

## Perceptual Fidelity Aware Mean Squared Error

Wufeng Xue<sup>1,2</sup>, Xuanqin Mou<sup>1</sup>, Lei Zhang<sup>2</sup>, and Xiangchu Feng<sup>3</sup>

<sup>1</sup>Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, CHINA

<sup>2</sup>Department of Computing, The Hong Kong Polytechnic University, Hong Kong, CHINA

<sup>3</sup>School of Science, XiDian University, CHINA

xwolfs@hotmail.com, xqmou@mail.xjtu.edu.cn

cslzhang@comp.polyu.edu.hk, xcfeng@mail.xidian.edu.cn

### Abstract

*How to measure the perceptual quality of natural images is an important problem in low level vision. It is known that the Mean Squared Error (MSE) is not an effective index to describe the perceptual fidelity of images. Numerous perceptual fidelity indices have been developed, while the representatives include the Structural SIMilarity (SSIM) index and its variants. However, most of those perceptual measures are nonlinear, and they cannot be easily adopted as an objective function to minimize in various low level vision tasks. Can MSE be perceptual fidelity aware after some minor adaptation?*

*In this paper we propose a simple framework to enhance the perceptual fidelity awareness of MSE by introducing an  $l_2$ -norm structural error term to it. Such a Structural MSE (SMSE) can lead to very competitive image quality assessment (IQA) results. More surprisingly, we show that by using certain structure extractors, SMSE can be further turned into a Gaussian smoothed MSE (i.e., the Euclidean distance between the original and distorted images after Gaussian smooth filtering), which is much simpler to calculate but achieves rather better IQA performance than SSIM. The so-called Perceptual-fidelity Aware MSE (PAMSE) can have great potentials in applications such as perceptual image coding and perceptual image restoration.*

### 1. Introduction

In many image processing and low level vision tasks, it is indispensable to find a suitable fidelity measure to measure the fidelity index of the underlying visual signal. The

use of fidelity measure can be generally classified into the following categories. First of all, the fidelity measure is ubiquitously used to evaluate the performance of competing algorithms and to guide the parameter selection, for example, in image denoising [9] and medical image reconstruction [29]. Second, the fidelity measure is used to design the objective function for minimization in applications such as image coding [12,23] and image restoration [17]. Third, the fidelity measure can be used as the rule to make decisions, for example, in content based image retrieval [21] and block matching [16,27].

The most popular and widely used signal fidelity measure may be the classical Mean Squared Error (MSE), which measures the fidelity of a signal  $\mathbf{d} \in \mathcal{R}^N$  by  $\|\mathbf{r} - \mathbf{d}\|_2^2/N$ , where  $\mathbf{r} \in \mathcal{R}^N$  is the reference counterpart of  $\mathbf{d}$ . MSE has many appealing mathematical properties. It is simply the  $l_2$ -norm of the difference between  $\mathbf{r}$  and  $\mathbf{d}$ , and is the Euclidean distance metric in the  $\mathcal{R}^N$  space. MSE is convex and differentiable, and can be easily minimized with an analytical solution. From another viewpoint, MSE measures the energy of the error signal, which can be preserved under orthonormal transforms such as Fourier transform, discrete cosine transform, orthonormal wavelet transform and principal component analysis, etc. For a more thorough analysis of MSE, please refer to [24].

Despite of the many desired mathematical properties, one fatal weakness of MSE in the context of vision applications is its inconsistency with human perception of image quality [24]. This is mainly because MSE is pixel-wise and ignores the structural relationship in a neighborhood. Meanwhile, it also ignores the correlation between the noise signal and the original image.

The past decades have witnessed the rapid development of image quality assessment (IQA) methods, which aim to gauge the perceptual quality of natural images like human

This work is supported by Natural Science Foundation of China (No. 90920003 and No. 61172163) and HK RGC General Research Fund (PolyU 5315/12E).

does. In the case that the reference image is accessible, the resulting full reference (FR) IQA model can be utilized as an image perceptual fidelity term to mimic the human perception of image quality. Numerous FR IQA models have been proposed by using mutual information [18, 19], structure comparison [25, 28, 30, 31], and human vision system simulation [10], etc. Though these IQA models can better predict the perceptual quality of a test image than MSE, they do not profit much the applications such as perceptual image coding, perceptual image representation and restoration, etc. The main reason lies in that those IQA models are nonlinear and lack the properties such as differentiability and convexity. Most of them are not a valid distance metric and have much higher computational complexity than MSE.

Among the existing FR IQA models, the Structural Similarity (SSIM) [25] index is probably the most influential one. It has been widely studied by researchers and many IQA models were developed based on the principle of SSIM [2, 30]. In perceptual image coding and restoration, more visually comfortable results have been reported by using SSIM to guide the algorithm design [16, 17, 23]. The underlying assumption of SSIM is that the human vision system is highly adapted to extract structural information in the viewing field. For a pair of reference image  $\mathbf{r}$  and distorted image  $\mathbf{d}$ , SSIM estimates the perceptual quality of  $\mathbf{d}$  from three aspects: luminance, contrast and structure. It uses the following formula to compute the perceptual fidelity index of  $\mathbf{d}$ :

$$\text{SSIM}(\mathbf{r}, \mathbf{d}) = \frac{2\mu_r\mu_d + c_1}{\mu_r^2 + \mu_d^2 + c_1} \cdot \frac{2\sigma_{r,d} + c_2}{\sigma_r^2 + \sigma_d^2 + c_2} \quad (1)$$

where  $\mu_r$  and  $\mu_d$  are the local mean luminance of  $\mathbf{r}$  and  $\mathbf{d}$ ;  $\sigma_r^2$  and  $\sigma_d^2$  are the local variance;  $\sigma_{r,d}$  is the local covariance between  $\mathbf{r}$  and  $\mathbf{d}$ .

The mathematical properties of SSIM have been carefully analyzed in [1], revealing some desirable properties of SSIM such as quasi-convexity, the triangular inequality, differentiability, etc. However, compared with the  $l_2$ -norm based MSE, SSIM is still much more difficult to be used as an objective function to minimize, due to its complicated gradient computation [1, 26]. It is highly demanded to find an MSE-like fidelity measure, which could inherit some of the appealing merits of MSE while being highly perceptual fidelity aware. This will not only simplify the computation in IQA applications, but also facilitate significantly the use of perceptual fidelity measure in applications such perceptual image coding and perceptual image restoration.

With the above considerations, in this work we aim to develop an MSE-like  $l_2$ -norm perceptual fidelity measure. We propose to introduce an  $l_2$ -norm structural error term to the original MSE so that the resulting measure can be more perceptual fidelity aware. The structural error term can be simply designed by using the linear gradient operator or Laplacian of Gaussian operator. Our experimental

results show that such a Structural MSE (SMSE) leads to very competitive IQA performance with SSIM.

More surprisingly, we show that by using certain linear structure extractors, the proposed SMSE can be turned into a Gaussian smoothed MSE, i.e., the Euclidean distance between reference image  $\mathbf{r}$  and distorted image  $\mathbf{d}$  after Gaussian smooth filtering. We call the resulting measure Perceptual-fidelity Aware MSE (PAMSE), which provides a very simple MSE-like formula to calculate the image perceptual fidelity and achieves rather better perceptual consistency than SSIM. In addition, PAMSE can have great potentials in perceptual image coding and restoration.

Section 2 presents the framework of SMSE and the setting of its structural term. Section 3 presents the derivation of PAMSE from SMSE. Section 4 presents the experimental results and Section 5 concludes the paper.

## 2. A Framework of Structural MSE

### 2.1. The framework

As discussed in the Introduction section, Mean Squared Error (MSE) possesses many desirable mathematical properties and it is easy to use. It would be very interesting if we could design an MSE-like  $l_2$ -norm distance metric, which could inherit some important advantages of MSE while improving the consistency with human perception of image quality. Since human visual system is sensitive to image local structures, we propose to amend MSE a little so that the modified MSE can count more the structural information in the fidelity estimation.

Figure 1(a) shows the proposed framework of SMSE for IQA. The original MSE term  $\|\mathbf{r} - \mathbf{d}\|_2^2/N$  is preserved in SMSE to measure the signal energy preservation. In parallel, a linear structure extractor, denoted by  $\mathbf{S}$ , is used to extract the structure features of  $\mathbf{r}$  and  $\mathbf{d}$ , and a term  $\|\mathbf{S}\mathbf{r} - \mathbf{S}\mathbf{d}\|_2^2/N$  is introduced to measure the structural error in image  $\mathbf{d}$ . The resulting SMSE index is:

$$\begin{aligned} \text{SMSE}(\mathbf{r}, \mathbf{d}) &= \frac{1}{N}(\|\mathbf{r} - \mathbf{d}\|_2^2 + \alpha\|\mathbf{S}\mathbf{r} - \mathbf{S}\mathbf{d}\|_2^2) \\ &= \frac{1}{N}(\|\mathbf{r} - \mathbf{d}\|_2^2 + \alpha\|\mathbf{S}(\mathbf{r} - \mathbf{d})\|_2^2) \end{aligned} \quad (2)$$

where  $\alpha$  is a (negative or positive) constant to adjust the contribution of structural error term  $\|\mathbf{S}\mathbf{r} - \mathbf{S}\mathbf{d}\|_2^2$  to the final index. Here we restrict the structure feature extractor  $\mathbf{S}$  to be linear in order to make the whole SMSE model linear and easy to apply.

SMSE is an amendment of MSE by introducing a structural error term. It is necessary to know if SMSE can still be a valid distance metric, like the original MSE does. Eq. 2 can be re-written as:

$$\text{SMSE}(\mathbf{r}, \mathbf{d}) = \frac{1}{N}(\mathbf{r} - \mathbf{d})^T(\mathbf{I} + \alpha\mathbf{S}^T\mathbf{S})(\mathbf{r} - \mathbf{d}) \quad (3)$$

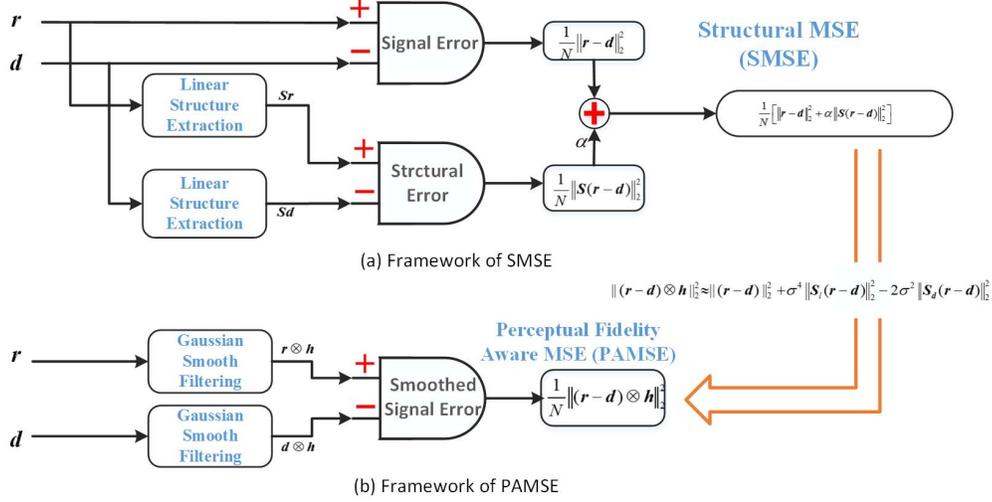


Figure 1. The framework of Structure MSE (SMSE) and Perceptual fidelity Aware MSE (PAMSE).

where superscript  $'T'$  means transpose and  $\mathbf{I}$  denotes the identity matrix. Let  $\mathbf{M} = \mathbf{I} + \alpha \mathbf{S}^T \mathbf{S}$ . If  $\mathbf{M}$  is positive semi-definite (PSD), SMSE will be a valid (pseudo-) distance metric. Denote by  $\lambda_i, i = 1 \dots N$ , the eigenvalues of matrix  $\mathbf{S}^T \mathbf{S}$ . Then the eigenvalues of  $\mathbf{M}$ , denoted by  $\gamma_i, i = 1 \dots N$ , can be written as  $\gamma_i = 1 + \alpha \lambda_i$ . Then the sufficient and necessary condition that SMSE is a distance metric is:

$$\gamma_i = 1 + \alpha \lambda_i \geq 0, i = 1 \dots N. \quad (4)$$

On the other hand, since  $\mathbf{M}$  is symmetric, when it is a PSD matrix with  $\gamma_i \geq 0$ , we can find a matrix  $\mathbf{P}$  such that

$$\mathbf{M} = \mathbf{P}^T \mathbf{P} \quad (5)$$

And then the SMSE metric can be re-written as:

$$\text{SMSE}(r, d) = \frac{1}{N} \|\mathbf{P}(r - d)\|_2^2 \quad (6)$$

The matrix  $\mathbf{P}$  can be viewed as a new feature extractor (a linear projection/transform) which is able to simultaneously measure the pixel-wise energy preservation and local neighborhood-wise image structure preservation.  $\mathbf{P}$  can also be viewed as a kernel to measure the similarity between  $r$  and  $d$ .

## 2.2. Linear structure extractor $\mathbf{S}$

There are many candidates for the linear structure operator  $\mathbf{S}$  in the proposed SMSE framework. For instances,  $\mathbf{S}$  can be chosen as the gradient operator which outputs the abrupt changes of image intensity;  $\mathbf{S}$  can also be chosen as the Laplacian or Laplacian of Gaussian (LOG) operator which mimics the receptive field of the ganglion cells and the lateral geniculate nucleus (LGN) cells [5]. The linear transforms such as wavelet transform and principle component analysis can also be employed as the feature extractor  $\mathbf{S}$ . For simplicity, in this paper we only consider the gradient operator and the Laplacian operator.

### 2.2.1 Gradient operators

Image gradient is a good feature for low level and higher level vision tasks, and gradient priors are widely used in image restoration. Image gradient has also been employed in IQA and led to good results [2]. Therefore, it is a good choice to employ gradient operators as  $\mathbf{S}$  in the proposed SMSE framework.

For continuous images, the gradient is computed as the first derivative along some direction. For discrete images, the simplest analog of derivative is the first order forward (or backward) difference operator. That is, we can apply a filter  $\mathbf{f} = [1, -1]$  horizontally and vertically to the images  $r$  and  $d$  to get their gradient maps. The forward difference filtering can be written as a matrix operator, denoted by  $\mathbf{S}_d = [\mathbf{S}_d^x; \mathbf{S}_d^y]$ , where  $\mathbf{S}_d^x$  and  $\mathbf{S}_d^y$  denote the corresponding operators along horizontal and vertical directions, respectively.

The forward difference operator can be too sensitive to small intensity changes. A more commonly used gradient operator is the Gaussian gradient operator, which actually smooths the image by using a Gaussian smooth filter  $\mathbf{h}$  before applying the forward difference filter  $\mathbf{f}$ . We use  $\mathbf{S}_g = [\mathbf{S}_g^x; \mathbf{S}_g^y]$  to denote the Gaussian gradient operator.

### 2.2.2 Laplacian operators

The gradient operator computes the first derivative of images. In comparison, the Laplacian operator, denoted by  $\mathbf{S}_l$ , exploits the second derivative of images by filtering the images with the Laplacian filter  $\mathbf{l} = [0, 1, 0; 1, -4, 1; 0, 1, 0]$ . The Laplacian operator  $\mathbf{S}_l$  is very sensitive to noise and image small changes, and hence the LOG operator, denoted by  $\mathbf{S}_{log}$ , is proposed to smooth the image by Gaussian smooth filter  $\mathbf{h}$  before applying  $\mathbf{l}$ . LOG has a clear biological explanation in that its response is similar to the receptive field of the neuron cells in human visual system [5]. The zero-

crossings of the LOG filtering response indicate the edge locations, and the LOG based edge map has been successfully used to predict image quality [28, 31].

### 2.2.3 The determination of $\alpha$

Given a linear structure extractor  $\mathbf{S}$ , we need to determine the parameter  $\alpha$  to make the proposed SMSE a valid distance metric. Refer to Eq. 4,  $\alpha$  should satisfy  $1 + \alpha\lambda_i \geq 0$  for all  $i = 1 \dots N$ . Therefore, we need to calculate  $\lambda_i$ , the eigenvalues of  $\mathbf{S}^T \mathbf{S}$ .

Suppose that the elements in vectorized images  $\mathbf{r}$  and  $\mathbf{d}$  are ordered column by column. Since the operators  $\mathbf{S}_d$ ,  $\mathbf{S}_g$ ,  $\mathbf{S}_l$ , and  $\mathbf{S}_{log}$  can be interpreted as filtering operations, they can be written as a circulant matrix (each row is a cyclic shift of another) of size  $N \times N$ , whose rows are repeated versions of the corresponding filter template. Let's denote by  $\mathbf{s} = [s_1, s_2, \dots, s_N]$  the first row of  $\mathbf{S}$ . The eigenvalue  $\beta_i$  of a circulant matrix  $\mathbf{S}$  can be easily calculated from the discrete Fourier transform of  $\mathbf{s}$  [7]:

$$\beta_i = \text{DFT}(\mathbf{s}) = \sum_{k=1}^N s_k \exp(-j \frac{2\pi k i}{N}) \quad (7)$$

For circulant matrix  $\mathbf{S}$ , its eigenvectors constitute a unitary matrix  $\mathbf{U}$ , and we have  $\mathbf{S} = \mathbf{U} \text{diag}(\beta_i) \mathbf{U}^*$ , where '\*' denotes conjugate transpose and  $\text{diag}(\beta_i)$  denotes a diagonal matrix with  $\beta_i, i = 1 \dots N$ , being its diagonal elements. We have:

$$\begin{aligned} \mathbf{S}^T \mathbf{S} &= \mathbf{U} \text{diag}(\beta_i^*) \mathbf{U}^* \mathbf{U} \text{diag}(\beta_i) \mathbf{U}^* \\ &= \mathbf{U} \text{diag}(\beta_i^* \beta_i) \mathbf{U}^* \end{aligned} \quad (8)$$

Therefore, the eigenvalues  $\lambda_i$  of matrix  $\mathbf{S}^T \mathbf{S}$  can be obtained as:  $\lambda_i = \beta_i^* \beta_i = |\beta_i|^2$ , where  $|\beta_i|$  denotes the modulus of  $\beta_i$ .

Let  $\beta_{max}$  be the eigenvalue of  $\mathbf{S}$  which has the largest modulus. From Eq. 4, we have  $\alpha \geq -\frac{1}{|\beta_{max}|^2}$ . We can then let

$$\alpha = c \frac{1}{|\beta_{max}|^2}, c \geq -1 \quad (9)$$

The constant  $c$  can be adjusted to make SMSE have good IQA performance while being a valid distance metric. In Section. 4.2, we will use experiments to investigate the selection of  $c$ , and it will be seen that  $c$  can be stably set as -1 across different IQA databases.

## 3. Gaussian smooth MSE: an inherent perceptual fidelity aware MSE

In Section. 2, we proposed a framework to make MSE be able to characterize image local structural changes while measuring the global energy of image pixel-wise error. With some linear structure extractor  $\mathbf{S}$  and the associated parameter  $\alpha$  determined in Eq. 9, the new image fidelity measure SMSE is however still a valid distance metric. The metric is actually characterized by the symmetric

PSD matrix  $\mathbf{M} = \mathbf{I} + \alpha \mathbf{S}^T \mathbf{S}$ , which can be written as  $\mathbf{M} = \mathbf{P}^T \mathbf{P}$ , and consequently the SMSE metric can be written as  $\text{SMSE}(\mathbf{r}, \mathbf{d}) = \frac{1}{N} \|\mathbf{P}(\mathbf{r} - \mathbf{d})\|_2^2$ . This hints use to make a very interesting hypothesis: if matrix  $\mathbf{P}$  is a circulant matrix and each row of it comes from a filter, then the feature extraction by  $\mathbf{P}\mathbf{r}$  and  $\mathbf{P}\mathbf{d}$  becomes the linear filtering of images  $\mathbf{r}$  and  $\mathbf{d}$  by this filter. As a result, SMSE can be further reduced to the MSE between the filtered images of  $\mathbf{r}$  and  $\mathbf{d}$ . Can we find such an operator  $\mathbf{P}$ ?

The operator  $\mathbf{P}$  depends on the used linear feature extractor  $\mathbf{S}$ . In Section 2, we discussed four commonly used operators,  $\mathbf{S}_d$ ,  $\mathbf{S}_g$ ,  $\mathbf{S}_l$  and  $\mathbf{S}_{log}$ . Using each one of them, much better IQA results than MSE and competitive results with SSIM can be obtained (please refer to Section 4 for details). We can also use more than one operator in the proposed SMSE framework. Suppose that we use the difference operator  $\mathbf{S}_d$  and the Laplacian operator  $\mathbf{S}_l$ , and hence the SMSE measure becomes

$$\begin{aligned} \text{SMSE}(\mathbf{r}, \mathbf{d}) &= \frac{1}{N} (\|\mathbf{r} - \mathbf{d}\|^2 + \alpha_d \|\mathbf{S}_d(\mathbf{r} - \mathbf{d})\|^2 \\ &\quad + \alpha_l \|\mathbf{S}_l(\mathbf{r} - \mathbf{d})\|^2) \end{aligned} \quad (10)$$

where  $\alpha_d$  and  $\alpha_l$  are constants. Interestingly, it can be proved that (please refer to **Appendix A**) if we set  $\alpha_d = -2\sigma^2$  and  $\alpha_l = \sigma^4$ , where  $\sigma$  is the scale parameter of a Gaussian smooth filter  $\mathbf{h}$  and  $\sigma$  is small, then we have

$$\begin{aligned} \text{SMSE}(\mathbf{r}, \mathbf{d}) &= \frac{1}{N} (\|\mathbf{r} - \mathbf{d}\|^2 - 2\sigma^2 \|\mathbf{S}_d(\mathbf{r} - \mathbf{d})\|^2 \\ &\quad + \sigma^4 \|\mathbf{S}_l(\mathbf{r} - \mathbf{d})\|^2) \approx \frac{1}{N} \|\mathbf{h} \otimes (\mathbf{r} - \mathbf{d})\|_2^2 \end{aligned} \quad (11)$$

where ' $\otimes$ ' means the convolution operators.

Eq. 11 raises a rather surprising conclusion: we can simply make MSE perceptual fidelity aware by filtering the images  $\mathbf{r}$  and  $\mathbf{d}$  with a Gaussian smooth filter. We call such an image fidelity measure Perceptual-fidelity Aware MSE (PAMSE), defined as

$$\text{PAMSE}(\mathbf{r}, \mathbf{d}) = \frac{1}{N} \|\mathbf{h} \otimes (\mathbf{r} - \mathbf{d})\|_2^2 \quad (12)$$

Figure. 1(b) illustrates the procedures of PAMSE. Since Eq. 12 can be written as  $\text{PAMSE}(\mathbf{r}, \mathbf{d}) = \frac{1}{N} (\mathbf{r} - \mathbf{d})^T \mathbf{P}_h^T \mathbf{P}_h (\mathbf{r} - \mathbf{d})$ , where  $\mathbf{P}_h$  is the matrix form of filtering by  $\mathbf{h}$ , PAMSE is always a (pseudo-)distance metric because  $\mathbf{P}_h^T \mathbf{P}_h$  is a PSD matrix.

## 4. Experiments and analysis

### 4.1. Databases and evaluation protocols

To validate the IQA performance of SMSE and PAMSE, we test them on three benchmark IQA databases: LIVE [20], CSIQ [10] and TID2008 [15]. The images in these databases are generated through different distortion channels and are all assigned with a subjective quality/distortion score. The consistency between the subjective

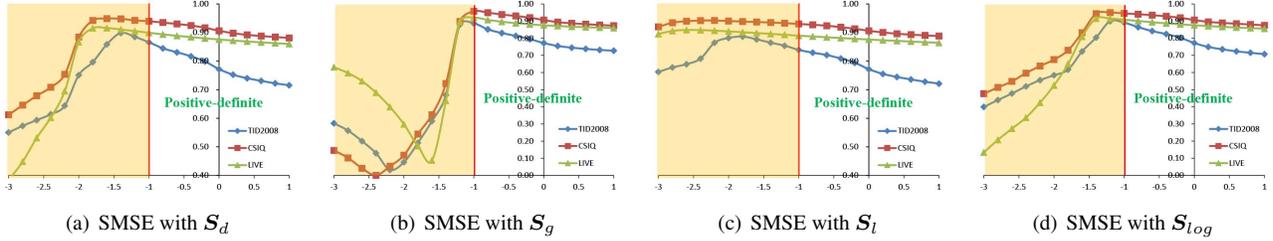


Figure 2. The IQA results (in terms of SRC) of SMSE versus  $c$  on the three databases. Note that when  $c$  is 0, SMSE reduces to MSE, and when  $c \geq -1$ , SMSE is a valid distance metric.

scores and the predicted scores can be examined in terms of Spear rank order correlation coefficient (SRC), Pearson correlation coefficient (PCC) and the root mean squared error (RMSE). Note that PCC and RMSE are calculated after a logistic regression between the predicted scores and the subjective scores [8]. A good IQA measure should have high SRC, PCC scores and low RMSE score.

The LIVE database consists of 29 reference images and 779 distorted images generated from JPEG compression (JPEG), JPEG2000 compression (JP2K), additive white noise (AWN), Gaussian blur (GB), and simulated fast fading Rayleigh channel. The CSIQ database consists of 30 reference images and 866 distorted images of JPEG2000, JPEG, AWN, GB, additive pink Gaussian noise (PGN) and contrast change. We exclude the the last type in experiments since it's not a structural distortion. The TID2008 database is the largest IQA dataset so far. There are 25 reference images and 1,700 distorted images of 17 distortion types. Since the last 4 types are non-structural distortions which are out the interest of this work, we only validate the IQA performance for the first 13 types of distorted images.

## 4.2. Implementation and results of SMSE

We implement the proposed SMSE by using the operators  $S_d$ ,  $S_l$ ,  $S_g$  and  $S_{log}$ , respectively. Since these operators are actually the matrix form of spatial filters, in implementation we use spatial convolution to compute the structural features  $Sr$  and  $Sd$ . As described in Section 2.2.3, we adjust  $c$  to find a good  $\alpha$ , and  $c$  should be no less than -1 to ensure SMSE a valid distance metric.

For SMSE with  $S_d$ , the corresponding spatial filter is  $f = [1, -1]$  along horizontal and vertical directions. It can be easily computed that the associated  $|\beta_{max}|^2$  is 8. In Figure 2(a), we plot the curves of  $S_d$ 's IQA results (in terms of SRC) versus  $c$  on the three databases. Note that we vary  $c$  from -3 to 1 with step-length 0.2 for a more comprehensive observation of the performance. The best SRC results of SMSE with  $S_d$  occur at  $c = -1.8$ ,  $c = -1.8$  and  $c = -1.4$ , respectively, for LIVE, CSIQ and TID2008 databases. However, those values will make SMSE an invalid distance metric. Therefore, we set  $c$  to -1, which is the critical point to ensure SMSE a distance metric. When

$c = 0$ , SMSE reduces to MSE. A clear performance improvement of SMSE over MSE on all the three databases can be observed from the curves. For instance, on the TID2008 database, the performance gain of SMSE (with  $S_d$ ) over MSE is 12.19%.

For SMSE with  $S_g$ , we will smooth the images  $r$  and  $d$  by using a Gaussian smooth filter  $h(i, j) = \frac{1}{2\pi\sigma^2} \exp(-\frac{i^2+j^2}{2\sigma^2})$  before applying the forward difference filter  $f$ . Through experiments, we found that setting  $\sigma = 0.5$  could lead to good results on all the three IQA databases. Then it can be computed that the  $|\beta_{max}|^2$  associated with  $S_g$  is 1.53. In Figure 2(b), we plot the SRC curves of SMSE (with  $S_g$ ) versus  $c$ . It can be seen that SMSE achieves its best or nearly best results at  $c = -1.2$  and  $c = -1$ . By letting  $c = -1$ , the performance improvements of SMSE over MSE are 14.85%, 5.62% and 5.04%, respectively, on TID2008, CSIQ and LIVE databases. Note that for SMSE with  $S_g$ , the performance will drops dramatically when  $c < -1.2$ .

For SMSE with  $S_l$ , the corresponding spatial filter is  $f = [0, 1, 0; 1, -4, 1; 0, 1, 0]$ . The associated  $|\beta_{max}|^2$  is 64. In Figure 2(c), we plot the curves of  $S_l$ 's IQA results. The best SRC results of SMSE with  $S_l$  occur at  $c = -2.6$ ,  $c = -2.2$  and  $c = -1.8$ . When  $c$  is set to the critical point -1, the performance improvement of SMSE (with  $S_l$ ) over MSE is 8.7%, 2.6% and 1.5%, respectively, on TID2008, CSIQ and LIVE databases.

For SMSE with  $S_{log}$ , we also need to set the scale parameter of the Gaussian smooth filter  $h$ . Through experiments, we set  $\sigma = 0.5$ , and the  $|\beta_{max}|^2$  associated with  $S_{log}$  is 41.88. The SRC curves of SMSE with  $S_{log}$  are shown in Figure 2(d). The best results occur at  $c = -1.4$ ,  $c = -1.4$  and  $c = -1.2$ , respectively, on LIVE, CSIQ and TID2008 databases. The results of SMSE at  $c = -1$  are close to the best results, and SMSE has 15.39%, 4.26%, 3.69% improvements over MSE on TID2008, CSIQ and LIVE, respectively.

In Table 1, we list the IQA results of SMSE with  $c = -1$  for all the four operators and compare them with MSE, S-SIM, as well as other popular IQA methods. More discussions about why  $c$  is negative will be given in Section 4.4.

Table 1. IQA performance comparison. In each column, the best three results are highlighted.

	LIVE (779 images)			CSIQ (750 images)			TID2008 (1300 images)			Weighted average	
	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC
MSE	0.8756	0.8739	13.283	0.9060	0.8882	0.125	0.7718	0.7649	0.852	0.8362	0.8279
SMSE with $S_d$	0.8992	0.8953	12.170	0.9394	0.9155	0.110	0.8659	0.8491	0.649	0.8946	0.8795
SMSE with $S_i$	0.8890	0.8890	12.647	0.9299	0.9049	0.116	0.8393	0.8230	0.751	0.8770	0.8629
SMSE with $S_g$	0.9197	0.9162	10.946	<b>0.9569</b>	0.9188	0.108	0.8864	0.8711	0.637	<b>0.9143</b>	0.8963
SMSE with $S_{log}$	0.9091	0.9046	11.644	0.9449	<b>0.9421</b>	<b>0.091</b>	<b>0.8934</b>	0.8764	0.890	0.9114	0.9016
PAMSE	0.9280	0.9243	10.428	<b>0.9565</b>	0.9163	0.109	<b>0.9162</b>	<b>0.9018</b>	<b>0.571</b>	<b>0.9301</b>	<b>0.9119</b>
IFC [19]	0.9259	0.9268	10.264	0.8827	0.8912	0.124	0.7589	0.8007	0.792	0.8383	0.8599
SSIM [25]	<b>0.9479</b>	<b>0.9451</b>	<b>8.927</b>	0.9247	0.9188	0.108	0.8742	0.8530	0.690	0.9082	0.8962
MAD [10]	0.9438	0.9394	9.368	<b>0.9604</b>	0.8881	0.125	0.8694	0.8306	0.736	0.9142	0.8762
VIF [18]	<b>0.9636</b>	<b>0.9604</b>	<b>7.614</b>	0.9282	0.9321	0.099	0.8731	<b>0.8938</b>	<b>0.593</b>	0.9130	<b>0.9226</b>
rNSE [28]	0.9242	0.9211	10.639	0.9405	<b>0.9495</b>	<b>0.086</b>	0.8622	0.8766	0.636	0.9002	0.9083
FSIM [30]	<b>0.9634</b>	<b>0.9597</b>	<b>7.674</b>	0.9544	<b>0.9541</b>	<b>0.082</b>	<b>0.9199</b>	<b>0.9068</b>	<b>0.557</b>	<b>0.9412</b>	<b>0.9341</b>

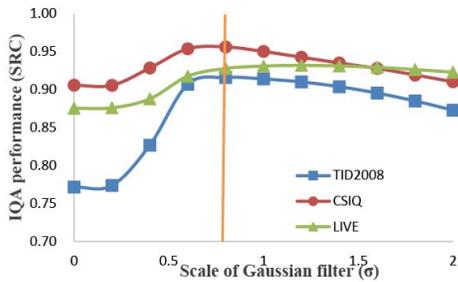


Figure 3. The IQA results (in terms of SRC) of PAMSE versus  $\sigma$  on the three databases.

### 4.3. Implementation and results of PAMSE

The implementation of PAMSE is even simpler than SMSE because we only need to set the scale parameter  $\sigma$  of the Gaussian smooth filter  $h$ . Refer to our derivation in **Appendix A**,  $\sigma$  cannot be big. To find a good value of  $\sigma$ , we tested the performance of PAMSE on the three IQA databases by varying  $\sigma$  from 0 to 2.0. Note that when  $\sigma = 0$ , PAMSE reduces to MSE. In Fig. 3, we plot the SRC curves of PAMSE versus  $\sigma$ . As  $\sigma$  increases, PAMSE shows much higher SRC than MSE for all the three databases. Based on these curves, we set  $\sigma$  to 0.8 for PAMSE on all the databases, and list the IQA results of PAMSE in Table. 1.

### 4.4. Comparison and discussions

In this sub-section, we compare the IQA results of SMSE and PAMSE with MSE and SSIM. The results of some state-of-the-art IQA methods (e.g., FSIM [30], VIF [18], IFC [19], rNSE [28] and MAD [10]) are also presented for reference. We would like to stress that the goal of this work is not to develop an IQA algorithm which outperforms all state-of-the-arts. The key message we would like to convey is that with some small adaptation, an MSE-like image perceptual fidelity measure can be obtained, which is very easy and efficient to implement while offering very competitive IQA results.

Table. 1 lists the results (in terms of SRC, PCC and

RMSE) of the competing methods. The results of SMSE and PAMSE are obtained with the parameters fixed in Sections 4.2 and 4.3, respectively. For all the other methods, we used the codes provided by the original authors with the default parameter setting.

From Table. 1, the following conclusions can be drawn. 1) On all the databases, SMSE and PAMSE outperform MSE with a large margin. 2) On databases CSIQ and TID2008, both SMSE and PAMSE have similar results to the best IQA methods. 3) By averaging the results over the three databases (weighted by the image number), PAMSE only lags behind FSIM, which is currently the best IQA method. 4) Compared with SSIM, which is the representative of modern IQA measures and has been extensively studied in the past decade, SMSE gives competitive performance in average, while PAMSE performs better than it. In summary, SMSE and PAMSE are among the state-of-the-art FR IQA measures.

Figure. 4 shows the scatter plots of the subjective score versus the predicted score by SSIM and PAMSE on the CSIQ database. Each point represents a distorted image. It can be clearly observed that the distributions by PAMSE is more consistent with the subjective score than SSIM (i.e., the predicted scores are more linear to the subjective scores). Similarity scatter plots are observed for SMSE models.

Though PAMSE is originated from SMSE, overall it shows better IQA performance than SMSE. The reason lies in two folds. Firstly, PAMSE implicitly uses two linear structure extractors, while SMSE uses only one linear structure extractor. Secondly, PAMSE is however free of the explicit use of any structure extractor, which is affected to some extent by the numerical implementation.

In spite of the promising performance of SMSE and PAMSE, at first sight the subtraction of the structure error term (i.e.,  $c = -1$ ) in SMSE and the smoothing Gaussian filter in PAMSE seem counterintuitive. It's widely accepted that it's the structures in natural images that attract hu-

Table 2. Runtime of the competing IQA methods.

Methods	Runtimes (s)	Ratio to MSE
MSE	0.0021	1
PAMSE	0.0056	2.66
SMSE with $S_d$	0.0092	4.35
SMSE with $S_l$	0.0057	2.72
SMSE with $S_g$	0.0103	4.84
SMSE with $S_{log}$	0.0071	3.32
SSIM [25]	0.0213	10.02
rNSE [28]	0.1167	54.97
FSIM [30]	0.4343	204.56
IFC [19]	0.9372	441.44
VIF [18]	0.9743	458.95
MAD [10]	2.0715	986.43

man’s attention. Intuitively, in SMSE the structure error term should be “added”, but not be “subtracted”. In PAMSE, the Gaussian filtering smoothes some texture information, which again seems unreasonable for the IQA tasks. Actually, both of them are plausible based on the contrast sensitivity property of human visual system (HVS). The contrast sensitivity function (CSF) describes the visibility of signals as a function of spatial frequency [3], and it shows the reduced sensitivity of HVS at lower and higher frequencies. MSE measures image distortion equally in all frequencies, and thus over-estimates the perceptual distortion. By subtracting the structure error term or using Gaussian filtering, the contribution of high frequency in measuring the difference between counterpart images is reduced, which is consistent with CSF. The low frequency information is remained for the sake of energy preservation. Actually, in many halftone image quality analysis methods [4, 13, 14], the CSF filter is used to weight the error in a transformed domain, and it is often approximated by Gaussian filters, which are believed to work better for suprathreshold distortions and be more robust to variations in viewing distances.

#### 4.5. Running time

At last, let’s discuss the complexity of SMSE and PAMSE. Clearly, one very attractive advantage of SMSE and PAMSE is their efficiency compared with other major IQA models such as SSIM, IFC, VIF, rNSE, MAD, and FSIM, etc. For SMSE, apart from the term  $\|r - d\|_2^2$ , we only need to compute  $\|S(r - d)\|_2^2$ , which can be obtained by convolving the error signal  $(r - d)$  with the filter corresponding to operator  $S$ , and then calculating the  $l_2$ -norm of the filtering response. Suppose that the size of the filter is  $k$ , then in SMSE the additional cost to MSE is just  $O(Nk)$ , which is significantly lower than other sophisticated IQA methods. Similar analysis goes to PAMSE.

In Table. 2, we list the running time of the competing methods to process an image of size  $512 \times 512$ . All algorithms were run on a desktop with Intel Core i5-2300 CPU @2.8GHz and 4G RAM. As can be seen, the time cost of

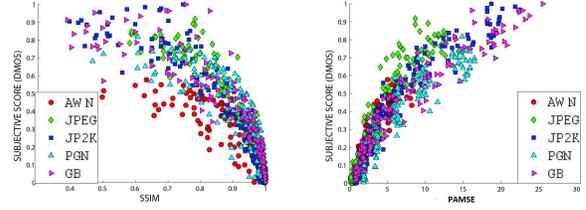


Figure 4. The scatter plots of the subjective score versus the results of IQA models on the CSIQ database. The different distortion types are indicated with various marker.

SMSE and PAMSE is only 2.6 ~ 4.8 times of that of MSE. The cost of SSIM is 10 times of that of MSE, while the other IQA methods are significantly slower than MSE.

## 5. Conclusions

By adding an  $l_2$ -norm structure error term to the original Mean Squared Error (MSE) index, in this paper we proposed a simple yet very effective framework, namely Structural MSE (SMSE), for image quality assessment (IQA). When using difference operator and Laplacian operator to extract the structure error, SMSE becomes the MSE between Gaussian smoothed reference and distorted images, and we call the corresponding SMSE measure Perceptual-fidelity Aware MSE (PAMSE). SMSE and PAMSE are very simple to implement, highly efficient, very effective, and they are valid distance metrics. In particular, PAMSE is among the best ones for IQA tasks, works much better than the well-known SSIM index. Meanwhile, SMSE and PAMSE have good potentials to be used as objective functions in perceptual quality based image processing tasks.

## Appendix A: Proof of Eq. 11

For the convenience of our following development, we denote  $r$  and  $d$  as two-dimensional (2D) continuous functions, and  $h$  the continuous 2D Gaussian smooth function with scale  $\sigma$ . We use  $\nabla$  and  $\Delta$  to denote the continuous counterpart of difference operator  $S_d$  and Laplacian operator  $S_l$ , respectively. Based on the Gabor formulas [6, 11], when  $\sigma$  is small, we have:

$$h \otimes r = r + \sigma^2 \Delta r + o(\sigma^2) \approx r + \sigma^2 \Delta r \quad (A1)$$

$$h \otimes d = d + \sigma^2 \Delta d + o(\sigma^2) \approx d + \sigma^2 \Delta d \quad (A2)$$

In (A1) and (A2), we ignored the higher order terms of  $\sigma$ . There is:

$$h \otimes (r - d) \approx (r - d) + \sigma^2 (\Delta r - \Delta d) \quad (A3)$$

We have:

$$\begin{aligned} \|\mathbf{h} \otimes (\mathbf{r} - \mathbf{d})\|_2^2 &\approx \|(\mathbf{r} - \mathbf{d}) + \sigma^2 (\Delta \mathbf{r} - \Delta \mathbf{d})\|_2^2 \\ &= \int_{\mathcal{V}} [(\mathbf{r} - \mathbf{d})^2 + \sigma^4 (\Delta \mathbf{r} - \Delta \mathbf{d})^2] \end{aligned}$$

$$\begin{aligned}
& + 2\sigma^2(\mathbf{r} - \mathbf{d})\Delta(\mathbf{r} - \mathbf{d})dV \\
= & \int_V [(\mathbf{r} - \mathbf{d})^2 + \sigma^4(\Delta\mathbf{r} - \Delta\mathbf{d})^2 \\
& - 2\sigma^2\nabla(\mathbf{r} - \mathbf{d}) \cdot \nabla(\mathbf{r} - \mathbf{d})]dV \\
= & \int_V [(\mathbf{r} - \mathbf{d})^2 + \sigma^4\Delta(\mathbf{r} - \mathbf{d})^2 - 2\sigma^2|\nabla(\mathbf{r} - \mathbf{d})|^2]dV \\
= & \|\mathbf{r} - \mathbf{d}\|_2^2 + \sigma^4\|\Delta(\mathbf{r} - \mathbf{d})\|_2^2 - 2\sigma^2\|\nabla(\mathbf{r} - \mathbf{d})\|_2^2 \tag{A4}
\end{aligned}$$

In the 3<sup>rd</sup> row of (A4), we use the Gauss theorem, i.e., the divergence theorem [22]:

$$\int_V (\vec{a} \cdot \nabla b + b \nabla \cdot \vec{a})dV = \int_{\partial V} b\vec{a} \cdot d\vec{s}$$

with  $\vec{a} = \nabla b$  and suppose that  $\int_{\partial V} b \nabla b \cdot d\vec{s} = 0$  on the boundary.

In the discrete case, by replacing  $\Delta$  and  $\nabla$  as  $S_l$  and  $S_d$ , respectively, we have Eq.11.

## References

- [1] D. Brunet, E. R. Vrscay, and Z. Wang. On the mathematical properties of the structural similarity index. *IEEE TIP*, 21:1488–1499, 2012. 2
- [2] G. Cheng, J. Huang, C. Zhu, Z. Liu, and L. Cheng. Perceptual image quality assessment using a geometric structural distortion model. In *IEEE ICIP*, pages 325–328, 2010. 2, 3
- [3] J. Daitch and D. Green. Contrast sensitivity of the human peripheral retina. *Vision Research*, 9(8):947–952, 1969. 7
- [4] S. J. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE/IS&T Electronic Imaging*, pages 2–15, 1992. 7
- [5] J. E. Dowling. The retina: an approachable part of the brain. *Optometry and Vision Science*, 65, 1988. 3
- [6] D. Gabor. Information theory in electron microscopy. *Laboratory Investigation*, 14:801–807, 1965. 7
- [7] R. M. Gray. Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory*, 2:155–239, 2005. 4
- [8] V. Q. E. Group et al. Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II. *VQEG*, Aug, 2003. 5
- [9] K. Hirakawa and T. Parks. Image denoising using total least squares. *TIP*, 15(9):2730–2742, 2006. 1
- [10] E. Larson and D. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1), 2010. 2, 4, 6, 7
- [11] M. Lindenbaum, M. Fischer, and A. M. Bruckstein. On Gabor’s contribution to image enhancement. *Pattern Recognition*, 27:1–8, 1994. 7
- [12] J. L. Mannos and D. J. Sakrison. The effects of a visual fidelity criterion of the encoding of images. *IEEE Transactions on Information Theory*, 20:525–536, 1974. 1
- [13] T. Mitsa, K. L. Varkur, and J. R. Alford. Frequency-channel-based visual models as quantitative quality measures in halftoning. In *IS&T/SPIE Electronic Imaging*, pages 390–401, 1993. 7
- [14] J. B. Mulligan and A. J. Ahumada Jr. Principled halftoning based on human vision models. In *SPIE/IS&T Electronic Imaging*, pages 109–121, 1992. 7
- [15] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(10):30–45, 2009. 4
- [16] A. Rehman and Z. Wang. SSIM-based non-local means image denoising. In *IEEE ICIP*, pages 217–220, 2011. 1, 2
- [17] A. Rehman, Z. Wang, D. Brunet, and E. R. Vrscay. SSIM-inspired image denoising using sparse representations. In *IEEE ICASSP*, pages 1121–1124, 2011. 1, 2
- [18] H. Sheikh and A. Bovik. Image information and visual quality. *TIP*, 15(2):430–444, 2006. 2, 6, 7
- [19] H. Sheikh, A. Bovik, and G. De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *TIP*, 14(12):2117–2128, 2005. 2, 6, 7
- [20] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik. Live image quality assessment database release 2 (2005). 4
- [21] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based image retrieval at the end of the early years. *IEEE TPAMI*, 22:1349–1380, 2000. 1
- [22] M. Spiegel, S. Lipschutz, and D. Spellman. Vector Analysis (2nd Edition). *Foundations and Trends in Communications and Information Theory*, 2009. 8
- [23] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao. Rate-SSIM optimization for video coding. In *IEEE ICASSP*, pages 833–836, 2011. 1, 2
- [24] Z. Wang and A. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, 26(1):98–117, 2009. 1
- [25] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 2, 6, 7
- [26] Z. Wang and E. P. Simoncelli. Maximum differentiation (mad) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12), 2008. 2
- [27] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPR*, pages 74–81, 2011. 1
- [28] W. Xue and X. Mou. An image quality assessment metric based on non-shift edge. In *ICIP*, pages 3309–3312, 2011. 2, 4, 6, 7
- [29] H. Yan, L. Cervino, X. Jia, and S. B. Jiang. A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT. *Physics in Medicine and Biology*, 57:2063–2080, 2012. 1
- [30] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: a feature similarity index for image quality assessment. *IEEE TIP*, 20(8):2378–2386, 2011. 2, 6, 7
- [31] M. Zhang, X. Mou, and L. Zhang. Non-shift edge based ratio (nser): An image quality assessment metric based on early vision features. *IEEE SPL*, (99):1–1, 2011. 2, 4