BLIND IMAGE QUALITY ASSESSMENT BASED ON VISUO-SPATIAL SERIES STATISTICS

Ziheng Zhou[⋆] *Wen Lu*[⋆] *Jiachen Yang*[†] *Lihuo He*[⋆] *Xinbo Gao*[⋆]

[⋆] School of Electronic Engineering, Xidian University, Xi'an, China
[†] School of Electronic Information Engineering, Tianjin University, Tianjin, China

ABSTRACT

Existing blind image quality assessment (BIQA) methods based on statistics attach limited attention to the relative position of pixels. Features in these BIQA methods are too flimsy to characterize quite a few distortions with strong locality or complexity. However, psychological studies have shown that according to the relative position within visual field, the cognitive system generates visuo-spatial serial memory used for cognitive tasks, e.g., subjective image quality assessment. Inspired by the visuo-spatial series generated by human visual system (HVS), we propose a BIQA method based on imitation Visuo-spatial Series Statistics (VSS). The proposed method simulates visual system to construct visuo-spatial series based on the relative position of pixels, and use statistical features of visuo-spatial series to predict image quality. Extensive experiments demonstrate the proposed method has a superior performance compared to the state-of-the-art BIQA methods.

Index Terms— Blind image quality assessment, pixel relation, visuo-spatial series statistics.

1. INTRODUCTION

Image quality assessment (IQA) algorithms are designed with the aim of evaluating the quality of distorted images without any prior information from the reference image, which has been receiving significant attention. There are two main schemata in IQA: full-reference (FR) IQA [1, 2] and blind/no-reference (NR) IQA [3, 4, 5, 6, 7].

FR-IQA requires a pristine reference image to evaluate the quality of the distorted image. The structural similarity index (SSIM) proposed by Wang et al. [1] can be considered as a milestone of the development of IQA models, which provides a good approximation of the perceived image quality. Following the framework of SSIM, Zhang et al. [2] proposed the feature similarity index (FSIM), which further improves the performance of IQA.

Relative to FR-IQA, NR-IQA takes only the distorted image to be assessed as input and thus is more applicable. NR-IQA algorithms based on statistics always employ natural scene statistics (NSS). NSS believes the hypothesis that natural scenes possess certain statistical properties which would be altered in the presence of distortion. Statistical models to characterize the NSS-based NR-IQA algorithms have been investigated in both the spatial domain such as BRISQUE [5] and the transform domain such as wavelet [3] and DCT domain [4]. They evaluate the unnaturalness extent to quantify possible degradation of image quality. Besides these opinion-aware NR-IQA models above-mentioned, another category [6, 7] trained without subject databases belongs to opinion-unaware. The database-independence in training process have an improvement on their generalization capability in some way. However, they also do not obtain satisfying results on quite a few distortions. There is still considerable room to improve on NR-IQA.

Psychological studies [8] have shown that processing ordered or structured information is an essential aspect of human cognition, a specific function of short-term memory. In accordance with the relative position within visual field, cognitive system generates memory for series of visuo-spatial stimuli. Inspired by the visuo-spatial series, we have a reason to believe that proper imitation visuo-spatial series based on the relative position of pixels could be useful when applied in IQA tasks. To a certain extent, some of the existing work has been partially validated the feasibility of the hypothesis . Features of the mean subtracted contrast normalized (MSCN) coefficients [5, 6, 7] and local binary pattern (LBP) [9] are used in NR-IQA. Those methods performed well in IQA , but they not explore further.

In view of this hypothesis, we proposed a BIQA method which attempts to mimic the statistical model of most representative statistics features of the relative position between pixels. This paper mainly includes two contributions. First, a new visuo-spatial series reflecting relative position of pixels called the neighborhood co-occurrence matrix (NCM) is proposed. Second, statistics of the imitation visuo-spatial series are employed as features to form a novel BIQA method. The experimental results on benchmark databases indicate that the proposed method, VSS, is effective and performs statistically better than other state-of-the-art NR-IQA approaches, and even be comparable to some conventional FR measure, e.g., FSIM [2].

This work was supported by the National Natural Science Foundation of China (61372130, 61432014, 61501349, 61571343), and the Key Industrial Innovation Chain Project in Industrial Domain (2016KTZDGY04-02).

2. VISUO-SPATIAL SERIES FOR IQA

This part illustrates how to construct the visuo-spatial series based on the relative position of pixels. There are two kinds of visuo-spatial series in the proposed method: one is the classical mean subtracted contrast normalized (MSCN) coefficient; the other is a new distribution called neighborhood co-occurrence matrix (NCM).

2.1. Normalized luminance and MSCN products

One visuo-spatial series derive from the distribution of locally MSCN coefficients and products of pairs of adjacent MSCN coefficients[5, 6].

It is pointed out that the locally normalized luminance of a natural gray-scale image I conforms to a Gaussian distribution. The so-called MSCN coefficients $\bar{I}(i,j)$ have been observed to follow a unit normal distribution on natural images that have not suffered noticeable quality distortions. This Gaussian model, however, is violated when images are subjected to quality degradations caused by common distortions. Measurements of the deviation of $\{\bar{I}(i,j)\}$from the Gaussian model are indicative of distortion severity. As visual spatial sequences used in our method, the statistics of MSCN coefficients are effective for IQA tasks and have been applied in several methods [5, 6, 7].

2.2. NCM and image perceptual quality

The co-occurrence matrix is defined to describe the relationship of co-occurring pixel gray-scale values at a given offset. However, the natural image is highly structured and the relative position of pixel-domain carriers extensive structural information. One pixel at a fixed distance is inadequate to reflect the relative position of each pixel. Hence, we replace the co-occurring pixel with several pixels in the local neighborhood, and call it neighborhood co-occurrence matrix (NCM).

Different to above series mainly referred to pairs of adjacent MSCN coefficients, sequences in NCM reflect a one-to-many relationship between each pixel and its neighborhood. The construction process of NCM is clarified as follow. For an image with p different pixel values, the $p \times p$ neighborhood co-occurrence matrix C is defined over an $m \times n$ image I, parameterized by the mean of partial neighborhood, as:

$$C(i,j) = -\sum_{x=1}^{m}\sum_{y=1}^{n}\begin{cases} 1 & I(x,y)=i \; and \; u(x,y)=j, \\ 0 & otherwise. \end{cases} \quad (1)$$

where $u(x,y)$ is rounded the mean of several neighborhood pixels in certain directions to integer. Fig. 1 shows the mapping process of an image to construct NCM. In the Fig. 1, it's worth noting that j is obtained from the mean of three pixels labeled as $2-4$ in the up-right corner of the neighborhood.

An elliptical nonuniform distribution of NCM is often deployed on a natural image. A high density gathers in the ellipse's major axis, and rapidly diminishing with increasing eccentricity. Distortions will change the pixel relation in a local neighborhood, and the distribution of NCM would also be changed with various degrees of distortion. As shown in Fig. 2, we re-arrange pixels of partial regions in a natural image with rising degrees, and the distributions of NCM in the next line corresponding to these images reflect the change of pixel relation intuitively. Fig. 3 presents five distorted versions of Fig. 1(a) by JPEG 2000. With the level of distortion rising, the distribution of corresponding NCM aggregates to the major axis of the ellipse by degrees. Consistent with our expectations, the elliptical shape and natural distribution of NCM would be altered in the presence of distortions. It illustrates that NCM is able to represent the degree of image distortion or quality degradation, and some statistics of NCM would be

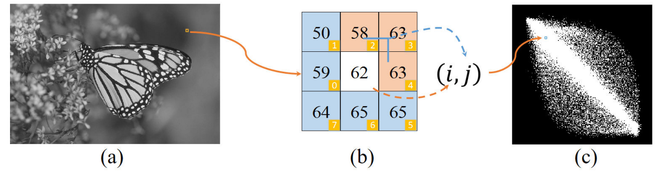

Fig. 1. The mapping process of an image to construct NCM: (a) a reference image from LIVE; (b) a pixel with its neighborhood as a sequence; (c) the NCM of (a).

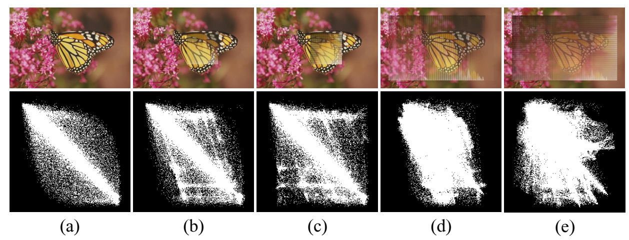

Fig. 2. Fig. 1(a) and its four pixels re-arranged versions in the top row, and their corresponding NCM in the next row: (a) natural image; (b) the partial region pixel re-arranged; (c) the partial region pixel further re-arranged; (d) and (e) similar to (b) and (c) but operated in a larger region.

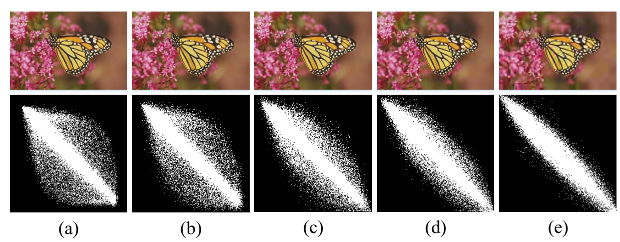

Fig. 3. Five distorted versions of Fig. 1(a) by JPEG 2000 in the top row and their corresponding NCM in the next row: the level of distortion rising from (a) to (e).

able to predict the quality of images.

3. THE PROPOSED METHOD BASED ON VISUO-SPATIAL SERIES STATISTICS (VSS)

The following part illustrates more details about features extracted from two kinds of imitation visuo-spatial series, and the framework of the method we proposed.

3.1. Statistics of visuo-spatial series

As to the first kind of visuo-spatial series, we extract features from multi-scale and multi-orientation following [5], and finally obtain a set of 36 dimensional features:

$$f_A = [f_1, f_2, f_3, \dots, f_{36}]^T \tag{2}$$

Psychological research [10] suggests information theory may provide a promising framework for the future investigation of the effect of path complexity in visuo-spatial serial memory. In particular, the concept of entropy is defined as a measure of disorder or randomness. Entropy has been used to measure the anisotropy of the image for BIQA [11]. Thus, we try to use statistical entropy of NCM to represent degree of quality degradation. Firstly, we obtain the entropy of the histogram obtained by summing the NCM along the direction of pixel magnitude, which is almost equal to the well-known information entropy. The histogram shows the number of pixels with same magnitude on different spatial positions. It is defined as:

$$e_1 = -\sum_{i=0}^{255} P_i \ln P_i \tag{3}$$

where P_i is the proportion of pixels whose magnitude is i in all pixels of NCM.

We also compute the entropy of NCM, which is called as structure entropy in this paper. If the information entropy described the aggregation property of pixels in the image, the structure entropy would represent the relative spatial position information between the pixels. It is defined as:

$$e_2 = -\sum_{i=0}^{255} \sum_{j=0}^{255} (P_{ij} \ln P_{ij}) \tag{4}$$

where i denotes the magnitude of the pixel, j denotes the mean of neighborhood and P_{ij} denotes the proportion of the element (i, j) in NCM.

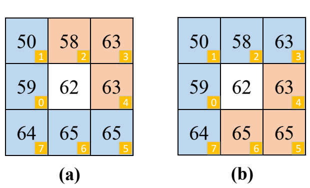

Fig. 4. Two orientation: (a) upper right; (b) lower right.

As shown in Fig. 4, the two orientations contain three pixels respectively. To reflect the change of the correlation structure in pixel-domain sensitively, we take three pixels from the above directions to conduct NCM, rather than all neighborhood pixels every time. In addition, it has been proved that image features derive from multi-scale and multi-orientation are beneficial to performance. Hereby, we build NCM from two orientations as Fig. 4 when calculate the mean of partial pixels in the neighborhood. There is no special meaning in the two orientations, just in consideration of the benefit from multi-orientation to the result and no more redundant computation. Other similar choices are equal, e.g., upper left and lower left.

What's more, the image gradient is a rich descriptor of local image structure and the local quality. Previous studies have found the distributions of its gradient components (partial derivatives) and gradient magnitudes are changed by introducing distortions to an image. NCM of gray-scale images reflect the relationship between pixels and their spatial position in luminance. And NCM of gradient map highlight the connection between pixels and their spatial position in structure. Hence, we conduct NCM and extract the same entropy features from gradient map. It's obvious that NCM obtained from two orientations have the same information entropy e_1 and different structure entropy e_{21}, e_{22}. Therefore, in terms of an image, we can extract a six-dimensional features vector from the gray-scale image and the gradient map:

$$f_B = [e_1, e_{21}, e_{22}, {e_1}', {e_{21}}', {e_{22}}']^T \tag{5}$$

3.2. Framework of the proposed method

The framework of VSS is illustrated in Fig. 5. For a test image, we extract statistics features of two kinds of visuo-spatial series for image quality prediction. And the final features can be obtained when we combine features as equation 4 and 5.

For each database, we partition it into a training subset and a testing subset. On training subset, the regression model is able to be trained to predict image quality scores with support vector machine (SVR) , and then we test the performance of our prediction model on the testing subset. The following experiments will prove the validity of the new features based on statistics of NCM for the quality prediction.

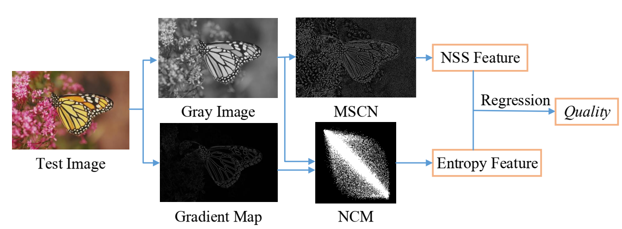

Fig. 5. The framework of the proposed method.

4. EXPERIMENTS AND RESULTS

Five common benchmark IQA datasets are used to evaluate the proposed index: LIVE [12], CSIQ [13], TID2013 [14], LIVE Multiply Distorted (MD) [15] and MDID [16]. The LIVEMD and MDID are multi-distortion image databases. Important informations about five databases are summarized in Table 1. These are 6715 distorted images in total.

For the visibility of all the tables, we only list one commonly metric to evaluate the performance of the competing BIQA methods. It is the Spearman Rank-order Correlation Coefficient (SRCC) which is computed between the objective scores predicted by the BIQA models and the subjective mean opinion scores (MOS) provided by the dataset. We compare the proposed VSS model with two conventional FR-IQA methods and five common BIQA methods as listed in Table 2. Because of the similar experimental settings, we have directly cited most results of those compared IQA methods from corresponding reference literature. If not find, we test them based on models provided by the original authors.

Firstly, we evaluate models on each individual dataset. We report results under the partition proportions that 80% of distorted images associated are used for training and the remaining 20% are used for testing. Each partition is randomly conducted 100 times on each dataset and the median SRCC is reported. Although several methods do not need training, we report their results on the partitioned test subset to make the comparison consistent. The results of performance on five individual datasets are shown in Table 2. The two best results for each column are highlighted. The results in Table 2 show the method we proposed in all test databases performs better than the other BIQA methods, even comparable with FR-IQA methods FSIM.

In order to evaluate the performance of IQA models adequately, we evaluate the various BIQA models on each individual distortion type in TID2013 dataset, which contains most types of distortion in five benchmark datasets. The SRCC on each individual distortion type is summarized in Table 3. The two best results for each row are also highlighted. According to Table 3, we can see that quite a few distortions in TID2013 are hard to characterize using the features of all the other BIQA models (the best SRCC under 0.6). However, the VSS method presents some excitement performance sometimes. Although on a few complex distortion types, none of the evaluated BIQA models could obtain satisfying results.

In VSS, we extract statistics features from two kinds of visuo-spatial sequences. In order to understand the contribution, we separately evaluate the performance of each type of features on TID2103. SRCC is used as the performance metric. When only using one kind, features based on MSCN perform 0.778, and features based on NCM perform 0.634. However, the SRCC of the VSS model is improved to 0.827 while combining statistics of two kinds of visuo-spatial series.

Table 1. The Number of Reference Images, Distortion Types and Distortion Images about Five Datasets

Datasets	Ref. Image	Dis. Type	Dis. Image
LIVE [12]	29	5	779
CSIQ [13]	30	5	886
TID2013 [14]	25	24	3000
LIVEMD [15]	15	2	450
MDID [16]	20	5	1600

Table 2. Performance (SRCC) on Each Individual Dataset

	LIVE	CSIQ	TID2013	LIVEMD	MDID
SSIM (FR) [1]	0.948	0.876	0.637	0.646	0.715
FSIM (FR) [2]	**0.963**	**0.924**	**0.802**	0.864	**0.896**
DIIVINE [3]	0.916	0.757	0.549	0.867	0.489
BLIINDS2 [4]	0.920	0.780	0.536	0.889	0.497
BRISQUE [5]	0.939	0.755	0.563	0.888	0.502
NIQE [6]	0.913	0.627	0.317	0.872	0.652
ILNIQE [7]	0.902	0.822	0.512	**0.917**	0.705
VSS	**0.953**	**0.891**	**0.827**	**0.926**	**0.905**

Table 3. Performance (SRCC) on Each Distortion Type in TID2013

	Ref. [3]	Ref. [4]	Ref. [5]	Ref. [6]	Ref. [7]	**VSS**
AGN	0.855	0.723	0.852	0.819	**0.876**	**0.916**
AGC	0.712	0.650	0.709	0.670	**0.816**	**0.888**
SCN	0.463	0.767	0.491	0.666	**0.923**	**0.926**
MN	0.675	0.513	0.575	**0.746**	0.512	**0.759**
HFN	**0.878**	0.825	0.753	0.845	0.869	**0.919**
IN	**0.806**	0.650	0.630	0.743	**0.755**	0.741
QN	0.165	0.782	0.798	**0.850**	**0.873**	0.779
GB	0.834	**0.856**	0.813	0.795	0.814	**0.892**
ID	0.723	0.712	0.586	0.590	**0.750**	**0.856**
JPEG	0.629	**0.864**	**0.852**	0.840	0.835	0.845
JP2K	0.853	**0.898**	0.893	**0.889**	0.858	0.888
JPEGTE	0.239	0.117	**0.315**	0.003	0.283	**0.699**
JP2KTE	0.061	**0.621**	0.359	0.510	0.525	**0.828**
NEPN	0.060	0.097	**0.145**	0.070	0.081	**0.317**
LBD	0.093	0.210	**0.224**	0.127	0.135	**0.742**
IS	0.010	0.128	0.124	0.163	**0.185**	**0.164**
CC	**0.460**	0.151	0.040	0.180	0.014	**0.618**
CCS	0.068	0.018	0.109	**0.246**	**0.163**	0.138
MGN	**0.787**	0.717	0.724	0.694	0.693	**0.885**
CN	0.116	0.018	0.008	0.155	**0.360**	**0.852**
LC	0.633	0.719	0.685	0.801	**0.829**	**0.919**
ICQ	0.436	0.736	0.764	**0.783**	0.749	**0.788**
CA	0.661	0.540	0.616	0.561	**0.679**	**0.860**
SSR	0.833	0.816	0.784	0.834	**0.865**	**0.899**

5. CONCLUSION

In conclusion, we propose an effective new BIQA method based on visuo-spatial series statistics. The proposed method simulates visual system to construct imitation visuo-spatial series based on the relative position of pixels, and use statistical features of visuo-spatial series to predict image quality. There are two kinds of imitation visuo-spatial series: the classical MSCN coefficient and a new distribution called NCM. In view of this hypothesis, NCM can be regarded as a significant extension and enhance to MSCN in some way. Convincing experiments show that VSS yields better performance and generalization capability in quality prediction.

6. REFERENCES

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 12, pp. 600–612, Apr. 2004.

[2] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: a feature similarity index for image quality as-sessment," *IEEE Trans. Image Process.*, vol. 20, pp. 2378–2385, Aug. 2011.

[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, pp. 3350–3364, Dec. 2011.

[4] M.A. 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, pp. 3339–3352, Aug. 2012.

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

[6] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, pp. 209–212, Mar. 2013.

[7] L. Zhang, L. Zhang, and A.C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, pp. 2579–2591, Aug. 2015.

[8] F.B. Parmentier, P. Andrs, G. Elford, and D.M. Jones, "Organization of visuo-spatial serial memory: interaction of temporal order with spatial and temporal grouping," *Psychological Research*, vol. 70, pp. 200–217, Apr. 2006.

[9] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, pp. 207–210, Feb. 2014.

[10] P. Andrs F.B. Parmentier and, "The impact of path crossing on visuo-spatial serial memory: Encoding or rehearsal effect?," *Quarterly Journal of Experimental Psychology*, vol. 59, pp. 1867–1874, Jun. 2006.

[11] S. Gabarda and G. Cristbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A.*, vol. 24, pp. B42–B51, Dec. 2007.

[12] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 2," *http://live.ece.utexas.edu/research/quality*.

[13] E.C. Larson and D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 011006, 2010.

[14] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.C.J. Kuo, "Color image database tid2013: peculiarities and preliminary results," in *European Workshop on Visual Information Processing*, 2013, pp. 106–111.

[15] D. Jayaraman, A.K. Moorthy, and A.C. Bovik, "Objective image quality assessment of multiply distorted images," in *Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 1693–1697.

[16] W. Sun, F. Zhou, and Q. Liao, "Mdid: a multiply distorted image database for image quality assessment," *Pattern Recognition*, vol. 61, pp. 153–168, Jul. 2016.