J. Vis. Commun. Image R. 28 (2015) 53–59

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

No-reference quality assessment for DCT-based compressed image ☆

Ci Wang [a], Minmin Shen [b,c], Chen Yao [d,e,*]

[a] East China Normal University, Department of Computer Science and Technology, China
[b] University of Konstanz, INCIDE Center, Germany
[c] South China University of Technology, School of Software Engineering, China
[d] The Third Research Institute of Ministry of Public Security, China
[e] Shanghai Key Laboratory of Digital Media Processing and Transmission, China

ARTICLE INFO

Article history:
Received 28 January 2014
Accepted 9 January 2015
Available online 28 January 2015

Keywords:
Compression distortion
Probability model
No-reference estimate
Objective quality assessment
Image quality assessment
Gaussian distribution
Uniform distribution
Noise variance

ABSTRACT

A blind/no-reference (NR) method is proposed in this paper for image quality assessment (IQA) of the images compressed in discrete cosine transform (DCT) domain. When an image is measured by structural similarity (SSIM), two variances, i.e. mean intensity and variance of the image, are used as features. However, the parameters of original copies are actually unavailable in NR applications; hence SSIM is not widely applicable. To extend SSIM in general cases, we apply Gaussian model to fit quantization noise in spatial domain, and directly estimate noise distribution from the compressed version. Benefit from this rearrangement, the revised SSIM does not require original image as the reference. Heavy compression always results in some zero-value DCT coefficients, which need to be compensated for more accurate parameter estimate. By studying the quantization process, a machine-learning based algorithm is proposed to estimate quantization noise taking image content into consideration. Compared with state-of-the-art algorithms, the proposed IQA is more heuristic and efficient. With some experimental results, we verify that the proposed algorithm (provided no reference image) achieves comparable efficacy to some full reference (FR) methods (provided the reference image), such as SSIM.

1. Introduction

Images and videos are often compressed for efficient storage and transmission, which inevitably introduces some distortions, such as blocking and ringing artifacts. These distortions degrade visual quality, and reduce the information validity. To ensure that the end users get satisfactory experience, the user experience is often fed back to produce better image/video quality or to further save bandwidth budget by parameter reconfiguration. User experience about image quality can be assessed by subjective or objective methods. For subjective evaluation, it is required to gather the opinions of human observers in given condition; hence it is cumbersome and inapplicable in real time. On the contrary, objective score is calculated from a designed algorithm with fast implementation speed, and its results should be consistent with subjective evaluation.

According to availability of original images, objective image quality metrics are classified into three categories: (1) full-reference (FR) schemes assume that the reference image is available to be compared; (2) reduced-reference (RR) schemes are set up in the condition that only partial information of the reference image is available, for example, some features of the reference image are embedded in the compressed image and can be extracted for analysis; and (3) no-reference methods are the ones that only the compressed image is available for quality assessment. Compared with the previous two, the third category is more challenging and is studied in this paper.

Image quality assessment (IQA) algorithm utilizes the statistical features [1] of natural images and distortions, which can be collected in spatial domain or transform domain. In the past decade, some efforts have been made to develop some objective IQA algorithms. The performance of IQA algorithms can be further improved if image features are extracted and cooperated with human visual system (HVS). A series of IQA metrics, such as noise quality measure [2], structural similarity [3], visual information fidelity (VIF) [4], PSNR-HVS model [5], visual signal-to-noise ratio [6], most apparent distortion (MAD) [7] and feature similarity (FSIM) [8], have been proposed, which fit with human opinion well. Recently, Lin et al. use the phase and magnitude of Fourier transform [9] and internal generative mechanism [10] to refine IQA schemes. Liu et al. indicate that gradient information [26] captures both image contrast and structure, and give it suitable

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: Yueyang Road 76, Shanghai 200030, China.
E-mail address: yaochensing@126.com (C. Yao).

http://dx.doi.org/10.1016/j.jvcir.2015.01.006
1047-3203/

masking as well as adaptive weighting by their proposed IQA. Wu et al. classify the visual content fidelities into the primary visual information and the residual uncertainty, and evaluate their information fidelities respectively [25]. It is noted that most of the above algorithms are FR and RR ones. Though some attractive results have been produced, their applications are still limited because extra information about the original image is required. Different from them, Bovik et al. propose a NR scheme, i.e. dubbed blind/no-reference image spatial quality evaluator (BRISQUE) [11], to quantify the possible missing information of "naturalness" in the image. For the other NR IQA, Saad [22] and Moorthy [23] study the natural scene statistics (NSS) model in discrete cosine transform (DCT) and wavelet domain respectively, and then use these models to evaluate image impairment. In NR IQAs [11,22,23], low level features of an image are distilled in spatial, DCT and wavelet domains respectively, and Gaussian model or asymmetric generalized Gaussian model (AGG) model is used to estimate these feature distribution. The changes of model parameters from the compression are selected as high level features and used to predict image quality scores. However, the performance improvement of these NR IQAs is limited because the physical process of compression is not considered. Recently, Wang et al. extract statistical features through the multi-scale and orientation transform, and then develop the distortion measure from SSIM [28]. Different from them, the decompressed blocks are used as low-level features in this paper, because they are strongly related to compression. Furthermore, the proposed algorithm estimates the quantization noise and uses this noise estimation in SSIM model, rather than directly estimates image quality score, which gives better physical interpretation than other popular NR IQA models.

With mobile devices, images or videos are often coded by DCT-based compression and transmitted to accommodate the bandwidth requirements. Considerable time and resources are taken to ensure that the end user has *constant* visual experience and satisfactory quality of experience (QoE). It is widely accepted that different scenes have different bit-rate requirements for comparable visual quality; hence the efficiency of encoding and multiplexing can be significantly optimized if image quality at the receiver is measured and fed back to the sender. NR IQA can also be used to control the post-processing to improve the quality of the decoded images. For these two applications, the proposed algorithm is competent due to its high execution speed and accuracy.

In the proposed algorithm, the particular information of the compressed stream helps us to develop IQA algorithm with less computational cost and better performance. First, we use different models to describe the noise distribution on large or small DCT coefficients. Second, a machine learning based method is used to estimate the noise on zero and nonzero DCT coefficients, respectively. Finally, we modify SSIM to be a function of noise variance.

In this paper, we first overview SSIM, as it is the basis of the proposed algorithm. Subsequently, a probabilistic model of the compressed noise on spatial and DCT domain is overviewed, and an algorithm for noise estimation is proposed. SSIM is modified to accommodate the proposed noise estimation. Finally, we examine the proposed IQA and demonstrate its efficiency.

2. Overview of SSIM

Fig. 1 illustrates a general scheme for IQA. A natural image C is degraded by some distortions, such as compression, to be its inferior copy D. This process is described as

$$D = C + n^{Spatial} \tag{1}$$

where $n^{Spatial}$ is the quantization noise. The natural image and its inferior copy are fed into human visual system for mutual information analysis.

In the above model, IQA is calculated as

$$Score = \frac{I(C,F|S)}{I(C,E|S)} \tag{2}$$

where $I(\,)$ is the mutual information operator. S is the model parameter for reference image C, associated with the normalization model of various visual neurons. C and D pass through HVS model [24] to form E and F, i.e. E and F are the brain reflection of C and D.

The denominator of (2) is irrelevant to the distorted image, and it merely reflects the interaction of image features and HVS characteristic. Natural images share some common features, thus the values of $I(C,E|S)$ for different C are similar. Therefore, IQA is dominated by the numerator of (2), i.e. the mutual information between the original image C and the contaminated image D, where S stands for the HVS features of the visual neurons. Including three HVS features, i.e. luminance, contrast and structure, SSIM [3] uses their mutual information to form IQA as

$$\mathrm{SSIM}(C,D) = [l(C,D)]^{\alpha}[c(C,D)]^{\beta}[s(C,D)]^{\gamma} \tag{3}$$

where α, β, γ are parameters to control the relative importance of the three components. $l(\,)$, $c(\,)$ and $s(\,)$ are luminance, contrast and structure comparison operators, respectively. $l(\,)$ is the function of mean intensities μ_C and μ_D of images; $c(\,)$ is a function of image standard deviations σ_C and σ_D. For structure comparison $s(\,)$, mutual information of C and D is represented by σ_{CD} as

$$\sigma_{CD} = \frac{1}{L}\sum_{i=1}^{L}(C_i - \mu_C)(D_i - \mu_D) \tag{4}$$

where L is the pixel number in image C.

SSIM is good at extracting structural information, such as edge or gradient features, which conveys important visual information for perception and understanding. It exploits low level HVS property and does not require complex transform or multi-scale decomposition, hence it is the most popular IQA and is applicable in real time.

In practical applications, the weight factors for three components (i.e. α, β, γ), are set to 1, and SSIM is simplified as

$$\mathrm{SSIM}(C,D) = \frac{(2\mu_C\mu_D + C_1)(2\sigma_{CD} + C_2)}{(\mu_C^2 + \mu_D^2 + C_1)(\sigma_C^2 + \sigma_D^2 + C_2)} \tag{5}$$

where C_1 and C_2 are constants. On image scale, SSIM is revised to its mean (MSSIM) as

$$\mathrm{MSSIM}(C,D) = \frac{1}{M}\sum_{i=1}^{M}\mathrm{SSIM}(C_i,D_i)$$

where M is the number of blocks. SSIM defined in (5) includes luminance, contrast and structure features, but its calculation requires some parameters, such as μ_C, μ_D, σ_C, σ_D and σ_{CD}. In the rest of this paper, we will explore how to estimate these parameters.

3. Quantization noise estimate

3.1. Quantization noise model

Transform-based coding algorithms are widely used to compress images or image sequences. Each image is firstly divided into equal-size blocks. Subsequently, linear transform (e.g. DCT) is applied on these blocks. The compressed copy of C is therefore expressed as in [12]

$$D = T^{-1}Q(T(C)) \tag{6}$$

where $Q(\,)$ is the quantization procedure, and T and T^{-1} are the forward and inverse transform. The quantization operator Q introduces compression noise on the decoded images. The linear

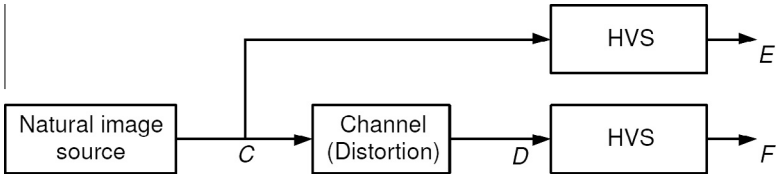

Fig. 1. A general IQA model [4].

relationship between the noise in transform and spatial domains is expressed as

$$n^{Spatial} = T^{-1} n^{DCT} \tag{7}$$

Because noise in spatial domain is the linear sum of the independent noise components in DCT domain, it can be assumed to be Gaussian distributed, according to the central limit theorem as

$$n^{Spatial} = N(0, K_Q) \tag{8}$$

where K_Q is the covariance matrix of quantization noise in spatial domain. The noise in spatial domain is related to the noise in transform domain with the IDCT operation as

$$K_Q = T^{-1} K_{Transform} (T^{-1})^T \tag{9}$$

$K_{Transform}$ is a diagonal matrix, whose diagonal elements are the expected noise power of each DCT coefficient. DCT is an orthogonal operator, which is used to completely de-correlate the pixels in spatial domain, and produce independent DCT coefficients. A common assumption is that an original DCT coefficient is independent uniform distributed within its quantization interval [12], if the quantization step size is small or the quantized DCT coefficient is large. Let G be the DCT of C and x_i be the value of the i-th element of G. Based on above assumption, the noise variance σ_i^2 of x_i is expressed as

$$\sigma_i^2 = q_i^2/12 \tag{10}$$

where q_i is the quantization step size for the coefficient.

Although uniform model simplifies the analysis, it is subjected to some limitations. Experimental results show that uniform noise model results in poor estimation of quantization noise. Prior knowledge about DCT coefficients of the natural images can be integrated into the noise model for better noise estimation. For example, a Laplacian model is widely used to describe DCT coefficient distribution [13,14]. The distribution of x_i is

$$p_{G(i)}(x) = \frac{\lambda_i}{2} \exp\{-\lambda_i |x_i|\} \tag{11}$$

When the quantized DCT coefficient x lies in the interval $[q_i^j, q_i^{j+1}]$, the variance estimated by the Laplacian model is computed in the closed form through integration

$$\sigma_G^2(i) = \kappa \int_{q_i^j}^{q_i^{j+1}} (\bar{x} - y) p_{G(i)}(y) dy \tag{12}$$

where κ is the normalization factor, j is the index of quantization interval of x_i, and $\bar{x}$ is the centroid of $[q_i^j, q_i^{j+1}]$. Experimental results indicate that the Laplacian model works well only if the quantized DCT coefficients are zero. On the contrary, the uniform model of nonzero coefficient approximates Laplacian tail better than the exponentially decaying tail. As a result, a noise model should be selected according to the DCT value for yielding more accurate noise estimation [27].

3.2. Prior knowledge about quantization noise

As stated in Section 3.1, the compressed noise in spatial domain follows Gaussian distribution. We consider the noise distribution of 8×8 block, which contains 64 independent elements. Its noise amplitude is the sum of these 64 noise components

$$n_{block}^{spatial} = N\left(0, \sum_{i=1}^{64} \sigma_i^{spatial}\right) \tag{13}$$

Quantization noise does not exist for specific DCT coefficients, such as zero-value DCT coefficients of the original image, hence noise estimate only considering the quantization scale [12] is not reasonable. This motivates us to propose a more accurate probabilistic model taking into account statistics of natural images.

We randomly select 13 raw images from LIVE database [17]. These images are compressed with quality score 40, and the compression distortion of their blocks is calculated. The distribution of their distortion is plotted in Fig. 2, which is well fitted with the Gaussian distribution with variance $\sigma_i = 0.028$. Therefore, the simple zero-mean Gaussian model is able to describe noise distribution precisely. This observation is very useful for analyzing the distortion in our noise estimation.

In [16], Pao et al. model the relationship of pixel values in DCT blocks as well as quantization noise in the DCT domain. They use the autocorrelation to roughly express the relation of DCT coefficients. Image content varies with their covariance differs.

The pixels in natural image are highly correlated, and their autocorrelation is computed as

$$r(p, q) = \sigma_f^2 \rho^{|p|} \rho^{|q|} \tag{14}$$

where p and q are the horizontal and vertical distances between two pixels, σ_f^2 is the variance of pixel values, and $|\rho| \leqslant 1$ is the correlation coefficients. After DCT operation, the variance of the (u, v) th DCT coefficients $\sigma_F^2(u, v)$ is written as

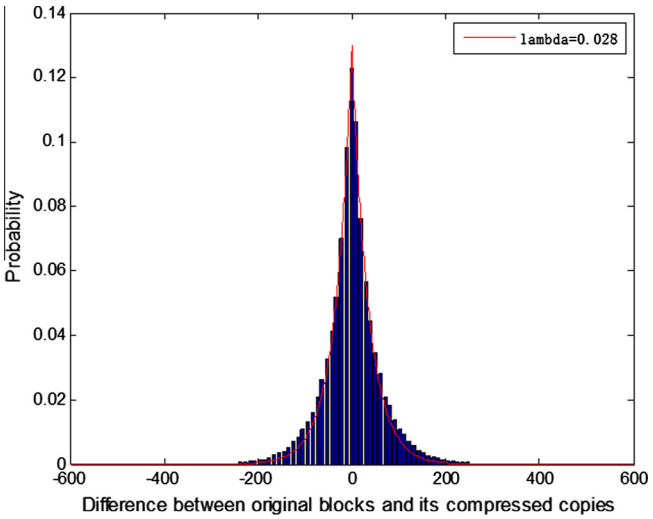

Fig. 2. Actual distribution of quantization noise of blocks vs Gaussian distribution.

$$\sigma_F^2(u,v) = \sigma_f^2 [ARA^T]_{u,u}[ARA^T]_{v,v} \tag{15}$$

where $R = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^7 \\ \rho & 1 & \rho & & \\ \rho^2 & \rho & 1 & & \\ \vdots & & & \ddots & \\ \rho^7 & & & & 1 \end{bmatrix}$, $[\,]_{u,u}$ is the (u,u) th component of the matrix, A is the basic vector of DCT.

If two blocks have similar quantization distortion distributions, i.e. with similar $\sigma_F^2(u,v)$, their σ_f^2 and ρ should be similar. As a special case, the similarity of two image blocks O_1 and O_2 can be defined in spatial domain as $SIM = \|O_1 - O_2\|^2$ or in DCT domain as $SIM = \|T(O_1) - T(O_2)\|^2$. Estimating the probabilistic model for an image from its inferior copy is an ill-pose problem and has an infinite number of solutions. Modeling the probability for each block is an over-sensitive problem and difficult to address. Moreover, the model precision is insufficient if we merely use one model to describe all types of blocks. According to the similarity defined above, we classify image blocks into several clusters, and assume that the blocks in the same cluster share similar mutual features. For parameter estimation, an original block is regarded as a random variance of the de-quantized block, which is a biased estimate from the cluster center with random variance following Gaussian distribution. If there are m blocks in cluster X, the distribution of its elements $X_{i\in[1,m]}$ is $N(b_i, \sigma_i)$, where b_i is the bias between i-th decoded block and the cluster center X_{center}, and σ_i is noise variance of i-th block. The probabilistic model of the cluster noise also follows Gaussian distribution $N(b_X, \sigma_X)$, where $b_X = \sum_{i=1}^m b_i$ and $\sigma_X^2 = \frac{1}{m}\sum_{i=1}^m \sigma_i^2$. It is obvious that b_X approaches zero, hence the cluster variance σ_X^2 can be used as the noise estimate for blocks in the cluster X.

There is a fundamental tradeoff between bias and variance in the parametric estimation problem [15]. We set Z as the cluster center to represent the general characteristics of similar compressed blocks. In this paper, the original image is an unknown and non-random variable in the form of a 2D vector C, and Z is the observed random variable. The vector C parameterizes the density $f(C; Z)$ of the observation Z. Let the estimator of C be $\widehat{C}$. Here, the "mismatch" between the mean of estimates and its true value is denoted as b_C, and the fluctuation of the statistical estimator is σ_C. They are defined as $b_C = E(\widehat{C}) - C$ and $\sigma_C = E[(\widehat{C} - C)^2]$, respectively. The mean square error (MSE) is widely used to measure the difference between the estimator and its ground-truth, and is related to b_C and σ_C with $\mathrm{MSE} = b_C^2 + \sigma_C^2$. In this application, the estimation of MSE is subjected to a tradeoff between b_C and σ_C. For instance, if only quantization noise on nonzero DCT coefficients is taken into consideration, the smooth estimation of the original image will reduce the variance of the image blocks at the expense of increasing bias (e.g. at lower spatial resolution). On the contrary, a biased shrinkage estimator reduces the variance of the ordinary squares estimator, while it enlarges the gap between the estimated variance and the true one.

General characteristics of blocks are used as prior knowledge for noise estimation. It is tightly associated with the clustered blocks, hence the number of clusters and clustering methods have some impacts on the accuracy of the prior knowledge. We adopt a standard method, i.e. fuzzy C means (FCM) clustering here. If a finite collection of m elements $X = \{X_1, \ldots, X_m\}$ is clustered into l fuzzy clusters with respect to the criterion

$$Cost = \sum_{i=1}^{l} \left(\sum_{k, X_k \in \tilde{X}_i} \|X_k - X_{i,center}\|^2 \right) \tag{16}$$

where $X_{i,center}$ is the cluster centroid of the i-th cluster $\tilde{X}_i$. FCM groups data points by populating some multidimensional space into a specific number of different clusters. With cluster number increasing, cluster diameter decreases and the data in the cluster is more compact. The blocks belonging to the same cluster will have small biases and MSE estimate errors. The computational cost of clustering is exponentially increasing with the number of clusters, and the accuracy of noise estimate increases slowly if the number of clusters is greater than a certain value. The influence of the number of clusters on MSE estimate distortion is beyond the scope of this paper, and will not be discussed here. Alternatively, the relation between the accuracy of IQA and the number of clusters is examined in Section 5.

4. NR SSIM

As quantization noise in spatial domain accords with Gaussian distribution, it is possible to simplify SSIM to evaluate the impact of compression distortion on HVS. Because $n^{Spatial}$ in (1) is zero-mean Gaussian variable, we get

$$\mu_C = \mu_D \tag{17}$$

where μ_C and μ_D are the means of original gray scale image and its compressed copy, respectively. The variance of the original image is calculated as

$$\sigma_C^2 = \sigma_D^2 + \sigma_n^2 - \frac{1}{N^2 - 1}\sum_{i=1}^{N^2} 2D_i \cdot n_i \tag{18}$$

where N is the block size in horizontal or vertical direction, i.e. $N = 8$ in JPEG. Because quantization noise is irrelative to the content of the quantized image, the third term $\frac{1}{N^2-1}\sum_{i=1}^{N^2} 2D_i \cdot n_i$ approaches zero. As a result, Eq. (5) is simplified as

$$\mathrm{SSIM}_{sim} = \frac{2\sigma_D^2 + C_2}{2\sigma_D^2 + \sigma_n^2 + C_2} \tag{19}$$

where σ_D^2 can be directly calculated from the decoded image. After simplification, the difficulty to estimate SSIM in NR condition lies in estimating σ_n^2. DCT is a bi-orthogonal transform where Parseval theorem holds. Therefore, we calculate σ_n^2 in DCT domain as it is straightforward. Experimental results show that the proposed quantization noise estimation on nonzero DCT coefficients is more reliable than that on zero DCT coefficients, and meet the requirement for most NR IQA applications. Therefore, we focus on the latter, i.e. noise variance estimation of zero-value DCT region.

If there are m blocks X_k in the cluster $\tilde{X}$, its cluster center is defined as

$$X_{center} = \frac{\sum_{k, X_k \in \tilde{X}} w_k X_k}{\sum_{k, X_k \in \tilde{X}} w_k} \tag{20}$$

w_k is the credibility that X_k belongs to $\tilde{X}$, which is related to the inverse of the distance between X_k and the cluster center of $\tilde{X}$. X_{center} is a vector with abundant frequency components, whose variance is denoted as σ_{center}^2. DCT transform, quantization and inverse quantization with the quantization scale of the tested image are applied on X_{center} to generate its compressed copy. This copy is then decomposed into two parts: the zero-value and the nonzero-value region. Uniform-distributed model is applied on nonzero-value area to estimate its noise variance $\sigma_{center\text{-}nonzero}^2$, and the noise variance estimate of the zero-value part will be $\sigma_{center\text{-}zero}^2 = \sigma_{center}^2 - \sigma_{center\text{-}nonzero}^2$, which serves as the general characteristic for noise estimation on the nonzero-value region with similar DCT distribution.

To estimate the compression noise of a compressed block, we firstly calculate the noise on nonzero DCT coefficients following a

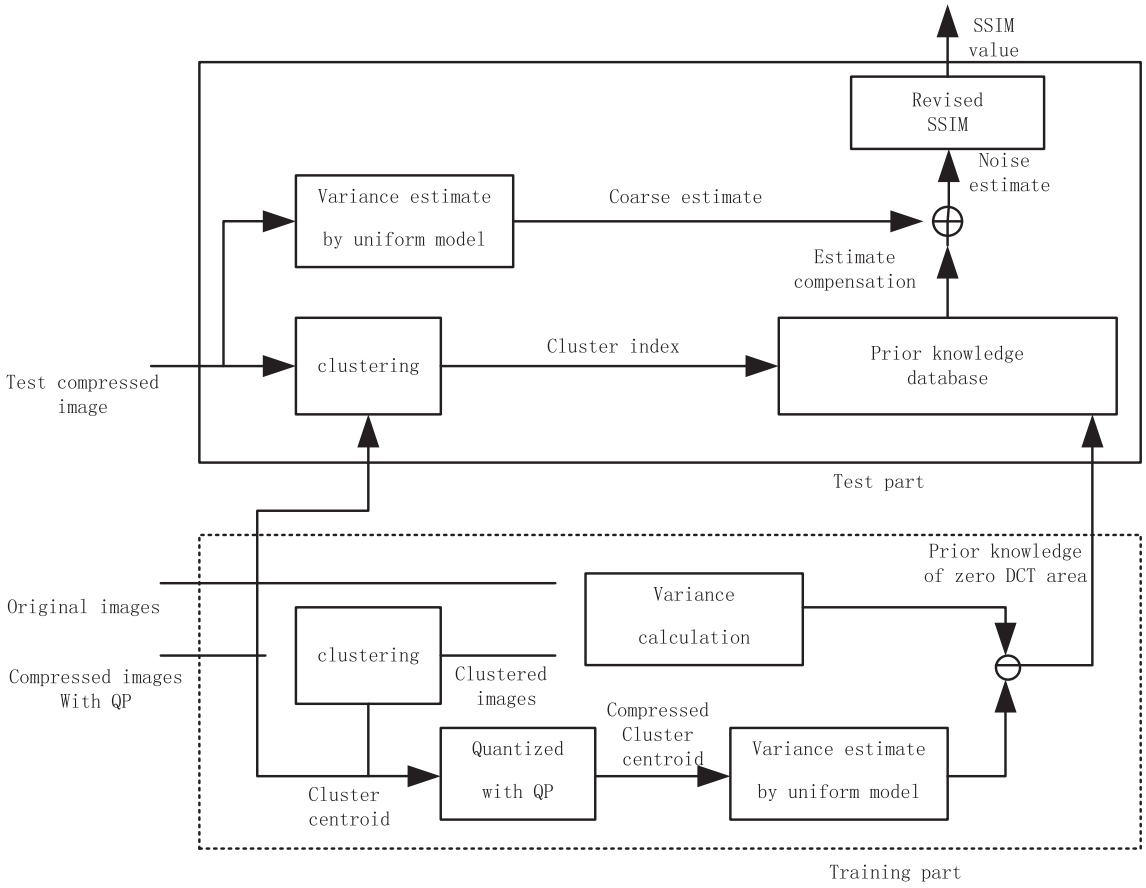

Fig. 3. Flow chart of the proposed IQA.

uniform probabilistic model, and then add this estimate to its $\sigma^2_{center\text{-}zero}$ to form the noise estimate σ^2_n in (19).

Fig. 3 illustrates the flow chart of the proposed IQA. In the training stage, some images are compressed with a QP to produce the degraded copies, and their PSNRs are calculated. These compressed copies are then classified into some clusters, and their centroids are regarded as representative image of the group and then requantized. Let $PSNR_{mean}$ be the mean of true PSNRs in the cluster. Based on the uniform distribution model, we estimate its quantization noise on nonzero DCT coefficients. The difference between $PSNR_{mean}$ and noise estimate is regarded as the noise estimate compensation, which corresponds to the quantization noise on the zero region of decoded DCT coefficients. As for testing, the decoded image is classified to the nearest group to find its estimate compensation, which is added on the rough noise estimate of nonzero area to produce the refined noise estimate σ^2_n in (19).

5. Simulation

We use LIVE [17] and TID2008 [18] databases to verify the performance of the proposed algorithm. In these two databases, images compressed with JPEG standard are used for testing, as each of them has an associated mean opinion score (MOS) and difference mean opinion score (DMOS) to represent its subjective quality.

5.1. Experimental implementation

The experiments are composed of three steps: training for quantization noise estimate; calculating NR SSIM and fitting it with subjective scores. To avoid the reported results being affected by known features, images in these three steps are completely not overlapped with each others. To demonstrate the robustness of the proposed algorithm, all experiments are repeated for 20 times. As shown in Table 1, the results of each experiment are highly correlated, which are located within narrow ranges as CC ∈ [0.969, 0.987], SROCC ∈ [0.93, 0.98], MAE ∈ [4.1, 5.1], RMSE ∈ [6.2, 7.5]. More comparisons are shown in Table 2.

To distill the common knowledge of quantization noise, we randomly select ten images from the databases, and compress them with different quality score (QS) = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100. The 8 × 8 blocks with the same QS are clustered, and the noise statistics of clusters are calculated. The statistics are used to calculate NR SSIM, which are fitted subsequently. After applying

Table 1
Performance comparison of image quality assessment models on LIVE database.

	CC	SROCC	MAE	RMSE
MSSIM	0.977	0.974	5.014	7.782
The proposed measure	0.980	0.960	4.839	7.685

Table 2
Performance comparison of image quality assessment models on TID2008 database.

	CC	SROCC	MAE	RMSE
MS_SSIM (FR)	0.916	0.899	0.507	0.685
VIF (FR)	0.955	0.917	0.333	0.507
The proposed measure (NR)	0.929	0.892	0.479	0.630
BLIINDS-II (NR)	0.873	0.835	0.533	0.772
DIIVINE (NR)	0.865	0.831	0.529	0.781
BRISQUE (NR)	0.894	0.868	0.521	0.764

logistic nonlinear operation, the NR SSIM value is fitted with subjective values, e.g. MOS or DMOS, to produce the fitting function. Finally, the NR SSIM of the tested image is fed into the fitting function to compute its NR IQA. For simplification, we use coarser QS in training to get the cluster centroids and the corresponding $\sigma^2_{center\text{-}zero}$s used for NR SSIM. If the QS of the tested image does not exactly match the QS in training, the parameters σ^2_{center} and $\sigma^2_{center\text{-}zero}$ of the training with the nearest QS will be used at the stage of NR SSIM.

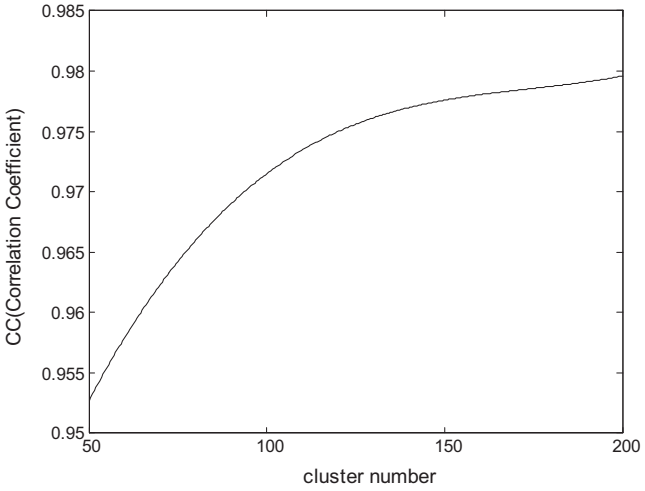

Fig. 4. Performance of the proposed algorithm vs. cluster number.

5.2. The effect of cluster number

The clustering performance affects the accuracy of MSE estimation, as well as objective quality assessment. Our IQA algorithm is independent with clustering method, thus we use a standard one (i.e. FCM) in this paper. We observe the influence of cluster number on the proposed algorithm in Fig. 4. The median linear correlation coefficient (CC) between the proposed NR IQA and the true DMOS is calculated to measure their correlation. A higher value of CC indicates a higher correlation with human opinion, and its maximum value is 1. This part of experiment is tested on LIVE database. It is shown that the performance of the proposed algorithm improves with cluster number increasing, but this improvement saturates when cluster number is larger than 100. Considering the computational cost, the number of clusters is set as 200 here to balance the performance improvement and the implementation speed.

5.3. Performance comparison

The proposed NR IQA is modified from SSIM, hence these two matrices are firstly compared in Table 1 to testify the efficacy of the proposed modification. This experiment is constructed on LIVE database. The correlation coefficient (CC), Spearman rank-order correlation coefficient (SROCC), Mean absolute prediction error (MAE) and Root mean square prediction error (RMSE) after nonlinear regression are used as criteria, where CC indicates the prediction accuracy and SROCC measures the prediction monotonicity. It is shown that the proposed algorithm achieves comparable performance to the MSSIM, even in the condition that original images are unavailable.

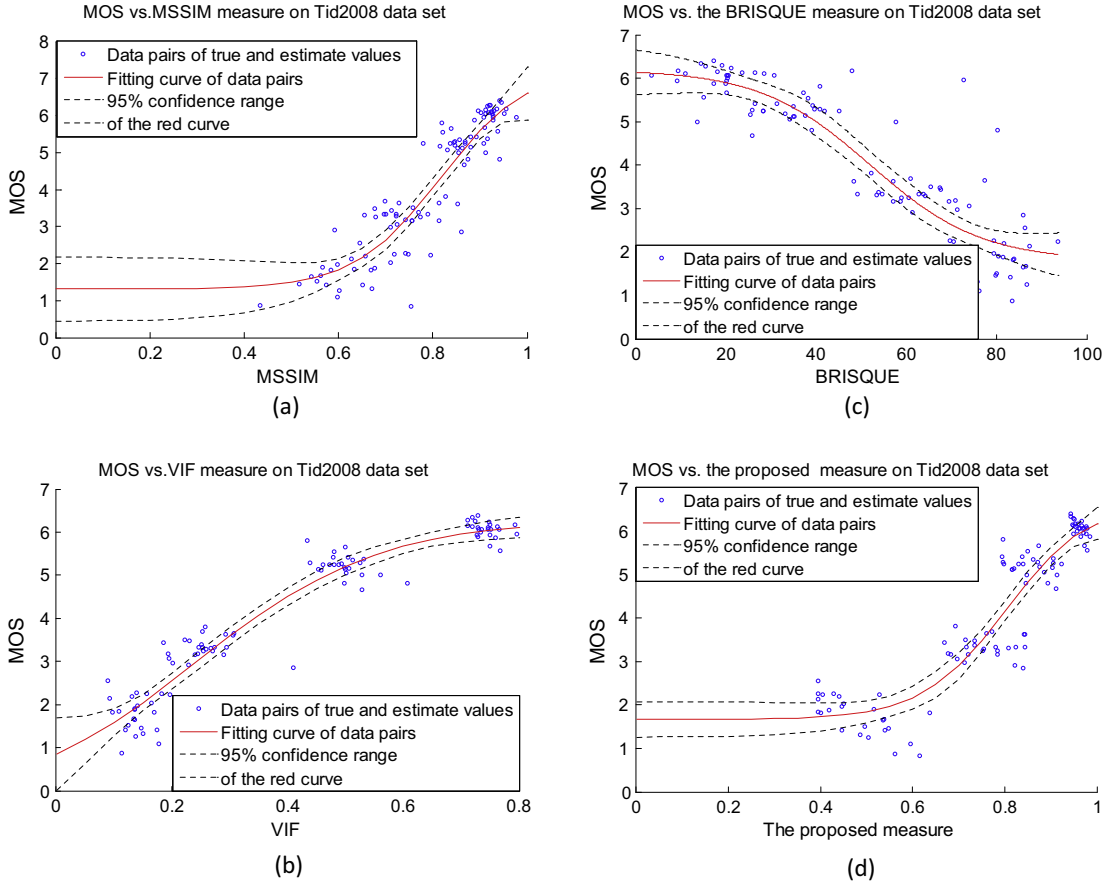

Fig. 5. Scatter plots of subject scores versus the different metric scores. (a) MS_SSIM; (b) VIF; (c) BRISQUE; and (d) proposed metric.

We further compare the performance of the proposed algorithm with some IQA algorithms, including two FR and three NR metrics. The compared FR IQA indices include multi-scale structural similarity index (MS_SSIM) [19] and visual information fidelity (VIF) [4], which are the most popular indices due to their good performance and simplicity. The referred NR indexes are BLIINDS-II [22], DIIVINE [23], dubbed blind/referenceless image spatial quality evaluator (BRISQUE) [11]. BLIINDS-II, DIIVINE, BRISQUE are compared because they are the up to date ones and their advantages over other benchmarks have been proven [11].

To verify the robustness of IQA algorithms, we use different data in training and testing, i.e. the prior knowledge is distilled from LIVE database and IQA is done on images in TID2008 database. IQA results are tabulated in Table 2. Among all IQA algorithms, MS_SSIM and VIF are FR algorithms, and do not depend on image statistical information. NR IQA is more difficult than FR ones, because they do not have original images as reference. The performance of the proposed algorithm is comparable to that of MS_SSIM to prove its robustness. Among all the NR algorithms, the proposed algorithm is the best one. It performs constantly better than BRISQUE on the aspects of CC, SROCC, MAE and RMSE, and these improvements are over 5% in most cases. BRISQUE is better than BLIINDS-II and DIIVINE, and its CC and SROCC reduce about 3%, because BRISQUE features are relatively high and is easily perturbed by image content.

The scatter plot of MOS versus metric prediction by different methods is shown in Fig. 5. The data is indicated as blue circle; the logistic function for data fitting is given by the red curve; 95% confidence range of the red curve is plotted by the blue lines of dashes. It shows that VIF performs the best and the proposed method is slightly better than MS_SSIM in its narrower range of double standard deviations. BRISQUE does not produce satisfactory results because some of its scatters are located far away from the red fitting curve, as shown in Fig. 5(c). In brief, Fig. 5 gives a similar conclusion as Table 2.

5.4. Computational Cost

The calculation of MS_SSIM and VIF is fast, as classification is not required. BRISQUE only uses small number of features (18 per scale and 2 scales) to represent an image, hence it is also faster than the proposed algorithm. Most of the computational cost of the proposed algorithm lies in block classification. For classification, differences of the tested block and the cluster center are examined, which requires 64 additions for each comparison. Therefore, its computational cost will be considerable if the tested image contains many blocks.

The complexity of various IQA metrics is compared, in terms of the processing time (in seconds) on TID2008. The IQA algorithms are implemented by MATLAB on a single-core CPU at 2.2 GHz with 2 GB RAM. It is shown that the proposed algorithm takes the longest processing time, i.e. about 9 seconds for an image. However, the processing time can be significantly reduced by three ways: (1) setting a small cluster number; (2) shortening the length of description vector for cluster center, because most high frequency DCT coefficients are quantized to zeros; and (3) using parallel computation for multiple comparisons within a cycle, e.g., a single Instruction and multiple data (SIMD) for 32 bits addition by an Instruction. It is expected that the implementation speed of the proposed algorithm will accelerate 10 times by these strategies.

6. Conclusion

In this paper, we firstly overview SSIM and the probabilistic model of quantization noise, and then propose a learning-based algorithm to accurately estimate the noise amplitude, which is used for the modified SSIM. The experimental results validate the efficiency of the proposed metrics over state-of-the-art NR IQA algorithms, such as BRISQUE. The proposed algorithm can be used to monitor video objective quality or to control the re-quantization in trans-coding.

Acknowledgments

This work is supported by National Nature Science Foundation of China, Nos. 60902072, 61302121, 61201446 as well as the Opening Project of Shanghai Key Laboratory of Digital Media Processing and Transmission.

References

[1] A.C. Bovik, Handbook of Image and Video Processing, Academic, New York, 2005.
[2] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, A.C. Bovik, Image quality assessment based on a degradation model, IEEE Trans. Image Process. 9 (4) (2000) 636–650.
[3] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[4] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[5] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, On between-coefficient contrast masking of DCT basis functions, in: Proc. 3rd Int. Workshop Video Process. Quality Metrics Consumer Electron., January 2007, pp. 1–10.
[6] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.
[7] E.C. Larson, D.M. Chandler, Most apparent distortion: full reference image quality assessment and the role of strategy, J. Electron. Imag. 19 (1) (2010) 011006-1–011006-21.
[8] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[9] M. Narwaria, W.S. Lin, I.V. McLoughlin, S. Emmanuel, L.T. Chia, Fourier transform-based scalable image quality measure, IEEE Trans. Image Process. 21 (8) (2012) 3364–3377.
[10] J.J. Wu, W.S. Lin, G.M. Shi, A.M. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 22 (1) (2013) 43–54.
[11] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[12] C.A. Segall, A.K. Katsaggelos, R. Molina, J. Mateos, Bayesian resolution enhancement of compressed video, IEEE Trans. Image Process. 13 (7) (2004) 898–911.
[13] R.C. Reininger, J.D. Gibson, Distributions of the two-dimensional DCT coefficients for images, IEEE Trans. Commun. COM-31 (6) (1983) 835–839.
[14] E.Y. Lam, J.W. Goodman, A mathematical analysis of the DCT coefficient distributions for images, IEEE Trans. Image Process. 9 (10) (2000) 1661–1666.
[15] A.O. Hero, J.A. Fessler, M. Usman, Exploring estimator bias-variance tradeoffs using the uniform CR bound, IEEE Trans. Signal Process. 44 (8) (1996) 2026–2041.
[16] I.-M. Pao, M.-T. Sun, Approximation of calculations for forward discrete cosine transform, IEEE Trans. Circ. Syst. Video Technol. 8 (1998) 264–268.
[17] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE Image Quality Assessment Database Release 2, 2005. <http://live.ece.utexas.edu/research/quality>.
[18] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008—a database for evaluation of full-reference visual quality assessment metrics, Adv. Modern Radioelectron. 10 (10) (2009) 30–45.
[19] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: Proc. Asilomar Conf. Signals, Syst. Comput., vol. 2, 2003, pp. 1398–1402.
[22] M. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[23] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.
[24] T.N. Pappas, R.J. Safranek, Perceptual criteria for image quality evaluation, in: A. Bovik (Ed.), Handbook of Image & Video Proc, Academic Press, 2000.
[25] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, IEEE Trans. Image Process. 21 (4) (2012) 1500–1512.
[26] J. Wu, W. Lin, G. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimedia 15 (7) (2013) 1700–1705.
[27] M.A. Robertson, R.L. Stevenson, DCT quantization noise in compressed images, IEEE Trans. Circ. Syst. Video Technol. 15 (1) (2005) 27–38.
[28] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (8) (2012) 3378–3389.