

No-Reference Image Quality Assessment through SIFT Intensity

Tongfeng Sun^{1,*}, Shifei Ding^{1,2} and Xinzheng Xu¹

¹ School of Computer Science and Technology, China University of Mining and Technology, Xuzhou Jiangsu 221116, China

² Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Science, Beijing 100190, China

Received: 20 Aug. 2013, Revised: 22 Nov. 2013, Accepted: 23 Nov. 2013

Published online: 1 Jul. 2014

Abstract: SIFT (Scale Invariant Feature Transform) points are scale-space extreme points, representing local minutiae structure features in the Gaussian scale space. SIFT intensity, as a novel no-reference metric, is feasible to assess various common distortions without the access to reference images. The metric introduces image preprocessing: neighborhood enhancement through contrast enhancement of adjacent pixels to reduce false SIFT points triggered by random signals; double-size image magnification through linear interpolation to amplify distortion effects to improve its sensitivity to image quality. SIFT intensity is defined as the number of SIFT points in a unit region and is calculated based on the first octave of the difference-of-Gaussian scale space. Experimental results demonstrate that SIFT intensity is superior to existing classic no-reference metrics and can be used to assess different distortions.

Keywords: Scale Invariant Feature Transform, minutiae structure features, SIFT intensity, neighborhood enhancement

1 Introduction

People often need some metrics to assess image quality in the fields of image preprocessing, fusion, transmission, etc. Image quality metrics can be categorized into two groups: subjective metrics and objective metrics. Subjective metrics are simple and accurate. But they can not work without human's participations. Furthermore, it is infeasible in the case that there are a mass of images. Therefore, it is indispensable to attach more importance to objective metrics.

Objective metrics are divided into three categories: full-reference (FR) metrics, reduced-reference (RF) metrics and no-reference (NR) metrics. FR metrics are based on the following assumptions: the reference image is an image with the highest quality and is available to access. A test image quality is assessed by calculating the similarity between the test image and its reference image. Common FR metrics include root mean square error (RMSE), peak signal to noise ratio (PNSR), mutual information (MI), etc. In recent years, a multitude of efforts have been made to develop new objective image/video quality metrics that consider human visual system (HVS) (actually, HVS is also widely investigated

in RF and NR metrics besides FR metrics). There are a variety of HVS models, each of which can model parts of human vision (e.g. spatial resolution, temporal motion, color fidelity, color resolution, contrast and orientation sensitivity, frequency selectivity) [1,2,3,4,5,6]. By integrating with some HVS characteristics, the performance of some present metrics can be improved. For instances, in [5], visual signal-to-noise ratio (VSNR), based on near-threshold and supra threshold properties of human vision, is presented to quantify the visual fidelity of natural images. In [6], sparse coding statistical models of natural images are used to simulate HVS. Based on the models, a sparse correlation coefficient (SCC) between two visual signals of images in a cortical visual space is investigated to assess image quality. Under the assumption that human visual system is highly adapted for extracting structural information, Structural Similarity (SSIM) has been growingly developed as a category to estimate image quality via the measurement of the retention of structural information in a distorted image with regard to its reference image [7,8,9]. In [7], Wang, et al. develop a structural similarity index for image quality measurement, modeling image degradation as the loss of structural information. It has proved that SSIM is

* Corresponding author e-mail: stfok@126.com

more efficient than MSE or PSNR. In [8], Santiago et al. propose image quality assessment based on local variance to improve SSIM's performance. In [9], modified versions of (G-)SSIM and MS-(G-)SSIM, called four-component (G-)SSIM and four-component MS-(G-)SSIM, classify image local regions according to edge and smoothness properties, determine the weights of SSIM scores by region types, and apply weighted SSIM scores to assess blurred and noisy images.

RF metrics only need access partial reference information instead of full reference images. Partial reference information should have been extracted in advance or is available through an ancillary data channel [11, 12, 13, 14]. Reference features are extracted from the reference image/video sequence and these are then compared with the same features extracted from the distorted images to obtain RR quality. The challenging task in the design of a RR metric is the selection of the smallest set of features that can support quality assessment effectively. In [12], a hybrid image quality metric (HIQM) combines five structural features: blocking, blur, edge-based image activity, gradient-based image activity and intensity masking. The metric makes a similarity assessment between the test image and the reference image to assess JPEG coded images. In [13], the differences of wavelet coefficients between two images based on coefficient histograms are used to assess distorted images. In [14], the Circular-ELM (C-ELM), an augmented version of the basic Extreme Learning Machine (ELM), defines a meaningful feature-based representation of the visual signal and handles the actual mapping of visual signals into quality scores.

NR metrics are not relative to reference images. From the view of applications, NR metrics are more adaptable to different conditions. Common NR metrics include variance metric, auto-correlation metric, frequency threshold metric, histogram threshold metric, histogram frequency metric, generalized block-edge impairment metrics, etc. [15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Distortions can be mainly classified into noise, blur, blocking artifact (including blocking blur). Many NR metrics can only measure some types of distortions, especially feasible for blocking or blur [15, 16, 17, 18, 19, 20, 21, 22]. In [23], a natural scene statistics (NSS) model of contourlet coefficients adopts an image-dependent threshold to assess different distortions, but its accuracy still need to be further improved and it needs a training database to determine proper related parameters of the approach. In [24], under the assumption that the images are affected by a single distortion, SA-based (structural-activity based) framework is proposed to assess different distortions with the prior knowledge of the distortion, but the framework also need a training database and is sensitive to the chosen SA indicator and SA weights. In [25], anisotropy entropy is proposed to assess noise, blur and blocking artifact. But for the images affected by small-scale smoothing, large noise or blocking artifact, the metric doesn't work well. Although

its improved algorithm can improve assessments on blocking artifact, its deficiencies aren't yet completely overcome. An ideal NR metric can be applied to assess various distortions without the prior knowledge of image distortion types. In the latest years, more efforts are made on the improved metrics based on traditional basic metrics through intelligent training on a sample database [23, 24, 26]. On the other hand, we still need to probe into novel basic NR metrics to provide more accurate and reliable assessment under certain application conditions.

In fact, NR metrics are often used in the case that each one of the images to be assessed is mainly corrupted by a distortion. So, under the assumption that each test image is affected by a distortion (the distortion types in test images may be different from each other), this paper proposes a novel NR metric through SIFT intensity to evaluate different distortions without the prior knowledge of distortion types and the training database. SIFT points are scale-space extreme points, widely used in image registration, but rarely in NR assessment. The richer the image textures are, the more SIFT points the image has. After proper preprocessing, the number of SIFT points changes with distortions, reflecting image quality. The rest of this paper is organized as follows. Section 2 introduces SIFT technology, investigates the relationship between the number of SIFT points and the octave in scale space, and introduces a concept of SIFT intensity. Section 3 discusses image preprocessing and proves into the relationships between SIFT intensity and distortions, such as noise, blur and blocking artifact. In Section 4, experiments are carried out to verify the performance of SIFT intensity compared with some classic basic NR metrics, including spatial frequency (SF), generalized block-edge impairment metric (GBIM), entropy, anisotropy entropy (AE). Finally, conclusions and discussions are given in Section 5.

2 SIFT intensity

2.1 SIFT technology

SIFT (Scale Invariant Feature Transform) are insensitive to light, rotation, scale, noise, etc., representing image minutiae structure features. To efficiently detect stable keypoints in scale space, Lowe proposed scale-space extrema in the difference-of-Gaussian scale space $D(x, y, \sigma)$ [27], which can be computed from the difference of two adjacent Gaussian-scale images separated by a multiplicative constant factor k :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (1)$$

where $*$ is the convolution operation and $L(x, y, \sigma)$ represents the Gaussian-scale-space image. $L(x, y, \sigma)$ is produced from the convolution of a variable-scale Gaussian function $G(x, y, \sigma)$ and an input image $I(x, y)$.

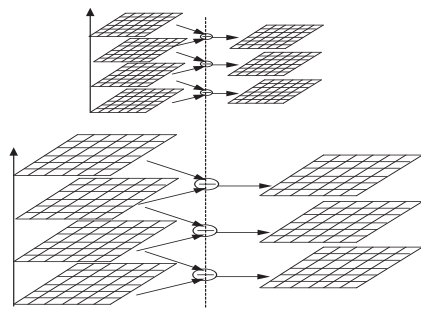


Fig. 1: Gaussian scale space and difference-of-Gaussian scale space

To detect more scale-space extreme points, Gaussian pyramid and difference-of-Gaussian pyramid are built, consisting of some octaves, each of which is further composed of some levels. In order to get more features, a test image is commonly twice magnified in size (double-size magnification) by linear interpolation as the first level of the first octave. The magnified image is repeatedly convolved with Gaussian functions, producing a set of scale-space images as shown on the left of dotted line in Fig. 1. Adjacent Gaussian images are subtracted to produce difference-of-Gaussian images on the right of dotted line. Then, the magnified image is down-sampled by 2 as the first level of the second octave. And Gaussian images and DOG images in the second octave are constructed similarly to those in the first octave. Above processings are repeated until image size is lower than a certain threshold. More details about SIFT are discussed in [27].

2.2 Relationship between the number of SIFT points and the octave in the Gaussian scale space

For three common types of distortions: noise, blur and blocking artifact, noise is inevitable, which may be brought about when shot and transferred; blur is often caused by different focuses, which can be simulated by regional image smoothing; blocking is mainly caused by compression, in which quantization introduces quantization errors.

In the implementation of SIFT, the scales are represented as different octaves and levels through down-sampling and Gaussian smoothing. The smallest scale is corresponding to the first level of first octave and the largest scale is corresponding to the last level of last octave in Gaussian pyramid. To keep the scale's continuity, k in equation (1) is defined as

$$k = 2^{1/m} \tag{2}$$

where m is the number of SIFT scale levels in each octave of the difference-of-Gaussian pyramid. Supposing that m

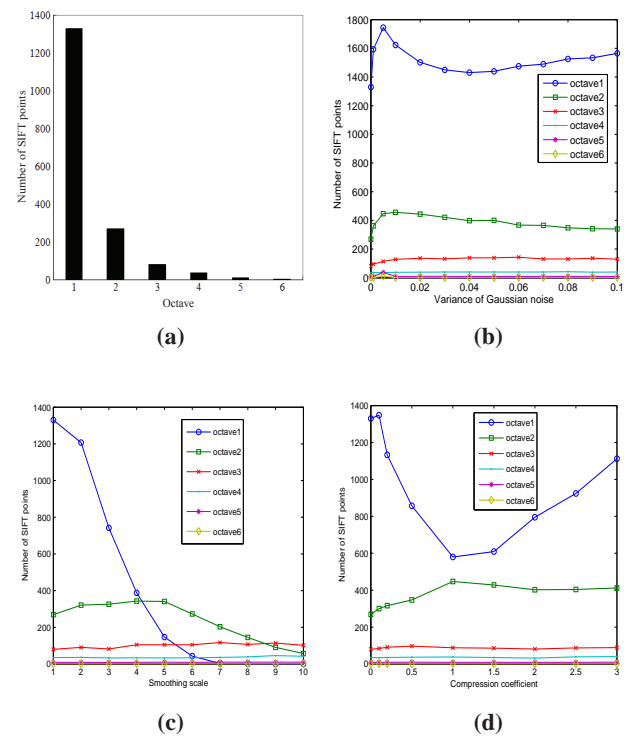


Fig. 2: Relationship between the number of SIFT points and the octave in the Gaussian scale space

is a predefined constant, $m + 2$ levels in a octave of the difference-of-Gaussian pyramid should be constructed because there is no SIFT point in the first and last level. And $m + 3$ levels in a octave of Gaussian pyramid should be constructed to build $m + 2$ levels in a octave of difference-of-Gaussian pyramid.

Standard images, including Lena, Peppers, Baboon and Barbara, are chosen as test images to investigate the effects of different distortions in different octaves. We choose the number of SIFT points as image feature to depict distortion's effect. Because the experiments on the test images show similar results, we only show the results of Lena image to keep the paper concise. Here, difference-of-Gaussian pyramid is composed of six octaves (each octave consisting of 5 levels and m equal to 3). As shown in Fig. 2, Fig. 2(a) indicates that the number of SIFT points in the first octave nearly accounts for 77% of the total SIFT points and drops down quickly with the octave increasing.

Fig. 2(b) shows the relationship between the number of SIFT points and noise in different octaves. In experiments, different Gaussian noise, whose mean is zero and variance changes from 0.00001 to 0.1, is added to the test image. Because noise is random signal, experimental result is the average of 10 experiments to ensure its validity. Distinctly, although the number of SIFT points in the first octave can't

assess image quality, SIFT feature in the first octave is the most active among different octaves.

Fig. 2(c) shows the relationship between the number of SIFT points and blur in different octaves. Image smoothing will result in the loss of image minutiae, making the image blurry. The greater the smoothing scale is, the blurrier the image is. It can be used for the simulations of multi-focus images. The Gaussian filtering sliding window is a kind of common image smoothing, whose pixels' weights in the smoothing window comply with two-dimensional Gaussian distribution. Gaussian filtering windows ($n \times n$, n ranging from 1 to 10) are chosen to probe into the relationship between the number of SIFT points and smoothing. It can be seen that the number of SIFT points in the first octave drops down with the smoothing scale increasing and can be used to assess blur distortion.

Fig. 2(d) shows the relationship between SIFT points and blocking artifact in different octaves. Image compression can effectively reduce required storage space and bandwidth in data storage and communication. Block compression algorithms, such as block-based discrete cosine transform (BDCT) and block-based discrete wavelet transform (BDWT), has a high compression efficiency and low computational complexity. They are core compression algorithms in most image/video compression standards, such as JPEG, JPEG 2000, MPEG, H.263. Due to separate processes of different image blocks at encoder and decoder sides, quantization error will be superposed, destroy correlations and continuities among image blocks and cause blocking artifact. Image compression experiments choose BDCT compression. The quantization matrix is $L \times \alpha$, where L refers to a luminance quantization matrix for JPEG, as shown in Table 1, and α is a coefficient ranging from 0.2 to 2. So, the relationship between the number of SIFT points and blocking artifact is transformed into the relationship between the number of SIFT points and α . Similarly to Fig 2(b), although SIFT features in the first octave can't assess blocking artifact, SIFT feature in the first octave is still the most active among different octaves.

The experiments show that distortions play different effects in different scale space and the number of SIFT points in the first octave is the most sensitive to different distortions. So, SIFT detection is based on double-size images convolved with different Gaussian scale smoothing in this paper.

2.3 SIFT intensity

SIFT points, in the first octaves of the difference-of-Gaussian scale space, reflect image minutiae: scale-space structural information of adjacent pixels (extreme points in scale space). As human visual system is sensitive for structural information, structural information has been widely used to assess image quality

Table 1: Luminance quantization matrix for JPEG

luminance quantization table							
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
79	92	95	98	112	100	103	99

in various metrics [7, 8, 9]. The more SIFT points are, the more structural information and higher quality the image has. So, SIFT intensity is introduced to assess image quality in this paper. It is defined as the number of SIFT points in a unit region, represented as:

$$s = 1/w \sum_{i=1}^w s_i \quad (3)$$

It describes the minutiae in certain regions, whose size and number depend on actual applications. If image size is not large, the entire image is chosen as a unit to calculate SIFT intensity and w is equal to 1. Otherwise, the image can be divided into several equal regions, SIFT points are detected respectively, and w is equal to the actual number of the regions. In this way, computational efficiency can be improved and required memory space be reduced. Furthermore, each pixel based on local region can be assessed to get local quality, which can be used in image local processing.

3 Image quality assessment through SIFT intensity

We hope the metric can accurately assess noise, blur and blocking, even though the distortion type is unknown in advance. Because noise assessment is the most difficult among these factors, noise assessment is a crucial problem for the metric. We still take Lena image as an example. Because its size is not large (512×512), SIFT intensity is calculated based on the whole image. All the experiments are carried out similarly to section 2.2. In the experiments for investigating the relationship between SIFT intensity and noise, the mean of noise is zero and the variance of noise changes from 0 to 0.1; in the experiments for investigating the relationship between SIFT intensity and smoothing, the Gaussian smoothing window scale n changes from 1 to 10; in the experiments for investigating the relationship between SIFT intensity and compression, the BDCT quantization coefficient α changes from 0.2 to 3.

3.1 Image preprocessing and relationship between SIFT intensity and noise

SIFT points are scale-space extreme points instead of gray extreme points. Noise, as random signal, can not only destroy original image minutiae, but also cause some false SIFT points, sometimes leading to the increase of SIFT intensity.

Because noise signals of adjacent pixels are unrelated to each other, a neighborhood enhancement filter can increase gray difference between adjacent pixels, causing a lot of noise-gray extreme points and reducing false SIFT points. Although neighborhood enhancement can also destroy the texture relationships of original images, the number of the original SIFT points are slightly suppressed from the view of statistics compared with that of false SIFT points caused by noise, due to the strong correlation of original image pixels.

In the paper, image enhancement adopts a Laplacian operator to enhance isolated points and thin lines, represented as in equation (4),

$$H_l = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (4)$$

Laplacian operator is a second-derivative operator, which is sensitive to isolated points and thin lines. The enhanced image is implemented through $I(x,y) - \eta \times I_l(x,y)$ (η is a fitting parameter, and $I_l(x,y)$ is produced from the convolution of H_l and an image $I(x,y)$), so the enhancement matrix is represented as:

$$H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \eta \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (5)$$

Through experiments over the LIVE database [28], enhancement matrix H as equation (6) ($\eta=0.09$) can achieve satisfying results, making SIFT intensity accurately depict image quality:

$$H = \begin{bmatrix} -0.09 & -0.09 & -0.09 \\ -0.09 & 1.72 & -0.09 \\ -0.09 & -0.09 & -0.09 \end{bmatrix} \quad (6)$$

The relationship between SIFT intensity and noise is shown in Fig. 3. Fig. 3(a) shows the relationship between SIFT intensity and noise only with double-size magnification preprocessing. It can be seen that SIFT intensity experiences a course of increasing, decreasing and increasing. With noise variance less than 0.003, SIFT intensity increases quickly with noise; with noise variance in the scope [0.003 0.03], SIFT intensity gradually decreases; with noise variance greater than 0.03, SIFT intensity increases again. So the metric can't accurately assess noise in this case. Fig. 3(b) shows the relationship between SIFT intensity and noise with neighborhood

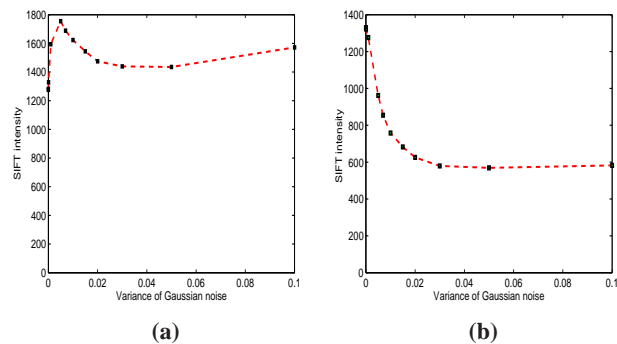


Fig. 3: Relationship between SIFT intensity and noise (a) The images only with double-size magnification (b) The images with neighborhood enhancement and double-size magnification

enhancement and double-size magnification preprocessing. SIFT intensity gradually decreases with the noise increasing in variance scope [0 0.03] and nearly keeps steady in variance scope [0.03 0.1]. Because noise variance is usually lower than 0.03, SIFT intensity can assess the image corrupted by noise.

In most cases, we don't know distortion types in advance. In order to make SIFT intensity more extensively adaptive and sensitive, image assessment based on SIFT intensity in the following experiments introduces neighborhood enhancement and double-size magnification preprocessing.

3.2 Relationship between SIFT intensity and smoothing

Fig. 4 represents the relationship between SIFT intensity and smoothing after the preprocessing of neighborhood enhancement and double-size magnification. Because each image when shot is inevitably subject to random outside interferences, small-scale smoothing may play a role of a noise filter, which can improve image quality. So as shown in Fig. 4, when n is equal to 2, SIFT intensity increases. Then it successively drops down to below 40. Experiments also show that SIFT intensity is sensitive when $n \in [1 10]$ and is close to zero when $n > 10$. According to the trend of SIFT intensity, it can assess images corrupted by smoothing.

3.3 Relationship between SIFT intensity and blocking artifact

Fig. 5 shows the relationship between SIFT intensity and blocking artifact after the preprocessing of neighborhood enhancement and double-size magnification. It can be

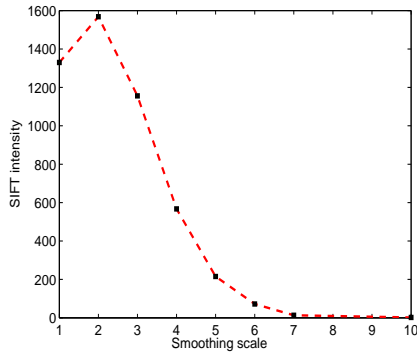


Fig. 4: Relationship between SIFT intensity and smoothing

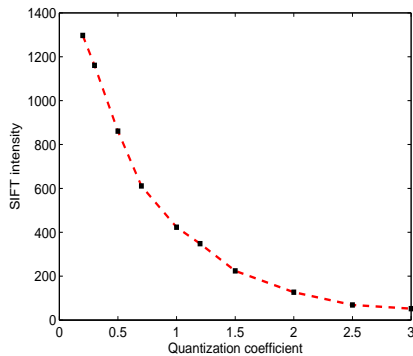


Fig. 5: Relationship between SIFT intensity and compression coefficients

seen that SIFT intensity decreases gradually from maximum 1300 to minimum below 100 with α increasing. The more obvious the blocking artifact is, the lower the SIFT intensity is. So, SIFT intensity can assess images corrupted by blocking artifact.

3.4 Image quality assessment through SIFT intensity

According to above analyses, the higher SIFT intensity is, the higher image quality is. According to Fig. 3,4,5, we can infer that SIFT intensity can assess noise, blur and blocking artifact even without the prior knowledge of distortion types. To make the metric concise, sensitive and reliable, image assessment through SIFT intensity is summarized as follows:

- (1) to preprocess images by neighborhood enhancement and double-size magnification (four times as large as the acreage of original image);
- (2) to calculate image SIFT intensity;
- (3) to assess image quality according to SIFT intensity.

The unit region in SIFT intensity is chosen according to actual applications. It can be a image or a local region. Image magnification can make SIFT intensity more sensitive to distortions, improving image assessment. But it doesn't work well for high noise and large smoothing scales, such as the Gaussian noise whose mean is zero and variance is greater than 0.1, the smoothing window whose scale n is greater than 10. In these cases, SIFT intensity is usually fixed within a certain small scope. In the cases, SIFT intensity of the original image with neighborhood enhancement preprocessing except double-size magnification can properly assess the noise with its variance ranging from 0.1 to 0.4 (its mean is zero) as well as the smoothing scale ranging from 10 to 15, but its sensitivity will be weakened.

4 Experiments

In this section, experiments are carried out to explore the performances of SIFT intensity. Because our goal is to investigate a novel basic NR metric, we choose some existing classic basic NR metrics for comparison to investigate the relationship between each metric and subjective assessment, as well as the relationships between each metric and distortion scale. The chosen metrics include SF, GBIM [18], entropy and AE [25], instead of the ones which are partly improved based on a basic metric, intelligently trained through a sample database, or designed based on multiple basic metrics.

4.1 Comparison of NR metrics based on subjective assessment

The experiments were performed on the LIVE database [28]. This database contains 29 high-resolution 24bits/pixel RGB color images as reference and five types of corresponding distorted images: JPEG, JPEG2000, White Noisy (WN), Gaussian blurred (Gblur) and Fast-Fading (FF) Rayleigh channel noisy images. Subjective quality scores are given in terms of the realigned difference mean opinion score (DMOS) on a scale of 0-100 with a large DMOS indicating poor visual quality. Although we only probe into the assessment of images distorted by noise, blur and blocking artifact in section 3, SIFT intensity can work well for above five distortions.

The relationship between predicted quality rating and DMOS (or MOS (mean opinion score), acquired according to DMOS [29]) can have nonlinear quality rating compression at the extremes of the test range. So, we establish the nonlinear mapping via the logistic function [29]

$$q = ax^3 + bx^2 + cx + d \quad (7)$$

where x is the objective quality score calculated by quality metrics, q is the predicted subjective quality score,

and a, b, c are the parameters. This function must be constrained to be monotonic in the range of possible values for x . We employ three statistics for evaluation after several runs of the metric: Pearson Correlation linear correlation coefficient (CC), Root-Mean Squared Error (RMSE) and Outlier Ratios (OR).

According reference images, all the images from the LIVE database are classified into 29 objects. To validate the proposed metric, we demonstrate the performance by running 15 times and choosing 25 images each time. Table 2 gives the mean and standard deviation (SD) of each criterion on the testing data for 15 runs against the MOS. SIFT intensity has the highest CC, the smallest RMSE and OR with relatively small SD. It can be seen that SIFT intensity is the most consistent with subjective assessment, orderly followed by AE, SF, GBIM and Entropy. Furthermore, under the assumption that the images are mainly corrupted by a single distortion, the mapping in equation (7) is not necessary and SIFT intensity can assess image quality even without the prior knowledge of image distortion types.

4.2 Comparisons of NR metrics on different-scale distortions

The experiments were performed on the Million Gallery [30], which includes natural images, bird/beast images, natural scenery images, cultural wonder images, etc. We randomly choose 15 natural scenery images and the images are resized to 500×500 . They are processed by noise-adding with noise's variance σ from 0 to 0.1 (noise mean is 0), window smoothing with smoothing scale n from 1 to 10 and BDCT compressing with compression coefficient α from 0.2 to 2 (as in section 3.3). Here noise's variance σ , smoothing windows n and compression coefficient α are seen as distortion scale. And each image is taken as a unit to calculate SIFT intensity, after enhanced through neighborhood enhancement filter and twice magnified through linear interpolation. For smoothing and compressing, the assessment of each image is the result over one experiment. For noise, the assessment of each image is an average over ten experiments to ensure experimental accuracy. Experiment results are the averages over 15 images. The comparison results are demonstrated in Table 3,4,5,6, which show the assessment results for SF, GBIM, Entropy, AE and SIFT intensity with the distortions increasing. Final statistical results over 15 images are demonstrated in Table 6, which shows the mean and standard deviation (SD) of Pearson rank order correlation coefficient (ROCC) according to the relationship between predicted quality rating and distortion scale. Pearson rank order correlation coefficient can describe nonlinear correlation relationship and the coefficient ranges from -1 to 1. In the light of the fact that a NR metric can be viewed as a excellent metric as long as the metric

monotonously changes with distortion increasing, the positive linear correlation relationship is unnecessary. What's more, from Fig. 3,4,5, we can find that the relationships between SIFT intensity and distortions are nonlinear. So, ROCC is more feasible to validate SIFT intensity. And we adopt the absolute value of ROCC in Table 6.

From Table 3,4,5, we can directly view the performances of above five NR metrics. SF can, as a whole, accurately assess smoothing (blur) and is not fit for noise and blocking artifact. When smoothing scale $n = 2$, SF is lower than original SF, which is not often consistent with actual applications because a small scale filter is generally beneficial to image quality. GBIM can assess image blocking artifact, not fit for noise and smoothing. Entropy can assess large-scale smoothing, not fit for noise, small-scale smoothing and blocking artifact. AE is, as a whole, superior to SF, Entropy and GBIM. But it still doesn't perform well for small noise, small-scale smoothing and blocking artifact. SIFT intensity works well for noise, smoothing and blocking artifact. It drops down with distortions increasing. When noise's variance is less than 0.0001, SIFT intensity is near to that of reference image. In the case, subjective assessment is also difficult to give an accurate judge. Therefore, the problem doesn't actually affect the approach's performances. When smoothing scale is 2, the increase of SIFT intensity is in line with the objective reality, which is because images, when shot from CCD devices, will be inevitably subject to outside interferences and a small window smoothing will help improve image quality.

From Table 6, we can view the statistical performances of above five NR metrics. If ROCC is near to 1, it indicates a high correlation. So, we choose ROCC to validate SIFT intensity. According to the mean and SD of ROCC, SIFT intensity keeps the highest performances with the scale of white noise, Gaussian smoothing and compression changing. SF is insufficient in the assessments of white noise and compression (blocking artifacts), GBIM insufficient in the assessments of white noise and Gaussian smoothing, Entropy insufficient in the assessments of all the three chosen distortions, and AE insufficient in the assessment of compression.

Based on above analyses, it can be inferred that SIFT intensity can perform well in the assessments of different distortions and is superior to SF, GBIM, Entropy and AE.

5 Conclusions and discussions

Under the assumption that images are mainly affected by a single distortion, SIFT intensity can accurately assess the distortions without the prior knowledge of distortion type and the training database. Experiments show that the metric can accurately assess different distortions, decreasing with noise, blur and blocking artifact increasing. Compared with SF, GBIM, entropy, AE, etc.,

Table 2: Performance of metrics on subjective assessment

	SF	GBIM	Entropy	AE	SIFT intensity
CC					
mean	0.3725	0.3848	0.1523	0.8214	0.9412
SD	0.1868	0.1491	0.1036	0.0315	0.0217
RMSE					
mean	26.15	25.72	25.36	10.61	6.821
SD	6.218	5.016	8.083	0.986	0.892
OR					
mean	0.7586	0.7137	0.7664	0.5131	0.3108
SD	0.1621	0.1391	0.1643	0.0905	0.0912

Table 3: Relationship between image quality assessment and noise's variance σ (noise's mean is 0)

	0	0.0001	0.001	0.005	0.007	0.01	0.015	0.02	0.03	0.05	0.1
SF	7.6763	7.7034	7.7245	7.7911	7.8564	7.8421	7.7539	7.8873	8.0638	8.4236	7.5227
GBIM	2.6558	2.6600	2.6554	2.6424	2.6414	2.6386	2.6406	2.6459	2.6414	2.6414	2.6420
Entropy	5.1609	5.2906	5.2448	5.3181	5.3421	5.3662	5.3878	5.3969	5.3939	5.3410	5.1153
AE	0.0037	0.0038	0.0036	0.0029	0.0026	0.0023	0.0020	0.0018	0.0016	0.0014	0.0010
SIFT Intensity	1330	1341.8	1306.4	981.8	864.4	767.9	692.7	635.8	598.20	581.9	603.2

Table 4: Relationship between image quality assessment and smoothing scale n

	1	2	3	4	5	6	7	10
SF	7.6763	6.4420	5.2036	4.6241	4.0523	3.7483	3.4201	0.0112
GBIM	2.6558	2.6727	2.6635	2.5321	2.4673	2.4224	2.3907	2.6899
Entropy	5.1609	5.2678	5.3178	5.2968	5.3040	5.3026	5.2987	4.9258
AE	0.0037	0.0032	0.0026	0.0021	0.0016	0.0013	0.0009952	0.000497
SIFT intensity	1330	1568	1156	567	215	72	13	2

Table 5: Relationship between image quality assessment and compression coefficient α

	0.2	0.3	0.5	0.7	1.0	1.2	1.5	2.0	2.5	3.0
SF	6.8150	6.5527	6.2710	6.0062	5.7944	5.6966	5.7825	6.2903	6.7015	7.1465
GBIM	1.7102	1.4397	1.0801	0.8020	0.5937	0.4573	0.3158	0.2040	0.1359	0.0965
Entropy	5.1696	5.0966	5.1015	5.0873	5.0155	5.0211	4.9795	4.9496	4.9568	4.9342
AE	0.0034	0.0033	0.0030	0.0028	0.0025	0.0025	0.0023	0.0023	0.0024	0.0034
SIFT Intensity	1297	1161	861	611	423	348	224	127	69	52

Table 6: ROCC correlation relationship between the metric and distortion scale

	SF	GBIM	Entropy	AE	SIFT intensity
White noise					
Mean of ROCC	0.0486	0.1192	0.1709	0.8690	0.8718
SD of of ROCC	0.0589	0.0278	0.1091	0.0621	0.0623
Gaussian Smoothing					
Mean of ROCC	0.9708	0.0728	0.12076	0.9683	0.9292
SD of of ROCC	0.0521	0.0821	0.1325	0.0561	0.0621
Compression					
Mean of ROCC	0.1325	0.8963	0.2606	0.2859	0.9587
SD of of ROCC	0.1034	0.0584	0.1462	0.1627	0.0607

this metric is more stable, adaptive and in line with human subjective judgment and distortion scale.

In the future, there are a number of issues that deserve further investigation: (a) further application research of SIFT intensity; (b) development of SIFT intensity to extend its assessment to multiple distortions, happening simultaneously; (c) development of the metric to provide online image/video quality assessment in video surveillance systems.

Acknowledgement

The authors would like to thank the editor and the anonymous reviewers for their valuable comments and constructive suggestions. This paper is jointly supported by the National Key Basic Research Program of China (National 973 Program, No. 2013CB329502), the National Natural Science Foundation of China (No.61379101, No.BK20130209).

References

- [1] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, <http://www.vqeg.org/>, (2000).
- [2] A. M. Eskicioglu and P.S. Fisher, IEEE Transactions on Communications, **43**, 2959-2965 (1995).
- [3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, et al., IEEE Transactions on Image Processing, **9**, 636-650 (2000).
- [4] J.-B. Martens and L. Meesters, Signal Processing, **70**, 155-176 (1997).
- [5] D. M. Chandler and S. S. Hemami, IEEE Transactions on Image Processing, **16**, 2284-2298 (2007).
- [6] H. Chang and M. Wang, Signal Processing: Image Communication, **26**, 577-588 (2011).
- [7] Z. Wang, A. C. Bovik, H. R. Sheikh, et al., IEEE Transactions on Image Processing, **13**, 600-612 (2004).
- [8] A. F. Santiago, S. J. E. Raul, A. L. Carlos, et al., //Proceeding of the 28th IEEE EMBS Annual International Conference, New York: IEEE, 4815-4818 (2006).
- [9] C. Li and A.C. Bovik, Signal Processing: Image Communication, **25**, 517-526 (2010).
- [10] S. Wolf and M. H. Pinson, //Proceeding of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ (2005).
- [11] L. Ma, S. Li, K. N. Ngan, Signal Processing-Image Communication, **28**, 884-902 (2012).
- [12] T. M. Kusuma and H.-J. Zepernick, //Proceeding of Joint First Workshop on IEEE Mobile Future and Symposium on Trends in Communications, Bratislava, Slovakia: IEEE, 71-74 (2003).
- [13] Z. Wang and E. P. Simoncelli, //Proceedings of SPIE Human Vision and Electronic Imaging, **5666**, 149-159 (2005).
- [14] S. Decherchi, P. Gastaldo, R. Zunino, et al., Neurocomputing, **102**, 78-90 (2012).
- [15] S. Liu and A. C. Bovik, IEEE Trans. on Circuits and Systems for Video Technology, **12**, 1139-1149 (2002).
- [16] R. Ferzli and L. J. Karam, //Proceedings of the IEEE International Conference on Image Processing, Genoa, Italy: IEEE, **1**, 405-408 (2005).
- [17] U. Engelke and H.-J. Zepernick, //Proceedings of the 3rd EuroNGI Conference on Next Generation Internet Networks, Trondheim, Norway: IEEE, 190-197 (2007).
- [18] H. R. Wu and M. Yuen, IEEE Signal Processing Letters, **4**, 317-320 (1997).
- [19] P. Marziliano, F. Dufaux, S. Winkler, et al., //Proceeding of IEEE Int. Conf. on Image Processing, **3**, 57-60 (2002).
- [20] X. Marichal, W. Y. Ma, H.J. Zhang, //Proceeding of IEEE Int. Conf. on Image Processing, **2**, 386-390 (1999).
- [21] J. Caviedes and S. Gurbuz, //Proceeding of IEEE Int. Conf. on Image Processing, **3**, 53-56 (2002).
- [22] J. Zhang, S. H. Ong, T. M. Le, Signal Processing: Image Communication, **26**, 13-23 (2011).
- [23] W. Lu, K. Zeng, D. Tao, et al., Neurocomputing, **73**, 784-794 (2010).
- [24] J. Zhang, T. M. Le, S. H. Ong, et al., Signal Processing, **91**, 2575-2588 (2011).
- [25] S. Gabarda and G. Cristóbal, Journal of the Optical Society America A., **A24**, B42-B51 (2007).
- [26] S. Suresh, R. Venkatesh Babu, H. J. Kim, Applied Soft Computing, **9**, 541-552 (2009).
- [27] D. G. Lowe, International Journal of Computer Vision, **60**, 91-110 (2004).
- [28] H. R. Sheikh, Z. Wang, L. Cormack, et al., LIVE Image Quality Assessment Database Release 2, <http://live.ece.utexas.edu/research/quality>, (2012).
- [29] VQEG, Final Report from the Video Quality Experts Group on the Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I. [Online]. 2009, <http://www.vqeg.org/S>, (2009).
- [30] Million gallery, <http://www.mypcera.com/PHOTO>, (2011).



Tongfeng Sun

received his B.S. degree in industrial automation from Northwestern Polytechnical University in 1999. He received his M.S. degree in computer application technology in 2004 and Ph.D. degree in detection technology and automation devices in 2012 from China University of Mining and Technology. He is currently a teacher in the university. His research interests are in the areas of data mining, machine learning and information system.



Shifei Ding received his B.S. degree and M.S. degree from Qufu Normal University in 1987 and 1998 respectively. He received his Ph.D. degree from Shandong University of Science and Technology in 2004. He received postdoctoral degree from Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences in 2006. And now, he works in China University of Mining and Technology as a professor and Ph.D. supervisor. His research interests include intelligent information processing, pattern recognition, machine learning, data mining and granular computing.



Xinzheng Xu received his B.S. degree from Shandong University of Science and Technology in 2002, and his M.S. degree from Xiamen University in 2005. He received his Ph.D. degree from China University of Mining and Technology in 2012. He is currently an associate professor at School of Computer Science and Technology, China University of Mining and Technology. He is a member of China Computer Federation and China Association for Artificial Intelligence. His research interests include intelligent information processing, pattern recognition, machine learning and granular computing.