No-reference Perception Based Image Quality Evaluation Analysis using Approximate Entropy

Ana Gavrovska
Dept. of Telecommunications
University of Belgrade - School of Electrical Engineering
Belgrade, Serbia
anaga777@gmail.com,
anaga777@etf.bg.ac.rs

Andreja Samčović
Dept. of Telecommunications and Networks
University of Belgrade – Faculty of Transport and Traffic Engineering
Belgrade, Serbia
andrej@sf.bg.ac.rs

Dragi Dujković
Dept. for Human Resources and General Affairs
School of Electrical Engineering - University of Belgrade
Belgrade, Serbia
dragi@etf.bg.ac.rs

***Abstract*—Due to extensive relevance across many disciplines, interest of no-reference image quality evaluation has been increased. The main goal is to assess the visual quality of an image using an objective metric that should be highly consistent with the subjective scores given by viewers. Well-known naturalness and perception based metrics include patch level distortion estimation and may show specific effects when comparing to high difference mean opinion scores. In this paper such effects are demonstrated, as well the possibility of using approximate entropy to overcome such manifestations. The obtained results show that approximate entropy technique can be used as an estimator in order to additionally distinguish image information related to subjective index.**

Keywords—approximate entropy, perception, distortion types, subjective score, opinion-unaware, image quality

I. Introduction

Perceptually optimized delivery of visual services is gaining prominence. Quality metrics have significantly improved recently, especially those that are no-reference (NR), meaning those that do not rely on reference presenting pristine image version [1-3]. Generally, objective image quality assessment (IQA) methods may be divided into three groups based on the amount of information that is accessible about the reference image: full-reference (FR), reduced-reference (RR) and no-reference (NR) [2]. In order to estimate the quality of distorted images, FR methods require the whole information about the reference. Reducing the amount of information about the original leads to RR approach. NR methods are considered of particular interest in practical implementations since it is valuable to automatically and objectively forecast the perceived quality of images without having access to reference image which may not be available. According to the prior knowledge of distortion type, methods can be divided into two groups: distortion related and without any knowledge about existing distortions. Moreover, in practical implementations it is difficult to assume how severe distortions present in an image are and how to connect them to the opinion index.

Since reference content is not available in the actual implementations of visual services, efficient and affordable approaches for NR-IQA are needed. Moreover, opinion-unaware (OU) blind methods have a wide range of potential applications. Mostly naturalness or natural scene statistics based features are included in model development, where method's high performance depends on specific local perception based framework created on knowledge about human visual system (HVS) [3-4]. Naturalness Image Quality Evaluator (NIQE) uses this framework for high correlation results with opinion scores relying on local patch-wise variation estimation [4]. Similar approach can be found in [1], with Perception Image Quality Evaluator (PIQE). PIQE represents typical evaluator in many applications like medical images [5-6] and UAV (Unmanned Aerial Vehicle) imagery [7]. Extraction of specific quality-aware local descriptors is expected to improve NR-IQA. Namely, statistical regularities in an image should be investigated since HVS is sensitive to them [8]. OU blind methods may combine both local framework and global evaluations in order to improve results [9].

Recent results presented in [10] show that traditional NR measure such as Shannon entropy (SE) may have similar behavior in fitting of objective scores with perception IQA. Similar dependency is obtained with NIQE and PIQE values even though direct correlation between entropy and opinion index is not present. In the measurement of signal complexity various entropy algorithms may be valuable, especially the ones related to regularity measurements, such as approximate entropy (AppE) [11].

This paper considers applying AppE in NR-IQA evaluation such as PIQE based one for the first time to the authors' knowledge. So far, AppE has been used as a measure of complexity in many applications like: evaluation of nonlinear dynamics of bio-movements [12], contour segmentation [13] or physiological activity [14]. In this work opinion scores are compared to PIQE by means of AppE for improved OU image quality evaluation.

The paper consists of six sections. Section II is dedicated to NR-IQA metrics and local feature extraction, as well as comparison between perception based evaluation and opinion scores. PIQE is chosen as a typical OU method, and a combination between local and global perception is considered. Common distortion types available in a publicly available dataset are explained in Section III. Also, SE and AppE as complexity measures are briefly described in the same section, where the AppE is tested. In Section IV, simulation steps of the experimental analysis are explained. The obtained results are presented in Section V, where distortion-specific effects with PIQE are demonstrated and possible filtering by means of AppE is performed. Finally, conclusions are mentioned in Section VI.

II. Perception based Image Quality Evaluation

A. No-reference quality assessment and blind models

Quality assessment that do not require reference is not an easy task. Moreover, blind IQA models do not utilize human-rated distorted images. This means that OU blind methods show less correlations to subjective ratings compared to methods that apply them for training purposes. Since variations exists in both machine and human ratings, where human judgments are applied as reference-like data, these blind models may enable researchers to understand opinion scores more thoroughly as well.

Local artifacts affect the human perception score, where they can be described as visible modifications from a pristine image. This artifact-less representation is associated with naturalness existing in standard photographies. Popular NR-IQA methods exploit naturalness based estimations. The main idea behind the successful naturalness estimations is to gather a collection of quality-aware features with the application of a proper model like multivariate Gaussian model in [4]. Image is firstly treated in the pre-processing phase, where local mean removal and divisive normalization are performed in the spatial domain. The obtained mean subtracted and contrast normalized (MSCN) coefficients are used for further processing. Domain division into local patches or blocks introduces a suitable local framework for further evaluations for measures such as NIQE. For each patch, local activity can be calculated, and with proper selected threshold only active patches/blocks are taken into account for Gaussian model fit. Distance between naturalness feature model and extracted features are proposed as a NIQE index. The dimension of patches can be selected in order to extract the same number of features (typically thirty-six). The computed distances can then be compared to means opinion scores (MOS) or difference MOS (DMOS). In OU PIQE method used in practical implementations [5-6] similar calculations are made. The method extracts local features and performs estimation based on selected active patches. The activity is considered as amount of distortion D_k related to present variations within a patch. The PIQE index is found by:

$$PIQE = \left(\left(\sum_{k=1}^{N_{SA}} D_k\right) + C_1\right) / (N_{SA} + C_1) \qquad (1)$$

where N_{SA} is the number of spatially active patches in an image and C_1 is a positive constant to prevent instability of calculation (usually has value 1).

B. Comparison between opinion-unaware quality evaluation and subjective scores

The end users who utilize metrics implementations like PIQE take into account developed models which are considered as appropriate judges of quality. When DMOS value is equal to zero, the corresponding image is found as a pristine one. High DMOS values means that analyzed image has a poor image quality from the perception point of view. These estimations are helpful for new metrics development, where in comparison to the opinion ratings the use of metrics is thought to be a suitable fit [10].

The comparison between OU quality evaluation and subjective scores can be applied in different manner. Typically, the scatter diagrams are generated to show points set by coordinates representing two variables (e.g. PIQE and DMOS) for a set of data, and to display strength and the relationship between the variables [8]. Also typically, Pearson correlation coefficient is calculated meaning that the relationship is linear [8, 10]. Nevertheless, each obvious trend between the variables is valuable. Also, there is a possibility when PIQE metric is not able to deal with a fitting model accurately enough.

Uncertainties may be found not only in images that do not seem distorted and that are pleasing to eyes, but also in images where distortions are obvious. In the absence of more certain reference than judgment rating these uncertainties may become obvious. In order to overcome time-consuming evaluation and provide even better fitting of OU methods besides patch-wise feature extraction global methods are introduced. For example, in [9] local information of perception and global object detection are combined into a new quality measure model.

It is proved that saliency or segmentation information may be suitable to perform tuning of quality results obtained via local patch-wise approach. Such tuning and calibration can be seen in quality assessments like in CT (Computed Tomography) images in [15], where PIQE method is applied for extracting local characteristics. Local lesion region measure and global image quality value construct a linear pooling of the two methods into a single score. Global image quality measure in [15] represents a combination of common FR techniques applied on the whole images. Functional HVS aspects and saliency are considered in [16] for blind IQA, where HVS-aware features are implemented through Shannon entropy (SE). Thus, information and content expressed using SE is also found useful as a global approach. In most of such state-of-the art methods LIVE dataset [17] has been applied as appropriate data for fitting tasks.

III. Distortion types and Approximate Entropy

High correlation between subjective assessment score and selected metric should be evident. General success of NR-IQA metric is based on the sensitivity to present artifacts when utilizing an objective method. The artifacts are result of made adjustments or manipulations to an image. Depending on a distortion various effects using a method may occur, where diverse factors can cause them.

In order to keep the experimental setup simple, a publicly available dataset called LIVE is applied. It is gathered of synthetically made distortions, where impaired images are accompanied with Difference Mean Opinion Scores (DMOS values). There are five distortion types of various levels within the set labeled as: jp2k, jpeg (JPEG - Joint Photographic Experts Group), wn, gblur, and ff. This means that different JPEG2000 and JPEG compression related distortion levels are depicted in jp2k and jpeg category, respectively. The third category labeled as wn represents white noise artifacts, while the fourth distortion type is Gaussian blur or gblur. The final category is Rayleigh fast-fading channel distortion (ff). In this collection there are 29 unaltered reference images that represent sources for distortion generation.

In addition to many forms of distortions for which OU method should be efficient, there are different kinds of complexity measures. Traditional statistical NR metric for measuring complex signals represents Shannon entropy denoted here as SE. It is well known for its applications in signal nonlinear modeling and information theory. Moreover, it enhances the comprehension of data being analyzed. For higher SE values greater information value is found within the content and vice versa. For the entire set of probabilities

p_i available for values x_i of a random variable, SE is calculated as:

$$SE = \sum_{i=1}^{N} p_i * log(p_i). \quad (2)$$

Image entropy is proved as suitable NR-IQA statistical quantifier, but SE does not represent the only approach for complexity computations. Additional entropy calculations are still possible [11]. Here, the focus is on approximate entropy denoted as AppE. It represents a different measure compared to SE used as a metric of system and signal complexity. Namely, it measures the degree of regularity and predictability of data fluctuations applicable in different fields [12-14]. Common statistical parameters like mean or variance, as well as rank order statistics, are not suitable for signal characterization and separating more and less regular sequences. For example, if two sequences are consisted of zeros and ones each with the same probabilities, approximate entropy of random signal compared to the regular one is much higher. So, the measure is useful to estimate the randomness found in data without any previous knowledge, and to compare signals that are of similar origin by focusing on correlation, persistence and regularity. If low values are obtained for data or system it can be stated that consistent and predictive patterns are found in the behavior. Also, if higher values are calculated for this type of entropy, this is not the case. The approximate entropy can be found for a number of measurements as:

$$AppE(m,r) = \lim_{N\to\infty} [\phi^m(r) - \phi^{m+1}(r)] \quad (3)$$

for selected *m* and *r* values representing dimension and radius, respectively, and where

$$\phi^m(r) = \left(\sum_{i=1}^{N-m+1} log\left(C_i^m(r)\right)\right)/(N-m+1) \quad (4)$$

for each count *C* defined as a ratio of the number of distances and *N*-*m*+1, as in implementation described in [11]. The number of distances is based only on distances between pairs of vectors constructed from data that are equal or below *r*. In this paper AppE is found using dimension of two and radius depending on the signal covariance [12-14].

IV. Simulation

Experimental analysis in this paper is based on PIQE metric and distortion based approach. For each image SE, AppE and PIQE values are calculated. AppE is found for histogram of MSCN coefficients. For performance evaluation two common measures are used SROCC (Spearman Rank Order Correlation Coefficient) and PLCC (Pearson Linear Correlation Coefficient) [8, 10]. In the first phase, effects for different DMOS values analyzed for PIQE quality assessment per distortion. In the second stage, AppE values are compared per image to observe SE and PIQE behavior compared to DMOS. The main goal is data filtering for random effects removal that may affect correlation results. This is done by AppE. The rest of data (and distortions) are analyzed separately. The applied framework is illustrated in Fig.1.

V. Experimental Results

Natural statistics in patch based local examination gives fine fitting and correlation results. Nevertheless, in the examinations it is not a rare case to observe saturation-like effects for high DMOS values meaning poor quality.

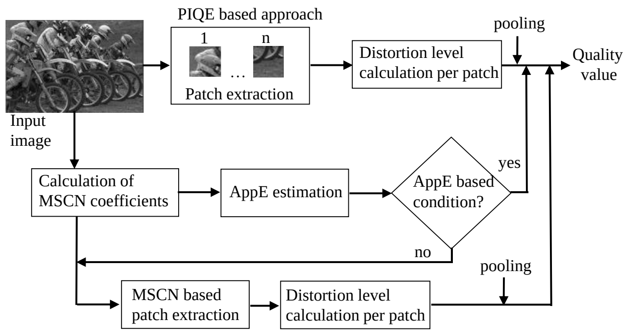

Fig. 1. AppE based perception framework for OU NR-IQA.

This is obtained here for PIQE metric and LIVE dataset, Fig.2. Higher concentration of points are visible. Also, this does not affect SROCC and PLCC results in a great manner which is presented in Table I. The highest results are obtained for wn, while ff distortion gives the lowest correlation results.

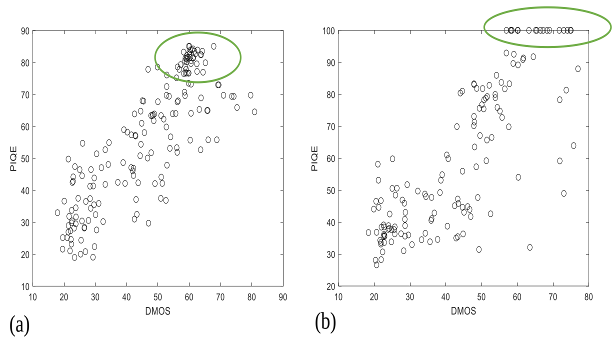

Fig. 2. High DMOS effects in PIQE versus DMOS comparison for: (a) jpeg and (b) ff impaired images.

TABLE I. Correlation results per distortion

No.	Dist. type	PLCC per image	SROCC per image	PLCC total	SROCC total
1	jp2k	Mean 0.9479 Std 0.0420	Mean 0.9328 Std 0.0781	0.8769	0.8830
2	jpeg	Mean 0.9218 Std 0.0779	Mean 0.8792 Std 0.1895	0.8330	0.8313
3	wn	Mean 0.9579 Std 0.0233	Mean 1 Std 0	0.9353	0.9853
4	gblur	Mean 0.9751 Std 0.0314	Mean 0.9732 Std 0.0782	0.8881	0.9104
5	ff	Mean 0.8673 Std 0.2459	Mean 0.8134 Std 0.2441	0.8014	0.7856

AppE is then calculated for MSCN coefficients, and this shows that there are impaired versions of each content per distortion that give higher values meaning unpredictive content. They generally seem to affect the trend in an image based analysis of SE (here zero mean is set for images) and the trend found in PIQE. This is illustrated in Fig.3.

Thus, instead of combining local and global perception in a traditional manner, only local analysis is performed for low AppE values. The correlation results per distortion using low

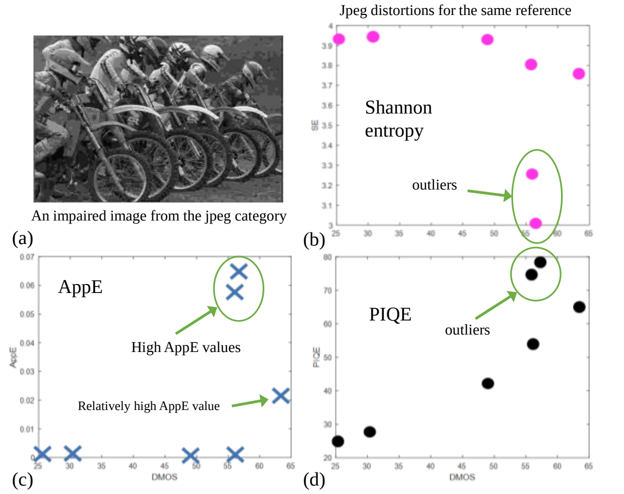

Fig. 3. (a) An example of distorted image and (b) SE, (c) AppE and (d) PIQE outliers per pristine image reference in the same jpeg category.

AppE are shown in Table II, where values less than 0.001 are used to show presence of patterns. Also, the correction of filtered points is given in Fig.4.

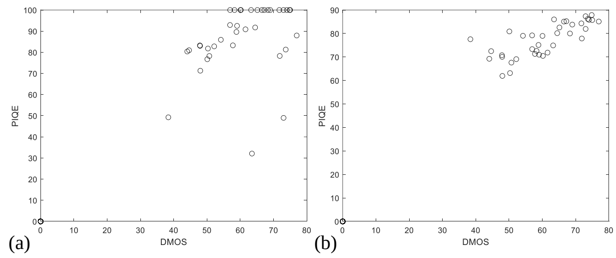

Fig. 4. Points of high AppE values (a) before and (b) after effect correction for ff category.

TABLE II. CORRELATION RESULTS PER DISTORTION USING LOW APPE

Dist. type	Images [%]	PLCC per image	SROCC per image	PLCC total	SROCC total
1 (jp2k)	75.7	Mean 0.9876 Std 0.0083	Mean 0.9996 Std 0.0012	0.9806	0.9998
2 (jpeg)	46.9	Mean 0.9883 Std 0.0168	Mean 0.9984 Std 0.0046	0.9643	0.9995
3 (wn)	100	Mean 0.9905 Std 0.0116	Mean 1 Std 0	0.9922	1
4 (gblur)	71	Mean 0.9935 Std 0.0079	Mean 0.9930 Std 0.0371	0.9825	0.9997
5 (ff)	74.5	Mean 0.9745 Std 0.0569	Mean 0.9841 Std 0.0742	0.9680	0.9995

It can be seen in Table II that wn is incorporated in PIQE local examination approach giving the highest results. The proposed filtering does not affect this category. SROCC and PLCC give satisfying results after filtering content with low AppE by covering about 72.7% of total amount of distorted images (779 images). Moreover, the proposed framework is able to correct and improve the correlation results as shown in Fig.4.

VI. Conclusion

The performed analysis show that applying well-known naturalness and metrics like PIQE may exhibit particular consequences when compared with DMOS. Moreover, the potential application of approximate entropy is considered showing its suitability to detect specific effects found while common fitting. So far, regularity from the aspect of this entropy has not been considered. The ability to employ this entropy as global evaluation may be useful for local estimation blind IQA tasks. Future experiments should be widened to different types of content, distortions, as well as entropies that seem to be highly consistent with perception.

Acknowledgment

The analysis in this paper is performed within bilateral cooperation supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia. The research is supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, no.: 451-03-47/2023-01/200103.

References

[1] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. on image processing*, *21*(12), pp. 4695-4708, 2012.

[2] R. A. Manap, and L. Shao, “Non-distortion-specific no-reference image quality assessment: A survey,” *Information Sciences*, *301*, pp. 141-160, 2015.

[3] B. Yan, B. Bare, and W. Tan, “Naturalness-aware deep no-reference image quality assessment,” *IEEE Trans. on Multimedia*, *21*(10), pp. 2603-2615, 2019.

[4] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal processing letters*, *20*(3), pp. 209-212, 2012.

[5] A. Pandey, et al., “Evaluation of Perception based Image Quality Evaluator (PIQE) no-reference image quality score for 99mTc-MDP bone scan images,” *The Journal of Nuclear Medicine*, pp. 1415, 2020.

[6] I. Stępień, and M. Oszust, “A brief survey on no-reference image quality assessment methods for magnetic resonance images,” *Journal of Imaging*, *8*(6), pp. 160, 2022.

[7] J. Wang, et al., “LighterGAN: an illumination enhancement method for urban UAV imagery,” *Remote Sensing*, *13*(7), pp.1371, 2021.

[8] D. Varga, “A Human Visual System Inspired No-Reference Image Quality Assessment Method Based on Local Feature Descriptors,” *Sensors*, *22*(18), pp. 6775, 2022.

[9] L. Wu, et al., “VP-NIQE: An opinion-unaware visual perception natural image quality evaluator,” *Neurocomputing*, *463*, pp. 17-28, 2021.

[10] A. Gavrovska, et al., “No-reference local image quality evaluation,” In *2022 30th Tel. Forum* (TELFOR), pp. 1-4, IEEE, 2022.

[11] A. Delgado-Bonal, and A. Marshak, “Approximate entropy and sample entropy: A comprehensive tutorial,” *Entropy*, *21*(6), pp. 541, 2019.

[12] S. Mejia-Romero, et al., “Evaluation of Bio-movements Using Nonlinear Dynamics,” In *Proc. of Int. Conf. on Data Science and Applications: ICDSA 2019*, pp. 197-208, Springer Singapore, 2021.

[13] A. Histace, et al., “Active Contour Segmentation based on Approximate Entropy,” In *Proc. of the Int. Joint Conf. on Biomedical Eng. Systems and Technologies*, 4, pp. 270-277, 2014.

[14] C. Pappalettera, et al., “Analysis of complexity in the EEG activity of Parkinson’s disease patients by means of approximate entropy,” *GeroScience*, *44*(3), pp.1599-1607, 2022.

[15] Q. Gao, et al., “Combined global and local information for blind CT image quality assessment via deep learning,” In *Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*, Vol. 11316, pp. 242-247, SPIE, 2020.

[16] X. Guan, et al., “Entropy based data expansion method for blind image quality assessment,” *Entropy*, *22*(1), pp. 60, 2019.

[17] H.R. Sheikh, M.F. Sabir and A.C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Trans. on Image Processing*, *15(11)*, pp. 3440-3451, 2006.