COMPLETELY BLIND IMAGE QUALITY ASSESSMENT USING LATENT QUALITY FACTOR FROM IMAGE LOCAL STRUCTURE REPRESENTATION

Min Zhang[1], *Yifan Li*[2,3], *Yu Chen*[2]

[1]Department of Mathematics, Northwest University, Xi'an, 710027, China
[2] School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[3]Shaanxi Electric Power Information & Telecommunication Co., Ltd., Xi'an 710048, China

ABSTRACT

Although opinion-unaware (OA) blind image quality assessment (BIQA) is the most difficult task, it is still very attractive because of its great potential for good generalization capability and practical usage. In this paper, a novel OA-BIQA algorithm is proposed. This algorithm is based on the construction of a 'quality aware' collection of statistical features based on a simple and successful local structure descriptor – generalized local binary pattern (GLBP). GLBP statistical features of images are regarded as the latent characteristics that can distinguish the visual distorted image from those of 'natural' or 'pristine' images. First, the statistical GLBP feature vectors learned from a corpus of the pristine images are divided into different subgroups. Second, a set of multivariate Gaussian (MVG) models are learned from each subgroup features fitting. Finally, the quality of a test image is evaluated by integrating the distances between its pair-wised parameters which describes each MVG models and that of each MVG models learned from a corpus of pristine images. Experimental results show that the proposed model delivers performance correlates well with human difference mean opinion scores on the LIVE IQA database.

Index Terms— Completely blind, image quality assessment, generalized local binary pattern

1. INTRODUCTION

With the popularity of multimedia and social network in our daily life, capturing and sharing images have been much more convenient and receptive than before. The digital visual content has surrounded people and image quality assessment (IQA) has an increasing desire. The purpose of the image quality assessment is looking for a computational model to evaluate the perceptual image quality automatically, making the evaluated result agree with the result of the subjective rating score[1].

If no image reference information is available during the quality evaluation task, then IQA is conducted without any prior knowledge. This so-called No reference (NR)/Blind IQA is the most attractive and of great importance because of its potential practical usage. With tighter conditions, general-purpose BIQA models can be classified into two subcategories. If a BIQA algorithm relies on human subjective opinion to train the quality prediction model, then it is an 'opinion-aware' (OA) BIQA algorithm.

The OA-BIQA algorithms are usually based on the train-test framework which requires a large number of distorted images with human judgments to learn the regression model. The limitation of OA-BIQA is in their generalization capability. Normally, their range of application is limited to the distortion types they have been trained on. When applying a model learned on a database to another database, or to real-world distorted images, the quality prediction performance may deteriorate significantly. Most existing general purpose BIQA algorithms belong to OA-BIQA, such as BRISQUE [2], BLIINDS-II [3], CBIQ [4] and NR-GLBP [5], etc. These algorithms usually deploy some suitable ''quality-aware'' features, such as natural scene statistics (NSS) in the wavelet domain [3], the DCT domain [6], or the spatial domain [2] , or using features that reflect or approximate NSS, such as image edges [7], image statistics from generalized local binary patterns [5]. Recently, more OU-BIQA algorithms appeared [8-10].

Recently, the deep neural network is introduced to IQA task, such as the implementation of Convolutional Neural Network (CNN) model in full reference (FR) and blind image quality assessment [11-14], etc. However, as the number of parameters to be trained in deep networks is usually very high, the training set has to have enough data samples to avoid overfitting. Until now, the generalization capability of these deep neural network based algorithms are not satisfying.

Considering the impracticality of obtaining collections of distorted images with co-registered human scores, it is desirable to develop 'opinion-unaware' (OU) BIQA models, which do not need any training samples of distortions nor human judgments [15]. However, OU-BIQA is considered

the most difficult quality evaluation tasks due to the limited available information.

Current progress in OU-BIQA algorithms is limited. In [15], Mittal et al. proposed an NSS-driven OU-BIQA model based on the construction of a set of 'quality aware' features from the spatial domain NSS statistics. With fitting the feature vectors to a multivariate Gaussian (MVG) model, the quality of a test image is evaluated by the distance between its parameters, which describes the MVG model and that of the MVG model learned from a corpus of pristine images. After that, Lin Zhang et al. improves this algorithm by integrating four more natural image statistics features derived from multiple cues to learn a multivariate Gaussian (MVG) model of pristine images [16]. This algorithm has a better generalization capability on different types of image distortions and various databases. In very recent days, more OU-BIQA algorithms started to appear [17, 18].

However, until now, no opinion-unaware algorithm has good quality prediction accuracy comparable with currently available FR IQA algorithms. It is a big challenge and of great significance to develop an opinion-unaware model that compete state-of-the-art OA-BIQA models.

Recently, we proposed a set of new local structure descriptors called generalized local binary patterns (GLBPs). Benefit from the statistics of image GLBP features, a novel and efficient OA-BIQA algorithm learned from a regression model has accomplished subsequently [5]. Our contribution in this paper is the development of a GLBP-based OU-BIQA algorithm which does not need training samples of distortions nor the human subjective score. The new OU-BIQA quality index performs better than the popular FR peak-signal-to-noise-ratio (PSNR) and structural similarity (SSIM) index and delivers performance at par with top performing OA-DA BIQA approaches. Our experiments demonstrate that the proposed model delivers performance correlates well with human difference mean opinion scores on the LIVE IQA database.

The rest of this paper is organized as follows. Section 2 presents the proposed OU-BIQA algorithm. Section 3 details the experimental setup. Section 4 presents the preliminary experimental results. Section 5 concludes the paper.

2. GLBP BASED OU-BIQA INDEX

Given the GLBP feature extraction method described below, we then derive a powerful OU-BIQA model from it.

2.1. Local Structure Feature Extraction

In [5], we proposed an OA-BIQA algorithm based on a novel feature descriptor which is called generalized local binary pattern (GLBP). GLBP is a reasonable extension of the traditional local binary pattern (LBP) to obtain more visual brightness contrast information as well as local structure information of image discriminant information by setting multi-threshold [19]. Hence, GLBP is quite efficient to approximate quality-aware visual words. It is defined as:

$$GLBP_{P,R,T}(t_c) = \sum_{p=0}^{P-1} S(t_p - t_c)2^p, \text{ where } S(t) = \begin{cases} 1, & t \geq T \\ 0, & t < T \end{cases} \quad (1)$$

The GLBP code is defined as follow: each contrast pixel t_p is transformed into a binary code of either "0" or "1", according to whether the difference between t_p and t_c is more than threshold T. Obviously, GLBP provides a more flexible option than LBP, and GLBP becomes LBP when $T = 0$. The parameter R is the radius of t_c neighbourhood and P is the number of pixels selected in the neighbourhood.

Similar to the definition of the LBP with different characteristics, the 'uniform' pattern of the GLBP and local rotation invariant pattern of the GLBP can be defined [5].

With multi-scale representation from Laplacian-of-Gaussian (LOG) filter set [5], GLBP statistical feature of each subband of a original image $\boldsymbol{X}$ is represented as:

$$GLBP_{P,R,\sigma,T}(\boldsymbol{X}_c) = \sum_{p=0}^{P-1} S'(\boldsymbol{F}(\boldsymbol{X}_p, \sigma) - \boldsymbol{F}(\boldsymbol{X}_c, \sigma))2^p \quad (2)$$

$\boldsymbol{F}_{subband}(\boldsymbol{X}, \sigma)$ is the subband image from LOG filter with the pixel size of $[M, N]$. And σ is the standard deviation of LOG filter.

For each subband image, the histogram with multi-threshold can be stacked and shown as follow:

$$\boldsymbol{J}_{\sigma}^{W} = (\boldsymbol{H}_{GLBP_{P,R,\sigma,T_0}}, \boldsymbol{H}_{GLBP_{P,R,\sigma,T_2}}, ..., \boldsymbol{H}_{GLBP_{P,R,\sigma,T_{W-1}}}) \quad (3)$$

where W is the number of threshold collection T.

Finally, we can obtain the quality-aware feature of whole image by stack each subband feature together:

$$\boldsymbol{J}_{GLBP}(\boldsymbol{I}) = (\boldsymbol{J}_{\sigma_1}^{W}, \boldsymbol{J}_{\sigma_2}^{W}, ..., \boldsymbol{J}_{\sigma_D}^{W}) \quad (4)$$

where D is the number of decomposed subband images.

The total number of GLBP feature PN is determined by the number of threshold collection W, the maximum value of $GLBP$ encoding map K and subband images D:

$$PN = W \times K \times D \quad (5)$$

2.2. Patch Selection

A plural number of pristine images are used and partitioned into sized patches, and GLBP features are extracted from selected image patches individually. Then the regularity of natural images is embodied from GLBP statistical features with a set of multivariate statistical models [15]. The reason for the patch selection is as follows.

Due to the restriction of photographic equipment or aesthetic appeal, every image is subject to some limiting distortion. For example, many images have some blur parts which are out of focus due to the inherent characteristic of single-lens camera's limited depth of field (DOF). Since defocus blur parts represents a potential loss of visual information, we use a simple way to differentiate patches

with rich visual information amongst a collection of natural patches from a corpus of natural images.

Index those $P \times P$ sized patches as $p = 1,2,..,n$, and compute the average local deviation field of each patch:

$$\delta(p) = \sum_{(i,j) \in patch_p} \sum \sigma(i,j) \quad (6)$$

where $\delta(p)$ indicates the local activity of patch p.

A single threshold ST is used to differentiate patches:

$$ST = sh * \delta_{\max} \quad (7)$$

Where sh is a fraction in the range [0, 1]; $\delta_{\max}$ is the peak sharpness among these patches. A subset of patches has suprathreshold sharpness $\delta \geq ST$ are selected as sharp ones.

2.3. Statistical Models from GLBP Feature of Natural Image

We consider that GLBP statistical features of images can be regarded as the inherent characteristics of image local structure representation that can distinguish the visual distorted image from those of 'natural' or 'pristine' images. Hence, GLBP statistical features extracted from the selection of patches of a corpus of pristine images are used for the construction of natural image statistical model.

Different from the traditional implementation of single global multivariate Gaussian (MVG) distribution fitting with NSS features in [15, 16], multi-MVG models are proposed for the construction of new natural pristine image statistics model in this paper. The single MVG density is defined as:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} \left|\mathbf{\Sigma}\right|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})) \quad (8)$$

Where $\mathbf{x}$ is the feature matrix. $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ denote the mean and covariance matrix of the MVG distribution, respectively, And the MVG model can be fully described by the pair $(\boldsymbol{\mu}, \mathbf{\Sigma})$. These two statistical parameters are estimated using the standard maximum likelihood estimation [20].

The computed GLBP features from each patch are classified into W subgroups depending on the threshold collection $\mathbf{T} = \{T_0, T_2, ..., T_{W-1}\}$ in Equation (3). Then each subgroup features are fitted to an MVG model and the corresponding model statistical parameter is denoted as $(\boldsymbol{\mu}, \mathbf{\Sigma})_{T_i}$, $i \in (0, W-1)$.

Hence, the natural image statistics are learned as a set of MVG models from a corpus of natural, undistorted images, which are described by the statistical parameter denoted as $(\boldsymbol{\mu}^{ref}, \mathbf{\Sigma}^{ref})_{T_i}$, $i \in (0, W-1)$. The MVG models' statistical parameter set of a test image is denoted as $(\boldsymbol{\mu}^{dist}, \mathbf{\Sigma}^{dist})_{T_i}$, $i \in (0, W-1)$.

The quality of a test image is then predicted by the Mahalanobis distance between its pair-wised parameters which describes each MVG models and that of each MVG models learned from a corpus of pristine naturalistic images.

$$D_{T_i}\left((\boldsymbol{\mu}^{ref}, \mathbf{\Sigma}^{ref})_{T_i}, (\boldsymbol{\mu}^{dist}, \mathbf{\Sigma}^{dist})_{T_i}\right) = \sqrt{\left((\boldsymbol{\mu}^{ref} - \boldsymbol{\mu}^{dist})^T \left(\frac{\mathbf{\Sigma}^{ref} + \mathbf{\Sigma}^{dist}}{2}\right)^{-1} (\boldsymbol{\mu}^{ref} - \boldsymbol{\mu}^{dist})\right)} \quad (9)$$

Finally, we can obtain the image quality predicted score D_{fusion}:

$$D_{fusion} = \sum_{i=0}^{W-1} \log D_{T_i} \quad (10)$$

3. EXPERIMENTAL SETUP

A corpus of 125 high-quality natural images with resolution ranged from 480×320 to 1280×720 is selected in this paper. These images are collected from copyright free Flickr data and Berkeley image segmentation database [21]. The image can be viewed and downloaded at http://live.ece.utexas.edu/research/ quality/pristinedata.zip.

The GLBP statistical features are extracted from each patch of the same size $P \times P$ of a test image. In our experiment, we chose patch size as 96×96. One noticeable point is that low sharpness patches of test image do not need to be removed from feature extraction patch collection since lost sharpness is a type of distortion called blur. Disregarding them will cause deviation in final image quality prediction. Four decomposition scales of LOG filters and three thresholds parameter are used for GLBP feature extraction in the proposed algorithm.

We set $sh = 0.5$. The performance is quite stable while sh changes in the range of [0.4, 0.6]. This subset of patches is then used to construct a model of the statistics of natural image patches.

To test the performance of the proposed algorithm, we used the LIVE IQA database [22]. LIVE IQA database includes 29 reference images and 779 distorted images of 5 distortion categories – JPEG, JP2k, white noise (WN), Gaussian blur (Blur) and fastfading (FF). It also provides a difference mean opinion score (DMOS) of each image obtained from human visual experiment to represents its subjective quality rating score.

We used Spearman's rank order correlation coefficient (SROCC), and Pearson's linear correlation coefficient (LCC) to test the proposed model. The predicted scores are passed through a logistic non-linearity [5] before computing LCC for mapping to DMOS space. The proposed algorithm is compared with two FR indices: PSNR and SSIM [23], Four general purpose OA-BIQA algorithms - CBIQ [3], BLIINDS-II [3], BRISQUE [2] and two OU-BIQA algorithms including NIQE [15] and IL-NIQE [16].

Table 1. SROCC comparison on the LIVE database

	Type	JP2K	JPEG	WN	Blur	FF	All
PSNR	FR	0.865	0.883	0.941	0.752	0.874	0.864
MS-SSIM		0.939	0.947	0.964	0.905	0.939	0.913
BRISQUE	OA-BIQA	0.914	**0.965**	**0.979**	0.951	0.877	0.940
BLIINDS-II		0.929	0.942	0.969	0.923	0.889	0.931
CBIQ		0.919	**0.965**	0.933	0.944	**0.912**	0.930
NR-GLBP		**0.947**	0.956	**0.979**	**0.954**	0.889	**0.951**
NIQE	OU-BIQA	0.908	0.934	0.965	0.924	0.858	0.906
IL-NIQE		0.894	0.942	0.970	0.915	0.832	0.897
The Proposed		**0.911**	**0.963**	**0.974**	**0.945**	**0.864**	**0.908**

Table 2. LCC comparison on the LIVE database

	Type	JP2K	JPEG	WN	Blur	FF	All
PSNR	FR	0.876	0.903	0.917	0.780	0.879	0.859
MS-SSIM		0.941	0.946	0.982	0.900	0.951	0.907
BRISQUE	OA-BIQA	0.923	**0.973**	**0.985**	0.951	0.903	0.923
BLIINDS-II		0.935	0.968	0.980	0.938	0.896	0.935
CBIQ		0.920	0.928	0.954	0.949	0.939	0.928
NR-GLBP		**0.956**	0.972	**0.985**	**0.954**	**0.912**	**0.954**
NIQE	OU-BIQA	0.916	0.945	0.973	0.938	**0.881**	0.915
IL-NIQE		0.905	0.959	**0.986**	0.932	0.864	0.902
The Proposed		**0.920**	**0.966**	0.978	**0.957**	0.878	**0.907**

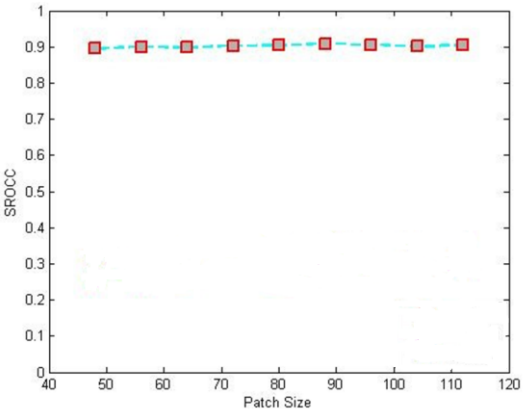

Fig. 1. Performance variation with different patch size on the LIVE database.

4. PERFORMANCE EVALUATION AND DISCUSSION

Tables 1 and 2 summarize both the performance of the overall prediction accuracy and the specific distortion prediction accuracy of the proposed algorithm with other state-of-the-art algorithms, as well as two FR indexes - PSNR and SSIM on the LIVE database. The best correlations are marked in bold in term of OA and OU algorithms, respectively. From Table 1 and 2, it is evident that the GLBP statistical features based BIQA algorithms are highly correlated to human perceptual quality. The proposed GLBP statistical feature based OU-BIQA algorithm is comparable with two state-of-the-art OU-BIQA algorithms on the LIVE database in both the individual distortion types and overall performance. From the comparison of results between OA-BIQA and OU-BIQA, it is also clear that the performance of OU-BIQA algorithms still cannot compare with OA-BIQA. There are still a lot of works need to be done in the future.

Figure 1 shows the SROCC performance against the size of image patches. It may be observed that a quite stable natural model can be obtained across a large range of the image patch size from 48×48 to 112×112. The SROCC performance has little difference in a very broad range of the size of the image patches. It should be noted that the parameters of the proposed algorithm have not been fully optimized. And with integrating more natural image statistics features, the performance may be improved further.

5. CONCLUSION

This paper offered a new OU-BIQA approach based on GLBP statistical feature. The image quality is expressed as a distance metric between the constructed 'pristine' model statistics from GLBP and those of the distorted image. The experimental results indicate that the performance of the proposed algorithm is comparable with the state-of-the-art. GLBP statistical features based OU-BIQA framework have much potential in practical usage and further research.

6. ACKNOWLEDGEMENT

This research work is financially supported by the National Natural Science Foundation of China (No. 61701404) and partially supported by Major Program of National Natural Science Foundation of China (No. 81727802), Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 17JK0769).

7. REFERENCES

[1] A. Bovik, “Automatic prediction of perceptual image and video quality,” Proceedings of the IEEE, vol. 101, no. 9, pp. 2008–2024, 2013.

[2] 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.

[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, 2012.

[3] A. K. Moorthy and A. C. Bovik, ‘‘Blind image quality assessment: From natural scene statistics to perceptual quality,’’ IEEE Trans. Image Process., vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[4] P. Ye and D. Doermann, “No-reference image quality assessment using visual codebook,” in IEEE Int. Conf. Image Process., 2011.

[5] Zhang M, Muramatsu C, Zhou X, et al. Blind Image Quality Assessment Using the Joint Statistics of Generalized Local Binary Pattern[J]. IEEE Signal Processing Letters, 2014, 22(2):207-210.

[6] M. A. Saad, A. C. Bovik, and C. Charrier, ‘‘A DCT statistics based blind image quality index,’’ IEEE Signal Process. Lett., vol. 17, no. 6, pp. 583–586, Jun. 2010.

[7] H. Tang, N. Joshi, and A. Kapoor, ‘‘Learning a blind measure of perceptual image quality,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Toronto, ON, Canada, Jun. 2011, pp. 305–312.

[8] Wu, Qingbo, et al. "Blind Image Quality Assessment Based on Multichannel Feature Fusion and Label Transfer." *IEEE Transactions on Circuits & Systems for Video Technology* 26.3(2016):425-440.

[9] Xu, J., et al. "Blind Image Quality Assessment based on High Order Statistics Aggregation. " *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* 25.9(2016):4444-4457.

[10] Siahaan, Ernestasia, A. Hanjalic, and J. A. Redi. "Augmenting Blind Image Quality Assessment Using Image Semantics." *IEEE International Symposium on Multimedia* IEEE, 2017:307-312.

[11] Bosse S, Maniry D, Muller K R, et al. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment.[J]. IEEE Transactions on Image Processing, 2017, PP(99):1-1.

[12] Fu, Jie, H. Wang, and L. Zuo. "Blind image quality assessment for multiply distorted images via convolutional neural networks." *IEEE International Conference on Acoustics, Speech and Signal Processing* IEEE, 2016:1075-1079.

[13] Bianco, Simone, et al. "On the use of deep learning for blind image quality assessment." *Signal Image & Video Processing* 3(2016):1-8.

[14] Kang, Le, et al. "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks." *IEEE International Conference on Image Processing* IEEE, 2015:2791-2795.

[15] Mittal A, Soundararajan R, Bovik A C. Making a “Completely Blind” Image Quality Analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3):209-212.

[16] Zhang L, Zhang L, Bovik A C. A Feature-Enriched Completely Blind Image Quality Evaluator[J]. IEEE Transactions on Image Processing, 2015, 24(8):2579-2591.

[17] Yang, Xichen, Q. Sun, and T. Wang. "Completely blind image quality assessment based on gray-scale fluctuations." *Proceedings of SPIE - The International Society for Optical Engineering* 9159.3(2014):243-251.

[18] Zhou, Wujie, et al. "Utilizing binocular vision to facilitate completely blind 3D image quality measurement." *Signal Processing* 129.C(2016):130-136.

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

[20] K. Sharifi and A. Leon-Garcia, “Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video,

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in Int. Conf. Comput. Vision, vol. 2, 2001, pp. 416–423.

[22] 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.

[23] Z. Wang, A. C. Bovik, H. R.Sheikh, and E. P. Simoncelli, “Image quality assessment: From error measurement to structural similarity,” IEEE Trans.ImageProcess.,vol.13,no.4,pp.600–612,2004.