or it is textured. Fig. 3 shows a high-quality image of chessboard. This image has a very bad quality according to [6] ($Q = -245.89$). NJQA [14] likewise assessed this image as being of bad quality ($Q = 0.3414$). GridSAR [17] was not able to provide a numerical value. MUG is equal to 0.8060 (very bad quality), which also provides wrong assessment. In contrast, $MUG^+ = 0.0448$, which truly means that the chessboard image has a very good quality. This is another advantage of the proposed index $MUG^+$. Note that the datasets used in this letter do not have any image sample with this behavior.

### III. Experimental Results

In the experiments, six standard datasets of natural images and a benchmark dataset of synthetic images are used. The TID2013 [21] dataset contains 125 JPEG compressed images in total. The CSIQ dataset [22] has 150, the LIVE dataset [19] has 175, the VCL dataset [23] has 138, and the MICT dataset [24] has 84 JPEG compressed images. The ESPL dataset [25] is a synthetic dataset that contains 100 JPEG compressed images. The TID2008 dataset [26] is another dataset with 100 JPEG compressed images, which is in fact a subset of TID2013.

For objective evaluation, two evaluation metrics were used in the experiments: the SRCC, and the Pearson linear correlation coefficient (PLCC). The SRCC and PLCC metrics measure prediction monotonicity and prediction linearity, respectively.

To get a visual observation, the scatter plots of the proposed NR-IQA models MUG and $MUG^+$ on the LIVE dataset are shown in Fig. 4. The logistic function suggested in [19] was used to fit a curve on each plot:

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

(6)
where $\beta_1, \beta_2, \beta_3, \beta_4$, and $\beta_5$ are fitting parameters computed by minimizing the mean square error between quality predictions $x$ and subjective scores MOS. 

SSIM [27] as an FR-IQA, as well as five NR-IQAs including [6], NJQA [14], GridSAR [17], and the proposed indices MUG and MUG$^+$ were used in the experiments. [6] was chosen because it shows outstanding performance, and NJQA because it follows a different approach with promising performance. GridSAR is recently introduced blockness metric, which is also able to the handle block misalignment. Table I provides a performance comparison between the six aforementioned FR/NR-IQAs in terms of SRCC and PLCC. The same experiment is repeated on JPEG compressed images with misaligned blocks. JPEG compressed images with misaligned blocks are generated by cutting one pixel from the borders of the images. Since only one pixel width is cropped from image borders, the MOS values should remain unchanged. When block positions are known beforehand, the NR-IQA of [6] shows the best overall performance for the seven datasets. The proposed indices show consistent prediction accuracy over different datasets and comparable to the GridSAR and SSIM. The proposed indices in general outperform NJQA [14]. When block positions are not known, it can be seen from Table II that the proposed indices, e.g., MUG and MUG$^+$, and GridSAR show almost the same prediction accuracy as in Table I. This means that they are robust to the block misalignment. In contrast, [6] provides predictions with low accuracy.

While GridSAR performs better than MUG$^+$ on more considered datasets, it should be noted that GridSAR is a complex metric with several parameters to set. It is also computationally inefficient and numerically unstable.

A. Complexity

To show the efficiency of the proposed indices, a run-time comparison between six IQAs is performed and shown in Table III. The experiments were performed on a Corei7 3.40 GHz CPU with 16 GB of RAM. The IQA model was implemented in MATLAB 2013b running on Windows 7. It can be seen that MUG and MUG$^+$ have satisfactory run-times. Compared to the competing metric GridSAR, the proposed metric is about 250 times faster.

IV. CONCLUSION

In this letter, two novel IQA models for JPEG compressed images were proposed. The proposed indices are very simple and do not need training. They are based on the two simple facts of gradient magnitude of JPEG compressed images. As a result of more JPEG compression, the number of unique gradient magnitude values decreases and the median value of unique gradient magnitude values increases. The extensive experimental results show that the proposed indices are robust to block misalignment and have consistent performance on seven benchmark datasets.

| Index Type | SSIM [6] | NJQA [14] | MUG | MUG$^+$ |
|------------|----------|-----------|-----|--------|
| TID 2008   | 0.9540   | 0.9518    | 0.9442 | 0.9511  | 0.9408  | 0.9529 |
| TID 2013   | 0.9544   | 0.9530    | 0.9477 | 0.9545  | 0.9419  | 0.9546 |
| CSIQ       | 0.9200   | 0.9267    | 0.8860 | 0.9309  | 0.9077  | 0.9185 |
| LIVE       | 0.9786   | 0.9751    | 0.9539 | 0.9788  | 0.9674  | 0.9717 |
| VCL        | 0.9764   | 0.9735    | 0.9562 | 0.9756  | 0.9649  | 0.9730 |
| MICT       | 0.8664   | 0.8876    | 0.8746 | 0.9304  | 0.9304  | 0.9372 |
| ESPL       | 0.9314   | 0.9412    | 0.9154 | 0.9331  | 0.9285  | 0.9265 |

B. Comparison of the IQA Models on JPEG Compression Distortion Type of Seven Datasets in Terms of SRCC and PLCC

| Index Type | TID 2008 | TID 2013 | CSIQ | LIVE | VCL | MICT | ESPL |
|------------|----------|----------|------|------|-----|------|------|
| Index      | SRCC     | 0.8989   | 0.8909 | 0.8996 | 0.8722 | 0.8043 | 0.8017 |
| Time (ms)  | 140.21   | 225.52   | 187.85 | 222.06 | 79.983 | 79.983 | 79.983 |

Table II

| Index      | TID 2008 | TID 2013 | CSIQ | LIVE | VCL | MICT | ESPL |
|------------|----------|----------|------|------|-----|------|------|
| Index      | SRCC     | 0.8989   | 0.8909 | 0.8996 | 0.8722 | 0.8043 | 0.8017 |
| Time (ms)  | 140.21   | 225.52   | 187.85 | 222.06 | 79.983 | 79.983 | 79.983 |

Table III

| Index Type | JPEGind | SSIM [27] | MUG | MUG$^+$ |
|------------|---------|-----------|-----|--------|
| Time (ms)  | 140.21  | 187.85    | 222.06 | 79.983 |

Fig. 4. Scatter plots of MUG and MUG$^+$ scores against the subjective MOS on the LIVE dataset. Left: MUG (PLCC = 0.9649), and right: MUG$^+$ (PLCC = 0.9730).
REFERENCES

[1] H. Wu and M. Yuen, “A generalized block-edge impairment metric for video coding,” IEEE Signal Process. Lett., vol. 4, no. 11, pp. 317–320, Nov. 1997.

[2] K. Tan and M. Ghanbari, “Frequency domain measurement of blockiness in MPEG-2 coded video,” in Proc. Int. Conf. Image Process., vol. 3, 2000, pp. 977–980.

[3] K. Tan and M. Ghanbari, “Blockiness detection for MPEG2-coded video,” IEEE Signal Process. Lett., vol. 7, no. 8, pp. 213–215, Aug. 2000.

[4] Z. Wang, A. Bovik, and B. Evan, “Blind measurement of blocking artifacts in images,” in Proc. Int. Conf. Image Process., vol. 3, 2000, pp. 981–984.

[5] A. Bovik and S. Liu, “DCT-domain blind measurement of blocking artifacts in DCT-coded images,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., vol. 3, 2001, pp. 1725–1728.

[6] Z. Wang, H. R. Sheikh, and A. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in Proc. Int. Conf. Image Process., vol. 1, 2002, pp. 1–477–1–480.

[7] F. Pan, X. Lin, S. Rahardja, E. Ong, and W. Lin, “Measuring blocking artifacts using edge direction information [image and video coding],” in Proc. IEEE Int. Conf. Multimedia Expo., vol. 2, Jun. 2004, pp. 1491–1494.

[8] C. Perrera, F. Massidda, and D. Giusto, “Image blockiness evaluation based on Sobel operator,” in Proc. IEEE Int. Conf. Image Process., vol. 1, Sep. 2005, pp. 1–389–1–392.

[9] C.-S. Park, J.-H. Kim, and S.-J. Ko, “Fast blind measurement of blocking artifacts in both pixel and DCT domains,” J. Math. Imag. Vis., vol. 28, no. 3, pp. 279–284, 2007.

[10] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, “No-reference noticeable blockiness estimation in images,” Signal Process.: Image Commun., vol. 23, no. 6, pp. 417–432, 2008.

[11] H. Liu and I. Heynderickx, “A perceptually relevant no-reference blockiness metric based on local image characteristics,” EURASIP J. Adv. Signal Process., vol. 2009, pp. 2:1–2:14, Jan. 2009.

[12] C. Chen and J. Bloom, “A blind reference-free blockiness measure,” in Advances in Multimedia Information Processing (Lecture Notes in Computer Science 6297), Berlin, Germany: Springer, 2010, pp. 112–123.

[13] S. Lee and S. J. Park, “A new image quality assessment method to detect and measure strength of blocking artifacts,” Signal Process.: Image Commun., vol. 27, no. 1, pp. 31–38, 2012.

[14] S. Golestaneh and D. Chandler, “No-Reference Quality Assessment of JPEG images via a quality relevance map,” IEEE Signal Process. Lett., vol. 21, no. 2, pp. 155–158, Feb. 2014.

[15] L. Li, W. Lin, and H. Zhu, “Learning structural regularity for evaluating blocking artifacts in JPEG images,” IEEE Signal Process. Lett., vol. 21, no. 8, pp. 918–922, Aug. 2014.

[16] L. Li, H. Zhu, G. Yang, and J. Qian, “Referenceless measure of blocking artifacts by tchebichef kernel analysis,” IEEE Signal Process. Letters, vol. 21, no. 1, pp. 122–125, Jan. 2014.

[17] L. Li, Y. Zhou, J. Wu, W. Lin, and H. Li, “GridSAR: Grid strength and regularity for robust evaluation of blocking artifacts in JPEG images,” J. Visual Commun. Image Representation, vol. 30, pp. 153–163, 2015.

[18] Z. Wang and A. Bovik, “Reduced- and No-Reference image quality assessment,” IEEE Signal Process. Mag., vol. 28, no. 6, pp. 29–40, Nov. 2011.

[19] H. Sheikh, M. Sabir, and A. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Trans. Image Process., vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[20] J.-M. Geusebroek, R. Van den Boomgaard, A. Smeulders, and H. Geerts, “Color invariance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 12, pp. 1338–1350, Dec. 2001.

[21] N. Ponomarenko et al., “Color image database TID2013: Peculiarities and preliminary results,” in Proc. 2013 4th Eur. Workshop Visual Inform. Process., Jun. 2013, pp. 106–111.

[22] E. C. Larson and D. M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” J. Electron. Imag., vol. 19, no. 1, 2010, Art. no. 011006.

[23] A. Zaric et al., “VCL@FER image quality assessment database,” AUTOMATIKA, vol. 53, pp. 344–354, 2012.

[24] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, MICT Image Quality Evaluation Database 2000. [Online]. Available: http://mitc.eng.u-toyama.ac.jp/mict/index2.html

[25] D. Kundu and B. L. Evans, “Full-reference visual quality assessment for synthetic images: A subjective study,” in Proc. IEEE Int. Conf. Image Process., 2015, pp. 2374–2378.

[26] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, “TID2008—A database for evaluation of Full-Reference visual quality assessment metrics,” Adv. Modern Radioelectronics, vol. 10, pp. 30–45, 2009.

[27] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.