

Research Article

Total Variation Based Perceptual Image Quality Assessment Modeling

Yadong Wu,^{1,2} Hongying Zhang,³ and Ran Duan³

¹ School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China

² Fundamental Science on Nuclear Wastes and Environmental Safety Laboratory, Southwest University of Science and Technology, Mianyang 621010, China

³ Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, School of Information and Engineering, Southwest University of Science and Technology, Mianyang 621010, China

Correspondence should be addressed to Yadong Wu; wyd028@163.com

Received 12 February 2014; Accepted 10 March 2014; Published 1 April 2014

Academic Editor: X. Song

Copyright © 2014 Yadong Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Visual quality measure is one of the fundamental and important issues to numerous applications of image and video processing. In this paper, based on the assumption that human visual system is sensitive to image structures (edges) and image local luminance (light stimulation), we propose a new perceptual image quality assessment (PIQA) measure based on total variation (TV) model (TVPIQA) in spatial domain. The proposed measure compares TVs between a distorted image and its reference image to represent the loss of image structural information. Because of the good performance of TV model in describing edges, the proposed TVPIQA measure can illustrate image structure information very well. In addition, the energy of enclosed regions in a difference image between the reference image and its distorted image is used to measure the missing luminance information which is sensitive to human visual system. Finally, we validate the performance of TVPIQA measure with Cornell-A57, IVC, TID2008, and CSIQ databases and show that TVPIQA measure outperforms recent state-of-the-art image quality assessment measures.

1. Introduction

Visual quality evaluation has numerous uses in practice and also plays a central role in shaping many visual processing algorithms and systems, as well as their implementation, optimization, and testing. As human being is end receiver of images, one straightforward way for evaluating image quality is subjective testing. The mean opinion score (MOS), subjective quality measurement, has been used for many years. However, it is very expensive and time consuming, which makes it impractical for image processing applications. These drawbacks lead to the development of perceptual image quality assessment (PIQA) metrics that can automatically evaluate the image perceptual quality.

An objective measurement of perceptual quality plays a very important role in many image processing tasks, such as image compression and enhancement. It can be used to dynamically monitor and adjust image quality, optimize algorithms, and benchmark image processing systems [1]. In

recent years, a great deal of effort has been made to develop objective image quality metrics that correlate with human visual behaviors in evaluating image quality [1–4].

Depending upon the availability of a “perfect quality” reference image, the image quality assessment (IQA) metrics are classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) [4]. FR metrics are those that need access to an original reference image to produce a quality score that predicts the subjective judgment of a distorted image. NR metrics only require distorted images to predict quality scores. RR metrics are between FR metrics and NR metrics, which require only partial information about the reference image [3]. In this paper, we focus on FR image quality assessment.

Generally, FR metrics measure the distance between a distorted image and its original image in a perceptually meaningful way. FR metrics can also be designed in two ways. One is modeling HVS, which has been regarded as the most appropriate way to measure and predict the perceptual

image quality. The underlying assumption is that HVS is sensitive to the differences of visual signals in some respects, such as brightness, contrast, and frequency content. Under this assumption, the strength of the difference between a distorted and its reference image reflects the different perceived sensitivities of HVS. The other way explores signal fidelity criteria that is not based on assumptions about HVS model but is motivated instead by the need to capture the loss of signal structure that HVS hypothetically extracts for cognitive understanding [5].

In this paper, we propose a new framework for PIQA based on TV model in spatial domain. In the proposed PIQA metric, two human visual sensitivity factors, image structures and luminance changes in enclosed regions, are considered. As far as natural image signal is concerned, an evaluation metric needs to consider the characteristics of the image itself, such as image structure and content, to reflect the image visual complexity. On the other hand, from HVS perspective, another important factor to be considered is luminance change of smooth and enclosed regions, as HVS is very sensitive to luminance change. Based on these ideas, we propose a TVPIQA metric in spatial domain, in which we use TV to describe the image structure and the energy of enclosed regions in the difference image to measure luminance changes. Then, TVPIQA is represented by the weighted sum of these two factors.

To successfully assess the image quality, there are two major contributions of the proposed metric. First, we introduce TV model to assess image's structure. The TV comparison between a distorted image and its reference image is applied to measure the distance of an image structural characteristic. Because of the good performance of TV in describing the edges, the proposed TVPIQA metric can describe the image structure information very well. Second, the luminance changes in enclosed regions are also considered in TVPIQA. The energy of enclosed regions in a difference image is used to measure the missing luminance information that is sensitive to human visual system. In addition, in order to make TVPIQA metric closer to perceptual feelings, isolated pixels' energy in difference image is removed based on the idea of just noticeable distortion (JND), and a fast approximation method of calculating difference image's energy is also proposed.

We demonstrate our TVPIQA metric by presenting performance results with extensive subjective databases (Cornell-A57 [6, 7], IVC [8, 9], TID2008 [10, 11], and CSIQ [12, 13]) and comparisons to seven often-used image metrics (PSNR, SSIM [1], IW-PSNR [4], IW-SSIM [4], MS-SSIM [14], VSNR [7], and VIF [5]). Experimental results demonstrate that the performance of TVPIQA metric outperforms other state-of-the-art metrics. It is worth noting that the proposed metric is easy to compute in spatial domain and does not need any other additional information.

The rest of paper is organized as follows. Section 2 presents some related work. The PIQA metric based on TV model is given in Section 3, and the implementation details of TVPIQA metric are also provided in this section. The characteristics of TVPIQA metric are analyzed in Section 4

and its performance is evaluated and discussed in Section 5. We conclude the paper in Section 6.

2. Related Work

According to different methodologies being considered, PIQA metrics can be divided into two categories: HVS features based modeling and signal driven approach [15]. For HVS features based modeling, PIQA metrics are developed based upon systematical modeling of relevant psychophysical properties and physiological knowledge, including temporal/spatial/color decomposition, contrast sensitivity function (CSF), luminance adaptation, and various masking effects [15]. A number of HVS based methods have been proposed in the literature [16–19]. Some have also considered JND model [20, 21]. HVS based methods extrapolate the vision models that have been proposed in the visual psychology literature to PIQA. However, HVS features based methods involve expensive computation and difficulties due to the gap between the knowledge for vision research and the need for engineering modeling [15].

Recently, a lot of research efforts have been concentrated on signal driven PIQA metrics, which are designed from the viewpoint of signal extraction and analysis, such as statistical features, structural distortion, and so forth [1, 4, 22, 23]. Signal driven methods do not attempt to build a comprehensive HVS model regarding quality evaluation. These metrics look at how to represent image features to estimate overall quality, and they often consider psychophysical effects as well, usually based on image content and distortion analysis. However, although some signal fidelity metrics reflect picture quality change, they fail to predict HVS perception because of some problems [15]. For example, not every image change is noticeable and leads to distortion. Therefore, signal driven methods need HVS features to help tackle these problems, so that they can better approximate perceptual quality evaluation.

Variational methods have been extremely successful in wide various fields in image processing and computer vision during last decades [24, 25]. TV model is first introduced by Rudin, Osher, and Fatemi (ROF) in their pioneering work on edge preserving [26]. For an image u , its TV can be formulated by

$$\text{TV}(u) = \int_{\Omega} |\nabla u|, \quad (1)$$

where Ω denotes the image domain and ∇ means the gradient operator.

Many research results have shown that the proper norm for an image is the total variation norm, which is essentially L_1 norm of derivative and is more appropriate for image estimation and description in discontinuities [24, 25]. The advantages of TV norm led us to consider using it to measure image structure change, which is the distance between a distorted image and its original image. The proposed TVPIQA metric will focus on two human visual sensitivity factors, that is, image structures and luminance changes in enclosed regions. A significant difference between TVPIQA metric and other PIQA metrics is that TV model is introduced to assess



FIGURE 1: Comparison of “Lena” images with different types of distortions. (a) Original image (512 × 512, 8 bits/pixel). (b) Blurring. (c) Gaussian noise contamination. (d) Impulsive noise contamination.

image structures in spatial domain. Moreover, luminance changes in enclosed regions are also considered.

3. PIQA Metric Based on TV Model

Many signal driven PIQA metrics have been investigated based on the assumption that the loss of perceptual quality is related to the visibility of error signals. Generally, the mean squared error (MSE) or the peak signal-to-noise ratio (PSNR) has been a popular and usual metric to evaluate image quality. But MSE measure exhibits weak performance in assessing perceptual image quality [1]. From Figure 1 where the original “Lena” image is degraded with different distortions, we can see that MSE cannot reflect an image PIQ. The motivation of

the paper is to design an appropriate measure for some HVS characteristics, especially image structures and luminance, and develop a novel PIQA metric.

3.1. Framework of Proposed TVPIQA Metric. When a natural image is observed through HVS, the subjective quality measure is affected by many factors. Because human eyes are sensitive to changes of image edges, especially the edge location information, and changes of luminance contrast [27], there are two main factors worthy of attention. One is image edge (structure) information, and the other is luminance information. Therefore, we propose a new TVPIQA metric to measure these two factors and provide a good approximation to perceived image distortion.

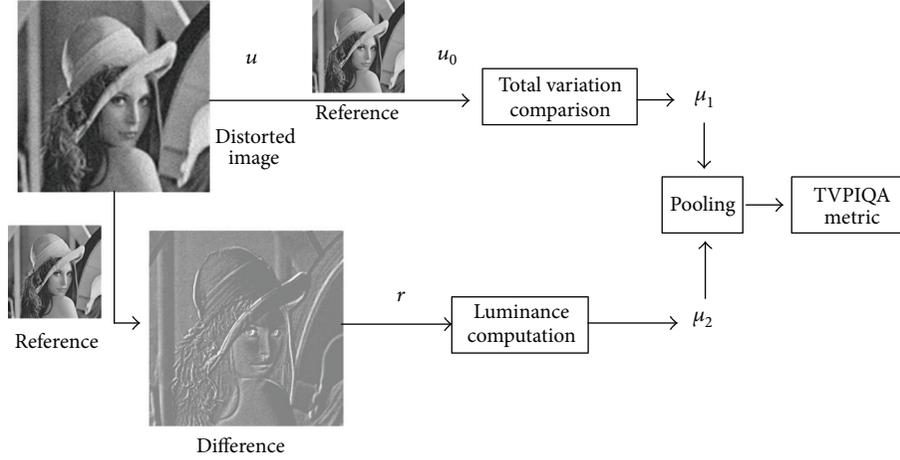


FIGURE 2: Framework of the proposed TVPIQA metric.

Figure 2 shows the framework of the proposed TVPIQA metric, which is separated into two parts: structure part and local region luminance part. In the structure part, the normalized TV comparison between the distorted image and the reference image, denoted by μ_1 , is applied to represent the image structure changes. On the other hand, the energy of the enclosed regions in a difference image is used to measure the luminance changes, which is also normalized and denoted by μ_2 . Then, the proposed TVPIQA metric is the mean of these two parts as follows:

$$\text{TVPIQA} = \frac{\mu_1 + \mu_2}{2}. \quad (2)$$

3.2. TV Based Structure Measure μ_1 . Let u and u_0 represent the degraded image and its reference image, respectively. Because TV norm is very appropriate for image description in discontinuities, the structure's change is measured by the TV difference between a reference image and its degraded image

$$\text{TV}_{\text{struct}} = \|\text{TV}(u) - \text{TV}(u_0)\|_1, \quad (3)$$

where $\|\cdot\|_1$ represents L_1 norm and $\text{TV}(u)$ denotes the total variation of the image u , expressed in discrete form

$$\text{TV}(u) = \sum_{(i,j) \in \Omega} \left(\sqrt{(u_{i,j} - u_{i+1,j})^2 + (u_{i,j} - u_{i,j+1})^2} \right), \quad (4)$$

where $u_{i,j}$ represents the intensity value at pixel (i, j) .

Although the above measurement can work to assess the structure change, it is not a normalized measure and cannot be used as an evaluation to describe subjective feelings about image's quality. Therefore, according to

$\|\text{TV}(u) - \text{TV}(u_0)\|_1^2 \geq 0$, the normalized perceptual distance for image structure is derived and defined by

$$\begin{aligned} \mu_1 = \frac{1}{N} \sum & \left(\left(2\sqrt{(u_{i,j} - u_{i+1,j})^2 + (u_{i,j} - u_{i,j+1})^2} \right. \right. \\ & \times \sqrt{(u_{0i,j} - u_{0i+1,j})^2 + (u_{0i,j} - u_{0i,j+1})^2 + c} \\ & \times \left. \left. \left((u_{i,j} - u_{i+1,j})^2 + (u_{i,j} - u_{i,j+1})^2 \right. \right. \right. \\ & \left. \left. \left. + (u_{0i,j} - u_{0i+1,j})^2 + (u_{0i,j} - u_{0i,j+1})^2 + c \right)^{-1} \right), \end{aligned} \quad (5)$$

where N is the image size and c is a constant and set to 75 according to our experiments. It is obvious that $\mu_1 \in (0, 1]$.

3.3. Local Region Luminance Measure μ_2 . A difference image represents the information loss in a distorted image and is defined by the difference between a reference image and its distorted image; that is, $r = u_0 - u$. Because HVS is also sensitive to luminance changes when observing an image, the energy of a difference image is used to measure the missing luminance information. Furthermore, based on the idea of JND modeling, that is, not every change in an image is noticeable [15], the isolated pixels' energy in a difference image is filtered out. The energy measure of a difference image E_r , which measures the luminance loss of a distorted image, is defined by

$$E_r = \frac{1}{N} \sum_{(i,j) \in \Omega'} r_{i,j}^2, \quad (6)$$

where Ω' represents the enclosed regions in a difference image and $r_{i,j}$ is the intensity value at pixel (i, j) in difference image. To compute the energy measure E_r efficiently and

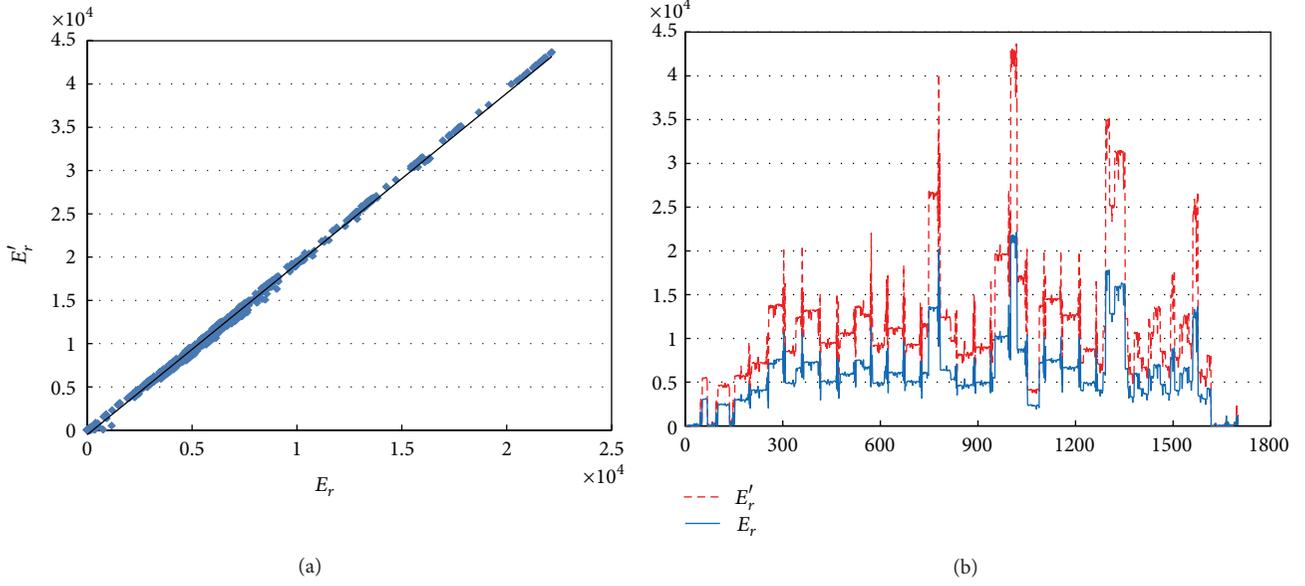


FIGURE 3: Comparison of E_r and E'_r for TID2008. (a) $E'_r - E_r$ curve. The correlation coefficient value between E'_r and E_r is 0.9988. (b) E_r and E'_r curves for different distorted images.

avoid the judgment of the enclosed regions, we proposed an approximation of the difference image's energy E'_r defined as

$$E'_r = \frac{1}{N} \sum_{(i,j) \in \Omega} (r_{i,j} r_{i+1,j} + r_{i,j} r_{i,j+1}), \quad (7)$$

where Ω represents the whole regions in a difference image.

To test the relationship between E_r and E'_r , we applied these two functions to TID2008 image database [10, 11], which includes 1700 distorted images generated from 25 reference images with 17 distortion types at four distortion levels. Figure 3(a) shows the relationship between E'_r and E_r . The value of correlation coefficient is 0.9988, indicating that E'_r is highly related to E_r . In Figure 3(b), the horizontal axis indicates the image number and the vertical axis indicates the energy value. Figure 3(b) shows the change curves of E'_r and E_r plotted by all distorted images in TID2008. From Figure 3(b), we can see that E'_r and E_r almost have the same change trend. Therefore, we can use E'_r , instead of E_r , to measure the energy of a difference image.

Considering that human perception is more sensitive to luminance contrast rather than to absolute luminance, we adjust E'_r according to the mean intensity of a difference image

$$E'_r = \frac{1}{N} \sum_{(i,j) \in \Omega} ((r_{i,j} - \bar{r})(r_{i+1,j} - \bar{r}) + (r_{i,j} - \bar{r})(r_{i,j+1} - \bar{r})), \quad (8)$$

where \bar{r} represents the mean intensity of a difference image; that is, $\bar{r} = (1/N) \sum_{(i,j) \in \Omega} r_{i,j}$.

In order to obtain the normalized measure for luminance change, we need to find the maximum difference image's energy according to the reference image. In the case of

consistency of the overall image energy, we assume that the distorted image, in which intensity values in all pixels are the same and equal to the mean intensity of the reference image, corresponds to the maximum luminance change. Based on this assumption, the maximum difference image is denoted by $r_{\max} = u_0 - \bar{u}_0$. It describes the maximum loss of luminance information in an original reference image. Then, the normalized perceptual distance measure for image luminance change is defined by

$$\mu_2 = 1 - \sqrt{\frac{E'_r}{E'_{r_{\max}}}}, \quad (9)$$

where $E'_{r_{\max}}$, computed by (8), represents the energy measure of the maximum difference image r_{\max} . Obviously, $\mu_2 \in (0, 1]$. The higher the value of μ_2 is, the less luminance information lose.

4. Analysis of the Proposed TVPIQA Measure

This section analyzes some properties of the proposed TVPIQA measure, such as symmetry, boundedness, and unique maximum. Meanwhile, the difference of TVPIQA measure is also discussed in evaluating different types (contrast change, noise contamination, and blurring) of distorted images in this section.

4.1. Properties of TVPIQA Measure

(1) *Symmetry*. Because the structure measure function $\mu_1(x, y)$ is derived by $\|TV(u) - TV(u_0)\|_1^2 \geq 0$, $\mu_1(x, y) = \mu_1(y, x)$ is obvious. On the other hand, the energy function satisfies the symmetry; that is, $E'_r(x, y) = E'_r(y, x)$, and the energy

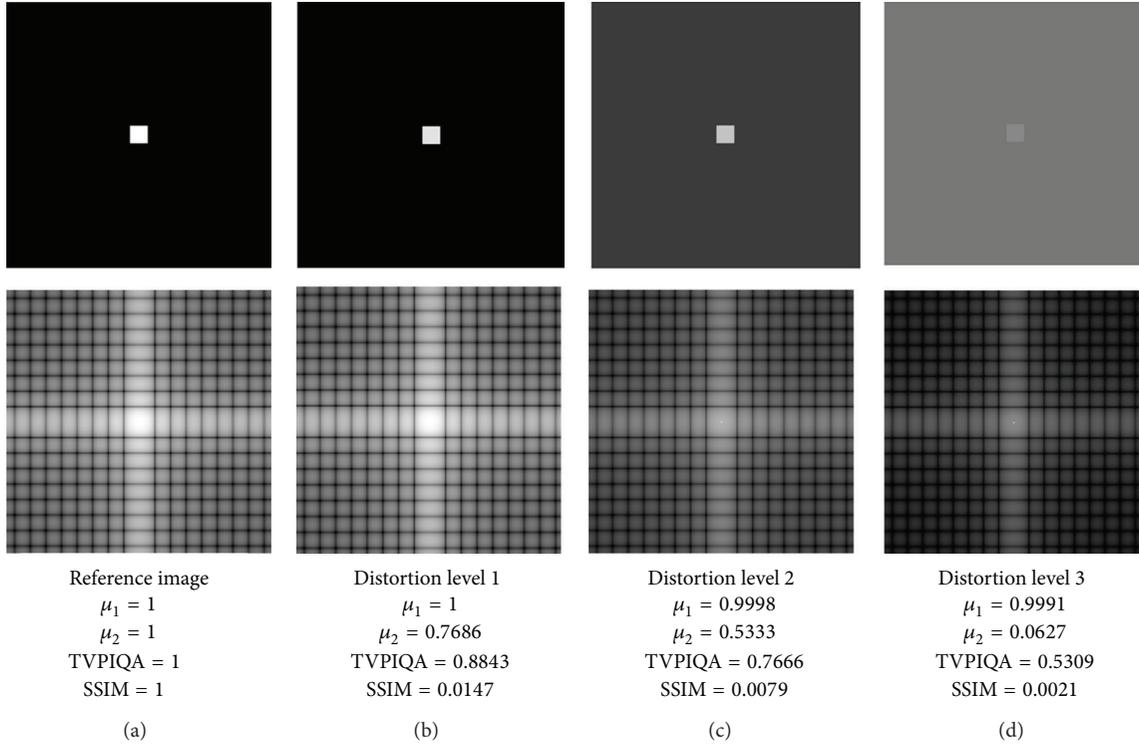


FIGURE 4: Contrast change distortion analysis in frequency domain. The first row images are the reference image and contrast change distorted images. The second row images are Fourier spectrums of the first row images. (a) Reference image. Background is black with a white square. The gray values of the background and the square are 0 and 255, respectively. (b) Contrast change distortion level 1. Background is 30, and square is 226. (c) Contrast change distortion level 2. Background is 60, and square is 196. (d) Contrast change distortion level 3. Background is 120, and square is 136.

measure function $\mu_2(x, y) = 1 - \sqrt{E'_r(x, y)/E'_{r_{\max}}}$ also satisfies the symmetry. Therefore, the proposed TVPIQA measure, $TVPIQA(x, y) = (\mu_1(x, y) + \mu_2(x, y))/2$, is symmetric.

(2) *Boundedness and Unique Maximum.* According to the definition of $\mu_1(x, y)$ and $\mu_2(x, y)$ in Section 3, measure $\mu_1(x, y) \in (0, 1]$ and measure $\mu_2(x, y) \in (0, 1]$. So, TVPIQA metric is bounded; that is, $TVPIQA(x, y) \in (0, 1]$. Only when a distorted image is the same as its reference image, the structure measure $\mu_1(x, y) = 1$ and energy measure $\mu_2(x, y) = 1$; that is, if and only if $x = y$, $TVPIQA(x, y) = 1$.

4.2. *TVPIQA Measure for Different Types of Distortions.* Due to the limited considered factors in designing PIQA metric and limited understanding of HVS, it is impossible for any PIQA metric to measure all kinds of distortions in the same measurement scale. To some distortions, a PIQA metric may be gentle, while being critical to other distortions. In our extensive experiments, the proposed TVPIQA metric is more critical in evaluating contrast distortions, compared with evaluating other distortions. This performance of TVPIQA exactly reflects HVS contrast sensitivity characteristic.

Figure 4 shows different levels of contrast distortions for an artificial image. In Figure 4, μ_1 measures structure changes in the distorted images, and μ_2 measures luminance changes. TVPIQA and SSIM measure the distances from

distorted images to the reference image. From Figure 4, we can see that structure measure in the distorted images, μ_1 , changes extremely slowly. However, the energy measure in the distorted images, μ_2 , declines rapidly. The Fourier spectrums of the distorted images also show the same changes. Figure 5 shows another example, and the test image is "1600" images in CSIQ database. The same situation can be observed from Figures 5(b), 5(c), and 5(d) that μ_1 changes slowly and μ_2 drops dramatically. For the same distortion level with different distortion types, shown in Figures 5(d), 5(e), and 5(f), the contrast distortion has the largest drop between luminance measure and structure measure.

According to Figures 4 and 5, TVPIQA measure shows different measure scales for contrast distortions and other distortions, such as noise contamination and blurring. However, this does not mean that the performance of TVPIQA measure is not good, because, for HVS perception, not every change yields the same extent of perceptual effect with the same magnitude of change [15]. Therefore, to evaluate TVPIQA measure performance, the contrast distortion will be discussed separately in the next section.

5. Experimental Results

In this section, we validate the performance of the proposed TVPIQA measure and compare it with other seven IQA measures, that is, PSNR, SSIM [1], IW-PSNR [4], IW-SSIM

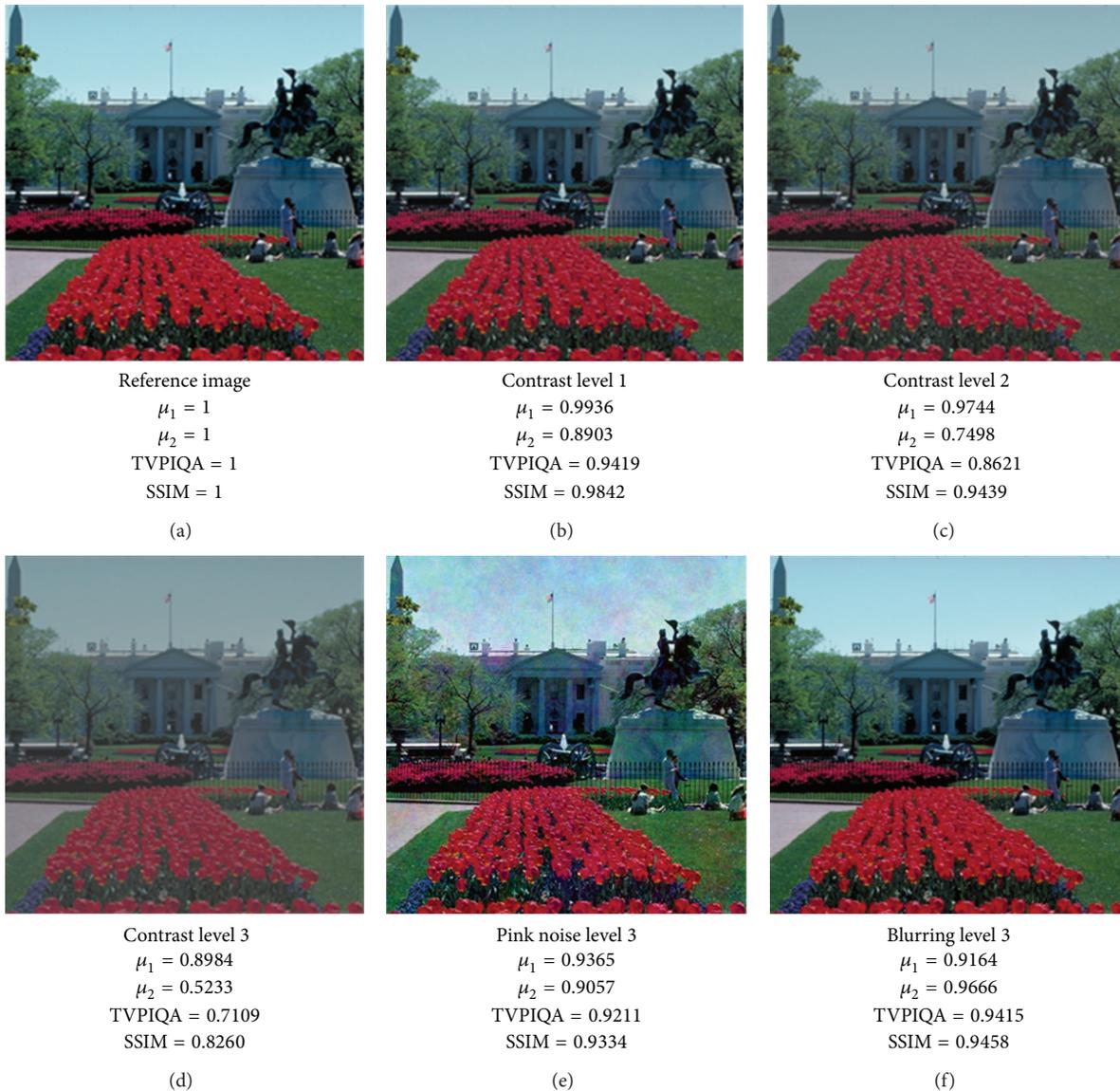


FIGURE 5: TVPIQA measure comparison for different type distortions. The test image is the “1600” images in CSIQ database. (a) Reference image. (b) Contrast change distortion level 1. (c) Contrast change distortion level 2. (d) Contrast change distortion level 3. (e) Additive Gaussian pink noise contamination level 3. (f) Blurring level 3.

[4], MS-SSIM [14], VSNR [7], and VIF [5]. PSNR is widely used in the image processing field and is also a useful baseline comparison. SSIM, MS-SSIM, visual signal-to-noise ratio (VSNR), and VIF are state-of-the-art measures that have demonstrated competitive performance. The information content weighted PSNR (IW-PSNR) and information content weighted SSIM (IW-SSIM) measures have been confirmed having the best overall performance compared with the previous measures. So, IW-PSNR and IW-SSIM are also good benchmarks to evaluate the new TVPIQA measure.

The proposed TVPIQA measure and other seven measures are evaluated on four publicly available subjective image databases that are widely recognized in the IQA research community, that is, Cornell-A57 [6, 7], IVC [8, 9], TID2008 [10, 11], and CSIQ [12, 13]. Two different types of subjective

quality scores have been used: MOS and differential MOS (DMOS) in these image databases.

The Cornell-A57 database [6, 7] was created at Cornell University. It contains 54 distorted images with six types of distortions including quantization distortion, noise contamination, and blurring. The IVC database [8, 9] includes 185 distorted images generated from ten original images. There are four types of distortions that are (a) JPEG compression, (b) JPEG2000 compression, (c) local adaptive resolution (LAR) coding, and (d) blurring. The Tampere Image Database 2008 (TID2008), introduced in the previous section, is intended to evaluate full-reference image visual quality assessment metrics. It has 17 types of distortions, such as noise distortion, blur distortion, and contrast change. The categorical image quality (CSIQ) database [12, 13] was developed at Oklahoma

State University. It consists of 30 original images and 866 distorted images using six different types of distortions at four to five different levels of distortion. The distortion types include JPEG compression, JPEG2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring.

The performance of any objective visual quality assessment metric is evaluated by measuring its correlation with human perception. In order to do so, the objective quality scores of IQA metrics are correlated with subjective scores using a variety of statistical measures such as the correlation coefficient (CC), Spearman's rank correlation coefficient (SRCC), and Kendall's rank correlation coefficient (KRCC). To compare performance of different IQA measures, three evaluation metrics, that is, CC, SRCC, and KRCC, are used in experiments as shown in the following.

Correlation coefficients evaluate the prediction accuracy and measure linear dependence between the subjective and the objective scores. CC is defined as

$$CC = \frac{\sum_{i=1}^N (s_i - \bar{s}) \cdot (o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (s_i - \bar{s})^2} \sqrt{\sum_{i=1}^N (o_i - \bar{o})^2}}, \quad (10)$$

where \bar{s} is the mean value of subjective scores s_i , $i = 1, \dots, N$ and \bar{o} is the mean value of objective scores o_i .

Spearman's rank correlation coefficients measure the prediction of monotonicity [28]. SRCC is given by

$$SRCC = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}, \quad (11)$$

$$d_i = R_{s_i} - R_{o_i},$$

where R_{s_i} and R_{o_i} are the i th image's ranks in subjective and objective evaluations, respectively. SRCC is a nonparametric rank-based correlation metric, independent of any monotonic nonlinear mapping between subjective and objective scores [4].

Kendall's rank correlation coefficient (KRCC) is another nonparametric rank correlation metric computed by

$$KRCC = \frac{N_c - N_d}{N(N-1)/2}, \quad (12)$$

where N_c and N_d are the numbers of concordant and discordant pairs in the data set, respectively.

According to the above definitions, larger CC, SRCC, and KRCC values, close to 1, indicate that the objective and subjective scores correlate better, that is to say, a better performance of IQA metric. In our performance comparisons, CC, SRCC, and KRCC of seven metrics are either computed or referenced from some research works [29–31]. Among these metrics, PSNR, IW-PSNR, SSIM, MS-SSIM, and IW-SSIM are referred to Zhou Wang's research [29], while VSNR and VIF are referred to Lin Zhang's research [30, 31].

Table 1 shows our test results of eight IQA metrics using four databases. As analyzed in Section 4, the contrast distorted images in TID2008 and CSIQ databases are discussed separately. For each evaluation metric in each test, we

TABLE 1: Performance comparisons of eight IQA metrics on four publicly available image databases.

Metric	CC	SRCC	KRCC
Cornell-A57 database (54 images) [6, 7]			
PSNR	0.6346	0.6189	0.4309
SSIM [1]	0.7531	0.8066	0.6058
TVPIQA	0.8797	0.8328	0.6510
IW-PSNR [4]	0.8752	0.8759	0.6967
IW-SSIM [4]	0.8918	0.8709	0.6842
MS-SSIM [14]	0.8396	0.8414	0.6478
VSNR [7]	0.9147	0.9355	0.8031
VIF [5]	0.6141	0.6223	0.4589
IVC database (185 images) [8, 9]			
PSNR	0.6705	0.6884	0.5218
SSIM [1]	0.8092	0.9018	0.7223
TVPIQA	0.8551	0.9043	0.7242
IW-PSNR [4]	0.8900	0.8998	0.7165
IW-SSIM [4]	0.7943	0.9125	0.7339
MS-SSIM [14]	0.7854	0.8980	0.7203
VSNR [7]	0.7824	0.7993	0.6053
VIF [5]	0.8800	0.8964	0.7158
TID2008 database without contrast distortion (1600 images) [10, 11]			
PSNR	0.5599	0.5974	0.4394
SSIM [1]	0.7519	0.7912	0.5912
TVPIQA	0.8592	0.8772	0.6869
IW-PSNR [4]	0.6353	0.7855	0.6074
IW-SSIM [4]	0.8217	0.8748	0.6868
MS-SSIM [14]	0.8031	0.8742	0.6798
VSNR [7]	0.3238	0.7992	0.6096
VIF [5]	0.7622	0.7383	0.5777
TID2008 database with only contrast distortion (100 images) [10, 11]			
PSNR	0.5770	0.5830	0.4178
SSIM [1]	0.5061	0.5204	0.3891
TVPIQA	0.6210	0.6054	0.4421
IW-PSNR [4]	0.5691	0.5627	0.4000
IW-SSIM [4]	0.7605	0.6251	0.4675
MS-SSIM [14]	0.7556	0.6379	0.4793
VSNR [7]	0.4048	0.4090	0.2686
VIF [5]	0.8706	0.8094	0.5678
CSIQ database without contrast distortion (750 images) [12, 13]			
PSNR	0.8544	0.9061	0.7237
SSIM [1]	0.8475	0.9247	0.7534
TVPIQA	0.9040	0.9583	0.8170
IW-PSNR [4]	0.9267	0.9523	0.8069
IW-SSIM [4]	0.8292	0.9544	0.8096
MS-SSIM [14]	0.8481	0.9506	0.7984
VSNR [7]	0.8471	0.9266	0.7575
VIF [5]	0.9137	0.9181	0.7561

TABLE 1: Continued.

Metric	CC	SRCC	KRCC
CSIQ database with only contrast distortion (116 images) [12, 13]			
PSNR	0.8887	0.8621	0.6449
SSIM [1]	0.7666	0.7922	0.5779
TVPIQA	0.9491	0.9543	0.8162
IW-PSNR [4]	0.9125	0.9230	0.7508
IW-SSIM [4]	0.9098	0.9539	0.8168
MS-SSIM [14]	0.9003	0.9526	0.8123
VSNR [7]	0.8680	0.8795	0.7022
VIF [5]	0.9336	0.9404	0.7901

TABLE 2: Average performance over four databases.

Metric	CC	SRCC	KRCC
Database size-weighted average without contrast distortion			
PSNR	0.6547	0.6938	0.5275
SSIM [1]	0.7837	0.8381	0.6479
TVPIQA	0.8723	0.9017	0.7265
IW-PSNR [4]	0.7429	0.8439	0.6748
IW-SSIM [4]	0.8234	0.9005	0.7257
MS-SSIM [14]	0.8156	0.8973	0.7164
VSNR [7]	0.5205	0.8389	0.6562
VIF [5]	0.8114	0.7993	0.6368
Database size-weighted average with all distorted images			
PSNR	0.6616	0.6968	0.5284
SSIM [1]	0.7731	0.8249	0.6358
TVPIQA	0.8666	0.8933	0.7201
IW-PSNR [4]	0.7437	0.8371	0.6682
IW-SSIM [4]	0.8247	0.8929	0.7203
MS-SSIM [14]	0.8170	0.8904	0.7119
VSNR [7]	0.5307	0.8253	0.6442
VIF [5]	0.8186	0.8055	0.6406

highlight our proposed TVPIQA metric and the best of other seven metrics with boldface.

From the experimental results in Table 1, we have two major observations. TVPIQA metric has similar performance to IW-SSIM on Cornell-A57 and IVC databases. Without considering the contrast distortion, the proposed TVPIQA metric has the best performance on TID2008 and CSIQ databases. On the other hand, TVPIQA metric has the highest SRCC value in evaluating the contrast distortion in CSIQ database. However, many metrics do not work well on the contrast distortion in TID2008 database. The main reason is that contrast enhancement images in TID2008 have high subjective scores for HVS perception, while they are measured as the distortion by some objective evaluation metrics.

To evaluate the overall performance of IQA metrics under comparison, Table 2 presents the average CC, SRCC, and KRCC results over four databases, where the average

values are computed in two cases. In the first case, without considering the contrast distortion, the correlation scores are computed by the size weighted average method. Different weights are given to different databases, depending upon their sizes (measured as the numbers of images, i.e., 54 for Cornell-A57, 185 for IVC, 1600 for TID2008, and 750 for CSIQ databases), while in the second case, the contrast distortions in TID2008 and CSIQ are also considered, and the correlation scores are still averaged according to the numbers of images, 100 for TID2008 contrast distortion and 116 for CSIQ contrast distortion.

From Table 2, it can be observed that our proposed TVPIQA metric has better overall performance than other IQA metrics. Although IW-SSIM and TVPIQA nearly have the same performance from test results, it is worth mentioning that, from the view point of computation complexity, TVPIQA metric achieves this excellent performance only by computing the image structure and energy information in spatial domain, which does not need any preprocessing, such as image transform operation, extra image analysis operation, and so forth. However, IW-SSIM needs more computation time to information content weights.

6. Conclusions

In this paper, we propose a new framework for TV based perceptual image quality assessment measure in spatial domain. The proposed TVPIQA measure focuses on two human visual sensitivity factors, image structures and luminance changes. A significant difference between TVPIQA measure and other IQA measures is that TV model is introduced to assess image structures. Meanwhile, the energy of enclosed regions in a difference image is used to measure the missing luminance information which is also sensitive to human visual system. Extensive experimental results with four publicly available independent image databases demonstrate that the proposed TVPIQA measure achieves the best overall performance when compared with other seven popular IQA measures.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (Grant no. 61303127), Western Light Talent Culture Project of Chinese Academy of Sciences (Grant no. 13ZS0106), Project of Science and Technology Department of Sichuan Province (Grant nos. 2011JQ0041 and 11ZS2009), and Key Program of Education Department of Sichuan Province (Grant nos. 11ZA130 and 13ZA0169).

References

- [1] 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.

- [2] “Perceptual criteria for image quality evaluation,” in *Handbook of Image and Video Processing*, T. N. Pappas, R. J. Safranek, J. Chen, and A. Bovik, Eds., Academic Press, New York, NY, USA, 2nd edition, 2005.
- [3] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan and Claypool Publishers, San Rafael, Calif, USA, 2006.
- [4] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [5] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
- [6] D. M. Chandler and S. S. Hemami, “Cornell-A57 Database,” <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [7] D. M. Chandler and S. S. Hemami, “VSNR: a wavelet-based visual signal-to-noise ratio for natural images,” *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [8] A. Ninassi, P. Le Callet, and F. Atrousseau, “Pseudo no reference image quality metric using perceptual data hiding,” in *Human Vision and Electronic Imaging*, vol. 6057 of *Proceedings of the SPIE*, San Jose, Calif, USA, January 2006.
- [9] A. Ninassi, P. Le Callet, and F. Atrousseau, “Subjective quality assessment: IVC database,” <http://www2.ircsyn.ec-nantes.fr/ivcdb>.
- [10] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, “Metrics performance comparison for color image database,” in *4th International Workshop on Video Processing and Quality Metrics*, Scottsdale, Ariz, USA, January 2009.
- [11] N. Ponomarenko and K. Egiazarian, “TAMPERE IMAGE DATABASE 2008 TID2008, version 1.0,” <http://www.ponomarenko.info/tid2008.htm>.
- [12] E. C. Larson and D. M. Chandler, “Categorical image quality (CSIQ) database,” <http://vision.okstate.edu/csiq>.
- [13] E. C. Larson and D. M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *Journal of Electronic Imaging*, vol. 19, no. 1, Article ID 011006, 2010.
- [14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” in *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402, Pacific Grove, Calif, USA, November 2003.
- [15] W. Lin and C.-C. Jay Kuo, “Perceptual visual quality metrics: a survey,” *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.
- [16] S. Daly, “The visible difference predictor: an algorithm for the assessment of image fidelity,” in *Human Vision, Visual Processing, and Digital Display*, vol. 1616 of *Proceedings of SPIE*, pp. 2–15, 1992.
- [17] O. D. Faugeras, “Digital color image processing within the framework of a human visual model,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 4, pp. 380–393, 1979.
- [18] F. X. Lukas and Z. L. Budrikis, “Picture quality prediction based on a visual model,” *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1679–1692, 1982.
- [19] A. B. Watson, “DCTune: a technique for visual optimization of DCT quantization matrices for individual images,” *Society for Information Display Digest of Technical Papers*, vol. 24, pp. 946–949, 1993.
- [20] L. Ma and K. N. Ngan, “Adaptive block-size transform based just-noticeable difference profile for images,” in *Proceedings of the 10th Pacific Rim Conference on Multimedia*, 2009.
- [21] W. Lin, L. Dong, and P. Xue, “Visual distortion gauge based on discrimination of noticeable contrast changes,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 900–909, 2005.
- [22] M. Miyahara, K. Kotani, and V. Ralph Algazi, “Objective picture quality scale (PQS) for image coding,” *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.
- [23] I. Avcibaş, B. Sankur, and K. Sayood, “Statistical evaluation of image quality measures,” *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 206–223, 2002.
- [24] T. F. Chan, S. Esedoglu, F. Park, and A. Yip, “Recent developments in total variation image restoration,” in *Handbook of Mathematical Models in Computer Vision*, Springer, 2005.
- [25] T. F. Chan, J. Shen, and L. Vese, “Variational PDE models in image processing,” *Notices of the American Mathematical Society*, vol. 50, no. 1, pp. 14–26, 2003.
- [26] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [27] T. T. Norton, D. A. Corliss, and J. E. Bailey, *Psychophysical Measurement of Visual Function*, Butterworth-Heinemann, Boston, Mass, USA, 2002.
- [28] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” 2000, <http://www.vqeg.org/>.
- [29] “IW-SSIM: Information Content Weighted Structural Similarity Index for Image Quality Assessment,” 2011, <https://ece.uwaterloo.ca/~z70wang/research/iwssim/>.
- [30] “Evaluation of VIF,” 2011, http://sse.tongji.edu.cn/linzhang/IQA/Evaluation_VIF/eva-VIF.htm.
- [31] “Evaluation of VSNR,” 2011, http://sse.tongji.edu.cn/linzhang/IQA/Evaluation_VSNR/eva-VSNR.htm.