

No-Reference Image Quality Assessment Based on Local Region Statistics

Qiaohong Li^{*}, Weisi Lin^{*†}, Yuming Fang[‡], Xinfeng Zhang[†], and Yabin Zhang[†]
^{*}*School of Computer Science and Engineering, Nanyang Technological University, Singapore*
[†]*Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore*
[‡]*School of Information Technology, Jiangxi University of Finance and Economics, China*
Email: {QLI013, WSLIN, FA0001NG, XFZHANG, ZHAN0398}@ntu.edu.sg

Abstract—In this paper, we propose an effective no-reference image quality assessment (IQA) method based on local region statistics (NRLRS). The proposed method is built on the hypothesis that image distortions may alter the local region statistics which can be well characterized by the inter-pixel relationship. Hence, by extracting perceptual features that describe the inter-pixel patterns of a distorted image, we can effectively quantify the impact of image degradation. For this purpose, the perceptual gray-level differences between neighboring pixels are extracted and a Gaussian Mixture Model (GMM) codebook is constructed as the generative model of extracted features. The Fisher vector representation is then derived to describe image as their derivations from the GMM model. Finally, partial least square regression is used to map the Fisher encodings to quality scores. Experimental results indicate that the proposed method achieves better performance in quality prediction as compared to relevant full-reference and no-reference IQA methods.

Index Terms—Gray-level differences, image quality assessment, no-reference, partial least square

I. INTRODUCTION

Image quality assessment (IQA) is an important research topic with wide applications in multimedia domain. Consumers always demand better image acquisition and delivery services, and thus it is highly desirable to automatically and accurately measure visual quality as perceived by these users. Objective IQA methods are developed to routinely and readily quantify the impact of various distortions through computational models. Based on the amount of reference information utilized in the model design, IQA methods are generally divided into three different groups: full-reference (FR), reduced-reference (RR) and no-reference (NR) methods [1]. In FR methods, the clean reference image is provided along with the distorted image for quality estimation. In RR methods, several features extracted from reference image sent through auxiliary channel or embedded into watermarking are required for comparison. For NR methods, only the distorted image is available for quality evaluation. Thus, NR methods are more versatile and challenging than other methods.

According to the application range, existing NR methods can be classified as distortion-specific and general-purpose

methods. The former ones target at specific application scenarios, such as Gaussian blur [2], JPEG 2000 (JP2K) [3] or JPEG compression [4], contrast change [5] and multiple distortions [6]. In such cases, the distortion type of the corrupted image is given as a priori. This constraint would limit the application domain of such methods, which also cannot handle online images whose distortion types are unknown. By contrast, general-purpose methods can evaluate perceptual quality of images degraded by various distortion types.

General-purpose NR-IQA methods can be classified into three categories according to the features representing image quality, i.e., natural scene statistics (NSS) based methods, perceptual features based methods and feature learning methods. BIQI [7] and DIIVINE [8] are wavelet NSS based methods with two-stage framework which first identifies distortion types and then evaluates the distortion-specific quality levels. BLINDS2 [9] is a DCT NSS based method that extracts statistical features from block DCT domain and estimates quality scores using a Bayesian inference approach. BRISQUE [10] is a spatial domain NSS based method that extracts curve parameters from the normalized luminance coefficients. Perceptual features sensitive to image distortions are also proved to be effective for NR-IQA task. For example, joint distributions of gradient and Laplacian of Gaussian (LOG) responses are employed as quality-aware feature by GMLOG method [11]. NRSL extracts structural and luminance histograms in spatial domain and validates their complementary roles for IQA [12]. Perceptual structural features extracted from multiple high order derivative images are used to build one NR-IQA method in [13]. Moreover, several NR methods adopt the feature learning strategy that generates high-dimensional representation from dictionary [14] or convolutional neural network [15] to predict visual quality.

In this work, we propose a novel general-purpose NR-IQA method based on local neighbourhood statistics. The proposed method is based on the hypothesis that the inter-pixel relationship can well capture the influence of image distortions on local region statistics. Unlike most previous NR methods that extract pixel-level statistical features from spatial or transform domain, we propose that statistical features characterizing image inter-pixel relationship are quite effective to describe the impact of various distortions, and IQA method built on it can achieve better performance than relevant NR

methods.

The paper is organized as follows. Section II describes the details of the proposed method. Extensive experimental results and analysis are provided in Section III. Section IV concludes the paper.

II. THE PROPOSED METHOD

The framework of the proposed method for NR-IQA is as follows. Given one image, we first extract the gray-level differences between neighboring pixels as the low-level features describing the inter-pixel relationship. Then we use GMM clustering to build a probabilistic codebook as the generative model of extracted low-level features. The GMM codebook serves as an intermediate step from which further features are computed as the Fisher vector representation, which describes the statistical differences between low-level features from one image and the GMM codebook. Finally, partial least square regression is utilized to estimate quality score for one image based on its Fisher vector representation.

A. Low-level feature extraction

In this work, we extract gray-level differences between neighboring pixels as the image low-level feature. For each pixel in one image, we first determine the circular neighboring pixels with d distance to it, and calculate the gray-level differences between the neighboring pixels and the central pixel. Suppose that x_c is one random pixel in an image, we denote its neighboring pixels with distance d as (x_1, x_2, \dots, x_P) . Then the gray level of the central pixel x_c is subtracted from its neighboring pixels, yielding the feature vector $\mathbf{y} = (y_1, y_2, \dots, y_P) = (x_1 - x_c, x_2 - x_c, \dots, x_P - x_c)$. \mathbf{y} is the gray-level differences vector of pixel x_c with distance of d . To account for the human nonlinear perception on physical stimuli, we use the following formula to compress the magnitude of the gray-level differences:

$$\mathbf{z} = \text{sign}(\mathbf{y}) \log(|\mathbf{y}| + 1) \quad (1)$$

\mathbf{z} is the extracted feature vector to describe the perceptual gray-level differences.

B. GMM codebook construction

We use images in CSIQ database to build a GMM codebook as the generative model of extracted low-level features. For each image in the database, we can get a set of low-level features $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N]$. The low-level features from all the images in CSIQ database form the training set Φ which is used to construct the GMM codebook. First, PCA whitening is applied to decorrelate the low-level features and to better fit the diagonal covariance matrix assumption of the GMM model. Then, a GMM model with K Gaussian components can be learnt on the training set using the expectation maximization (EM) algorithm. We denote the parameters of the K -component GMM by $\lambda = \{\omega_k, \mu_k, \Sigma_k, k = 1, \dots, K\}$, where ω_k , μ_k and Σ_k denote the mixture weight, mean vector and covariance matrix of the k -th component Gaussian. Each Gaussian component represents one codeword in the

codebook. In this work, we use diagonal covariance matrices to simplify the GMM model, and the GMM is fully specified by $(2D + 1)K$ scalar parameters, where D is the dimension of each low-level feature after PCA whitening.

C. Fisher vector representation

The generative GMM represents the distribution of low-level features in the training set. Fisher vector encoding captures the average first and second order differences between image low-level features and the GMM centers. Given a set of low-level features from an image, $\mathbf{z}_1, \dots, \mathbf{z}_N$, we first apply the PCA whitening to reduce the correlations and get $\mathbf{z}'_1, \dots, \mathbf{z}'_N$. The log-likelihood of \mathbf{z}'_i from this GMM is given by:

$$\mathcal{L}(\mathbf{z}'_i) = \log \sum_{k=1}^K \omega_k \mathcal{N}(\mathbf{z}'_i; \mu_k, \Sigma_k) \quad (2)$$

Assuming that low-level features in an image are independent, the log-likelihood of a set \mathbf{Z}' of N features is then represented as:

$$\mathcal{L}(\mathbf{Z}') = \sum_{i=1}^N \log \left(\sum_{k=1}^K \omega_k p_k(\mathbf{z}'_i) \right) \quad (3)$$

Fisher encoding [16] calculates the gradient of this log-likelihood with respect to GMM model parameters for all the feature vectors from one image:

$$\nabla \mathcal{L}(\mathbf{Z}') = \sum_{i=1}^N \nabla \log \left(\sum_{k=1}^K \omega_k \mathcal{N}(\mathbf{z}'_i; \mu_k, \Sigma_k) \right) \quad (4)$$

These gradients can be derived as follows for Gaussian model [16]:

$$\mathcal{G}_k^\mu = \frac{1}{\sqrt{\omega_k}} \sum_{i=1}^N \gamma_{ik} \left(\frac{\mathbf{z}'_i - \mu_k}{\sigma_k} \right) \quad (5)$$

$$\mathcal{G}_k^\sigma = \frac{1}{\sqrt{2\omega_k}} \sum_{i=1}^N \gamma_{ik} \left(\frac{(\mathbf{z}'_i - \mu_k)^2}{\sigma_k^2} - 1 \right) \quad (6)$$

where γ_{ik} is the probability that local descriptor \mathbf{z}'_i has been generated by the k -th Gaussian component. The final Fisher vector representation for the set of descriptors \mathbf{Z}' of one image is the concatenation of \mathcal{G}_k^μ and \mathcal{G}_k^σ .

D. Partial least square regression

By applying Fisher vector encoding to a K -component GMM with D -dimensional features, the final image representation is of length $2KD$. For a GMM with 1024 components and features of 14 dimensions, this yields descriptors of approximately 28,672 dimensions. Considering the high-dimensional image representation, partial least square (PLS) regression is adopted here to map from this high-dimensional feature space to quality scores due to its low-complexity and remarkable capability to handle high-dimensional data [17]. PLS regression works by reducing the input high-dimensional features to several uncorrelated latent components and performing least squares regression on these components. There is only one parameter (the number of components) in this regression model, which can be determined through cross validation on the training set.

III. EXPERIMENT

A. Experimental protocol

LIVE [18] and TID2013 [19] databases are used to conduct experiments to show the performance of the proposed method. LIVE database is generated by introducing five distortion types to 29 reference images: JP2K compression, JPEG compression, Gaussian white noise (WN), Gaussian blur (GB) and simulated fast-fading Rayleigh channel (FF). TID2013 database is generated by introducing 24 distortion types to 25 reference images. In our experiments, we consider only the four common distortions (i.e., JP2K, JPEG, WN, GB) that are shared by the LIVE database.

The performance is evaluated by three criteria: Spearman rank order correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC) and root mean squared error (RMSE). The latter two criteria are calculated after the monotonic logistic mapping recommended by VQEG [20].

B. Implementation details

The implementation details are as follows: the number of GMM components is 1024; the neighboring pixel distance is 2; the number of neighboring pixels is 16; PCA reduced dimension of low-level feature is 14. We also adopt power normalization followed by L2 normalization before PLS regression to further boost the performance of Fisher vector representation [21], and the power normalization parameter is 0.2. For the PLS regression, the number of latent components is 7. We have also tried other parameters and found that these provided ones yield the optimal performance.

C. Evaluation on LIVE

On LIVE database, we first divide the whole database into training and test sets. Distorted versions of 23 (80%) random-chosen reference images are used for training, and the rest are used for performance evaluation. The median results of 1000 iterations of training-test split are reported as the final performance.

Table I to Table III shows the performance of state-of-the-art FR and NR IQA methods on LIVE database, including performance on specific distortion types as well as the whole database. The best NR method for each case is shown in boldface. NRLRS is the proposed method. From these tables we can observe that NRLRS delivers outstanding performance on all five distortion types as well as the overall performance. It outperforms previous state-of-the-art NR methods on JP2K, JPEG, GB, and FF distortion types. For the remaining WN distortion, its performance is quite similar to the best performing one. Furthermore, when all the distortion types are mixed together, NRLRS also achieves better performance than other NR methods.

Figure 1 shows the box plots of SRCC distributions of different NR-IQA methods across 1000 iterations on LIVE database. It can be clearly seen that the proposed method has the highest median value with the smallest standard deviation. It shows the content robustness of the proposed method,

TABLE I
SRCC COMPARISON ON THE LIVE DATABASE.

IQA model	JP2K	JPEG	WN	GB	FF	ALL
PSNR	0.9059	0.8946	0.9835	0.8078	0.8945	0.8845
SSIM	0.9602	0.9753	0.9769	0.9613	0.8879	0.9402
DIIVINE	0.9066	0.8967	0.9818	0.9324	0.8543	0.9107
BLIINDS2	0.9314	0.9494	0.9471	0.9159	0.8723	0.9309
BRISQUE	0.9170	0.9658	0.9800	0.9528	0.8865	0.9450
GMLOG	0.9258	0.9632	0.9831	0.9293	0.9012	0.9501
NRSL	0.9433	0.9597	0.9844	0.9591	0.8803	0.9517
NRLRS	0.9554	0.9701	0.9746	0.9595	0.9179	0.9611

TABLE II
PLCC COMPARISON ON THE LIVE DATABASE.

IQA model	JP2K	JPEG	WN	GB	FF	ALL
PSNR	0.9161	0.9136	0.9903	0.8230	0.9158	0.8823
SSIM	0.9723	0.9807	0.9871	0.9696	0.9032	0.9345
DIIVINE	0.9244	0.9158	0.9893	0.9421	0.8852	0.9114
BLIINDS2	0.9428	0.9653	0.9677	0.9402	0.9144	0.9369
BRISQUE	0.9362	0.9792	0.9887	0.9585	0.9174	0.9488
GMLOG	0.9412	0.9779	0.9912	0.9427	0.9365	0.9536
NRSL	0.9536	0.9763	0.9908	0.9668	0.9143	0.9557
NRLRS	0.9702	0.9839	0.9839	0.9680	0.9418	0.9616

as different content partitions would result in similar SRCC performance.

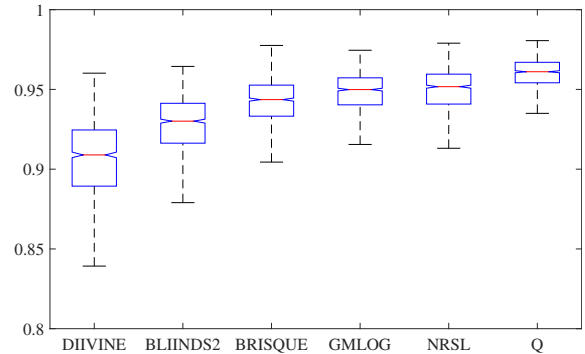


Fig. 1. The SRCC distributions across 1000 training-test iterations on the LIVE database.

D. Cross database validation

Cross database validation is conducted to test the generalization ability of our method. We train our model using all the images from LIVE database, and then test its performance on TID2013 database. Only the four common distortion types that are shared by LIVE and TID2013 databases are included

TABLE III
RMSE COMPARISON ON THE LIVE DATABASE.

IQA model	JP2K	JPEG	WN	GB	FF	ALL
PSNR	9.945	12.877	3.858	10.386	11.126	12.799
SSIM	5.854	6.184	4.490	4.470	12.153	9.697
DIIVINE	9.577	12.628	4.031	6.097	13.104	11.248
BLIINDS2	8.326	8.292	7.073	6.237	11.312	9.491
BRISQUE	8.838	6.437	4.195	5.166	11.138	8.583
GMLOG	8.437	6.592	3.719	6.096	9.736	8.207
NRSL	7.509	6.882	3.770	4.683	11.248	8.018
NRLRS	6.118	5.698	5.014	4.549	9.505	7.516

in this experiment. The SRCC comparison between state-of-the-art NR-IQA methods is shown in Table IV. Since PLCC and RMSE follow the same trend as SRCC, here we omit them for brevity. Results of the cross database test show that the proposed method is very competitive to state-of-the-art NR methods and shows promising generalization capability to unseen database.

TABLE IV
SRCC COMPARISON ON THE TID2013 DATABASE. ALL NR METHODS ARE TRAINED USING THE WHOLE LIVE DATABASE.

IQA model	JP2K	JPEG	WN	GB	ALL
DIIVINE	0.8630	0.8302	0.8607	0.8749	0.8661
BLINDS2	0.8804	0.8695	0.7539	0.8200	0.8554
BRISQUE	0.9079	0.8896	0.8209	0.8724	0.8776
GMLOG	0.9361	0.8994	0.9081	0.8963	0.9072
NRSL	0.9370	0.8992	0.9034	0.9075	0.9159
NRLRS	0.9481	0.9216	0.8598	0.9075	0.9161

E. Discussions

In this work, by characterizing inter-pixel patterns in one image, an effective NR-IQA method has been built. The outstanding performance of the proposed method has validated the hypothesis that inter-pixel statistical features are effective to describe the impact of various distortions. Most previous NR methods are developed from pixel-level statistical features in spatial or transform domain. There are several existing methods which also utilize image inter-pixel features in a different manner. BRISQUE extracts curve parameters from pixel-pair products of normalized luminance coefficients. NRSL utilizes the local binary pattern (LBP) descriptor to capture the inter-pixel patterns. Our method is more advanced and comprehensive in several aspects. In BRISQUE, the 1D distribution of pixel-pair products only considers the relationship between one pixel pair at each time. While image distortions generally degrade all pixel pairs in one local region simultaneously, such degradation cannot be well captured by considering only one pixel pair. LBP captures the local pattern simultaneously, but it encodes only the sign information of gray-level differences, and the neglected magnitude information is also vital for quality perception. Our method has addressed the problems with both BRISQUE and NRSL, and provides an effective way to representing local neighbourhood statistics.

IV. CONCLUSION

In this paper, we propose a novel NR-IQA method based on perceptual features that describe image inter-pixel patterns. The Fisher vector representation extracted from GMM codebook generated from perceptual gray-level differences is combined with partial least square regression to build the proposed method. Extensive experiments demonstrate that the proposed method yields better performance than previous NR methods for perceptual quality estimation.

REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[2] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *2003. Proceedings. Seventh International Symposium on Signal Processing and Its Applications*, vol. 1, July 2003, pp. 469–472 vol.1.

[3] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[4] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *2002 International Conference on Image Processing*, vol. 1. IEEE, 2002, pp. I–477.

[5] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2015.

[6] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, April 2016.

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

[8] —, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

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

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

[11] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[12] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2016.

[13] Q. Li, W. Lin, and Y. Fang, "No-reference image quality assessment based on high order derivatives," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, July 2016, pp. 1–6.

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

[15] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 1733–1740.

[16] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007, pp. 1–8.

[17] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop*, Berlin, Heidelberg, 2006, pp. 34–51.

[18] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," 2005.

[19] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[20] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *VQEG*, Mar, 2000.

[21] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *2010 11th European Conference on Computer Vision (ECCV)*, Berlin, Heidelberg, 2010, pp. 143–156.