GIP: Generic Image Prior for No Reference Image Quality Assessment

Qingbo Wu[1], Hongliang Li[1], and King N. Ngan[1,2]

[1] University of Electronic Science and Technology of China, Chengdu 611731, China,
{qbwu,hlli}@uestc.edu.cn,
[2] Chinese University of Hong Kong, ShaTin, Hong Kong,
knngan@ee.cuhk.edu.hk

Abstract. No reference image quality assessment (NR-IQA) has attracted great attention due to the increasing demand in developing perceptually friendly applications. The crucial challenge of this task is how to accurately measure the naturalness of an image. In this paper, we propose a novel parametric image representation which is derived from the generic image prior (GIP). More specifically, we utilize the classic fields of experts model to capture the prior distribution of an image with respect to a random field, which is learned from a great deal of natural images. Then, the parameters in modeling this prior distribution are used as the quality-relevant image feature, which is represented by a simple two-dimension vector. Experimental results show that the proposed method achieves competitive quality prediction accuracy in comparison with the state-of-the-art NR-IQA algorithms at the expense of much less memory usage and computational complexity.

Keywords: generic image prior, no reference image quality assessment, human perception

1 Introduction

No reference image quality assessment (NR-IQA) aims to estimate the perceptual quality of a test image without accessing to its reference version. Due to the high demand in evaluating and improving the experience of user experience, NR-IQA has attracted a great deal of attentions in the past few years [18]. To adapt to various real-world applications under different platforms, a NR-IQA metric is expected to accurately estimate the image quality at the lowest overhead in terms of memory and computations. Therefore, it is highly desired to develop compact image representation for balancing the quality prediction accuracy against the computational cost.

Recently, the research on NR-IQA has made significant process [12, 11, 15, 9, 24, 4, 23, 5]. By incorporating the quality-aware features into a machine learning framework, many existing algorithms are reported to deliver state-of-the-art performance. An interesting development in this process is the ever-growing dimension of the image feature, where a higher dimensional feature space is usually believed to be beneficial for achieving good classification or regression performance [6, 2, 1]. In [12, 11, 15, 9], only tens of statistical indexes are extracted from the transform coefficients on wavelet, DCT and normalized spatial domains. Then, the image feature dimension increases to thousands in [10, 21]. By means of unsupervised learning, Ye *et al.* [24] proposed a state-of-the-art NR-IQA method, whose feature dimension is up to twenty thousand. Although a better prediction performance is achieved by extracting richer image information, the extremely increased memorial and computational cost impede their application in many platforms with limited computation resources.

In this paper, we aim to propose a compact image representation which could deliver good prediction performance with limited storage and computation resources. Unlike conventional natural scene statistics [12, 11, 15, 9] which are derived from handcrafted transform basis, we utilize the unsupervised learning method in [14] to train a set of filters whose responses on the natural images present regular student t-distribution [19]. Then, two parameters in modeling the raw distribution of the responses on these learned filters are used as the natural scene statistics. Except for the statistics on the real-value filter response, a statistical index derived from the response on specific local binary pattern [22] is also included in our image representation, which builds a three dimension feature referred to as generic image prior (GIP). Finally, by means of the support vector regression (SVR) [16], we can map the GIP feature to a quantitative image quality score. Experimental results on the popular IQA database – LIVE II [17] demonstrate the effectiveness and high-efficiency of the proposed NR-IQA method.

2 Proposed method

Accurately measuring the naturalness of an image is crucial for the NR-IQA task. Inspired by the natural scene statistics study in [14], we employ the unsupervised learning method to train a set of filters from a large number of natural images, which contain inherent prior information in measuring the naturalness. Let $\mathbf{x}$ denote the whole input image and $\mathbf{x_{(k)}}$ denote the kth cropped patches from $\mathbf{x}$. Let J_i and α_i denote the ith linear filter and model parameter to be learned. Then, the natural image prior $p(\cdot)$ can be presented as the full product of t-distribution model [20], i.e.,

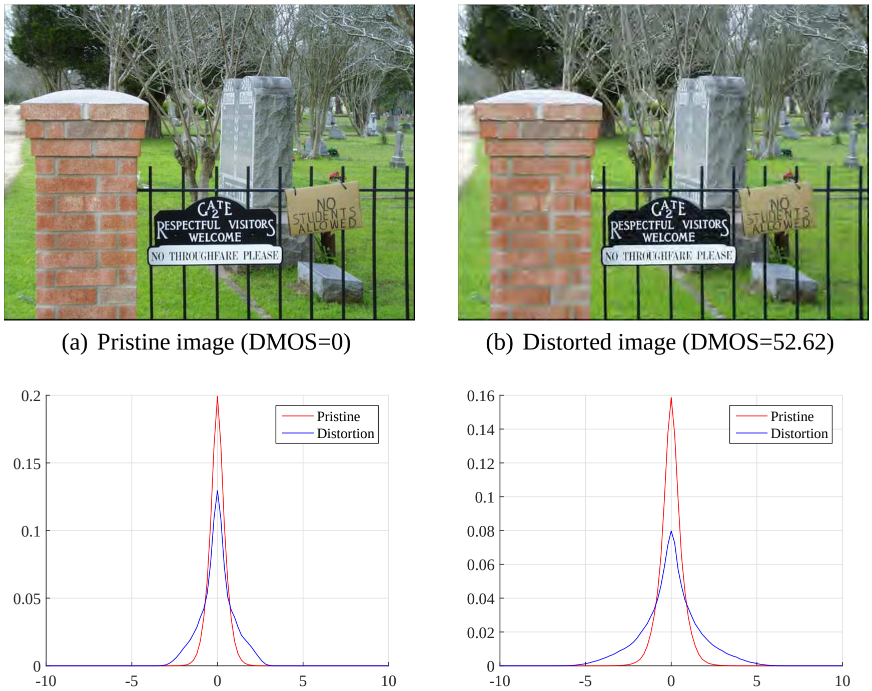
$$p(x;\Theta) = \frac{1}{Z(\Theta)} \prod_{k=1}^{K} \prod_{i=1}^{N} \phi_i(J_i^T x_{(k)}; \alpha_i) \tag{1}$$

where $\Theta = \{\theta_1, \cdots, \theta_N\}$ and $\theta_i = \{\alpha_i, J_i\}$. The experts ϕ_i can be given by

$$\phi_i(J_i^T x_{(k)}; \alpha_i) = \Big(1 + \frac{1}{2}(J_i^T x)^2\Big)^{-\alpha_i} \tag{2}$$

where the linear filters J_i and parameters α_i are learned from a set of generic natural images by maximizing its likelihood. It should be noted that there is no any annotation information needed in this training process. Following the configurations in [14], we train the image filers in two sizes, i.e., 3×3 and 5×5. All training images are collected from the Berkeley database [8].

After training the linear filters and parameters set, we can analyze the statistical property of an image from its response on the learned filters. More specifically, the



(a) Pristine image (DMOS=0)

(b) Distorted image (DMOS=52.62)

(c) Distribution of 3×3 filter response

(d) Distribution of 5×5 filter response

Fig. 1. The comparison of distributions between the pristine image and its JP2K version.

image response is derived from the gradient of the log-prior, i.e.,

$$\nabla_x \log p(x) = \sum_{i=1}^{N} J_i^- * \psi_i(J_i * x) \tag{3}$$

where $*$ denotes the convolution operation and J_i^- denotes the mirror filter of J_i. The function ψ_i represents the partial differential operator, i.e,

$$\psi_i(y) = \frac{\partial \log \phi_i(y; \alpha_i)}{\partial y} \tag{4}$$

To highlight the discriminating power of the learned naturalness-aware filters, the filter responses comparison between a pristine image and its distorted version is illustrated in Fig. 1. From the observation in Figs. 1 (c) and (d), it can be seen that the filter response of a natural image exhibits clear non-gaussian and heavy tailed distribution under different filter sizes. When the distortion is present, the distribution of the filter response would become more dispersed and deviate from the image with pristine quality. In order to quantify this variation, we model the filter response of an image with a standard t-distribution, whose probability density function (PDF) is given by,

$$f(t|n) = c(1 + \frac{t^2}{n})^{-(n+1)/2} \tag{5}$$

where c is a constant and $t = \nabla_x \log p(x)$. n is the degree of freedom of a standard t-distribution.

In this paper, the maximum likelihood estimation (MLE) [13] is utilized to solve the parameter n, whose PDF is most likely to approximate the raw density of t. Let $\mathcal{L}(n|t)$ denote the likelihood function, which is defined by reversing the roles of the observed data t and the parameter n in $f(t|n)$, i.e.,

$$\mathcal{L}(n|t) = f(t|n) \tag{6}$$

For computational convenience, the likelihood function is usually converted to its logarithm version, i.e., $\ln \mathcal{L}(n|t)$. Then, the MLE is formulated as

$$\hat{n} = \arg\max_{n \in \Omega} \ln \mathcal{L}(n|t) \tag{7}$$

where Ω is the parameter set. The Quasi-Newton algorithm [3] is employed to solve the optimization problem in (7).

Since two sizes of filters (i.e., 3×3 and 5×5) are learned in our simulation, we compute two distribution parameters $\hat{n}_{3\times3}$ and $\hat{n}_{5\times5}$ to capture the multi-scale natural scene statistics. Besides the real-value filter responses derived from the original image intensities, we also compute a local pattern statistical index (*LPSI*) [22] which captures the response on specific local binary pattern. By combing these three statistical indexes, we can obtain our GIP feature s, i.e.,

$$s = [\hat{n}_{3\times3}, \hat{n}_{5\times5}, LPSI] \tag{8}$$

In order to map the proposed GIP feature s to a perceptual quality score Q, the SVR is used to train the image quality regression function $R(s)$, i.e.,

$$R(s) = \sum_{k=1}^{K} \alpha_k \mathbb{K}(s, \hat{s}_k) + \beta \tag{9}$$

where α_k and β denote the SVR parameters to be learned from the training samples. $\hat{s}_k$ is the GIP feature of the kth support vector. $\mathbb{K}(\cdot,\cdot)$ is the kernel function, which is used to measure the similarity of two samples. Similar with previous works [12, 11, 9], the radial basis function kernel is used here, i.e.,

$$\mathbb{K}(s, \hat{s}_k) = \exp(-\gamma \|s - \hat{s}_k\|^2) \tag{10}$$

where γ is the custom parameter which is determined by cross validation.

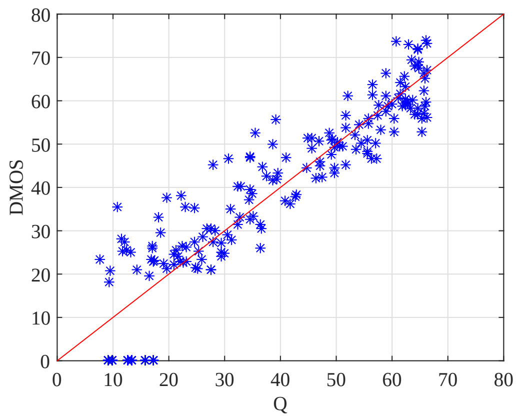


Fig. 2. The scatter plot of the predicted image quality Q versus the ground-truth human opinion scores DMOS on LIVE II database.

3 Experimental results

To verify the performance of the proposed NR-IQA method, we implement the image quality estimation on the LIVE II database [17], which is consist of 29 pristine images and 779 distorted images across 5 distortion types (i.e., JP2K, JPEG, WN, Blur and FF). In our simulation, two metrics are employed to evaluate the performance of the proposed method, which include the Pearson's linear correlation coefficient (LCC) and Spearman's rank ordered correlation coefficient (SROCC). For both metrics, a value which is more close to 1 indicates a better performance in predicting the human perception.

Following the criterion in [11, 15, 9], we randomly divide the LIVE II database into two non-overlap subsets, where the 80% reference images and their associated distorted versions are used to construct the training set, and the left 20% images are used for testing. The splitting procedure is repeated 1000 times and the median LCC and SROCC are reported for evaluating the final performance. Seven state-of-the-art NR-IQA algorithms are included in this comparison, i.e., BIQI [12], DIIVINE [11], BLIINDS-II [15], BRISQUE [9], NIQE [10], CORNIA [25] and TCLT [21].

Fig. 2 first shows the scatter plot of predicted image quality via our GIP method against the DMOS on LIVE II database. It is clear that the predicted image quality presents nearly linear relationship with DMOS, which indicates that the proposed objective metric is highly consistent with the subjective perception in terms of the image quality. The median LCC and SROCC results are listed in Tables 1 and 2, respectively. For clarity, the best metric in each column is highlighted by boldface. Although a very compact image representation, which only includes three indexes, is used to capture the perceptual quality, the proposed GIP method delivers highly competitive performance in comparison with the state-of-the-art NR-IQA algorithms.

To measure the memory overhead of different NR-IQA algorithms, we summarize their feature dimensions and the corresponding memory usage of storing the feature

Table 1. The median LCC performance on the LIVE II database

Method	JP2K	JPEG	WN	Blur	FF	All
BIQI	0.750	0.630	0.968	0.800	0.722	0.740
DIIVINE	0.922	0.921	0.988	0.923	0.888	0.917
BLIINDS-II	0.963	**0.979**	0.985	0.948	**0.944**	0.923
BRISQUE	0.923	0.974	0.985	0.951	0.903	0.942
CORNIA	0.951	0.965	0.987	**0.968**	0.917	**0.935**
NIQE	0.937	0.956	0.977	0.953	0.913	0.915
TCLT	0.902	0.946	**0.989**	0.954	0.923	0.935
GIP	**0.977**	0.958	0.883	0.931	0.920	0.923

Table 2. The median SROCC performance on the LIVE II database

Method	JP2K	JPEG	WN	Blur	FF	All
BIQI	0.736	0.591	0.958	0.778	0.700	0.726
DIIVINE	0.913	0.910	**0.984**	0.921	0.863	0.916
BLIINDS-II	0.951	0.942	0.978	0.944	0.927	0.920
BRISQUE	0.914	**0.965**	0.979	0.951	0.877	0.940
CORNIA	0.943	0.955	0.976	**0.969**	0.906	**0.942**
NIQE	0.917	0.938	0.966	0.934	0.859	0.914
TCLT	0.898	0.923	0.979	0.940	0.903	0.934
GIP	**0.978**	0.935	0.897	0.965	**0.929**	0.936

of an image in Table 3, where the smallest value in each column is highlighted by boldface. It is seen that the proposed GIP method spends smallest memory for sorting the image feature, which is beneficial for saving the storage space and bandwidth in many real-world applications.

To analyze the computational complexity of different NR-IQA algorithms, we compute their running time of training a support vector regressor using all images in the LIVE II database. It is noted that the parameter tuning procedure is included in this evaluation, where two parameters (C,γ) are determined by grid search [7]. The candidate parameter space is 2^S, where S ranges from 0 to 5 with the interval 0.5. The system platform is Intel Core 2 Duo processor of speed 2.0GHz, 6GB RAM and Windows 7 64-bit version. All methods are tested using the MATLAB2015a software. Table 4 reports the detailed running time in terms of seconds, where the smallest value is highlighted by boldface. It is seen that the proposed GIP method spends much less running time in comparison with other algorithms, which is highly desired for many low power devices.

4 Conclusion

In this paper, we proposed a GIP based no reference image quality assessment method. Specifically, a set of naturalness-aware filters are learned from large number of pristine images. The distribution of filter responses is used to measure the perceptual quality variation. Meanwhile, the LPSI is used to compensate the image information derived from these real-value filter responses. Finally, a compact image representation, which only computes three statistical indexes, is used as the quality-aware feature. Experi-

Table 3. The memory usage for storing the feature of an image in terms of Kilobyte (KB)

Method	Feature dimension	Memory usage
BIQI	18	0.325
DIIVINE	88	0.880
BLIINDS-II	24	0.368
BRISQUE	36	0.466
CORNIA	20000	149.591
NIQE	1332	8.162
TCLT	4413	9.172
GIP	**3**	**0.198**

Table 4. Running time of different NR-IQA algorithms for training SVR (seconds)

Method	Running time
BIQI	87.821
DIIVINE	214.956
BLIINDS-II	94.121
BRISQUE	117.997
CORNIA	5.243×10^4
NIQE	3.492×10^3
TCLT	1.094×10^4
GIP	**45.716**

mental results demonstrate its effectiveness and high-efficiency in evaluating the image quality.

5 Acknowledgement

This work was supported in part by National Natural Science Foundation of China (No. 61525102, No. 61271289), and by The program for Science and Technology Innovative Research Team for Young Scholars in Sichuan Province, China (No. 2014TD0006).

References

1. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3025–3032 (2013)
2. Clarke, B., Fokoue, E., Zhang, H.: Principles and Theory for Data Mining and Machine Learning. Springer Series in Statistics, Springer New York (2009)
3. Dennis, Jr, J.E., Moré, J.J.: Quasi-newton methods, motivation and theory. SIAM review 19(1), 46–89 (1977)
4. Fang, Y., Ma, K., Wang, Z., Lin, W., Fang, Z., Zhai, G.: No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Processing Letters 22(7), 838–842 (July 2015)

5. Fang, Y., Zeng, K., Wang, Z., Lin, W., Fang, Z., Lin, C.W.: Objective quality assessment for image retargeting based on structural similarity. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 4(1), 95–105 (March 2014)
6. Han, J., Pei, J., Kamber, M.: Data Mining, Southeast Asia Edition. The Morgan Kaufmann Series in Data Management Systems, Elsevier Science (2006)
7. Hsu, C.W., Chang, C.C., Lin, C.J., et al.: A practical guide to support vector classification. Tech. rep. (2003)
8. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: IEEE International Conference on Computer Vision. vol. 2, pp. 416–423 (2001)
9. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21(12), 4695–4708 (2012)
10. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters 20(3), 209–212 (2013)
11. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Transactions on Image Processing 20(12), 3350–3364 (2011)
12. Moorthy, A., Bovik, A.: A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters 17(5), 513–516 (2010)
13. Redner, R.A., Walker, H.F.: Mixture densities, maximum likelihood and the em algorithm. SIAM review 26(2), 195–239 (1984)
14. Roth, S., Black, M.: Fields of experts: a framework for learning image priors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 860–867 vol. 2 (2005)
15. Saad, M., Bovik, A.C., Charrier, C.: Blind image quality assessment: A natural scene statistics approach in the dct domain. IEEE Transactions on Image Processing 21(8), 3339–3352 (2012)
16. Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural Comput. 12(5), 1207–1245 (2000)
17. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE Image Quality Assessment Database Release 2, [Online]. Available: http://live.ece.utexas.edu/research/quality
18. Wang, Z.: Applications of objective image quality assessment methods [applications corner]. IEEE Signal Processing Magazine 28(6), 137–142 (2011)
19. Welling, M., Osindero, S., Hinton, G.E.: Learning sparse topographic representations with products of student-t distributions. In: Advances in neural information processing systems. pp. 1359–1366 (2002)
20. Welling, M., Osindero, S., Hinton, G.E.: Learning sparse topographic representations with products of student-t distributions. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, pp. 1383–1390 (2003)
21. Wu, Q., Li, H., Meng, F., Ngan, K.N., Luo, B., Huang, C., Zeng, B.: Blind image quality assessment based on multichannel feature fusion and label transfer. IEEE Transactions on Circuits and Systems for Video Technology 26(3), 425–440 (2016)
22. Wu, Q., Wang, Z., Li, H.: A highly efficient method for blind image quality assessment. In: IEEE International Conference on Image Processing. pp. 339–343 (2015)
23. Wu, Q., Li, H., Meng, F., Ngan, K.N., Zhu, S.: No reference image quality assessment metric via multi-domain structural information and piecewise regression. Journal of Visual Communication and Image Representation 32, 205 – 216 (2015)
24. Ye, P., Doermann, D.: No-reference image quality assessment using visual codebooks. IEEE Transactions on Image Processing 21(7), 3129–3138 (2012)
25. Ye, P., Kumar, J., Kang, L., Doermann, D.: Unsupervised feature learning framework for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1098–1105 (2012)