A New Image Quality Metric for Image Auto-Denoising[*]

Xiangfei Kong
City University of Hong Kong

Kuan Li
National Defense University of China

Qingxiong Yang[†]
City University of Hong Kong

Liu Wenyin
Shanghai University of Electric Power

Ming-Hsuan Yang
University of California at Merced

http://www.cs.cityu.edu.hk/~qiyang/publications/iccv-13/

Abstract

This paper proposes a new non-reference image quality metric that can be adopted by the state-of-the-art image/video denoising algorithms for auto-denoising. The proposed metric is extremely simple and can be implemented in four lines of Matlab code[1]. The basic assumption employed by the proposed metric is that the noise should be independent of the original image. A direct measurement of this dependence is, however, impractical due to the relatively low accuracy of existing denoising method. The proposed metric thus aims at maximizing the structure similarity between the input noisy image and the estimated image noise around homogeneous regions and the structure similarity between the input noisy image and the denoised image around highly-structured regions, and is computed as the linear correlation coefficient of the two corresponding structure similarity maps. Numerous experimental results demonstrate that the proposed metric not only outperforms the current state-of-the-art non-reference quality metric quantitatively and qualitatively, but also better maintains temporal coherence when used for video denoising.

1. Introduction

Image denoising is one of the most fundamental tasks that finds numerous applications. It aims at recovering the original image signal as much as possible from its noise-corrupted version. Numerous denoising algorithms have been proposed in the literature. Notwithstanding the demonstrated success, these algorithms all entail tedious manual parameterizations and prior knowledge in order to obtain the best results.

As the distortion-free reference image is not available, typical image quality assessment (**IQA**) metrics such as the mean squared error (**MSE**) and peak signal to noise ratio (**PSNR**) cannot be used to assess the denoised image quality. No-reference IQA metrics that do not use the reference image is emerging. However, most of the existing metrics [9, 6] are built based on a computationally expensive training process which requires different mean opinion scores collected from human observers.

The most related work to this one is the Q-metric by Zhu and Milanfar [28]. It selects sparse patches that have strong structure from the input noisy image with a fixed threshold, and a score will be computed at each patch based on statistical properties of the singular value decomposition. The mean of all scores is used as the metric for IQA. Nevertheless, this metric excludes contributions from homogeneous regions.

Unlike Q-metric, the proposed metric takes into account every image pixels; thus is more robust and accurate. The proposed metric is extremely simple. It is inspired by the fact that many types of the image noise (*e.g.*, photon shot noise, Gaussian noise) are independent of the original image. With the assumption of the availability of a perfect image denoising algorithm that can be used to separate a noisy image into an image containing only the noise named "method noise image" (**MNI**) [2] and a denoised image, the dependence of the image noise and the original image can be computed and used as an IQA metric. However, this is impractical due to the relatively low accuracy of existing denoising method (except when the noise level is extremely low).

The proposed metric aims at maximizing the structure similarity between the input noisy image and the extracted MNI (which corresponds to the maximization of noise reduction) around homogeneous regions and the structure similarity between the input noisy image and the denoised image (which corresponds to the maximization of structure preservation) around highly-structured regions. This paper proposes to use a high-quality denoising algorithm (*e.g.*,

---

[†]Correspondence author. The source code is available on the authors' webpage.

[1]With the availability of an implementation of the SSIM metric.

BM3D [4] or SKR [18]) to compute two structure similarity maps 1) between the input noisy image and the extracted MNI and 2) between the input noisy image and the denoised image. The linear correlation coefficient of the two structure similarity maps is used as an IQA metric. Linear correlation coefficient is a very simple solution. It is obviously not the optimal solution. However, its computational complexity is very low and has been demonstrated to be very effective and robust for a linear relationship between two variables.

Numerous experiments have been conducted to evaluate the effectiveness and robustness of the proposed metric, including both visual and numerical evaluations, real and synthetic noise, image and video noise. The experimental results demonstrate that the proposed metric not only outperforms the current state-of-the-art non-reference quality metric quantitatively and qualitatively, but also better maintains temporal coherence when used for video denoising. Human subject study is also employed to demonstrate that the proposed metric perceptually outperforms Q-metric when the obtained PSNR values are very close while the denoised images are visually different.

Although the proposed metric uses the entire input image, its computational complexity is very low because it can be decomposed into a number of box filters that can be computed very efficiently (in time linear in the number of image pixels). It is indeed even faster than Q-metric which uses very sparse local patches. It only takes around 55 ms (using Matlab) to process a 512×314 image on a 3.40GHz i7-2600 CPU and 12 GB RAM memory.

2. Related Work

The IQA metrics can be normally categorized, based on the existence of reference image, into full-reference, reduced-reference and no-reference metrics. Full-reference IQA metrics compare the processed frame with the original one free of any distortion. The IQA metrics include the (root) mean square error (MSE or RMSE) and peak signal noise ration (PSNR), and they can be computed efficiently with clear physical indications and desirable mathematical properties [19]. While these metrics are well accepted and are heavily used in some applications, they are not correlated well with the visual perception of human vision system (**HVS**), which is highly non linear and complex [19, 21, 8]. The structure similarity (**SSIM**) metric [21] makes a significant progress compared to PSNR and MSE. It is based on the hypothesis that the HVS is highly adapted for structures and less sensitive to the variance of the luminance and contrast. Variants of the SSIM metric including multi-scale SSIM [24] and information content-weighted SSIM [22] have made further progress based on perceptual preference of HVS. In addition, other metrics that exploits image structure have been proposed based on feature similarity index [26], analysis with singular vector decomposition [17, 14], and assessment on image gradient [27, 3]. Numerous findings have demonstrated the merits of using SSIM for effective image assessment [21, 26, 17, 14, 27, 3].

The proposed metric also exploits image structure for quality assessment of image denoising algorithms.

Reduced-reference IQA metrics utilize only partial information of the reference image in terms of features [21]. These features are extracted using certain models and compared to those extracted from some specific representation of the distorted images. These representations range from the wavelet coefficients [25, 13] to divisive normalization [12] and statistical distortion models [11].

On the other hand, no-reference IQA metrics do not use the reference image and the image quality is assessed blindly. Early attempt is developed for JPEG compression evaluation [23], and other extensions include just noticeable blur [7] that compares edge width and the kurtosis measurements on the transformed coefficients. More recent algorithms are developed based on the feature encoding method [9, 6] and receive more attention in recent years. The features of the training images along with different mean opinion score (DMOS) collected from human observers are coded and trained to form a dictionary. Test images are encoded via sparse coding based on the dictionary. The coding vector of a test image is used to facilitate mapping the test image quality score to the DMOS computed from the training images. While this scheme demonstrates its effectiveness, the training phase is computationally expensive.

3. Our Metric

Good parameter setting is important to guide the denoising algorithm to process a noisy image with proper balance between preserving the informative structural details and the reduction of the noise. For such purposes, the proposed method evaluates the denoised images with two measurements: (1) the noise reduction, and (2) the structure preservation. Both of these measurements are computed by using the similarity comparison from the SSIM metric. However, different from SSIM, the proposed metric operates without the reference (noise-free) image.

3.1. Overview

The proposed metric is very simple and straightforward as summarized in Algorithm 1.

Let I denote the input noisy image and $\hat{I}_h$ denote the denoised image obtained from a state-of-the-art denoising algorithm with parameter configuration h. The difference of the two is the MNI which corresponds to the estimated image noise. Let M_h denote the MNI obtained with parameter configuration h. Two maps N and P measuring the local structure similarity between the noisy image I and M_h and I and $\hat{I}_h$ are then computed based on SSIM, and the linear correlation coefficient of the two maps is used as an IQA metric. The detailed description is presented in Sec. 3.2 to 3.4.

This IQA metric can be employed by a parametric denoising algorithm for image auto-denoising. Auto-denoising is formulated as a parameter selection problem which aims at selecting the optimal parameter configuration

Algorithm 1 A Non-reference Metric for Image Denoising

Input: the noisy image I and the denoised image $\hat{I}_h$.
Output: the image quality score e.

---

1. Compute the MNI which is the difference of the input noisy image I and the denoised image $\hat{I}_h$: $M_h = I - \hat{I}_h$;
2. Compute structure similarity map N between the input noisy image I and the MNI M_h via SSIM metric (Eq. 3);
3. Compute structure similarity map P between the input noisy image I and the denoised image $\hat{I}_h$ via SSIM metric (Eq. 4);
4. Compute image quality score e as the linear correlation coefficient of the two structure similarity maps N and P.

---

h so that the denoised image $\hat{I}_h$ has the best visual quality with respect to the input noisy image I:

$$\hat{I}_h = \underset{\hat{I}^{h_i}}{\operatorname{argmax}} \, e(\hat{I}^{h_i}, I), \qquad (1)$$

where $h_i \in (h_1, h_2, ...h_K)$ representing K possible parameter configurations for the selected denoising algorithm and $e(\cdot)$ is the proposed IQA metric.

3.2. Structure Comparison

The computation of the structure comparison is first introduced by Wang et al. with the SSIM metric [21]. In our problem, we assume that a denoising algorithm does not change the luminance nor the contrast of a noisy image (which is true most of the time) and estimate the visual quality of a denoised image only with the structure comparison term. In the adopted structure comparison term, the compared digital images are firstly removed with local luminance difference and local contrast difference. Let A and B denote two local image patches, and $|A| = |B| = n$. The luminance is quantified as the mean intensity value of a local image patch $\mu_A = \frac{1}{n} \sum_{i=1}^{n} A_i$, and the contrast (*i.e.*, standard deviation) $\sigma_A = (\frac{1}{n-1} \sum_{i=1}^{n} (A_i - \mu_A)^2)^{\frac{1}{2}}$. Structure comparison of two local image patches is modeled as the correlation coefficient of the two patches with luminance and contrast normalized $(A - \mu_A)/\sigma_A$ and $(B - \mu_B)/\sigma_B$, which is equally as:

$$\mathcal{S}(A, B) = \frac{\sigma_{AB} + c}{\sigma_A \sigma_B + c}, \qquad (2)$$

where c is a small constant to ensure stability when the denominator is too small. It is the most important term in SSIM since it incorporates the comparison of spatial distribution of the image structure. The image structure compared here is independent of luminance and contrast, both of which affect the visual quality of an image less than the structure does [20].

3.3. Noise and Structure Measurements

The MNI is the difference between the input noisy image and the denoised image: $M_h = I - \hat{I}_h$. Comparing to the
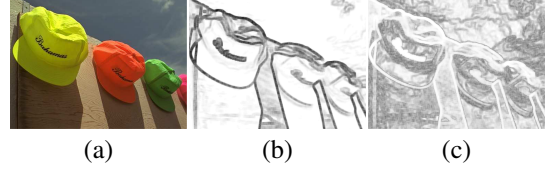

(a)   (b)   (c)

Figure 1. An example of the noise reduction and structure preservation maps. (a) is an input noisy image corrupted by WGN (with σ=10). (b) and (c) are two maps for measuring the noise reduction and structure preservation, respectively. BM3D denoising algorithm is used to obtained the denoised image with parameter σ_{est} set to σ.

MNI M_h, the noisy image I and the denoised images $\hat{I}_h$ are rich in image contents. Thus, from the MNI, the HVS can find what has been eliminated from the noisy image I easily. This property makes the MNI potentially helpful to evaluate the nature of the denoising algorithms. Note that MNI is also used by Buades et al. [1] for analyzing the image denoising algorithms.

In the proposed metric, the noise reduction measurement is designed as a map of local structure similarity measurement N computed from the noisy image I and the MNI M_h. Let I^p and M_h^p denote two local image patches of image I and M_h centered at pixel p, respectively. The noise reduction measurement at p is then computed as follows:

$$N_p = \mathcal{S}(I^p, M_h^p). \qquad (3)$$

Figure 1(b) shows an example of the noise reduction measurement computed using Eq. (3). The main motivation to use this measurement is that in homogeneous regions, a good denoising algorithm should reduce the image noise as much as possible, and the removed noise should present in the MNI at the same location. Thus, the structure of the noisy image I and the MNI M_h should be locally similar. On the other hand, if the denoising algorithm fails, the structure should be dissimilar.

Same as the noise reduction measurement, the structure preservation measurement is also a local structure similarity map P which is computed from the input noisy image I and the denoised image $\hat{I}_h$:

$$P_p = \mathcal{S}(I^p, \hat{I}_h^p). \qquad (4)$$

Figure 1 (c) presents an example of the maps computed from Eq. (4). Note that the similarity is high around highly-textured regions and low around homogeneous regions.

3.4. Integration of Measurements

The two measurements presented in Sec. 3.3 incorporate not only the spatial information of the noise reduction and structure preservation but also their energy/strength. A good denoising algorithm should maintain a good balance and maximize both terms. In regions with large N values (i.e., homogeneous regions that are not dominated by image structures), the other term P should be as small as possible, and vice versa. Considering those terms as two random variables, a natural choice for modeling such observation is the correlation coefficient, which computes the dependency

relation between them. The proposed method utilizes the simplest Pearson's linear correlation coefficient [21] to capture the linear dependency of N and P. Other rank-order based correlation such as Spearman rank-order correlation coefficient are not suitable since they change the order of the elements and thus change the spatial distribution of the measurements.

4. Experimental Results

To demonstrate the effectiveness of the proposed metric, visual and numerical evaluations are conducted on both real and synthetic noisy images and videos. A human subject study is also conducted to show that the proposed metrics can better match human perception than Q-metric. All the experiments are conducted on a 3.40GHz i7-2600 CPU and 12 GB RAM memory. The proposed metric takes around 55 ms to process a 512×314 image using a Matlab implementation.

4.1. Denoising with Real Noisy Images

This section presents experimental results to demonstrate the effectiveness of the proposed metric when real noise is presented. Two JPEG format images *Penguin* are captured with a Nikon D90 digital camera with noise generated by the filters of the complementary metal-oxide-semiconductor (CMOS) image sensor. The CMOS noise is known to be much more complicated than WGN noise [10]. The ISO value is manually set to be 6400 and 200 to obtain both the noisy and clean version of the image, respectively. The high ISO noise reduction function of the camera is turned off and the output image quality is set to be *JPEG fine*. These settings guarantee the high frequency information (including noise and details of the image content) produced by the filters on sensor is mostly retained after the demosaicking process by the camera system. The resolution of the captured JPEG image is 2144×1244 pixels.

Figure 2 demonstrates that using the proposed metric, the CBM3D [5] filter (a generalized version of the BM3D [4] algorithm for WGN denoising on color images) handles this noise well in practice. The denoised images in Figure 2 (c) and (d) are obtained from the proposed metric and Q-metric, respectively. While the noise is reduced effectively in both images, the visual quality of the image using the proposed metric can better preserve images details. More evaluations on denoising using real images are available in the supplementary material.

4.2. Denoising with Synthetic Noisy Images

The quantitative evaluation of the proposed metric and Q-metric is conducted on two image benchmark datasets *TID 2008* [15] (containing 25 images) and LIVE 2 [16] (containing 29 images), and two state-of-the-art image denoising algorithms, BM3D [4] and SKR [18], are used to compute the denoised images. The WGN is added to the test images with standard deviation σ from 5 to 20. The proposed metric and Q-metric are used to estimate the

noise level σ_{est} for the BM3D algorithm using Eq. 1, and the number of iterations itr for the SKR method. We note that incorrect parameter setting of these two denoising algorithms likely leads to either insufficient noise reduction or loss of details.

We use the PSNR metric for evaluating the quality of the denoised image in these experiments. The denoised image obtained with parameter setting optimized using the PSNR metric (which requires the ground truth image) is used as the "optimal" solution, and the PSNR value obtained from this denoised image as well as the ground truth is considered as the "optimal" PSNR value. The overall performance is then evaluated in terms of the **PSNR error**, which is defined as the absolute difference between this "optimal" PSNR value and the PSNR value of the denoised image obtained with parameter setting optimized using another metric (e.g., the proposed metric or Q-metric).

(a) BM3D on TID.   (b) BM3D on LIVE.

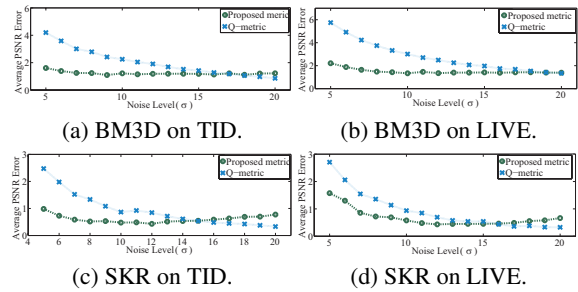(c) SKR on TID.   (d) SKR on LIVE.

Figure 3. Numerical comparison of the proposed metric and Q-metric using PSNR error. From left to right: average PSNR error on TID and LIVE database, respectively; from top to bottom: average PSNR error using BM3D and SKR image denoising algorithm, respectively. Note that the proposed metric clearly outperforms Q-metric except when the noise level is high.

The overall performance in terms of average PSNR error is presented in Figure 3. The results in Figure 3(a) show that the proposed metric has a lower PSNR error than Q-metric especially when the BM3D denoising algorithm is employed (see Figure 3(a) and (b)). Figure 4 presents three examples obtained from the BM3D algorithm with relatively low noise level ($\sigma \leq 10$) for visual comparison. We note that the proposed metric can better preserve visual details.

According to the curves reported in Figure 3, the over performance of the proposed metric is indeed a bit lower than Q-metric when the noise level is high, especially when the SKR image denoising algorithm is used. However, we note that the average PSNR errors in Figure 3(c) and (d) are both lower than 1 dB when $\sigma \geq 15$. Thus the performance of the proposed metric and Q-metric is very close to the PSNR metric. Figure 6 presents two examples obtained from SKR with relatively high noise level ($\sigma = 15$ and $\sigma = 20$) for visual comparison. In addition noise, textures details are usually missing in the results obtained by using the Q-metric. In fact, the denoised images that have higher PSNR value are not visually superior to denoised images that have lower PSNR in this case. Figure 7 presents

(a) ISO = 200 (b) ISO = 6400 (c) Proposed ($\sigma_{est} = 11$) (d) Q-metric ($\sigma_{est} = 28$)
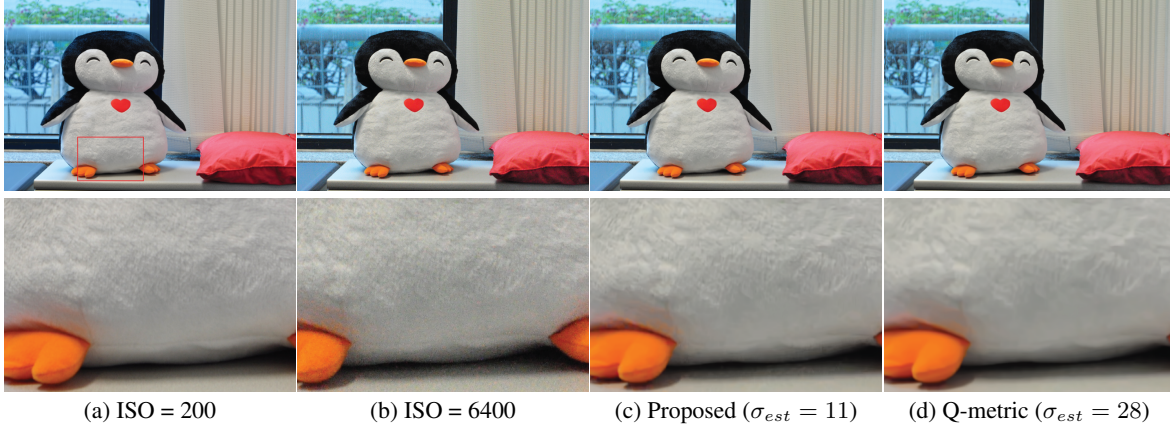
Figure 2. Evaluation using real CMOS noise. (a) is a real image captured with very little noise (when ISO is set to 200) for visual evaluation, while (b) is a noisy version of (a) captured with high ISO value (set to 6400) . (c) Denoised image obtained from CBM3D [5] with noise level estimated using the proposed metric and Q-metric. The optimal noise standard deviation values estimated for CBM3D are presented under the corresponding denoised images. Best viewed on high-resolution displays.

denoising results obtained with high noise levels ($\sigma = 19$) using the SKR denoising algorithm. The images on the left hand side have relatively lower PSNR values than the values of the images on the right hand side. However, it is hard to confirm this performance visually. Nevertheless, numerical comparison of the two metrics using PSNR error with respect to large noise levels ($\sigma \geq 25$) based on the BM3D denoising algorithm[2] is presented in Figure 5.
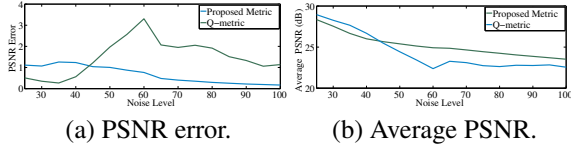

(a) PSNR error. (b) Average PSNR.

Figure 5. Numerical comparison of the proposed metric and Q-metric when the noise level is high ($\sigma \geq 25$). As can be seen in (a), the performance of the proposed metric is also close to Q-metric when evaluated using PSNR error. However, even state-of-the-art denoising algorithm (BM3D) is weak when the noise level is high (see average PSNR value in (b)); thus evaluation using PSNR error is not that suitable.

4.3. Video Denoising

This section evaluates the proposed metric with the BM3D algorithm for video denoising. The first 100 frames of the *BasketballPass* video is used in two experiments conducted (evaluation using another video is presented in the supplementary material). The images are corrupted with WGN with a constant noise level ($\sigma = 15$) in the first experiment. The PSNR errors and the estimated σ_{est} parameter values are presented in Figure 8(a) and (c), respectively. Note that the curve of the proposed metric in Figure 8(c) is flatter than that by the Q-metric, which demonstrates the temporal consistency of the proposed metric. In the second experiment, the images are also corrupted with WGN

---

[2]Results obtained using SKR algorithm is presented in the supplementary material due to page limit.

bu the noise level is changed dynamically with respect to the time domain. The PSNR error curves presented in Figure 8(b) demonstrates that the proposed metric outperforms the one by the Q-metric when the noise level is relatively low. Figure 8(d) shows that the shape of the noise levels estimated using the proposed metric better agrees with the shape of the synthetic noise levels. The Pearson correlation is used for numerical comparison. The correlation of the green curve (noise level estimated from the proposed metric) and the dark curve (synthetic noise level) in Figure 8(d) is 0.989 which are higher than the correlation of the blue curve (noise level estimated from Q-metric) and the dark curve which is only 0.937.

4.4. Human Subject Study

This section evaluates the perceptual performance of the proposed metric against the Q-metric and PSNR metric using human subject study. The TID and LIVE2 databases are used in this experiment. All the tested images are corrupted by WGN with four noise levels in {5, 10, 15, 20}. Each time, two denoised images obtained with two different IQA metrics (including the proposed metric, Q-metric and PSNR metric) are displayed on the left and right side of the WGN corrupted image, and the participant is asked to vote for the one with better visual quality. A total of 16 graduate and undergraduate students participate in this experiment. On average, 71.56% of the participants prefer the results obtained from the proposed metric to those from the Q-metric, 29.68% prefer the proposed metric to PSNR metric, and 17.5% prefer Q-metric to PSNR metric. Figure 9 shows the detailed performance with respect to different noise levels. The results show that the proposed metric outperforms Q-metric, and is comparable to PSNR metric when the noise level is relatively low.

5. Concluding Remarks

This paper proposes a new metric for automatizing existing state-of-the-art image/video denoising algorithms.

(34.15 dB, $\sigma = 5$)   (38.23 dB, $\sigma_{est} = 6$)   (33.10 dB, $\sigma_{est} = 20$)   (38. 83 dB, $\sigma_{est} = 5$)

(34.15 dB, $\sigma = 5$)   (34.6 dB, $\sigma_{est} = 9$)   (27.94 dB, $\sigma_{est} = 29$)   (36.58 dB, $\sigma_{est} = 5$)

(28.13 dB, $\sigma = 10$)   (32.8 dB, $\sigma_{est} = 13$)   (25.91 dB, $\sigma_{est} = 20$)   (33.05 dB, $\sigma_{est} = 14$)

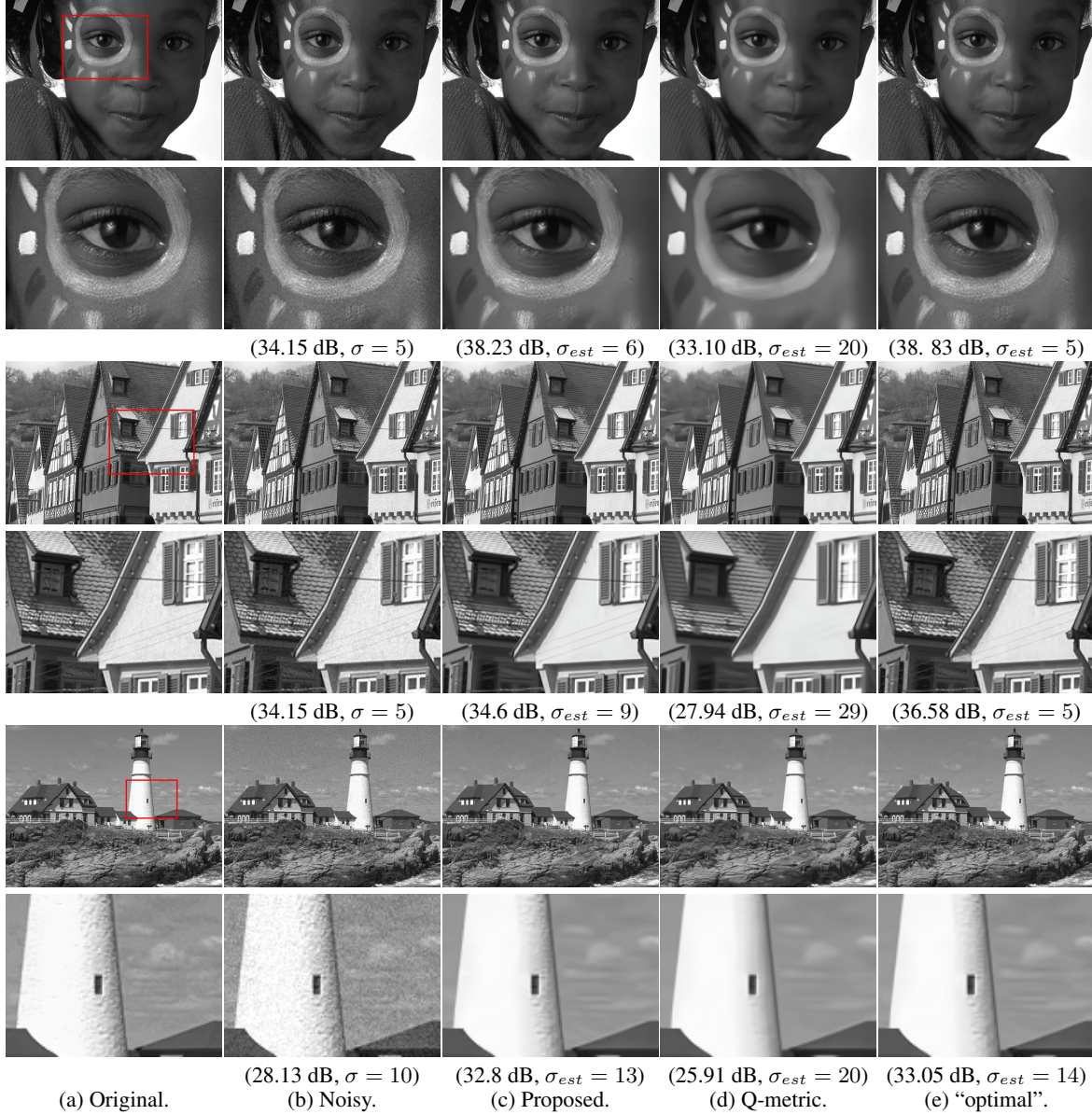(a) Original.   (b) Noisy.   (c) Proposed.   (d) Q-metric.   (e) "optimal".

Figure 4. Visual evaluation using BM3D with relatively low synthetic noise levels ($\sigma \leq 10$). (a): the original image; (b): the noisy image (corrupted using WGN with standard deviation σ); (c)-(e): denoised images obtained from the proposed metric, Q-metric and PSNR metric. Note that the proposed metric visually outperforms Q-metric for preserving structure details. Best viewed on high-resolution displays.

Specifically, the proposed metric is used to search for the optimal parameter setting of a denoising algorithm by evaluating the quality of the denoised images. The propose metric is extremely simple (can be implemented in four lines of Matlab code) and yet very robust and efficient. Experimental results demonstrate that the proposed metric outperforms the current state-of-the-art Q-metric method on two popular image quality assessment data sets and a video sequence. Our future work will extend the proposed work to other types of noise and distortion including spatially correlated noise and JPEG compression.

References

[1] A.Buades, B. Coll, and J. Morel. Self-similarity-based image denoising. *CACM*, 54(5):109–117, 2011.

[2] A. Buades, B. Coll, and J. Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005.

[3] G. H. Chen, C. L. Yang, and S. L. Xie. Gradient-based structural similarity for image quality assessment. In *ICIP*, pages 2929–2932, 2006.

[4] K. Dabo, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *TIP*, 16(8):2080–2095, 2007.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filtering with

| | (24.8 dB, $\sigma = 15$) | (29.05 dB, $itr = 8$) | (27.72 dB, $itr = 19$) | (29.05 dB, $itr = 8$) |

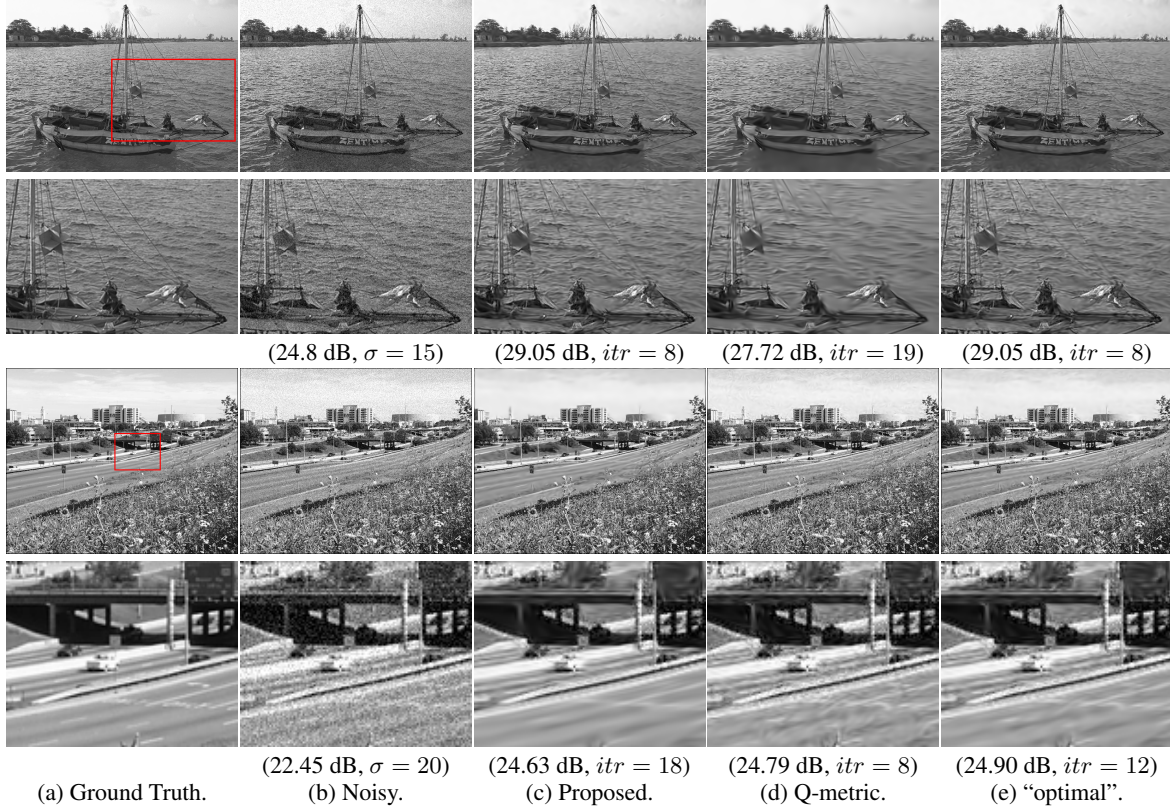| | (22.45 dB, $\sigma = 20$) | (24.63 dB, $itr = 18$) | (24.79 dB, $itr = 8$) | (24.90 dB, $itr = 12$) |
| (a) Ground Truth. | (b) Noisy. | (c) Proposed. | (d) Q-metric. | (e) "optimal". |

Figure 6. Visual evaluation using SKR with relatively high synthetic noise levels ($\sigma \geq 15$). Note that Q-metric tends to remove textures besides noise. Best viewed on high-resolution displays.

(a) Constant noise level.

(b) Varying noise level.

Evaluation using PSNR error.

(c) Constant noise level.

(d) Varying noise level.

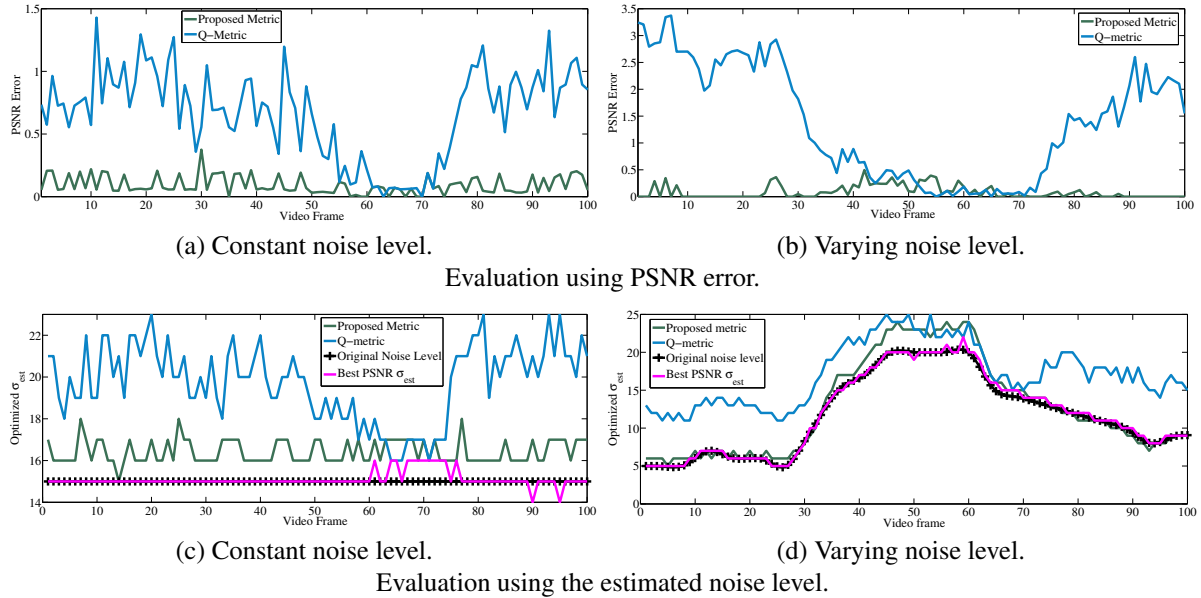Evaluation using the estimated noise level.

Figure 8. Video denoising. From left to right: (i) experimental results for synthetic WGN that has a constant noise level ($\sigma = 15$); (ii) experimental results for synthetic WGN that has dynamic changing noise levels with respect to the time domain. Note that the performance of the proposed metric is higher than the Q-metric for both situations.

grouping constraint in luminance-Chrominance space. In *ICIP*, pages I–313–I–316, 2007.

[6] D. Doermann. Unsupervised feature learning framework

for no-reference image quality assessment. In *CVPR*, pages 1098–1105, 2012.

[7] R. Ferzli and L. J. Karam. A no-reference objective image

(26.15 dB, $itr = 8$)   (27.29 dB, $itr = 30$)

(29.71 dB, $itr = 13$)   (31.00 dB, $itr = 21$)
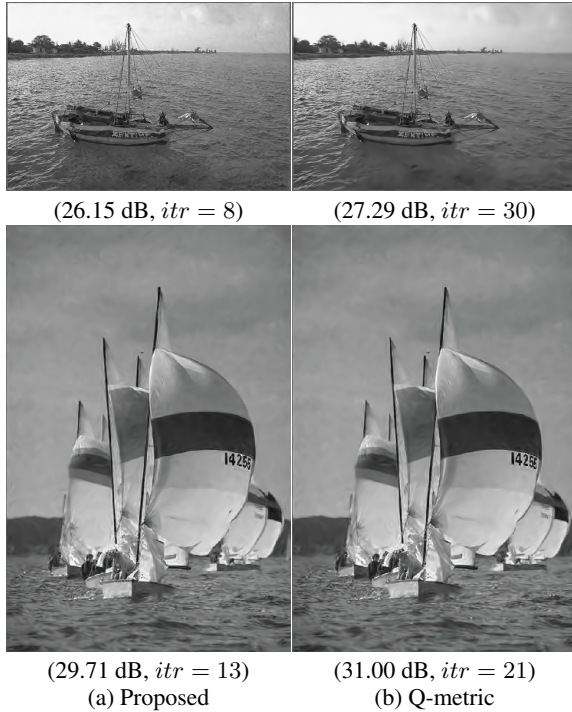(a) Proposed   (b) Q-metric

Figure 7. Visual evaluation using BM3D and SKR with high synthetic noise ($\sigma = 19$). Note that the visual perception does not always agree with the PSNR metric (which shows that the left image should have a lower performance). Best viewed on high-resolution displays.
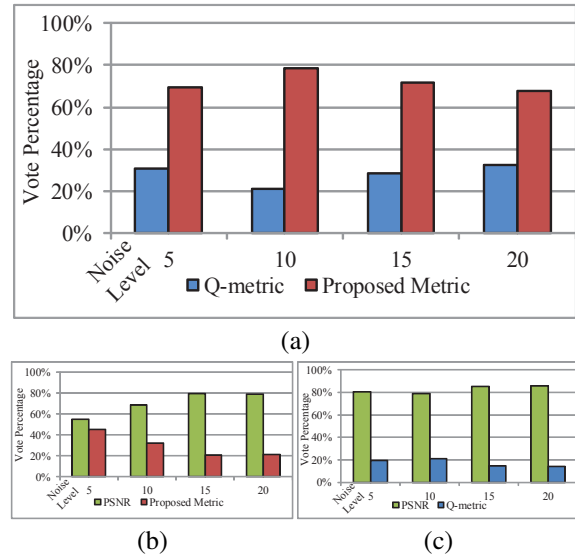

(a)

(b)   (c)

Figure 9. Human subject study. (a) subjective comparison between the proposed metric and Q-metric; (b) subjective comparison between the proposed metric and PSNR metric; (c) subjective comparison between Q-metric and PSNR metric.

sharpness metric based on the notion of just noticeable blur. *TIP*, 18(4):717–728, 2009.

[8] B. Girod. What's wrong with mean-squared error? In A. B. Watson, editor, *Digital images and human vision*, pages 207–220. MIT Press, 1993.

[9] L. He, D. Tao, X. Li, and X. Gao. Sparse representation for blind image quality assessment. In *CVPR*, pages 1146–1153, 2012.

[10] K. Hirakawa and T. W. Parks. Joint demosaicing and denoising. *TIP*, 15(8):2146–2157, 2006.

[11] T. M. Kusuma and H. J. Zepernick. A reduced-reference perceptual quality metric for in-service image quality assessment. In *Joint Workshop on Mobile Future and Symposium on Trends in Communications*, pages 71–74, 2003.

[12] Q. Li and Z. Wang. General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation. In *ICIP*, pages 1192–1195, 2008.

[13] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *JSTP*, 3(2):202–211, 2009.

[14] M. Narwaria and W. Lin. SVD-based quality metric for image and video using machine learning. *SMC-B*, 42(2):347–364, 2012.

[15] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, and F. B. M. Carli. TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Adv. Modern Radioelectron.*, 10:30–45, 2009.

[16] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik. LIVE image quality assessment database release 2.

[17] A. Shnayderman, A. Gusev, and A. Eskicioglu. An SVD-based grayscale image quality measure for local and global assessment. *TIP*, 15(2):422–429, 2006.

[18] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *TIP*, 16(2):349–399, 2007.

[19] Z. Wang and A. Bovik. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *Signal Processing Magazine*, 26(1):98 –117, 2009.

[20] Z. Wang and A. C. Bovik. A universal image quality index. *SPL*, 9(3):81–84, 2002.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural smilarity. *TIP*, 13(4):600–612, 2004.

[22] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. *TIP*, 20(5):1185 –1198, 2011.

[23] Z. Wang, H. Sheikh, and A. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *ICIP*, pages I–477 – I–480 vol.1, 2002.

[24] Z. Wang, E. Simoncelli, and A. Bovik. Multiscale structural similarity for image quality assessment. In *Asilomar Conf. on Signals, Systems and Computers*, pages 1398–1402, 2003.

[25] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *SPIE Human Vision and Electronic Imaging*, pages 149–159, 2005.

[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: A feature similarity index for image quality assessment. *TIP*, 20(8):2378–2386, 2011.

[27] J. Zhu and N. Wang. Image quality assessment by visual gradient similarity. *TIP*, 21(3):919–933, 2012.

[28] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *TIP*, 19(12):3116–3132, 2010.