Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 1243410, 10 pages
http://dx.doi.org/10.1155/2016/1243410

Research Article

Subjective Score Predictor: A New Evaluation Function of Distorted Image Quality

Xiaoyan Luo,[1] Shining Wang,[2] and Ding Yuan[1]

[1]*Image Processing Center, Beihang University, Beijing 100191, China*
[2]*China Waterborne Transport Research Institute, Beijing 100088, China*

Correspondence should be addressed to Xiaoyan Luo; luoxy@buaa.edu.cn and Ding Yuan; dyuan@buaa.edu.cn

Received 23 March 2016; Revised 5 July 2016; Accepted 11 July 2016

Academic Editor: Mitsuhiro Okayasu

Image quality assessment (IQA) is a method to evaluate the perceptual performance of image. Many objective IQA algorithms are developed from the objective comparison of image features, which are mainly trained and evaluated from the ground truth of subjective scores. Due to the inconsistent experiment conditions and cumbersome observing processes of subjective experiments, it is imperative to generate the ground truth for IQA research via objective computation methods. In this paper, we propose a subjective score predictor (SSP) aiming to provide the ground truth of IQA datasets. In perfect accord with distortion information, the distortion strength of distorted image is employed as a dependent parameter. To further be consistent with subjective opinion, on the one hand, the subjective score of source image is viewed as a quality base value, and, on the other hand, we integrate the distortion parameter and the quality base value into a human visual model function to obtain the final SSP value. Experimental results demonstrate the advantages of the proposed SSP in the following aspects: effective performance to reflect the distortion strength, competitive ground truth, and valid evaluation for objective IQA methods as well as subjective scores.

1. Introduction

Image quality assessment (IQA) is fundamental and important in evaluating and improving the perceptual quality of images, which is widely applied in image-based instrumentation [1, 2]. The IQA problem can be addressed in objective and subjective classes.

Over the years, various objective IQA methods have been proposed. According to the availability of reference image, the objective image quality metrics is classified as full-reference (FR), reduced-reference (RR), and no-reference (NR) [3]. Most existing methods are FR via comparing the distorted image with a complete reference image. The simplest FR metrics is the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR). However, they both cast the image quality on the pixel values without the structure of image and human visual system (HVS). To overcome these drawbacks, Wang et al. proposed the structural similarity index (SSIM) to compare the local patterns on luminance and contrast [4]. Deriving from SSIM, some researchers developed a gradient-based SSIM [5] and a multiscale SSIM [6]. However, the SSIM and the modified SSIM IQA methods compare the features within corresponding patches of reference and distorted images, so they ignore the image curvature information [7]. To adequately investigate the visual disparity between a center patch and its spatial neighborhoods, Zhou et al. proposed an FR IQA scheme via comparing visual similarity in both the interpatch and intrapatch ways in [7]. Since the reference image is often not available in practice, NR-IQA methods have become a good alternative to evaluating the distorted image quality. To remedy the lack of prior knowledge of reference images, most of NR-IQA algorithms are limited to some specific distortions, dataset training, or IQA models. For example, Jiang et al. designed the image quality metric for the compressed remote sensing image assessments [8]. Based on training on dataset, Tang et al. learned a map model from low-level features to image quality scores from human observers [9]. With a multivariate Gaussian model (MVG) of IQA, Mittal et al. proposed a blind image quality analyzer via measuring the distance between the MVG fit of image features from test image and a MVG model of natural images [10]. In practice, their assumption is not satisfied

well, so they failed in the various real applications. Therefore, RR-IQA methods provide a lying solution between FR and NR-IQA methods, which is based on partial information about the reference image. For instance, Wang and Li et al., respectively, developed the RR image quality metrics using a wavelet-domain natural image statistic model and a divisive normalization-based image representation [11, 12].

Generally, all objective IQA metrics is to evaluate the image quality in agreement with the subjective opinion of human observers [13], so their performances are validated via comparing with the subjective scores in open IQA datasets such as LIVE [14], A57 [15], CSIQ [16], IVC [17], TID2008 [18], and Toyoma [19]. However, these datasets only have a limited amount of images since the subjective experiments are time-consuming and expensive [20]. Among the above six datasets, the number of natural reference images is only up to 30 within Toyoma. In practice, this is not enough to generally cover the vast proliferation of image data, and it is really not enough to reflect the performance of objective IQA algorithms. In addition, the subjective scores could be unfaithful to reflecting the ground truth of image quality because of the various experimental conditions, individual observers, and different processing methods of raw scores. Thus, it is necessary to develop an objective predictor to correctly represent the real quality of distortion image.

To predict the subjective scores of image quality, Kaya et al. imitate the human observers with a trained multilayer neural network based on extracted statistical features [21]. It requires many training samples including not only image features but also subjective scores. To address the lack of ground truth of image quality, Lu et al. calculate the objective distortion score (ODS) from the logarithm of distortion parameter p_i, that is, $\ln(p_i)$ [22]. The larger ODS value means the worse image quality. This is the first work to provide ground truth for IQA research according to a simple function of the impaired distortion. However, referring to the analyzed criteria of objective IQA [23], ODS has some problems in terms of consistency, discrimination, and convergence. For example, since $\ln(\cdot)$ is a monotonically increased function, the ODS is still increasing on larger p_i no matter that the image quality becomes better or worse. On the other hand, different distortions could result in same ODS values at same distortion strengths, while the impaired degrees are obviously different. To remain consistent with the subjective scores (NODS), the ODS are mapped to the interval $[0, 100]$, but the normalization is limited by the difference of the evaluated maximum and minimum values. More seriously, when the distortion parameter is quite small, the ODS value is close to minus infinity, which leads to failure of this normalizing.

As the evaluation of subjective scores faces the aforementioned problems, we design a subjective score predictor (SSP) function to calculate the scores of distorted images. To reflect the distortion strength in agreement with ground truth of subjective scores, we combine the necessary knowledge of a successful IQA algorithm design, including the information of source image, distortion, and the human visual system (HVS) [23]. In this paper, the SSP function is demonstrated with three advantages. First, SSP unifies a visual perception based model, so it partially avoids the individual differences between observers on different distortions with different levels. Second, the distortion information and subjective score of source image are both integrated to make the subjective score prediction more consistent with human-based subjective scores. Third, an objective function is used to calculate the final value. It is efficient and feasible, which can reduce the cost and manpower. Therefore, the proposed approach has stronger application in constructing various IQA datasets to cover more source images such as high-resolution remote sensing images and disgusting medical disease images, which are difficult to test on human observers in subjective experiments.

The rest of this paper is organized as follows: Section 2 introduces the SSP model, and Section 3 analyzes its characteristics and extension. Its effectiveness is validated in Section 4. Section 5 concludes this paper.

2. The Proposed Model

To reflect the subjective scores of image quality by an objective function, we should firstly consider the requirement of a successful objective IQA algorithm. Inspired from the recent work, which suggested that a successful IQA algorithm design should combine the knowledge of source image, distortion, and HVS [23], we design our SSP function on information of the source image, distortion, and HVS characteristic, which are prior knowledge on constructing an IQA dataset.

2.1. The Source Image: Reference. In all open IQA datasets, the entire image databases are derived from some selected source images with given distortions. Referring to the source images, the distorted images are subjectively evaluated by human observers, and then the raw scores are adjusted to obtain the final subjective quality scores. From the perspective of objective IQA, the source images are treated as reference images, and the subjective observing is an FR IQA processing in human eyes. Therefore, the information of source image I_r including its corresponding distortion level p_r and subjective score S_r can be considered as reference information in SSP. For all the public datasets, the source images are given with perfect image quality, so subjective score S_r equals 100, and distortion level p_r is at zero-distortion.

2.2. Distortion: Parameter. Usually, the distortion types include noise, blur, compression, transmission, and intensity deviation, which are involved in the open IQA datasets [24]. Their parameters are standard deviation of noise σ_{N}, standard deviation of blur kernel σ_{B}, bit rate of per pixel r_{BPP}, transmission signal-to-noise ratio r_{SNR}, and intensity change value v_{I}, respectively.

For each distortion type, the image quality could monotonically vary according to the strength of the distortion level. Intuitively, if the distortion strength reaches one extreme, the distortion is very little or even inexistent. This means that the distortion has zero-distortion level, denoted as p_0, such as $\sigma_{\mathrm{N}} = 0$ and $\sigma_{\mathrm{B}} = 0$ for noise and blur, respectively. Ideally, the distortion level of ground truth image is at zero-distortion level p_0. Following the psychological mechanism of image

understanding, it is difficult to distinguish the levels of the evaluated images by human when they are impaired seriously beyond a certain threshold. This threshold can be marked as zero-score distortion level p_t. In a word, zero-distortion level p_0 and zero-score distortion level p_t can portray the distortion property in SSP.

2.3. HVS: Mechanism. HVS has complicated psychological inferences and is not a direct translation of information [25]. A vision system is subdivided into multiple parts with quite distinctive functions [26]. Based on the contrast sensitivity function (CSF), HVS models are developed as a band-pass or low-pass filter, which are reviewed by Kim and Allebach in halftoning [27]. In the point of spatial frequency, the most popular modulation transfer functions (MTF) of HVS are Gaussian model, exponent model, Barten model, and compound model [28, 29]. Recently, Bayesian brain theory suggests that the brain works with an internal generative mechanism for visual perception and image understanding [30]. Therefore, we are devoted to look for a function to deal with the signal processing approach of IQA and signify the physiological and psychological mechanisms of perception. Observing the popular CSF and MTF models in [27–29], the exponential function is employed as a prototype. In the same way, we depict the SSP function based on an exponential function with the prior knowledge of the reference information (distortion parameter p_r and subjective score S_r) and the distortion information (zero-distortion parameter p_0 and zero-score threshold p_t).

2.4. SSP: Function. Firstly, zero-distortion parameter p_0 and zero-score threshold p_t are treated as a zero-point (high score) and one-point (low score) of distortion strength. To limit the strength of different distortions to a uniform interval, input parameter p_i at ith-level distortion is normalized by the difference between one-point p_t and zero-point p_0. Secondly, to ensure that the SSP has a corresponding score referring to the subjective opinion of source image, the distortion strength is considered as a relative one from input parameter p_i to distortion level p_r of source image. Thirdly, the power of the SSP function is adjusted by a positive fading factor k depending on the distortion type. Finally, the SSP is defined as an exponential function in the following way:

$$\text{SSP: } S_i = S_r \cdot e^{-k\cdot((p_i-p_r)/(p_t-p_0))}. \tag{1}$$

Analyzing on this definition, the more seriously the distortion degree is impaired, the less the SSP value becomes, which can conform to the subjective expectation of human on image quality. When the reference image is the ground truth with zero-distortion p_0, the formula can be simplified as

$$\text{SSP: } S_i = 100 \cdot e^{-k\cdot((p_i-p_0)/(p_t-p_0))}. \tag{2}$$

As shown in (2), the SSP is only faithful to the actual distortion information of test images. In particular, for IQA datasets, this can avoid the individual differences of observers and the unfaithful subjective scores because of complex experimental conditions.

3. The Analysis of SSP

The SSP function can satisfy the following mathematical conditions.

(i) Bounded. To make the subjective experiments significant, the impaired distortion level is always between p_0 and p_t. Therefore, the input parameter have two extreme values p_0 and p_t. Based on simplified formula (2), when $p_i = p_0$, we can get $S_i = S_r = 100$. When $p_i = p_t$, $S_i = 100/e^k$. Therefore, the SSP can be bounded in $[100/e^k, 100]$.

(ii) Monotonous. According to the variation trend of image quality with the increasing numerical value of distortion strength, the distortions can be divided into increasing-type distortion (the larger numerical value of distortion strength reflecting less distortion can generate better image) and decreasing-type distortion (the larger numerical value of distortion strength reflecting more distortion can generate worse image).

For increasing-type distortion, zero-point distortion p_0 with higher quality score should have a larger numerical value than one-point distortion p_t with less quality score, that is, $p_t < p_0$. Then, the distortion parameter is $p_t \le p_i \le p_0$. Given two distortion parameters $p_t \le p_{i1} < p_{i2} \le p_0$, their relative distances satisfy $(p_{i1} - p_0)/(p_t - p_0) > (p_{i2} - p_0)/(p_t - p_0)$. Therefore, according to the monotonically decreasing property of e^{-x}, the final results of SSP meet $Sp_{i1} < Sp_{i2}$.

In contrast, for decreasing-type distortion, since the image quality score holds decreasing change regulation to the larger numerical value of distortion strength, we can get $p_0 < p_t$. If two input parameters are $p_0 \le p_{i1} < p_{i2} \le p_t$, their relative distances satisfy $(p_{i1} - p_0)/(p_t - p_0) < (p_{i2} - p_0)/(p_t - p_0)$, and then $Sp_{i1} > Sp_{i2}$.

(iii) Invertible. From formula (1), we can obtain the image quality score of reference image S_r via transposition operation:

$$\begin{aligned} S_r &= S_i \cdot e^{k\cdot((p_i-p_r)/(p_t-p_0))} = S_i \cdot e^{k\cdot(-(p_r-p_i)/(p_t-p_0))} \\ &= S_i \cdot e^{-k\cdot((p_r-p_i)/(p_t-p_0))}. \end{aligned} \tag{3}$$

Similarly, it obeys the basic prototype of formula (1), only exchanging the information of the reference image and test image.

(iv) Recursive. If objective assessment Sp_{i1} at parameter p_{i1} is known, the subjective score of reference image can be obtained according formula (3) as $S_r = Sp_{i1} \cdot e^{-k\cdot((p_r-p_{i1})/(p_t-p_0))}$. Thus, at parameter p_{i2} of the same distortion, we can calculate the objective assessment as

$$\begin{aligned} Sp_{i2} &= S_r \cdot e^{-k\cdot((p_{i2}-p_r)/(p_t-p_0))} \\ &= Sp_{i1} \cdot e^{-k\cdot((p_r-p_{i1})/(p_t-p_0))} \cdot e^{-k\cdot((p_{i2}-p_r)/(p_t-p_0))} \\ &= Sp_{i1} \cdot e^{-k\cdot((p_{i2}-p_{i1})/(p_t-p_0))}. \end{aligned} \tag{4}$$

Table 1: The characteristic information of LIVE II and the parameter setting of our proposed SSP function.

Distortion type	JPEG 2000	JPEG	White noise	Gaussian blur	Fast fading
Variable p_i	r_{BPP}	r_{BPP}	σ_{N}	σ_{B}	r_{SNR}
Min(p_i)	0.03	0.15	0.01	0.41	15.50
Max(p_i)	3.15	3.33	1.99	14.99	26.10
Max(DMOS)	74.71	80.88	75.67	84.48	76.97
Min(DMOS)	19.96	17.90	18.17	19.71	18.34
Range(DMOS)	54.75	62.98	57.50	64.77	58.63
DMOS(p_i) trend	↓	↓	↑	↑	↓
MOS(p_i) trend	↑	↑	↓	↓	↑
p_0	*3.5*	*4*	*0*	*0*	*45*
p_t	*0.01*	*0.1*	*5*	*20*	*1*
k	*1.4*	*1.7*	*3.5*	*2.5*	*1.8*

This formula also follows the definition of SSP in formula (1), which refers to the distorted image with Sp_{i1} score of image quality at parameter p_{i1}.

(v) Multiple Distortions. Our SSP can be extended into multiple distortions easily. Constructing one IQA dataset with multiple distortions, the reference image is sequentially polluted by different distortions. Given that the source image is ground truth and two distortion parameters p_{i1} and p_{i2} on different distortions with factors k_1 and k_2, respectively, multiple distorted image can be obtained from first distortion with factor k_1 and then the distortion with k_2. Therefore, after the first distortion is finished, the SSP with reference to the ground truth for the captured distorted image is

$$Sp_{i1} = S_r \cdot e^{-k_1 \cdot ((p_{i1} - p_{01})/(p_{t1} - p_{01}))}. \tag{5}$$

After the first distortion is finished, Sp_{i1} is the reference for the second distortion. Therefore, the SSP with reference to Sp_{i1} is

$$Sp_{i2} = Sp_{i1} \cdot e^{-k_2 \cdot ((p_{i2} - p_{02})/(p_{t2} - p_{02}))}. \tag{6}$$

Combining formula (5) into (6), we can get the final SSP to the ground truth as

$$\begin{aligned} Sp_{i2} &= Sp_{i1} \cdot e^{-k_2 \cdot ((p_{i2} - p_{02})/(p_{t2} - p_{02}))} \\ &= S_r \cdot e^{-k_1 \cdot ((p_{i1} - p_{01})/(p_{t1} - p_{01}))} \cdot e^{-k_2 \cdot ((p_{i2} - p_{02})/(p_{t2} - p_{02}))} \\ &= S_r \cdot e^{-k_1 \cdot ((p_{i1} - p_{01})/(p_{t1} - p_{01})) - k_2 \cdot ((p_{i2} - p_{02})/(p_{t2} - p_{02}))}. \end{aligned} \tag{7}$$

Overall, the SSP model can obtain bounded, monotonous, invertible, and recursive properties, so that it is easy to extend for multiple distortions. The SSP function is mainly dependent on objective distortion parameters, so it usually keeps the same order as the distortion degree order. In addition, SSP avoids the subjective experiments, which has diversity in experimental circumstances, individual differences, and different scenes, and it is easily obtained and faithful to the real image quality.

4. Experiments

This section presents the experiments on dataset LIVE II [14] owing to the given distortion parameters, in which 29 high-resolution 24 bits/pixel color images are distorted by five distortion types: JPEG 2000, JPEG, white noise, Gaussian blur, and transmission errors using a fast fading Rayleigh channel model [31]. The subjective scores are reflected by the difference of the Mean Opinion Score (MOS) between reference image and distorted image, naming it as Difference Mean Opinion Score (DMOS).

Initialization (the setting of the parameters). For SSP, three parameters including zero-distortion parameter p_0, zero-score threshold parameter p_t, and fading factor k are required to initialize at first for each distortion.

To obtain p_0 and p_t in SSP, we tabulate the distortion information, the subjective opinions, and their corresponding relationship in LIVE II dataset in Table 1. Considering the human distinctive characteristic, prior variables p_0 and p_t are empirically initialized as the values in italic type at the bottom of Table 1.

Fading factor k is based on the expectation that SSP and MOS are equivalent; that is, we expect that the SSP can hold consistency with MOS scores in terms of the upper limit, the lower limit, and the range of values. Since the clear reference image quality is viewed as 100, the MOS can be simply equal to 100-DMOS. To find the appropriate k for each distortion type, we adjusted fading factor k from 0.5 to 20 by step 0.1 in our experiment. Figure 1 illuminates the average of the differences in the upper limit, the lower limit, and the range of SSP values at each k sample. As shown in Figure 1, there exists only one point with least average difference for each distortion, which indicates the SSP fitting MOS well. Therefore, we can set k = 1.4, 1.7, 3.5, 2.5, and 1.8 as the fixed fading factors of JPEG 2000, JPEG, white noise, Gaussian blur, and transmission errors, respectively, also recorded at the bottom of Table 1.

Experiment 1 (the effectiveness of the proposed SSP). To measure the effectiveness of our SSP function, we compare SSP with the subjective scores DMOS in dataset and recent NODS index [22]. In this experiment, the linear correlation

Table 2: PLCC and SROCC between distortion strengths and image quality scores including DMOS, NODS, and SSP.

Distortion type	Image number	PLCC			SROCC		
		DMOS	NODS	SSP	DMOS	NODS	SSP
JPEG 2000	169	−0.7830	0.8642	0.9872	−0.8949	1	1
JPEG	175	−0.8019	0.9227	0.9814	−0.8680	1	1
White noise	145	0.7947	0.8414	−0.9864	0.9841	1	−1
Gaussian blur	145	0.7808	0.8792	−0.9727	0.9583	1	−1
Fast fading	145	−0.6438	0.9975	0.9983	−0.6486	1	1

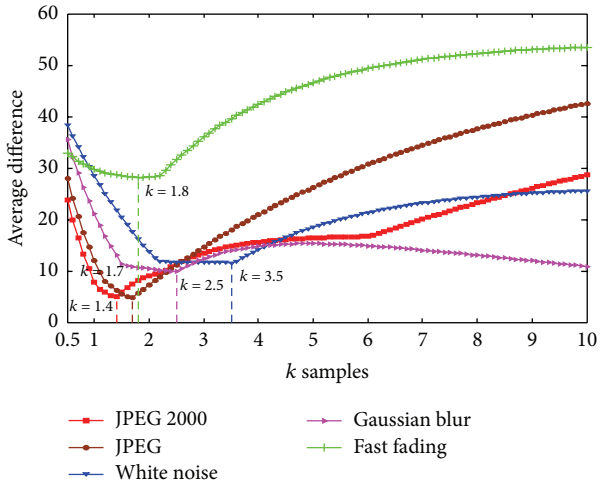


Figure 1: The average differences between SSP and 100-DMOS at k samples.

is tested between distortion strengths and image quality scores (i.e., SSP, DMOS, and NODS) using Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficient (SROCC).

One thing needs to be paid attention here. Our method directly evaluates the image quality, similar to the subjective MOS opinion of human expectation. In contrast, the NODS index is a distortion score more like the image quality difference DMOS from the reference image. Therefore, the correlation between proposed SSP function and distortion is opposite to that between DMOS and distortion, while this correlation of NODS should be the same as DMOS. Intuitively, for the increasing-type distortions such as JPEG 2000, JPEG, and fast fading, SSP should generate positive correlation values since the better image quality is from the larger distortion parameter, while the DMOS and NODS have negative values in both PLCC and SROCC. In contrast, for the decreasing-type distortions such as white noise and Gaussian blur, the PLCC and SROCC values are negative in SSP but positive in DMOS and NODS, since the distortion parameter increasing causes the image quality to be worse.

Table 2 tabulates the values of PLCC and SROCC between distortion strengths and the DMOS, NODS, and SSP scores, respectively. From it, we can observe that the DMOS expresses positive correlation for the decreasing-type distortions such as white noise and Gaussian blur, while negative correlation is responded for the increasing-type distortions such as JPEG 2000, JPEG, and fast fading. SSP holds opposite correlation compared with DMOS, which is consistent with the previous analysis. However, the NODS demonstrate an invariable positive correlation no matter what the distortion type is. Therefore, SSP can reasonably catch the human expectation about image quality with the strength changing for different distortions, similar to DMOS, but the NODS index has failed on the increasing-type ones. In addition, both of SSP and NODS show stronger linear correlation in PLCC and SROCC than DMOS, since they exclude the various changes in experimental conditions and raw score processing error of DMOS. Furthermore, since SSP is objectively calculated from the relative degree instead of the direct numerical value of distortion parameter, and SSP has better linear correlation than NODS in PLCC.

Experiment 2 (the comparison as ground truth). SSP is proposed to provide the ground truth of distorted image quality, similar to DMOS scores. To compare the SSP scores with the DMOS more distinctly, Figure 2 scatters the scores of DMOS and SSP for different distortions and gives the linear trends via $x - \text{detrend}(x)$ in MATLAB.

Comparing with DMOS, we can get three appearances from Figure 2: (i) similar effectivity: the linear trend line of SSP holds a complementary changing characteristic well with DMOS, so that it can reflect the overall variation trend of DMOS. (ii) Obvious discriminability: SSP is monotonically changes with the increasing of distortion numerical value. For one certain distortion strength, SSP has unique value, while the DMOS values present some shock changes. (iii) Objective quantitation: SSP only depends on the distortion information without the influence of the image scene and human subjective factors, so it demonstrates that the proposed SSP function is seamlessly meets the ground truth of IQA.

In the distortion distinction, we test whether the SSP changes reasonably with the real image quality. Figure 3 shows a group of distorted images for a house. The corresponding values of DMOS and SSP are shown in Table 3. Figures 3(a)–3(c) are the distorted images of fast fading, and their signal-to-noise ratio are 17.9, 20.3, and 22.7, respectively. In theory, the quality scores should increase from (a) to (c). The SSP values in Table 3 follow this rule, while the DMOS values do not make sense for Figure 3(b). Figures 3(d)-3(e) are both compressed images with similar bit rate per pixel via JPEG and JPEG 2000. As we well know, JPEG 2000 has better compression property than JPEG, so the image in Figure 3(e) should have high score compared to Figure 3(d). It is logical

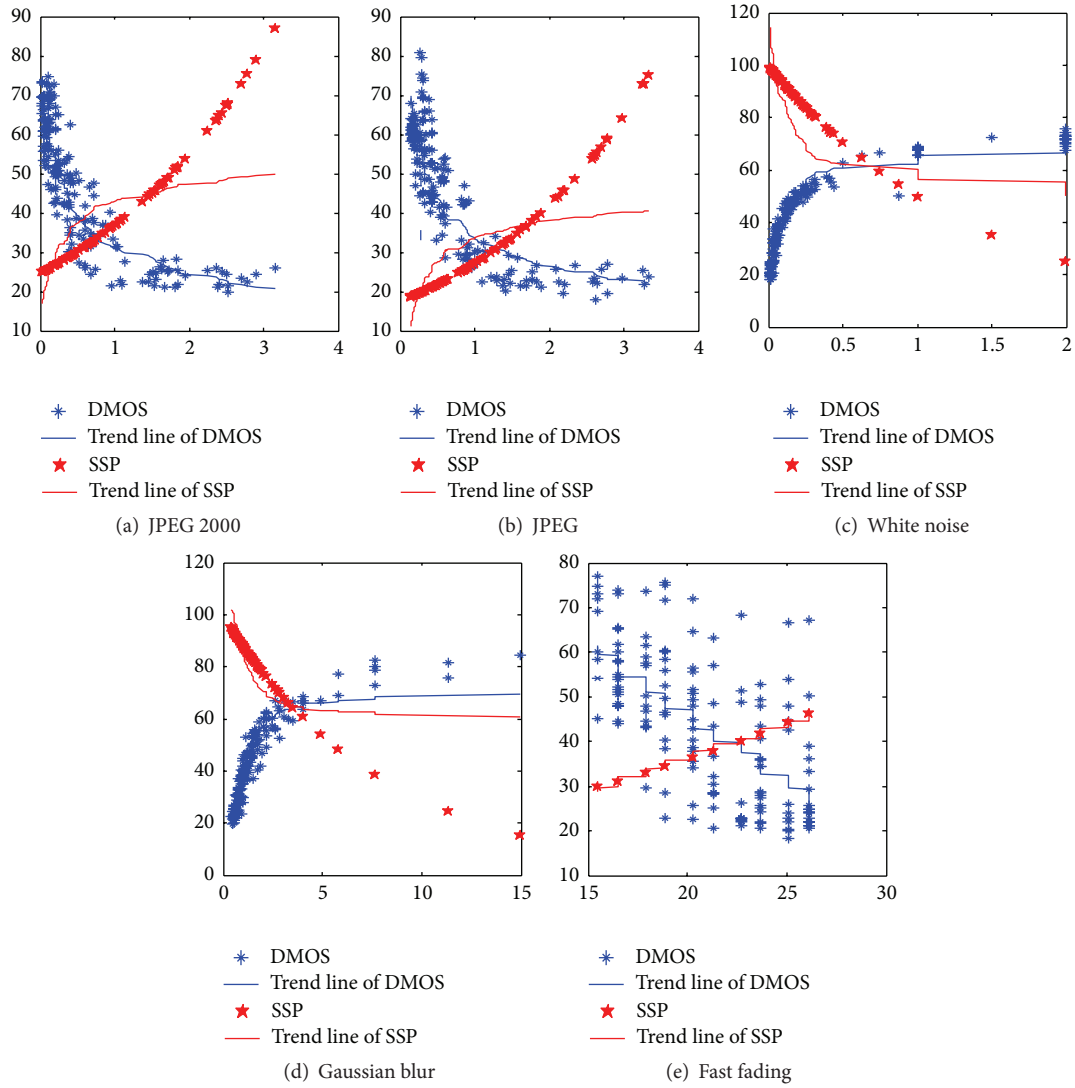


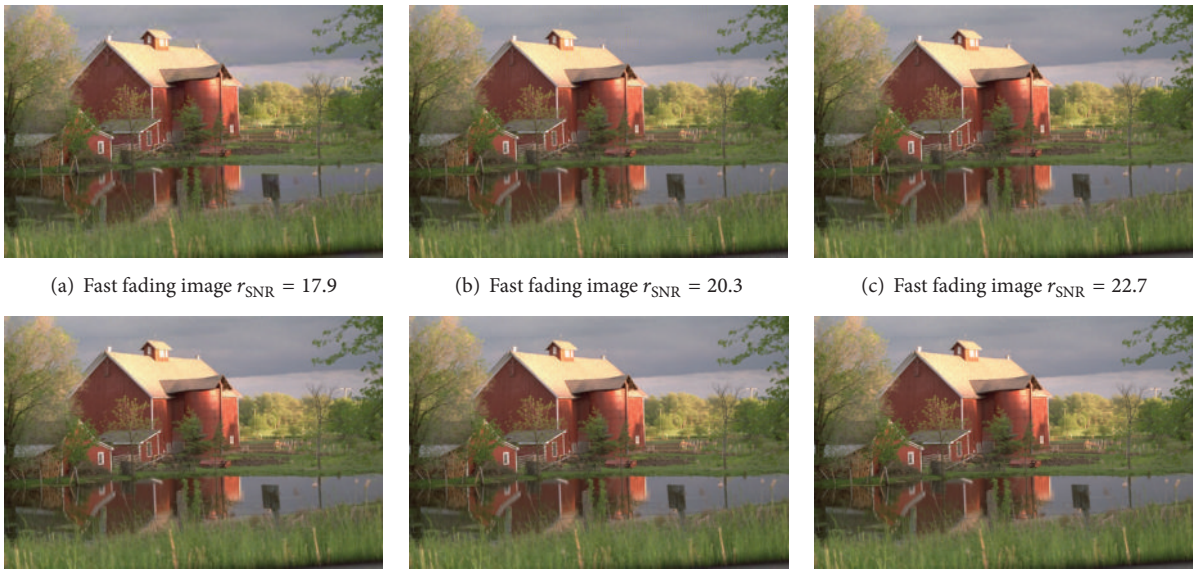
Figure 2: The scores and trend lines of DMOS and SSP for different distortion categories.

Table 3: DMOS and SSP values for the distorted images in Figure 3.

Items	Subfigures in Figure 3				
	Figure 3(b)	Figure 3(c)	Figure 3(d)	Figure 3(e)	Figure 3(f)
Distortion		Fast fading		JPEG	JPEG 2000
Strength	17.9	20.3	22.7	1.8851	1.8156
DMOS	29.5508	37.7149	22.8825	21.0245	21.2732
SSP	33.0009	36.4053	40.1610	39.7771	50.8805

for SSP values in Table 3, but the DMOS has almost the same value.

Experiment 3 (the validation for IQA methods). The ground truth of image quality is always used to evaluate the performance of IQA algorithms. In this experiment, we test whether SSP scores can evaluate the IQA algorithms. As mentioned before, the objective IQA methods are classified as FR, RR, and NR. Therefore, we representatively select the popular FR multiscale structural similarity (MS-SSIM) [6], the recent RR entropic differencing (RRED) [32], and NR spatial-spectral entropy-based quality (SSEQ) [33] to test on the total 634 distorted images in LIVE II dataset. Then, we calculate the PLCC and SROCC between the values of IQA algorithms and the ground truth DMOS or SSP scores. Table 4 records the average values of PLCC and SROCC for each IQA index. Regardless of the positive and negative direction, the IQA algorithms of the PLCC and SROCC values are ranked in



(a) Fast fading image $r_{SNR} = 17.9$ (b) Fast fading image $r_{SNR} = 20.3$ (c) Fast fading image $r_{SNR} = 22.7$

(d) JEPG image $r_{BPP} = 1.8851$ (e) JEPG 2000 image $r_{BPP} = 1.8156$ (f) Ground truth image

Figure 3: Group of distorted images for a house image (f).

Table 4: The average and rank of PLCC and SROCC between ground truth and IQA results.

IQA algorithms	PLCC (rank)		SROCC (rank)	
	DMOS	SSP	DMOS	SSP
MS-SSIM	−0.8077 (3)	0.6640 (3)	−0.9462 (2)	0.8883 (2)
RRED	0.8797 (2)	−0.6986 (2)	0.9478 (1)	−0.8899 (1)
SSEQ	0.9178 (1)	−0.7643 (1)	0.9152 (3)	−0.8821 (3)

the parentheses following the correlation strength. It can be seen from that the IQA algorithms have the same rank evaluated by SSP scores as the DMOS scores. Thus, the SSP is applicable as ground truth to compare the performance of different IQA algorithms.

Experiment 4 (the stabilization for multiple distortions). For many open IQA datasets, the subjective studies obtain MOS or DMOS on corrupted images by only one distortion. However, the majority of images could be corrupted by multiple distortions in practical consumption [34]. It motivated us to extend our SSP as formula (6) for the ground truth of multiple distortion dataset.

We simulate the modified SSP with LIVE Multiply Distorted Image Quality Database (LIVE MD) [35] to evaluate the usability of our extended method. There exist two kinds of multiple distortions firstly blur followed by JPEG or noise. The detailed information about distortion is shown in Table 5. Referring to Table 1 of LIVE II, the fading factors are selected as 2.5, 1.7, and 3.5 for blur, JPEG, and noise, respectively. Also, the zero-distortion and zero-score parameters of blur and noise are set to be the same as Table 1 because of same quantitative indicators of distortion strengths, while the JPEG with different distortion indicator are set according to the parameter setting limitations of Matlab *imwrite* function.

To evaluate the stabilization of our SSP method, the LIVE MD dataset is partitioned into 15 groups with the same combination of two distortion parameters on 15 source images. Tables 6 and 7 give the mean and deviation values of DMOS and SSP (SSP1: the reference score is the MOS of source images; SSP2: the reference score is 100) for blur + JPEG and blur + noise, respectively. Compared with DMOS, the deviation of SSP is less considering the changes of MOS values for source images. This cannot be avoided in subjective experiments, since the observers are always influenced by the contents of the images [35]. In contrast, the SSP from the real distortion parameters is objective and stable, which can be easily observed from SSP2 in Tables 6 and 7.

More directly, Figure 4 records the mean scores of 100-DMOS and SSP. It demonstrates that the 100-DMOS values are almost the same for blur + JPEG and blur + noise, while the larger differences appear in level 0 of JPEG and Noise. For the proposed SSP, although it has small values for blur + JPEG owing to large parameter p_0 of JPEG distortion, it is effective in distinguishing the qualities of different distortion levels, integrating the multiple distortions, and keeping constant at same distortion degree.

5. Conclusion

A novel metric SSP for distorted image quality has been proposed in this paper, which is derived from distortion information based on an exponent prototype of human visual model. Tested on LIVE II dataset, the SSP shows strong correlation with the distortion strength, effective consistency with the subjective score DMOS, and reasonable ground truth to evaluate the IQA algorithms. Moreover, the proposed SSP has more stabilization than DMOS to handle the combination of the multiple distortions on LIVE MD dataset. Therefore,

Table 5: The distortion information in the LIVE MD Database.

Distortion level	First distortion	Second distortion	
	Gaussian blur (σ_B)	JPEG (Q)	Noise (σ_N)
0	0	—	—
1	3.2	27	0.0447
2	3.9	18	0.0894
3	4.6	12	0.1789
p_0	*0*	*100*	*0*
p_t	*20*	*0*	*5*
k	*2.5*	*1.7*	*3.5*

Table 6: The mean (deviation) statistics of DMOS and SSP scores for blur + JPEG.

Number	Blur level	SSP1	SSP2	JPEG level	DMOS	SSP1	SSP2
1				1	11.16 (4.12)	25.88 (1.26)	28.91 (0)
2	0	89.51 (4.36)	100 (0)	2	16.89 (4.52)	22.21 (1.08)	24.81 (0)
3				3	25.08 (8.35)	20.05 (0.98)	22.40 (0)
4				0	34.85 (9.16)	60.00 (2.93)	67.03 (0)
5	1	60.00 (2.93)	67.03 (0)	1	40.23 (7.07)	17.35 (0.85)	19.38 (0)
6				2	42.97 (7.13)	14.88 (0.73)	16.63 (0)
7				3	48.86 (8.15)	13.44 (0.66)	15.02 (0)
8				0	52.81 (8.69)	54.97 (2.68)	61.42 (0)
9	2	54.97 (2.68)	61.42 (0)	1	54.36 (6.39)	15.89 (0.77)	17.76 (0)
10				2	56.52 (6.93)	13.64 (0.67)	15.24 (0)
11				3	59.64 (7.55)	12.32 (0.6)	13.76 (0)
12				0	65.11 (7.25)	50.37 (2.46)	56.27 (0)
13	3	50.37 (2.46)	56.27 (0)	1	65.2 (6.34)	14.56 (0.71)	16.27 (0)
14				2	65.68 (7.03)	12.49 (0.61)	13.96 (0)
15				3	69.58 (6.71)	11.28 (0.55)	12.61 (0)
Average		61.99 (3.02)	69.26 (0)		47.26 (7.03)	23.96 (1.17)	26.76 (0)

The source images: mean (DMOS) is 89.508; deviation is 4.3648.
SSP1: using the MOS values in LIVE MD dataset as reference scores S_r in formula (6).
SSP2: setting reference scores S_r in formula (6) as 100.

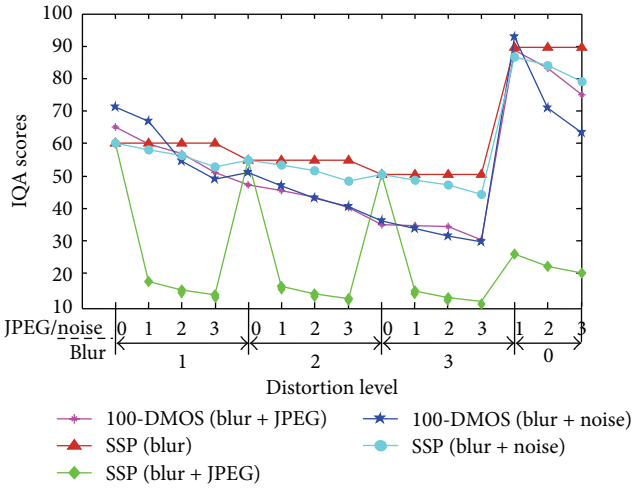


Figure 4: The average scores of SSP and 100-DMOS for LIVE MD database.

Table 7: The mean (deviation) statistics of DMOS and SSP scores for blur + noise.

Number	Blur level	SSP1	SSP2	Noise level	DMOS	SSP1	SSP2
1	0	89.51 (4.36)	100 (0)	1	7.18 (3.26)	86.75 (4.23)	96.92 (0)
2				2	29.14 (5.97)	84.08 (4.1)	93.93 (0)
3				3	36.55 (6.42)	78.97 (3.85)	88.23 (0)
4	1	60.00 (2.93)	67.03 (0)	0	28.91 (8.11)	60.00 (2.93)	67.03 (0)
5				1	33.09 (7.17)	58.15 (2.84)	64.97 (0)
6				2	45.46 (7.12)	56.36 (2.75)	62.97 (0)
7				3	50.93 (5.6)	52.94 (2.58)	59.14 (0)
8	2	54.97 (2.68)	61.42 (0)	0	49.04 (8.95)	54.97 (2.68)	61.42 (0)
9				1	52.89 (7.92)	53.28 (2.6)	59.52 (0)
10				2	56.86 (7.12)	51.64 (2.52)	57.69 (0)
11				3	59.55 (6.46)	48.5 (2.37)	54.19 (0)
12	3	50.37 (2.46)	56.27 (0)	0	63.77 (7.5)	50.37 (2.46)	56.27 (0)
13				1	66.33 (6.92)	48.82 (2.38)	54.54 (0)
14				2	68.52 (7.86)	47.31 (2.31)	52.86 (0)
15				3	70.23 (6.39)	44.44 (2.17)	49.65 (0)
Average		61.99 (3.02)	69.26 (0)		47.9 (6.85)	58.44 (2.85)	65.29 (0)

The source images: mean (DMOS) is 89.508; deviation is 4.3648.
SSP1: using the MOS values in LIVE MD dataset as reference scores S_r in formula (6).
SSP2: setting reference scores S_r in formula (6) as 100.

it is feasible to integrate the proposed SSP into IQA dataset to not only directly generate the ground truth of various images but also improve the subjective experiments with more distortion categories.

Competing Interests

The authors declare that there are no competing interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 61201315) and the Fundamental Research Funds for the Central Universities (YWF-13-T-RSC-089 and YWF-14-YHXY-13).

References

[1] F. Russo, "Automatic enhancement of noisy images using objective evaluation of image quality," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1600–1606, 2005.

[2] W. Zou, J. Song, and F. Yang, "Perceived image quality on mobile phones with different screen resolution," *Mobile Information Systems*, vol. 2016, Article ID 9621925, 17 pages, 2016.

[3] A. C. Bovik and Z. Wang, *Modern Image Quality Assessment*, Morgan and Claypool, New York, NY, USA, 2006.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 2929–2932, October 2006.

[6] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proceedings of the Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398–1402, IEEE, November 2003.

[7] F. Zhou, Z. Lu, C. Wang, W. Sun, S.-T. Xia, and Q. Liao, "Image quality assessment based on inter-patch and intra-patch similarity," *PLoS ONE*, vol. 10, no. 3, Article ID e0116312, 2015.

[8] H. Jiang, K. Yang, T. Liu, and Y. Zhang, "Quality prediction of DWT-based compression for remote sensing image using multiscale and multilevel differences assessment metric," *Mathematical Problems in Engineering*, vol. 2014, Article ID 593213, 15 pages, 2014.

[9] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 305–312, Providence, RI, USA, June 2011.

[10] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[11] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *IS and T Electronic Imaging—Human Vision and Electronic Imaging X*, Proceedings of SPIE, pp. 149–159, International Society for Optics and Photonics, January 2005.

[12] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 202–211, 2009.

[13] A. De Angelis, A. Moschitta, F. Russo, and P. Carbone, "A vector approach for image quality assessment and some metrological considerations," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 14–25, 2009.

[14] LIVE Image Quality Assessment Database, http://live.ece.utexas.edu/research/quality/subjective.htm.

[15] D. M. Chandler and S. S. Hemami, "Online supplement to 'VSNR: a visual signal-to-noise ratio for natural images based on near-threshold and suprathreshold vision'," 2007, https://www.researchgate.net/publication/267418084_Online_Supplement_to_VSNR_A_Visual_Signal-to-Noise_Ratio_for_Natural_Images_Based_on_Near-Threshold_and_Suprathreshold_Vision.

[16] 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. 1, Article ID 011006, 21 pages, 2010.

[17] IVC Image Quality Database, http://www2.irccyn.ec-nantes.fr/ivcdb.

[18] Tampere Image Database, http://www.ponomarenko.info/tid2008.htm.

[19] Y. Horita, K. Shibata, and Y. Kawayoka, "Toyama Image quality evaluation database," https://www.researchgate.net/publication/221678019_Toyama_Image_quality_evaluation_database.

[20] J. Redi, H. Liu, H. Alers, R. Zunino, and I. Heynderickx, "Comparing subjective image quality measurement methods for the creation of public databases," in *Image Quality and System Performance VII, 752903*, vol. 7529 of *Proceedings of SPIE*, International Society for Optics and Photonics, January 2010.

[21] S. Kaya, M. Milanova, J. Talburt, B. Tsou, and M. Altynova, "Subjective image quality prediction based on neural network," in *Proceedings of the 16th International Conference on Information Quality (ICIQ '11)*, Lecture Notes, pp. 259–266, 2011.

[22] Y. Lu, F. Xie, Z. Jiang, and R. Meng, "Objective method to provide ground truth for IQA research," *Electronics Letters*, vol. 49, no. 16, pp. 987–989, 2013.

[23] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment," *Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, 2011.

[24] T. J. Liu, W. Lin, and C. C. J. Kuo, "Image quality assessment using multi-method fusion," *Transactions on Image Processing*, vol. 22, no. 5, pp. 1793–1807, 2013.

[25] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 41–52, 2012.

[26] M. Livingstone and D. Hubel, "Segregation of form, color, movement, and depth: anatomy, physiology, and perception," *Science*, vol. 240, no. 4853, pp. 740–749, 1988.

[27] S. H. Kim and J. P. Allebach, "Impact of HVS models on model-based halftoning," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 258–269, 2002.

[28] B. W. Wu, Y. C. Fang, and L. S. Chang, "Study on human vision model of the multi-parameter correction factor," in *MIPPR 2009: Pattern Recognition and Computer Vision, 74960E*, vol. 7496 of *Proceedings of SPIE*, International Society for Optics and Photonics, October 2009.

[29] C. Shi, K. Xu, J. Peng, and L. Ren, "Architecture of vision enhancement system for maritime search and rescue," in *Proceedings of the 8th International Conference on Intelligent Transport System Telecommunications (ITST '08)*, pp. 12–17, Phuket, Thailand, October 2008.

[30] D. C. Knill and A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation," *Trends in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.

[31] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[32] R. Soundararajan and A. C. Bovik, "RRED indices: reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[33] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[34] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proceedings of the IEEE Conference Record of the 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR '12)*, pp. 1693–1697, Pacific Grove, Calif, USA, November 2012.

[35] LIVE Multiply Distorted Image Quality Database, http://live.ece.utexas.edu/research/quality/live_multidistortedimage.html.