A new methodology in constructing no-reference focus quality assessment metrics

Song Jie (✉ so0001ie@gmail.com)  Advanced Remanufacturing Technology Centre  https://orcid.org/0000-0002-2243-3902
Liu Mengjun  Advanced Remanufacturing Technology Centre

Research Article

Keywords: Autofocus, Microscopy image analysis, No-reference image quality assessment

Posted Date: November 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2226467/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
A new methodology in constructing no-reference focus quality assessment metrics

Song Jie*, Liu Mengjun

1Intelligent Product Verification, Advanced Remanufacturing Technology Centre, 3 Cleantech Loop, CleanTech Two, 637143, Singapore

*Corresponding Authors: song_jie@artc.a-star.edu.sg ; so0001ie@gmail.com (J. Song)

Abstract

This paper proposed a new methodology which converts a full-reference focus quality assessment metric into a no-reference one. The methodology consists of three hypotheses which describe the relationship in focus quality between the original image and its variants. Using the proposed methodology, two no-reference metrics were constructed. The first used Brenner Gradient and the second used a full-reference metric proposed by ourselves. Evaluation was conducted on a public dataset and our own proposed dataset. Comparing with other no-reference metrics, our second one exhibited best performance on both datasets, with calculation time comparable to some fastest metrics considered.

Keywords: Autofocus, Microscopy image analysis, No-reference image quality assessment

1. Introduction

Autofocus (AF) is essential for achieving automated microscopy image analysis without visual inspection. However, focusing under microscope can be challenging given the small depth of field of the lens. Through decades, various AF methods have been proposed, including gradient-based approach [1], patching and interpolation [2], wavelet-based approach [3], depth from
focus/defocus based algorithm [4], learning based approach [5], focus measure curve estimation [6], among many others, and many comparison studies have been conducted on them [7-11]. Focus measurement is also the foundation in focus-based applications. For example, in shape from focus, a focus metric is applied locally on a small window surrounding each pixel. Studies by Malik and Choi [12] and Pertuz et al. [13] compared various focus metrics on their performance for shape from focus under various imaging conditions and different window sizes. Depth from focus/defocus is another application of focus metrics [14-16], which estimated the focus level of each pixel and associated it with the scene depth.

Most focus assessment metrics need either a score threshold to distinguish in-focus from out-of-focus cases or referring prior / subsequent images, through full-reference (FR) or reduced-reference (RR), in order to determine the maximum score. This has several drawbacks. Firstly, the “ideal” threshold for a given batch of images (from which the in-focus images are selected) may be different if the imaging conditions, e.g. illumination, vary from batch to batch, making it difficult or even impossible to choose a global optimal threshold. Secondly, the local maxima in some focus metrics may result in false identification of the optimal focus. To prevent this, exhaustive search techniques need to be performed in order to determine the global maximum. Thirdly, given an image batch not containing any focused image, the focus metrics will still produce a score for each image, in which case the global maximum is misleading as it does not guarantee focusing. On the other hand, for some metrics, e.g., the Wavelet_2 in Ref. [17], even the global maximum itself may not correspond to the in-focus position. It is therefore desirable to have focus metrics working on isolated images without inter-referencing and being free from dataset-
specific threshold. Such metrics are known as no-reference (NR) focus quality assessment (FQA) metric.

NR FQA belongs to a more general category called NR image quality assessment (IQA). NR IQA is of great interest in applications of image acquisition, transiting, compression and recording, where images inevitably suffer from some types of damage (e.g. white noise or Gaussian blur) and need to be recognized and removed. Typical applications of NR IQA include noise estimation [18], blur/sharpness assessment [19, 20], distortion evaluation [21], restoration [22], etc. As in-focus and out-of-focus images have different blur level, NR IQA metrics for defocus identification, i.e., NR FQA, have been proposed. Most NR FQA metrics are either sophisticatedly hand-crafted utilizing spatial and spectral characteristics of the images [23-26], or purely learning-based through AI approaches [27, 28].

In this paper, we aim to develop a new methodology in constructing NR FQA metric. The methodology consists of a set of hypotheses and can easily convert a FR metric into a NR one. However, to ensure the methodology works properly, the FR metric should possess certain characteristic. One suitable FR metric is Brenner gradient (BG) [1]. According to Refs. [17, 29], BG exhibits a sharp, distinct peak at the focus range, and decays rapidly away from focus. Such characteristic makes BG highly sensitive to the change in an image’s sharpness level, when the image goes from in-focus to out-of-focus, while blunt to sharpness change from one location to another within the out-of-focus range. To demonstrate the effectiveness of the methodology, we proposed two NR focus metrics based on BG and another FR metric we designed in this study, respectively. Their performances were evaluated and compared with state-of-art NR focus metrics
on our own datasets as well as a public dataset, FocusPath [30]. In summary, this paper makes the following contributions in focus assessment:

- A new methodology in constructing NR FQA metric is proposed
- A new FR metric, called sum of gradient (SoG), is proposed
- Two NR FQA metrics, NRBG and NRSoG, constructed based on the methodology, are proposed
- A new dataset, SS316_ShotPeen\(^1\), is proposed

The remainder of this paper is as follows: Section 2 reviews previous work related to NR FQA. Section 3 describes the methodology and its hypotheses as well as the proposed NR metrics. Comparisons between our metrics and other state-of-art methods on our own dataset and FocusPath are presented in Section 4. Section 5 concludes the paper.

2. Previous work

There have been many studies on NR FQA in past decades. For example, Marais and Steyn [23] proposed a metric differentiating between in-focus and out-of-focus blur using a variation of the spectral subtraction method. Wu et al. [24] proposed a NR method for defocus blur measurement in which Sobel operator is used for edge detection and Radon transform is applied to locate line features. Bahrami and Kot [31] defined the maximum local variation (MLV) of each pixel and used the standard deviation of the weighted MLV distribution as a metric to measure sharpness. In another study by Hassen et al. [26], the sharpness metric was developed based on local phase coherence (LPC) near distinctive image features evaluated in the complex wavelet transform domain. Recently, Hosseini and Plataniotis [32] proposed MaxPol convolution kernels. The
kernels are close approximation to a visual sensitivity model, and with the model, they built their NR image sharpness metric. In a further study, Hosseini et al. [30] proposed a novel NR sharpness metric using MaxPol kernels and based on the human vision system response. Their test results showed that the proposed metric significantly outperformed other metrics over both synthetic and natural blur databases. In Ref. [33], Hosseini et al. further tailored their method in Ref. [30] for digital pathology images, resulting an NR FQA metric called FQPath, which showed the best overall performance among eleven state-of-art methods considered in their study.

With the advances in AI technologies, in particular deep learning, researches on learning-based focus metric have been conducted. For example, Senaras et al. [27] proposed a novel deep learning framework, DeepFocus, to identify blurry regions in digital slides. In another study, Yang et al. [28] trained a deep learning model using synthetically defocused images generated from natural in-focus microscope images. The model can work in no-reference way by giving prediction on the level of focus for isolated images. Wang et al. [34] proposed a highly efficient CNN-based model FocusLiteNN which has 148 parameters for FQA. The model was trained specifically for assessing pathological images. For AI approach, although no user-specified parameters are needed, model training can be time-consuming and the performance of model depends on the quality of dataset, which may not always be good or enough. Moreover, the transferability of AI model may also be a problem, especially when the images encountered in real applications are significantly different from those in the training set.

3. Methodology and metrics

3.1 Methodology and hypotheses
The key idea of the proposed methodology which converts a FR metric to a NR one is the sharpness comparison between the original image ($f$) and its three variants, namely, blurred image ($g$), downsampled image ($h$), and blurred downsampled image ($j$), using the scores given by the FR metric. For a suitable FR metric, denoted as $M$, which gives higher score for more focused image, three hypotheses are made as below:

a) For an image already in-focus, blurring shall reduce its sharpness. With $M$, it means $M(g) \leq M(f)$, if $f$ is focused.

b) For an image already in-focus, downsampling shall not reduce its sharpness. With $M$, it means $M(h) \geq M(f)$, if $f$ is focused.

c) For two in-focus images having different focus level, downsampling shall increase the relative sharpness, defined as the ratio of sharpnesses between downsampled and original images, more for the less focused image than for the more focused one. With $M$, it means $\frac{M(h_2)}{M(f_2)} > \frac{M(h_1)}{M(f_1)}$, if $f_1$ and $f_2$ are focused, and $M(f_1) > M(f_2)$. For future reference, we rewrite the condition as $\frac{M(f_1)}{M(h_1)} > \frac{M(f_2)}{M(h_2)}$.

Hypothesis c) can be interpreted that for an image already very sharp, further downsampling will only increase its sharpness slightly, so its relative sharpness increment is small, on the other hand, an image not very sharp will have more sharpness increment after downsampling, so its relative sharpness increment is large. A suitable FR metric means the score it produces satisfies the above-mentioned hypotheses. We further prescribe that the FE metric is a convex function of image sharpness, i.e. the metric score acceleratingly increases as the image becomes more and more sharp.

For the given $M$, the construction of the corresponding NR metric can be described as follows. First, we calculate the ratio of $M$ scores between $f$ and $h$ and between $g$ and $j$, respectively, i.e.
\[ r_{fh} = \frac{M(f)}{M(h)} \quad (1) \]

\[ r_{gj} = \frac{M(g)}{M(j)} \quad (2) \]

We further prescribe that

\[ r_{fh} < r_{gj} \quad (3) \]

Finally, it is required that

\[ r_{gj} < 1 \quad (4) \]

If \( f \) is focused, from (3&4) we can derive that \( r_{fh} < 1 \), thus Hypothesis b) is satisfied.

The conditions given by (3&4) can be qualitatively explained with Fig. 1, which shows \( M \) as a convex function of image sharpness. In microscope imaging, a FR metric, e.g. BG, is a function of z-stack position. Here the general term “sharpness” is used in order to compare the focus level between the original image and its variants, as there is no corresponding z-stack location associated with them. The score of \( f \) and \( h \) is marked by green lines. The sharpness of \( g \) and \( j \) is smaller than that of \( f \) and \( h \) respectively according to Hypothesis a). For easy explanation, we further assume that the effect of blurring is infinitesimal, therefore \( g \) and \( j \) are to the left of \( f \) and \( h \) with a small shift \( \Delta \) in the axis of sharpness and their \( M \) scores are marked with red lines. Expand \( M(f) \) and \( M(h) \) at \( g \) and \( j \) up to first order in Taylor series, and if the condition of (3) is satisfied, we have,

\[
\frac{M(f)}{M(h)} \approx \frac{M(g) + t_1 \Delta}{M(j) + t_2 \Delta} < \frac{M(g)}{M(j)}
\]

in which \( t_1 \) and \( t_2 \) are the derivative of \( M \) with respective to sharpness, or more explicitly, the slope of the two blue lines in Fig. 1. From (4&5) we can get,

\[
\frac{t_1}{t_2} < \frac{M(g)}{M(j)} = r_{gj} < 1
\]
Hence, if $(3\&4)$ are satisfied, $t_1 < t_2$, which means image $f$ is in the region of the sharpness axis where the score of $M$ is acceleratingly increasing. For a FR metric like BG, which only exhibits rapid increase within a small neighborhood of focus location, the acceleratingly increasing region where $f$ is ensures that it is or very near to the focus location.

According to Hypothesis c), $r_{fh}$ is bigger for a more focused image, providing the image satisfies $(3\&4)$. Hence, it can be used to measure focus level. The whole methodology is schematized in Fig. 2. Blurring is conducted through a $3 \times 3$ convolution. The true or false conditions by $(3\&4)$ are softened by two sigmoid functions, each giving a value close to 1 or 0 if the condition is satisfied or unsatisfied. We use ratio of $r_{gj}$ to $r_{fh}$ and reciprocal of $r_{gj}$ to implement conditions $(3\&4)$ which showed better performance than simply taking subtraction. Two weight factors, $W_1$ and $W_2$, are applied to amplify the input strength. If both conditions are satisfied, the sum of their outputs will be close to 2, and the output of the ReLU (with bias -1) will be close to 1, otherwise the output of ReLU will be close to 0. The ReLU output will multiply $r_{fh}$ to produce the final NR score.

Fig. 1. The illustration of conditions given by $(3\&4)$
3.2 NR metrics

Two FR metrics are adopted in this paper to construct their corresponding NR metrics. One is BG, of which the score is calculated as

$$BG_1(f) = \frac{\text{reduce}_\text{sum}(K_{BG} \otimes f \circ K_{BG} \otimes f)}{H \times W}$$  \hspace{1cm} (7)$$

in which $H$ and $W$ are the height and width of the image, $K_{BG} = [1,0,-1]$, subscript 1 indicates the convolution is along width direction (dimension 1), and $\otimes$ and $\circ$ denote convolution and element-wise multiplication, respectively. The other FR metric, SoG, is proposed by this paper. Its score is given by the following equation
\[ \text{SoG}_1(f) = \frac{\text{reduce}_\text{sum}(K_{\text{SoG}} \otimes f + K_{\text{SoG}} \otimes f)}{H \times W} \]  

in which \( K_{\text{SoG}} = [1, 0, 0, -1, 1, 1, -1, 1, -1, 1, 1, -1, 1, 0, 0, -1] \). The metric’s name is based on \( K_{\text{SoG}} \) which calculates the sum of gradients with different interval length. It should be noted that both BG and SoG can do convolution along height direction (dimension 0) by transposing the kernels in (7&8). We use \( BG_0 \) and \( SoG_0 \) to denote such cases. If referring the general forms regardless of convolution direction, we use BG and SoG.

The performance of proposed metrics is analyzed based on a small validation dataset with three groups. Each group has 40 images (in png format) captured at different z-stack positions. Images in Groups 1&2 were taken at two locations of a metal sample. Location of Group 1 has larger variation in grey scale intensities than location of Group 2. Images from Group 3 are all black, i.e., just capturing the noise of the microscope. The motivation is to test the metrics at the case where it is completely featureless. The sample patches from images in each group are shown in Fig. 3.

![Fig. 3. Sample patches from different images in each group of the validation dataset](image-url)
**Fig. 4** shows the SoG and BG scores at different z-stack positions for all groups. For Groups 1&2, both scores increase as the z-stack moves closer to the focus location, however, from bottom to peak, the relative change of SoG is more than two times larger than that of BG, meaning that the former is more sensitive to focus change. For Group 3, as expected, both metrics have small values and low variation in their scores.

Using SoG and BG, we can design their corresponding NR metrics, named as NRSOG and NRBG, based on the methodology in Fig. 2. The following kernel, which is a finite discrete approximation of Gaussian,

\[
K = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}
\]

is adopted in the methodology’s 3 × 3 convolutions for image blurring. The reason of choosing Gaussian kernel is that it is commonly used for blurring [13, 35, 36]. Other kernels, e.g. uniform kernel, have also been tried and the difference is not significant. The weights, \( W_1 \) and \( W_2 \), both are equal to 10 in our design so as to enhance the sigmoid input by one order of magnitude.
Fig. 5 shows the results of NRSoG\(_1\) and NRBG\(_1\) (calculated with SoG\(_1\) and BG\(_1\)) as well as their \(r_{fh}\) and \(r_{gf}\) values for all the images in each group. In the interest of saving space, values obtained by SoG\(_0\) and BG\(_0\) are not presented as they have similar pattern to those presented in Fig. 5.

![Graphs showing NRSoG\(_1\) and NRBG\(_1\) scores and \(r_{fh}\) and \(r_{gf}\) values for all the images in Group 1, Group 2, and Group 3.](image)

**Fig. 5.** NRSoG\(_1\) and NRBG\(_1\) scores and \(r_{fh}\) and \(r_{gf}\) values for all the images in (a&d) Group 1, (b&e) Group 2, and (c&f) Group 3

For Groups 1&2, NRSoG\(_1\) gives non-zero scores on a wider range of z-stack than NRBG\(_1\). This is because \(r_{fh}\) of NRSoG\(_1\) is lower and changes more slowly than that of NRBG\(_1\), as can been seen, e.g., in Group 1. Since the conditions in (3&4) are softened by sigmoid functions, even if they are slightly violated, the scores of both metrics may still be larger than zero. This gives the metrics more flexibility to assess images with few features or low pixel intensity variation, such as those in Group 2. However, if \(r_{fh}\) or \(r_{gf}\) is significantly larger than \(r_{gf}\) or 1, the metric scores will be zero. This can be observed clearly in Group 3. Furthermore, for both metrics, the non-zero score
is higher for more focused image, which demonstrates the effectiveness of the proposed methodology.

It is worth mentioning that the sharpness measured by the metrics in this paper is the one felt by human’s visual sensory, which is also what the term “sharpness” in Fig. 1 means. For an in-focus image of blurry or featureless scene, for example, a purely black surface, its appearance may resemble those in Group 3, and the metrics will score zero on that image.

Finally, the definition of NRSoG and NRBG, which considers both height and width convolutions in the corresponding FR metrics, is given as follows

\[
NRSoG(f) = \max(NRSoG_0(f), NRSoG_1(f))
\]  \hspace{1cm} (9)

\[
NRBG(f) = \max(NRBG_0(f), NRBG_1(f))
\]  \hspace{1cm} (10)

The max operation is adopted as a way to ensure that for a sharp image with pixel intensity varying only in one direction, either height or width, the NR metrics will give non-zero score.

4. Experiments & discussion

We evaluated our proposed metrics on two blur image datasets, both in png format. One is FocusPath, which contains 864 pathological images with size of $1024 \times 1024$. The other, called SS316_ShotPeen, contains 2772 images (44 specimens, 3 positions, 21 slices) of stainless-steel 316 after shot peening treatment, with size of $2048 \times 2448$. This dataset was built by the authors in order to compare different metrics in assessing the focus level of shot peened surface, which is critical to the quality inspection on surface finishing. As an example, Fig. 6 shows the patches
from 21 slices of specimen 10 at position 1. Detailed description of SS316_ShotPeen can be found from its homepage.

Fig. 6. Patches from 21 slices of specimen 10 at position 1 in SS316_ShotPeen, with the most focused one bounded by red box

Seven performance metrics, Pearson linear correlation coefficient (PLCC), Kendall rank correlation coefficient (KRCC), Spearman rank order correlation (SRCC), root mean square error (RMSE), and three measurements of Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) proposed in [37], i.e., AUC$_{DS}$, AUC$_{BW}$ and C$_0$, were considered. The calculation time for each FQA metric was also reported. All the FQA metrics were implemented with Matlab without optimization and were run on a PC with an Intel i9-9940X CPU and 128 GB RAM. The performance of each FQA metric on the datasets is listed in Table 1, with the top three best scores highlighted in bold.

It is seen that all the scores of NRSoG, except SRCC in FocusPath and calculation time, are among the best for both datasets. Most scores of FocusLiteNN and HVS-MaxPol-2 (both were trained or fine-tuned based on FocusPath) are also among the best for FocusPath. But for SS316_ShotPeen, some scores, e.g. PLCC, KRCC and SRCC, of FocusLiteNN and HVS-MaxPol-2 are significantly lower than those of NRSoG. For MLV and ARISMC, each has some scores among the top three for one dataset, but has relatively poor performance for the other. Therefore, NRSoG is of higher
Table 1 Performance evaluation of different FQA metrics on FocusPath, SS316_ShotPeen

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Metric</th>
<th>PLCC</th>
<th>KRCC</th>
<th>SRCC</th>
<th>RMSE</th>
<th>AUC_DS</th>
<th>AUC_RW</th>
<th>Co</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FocusLiteNN [34]*</td>
<td>0.8516</td>
<td>0.7422</td>
<td>0.8960</td>
<td>1.3467</td>
<td>0.7400</td>
<td>0.9588</td>
<td>0.8934</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0.8027</td>
<td>0.7910</td>
<td>0.9248</td>
<td>1.9072</td>
<td>0.7567</td>
<td>0.9757</td>
<td>0.9153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FQPath [33]</td>
<td>0.8438</td>
<td>0.6821</td>
<td>0.8550</td>
<td>1.3818</td>
<td>0.7238</td>
<td>0.9428</td>
<td>0.8616</td>
<td>0.0752</td>
<td>0.3088</td>
</tr>
<tr>
<td></td>
<td>0.9175</td>
<td>0.7783</td>
<td>0.9175</td>
<td>1.2324</td>
<td>0.8210</td>
<td>0.9749</td>
<td>0.9084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HVS-MaxPol-1 [30]</td>
<td>0.7900</td>
<td>0.6392</td>
<td>0.8154</td>
<td>1.6006</td>
<td>0.6907</td>
<td>0.9238</td>
<td>0.8388</td>
<td>0.0539</td>
<td>0.2211</td>
</tr>
<tr>
<td></td>
<td>0.8066</td>
<td>0.6758</td>
<td>0.8364</td>
<td>1.8877</td>
<td>0.7220</td>
<td>0.9437</td>
<td>0.8548</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HVS-MaxPol-2 [30]</td>
<td>0.8467</td>
<td>0.6855</td>
<td>0.8579</td>
<td>1.3672</td>
<td>0.7281</td>
<td>0.9453</td>
<td>0.8634</td>
<td>0.1036</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8940</td>
<td>0.7739</td>
<td>0.9155</td>
<td>1.3994</td>
<td>0.8019</td>
<td>0.9743</td>
<td>0.9061</td>
<td>0.4185</td>
<td></td>
</tr>
<tr>
<td>RISE [25]</td>
<td>0.6177</td>
<td>0.4792</td>
<td>0.6489</td>
<td>2.1602</td>
<td>0.6136</td>
<td>0.8307</td>
<td>0.7540</td>
<td>4.5845</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3582</td>
<td>0.1778</td>
<td>0.2656</td>
<td>3.4414</td>
<td>0.7268</td>
<td>0.6605</td>
<td>0.5933</td>
<td>37.5379</td>
<td></td>
</tr>
<tr>
<td>FocusPath/SS316_ShotPeen</td>
<td>GPC [38]</td>
<td>0.5234</td>
<td>0.6021</td>
<td>0.7813</td>
<td>2.4121</td>
<td>0.5817</td>
<td>0.8803</td>
<td>0.8191</td>
<td>0.0870</td>
</tr>
<tr>
<td></td>
<td>0.5161</td>
<td>0.6182</td>
<td>0.7778</td>
<td>2.9883</td>
<td>0.6534</td>
<td>0.9321</td>
<td>0.8244</td>
<td>0.3900</td>
<td></td>
</tr>
<tr>
<td>ARISMc [20]</td>
<td>0.4360</td>
<td>0.3115</td>
<td>0.4355</td>
<td>2.6250</td>
<td>0.5406</td>
<td>0.7271</td>
<td>0.6651</td>
<td>44.7555</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9116</td>
<td>0.7778</td>
<td>0.9268</td>
<td>1.2773</td>
<td>0.8039</td>
<td>0.9716</td>
<td>0.9083</td>
<td>100.9768</td>
<td></td>
</tr>
<tr>
<td>MLV [31]</td>
<td>0.8032</td>
<td>0.6992</td>
<td>0.8638</td>
<td>1.5498</td>
<td>0.6968</td>
<td>0.9388</td>
<td>0.8705</td>
<td>0.1367</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7695</td>
<td>0.7300</td>
<td>0.8813</td>
<td>2.0625</td>
<td>0.7231</td>
<td>0.9584</td>
<td>0.8832</td>
<td>0.5637</td>
<td></td>
</tr>
<tr>
<td>LPC [26]</td>
<td>0.8120</td>
<td>0.6704</td>
<td>0.8359</td>
<td>1.5146</td>
<td>0.7070</td>
<td>0.9325</td>
<td>0.8553</td>
<td>1.2609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8682</td>
<td>0.7246</td>
<td>0.8813</td>
<td>1.5596</td>
<td>0.7575</td>
<td>0.9459</td>
<td>0.8803</td>
<td>5.8627</td>
<td></td>
</tr>
<tr>
<td>S3 [39]</td>
<td>0.7441</td>
<td>0.6206</td>
<td>0.7959</td>
<td>1.7666</td>
<td>0.6497</td>
<td>0.9151</td>
<td>0.8288</td>
<td>24.1172</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.6338</td>
<td>0.7153</td>
<td>0.8643</td>
<td>2.5996</td>
<td>0.6941</td>
<td>0.9582</td>
<td>0.8756</td>
<td>83.9867</td>
<td></td>
</tr>
<tr>
<td>NRBG</td>
<td>0.8149</td>
<td>0.6650</td>
<td>0.8384</td>
<td>1.5020</td>
<td>0.7037</td>
<td>0.9304</td>
<td>0.8525</td>
<td>0.0739</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8809</td>
<td>0.8921</td>
<td>0.9756</td>
<td>1.4834</td>
<td>0.8615</td>
<td>0.9959</td>
<td>0.9682</td>
<td>0.3054</td>
<td></td>
</tr>
<tr>
<td>NRSoG</td>
<td>0.8481</td>
<td>0.6685</td>
<td>0.8555</td>
<td>1.3623</td>
<td>0.7287</td>
<td>0.9439</td>
<td>0.8639</td>
<td>0.0861</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9385</td>
<td>0.8579</td>
<td>0.9624</td>
<td>1.0645</td>
<td>0.8951</td>
<td>0.9862</td>
<td>0.9504</td>
<td>0.3616</td>
<td></td>
</tr>
</tbody>
</table>

* The processing time of FocusLiteNN is not included as it is AI based method which utilizes GPU

generality than the rest FQA metrics. For calculation time, NRSoG is comparable to FQPath and NRBG, but slower than HVS-MaxPol-1. However, NRSoG surpasses HVS-MaxPol-1 on all the seven performance metrics. NRBG, which is a simpler implementation of the proposed methodology than NRSoG, also has good performance on both datasets, with some scores among the best in SS316_ShotPeen. Compared with some sophisticated hand-crafted metrics, including MLV, ARISMc and LPC, NRBG has better or comparable performance but with more generality and higher processing speed. It is worth mentioning that both NRSoG and NRBG can be boosted
in speed if implementing them on GPU, since both convolution and average pooling are GPU compatible.

The fact that some metrics show good performance on one dataset but moderate, if not poor, on the other indicates that the two datasets contain images of different focus quality characteristics, and those metrics captured characteristics of either but not both datasets. Therefore, the proposed dataset, SS316_ShotPeen, widens the scope for FQA metric validation. On the other hand, one of the most important virtues of a NR metric is the ability to give absolute score on image sharpness. The score should be universally meaningful and irrespective of the image contents. Therefore, a dataset with ground truth focus quality score for each image is very useful to test the no-referenceness. This is another motivation of proposing SS316_ShotPeen. As seen in Table 1, the RMSE against ground truth for NRSoS is the lowest on both datasets, suggesting NRSoS has the best overall no-referenceness.

NRBG and NRSoS are just two possible NR metrics that can be constructed using the methodology we proposed. By incorporating different FR metrics into the methodology in Fig. 2, new NR metrics can be generated. Therefore, developing NR metric can be reduced from sophisticatedly handcrafting the power spectrum/frequency analyzers to looking for suitable FR metrics, which can be much easier (for example, neither NRBG nor NRSoS includes any power spectrum/frequency analysis). The methodology is based on the three hypotheses (we choose not using the term “axioms” to avoid overstatement), hence this work is the first attempt on building an axiomatic system for NR FQA.
The proposed methodology goes beyond existing NR metrics in several aspects. First, the methodology does not require analysis on the frequency components, which is simpler in concept and faster in processing speed than those sophisticatedly designed NR metrics. Second, the methodology is based on hypotheses which can be quantitively tested and refined. NR metrics with power spectrum/frequency analysis utilize the fact that images with high-frequency contents are sharp. However, since there is no definitive boundary between high and low frequencies and the concept of high/low frequency is scene-dependent, it is difficult to set up quantitive rules on the frequency components. Therefore, all frequency-based metrics have to rely on empirical assumptions which estimate whether a frequency component is high or low in the given image. Third, compared to AI-based metrics, the methodology has explainable architecture and better generality.

5. Conclusion

A new methodology in constructing the NR focus quality metric is proposed. The methodology consists of three hypotheses which describe the relationship in focus quality between the original image and its blurred, downsampled, and blurred downsampled versions. With those hypotheses, two NR metrics, NRBG and NRSoG, were constructed, using BG and our proposed SoG, respectively, as kernels.

Validation was conducted on a small dataset consists of three groups of images, i.e., those with large pixel value variation, small pixel value variation, and black images. The behaviours of both BG and SoG were analysed. The results showed that SoG has a wider range for non-zero values and a higher peak than BG, hence more sensitive to focus quality change. The NRBG and NRSoG,
constructed from BG and SoG, respectively, showed desirable behaviour which yielded high value for more focused image and zero value for the black images.

The performance of NRBG and NRSOG was evaluated and compared with that of other NR metrics on two datasets, FocusPath and our proposed SS316_ShotPeen. The results showed that NRSOG has best performance for both datasets, showing much higher generality than the rest metrics, and its calculation time is comparable to some fastest metrics considered in the present study. NRBG, which is simpler than NRSOG, is one of the fastest metrics, and it has good performance on both datasets. It has also been shown that some metrics performed well on either one dataset but not both, indicating that SS316_ShotPeen has characteristics different from FocusPath. Therefore, the proposed dataset widens the scope for FQA metric validation.

Compared to state-of-art NR FQA metrics, the proposed methodology has advantages including simplicity in concept, hypotheses-based approaches that can be quantitively tested, explainable architecture and good generality. Most importantly, it is an alternative approach which can provide new insights and deepen the understanding of the NR FQA problem.

In the future, several studies will be conducted. The first will be a full analysis and performance evaluation on SoG. The main focus of the current study is to present the methodology and demonstrate its effectiveness. The SoG serves as a kernel for building the best NR metric, NRSOG. However, the SoG deserves a thorough study on its own. From validation study in this paper, it has been observed that SoG shows better performance than Brenner gradient. Therefore, it can be an effective FR FQA metric for various focus/defocus based applications, and a potential
replacement of Brenner gradient. The second study will be on the underlying mechanism of the proposed methodology. Since both the methodology and the frequency-based approaches address the same phenomenon, and the frequency analysis is a more fundamental approach, the former should be able to be explained by the latter. The third study will be on the validation of the hypotheses we proposed by using various metrics and datasets. Finally, we will refine the methodology and will look for new kernels for constructing no-reference metrics.

Acknowledgment

The authors would like to greatly thank Miss Angel Clarita Juspi from National University of Singapore, for data processing and plotting the figures. The authors also thank Dr. Zhang Jie, Dr. Wang Zhengbiao and Dr. Fang Hongchao for their valuable suggestions.

References


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- declarationofcompetinginterests.pdf