An Opinion-unaware Blind Quality Assessment Algorithm for Multiply Distorted Images

Tongle Wang and Junchen Deng
Department of Computer Science and Technology, East China Normal University,
No. 3663, Zhongshan North Road, Putuo District, Shanghai, China

ABSTRACT

The blind image quality assessment algorithms produced every year are mostly “opinion-aware” (OA). It means that they require large numbers of subjective quality scores for regression model training. Subjective quality scores are not easily available, so people are eager to design an opinion-unaware (OU) algorithm which has free subjective quality scores. Besides, the OU algorithm has greater generalization capability than the OA algorithm. Therefore, we propose an OU algorithm based on a visual codebook for multiply distorted image quality assessment. Extensive experiments conducted on the three databases demonstrate that the proposed method is superior to the existing five OU methods in terms of the coherence with the human subjective rating. The MATLAB code is available at https://tonglewang.github.io.

Keywords: Image quality assessment, human visual system, multiply distorted, blind/no-reference

1. INTRODUCTION

Image quality assessment (IQA) is used to measure the quality degradation caused by multimedia acquisition, compression, and transmission systems. It plays an important role in many image processing tasks such as image denoising, restoration, and enhancement. Generally, the IQA method can be categorized into full-reference, reduced-reference, and no/blind-reference according to the availability of the reference image. Since the blind-reference algorithm does not depend on the reference image, it attracts widespread attention in recent years than the former two. In particular, the opinion-unaware blind image quality assessment method (BIQA) also eliminates the time-consuming and laborious subjective quality scoring process, so it is considered as the most challenging and promising IQA method.

To date, there have been a few methods devote to opinion-unaware BIQA. The following are the state-of-the-art ones among them. Mittal et al.[1] presented the first opinion-unaware BIQA framework using probabilistic latent semantic analysis techniques. Then, Xue et al.[2] proposed an algorithm called QAC, which employed quality-aware clustering and codebook to infer image quality. In literature 3, the authors fitted a multivariate Gaussian (MVG) model by local features of natural images. The quality metric of the test image was given by its distance from the MVG model. Zhang et al.[4] more comprehensively expressed the local features used to fit the MVG model based on the literature 3. For example, they extracted the natural scene, gradient, Gabor, and color statistical features.

The above methods have achieved good results on the singly distorted IQA databases, but it is not very well for multiply distorted databases such as MDID2013[5] and MDID2017[6]. The authors[5] implemented an opinion-unaware algorithm that is effective for both singly and multiply distorted images, but its performance still has room for improvement. Thus, this paper proposes an opinion-unaware IQA method for multiply images, which has higher accuracy but lower computational complexity than the opinion-unaware algorithms mentioned above.

The proposed algorithm first constructs a multiply distorted image block set. We extract the multiply distorted-related local binary mode (LBP) features for each block and use the excellent full-reference algorithm to calibrate its proxy quality score. Then, the LBP features are clustered to obtain a codebook, and the proxy quality score of each visual word (cluster center) is weighted by the proxy scores of those blocks belonging to the cluster. Finally, the test image block is assigned to different quality-aware visual words to infer quality scores by linear weighting. The flowchart of the algorithm is shown in Fig.1.

*tongle_wang@yeah.net; phone 18918079120

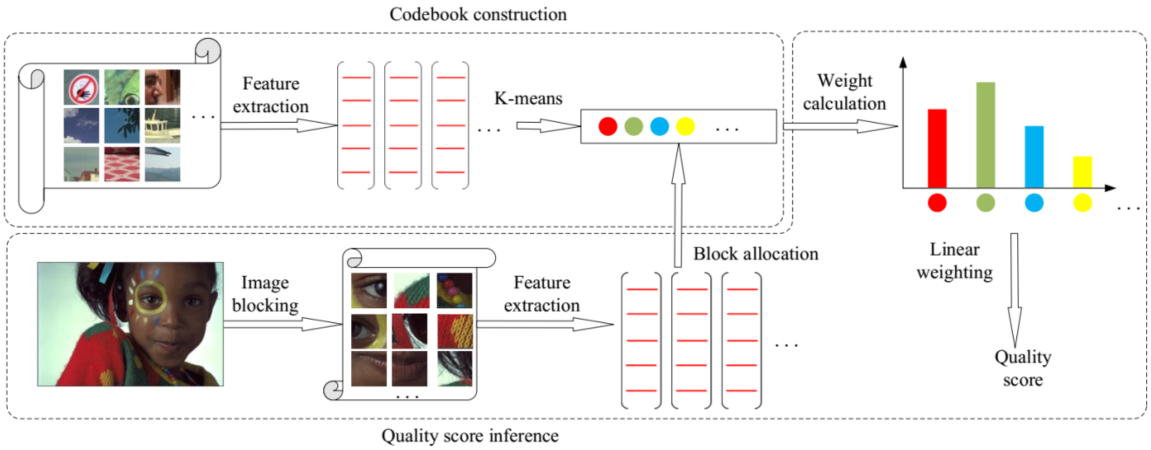

Figure 1. Flowchart of the proposed algorithm.

2. PROPOSED METHOD

2.1 Feature extraction

The mean subtracted contrast normalized coefficient (MSCN)[7] often appears in the IQA preprocessing steps, so we first calculate the MSCN of image I. The subsequent feature extraction is performed on the MSCN.

The human vision system (HVS) is highly sensitive to the image structure damage. LBP can approximate the structural primitives of the early cognitive phase of HVS, so it is very useful for simulating the structural damage of images. For an image I of size $H \times W$, its LBP is defined as follows

$$LBP_{P,R} = \sum_{i=0}^{P-1} E(o_i - o_c)2^i, \tag{1}$$

where o_c represents a coefficient, o_i is its neighbor coefficient, R is the neighborhood radius, and P is the number of neighborhood coefficients of o_c. The binary function $E(o_i - o_c)$ is given by

$$E(o_i - o_c) = \begin{cases} 1, & o_i - o_c \geq 0 \\ 0, & o_i - o_c < 0, \end{cases} \tag{2}$$

In this paper, we use the rotation-invariant uniform LBP ($LBP_{P,R}^{riu2}$) for feature extraction. The $LBP_{P,R}^{riu2}$ adds rotation invariant constraints and uniform constraints to the basic LBP, which is formulated as follows

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{i=0}^{P-1} E(o_i - o_c), & if \ U(LBP_{P,R}) \leq 2 \\ P+1, & else, \end{cases} \tag{3}$$

where "riu2" means that LBP is rotation-invariant and uniform. $U(LBP_{P,R})$ is a uniformity constraint, which is given by

$$U(LBP_{P,R}) = |E(o_0 - o_c) - E(o_{P-1} - o_c)| + \sum_{i=0}^{P-1} |E(o_i - o_c) - E(o_{i-1} - o_c)|, \tag{4}$$

where $|\cdot|$ indicates absolute value operation. $U(LBP_{P,R})$ defines the number of hops for LBP encoding from $0 \rightarrow 1$ or from $1 \rightarrow 0$. The fewer the LBP code hops, the more uniform the encoding. The coding mode with $U(LBP_{P,R}) \leq 2$ is considered as a uniform mode, and the rest is a non-uniform mode.

$LBP_{P,R}^{riu2}$ has a total of $P+1$ uniform modes and one non-uniform mode. Let $P=8$ and $R=1$, so there are nine uniform modes and one non-uniform mode. We use the number of occurrences of each mode in the histogram as the feature vector denoting by $f=[f_0, f_1, \cdots, f_9]$, where the item of f is

$$f_{mode} = \sum_{i=1}^{H}\sum_{j=1}^{W} F(LBP_{8,1}^{riu2}(i,j), mode) \quad for\ mode = 0,1,\cdots,9, \tag{5}$$

where $LBP_{8,1}^{riu2}(i,j)$ is the $LBP_{8,1}^{riu2}$ mode at coordinates (i,j), and the binary $F(LBP_{8,1}^{riu2}(i,j), mode)$ function is defined as

$$F(LBP_{8,1}^{riu2}(i,j), mode) = \begin{cases} 1, & LBP_{8,1}^{riu2}(i,j) = mode \\ 0, & otherwise. \end{cases} \tag{6}$$

The image I needs to be divided into a set of blocks, ie $I=\{p_1, p_2, \cdots, p_N\}$, where N is block number. We extract feature vector f for each block p_i on three scales, and the halved scale is obtained via $1/2$ down sampling. Hence the feature vector of p_i is represented by a 30-dimensional vector $f(p_i) = [f^1, f^2, f^3]$, where superscript is scale index.

2.2 Codebook construction

To construct the codebook, we collect 300 images from the Berkely[8] database and then divide them into 96×96 blocks. Each block is corrupted by seven different types of distortion, which are Blur, Jpeg, Noise, Blur + Jpeg, Blur + Noise, Jpeg + Noise, and Blur + Jpeg + Noise. Blur is obtained by convolving block using a $3\sigma_G$ Gaussian window, where $\sigma_G = 3.2, 3.9, 4.6$ indicate different distortion levels. Jpeg is generated by setting different quality parameters Q_J of MATLAB imwrite function, where $Q_J = 12, 18, 27$ represents different compression levels. Noise is created by adding Gaussian white noise, and the noise level is $\sigma_N = 0.002, 0.008, 0.032$. Blur + Jpeg, Blur + Noise, and Jpeg + Noise are respectively injected with the above two corresponding single distortions. Blur + Jpeg + Noise is successively corrupted by three types of distortion.

There are different methods for calibrating the proxy quality score for the synthesized distortion blocks. We mimic the literature[2] using a full-reference algorithm for calibration. After comparing the performance of two excellent full-reference algorithms (FSIM[9] and VIF[10]) on the three multiply distorted databases (MDID2013[5], MDID2017[6], and MLIVE[8]), we decide to adopt the VIF algorithm. Therefore, the proxy quality score of the distortion block p_i is finally denoted by $VIF(p_i)$.

The codebook construction needs to cluster distortion blocks, and K-means is used for this purpose. The K-means clustering is formulated as follows

$$min \sum_{k=1}^{K}\sum_{p_i\in\Omega_k} \left\| f(p_i) - c_k \right\|_F^2, \tag{7}$$

where c_k is the kth cluster center, Ω_k is a set of those image blocks p_i which are closest to c_k, and "F" is the notation of Frobenius norm.

Solving Equation (7) yields K cluster centers, and each center corresponds to a visual word. The proxy quality score for each visual word is given by

$$sc_k = \sum \frac{d_{i,k}}{\sum_{p_i\in\Omega_k} d_{i,k}} VIF(p_i) \quad for\ k = 1,2,\cdots,K, \tag{8}$$

where $d_{i,k}$ represents the Euclidean distance of p_i from its nearest visual word, and $VIF(p_i)$ is the proxy quality score calibrated for the distortion block p_i.

After block collection, calibration, and clustering, we get a visual codebook $C=\{c_1,c_2,\cdots,c_K\}$ containing K visual words. This codebook will be used for subsequent quality score inference.

2.3 Quality score inference

Given image $I=\{p_1,p_2,\cdots,p_N\}$ and codebook $C=\{c_1,c_2,\cdots,c_K\}$, we query the nearest top T blocks for each visual word c_k in the I. Suppose $d_{i,k}^{(t)}$ is the distance of c_k from its nearest top tth $(1\le t\le T)$ block, and the exponential attenuation of $d_{i,k}^{(t)}$ is expressed as $\tilde{d}_{i,k}^{(t)}=exp(-\gamma d_{i,k}^{(t)})\Big/\sum_{t=1}^{T}exp(-\gamma d_{i,k}^{(t)})$, where γ is the decay rate. In the implementation, we set $T=5$ and $\gamma=0.05$.

When the attenuation distance of the visual word from its nearest top T block is available, its distance from the entire image I is given by $d_{I,k}=\sum_{i=1}^{N}\sum_{t=1}^{T}g(p_i,c_k)\tilde{d}_{i,k}^{(t)}$, where $g(p_i,c_k)$=1 means that p_i is among the nearest T blocks from c_k, otherwise $g(p_i,c_k)$=0.

The image quality score is contributed by the visual word proxy score. Since the proxy score sc_k of c_k is known, only the contribution weight of c_k is needed. The weight is given by

$$ws_k=\frac{d_{I,k}}{\sum_{k=1}^{K}d_{I,k}}. \tag{9}$$

Therefore, the quality score Q(I) of image I is calculated by

$$Q(I)=\sum_{k=1}^{K}ws_k*sc_k. \tag{10}$$

3. EXPERIMENTS AND DISCUSSIONS

The experiments are conducted on the three multiply distorted databases including MDID2013[5], MDID2017[6], and MLIVE[8]. To measure the consistency of objective quality and subjective quality evaluation, we use three common IQA indicators such as SROCC (Spearman Rank Order Correlation Coefficient), PLCC (Pearson Linear Correlation Coefficient), and RMSE (Root Mean Square Error). We report the median of the three indicators when the algorithm is run 1000 times on the database.

In general, the codebook size has a significant impact on the performance of the codebook based algorithm. To make our method work well, we choose the optimal K value before the experiment. It can be seen from Fig.2 that SROCC and PLCC are not consistent with the change of K, so it is difficult to select an optimal K value. To maximize the performance of our algorithm on as many databases as possible, the K is set to 500.

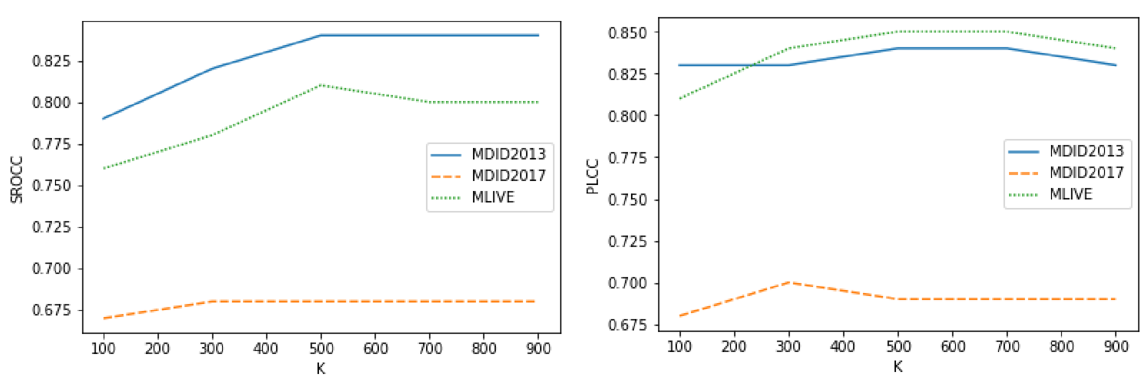

Figure 2. Performance on three databases varies with K value.

We select the state-of-the-art OU algorithms including NIQE[3], ILNIQE[4], QAC[2], BIQES[11], and SISBLITM[5] to compare with our algorithm. The results of the comparison are presented in Table 1.

Table 1. The performance comparison of the proposed algorithm and the other five IQAs on the three multiply distorted databases. Italic bolding indicates the row optimal value.

		NIQE	**ILNIQE**	**QAC**	**BIQES**	**SISBLIM**	**Our**
MDID2013[5]	SROCC	0.641	0.708	0.098	0.319	0.814	***0.838***
	PLCC	0.644	0.709	0.157	0.451	0.810	***0.843***
	RMSE	0.041	0.040	0.050	0.045	0.030	***0.027***
MDID2017[6]	SROCC	0.649	0.670	0.429	0.646	0.655	***0.683***
	PLCC	0.670	0.690	0.500	0.672	0.631	***0.693***
	RMSE	1.635	1.572	1.899	1.617	1.709	***1.562***
MLIVE[8]	SROCC	0.771	0.877	0.245	0.683	***0.878***	0.806
	PLCC	0.839	***0.898***	0.338	0.767	0.895	0.852
	RMSE	10.318	***8.333***	17.555	12.094	8.439	9.866
Time		0.453	4.296	3.630	0.252	0.569	***0.191***

From Table 1, we can see that our algorithm outperforms the five competitors on the MDID2013 and MDID2017 in terms of SROCC, PLCC, and RMSE. Specifically, SROCC increases by 2.4% and 2.8% compared to the second-ranked SISBLIM on MDID2013 and MDID2017, respectively. Our algorithm does not perform well on the MLIVE, because MLIVE's single distorted image accounts for 40% (180/450), while the single distorted block in our codebook accounts for only 14.3%. Poor single distorted samples result in poor performance.

Also, we compare the computational complexity. We take the elapsed time from the feature extraction to the quality score acquisition for a 512×512 image as a computational complexity metric. Experiments are conducted on the Dell Inspiron 7000 notebook with Intel Core i5-8250U CPU @1.8 GHz and the software platform of MATLAB2015b. We observe that our algorithm is faster than other competitive IQAs if without optimizing codes. For example, it is more than twice as fast as the QAC and more than 20 times faster than the slowest ILNIQE.

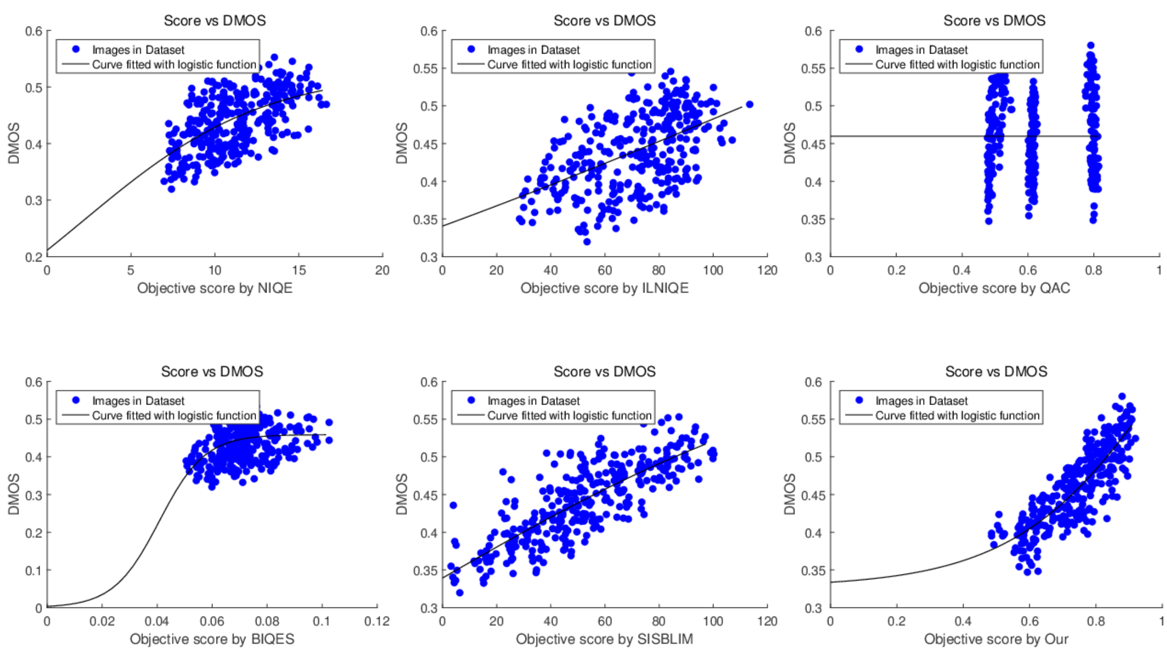

Figure 3. Scatter plots of objective scores predicted by six algorithms versus subjective scores on the MDID2013 database. It should be noted that the x-axis is the predicted value of the algorithm, and the y-axis is the subjective score of the MDID2013.

To further illustrate the superiority of the proposed method, in Fig.3, we draw the scatter plots of subjective DMOS against quality scores predicted by six IQAs on the MDID2013 database. The curve is obtained by the nonlinear fitting function given in literature 12. The closer the curve is to the diagonal and the denser the point distribution is, the more accurate the prediction is. So, it is observable that our method has a higher correlation with human subjective ratings than other evaluators.

4. CONCLUSION

In this paper, we propose a novel multiply distorted image quality assessment algorithm through a visual codebook. The algorithm is "opinion-unaware", so it can work without subjective quality scores. Furthermore, the experiments attached to the paper also prove that the proposed algorithm has higher accuracy and lower computational complexity than the existing approaches.

REFERENCES

[1] Mittal, A., Muralidhar, G. S., Ghosh, J., and Bovik, A. C., "Blind image quality assessment without human training using latent quality factors," IEEE Signal Processing Letters 19(2), 75–78 (2011).

[2] Xue, W., Zhang, L., and Mou, X., "Learning without human scores for blind image quality assessment," in [Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition], 995–1002 (2013).

[3] Mittal, A., Soundararajan, R., and Bovik, A. C., "Making a completely blind image quality analyzer," IEEE Signal Processing Letters 20(3), 209–212 (2012).

[4] Zhang, L., Zhang, L., and Bovik, A. C., "A feature-enriched completely blind image quality evaluator," IEEE Transactions on Image Processing 24(8), 2579–2591 (2015).

[5] Gu, K., Zhai, G., Yang, X., and Zhang, W., "Hybrid no-reference quality metric for singly and multiply distorted images," IEEE Transactions on Broadcasting 60(3), 555–567 (2014).

[6] Sun, W., Zhou, F., and Liao, Q., "Mdid: A multiply distorted image database for image quality assessment," Pattern Recognition 61, 153–168 (2017).

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

[8] Jayaraman, D., Mittal, A., Moorthy, A. K., and Bovik, A. C., "Objective quality assessment of multiply distorted images," in [2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR)], 1693–1697, IEEE (2012).

[9] Zhang, L., Zhang, L., Mou, X., and Zhang, D., "Fsim: A feature similarity index for image quality assessment," IEEE transactions on Image Processing 20(8), 2378–2386 (2011).

[10] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," in [2004 IEEE International Conference on Acoustics, Speech, and Signal Processing], 3, iii–709, IEEE (2004).

[11] Saha, A. and Wu, Q. M. J., "Utilizing image scales towards totally training free blind image quality assessment," IEEE Transactions on Image Processing 24(6), 1879–1892 (2015).

[12] Zhang, Y. and Chandler, D. M., "Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion parameter estimation," IEEE Transactions on Image Processing 27(11), 5433–5448 (2018).