Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment

Tonghan Wang, Lu Zhang, Huizhen Jia, Baosheng Li, Huazhong Shu

HAL Id: hal-01343532

https://univ-rennes.hal.science/hal-01343532v1

Submitted on 7 Oct 2016

Multiscale contrast similarity deviation: an effective and efficient index for perceptual image quality assessment

Tonghan Wang, [a] Lu Zhang,[b] Huizhen Jia,[c] Baosheng Li, [a,d] and Huazhong Shu[a]

[a]Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu, P.R. China

[b]IETR lab (UMR CNRS 6164), INSA de Rennes, 20 Avenue des Buttes de Coesmes, CS 70839F – 35708 Rennes Cedex 7, France

[c]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, P.R. China

[d]Shandong Cancer Hospital, Ji-Nan, Shandong, P.R. China

Abstract. Perceptual image quality assessment (IQA) uses a computational model to assess the image quality in a fashion consistent with human opinions. A good IQA model should consider both the effectiveness and efficiency. To meet this need, a new model called multiscale contrast similarity deviation (MCSD) is developed in this paper. Contrast is a distinctive visual attribute closely related to the quality of an image. To further explore the contrast features, we resort to the multiscale representation. Although the contrast and the multiscale representation have already been used by other IQA indices, few have reached the goals of effectiveness and efficiency simultaneously. We compared our method with other state-of-the-art methods using six well-known databases. The experimental results showed that the proposed method yielded the best performance in terms of correlation with human judgments. Furthermore, it is also efficient when compared with other competing IQA models.

The MATLAB source code of MCSD is public available online at http://www.mathworks.com/matlabcentral/fileexchange/56604-multiscale-contrast-similarity-deviation/content/mcsd.m

*Tonghan Wang, E-mail: thwang_seu@163.com

1 Introduction

Image quality assessment occupies a very important position in numerous fields and applications, such as image acquisition, compression, transmission and restoration. Since human beings are the ultimate receivers of the visual stimulus, it is essential to develop a perceptual model to closely correlate with the human visual system (HVS).

Objective quality assessment methods can be classified into three types [1]: (1) full-reference(FR), where an ideal "reference" image is available for comparison; (2) reduced-reference (RR), where partial information about the reference image is available; and (3) no-reference (NR), where the reference image is not accessible. This paper focuses on the FR methods. In the past decades, great efforts and huge advances have been made in FR methods. Here we briefly review some representative ones. More comprehensive surveys on FR-IQA metrics can be found in [16] , [17] and [18]. The traditional metrics such as the peak signal-to-noise ratio (PSNR) and the

mean squared error (MSE) did not correlate well with human opinions [2]. The later developed FR methods could be generally classified into three types of approaches: the HVS model based ones, the information theoretic ones and the structural ones.

The noise quality measure index (NQM) [7], the visual signal-to-noise ratio index (VSNR) [8] and most apparent distortion (MAD) [12] are the three representatives HVS based FR methods. The NQM and the VSNR quantified the effects of different visual signals (e.g. the luminance, the contrast, the frequency content, the interaction between them) on the HVS. The MAD was proposed by Larson and Chandler based on the hypothesis that the HVS used different strategies for high quality and low quality images. However this kind of FR methods is usually not computationally efficient.

The information theoretic approaches include the visual information fidelity (VIF)[9] and the information fidelity criteria (IFC) [10]. The VIF took the FR IQA problem as an information fidelity problem and chose the amount of information shared by the reference image and the distorted one as the similarity, which was an extented version of the IFC.

The structural approaches are based on the assumption that the HVS is highly adapted for extracting structural information from the visual scene. As a milestone in the development of IQA models, the structural similarity (SSIM) [3] surpassed the previous ones since it showed a better correlation with the human perception. Then a lot of SSIM-based metrics have been proposed in the literature [4-6]. In [11], Gao et al. presented a content-based metric (CBM), which divided the structural information into edges, textures and flat regions in accordance with the content and then pooled the three parts with different weights to obtain the final image quality. Based on the observation that the visual information in an image is often redundant and the HVS understands an image mainly based on its low-level features, Zhang et al. proposed the feature-similarity (FSIM) index [14] which employed two features (the phase congruency and the gradient magnitude) to compute the local similarity map. Considering the gradients' sensitivity to structure and contrast changes, Liu et al. proposed GSM [13] based on a gradient similarity. Unlike the SSIM's average pooling, the CBM, FSIM and the GSM adopted a weighting strategy for the pooling. Note that the weighting pooling may gain more IQA accuracy than those with average pooling to some degree, but it may increase the computational complexity. In addition, this pooling could make the predicted quality scores nonlinear to human opinions [15]. Based on these considerations, Xue et al. [15] proposed a gradient based model, based on the observation that the image gradient can effectively capture image local structures to which the HVS is highly sensitive. The gradient magnitude similarity deviation (GMSD) index [15] firstly the generated image gradient magnitude maps of the reference image and the distorted one, and then computed the similarity map of them, finally took the standard deviation of similarity map as the overall image quality score.

Generally, the effectiveness, namely, high correlation with the human subjective score, is the prerequisite of a good IQA model. The efficiency (low computation cost), however, is the second most important criterion of a good IQA model. Thus the

effectiveness and efficiency are two ultimate goals for the design of IQA models, but unfortunately it is hard to reach these two goals simultaneously. Among the above-mentioned methods, the GMSD [15] had a big success in terms of the two goals, but its performances were a little bit low for certain distortion types (such as the contrast distortion). In this paper, we attempt to make another effort to fill this need and to overcome the problem of the GMSD by proposing an effective and efficient FR-IQA model called multiscale contrast similarity deviation (MCSD). The MCSD is also a structural approach.

The rest of this paper is organized as follows. In Section 2, we present the proposed model in detail. Section 3 shows and discusses the results. We conclude this work in Section 4.

2 Multiscale contrast similarity deviation (MCSD)

Contrast is a good attribute for characterizing the quality of an image [19]. Proper contrast change may even improve the perceptual quality of images. In fact, we can define "high quality" as appropriate contrast and little distortion. The contrast has been widely used in the area of image enhancement [20-25]. Contrast has previously been utilized in SSIM [3], where it was used as one of the three features - luminance, contrast and structure. Here we use the contrast feature alone to design our IQA model. Furthermore, the contrast is sensitive to the spatiotemporal frequency and viewing distance [26], [27], which are related to the multi-scale representation to some extent. In fact, the multi-scale method is a convenient way to incorporate image details at different resolutions. Perceptibility of image details depends on viewing distance and sampling density of an image. Furthermore, a natural image might have objects and structures that are relevant at different scales, but the human eye is readily able to identify and process the information presented by it [28]. Thus, processing an image at various scales adds flexibility to the processing technique, and image scales play a very important role in IQA [4], [29]. In [4], five scales, namely, the original image scale plus the other four reduced resolution (each down sampled by a factor of 2), were utilized to design the multi-scale SSIM. Very recently, by using proper image scales, the authors of [29] designed a totally training free NR-IQA. For these reasons, we combined the contrast with the multi-scale representation to design our model.

2.1 Structure diagram of the MCSD

The structure diagram of the MCSD is shown in Fig.1. The design choices (all the parameters' settings) for MCSD will be introduced in section 3.2.
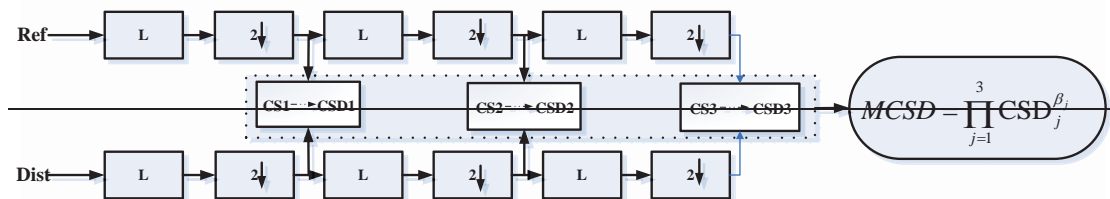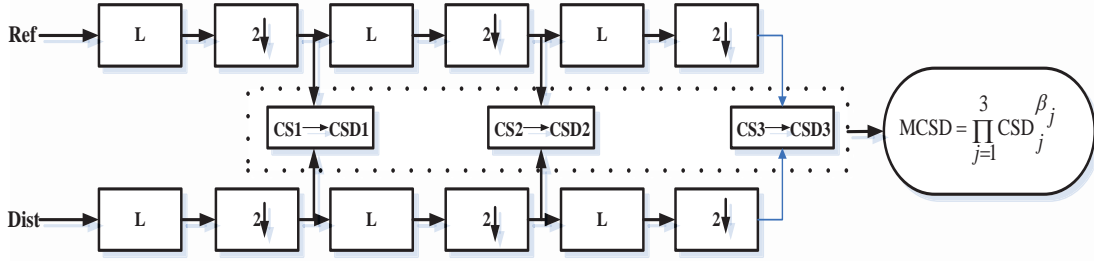
Fig.1. Structure diagram of the MCSD. Ref indicates a reference image and Dist indicates the corresponding distorted one, L: low-pass filtering; 2↓: down sampling by 2; CS1,CS2 and CS3 mean contrast similarity map1, contrast similarity map2 and contrast similarity map3, respectively; CSD1, CSD2 and CSD3 mean the standard deviation of CS1, CS2 and CS3, respectively. The general idea of the MCSD is to decompose the reference and the distorted images into three scales, then calculate their contrast similarity deviation for each scale, and get the final note via a pooling of the CSD values on the three scales.

The MCSD explores the contrast features by resorting to the multi-scale representation. The reason is that multiscale method incorporates image details at different resolutions, and contrast is relevant to the viewing distance. The contrast combined with multiscale representation were widely used in the literature of image enhancement [32-34,38,39] and image fusion [35-37], where the multiscale representation was implemented by pyramid decomposition [34], [36], wavelet decompositions [32, 34, 36], scale-space approach [33] and homomorphic transform [39]. In view of efficiency, the multiscale representation in the proposed model, however, was operated by iteratively applying a low-pass filter and downsampling the filtered image by a factor of 2, as did in MS-SSIM [4].

Considering the computational cost, we do not use the original image scale but three reduced resolutions (each down sampled by a factor of 2). For the reference image and distorted image at each scale, we calculate their contrast similarity deviation (CSD), sort of contrast similarity map between them (the derivation of the CSD will be detailed in the next subsection). Then the CSDs at three scales are pooled to have the final score MCSD using the following equation:

$$MCSD = \prod_{j=1}^{\text{nscales}} \text{CSD}_j^{\beta_j} \tag{1}$$

where $\prod$ means product, nscales is the total scale we use, j is the jth scale, β_j is the corresponding weight of the jth scale and $\sum_{j=1}^{\text{nscales}} \beta_j = 1$.

2.2 Contrast similarity deviation

There are different definitions for contrast [30], [31], such as Weber contrast, Michelson contrast, RMS (root-mean-square) contrast, etc. Weber contrast is mainly for characters, Michelson contrast is mainly for depicting the gratings, and RMS

contrast for natural stimuli [30]. Thus we chose the RMS contrast in this paper, as adopted by the SSIM [3]. Similar to [3], we use the standard deviation (the square root of variance) as an estimate of the signal contrast:

$$c = \left[\frac{1}{(M-1)(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_{ij} - \bar{I})^2 \right]^{1/2} \tag{2}$$

where intensities I_{ij} are the i-th j-th element of the 2D image of size M by N, and $\bar{I}$ is the average intensity of all pixel values in the image.

Whereas, the contrast is computed locally in the paper and it can be written in the following form.

$$c = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} (I_{ij} - \mu_I)^2} \tag{3}$$

and

$$\mu_I = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} I_{ij} \tag{4}$$

where w_{ij} is local window, namely, a circular-symmetric Gaussian weighting function

and then rescaled to unit volume, μ_I is the local mean of image I. The local window size plays an important role in the performance of the IQA model as it will be shown in section 3.2.

Contrast maps are computed locally for reference image and its distorted one using Eq. (3). We denote by CM_r and CM_d the contrast maps for the reference image and the distorted one, respectively. Then contrast maps similarity contrast similarity map(CS , which is also a map) for the two images being compared is defined as:

$$CS = \frac{(2CM_r \circ CM_d + a)}{(CMr.^2 + CMr.^2 + a)} \tag{5}$$

where "∘" indicates element-wise multiplication of two matrices, ".^{2}" indicates element-wise square of the matrix, and a is a positive constant to avoid division by zero. Note that CM_r and CM_d are computed from a local patch of reference image r and the distorted one d, respectively. For grayscale image, contrast is the difference in luminance that makes an object distinguishable. To reflect the contrast change as accurately as possible, we resort to the standard deviation as the pooling strategy since it can reflect the range of image local contrast degradation. This is in line with the contrast since both of them indicate "change". Thus the contrast similarity deviation (CSD) is defined as follows:

$$CSD = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (CS(i,j) - CSM)^2} \tag{6}$$

where CSM (contrast similarity mean) is the mean pooling of contrast similarity map:

$$CSM = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}CS(i,j) \qquad (7)$$

2.3 Algorithm description of the MCSD

To clearly state the workflow of the MCSD, we give its algorithmic description in Algorithm 1.

Algorithm 1
Initialization: Given the reference image Ref and its distorted image Dist, compute the contrast maps of Ref and Dist, respectively, by using Eq. (3)

Repeat:

 for j=1:nscales do

 1) Apply a 2×2 average filter on both Ref and Dist, and then down-sample the outputs by a factor of 2.

 2) Compute the contrast maps of Ref and Dist, respectively, by using Eq. (3);

 3) Compute the quality map of Ref and Dist by Eq. (5);

 4) Adopt the standard deviation as the the pooling strategy to obtain the image quality through Eq. (6);

 end for

 beta=[beta1 beta2 beta3];

 MCSD=CSD(1) ^{beta1} ×CSD(2) ^{beta2} ×CSD(3) ^{beta3}

3   Performance evaluation

3.1 databases and evaluation protocols

To test the performance of the proposed MCSD, we used the six well-known databases: LIVE [40], CSIQ [41], TID2008 [42], TID2013 [43], MICT [44] and IVC [45]. The information about these databases is summarized in Table 1.

- 1) LIVE database II [40]: It consists of 29 reference images and their degraded version with five types of distortions, i.e., JPEG 2000 compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blurring (Gblur), and fast fading (FF). The human subject score (DMOS) associated with distorted images are provided.
- 2) CSIQ IQA database [41]: It consists of 30 reference images and their degraded versions with six different types of distortions at four to five different levels: JPEG2000, JPEG, Additive Gaussian white noise (AGWN), Gaussian Blurring (GB), global contrast decrements (CTD), and additive pink Gaussian noise (APGN). DMOS associated with distorted images is provided and is in the range [0, 1], where lower DMOS indicates higher quality.

6

- 3) The TID2008 database [42]: It contains 1700 test images, which are derived from 25 reference images. There are 17 types of distortions for each reference image and four different scales of each type of distortion.
- 4) The TID2013 database [43]: TID2013 is an acronym for Tampere image database 2013. It is a newly arrived with very huge number of images are present among the IQA databases presented among the IQA databases. It contains 25 original images, 3000 distorted images and having 24 different types of distortions with 5 levels of variations in each distortion.
- 5) MICT [44]: There are 98 images of 768×512 pixels for both JPEG and JP2K group. Six quality scales are selected for each distortion type.
- 6) IVC [45]: It consists of 10 original images and 235 distorted images generated from four different processing: JPEG, JPEG2000, LAR coding and blurring.

The relevant information about these databases is summarized in Table 1.

Table 1 Information about the databases.

database	reference image	distorted image	distortion types	Observer	Image type	Image format	MOS/DMOS range
LIVE	29	779	5	161	color	bmp	[0,100] (DMOS)
CSIQ	30	866	6	35	color	png	[0,1] (DMOS)
TID2008	25	1700	17	838	color	bmp	[0,9] (MOS)
TID2013	25	3000	24	971	color	bmp	[0,9] (MOS)
Toyama-MICT	14	168	2	14	color	bmp	[1,5] (MOS)
IVC	10	185	4	15	grayscale	bmp	[1,5] (MOS)

We calculated four commonly used performance indices, i.e. the Spearman's rank ordered correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC) which measure the prediction monotonicity; the Pearson's linear correlation coefficient (PLCC) which is related to the prediction linearity (considered as the measure of prediction accuracy); and the root mean square error (RMSE) which evaluates the prediction consistency.

Let the vector s denote subjective scores, the vector x denote the original IQA scores predicted by an objective metric, d_i denote the difference between the ranks of each pair in x and s, both s and x have n elements, then:

$$SROCC = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \qquad (8)$$

$$KROCC = \frac{n_c - n_d}{0.5n(n-1)} \qquad (9)$$

where n_c is the number of concordant pairs in the data set and n_d is the number of discordant pairs in the data set. Let $(x_1, s_1), (x_2, s_2),...(x_n, s_n)$ be a set of joint observations from two random variables (IQA scores X and subjective scores S, respectively). For any pair of observations (x_i, s_i) and (x_j, s_j), if both $x_i > x_j$ and $s_i > s_j$ or if $x_i < x_j$ and $s_i < s_j$, then they are concordant; else if $x_i = x_j$ or $s_i = s_j$, then they are neither concordant nor discordant; else they are discordant.

To compute the PLCC and the RMSE, we used a logistic regression function to reduce the nonlinearity of predicted scores [46].

$$p_i = \alpha_1 (\frac{1}{2} - \frac{1}{1 + e^{\alpha_2(x_i - \alpha_3)}}) + \alpha_4 x_i + \alpha_5 \qquad i = 1,2,...,n \qquad (10)$$

where α_i (i=1, 2, …, 5) are parameters to be fitted and p_i is the element in the vector p that denotes the IQA score after regression. Then we have:

$$PLCC = \frac{\sum_{i=1}^{n}(p_i - \overline{p})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{n}(p_i - \overline{p})^2}\sqrt{\sum_{i=1}^{n}(s_i - \overline{s})^2}} \qquad (11)$$

where $\overline{p}$ and $\overline{s}$ are the mean of p and s, respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_i - p_i)^2} \qquad (12)$$

A value being close to 1 for SROCC, KROCC and PLCC indicates a good performance for quality prediction. Whereas, for RMSE, the smaller the value is, the better prediction consistency it yields.

3.2 Implementation of MCSD

Note that several parameters are involved in MCSD design. We thus tested their impact on the overall performance of the proposed model in this section. We draw the plot of the SROCC vs. the local window size and the SROCC vs. the constant a in Eq. (5) in Fig. 2 and Fig.3, respectively. We investigated their impacts on the performance on CSIQ database. We can see from Fig.2 that the performance is on a downward trend when the local window size increased. Thus we set it to 2. In other words, we use a 2×2 circular-symmetric Gaussian weighting function with its standard deviation of 1.5 and then rescaled to unit volume.
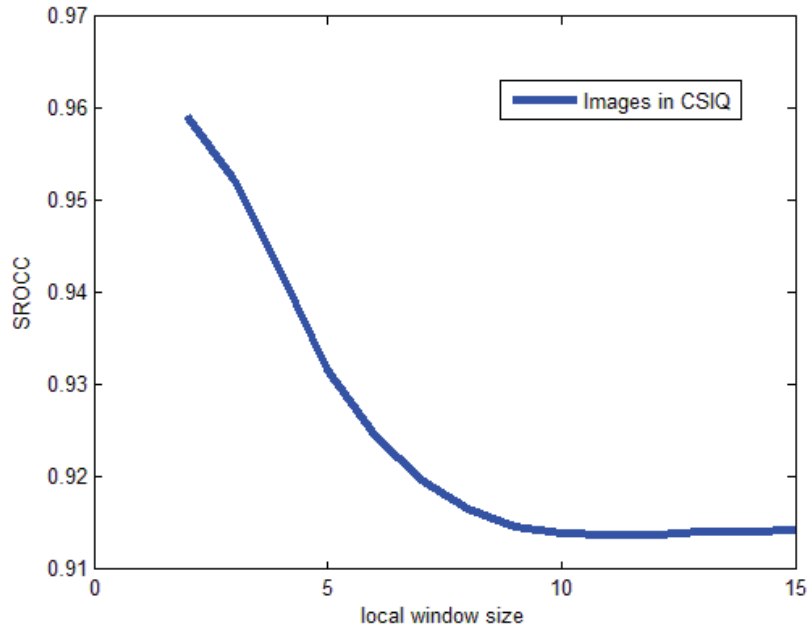
Fig. 2. SROCC with respect to the local window size

We investigated the constant a from 1 to 130 with a step of 10. From Fig.3 we can see that the performance reaches its maximum value at the point of 40 on CSIQ database.
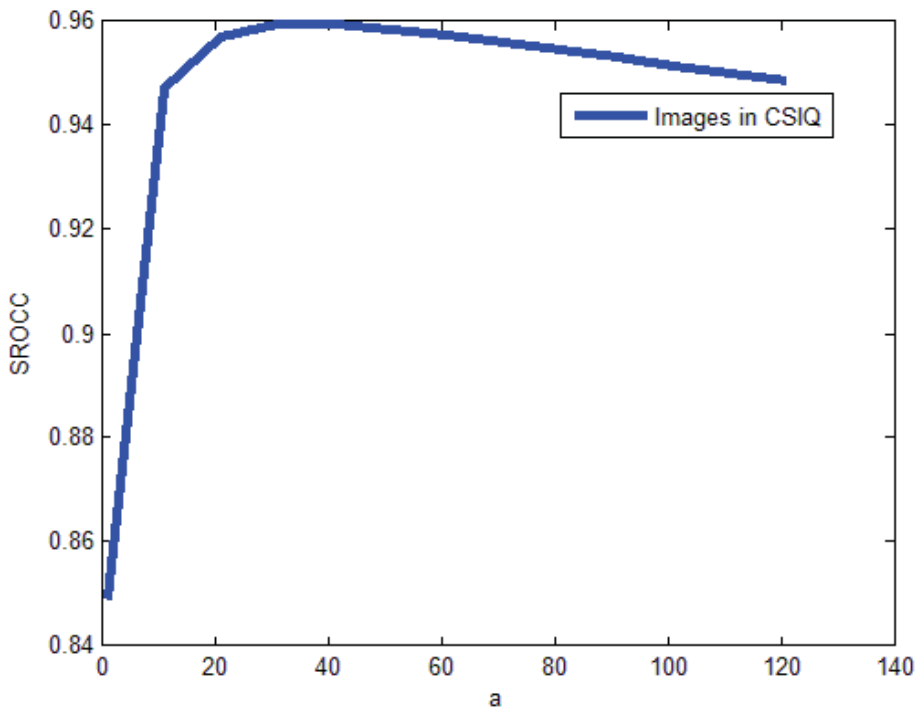
Fig.3.SROCC with respect to constant a

In our experiment, we set it to 45 considering its overall performance on all the benchmark databases. We set nscales=3, $\beta_1 = 0.65$, $\beta_2 = 0.1$ and $\beta_3 = 0.25$ based on experiments conducted in the same way as the above procedure.

9

To further investigate the influences of different pooling strategies on the performance, the simple average (MEAN) pooling strategy is compared with the standard deviation pooling strategy. We conducted an experiment on three six large databases-LIVE [40], CSIQ [41], TID2008 [42], TID2013 [43], MICT [44] and IVC [45]. The Spearman's rank ordered correlation coefficient (SROCC), defined in section 3, is used as the evaluation criterion here. The result is shown in Fig.4, from which we can see that the standard deviation pooling for our model is better than the mean pooling on all of the three six databases.
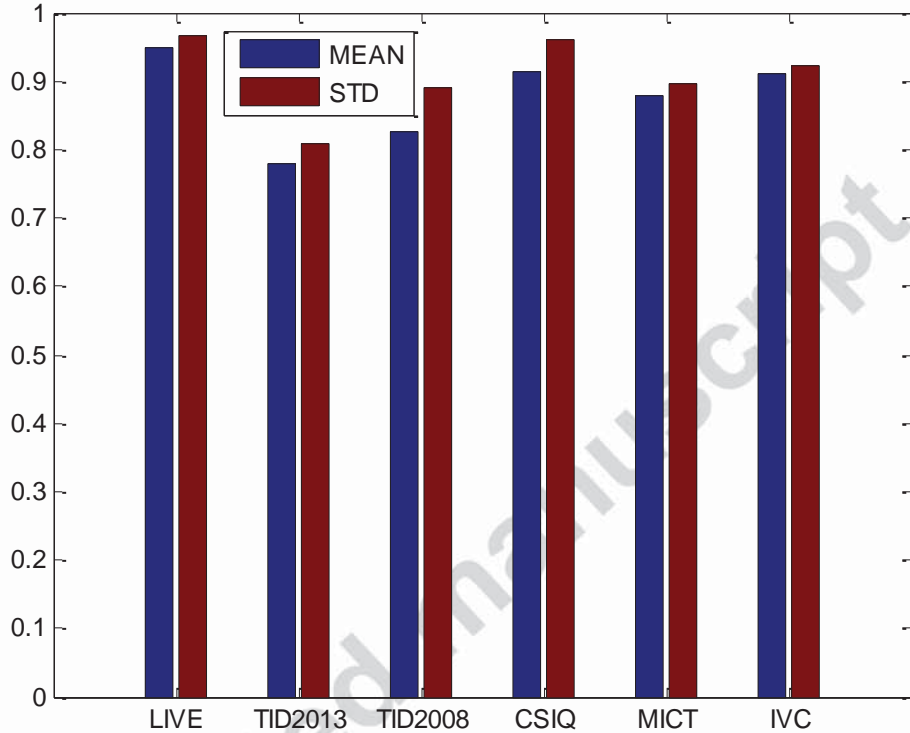
Fig.4. The SROCC of using mean and STD to evaluate on LIVE, TID2008, TID2013, CSIQ, MICT and IVC databases. STD denotes the standard deviation

3.3 Performance comparisons

We compared the proposed method MCSD to seven state-of-the-art and representative FR-IQA models, including SSIM [3], MS-SSIM [4] IW-SSIM [6] VIF [9], MAD [12], FSIM [14], GMSD [15]. All the IQA models were designed for grayscale image except the FSIM which could handle both color and grayscale image. Note that GMSD is the recently proposed top-performance IQA model. In Table 2, the best three ones are highlighted in boldface for each distortion for SROCC, KROCC, PLCC and RMSE. The source codes of all the other metrics were obtained from the original authors.

Table 2 Performances of the proposed metric and the other eight competing FR-IQA metrics on six benchmark databases. The top three metrics for each criterion are highlighted in boldface

		SSIM	MS-SSIM	IW-SSIM	VIF	MAD	FSIM	GMSD	MCSD
TID2013	SROCC	0.7417	0.7859	0.7779	0.6769	0.7807	0.8015	0.8044	0.8089
	KROCC	0.5588	0.6047	0.5977	0.5147	0.6035	0.6289	0.6339	0.6385
	PLCC	0.7895	0.8329	0.8319	0.7720	0.8267	0.8589	0.8553	0.8565
	RMSE	0.7608	0.6861	0.6880	0.7880	0.6975	0.6349	0.6423	0.6399
TID2008	SROCC	0.7749	0.8542	0.8559	0.7491	0.8340	0.8804	0.8907	0.8911
	KROCC	0.5768	0.6568	0.6636	0.5861	0.6445	0.6945	0.7092	0.7133
	PLCC	0.7732	0.8451	0.8579	0.8084	0.8308	0.8738	0.8788	0.8844
	RMSE	0.8511	0.7173	0.6895	0.7899	0.7468	0.6527	0.6404	0.6263
CSIQ	SROCC	0.8756	0.9133	0.9213	0.9195	0.9466	0.9310	0.9570	0.9592
	KROCC	0.6907	0.7393	0.7529	0.7537	0.7970	0.7690	0.8122	0.8171
	PLCC	0.8613	0.8991	0.9144	0.9277	0.9502	0.9192	0.9541	0.9560
	RMSE	0.1334	0.1149	0.1063	0.0980	0.0818	0.1034	0.0786	0.0770
LIVE	SROCC	0.9479	0.9513	0.9567	0.9636	0.9672	0.9634	0.9603	0.9668
	KROCC	0.7963	0.8045	0.8175	0.8282	0.8427	0.8337	0.8268	0.8407
	PLCC	0.9449	0.9489	0.9522	0.9604	0.9688	0.9597	0.9603	0.9675
	RMSE	8.9455	8.6188	8.3473	7.6137	6.7672	7.6780	7.6214	6.9079
IVC	SROCC	0.9018	0.8847	0.9125	0.8966	0.9147	0.9262	0.9146	0.9231
	KROCC	0.7223	0.7012	0.7339	0.7165	0.7406	0.7564	0.7371	0.7522
	PLCC	0.9119	0.8934	0.9231	0.9028	0.9210	0.9376	0.9235	0.9292
	RMSE	0.4999	0.5474	0.4686	0.5239	0.4746	0.4236	0.4675	0.4503
MICT	SROCC	0.8794	0.8874	0.9202	0.9077	0.9362	0.9050	0.8528	0.8973
	KROCC	0.6939	0.7029	0.7537	0.7315	0.7823	0.7280	0.6588	0.7180
	PLCC	0.8887	0.8927	0.9249	0.9138	0.9367	0.9065	0.8582	0.8971
	RMSE	0.5738	0.5639	0.4759	0.5084	0.4381	0.5283	0.6424	0.5531

In Table 3, we show the weighted-average SROCC, KROCC, PLCC, and RMSE results over six databases, where the weight assigned to each database linearly depends on the number of distorted images contained in that dataset. This can provide an overall performance evaluation of the tested IQA indices [6].

Table 3 Overall performance of IQA models over six databases

IQA models	SROCC	KROCC	PLCC	RMSE
SSIM	0.7993	0.6159	0.8186	1.6426
MS-SSIM	0.8442	0.6637	0.8612	1.5358
IW-SSIM	0.8443	0.6677	0.8680	1.4926
VIF	0.7718	0.6112	0.8305	1.4788
MAD	0.8450	0.6750	0.8656	1.3237
FSIM	0.8631	0.6935	0.8856	1.3813
GMSD	0.8684	0.7020	0.8874	1.3781
MCSD	0.8729	0.7092	0.8915	1.2877

The ranking of the weighted-average performances of the evaluated IQA indices based on four different performance metrics, SROCC, KROCC, PLCC, and RMSE, is presented in Table 4.

Table 4 Ranking of overall performance of IQA models

IQA models	SROCC	KROCC	PLCC	RMSE
SSIM	7	7	8	8
MS-SSIM	4	6	6	7
IW-SSIM	6	5	4	6
VIF	8	8	7	5
MAD	5	4	5	2
FSIM	3	3	3	4
GMSD	2	2	2	3
MCSD	1	1	1	1

From Table 2, it can be seen that the proposed MCSD performs consistently well on each individual database. It achieves the best on the four largest databases (TID2013, TID2008, CSIQ and LIVE). On IVC and MICT, even though it is not the best, it performs only slightly worse than the best one. While for other competitors, they either work well on one database but poor on other databases. For example, the MAD performs quite well on CSIQ, LIVE, IVC and MICT, but it works quite poor on the two largest databases, TID2013 and TID 2008. The GMSD performs only slightly worse than the proposed MCSD, but it lags the MCSD on MICT, IVC and LIVE databases a lot. From Tables 3 and 4 we can see that the MCSD has the best overall performance weighted on the six benchmark databases. The relative bad performance of the MCSD on the MICT will be discussed in the next subsection (after Table 6).

3.4 Performance comparisons on individual distortion types

To fully evaluate the IQA models, we also compared their abilities of predicting image quality degradations caused by different types of distortions. We list the results in Table 5 and Table 6. Like the GMSD [15], we also only showed the SROCC scores for space saving purpose. In Table 5 and Table 6, we use boldface font to highlight the top 3 models in each group.

Table 5 SROCC values of IQA model for each type of distortions in TID2013.

AGN, ANC, SCN, MN, HFN, IN, QN, GB, DEN, JPEG, JP2K, JGTE, J2TE, NEPN, Block, MS, CTC, CCS, MGN, CN, LCNI, ICQD, CHA and SSR in TID database are the distortions additive Gaussian noise, additive noise in color components, spatially correlated noise, masked noise, high frequency noise, impulse noise, quantization noise, Gaussian blur, image denoising, JPEG compression, JPEG2000 compression, JPEG transmission errors, JPEG2000 transmission errors, non-eccentricity pattern noise, local block-wise distortions of different intensity, mean shift, contrast change, change of color saturation, multiplicative Gaussian noise, comfort noise, lossy compression of noisy images, image color quantization with dither, chromatic aberrations and sparse sampling and reconstruction, respectively.

		SSIM	MS-SSIM	IW-SSIM	VIF	MAD	FSIM	GMSD	MCSD
TID2013	AGN	0.8671	0.8646	0.8438	0.8994	0.8843	0.8973	0.9462	0.9451
	ANC	0.7726	0.7730	0.7515	0.8299	0.8019	0.8208	0.8684	0.8671
	SCN	0.8515	0.8544	0.8167	0.8835	0.8911	0.8750	0.9350	0.9404
	MN	0.7767	0.8073	0.8020	0.8450	0.7380	0.7944	0.7075	0.7236
	HFN	0.8634	0.8604	0.8553	0.8972	0.8876	0.8984	0.9162	0.9166
	IN	0.7503	0.7629	0.7281	0.8537	0.2769	0.8072	0.7637	0.7769
	QN	0.8657	0.8706	0.8468	0.7854	0.8514	0.8719	0.9049	0.9103
	GB	0.9668	0.9673	0.9701	0.9650	0.9319	0.9551	0.9113	0.9095
	DEN	0.9254	0.9268	0.9152	0.8911	0.9252	0.9302	0.9525	0.9537
	JPEG	0.9200	0.9265	0.9187	0.9192	0.9217	0.9324	0.9507	0.9431
	JP2K	0.9468	0.9504	0.9506	0.9516	0.9511	0.9577	0.9657	0.9626
	JGTE	0.8493	0.8475	0.8388	0.8409	0.8283	0.8464	0.8403	0.7466
	J2TE	0.8828	0.8889	0.8656	0.8761	0.8788	0.8913	0.9136	0.9230
	NEPN	0.7821	0.7968	0.8011	0.7720	0.8315	0.7917	0.8140	0.8309
	Block	0.5720	0.4801	0.3717	0.5306	0.2812	0.5489	0.6625	0.6675
	MS	0.7752	0.7906	0.7833	0.6276	0.6450	0.7531	0.7351	0.5961
	CTC	0.3775	0.4634	0.4593	0.8386	0.1972	0.4686	0.3235	0.3584
	CCS	0.4141	0.4099	0.4196	0.3099	0.0575	0.2748	0.2948	0.2715
	MGN	0.7803	0.7786	0.7728	0.8468	0.8409	0.8469	0.8886	0.8857
	CN	0.8566	0.8528	0.8762	0.8946	0.9064	0.9121	0.9298	0.9333
	LCNI	0.9057	0.9068	0.9037	0.9204	0.9443	0.9466	0.9629	0.9697
	ICQD	0.8542	0.8555	0.8401	0.8414	0.8745	0.8760	0.9102	0.9172
	CHA	0.8775	0.8784	0.8682	0.8848	0.8310	0.8715	0.8530	0.8390
	SSR	0.9461	0.9483	0.9474	0.9353	0.9567	0.9565	0.9683	0.9678

Table 6 SROCCs of IQA models for each type of distortions in TID20008, CSIQ, LIVE and MICT. The abbreviations for distortion types in TID2008 are the same with those in TID2013 as given in the caption of Table 5. The AGWN, JPEG, JP2K, AGPN, GB and GCD in CSIQ database denote the distortion types of additive Gaussian white noise, JPEG compression, JPEG2000 compression, additive pink Gaussian noise, Gaussian blur and global contrast decrements, respectively. The JP2K, JPEG, AWGN, GB and FF correspond to JPEG2000 compression, JPEG compression, additive white Gaussian noise, Gaussian blur and fast fading, respectively.

		SSIM	MS-SSIM	IW-SSIM	VIF	MAD	FSIM	GMSD	MCSD
TID2008	AGN	0.8107	0.8086	0.7869	0.8804	0.8388	0.8574	0.9180	0.9187
	ANC	0.8029	0.8054	0.7920	0.8768	0.8258	0.8515	0.8977	0.8898
	SCN	0.8145	0.8209	0.7714	0.8709	0.8678	0.8485	0.9132	0.9210
	MN	0.7795	0.8107	0.8087	0.8683	0.7336	0.8023	0.7087	0.7207
	HFN	0.8729	0.8694	0.8662	0.9075	0.8864	0.9093	0.9189	0.9184
	IN	0.6732	0.6907	0.6465	0.8326	0.0650	0.7456	0.6611	0.6893
	QN	0.8531	0.8589	0.8177	0.7970	0.8160	0.8555	0.8875	0.8952
	GB	0.9544	0.9563	0.9636	0.9540	0.9197	0.9472	0.8968	0.8880
	DEN	0.9530	0.9582	0.9473	0.9161	0.9434	0.9604	0.9752	0.9766
	JPEG	0.9252	0.9322	0.9184	0.9168	0.9275	0.9282	0.9525	0.9486
	JP2K	0.9625	0.9700	0.9738	0.9709	0.9707	0.9775	0.9795	0.9787
	JGTE	0.8678	0.8681	0.8588	0.8583	0.8661	0.8708	0.8621	0.7681
	J2TE	0.8577	0.8606	0.8203	0.8501	0.8394	0.8542	0.8825	0.8946
	NEPN	0.7107	0.7377	0.7724	0.7619	0.8287	0.7494	0.7601	0.7990
	Block	0.8462	0.7546	0.7623	0.8324	0.7970	0.8489	0.8967	0.8933
	MS	0.7231	0.7338	0.7067	0.5102	0.5161	0.6695	0.6486	0.5350
	CTC	0.5246	0.6381	0.6301	0.8188	0.2723	0.6480	0.4659	0.5073
CSIQ	AGWN	0.8974	0.9471	0.9380	0.9575	0.9541	0.9262	0.9676	0.9674
	JPEG	0.9546	0.9634	0.9662	0.9705	0.9615	0.9654	0.9651	0.9670
	JP2K	0.9606	0.9683	0.9683	0.9672	0.9752	0.9685	0.9717	0.9746
	AGPN	0.8922	0.9331	0.9059	0.9511	0.9570	0.9234	0.9502	0.9479
	GB	0.9609	0.9711	0.9782	0.9745	0.9602	0.9729	0.9712	0.9747
	GCD	0.7922	0.9526	0.9539	0.9345	0.9207	0.9420	0.9037	0.9509
LIVE	JP2K	0.9614	0.9627	0.9649	0.9696	0.9692	0.9717	0.9711	0.9720
	JPEG	0.9764	0.9815	0.9808	0.9846	0.9786	0.9834	0.9782	0.9790
	AWGN	0.9694	0.9733	0.9667	0.9858	0.9873	0.9652	0.9737	0.9809
	GB	0.9517	0.9542	0.9720	0.9728	0.9510	0.9708	0.9567	0.9528
	FF	0.9556	0.9471	0.9442	0.9650	0.9589	0.9499	0.9416	0.9527
MICT	JPEG	0.8590	0.8618	0.9204	0.9061	0.9165	0.8989	0.8235	0.8730
	JP2K	0.9399	0.9377	0.9549	0.9559	0.9548	0.9561	0.9369	0.9442

It can be seen from Table 5 and Table 6 that the proposed MCSD is among the top 3 models 35 times, followed by the GMSD and the FSIM, which are among the top 3 models 31 times and 23 times, respectively.

We observe in Table 2 that the MCSD has a relatively poor performance on the MICT database. Here from Table 6 we can see that this is because the MCSD has a bad performance on the images with JPEG distortions in the MICT database. Since all the 14 reference images used in the MICT were also used in the TID 2008, TID 2013 and LIVE databases, we also calculated the SROCC values of the MCSD by only using the JPEG distorted images of the 14 images in these three databases and we got good results (e.g. 0.9400 for TID 2013). Since we cannot find all the compression ratios used in the four databases, we cannot verify if the different performances of the MCSD on the JPEG distortions in different databases are due to different compression ratios or different subjective testing methods.

One limitation of the MCSD is that it does not take advantage of the phase information, thus it is unable to catch the phase distortion accurately. The fact can be drawn from Table 5 that the proposed MCSD does not perform quite well for those distortions like transmission errors and Gaussian blur. The transmission errors like JGTE (JPEG transmission errors), J2TE (JP2K transmission errors) and fast fading (a JP2K codec passed through a Rayleigh fast fading channel to simulate packet loss) will lead to the magnitude degradation and phase distortion of the image. The Gaussian blur not only affects the luminance change and the texture shape, but also causes phase shifting and loss.

3.5 Performance comparisons across different distortion types

A good IQA model should also predict the image quality consistently across different distortion types. Thus we also show the scatter plots in Fig.5 for some representative models (SSIM, MS-SSIM, IW-SSIM, MAD, VIF, FSIM, GMSD and MCSD) on the TID2008 database. The curves were obtained by a nonlinear fitting using Eq. (10).
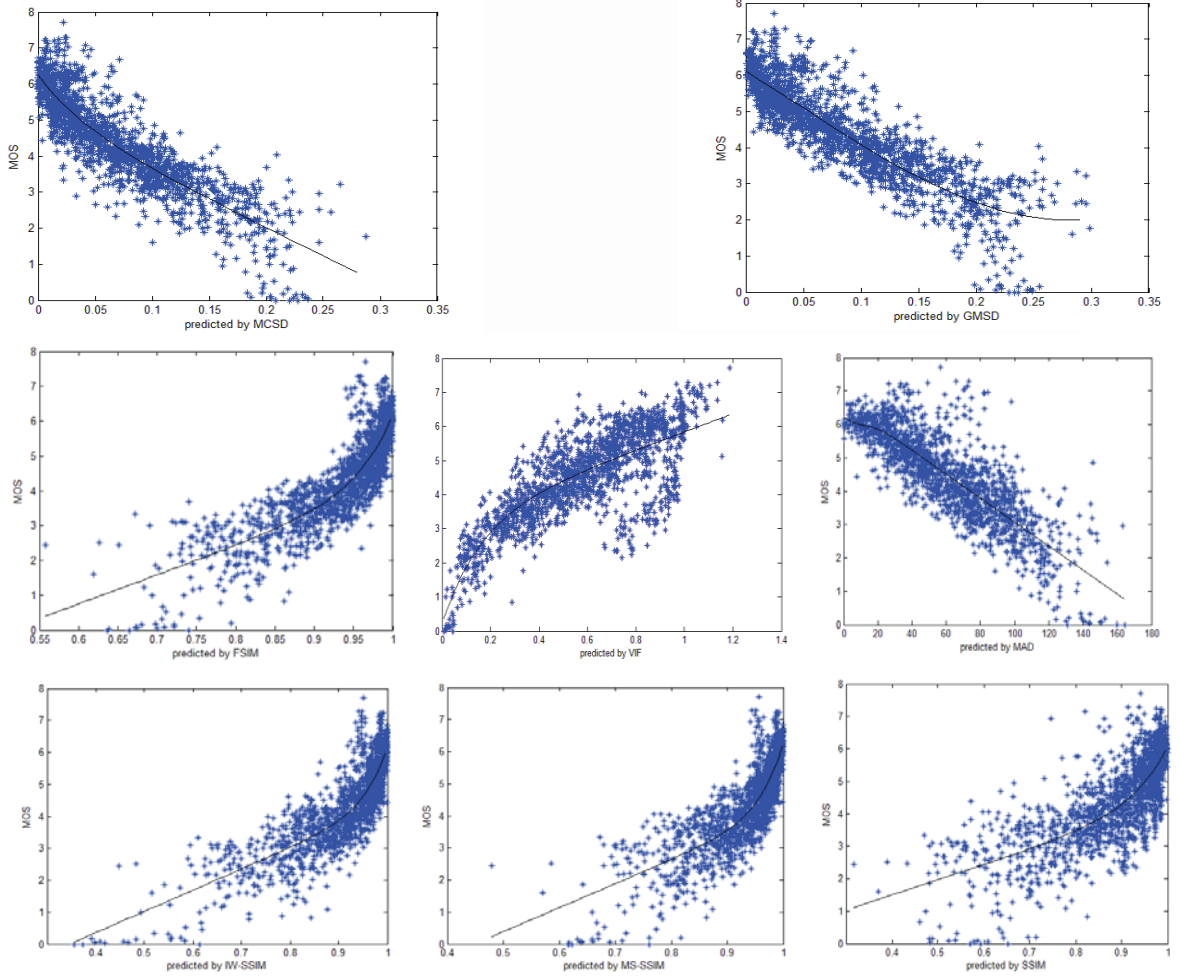
Fig.5. Scatter plots of predicted quality scores against the subjective quality scores (MOS) by representative FR-IQA models on the TID2008 database.

It can be seen that our MCSD is more concentrated across different groups of distortion types than the other competitors. It also shows that objective scores predicted by the MCSD correlate better with human judgments than the other competitors. From above analyses, we can safely reach a conclusion that the proposed model MCSD is a very effective model when compared with other state-of-the-art methods.

3.6 Computational cost

Besides the effectiveness, we will show the efficiency (in term of computation cost) of the proposed MCSD. We first analyze the computational complexity of the MCSD and the GMSD, and then give the computational cost (running time) of the competing IQA models. For each scale, the MCSD first convoluted the reference image and the distorted one with a 2×2 average filter①, and then computed the quality map② using Eq. (3), finally, the deviation pooling③ was used to yield a quality score. Suppose the total number of pixels in an image is m. The number of multiplications for step ① , ② and ③ is 8m (m×2×2×2), 21m (4m×2+m×2+4m×2+3×m) and m, respectively. For each scale, the MCSD needs 30m multiplications, thus the total number of multiplications for the MCSD is 90m

16

(30m×3). By contrast, the multiplications for the GMSD are 8m for pre-processing (m×2×2×2), 43m for the quality map generation ((m×3×3×2+2m) ×2+3m) and m for the deviation pooling. Thus the total number of multiplications for the GMSD is 52m (8m+43m+m). In all, the computational complexity for both the MCSD and the GMSD is O (m) and the computational complexity for our MCSD is a litter higher than the GMSD. In addition, the memory complexities for our MCSD and the GMSD are all in O (m). In other words, the time and memory complexities for our MCSD and the GMSD have a linear relation with the image size. In Table ~~6~~ 7 we list the amount of time (in seconds) to compute each quality measure on a color image of resolution 512×512 (taken from CSIQ database) on a 2.66 GHz Intel Core2 Quad CPU with 5 GB of RAM. Note that all the algorithms were not run through parallelization. The software platform for all algorithms was MATLAB R2013a. The computational costs for all the IQA models were done using 100 repetitions and the first 30 ~~largest~~ values of which were discarded to account for caching effects. The average of the remaining ones was used as the computational cost. In addition, to show how accurate the measurements are, we also list their standard deviation. To reduce the impact of the operating system's scheduler, we terminated as many background processes as possible.

Table ~~6~~ 7 Running time of the competing IQA models

IQA models	~~Running time (s)~~	~~standard deviation~~	Running time(s)	standard deviation
SSIM	~~0.035~~	~~2.3504×10^{-4}~~	0.0363	1.5×10^{-3}
MS-SSIM	~~0.160~~	~~5.7766×10^{-4}~~	0.1591	1.7×10^{-3}
IW-SSIM	~~0.838~~	~~4.5×10^{-3}~~	0.8481	2.78×10^{-2}
VIF	~~1.842~~	~~7.2×10^{-3}~~	1.8354	1.02×10^{-2}
MAD	~~2.674~~	~~1.64×10^{-2}~~	2.6820	5.09×10^{-2}
FSIM	~~0.510~~	~~2.3×10^{-3}~~	0.5090	6×10^{-3}
GMSD	~~0.012~~	~~2.4401×10^{-4}~~	0.0144	1.4×10^{-3}
MCSD	~~0.020~~	~~4.7771×10^{-4}~~	0.0203	1.3×10^{-3}

We can see from Table ~~6~~ 7 that our MCSD, the GMSD and the SSIM are the top three efficient IQA models and run much faster than other competing IQA models. For example, the MCSD is 25 times faster than the excellent model FSIM achieving the state-of-the-art prediction performance. In a word, there is no significant difference between the proposed MCSD and the GMSD. One can select the MCSD when considering the effectiveness (correlation with human judgment), and the GMSD when he processes images with huge size (e.g. remote sensing images).

4 Conclusion

We have presented a highly effective and efficient FR-IQA using the contrast feature and the multi-scale representation. For each image scale, we adopt the standard deviation as the pooling strategy to score the contrast similarity map. Our consideration is that the contrast and the standard deviation have instinctive attributes to some degree; namely, they are both "range" indicators. The contrast indicates the

range of image luminance, whereas the standard deviation indicates the range of distortion severity in an image. Using six databases, we compared the proposed MCSD model with other eight state-of-the-art FR-IQA models. The results showed that the MCSD performs better in terms of both accuracy and efficiency, making MCSD a good candidate for high performance IQA applications. To further ameliorate the proposed metric, we will consider how to integrate the phase information into the MCSD to make it more robust to different distortion types in our future work.

References

[1] Z. Wang and A. C. Bovik, Modern Image Quality Assessment. San Rafael, CA, USA: Morgan & Claypool (2006).

[2] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing , Orlando, FL, vol. 4,3313–3316(2002)

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., 13(4), 600–612(2004).

[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in Proc. 37th Asilomar Conf. Signals, Syst., Comput., 1398–1402(2003).

[5] C. Li and A.C. Bovik, "Three-component weighted structural similarity index", in Proc. SPIE, vol. 7242(2009).

[6] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Trans. Image Process. 20(5), 1185–1198(2011).

[7] N.Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," IEEE Trans. Image Process.9(4), 636–650(2000).

[8] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Trans. Image Process.16 (9), 2284–2298(2007).

[9] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process. 15(2), 430–444(2006).

[10] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Process.14 (12), 2117–2128(2005).

[11] X.B. Gao, T. Wang, and J. Li, "A content-based image quality metric," Lecture Notes in Artificial Intelligence. Vol. 3642, 231–240(2005).

18

[12]   Eric C. Larson and Damon M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," J. Electron. Imaging 19(1), 011006(2010).

[13]   A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," IEEE Trans. Image Process. 21(4), 1500–1512(2012).

[14]   L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process. 20(8), 2378-2386(2011).

[15]   W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," IEEE Trans. Image Process.23 (2), 684-695(2014).

[16]   L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in Proc. 19th IEEE ICIP, 1477-1480(2012).

[17]   W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," J. Vis. Commun. Image Represent., 22(4), 297–312(2011).

[18]   L. He, F.Gao , W. Hou, L. Hao , "Objective image quality assessment: a survey," Int J Comp Math.91(11), 2374-2388(2014).

[19]   A. B. Watson, "Visual detection of spatial contrast patterns: evaluation of five simple models," Opt. Exp., vol. 6, pp. 12–33, Jan. 2000.

[20]   M. Ferraro, G. Boccignone, Image contrast enhancement via entropy production, Real-Time Imaging 10 (2004) 229-238.[scale-space]

[21]   Stephen DelMarco and Sos Agaian,"The design of wavelets for image enhancement and target detection", Proc. SPIE 7351, 735103, 9 pages,   2009.

[22]   Panetta, K.; Agaian, S.; Yicong Zhou; Wharton, E.J.; "Parameterized Logarithmic Framework for Image Enhancement Systems," IEEE Trans. Syst. Man Cybern. Part B-Cybern, Volume: 41, Issue: 2, 2011, Page(s): 460 – 473.

[23]   Karen Panetta, Eric Wharton, and Sos Agaian, "Human visual System Based Image Enhancement and Measure of Image Enhancement," IEEE Trans. Syst. Man Cybern. Paart A-Syst. Hum., vol. 38, no.1, pp.174-188, February 2008.

[24]   S. Agaian, "Visual Morphology" Proceedings, IS&T/ SPIE's Symposium on Electronic Imaging Science &Technology, San Jose. CA, January, vol. 3304, pp.153-163, 1999.

[25]   Agaian, S.S.; Silver, B.; Panetta, K.A.; Transform Coefficient Histogram-Based Image Enhancement Algorithms Using Contrast Entropy, IEEE Trans. Image Process, Volume 16, Issue 3, March 2007 Page(s):741 – 758.

[26]   A. M. Haun and E. Peli, "Is image quality a function of contrast perception?" in IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 86510C-86510C (2013).

[27]   A. B. Watson, H. A. Solomon, A. Ahumada, and A. Gale, "DCT basis function visibility: Effects of viewing distance and contrast masking," in Human Vision, Visual Processing, and Digital Display IV, B.E. Rogowitz, Ed., pp.99–108(1994).

[28]   T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," J. Appl. Statist., 21(2), 224–270(1994).

[29]   A Saha, Q M J, Wu, "Utilizing image scales towards totally training free blind Image quality assessment," IEEE Trans. Image Process. 24(6), 1879-1892(2015).

[30] Pelli DG, Bex P, "Measuring contrast sensitivity," Vis Res (90), 10-14(2013).

[31] E. Peli, "Contrast in complex images," J. Opt. Soc. Amer. A.vol. 7, 2032–2039(1990).

[32] J Lu, D. M. Healy, J.B. Weaver, "Contrast enhancement of medical images using multiscale edge representation," Opt. Eng. 33(7), 2151-2161(1994).

[33] G. Boccignone, "A multiscale contrast enhancement method," In Proc. IEEE ICIP, 306-309, (1997).

[34] S.C. Nercessian, K.A. Panetta, and S. S. Agaian. "Non-linear direct multi-scale image enhancement based on the luminance and contrast masking characteristics of the human visual system," IEEE Trans. Image Process, 22(9), 3549-3561(2013).

[35] Toet, "Multiscale contrast enhancement with applications to image fusion," Opt. Eng., 31(5), 1026-1031(1992).

[36] S.C. Nercessian, K. A. Panetta, and S.S. Agaian, "Multiresolution Decomposition Schemes Using the Parameterized Logarithmic Image Processing Model with Application to Image Fusion," EURASIP J. Adv. Signal Process,(2011), http://dx.doi.org/10. 1155/2011/515084

[37] S. C. Nercessian, K. A. Panetta, and S.S. Agaian, "Multiscale image fusion using an adaptive similarity-based sensor weighting scheme and human visual system-inspired contrast measure," J. Electron. Imaging, 21(2), 021112(2012).

[38] Mukhopadhyay S, Chanda B. A multiscale morphological approach to local contrast enhancement. Signal Process, 2000; 80:685-96.

[39] Z. Rahman, G. A. Woodell, and D. J. Jobson, "A Comparison of the Multiscale Retinex with Other Image Enhancement Techniques," Proceedings of the IS&T 50th Anniversary Conference, May 1997.

[40] 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>.

[41] E. C. Larson and D. M. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy, Journal of Electronic Imaging .19 (1), 1-21 (2010).

[42] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, TID2008—A database for evaluation of full-reference visual quality assessment metrics, Advances of Modern Radioelectronics.10 (4) 30–45(2009).

[43] N. Ponomarenko et al., "Color image database TID2013: Peculiarities and preliminary results," in Proc. 4th Eur. Workshop Vis. Inf. Process.,Jun. 2013, pp. 106–111.

[44] Y. Horita, K. Shibata, Y. Kawayoke, et al., MICT image quality evaluation database [Online]. Available: <http://mict.eng.u-toyama.ac.jp/mictdb.html.2000>.

[45] A. Ninassi, P.L. Calet, F. Autrusseau, Pseudo no reference image quality metric using perceptual data hiding, SPIE Human Vision and Electronic Imaging 6057(08) (2006)146-157.

[46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15(11), 3440–3451 (2006).

Highlights

- Perceptual image quality assessment (IQA) uses a computational model to assess the image quality in consistent with human opinions. Generally, the effectiveness, namely, high correlation with the human subjective score, is the prerequisite of a good IQA model, it is of first importance to a IQA model.

- The efficiency (the least computation cost), however, is the second requirement of a good IQA model which become important under the premise that an IQA model meet the condition of "effectiveness".

- Thus the effectiveness and efficiency are two ultimate goals for the design of IQA models, but unfortunately it is hard to reach well these goals simultaneously. In this paper, we developed an effective and efficient IQA model called multiscale contrast similarity deviation (MCSD) which explores the contrast features by resorting to multiscale representation.

- Performances on six benchmark databases demonstrate its effectiveness and efficiency.