A HIGHLY EFFICIENT METHOD FOR BLIND IMAGE QUALITY ASSESSMENT

Qingbo Wu[1,2], Zhou Wang[2] and Hongliang Li[1]

[1]Schl. of Electronic Engineering, University of Electronic Science and Technology of China, China
[2]Dept. of Electrical & Computer Engineering, University of Waterloo, Canada

ABSTRACT

Blind image quality assessment (BIQA) has attracted a great deal of attention due to the increasing demand in industry and the promising recent progress in academia. To bridge the gap between academic research accomplishment and industrial needs, high efficiency BIQA approaches that allow for real-time computation are highly desirable. In this paper, we propose a novel BIQA method by selecting statistical features extracted from binary patterns of local image structures. This allows us to largely reduce the feature space to eventually one dimension. Somewhat surprisingly, such a single feature, faster-than-real-time approach named local pattern statistics index (LPSI) exhibits impressive generalization ability across different distortion types and achieves competitive quality prediction performance in comparison with state-of-the-art approaches on public databases such as LIVE II and TID2008.

Index Terms— Image quality assessment, blind image quality assessment, local binary pattern.

1. INTRODUCTION

No-reference or blind image quality assessment (BIQA) approaches predict perceived quality of a test image without referencing to an original image that is assumed to have pristine quality [1]. BIQA has become an active research area in the past few years due to the high demand in real-world applications and the significant research progress that has been made recently [2, 3, 4, 5, 6, 7]. BIQA is highly challenging not only because of the difficulty in accurately estimating human behaviors in evaluating image quality across different visual content, distortion types and distortion levels, but also because real-world applications such as online quality monitoring often require the image and video signals to be evaluated at high speed, ideally in real-time. Consequently, there is a strong practical need of highly efficient BIQA algorithms.

An interesting recent development in BIQA research is the incorporation of machine learning techniques, which have resulted in a number of state-of-the-art algorithms that report highly competitive performance [8, 9, 10, 11, 12]. The development of these algorithms basically follows a two-stage framework —- a feature extraction stage where a large number of features (possibly in the order of thousands) that are believed to be possibly relevant to image quality are extracted from the image, and a machine learning stage where image examples with different distortion types and levels are employed to train a quality prediction model. There are three limitations of this general approach. First, because of the high dimensionality of the feature space and the high complexity of the machine learning model, it is hard to deduce the actual contributions of individual features in the complete model, making it difficult to obtain a better understanding regarding human visual perception. Second, regardless of the performance on the training set, the generalization capability of these approaches is largely constrained by the training image samples, which live in an extremely high dimensional space (the space of all possible images) where it is difficult to justify how many samples are sufficient. Third, the large number of features require significant computation and storage resources, making them too slow and expensive in real-world applications. A recent study on opinion-free (OF) BIQA [13] has made an interesting attempt to avoid the supervised learning process and demonstrated great potentials, though its natural scene statistics (NSS) feature extraction process impedes it from real-time implementations.

In this paper, we aim to develop highly efficient method for BIQA that avoids any training process but delivers good generalization capability. Such an approach belongs to the OF category and relies heavily on the features being employed. In [14], Ojala et al. introduced an efficient local structural information descriptor named local binary pattern (LBP). By encoding the signs of the differences between a central pixel and its neighbors, the complex local structure is summarized by one of a limited set of binary patterns. By making use of LBP as the basis for our local feature extraction and modifying upon the statistics built on top of it, we obtain a handful of statistics on local image patterns. Careful inspection of these patterns on a large number of images suggests that some of them have significant discriminating power between high quality natural images and distorted ones. Somewhat surprisingly, we show that even by picking the statistic of one of the patterns, we can obtain a BIQA measure that is competitive with state-of-the-art approaches. Experiments with LIVE II and TID2008 databases demonstrate impressive generalization capability of such a training-free method.

2. PROPOSED METHOD

Let g_i and g_p denote the gray levels of the ith pixel and its circularly symmetric neighbor in an image, respectively. Let $t(\cdot)$ be the step function

$$t(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (1)$$

Then the uniform rotation invariant LBP code of the ith pixel can be expressed as

$$LBP_{P,R}(i) = \begin{cases} \sum_{p=0}^{P-1} t(g_p - g_i), & \text{if } U\big(LBP_{P,R}(i)\big) \leq 2 \\ P+1, & \text{otherwise} \end{cases}, \quad (2)$$

where the subscripts P and R represent the neighbors' number and radius, respectively. $U(\cdot)$ denotes the discontinuities of the circular binary presentation [14]

$$U\big(LBP_{P,R}(i)\big) = |t(g_{P-1} - g_i) - t(g_0 - g_i)| + \sum_{j=1}^{P-1} |t(g_j - g_i) - t(g_{j-1} - g_i)|. \quad (3)$$

Instead of using the equal voting scheme in accumulating the LBP code at each pixel [14], we employ a locally weighted scheme to refine the statistics. First, we normalize the gray level in the image by

$$\hat{g}_i = \frac{g_i - g_{min}}{g_{max} - g_{min}}, \quad (4)$$

where $\hat{g}_i$ denotes the normalized gray level. g_{max} and g_{min} are the max and min gray values in the input image. We then compute the local variance σ_i^2 at each pixel, and the vote of the pixel belonging to the LBP code l is computed by

$$v_l(i) = \frac{I_l\big(LBP_{P,R}(i)\big)}{\sigma_i^2 + c}, \quad (5)$$

where c is a small constant to avoid instability. $I_l(\cdot)$ is the indicator function

$$I_l\big(LBP_{P,R}(i)\big) = \begin{cases} 1, & \text{if } LBP_{P,R}(i) = l \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Let N be the number of the pixels in an image. The locally weighted statistic for the LBP code l is obtained by

$$s_l = \frac{1}{N} \sum_{i=1}^{N} v_l(i), \quad (7)$$

The dynamic range of s_l is $[0,+\infty)$. To facilitate observation, we scale the s_l to $[0,1)$ using a monotonic nonlinear mapping

$$\hat{s}_l = \frac{s_l}{s_l + \alpha}, \quad (8)$$
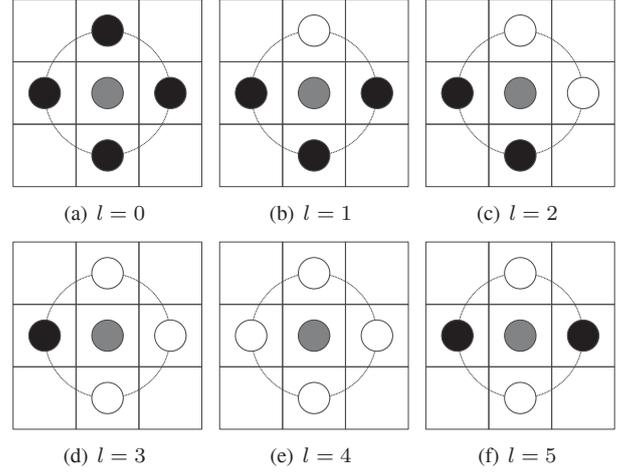
Fig. 1. Diagram of the local structures for $LBP_{4,1}$. The central pixel is labeled as a gray circle. The black and white circles denote the neighbors which have lower and larger intensities than the central pixel, respectively.

where α is a constant used to adjust the nonlinearity of the mapping. In the extreme case, when s_l equals to 0, $\hat{s}_l$ becomes 0. At the other extreme, when $s_l \gg \alpha$, $\hat{s}_l$ approximates 1.

As mentioned earlier, we aim for high efficiency BIQA method whose complexity is as low as possible. Thus, we only use four neighbors to compute the LBP codes (i.e., $LBP_{4,1}$) at each pixel, resulting in six distinct binary patterns [14]. The diagram of all possible LBP structures with four neighbors are shown in Fig. 1.

To investigate the discriminant ability of $\hat{s}_l$, we select 1000 high quality natural images from the VOC2012 database [15]. Four types of common distortions are simulated, i.e., JP2K, JPEG, WN and Blur. For each distortion type, we introduce five distortion levels, whose SSIM [16] scores are approximately $\{0.7, 0.75, 0.8, 0.85, 0.9\}$, respectively. As such, we obtain 5000 distorted images for each distortion type. For clarity, the $\hat{s}_l$ distributions across all natural images and their distorted versions are shown in Fig. 2, where the labels on the left denote the distortion type and those on the top denote our statistics for specific binary patterns. Overall, $\hat{s}_0$ presents the smallest histogram intersections for most distortion types. In addition, it is interesting to find that natural images' $\hat{s}_0$ concentrate on a very narrow range between 0.9 to 1. By contrast, the $\hat{s}_0$ values of the degraded images span a larger dynamic range, and are typically smaller than 0.9. Following the criterion in [17], the $\hat{s}_l$ histogram intersection $K_\cap(\cdot, \cdot)$ between the natural images and their degraded ones is used to measure the discriminant ability of each $\hat{s}_l$

$$K_\cap(h_r^l, h_d^l) = \sum_{i=1}^{n} \min\big(h_r^l(i), h_d^l(i)\big), \quad (9)$$
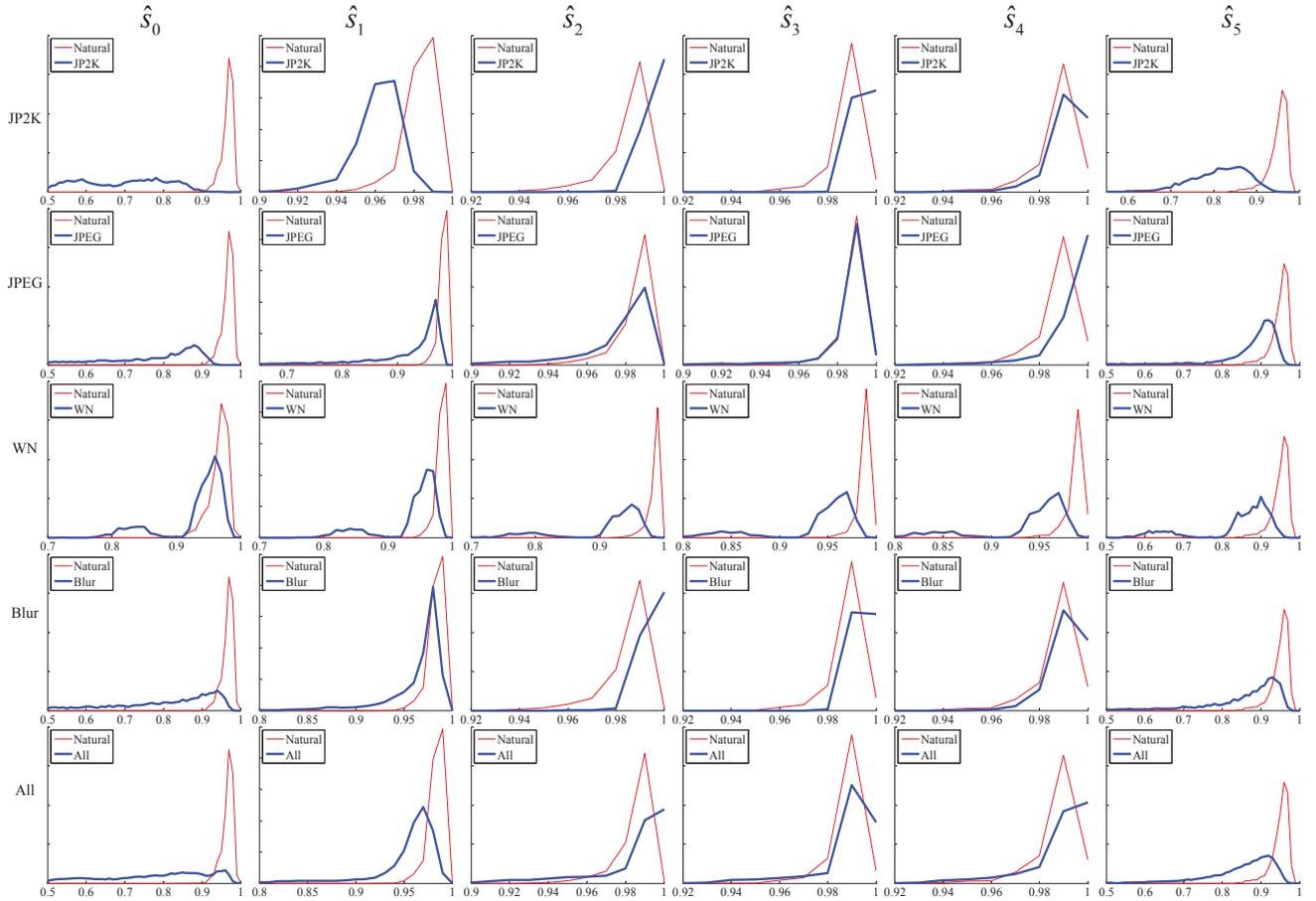
340

Fig. 2. Local pattern statistics of high quality natural images and distorted images.

where h_r^l and h_d^l denote the normalized $\hat{s}_l$ histograms of the natural images and their distorted versions, respectively. A smaller $K_{\bigcap}(h_r^l, h_d^l)$ indicates a better ability in differentiating distorted images from natural images.

The histogram intersections for each $\hat{s}_l$ across different distortion types are shown in Table 1, where the smallest $K_{\bigcap}(h_r^l, h_d^l)$ in each row is highlighted by boldface. It can be seen that the statistic $\hat{s}_0$ presents the best discriminant ability for JP2K, JPEG and Blur. When we mix the degraded samples across all four distortion types, $\hat{s}_0$ also achieves the smallest histogram intersection, as shown in the last row of Table 1. Based on this observation, we select only one dimension statistic $\hat{s}_0$ and define $Q = \hat{s}_0$ as our local pattern statistics index (LPSI).

3. EXPERIMENTAL RESULTS

We test the performance of the BIQA algorithms on LIVE II database [18], which includes 5 distortion types (i.e., JP2K, JPEG, WN, Blur, FF) and 779 distorted images. Since our experiment involves several opinion-aware (OA) BIQA methods which require training, we randomly divide the LIVE II

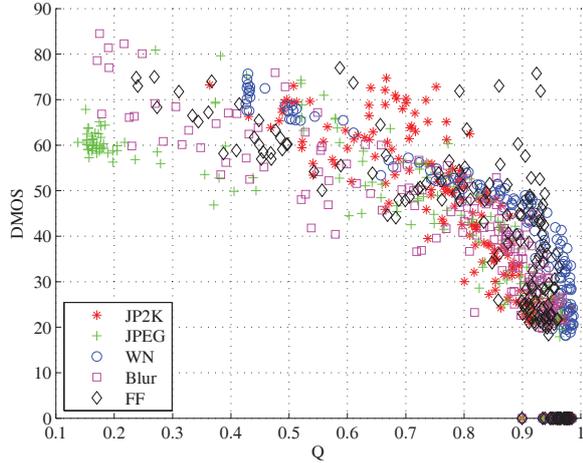**Table 1**. Histogram intersection of $\hat{s}_l$ between the natural images and their distorted versions.

Type	$\hat{s}_0$	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$	$\hat{s}_4$	$\hat{s}_5$
JP2K	**0.004**	0.178	0.336	0.548	0.743	0.098
JPEG	**0.019**	0.192	0.718	0.941	0.453	0.309
WN	0.567	0.183	**0.123**	0.144	0.196	0.209
Blur	**0.171**	0.617	0.408	0.573	0.762	0.365
All	**0.161**	0.312	0.500	0.668	0.645	0.276

database into two non-overlapping subsets for training and testing. Particularly, we use 23 of the 29 reference images and their distorted versions as the training set, and the remaining images constitute the testing set. We repeat this random splitting procedure 1000 times on the LIVE II database. The median Spearman's rank order correlation coefficient (SROCC) between the predicted quality and the DMOS across 1000 trials is used to evaluate a BIQA metric. Both the OA-BIQA (BIQI [2], BLINDS-II [5], BRISQUE [6] and CORNIA [7]) and OF-BIQA (NIQE [13]) metrics are included in this comparison.

The median SROCC results are reported in Table 2, where the best metric in each column is highlighted by boldface.

Table 2. SROCC performance on the LIVE II database

Distortion		JP2K	JPEG	WN	Blur	FF	All
Metric	Type						
BIQI	OA	0.736	0.591	0.958	0.778	0.700	0.726
BLINDS-II	OA	0.951	0.942	0.978	0.944	**0.927**	0.920
BRISQUE	OA	0.914	**0.965**	**0.979**	0.951	0.877	0.940
CORNIA	OA	0.943	0.955	0.976	**0.969**	0.906	**0.942**
NIQE	OF	0.917	0.938	0.966	0.934	0.859	0.914
LPSI	OF	**0.962**	0.937	0.892	0.962	0.923	0.921

Table 3. SROCC performance on the TID2008 database.

	OA				OF	
	BIQI	BLINDS-II	BRISQUE	CORNIA	NIQE	LPSI
SROCC	0.352	0.387	0.322	0.388	0.242	**0.449**

Table 4. Comparison of average running time (seconds)

	OA				OF	
	BIQI	BLINDS-II	BRISQUE	CORNIA	NIQE	LPSI
LIVE II	1.204	167.295	0.466	7.855	0.850	**0.073**
TID2008	1.028	100.562	0.361	7.427	0.494	**0.038**

We use program running time as a rough estimate of computational complexity. We compare the average running time for an image on LIVE II and TID2008 databases, respectively. The system platform is Intel Core 2 processor of speed 2.0GHz, 6GB RAM and Windows 7 64-bit version. All methods are tested with the MATLAB2013a software. The results are presented in Table 4, where the fastest results are highlighted by boldface in each row. Since the proposed measure only needs to compute a simple one dimension statistic, it is much faster than all existing BIQA algorithms, allowing it to be easily extended to real-time applications.

Fig. 3. The scatter plot of the predicted quality vs. the ground-truth subjective quality on LIVE II database. The x-axis denotes the predicted quality Q and y-axis denotes the $DMOS$.

The proposed LPSI algorithm, which only employs a single feature and does not require any training, provides highly competitive performance, even when comparing with OA algorithms that may involve sophisticated training procedures. Fig. 3 shows the scatter plot of the LPSI measure against DMOS for all images in the LIVE II database. It can be seen that LPSI exhibits a nearly linear relationship with DMOS, and achieves a reasonable balance across different distortion types.

To test the generalization capability of different BIQA metrics, we perform an experiment on the TID2008 database [19], which includes 17 distortion types and 1700 distorted images. For all OA-BIQA metrics, we use the entire LIVE II database to train the quality prediction function. Table 3 presents the SROCC results on the whole TID2008 databases across 17 distortion types, where the best metric is highlighted by boldface. It can be seen that the proposed LPSI method significantly outperforms all the other BIQA algorithms. It is worth nothing that the performance drops from LIVE II to TID2008 databases are significantly larger in all other BIQA algorithms being tested than in the proposed LPSI method, which, unlike all the OA methods, does not involve any training process. This demonstrates that LPSI has better generalization capability than existing BIQA methods.

4. CONCLUSION

We propose a highly efficient BIQA approach based on statistics of local binary patterns. The proposed LPSI algorithm eventually uses a single feature and does not involve any training process, but exhibits surprisingly good performance and generalization capability in comparison with state-of-the-art BIQA algorithms. We believe these are critical properties that are essential in real-world online image and video quality monitoring applications.

5. REFERENCES

[1] Zhou Wang and A.C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, Nov 2011.

[2] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, 2010.

[3] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[4] Tsung-Jung Liu, Weisi Lin, and C.-C.J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1793–1807, May 2013.

[5] M. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[6] A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[7] Peng Ye, J. Kumar, Le Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 1098–1105.

[8] Chaofeng Li, A.C. Bovik, and Xiaojun Wu, "Blind image quality assessment using a general regression neural network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, 2011.

[9] Lihuo He, Dacheng Tao, Xuelong Li, and Xinbo Gao, "Sparse representation for blind image quality assessment," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 1146–1153.

[10] Xinbo Gao, Fei Gao, Dacheng Tao, and Xuelong Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 12, pp. 2013–2026, Dec 2013.

[11] Le Kang, Peng Ye, Yi Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 1733–1740.

[12] Huixuan Tang, Neel Joshi, and Ashish Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2877–2884.

[13] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, March 2013.

[14] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[16] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

[17] K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," in *IEEE Int. Conf. on Computer Vision (ICCV)*, Oct 2005, vol. 2, pp. 1458–1465.

[18] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2*, [Online]. Available: http://live.ece.utexas.edu/research/quality.

[19] 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," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.