Practical Image Quality Metric Applied to Image Coding

Fan Zhang, *Member, IEEE*, Lin Ma, *Student Member, IEEE*, Songnan Li, *Student Member, IEEE*, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—Perceptual image coding requires an effective image quality metric, yet most of the existing metrics are complex and can hardly guide the compression effectively. This paper proposes a practical full-reference metric with consideration of the texture masking effect and contrast sensitivity function. The metric is capable of evaluating typical image impairments in real-world applications and can achieve the comparable performance as the state-of-the-art metrics on the publicly available subjectively-rated image databases. Due to its simplicity, the metric is embedded into JPEG image coding to ensure a better perceptual rate-distortion performance.**

***Index Terms*—Image quality metric, perceptual coding.**

I. Introduction

IMAGE quality assessment is fundamental to the performance optimization of imaging systems including the capture, display, storage, and transmission of images. Quality assessment is closely related to human perception. A quality metric is considered accurate if its results are consistent with the subjective evaluation. To evaluate the metric performance, subjectively-rated databases are used as "ground truths". Generally, the impairments, such as blocking artifacts, blur, additive noise, etc., are frequently involved in real-world applications. The databases provide the reference images, the distorted images, and the subjectively-rated scores of the distorted images. As far as we know, at least eight subjectively-rated image databases for mainly coding applications are available on the internet, including LIVE [1], IRCCyN/IVC [2], Toyama [3], TID2008 [4], A57 [5], CSIQ [6], WIQ [7], and LAR [8].

In most cases, the quality assessment is not the ultimate goal. As worthwhile directions, many image processing are posed as optimization problem, where an image quality measurement is deployed in the objective function. We pay attention to the optimal solution which achieves the highest image quality with the least cost for a particular application. Such attempts have already been done in a number of fields, including image halftoning [9], image segmentation and classification [10], and image watermarking [11]. Such attempts also require that the quality metric can be optimized easily. The statement of "a metric is easy to optimize" means that when being embedded into an optimization problem, the metric does not bring significant burden to find the optimal solution. For example, a popular metric is convex with respect to the distorted image and has a simple gradient and Hessian matrix, such that the problem is likely to have a closed-form analytic solution or at least can be approached by a fast gradient-descent method.

The mean square error (MSE) has been the primary choice for image quality assessment for many years, partly because the MSE is easy to optimize. MSE predicts quality of white-noise distorted image well, but fails to cope with other distortion types and cross-artifacts measurement. Seeking the substitutes for the MSE, the researchers have paid a lot of attention to the properties of the human visual system (HVS) [12]. Quite a few of the perceptual properties were exploited. Existing studies show that our perception on image differences exhibits complexity, since: 1) the difference below a certain threshold is imperceptible [13]; 2) neither the visual pathway can be divided into *absolutely* independent sub-channels, nor natural images can be transformed into *real* independent subspaces. Therefore, the distortions being decomposed into the frequency sub-bands (e.g., discrete cosine transform coefficients) still have high-order correlations and interaction with each other [14], [15]; 3) the high-order Minkowski norm is regarded as a better pooling strategy than the Euclidean norm [16]; 4) a lot of factors need to be determined so as to accurately simulate the cases of diverse display media, varying illumination conditions, and different viewing distances. Due to these reasons, the HVS-based metrics are often complicated and difficult to optimize, and cannot guide the application effectively.

This study deals with the quality assessment by a tradeoff between the practicability and the accuracy. Our comparative analysis indicates that two HVS properties, the texture masking effect and the contrast sensitivity function (CSF), play an important role in image quality assessment. Taking into account these two properties, we propose a simple yet effective full-reference image quality metric with the advantages of: 1) the metric has comparable performance compared to the state-of-the-art metrics; 2) with the explicit distortion formulation, perceptual coding is posed as a problem of perceptual rate-distortion optimization block by block and subband by subband; 3) the metric is easy to optimize and can be integrated into the JPEG standard without difficulty.

The rest of this paper is organized as follows. After a brief description of the HVS properties in visual perception and the typical metrics in Section II, the proposed metric is introduced in

Manuscript received August 23, 2010; revised December 20, 2010 and March 14, 2011; accepted March 14, 2011. Date of publication April 05, 2011; date of current version July 20, 2011. This work was supported in part by a grant from the Chinese University of Hong Kong under the Focused Investment Scheme (Project 1903003). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ton Kalker.

The authors are with the Department of Electronic engineering, Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: fan.zhang@technicolor.com; lma@ee.cuhk.edu.hk; snli@ee.cuhk.edu.hk; knngan@ee.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2134079

Section III. The metric performance is compared in Section IV. The corresponding perceptual coding scheme is proposed in Section V, followed by the conclusion.

II. Related Works

Image quality metrics can be classified from different viewpoints into: metrics designed by psychophysical approach and by engineering approach [17], error sensitivity and structural sensitivity metrics [18], and bottom-up and top-down metrics [16].

Among the existing metrics, many HVS properties are taken into account or implicitly considered. One of them is the texture masking effect. Masking effect refers to the reduction of the visibility of image distortion due to the presence of the original content of the reference image [16]. In general, the distortion is less visible where the reference image region presents rough texture. Therefore, texture masking is also termed as contrast masking [16] or noise visibility function (NVF) [19] in the watermarking community. Strictly speaking, texture masking can be divided into intra-band masking and inter-band masking [20]. Intra-band masking refers to the error tolerance due to the original content in the subband itself, while inter-band masking involves multiple subbands and is affected by their difference in phase, orientation, spatial frequency, and intensity. Since it requires setting a host of parameters to precisely model the texture masking, we turn to an approximate model based on local variances in this study.

Human sensitivity to sine-wave grating in luminance has a classic inverted-U shape, namely contrast sensitivity function (CSF). The psychophysical experiments that underlie many CSF models are to estimate the threshold at which a stimulus is just visible. At the suprathreshold level, traditional findings report that apparent contrast of sine-wave grating is relatively independent of spatial frequency, i.e., the phenomenon termed "contrast constancy" [21]. A recent study yet reported that suprathreshold CSF is still of an inverted-U shape as a function of frequency [22]. Existing HVS-based metrics [23], [24] often regard the CSF at threshold level as human sensitivity to image difference and use them to weigh the subbands' distortions, which sometimes are of suprathreshold. With this gap in mind, this paper refines the shape of suprathreshold CSF according to subjective databases. With the modified CSF, the quality measure keeps more consistent with subjective evaluations as explained in Section IV-A.

Many other HVS properties have also been exploited, such as light adaption (luminance masking) [23], perceptual color space [25], visual attention (region of interest) [26], etc. Each individual effect is supported by visual psychological experiments. However, whether or not their influences on visual quality measure are significant remains debated.

The above-mentioned HVS properties can be modeled by only the information of the reference image. Recently, the features of both the reference and the distorted images are employed for image quality measure to improve the metric accuracy, including the singular vector of images [27], the statistical moments on the Log-Gabor filter responses of images [28], and mutual information [29]. However, these metrics are difficult to optimize since they are nonlinearly dependent of the distorted image (i.e., the processed image).

In this study, five typical metrics are compared whose codes are publicly available: 1) MSE and DCTune [23], [30], which are commonly used for perceptual coding and watermarking; 2) MSSIM (the multi-scale version of SSIM [18]) and VIF [31], which exhibit an outstanding performance in previous study [32]; 3) PSNR-HVS-M [33] and VSNR [24], which are designed based on the databases [4] and [5], respectively, and performs quite well on them.

Among all the metrics above, PSNR-HVS-M and DCTune are quite similar to the proposed metric, but their models of the CSF and the texture masking are different from the proposed one. More than a quality metric, DCTune was originally developed for optimizing JPEG image compression. We will also compare it with our perceptual coding method.

III. Proposed DCTex Metric

The proposed metric is based on a key assumption that the signal error in each subband and each local region contributes to the entire distortion independently. Although over-simplistic, this assumption is still reasonable since most typical distortions have few (linear) correlation both between the subbands and between the neighborhoods at large spatial scales. On the other hand, the subband error and the local error contribute to the entire distortion unequally due to the CSF and the texture masking effect, respectively. Therefore, we propose a quality metric named as DCTex with the following definition:

$$D(X,Y) = g\sum_{i=1}^{n}\sum_{j=1}^{64}\frac{c_j\cdot(u_{i,j}-v_{i,j})^2}{l_i}. \tag{1}$$

In this paper, X and Y represent the reference and the distorted images, respectively. Both images are divided into 8×8 non-overlapping blocks, denoted by x and y, respectively. The images are supposed to have a total number of n blocks, and i indexes the variables associated with each block $(i=1,2,\ldots,n)$. $u_{i,j}$ and $v_{i,j}$ is the jth DCT coefficient of the block x_i and y_i, respectively. j indexes the variables associated with each DCT subband $(j=1,2,\ldots,64)$. The scores provided by the metric are further normalized by the number of image pixels, since the image resolutions may not be identical.

The DCTex metric combines the texture masking and the CSF. g and l_i are two types of texture masking parameters. g, named as global smoothness, is quantified by

$$g = \frac{\mathrm{var}(\overline{x})}{\mathrm{var}(X)} \tag{2}$$

where the denominator is the luminance variance of the whole image and the numerator is the variance of the block luminance means:

$$\mathrm{var}(\overline{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(\overline{x}_i - \sum_{i=1}^{n}\frac{\overline{x}_i}{n}\right)^2. \tag{3}$$

In (3), $\overline{x}_i$ is the block mean of block i. Global smoothness affects the overall distortion visibility of an image. Let us demonstrate the case of 1-D signal in Fig. 1. The left figure in Fig. 1 is a descending staircase signal with two smooth steps and the right

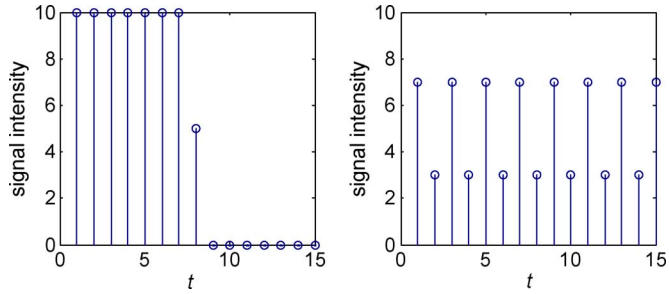


Fig. 1. Global smoothness of 1-D signals. Left: 1; right: 0.05. Window width = 5.

1.00	0.405	0.162	0.0647	0.0256	0.0101	0.0040	0.0016
0.405	0.278	0.131	0.0557	0.0229	0.0092	0.0037	0.0014
0.162	0.131	0.0758	0.0370	0.0165	0.0071	0.0029	0.0012
0.0647	0.0557	0.0370	0.0205	0.0101	0.0046	0.0020	0.0009
0.0256	0.0229	0.0165	0.0101	0.0055	0.0027	0.0013	0.0006
0.0101	0.0092	0.0071	0.0046	0.0027	0.0014	0.0007	0.0003
0.040	0.0037	0.0029	0.0020	0.0013	0.0007	0.0004	0.0002
0.016	0.0014	0.0012	0.0009	0.0006	0.0003	0.0002	0.0001

Fig. 2. Proposed mCSF, $\{c_j\}$, associated with 8 × 8 DCT coefficients.

figure is a sawtooth signal. The left signal looks smoother in a local scale than the right signal, despite its larger variance in overall. This is confirmed by the global smoothness with respect to a window width of 5: the left signal has a global smoothness of 1 and the right signal of 0.05. Similarly, images full of textures have the lower global smoothness and thus can tolerate more distortions than smooth images.

The other texture masking parameter, l_i, is called local roughness, since it is calculated block by block as

$$l_i = \sqrt{\mathrm{var}(x_i)} + \varepsilon \quad (i = 1, 2, \cdots, n) \tag{4}$$

where $\mathrm{var}(x_i)$ is the luminance variance in block i. Parameter ε in (4) controls the relative deviation of the local roughness. The larger the parameter, the smaller is the relative deviation of the local roughness and thus the weaker texture masking effect is simulated. With ε, dividing by zero in (1) is also avoided. The metric performance is not sensitive to this parameter and we set ε as 20 empirically for 8-bit images. More details about the global smoothness and the local roughness can be found in our previous work [34].

The DCTex metric employs a modified CSF (mCSF) associated with the DCT. Although CSF can be modeled for many spatial-frequency transforms, (e.g., wavelet in [24] and [35]), it is reported that the types of the spatial-frequency transforms are not of paramount importance [25], [36]. This paper selects the DCT domain so as to adapt to the wide DCT-based image applications. The mCSF, $\{c_j\}$ $(j = 1, 2, \ldots, 64)$, which is normalized by c_1, is shown in Fig. 2. How we obtain the mCSF will be explained in Section IV-C. The high-frequency components of mCSF are tiny in comparison with the CSF at the threshold level [37], [38]. As a result, the distortions in high-frequency subbands have very low weights and can be ignored.

The modified mCSF has a steeper slope with respect to the frequency than the traditional CSF at the just noticeable level (e.g., Watson's CSF). Our CSF leads to the performance improvement as demonstrated in Section IV. This may be attributed to the fact that Watson's CSF indicates the human sensitivity to distortion at the threshold level, while our mCSF is able to capture the human sensitivity to distortion at both the threshold and supra-threshold level. Another reason is that: in most of the typical distortions, the high-frequency distortions are always bounded by their intrinsic spectra (e.g., often conforming to the $1/f$ power law), so it will not affect the performance even when they are ignored.

IV. Metric Performance

A. Performance Comparison

It is important to make sure that the proposed metric works well before applying it. This section is dedicated to comparing the proposed metric with the state-of-the-art metrics on diverse distortions. All kinds of distortions in existing subjective databases are taken into account for a comprehensive justification. The performances of the metrics are evaluated by the correlation between the two groups of scores which indicate the quality of distorted images in the subjective databases. One group is the objective scores predicted by the metrics and the other is the subjective scores rated by the subjects. Correlation of Spearman rank order correlation coefficient (SROCC) is used to assess *prediction monotonicity* [39]:

$$\mathrm{SROCC} = 1 - \frac{6\sum_{m=1}^{M} d_m^2}{M(M^2-1)} \tag{5}$$

where d_m is the difference between the mth image's rank in subjective and its objective score. SROCC is independent of a monotonic regression between the subjective and the objective scores. A higher SROCC score indicates a better performance for a metric. The best SROCC of 1 will occur when the objective scores is the perfect monotonic function of the subjective ones.

From the view point of practicality, we compare MSE, DCTune, MSSIM, PSNR-HVS-M, VSNR, VIF, and the proposed metrics on the distortion subsets below. 1) Table I, JPEG and JPEG 2000. These two distortion types are covered by all the databases except WIQ, and represent the most typical coding artifacts. 2) Table II, additive Gaussian noise and Gaussian blur. They also commonly occur and are contained in most databases except Toyama, IVC, and WIQ. 3) Table III, transmission error. It is simulated by LIVE, TID, and WIQ. Note that two separate sessions were conducted in WIQ, and thus, we use them as two different test sets. 4) Table IV, combinations of all the distortions except the three types of "difficult" distortions. 5) Table V, the "difficult" distortions including "local block-wise distortions of different intensity" and "mean shift" in TID, as well as "contrast change" in TID and CSIQ. We isolate the three "difficult" distortions because not only our metric but also PSNR-HVS-M, VSNR, and DCTune perform poorly on them, and even worse than MSE does. All above metrics only consider the luminance information of images. For each dataset, we highlight the best two results with boldface. We have two major observations based on the results shown in Tables I–IV.

First, the proposed Gray DCTex, as a weighted MSE, converts the MSE into a quite competitive metric for the most typical

TABLE I
METRIC SROCC ON DISTORTION SUBSET OF JPEG AND JPEG2000 COMPRESSION

	Number of images	MSE	MSSIM	VIF	PSNR-HVS-M	VSNR	DCTune	Gray DCTex	M1	M2	M3	Color DCTex
LIVE	344	0.891	**0.970**	0.967	0.954	0.590	0.815	**0.975**	0.924	0.942	0.963	0.960
IVC	100	0.768	**0.910**	0.903	0.900	0.823	0.740	**0.950**	0.794	0.843	0.917	**0.955**
Toyama (full set)	168	0.613	0.886	**0.908**	0.848	0.799	0.711	**0.899**	0.661	0.711	0.832	**0.913**
TID	200	0.866	0.957	0.957	**0.971**	0.931	0.838	**0.970**	0.926	0.953	**0.976**	0.962
A57	18	0.810	**0.926**	0.882	0.897	**0.951**	0.781	0.905	0.742	0.752	0.901	-
CSIQ	300	0.916	**0.963**	0.959	0.956	0.929	0.893	**0.967**	0.940	0.949	0.965	**0.970**
LAR	80	0.806	0.904	**0.931**	0.916	0.870	0.849	**0.944**	0.852	0.868	0.942	**0.951**

TABLE II
METRIC SROCC ON DISTORTION SUBSET OF ADDITIVE GAUSSIAN WHITE NOISE AND GAUSSIAN BLUR

	Number of images	MSE	MSSIM	VIF	PSNR-HVS-M	VSNR	DCTune	Gray DCTex	M1	M2	M3	Color DCTex
LIVE	290	0.915	0.960	**0.977**	0.953	0.733	0.926	**0.968**	0.933	0.942	0.955	0.963
TID	200	0.903	0.799	**0.922**	**0.922**	0.888	0.883	0.891	0.863	0.857	**0.928**	**0.892**
A57	18	0.572	0.661	0.854	0.820	**0.905**	0.636	0.733	0.564	0.558	0.826	-
CSIQ	300	0.941	0.946	**0.969**	**0.962**	0.922	0.893	0.954	0.953	0.951	0.961	0.954

TABLE III
METRIC SROCC ON TRANSMISSION ERROR

	Number of images	MSE	MSSIM	VIF	PSNR-HVS-M	VSNR	DCTune	Gray DCTex	M1	M2	M3	Color DCTex
LIVE (fast fading)	145	0.893	**0.932**	**0.965**	0.914	0.681	0.824	0.910	0.911	0.917	0.894	0.898
TID 12 13	200	0.741	**0.876**	**0.888**	0.822	0.769	0.744	0.847	0.831	0.841	0.840	**0.850**
WIQ t1	40	0.632	**0.754**	**0.760**	0.739	0.612	0.499	0.673	0.658	0.693	**0.681**	-
WIQ t2	40	0.644	0.750	0.623	**0.791**	0.461	0.723	**0.791**	0.656	0.653	0.773	-

TABLE IV
METRIC SROCC ON DATASET EXCLUDING "DIFFICULT" DISTORTIONS

	Number of images	MSE	MSSIM	VIF	PSNR-HVS-M	VSNR	DCTune	Gray-DCTex	M1	M2	M3	Color DCTex
LIVE (full set)	779	0.856	0.945	**0.964**	0.930	0.648	0.872	**0.951**	0.900	0.913	0.937	0.940
IVC (full set)	185	0.679	0.885	**0.896**	0.883	0.799	0.715	**0.921**	0.722	0.774	0.904	**0.931**
TID 1~14	1400	0.740	**0.866**	**0.864**	0.830	0.793	0.664	0.857	0.791	0.787	0.850	**0.861**
A57 (Full set)	54	0.570	0.839	0.622	**0.896**	**0.935**	0.773	0.851	0.630	0.672	**0.896**	-
CSIQ 1~5	750	0.906	**0.950**	0.928	**0.948**	0.930	0.870	0.941	0.923	0.927	**0.946**	0.931
LAR (Full set)	120	0.819	0.886	**0.916**	0.907	0.863	0.834	**0.921**	0.860	0.874	0.921	**0.923**

distortions. Actually, the performance of Gray DCTex is often comparable to many state-of-the-art algorithms in Column 3 $\sim$ 9 of Tables I–IV.

Second, Gray DCTex fails to predict the difficult distortions as shown in Table V. The distortions of "mean shift" and "contrast change" do not change the structural information of images, so they are not that annoying. The metrics that measure the signal difference between the reference and the distorted image often cannot tolerate such distortions VIF, which evaluates the statistical difference of image variance, can tackle "contrast change". SSIM, which evaluates the similarity of image mean and image variance separately, can handle both "mean shift" and "contrast change". In fact, "mean shift" causes few distortions in AC subbands and "contrast change" changes every AC subband at the same rate, so a host of weights to AC subband distortions (e.g., CSF) will not influence the final image quality and thus will not improve the metric performance. Another difficult distortion, "Local block-wise distortions of different intensity", overlays smooth patches on the images [4]. It seems that such artifacts do not affect the perceptual quality of smooth images but the images full of texture. This effect conflicts with what we assume about the global smoothness—textural images are more likely to hide distortions than the smooth images. It is also reported that "Local block-wise distortions of different intensity" is difficult to handle due to its "nonuniform distortion" [28]. This distortion may challenge a smarter pooling strategy in line with cognitive theory. Luckily, those "difficult" distortions seldom happen in the applications such as image coding. Therefore, DCTex is still practical and reliable for the most typical distortions.

B. Factor Analysis

Gray DCTex takes account of the CSF and the texture masking effect. To identify the contribution of each effect model to the metric performance, we compare three candidate metrics in Tables I–V (Column 9 $\sim$ 12):

- M1) with the local roughness

$$D(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{64}\frac{(u_{i,j}-v_{i,j})^2}{l_i} \tag{6}$$

TABLE V
METRIC SROCC ON "DIFFICULT" DISTORTIONS

	Number of images	MSE	MSSIM	VIF	PSNR-HVS-M	VSNR	DCTune	Gray-DCTex	M1	M2	M3	Color DCTex
TID 15 block error	100	0.619	**0.761**	**0.832**	0.624	0.195	0.231	0.206	0.507	0.388	0.436	0.158
TID 16 mean shift	100	**0.696**	**0.736**	0.446	0.693	0.335	0.629	0.612	0.637	0.608	0.694	0.612
TID 17 contrast	100	0.586	**0.641**	**0.819**	0.582	0.425	0.443	0.469	0.542	0.497	0.555	**0.489**
CSIQ 6 contrast	116	0.862	**0.952**	**0.935**	0.880	0.872	0.569	0.801	0.827	0.811	0.839	**0.818**

- M2) with the local roughness and the global smoothness

$$D(X,Y) = g\sum_{i=1}^{n}\sum_{j=1}^{64}\frac{(u_{i,j}-v_{i,j})^2}{l_i} \tag{7}$$

- M3) with the modified CSF

$$D(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{64}\left[c_j(u_{i,j}-v_{i,j})^2\right]. \tag{8}$$

In Tables I–IV, each of the local roughness, the global smoothness and the modified CSF assists in improving the metric performance. This can be seen consistently by comparing the performance {MSE, M1}, {M1, M2}, and {MSE, M3}. By comparing {M2, M3, DCTex}, we find that combining the texture masking and the CSF is helpful in most cases. The counterexamples occur for the difficult distortions in Table V: neither individual factor nor the combination improves the metric performance. As analyzed in Section IV-A, the texture masking effect and the CSF cannot help in measuring such difficult distortions.

C. CSF Determination

In this section, we explain how to determine the mCSF ($\{c_j\}$). Human sensitivity to visual stimuli shows a band-pass property in the spatial frequency domain. Nill [40] and Ngan [41] showed that the generalized HVS models can be expressed by

$$\mathrm{CSF}(\omega) = (a_1 + a_2\omega)\exp(-a_3\omega) \tag{9}$$

where a_1, a_2, and a_3 are constants; ω is the spatial frequency (cycle/visual degree). For the (jr, jc)th subband in an 8×8 DCT block ($jr, jc = 0, 1, \ldots, 7$), the corresponding frequency $\omega_{jr,jc}$ can be calculated by

$$\omega_{jv,jh} = \frac{f_{jr,jc}}{32\cdot\arctan\left[1/(2\times R_{vd}H)\right]} \tag{10}$$

$$f_{jr,jc} = \sqrt{jr^2+jc^2} \tag{11}$$

where the pixel aspect ratio (PAR) of the monitors is assumed to be 1, R_{vd} stands for the ratio of viewing distance to picture height, and H is the number of pixels in picture height.

Due to the gap between the threshold and the suprathreshold effect, parameters a_1, a_2, and a_3 need be carefully selected. We make selection by optimizing M3's performance on subjective databases. Here, M3 defined in (8) uses the normalized CSF and can be rewritten as (12)–(15):

$$D(X,Y) = \sum_{i=1}^{n}\sum_{jr=0}^{7}\sum_{jc=0}^{7}\left[c_{jr,jc}(u_{i,jr,jc}-v_{i,jr,jc})^2\right] \tag{12}$$

$$c_{jr,jc} = (a_4 + f_{jr,jc})\exp(-a_5 f_{jr,jc}) \tag{13}$$

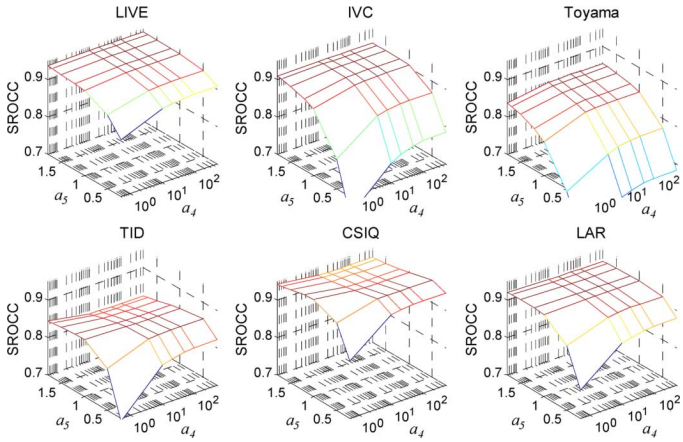


Fig. 3. Metrics performance on subjective databases as a function with respect to parameter a_4 and a_5.

where

$$a_4 = 32\times\arctan\left[1/(2\times R_{vd}H)\right]\times a_1/a_2 \tag{14}$$

$$a_5 = a_3/\left\{32\times\arctan\left[1/2\times R_{vd}H)\right]\right\}. \tag{15}$$

M3's SROCC performance using different a_4 and a_5 is shown in Fig. 3, where the full set of LIVE, IVC, Toyama, and LAR, as well as the subset of TID and CSIQ are tested. The difficult distortions are excluded from TID and CSIQ due to the reasons aforementioned. In Fig. 3, M3's performances always drop at small a_4 and a_5. Meanwhile, M3's performance at large a_4 or a_5 is bounded by the performance of the MSE, because: 1) when a_4 exceeds 100 and a_5 keeps positive and small, the CSF becomes "uniform" for all the subbands and thus the metric is equivalent to the MSE; 2) when a_5 exceeds 2 and a_4 keeps positive, the CSF has nonzero value only at $c_{0,0}$ and the metric is equivalent to the MSE on 1/8 downsampled images.

It is also found that the performance peaks occur at slightly different (a_4, a_5)-positions for different datasets. One reason is that the viewing conditions (e.g., view distance and pixel pitch of the monitor) are not identical in those subjective tests. The more important reason might be that the subjective quality assessment takes account of multiple factors more than the CSF and the metric performances in Fig. 3 are probably affected by the bias of other factors. Consequently, parameter selection from one dataset has the risk of overfitting. Nevertheless, it is reasonable to make selection according to multiple independent datasets. Another observation is that the performance surfaces have simple shape, since only two degrees of freedom are used to control the CSF and the DCTex metric has a simple form. Therefore, our model has fewer risks of overfitting than the other

complex nonlinear models. Finally, we choose $a_4 = 10$ and $a_5 = 1$; this configuration results in the CSF shown in Fig. 2.

D. Color DCTex

The gray DCTex metric can be extended for chromatic channels so as to guide the applications of color images. The color DCTex metric takes account of both the luminance channel and the chromatic channels as

$$D(X,Y) = g\sum_{k=1}^{K}\left[q_k\sum_{i=1}^{n_k}\sum_{j=1}^{64}\frac{c_j\cdot(u_{i,j}-v_{i,j})^2}{l_i}\right]. \quad (16)$$

A color image (e.g., RGB) is firstly transformed into a K-channel color space. In each channel k, the distortion is predicted by the luminance DCTex. After being weighted by q_k, the channel distortions are pooled. The texture masking parameters g and l for chromatic channels are still calculated from the corresponding luminance information.

In JPEG standard, color images are transformed to YUV space and the U and V components are often 1:2 down-sampled before frequency transform. Here, we give an instance of color DCTex, which adapts to the JPEG standard by that: 1) YUV space is used, indexed by $k = 1, 2, 3$, respectively; 2) the U and V components are 1:2 down-sampled; 3) the local roughness l for U and V channels is calculated from the corresponding down-sampled bock in Y channel; 4) the chromatic CSF for U and V channel is learned from subjective databases. At a common viewing distance, AC subbands of the 8×8 DCT on a downscaled 16×16 block usually present about $1.18 \sim 11.7$ cyc/deg, at which spatial frequency human sensitivity to the red-green and blue-yellow channels show a low-pass property [42]. We use the same a_4 in (13) for chromatic channels and Y channel, and only estimate a_5 in (13) and q in (16) for chromatic channels. Moreover, we use the identical q, a_4, and a_5 for U and Y channels. Given $q = 1$ for Y channel, we finally select $q = 0.8$ and $a_5 = 1.5$ for U and V channels, such that the performance of color DCTex metric is improved over that of luminance DCTex metric for most of the datasets.

Color DCTex's performance is shown in Column 13 of Tables I–V. The color DCTex does not always outperform the gray DCTex, yet also provides a consistent performance. Note that IVC database contains the compression in chromatic components, TID covers the additive chromatic noise, and CSIQ includes the additive pink noise; gray DCTex still does well in those chromatic distortions. This is probably another proof to the observation that perceptual color space is not a key factor for measuring the typical distortions [25].

E. Complexity Analysis

The DCTex metric is fast, since it is based on the orthogonal transform (DCT) on non-overlapping blocks at single scale, involving much less information than the case of overcomplete transform, overlapping blocks, or multi-scale analysis. DCTex metric can be embedded into the objective function of an optimization problem and keep the objective easy to be optimized. Note that DCTex is a weighted Euclidean distance in the DCT domain and the pixel domain. Given the reference images, there is a closed form of the gradient-descent solution to shorten the DCTex distance from the processed image to the reference. Therefore, DCTex metric is able to function in loop during optimization and ensures a fast solution. Comparatively speaking, DCTune is more complex to optimize due to its 4th order norm.

Another advantage of DCTex is that it decouples the image distortions into each block and each DCT subband. This assumption seems to be over-simplistic, but practical indeed. To be specific, it is reasonable to optimize the quantizer for each DCT subband one by one for image coding (as discussed in Section V). But for a metric which measures local distortions on overlapping blocks, the distortion in separated blocks (e.g., changing the DCT coefficient in two adjacent blocks) might affect the quality in a dependent way. Similarly, for a metric which measures the subband distortions on overcomplete transform domain, two modifications to different subbands (e.g., to modify two DCT coefficients of the same block) might influence the quality jointly. As a result, we have to optimize coding parameters (e.g., quantizers) together rather than one by one. Actually, minimization of SSIM in the coding scenario is nontrivial [43]. To the best of our knowledge, no literature has discussed how to optimize MSSIM and VIF. This issue will be further discussed in Section V-C.

V. Application in Perceptual Coding

A. Perceptual Coding Scheme

The DCTex metric is employed to improve JPEG image coding by means of a better rate-distortion (RD) optimization which would be more consistent with subjective evaluation than the traditional rate-MSE optimization. To minimize the RD function, the perceptual coding scheme optimizes the quantization table for the entire image and the quantizers for the transform blocks. Since JPEG has reserved bits to store user-configured quantization tables and the quantizer parameters can be recovered at the decoder, there is no extra overhead to transmit.

The RD objective function is the linear combination of the number of the bits B to code the image and the distortion D measured by DCTex, defined as

$$R = B + \lambda D. \quad (17)$$

The positive multiplier λ is chosen based on the theoretical rate-distortion curve of the image. We try to code the image several times with slightly different parameters to get several combinations of $\{(B_t, D_t)\}$ where t indexes the tth trial, and set λ to be equal to the slope of the normal to the distortion versus rate curve at the desired operating point:

$$\lambda = -\frac{\Delta B}{\Delta D} = -\arg\min_{\lambda,\tau}\sum_t |B_t - \lambda D_t - \tau|^2. \quad (18)$$

In (18), (λ, τ) can be regarded as the polynomial parameters to fit the linear regression of $B_t = \lambda \cdot D_t + \tau$. That is, $1/\lambda$ indicates the tangent direction of the local RD curve, and λ for the normal direction we are seeking. This way, the RD minimization is made along the normal of the current RD curve.

The RD minimization is carried out in two steps. The first step is searching for the optimal quantization table, and the second

one is choosing the optimal quantizer offsets for the transform blocks.

First, we try to optimize the DC components and the lowest five AC components (in the zigzag order) by adding an integer to the default quantization table, since the distortions measured by DCTex are most sensitive to the six coefficients in a transform block. The components that minimize the RD function will be chosen for the quantization table. After offline training on a host of images, we get the statistically optimal quantization table for the six coefficients, and then the encoder tries to modify the trained table within an integer range of ± 2. Note that JPEG standard has reserved the bits to transmit the quantization table, so no extra overhead is required.

Next, we try to optimize the high-frequency quantizers for the transform blocks. Due to the texture masking effect, subjects can tolerate more distortions in the textured blocks than in the smooth blocks. Uniform quantizers are probably not the best solution. As a result, we try the quantizers for the last 54 AC coefficients (in the zigzag order) by adding integer quantizer offset Δ as

$$Q_j = \begin{cases} Q_j & j \in \{1,2,9,17,10,3,4,11,18,25\} \\ Q_j + \text{round}\left(Q_j / \min\{Q\}\right) \cdot \Delta & \text{others.} \end{cases} \tag{19}$$

Here, j indexes the DCT coefficients and the corresponding quantizers in the raster scanning order, and {1, 2, 9, 17, 10, 3, 4, 11, 18, 25} corresponds to the lowest 10 coefficients in the zig-zag order. The quantizer for the DC coefficient ($j = 1$) will not be modified, because it probably causes a fluctuation in DPCM codes for the DC coefficients (i.e., the differences of every two consecutive DC coefficients after quantization) and thus inefficient entropy coding later. Only high-frequency quantizers are modified, since they are more probably quantized to zero so as to save bits in run-length coding. After trying a few quantizer offsets, the one that minimizes the RD function for the current block will be chosen. The quantizer offset value is tried within the range of [−4, 12].

To avoid transmitting overheads for the quantizers, we select particular blocks to be quantized in a checkerboard pattern, i.e., only the white blocks will be quantized by the optimized quantizer while its four black neighbors will be quantized by the default quantizer. At the decoder, the black blocks can be dequantized using the default quantizer, while the quantizer for the white block needs to be predicted. First, the decoder tries to dequantize the white block using the same range of quantizer offsets as the encoder (i.e. [−4, 12]). Then as a post-processing process, the true offset is predicted based on a maximum a posterior (MAP) criterion. The quantizer offset is expected to keep the gradients across the block borders smooth as

$$\widetilde{\Delta} = \arg\min_{\Delta} \left\{ \sum_{c=0}^{7} \left[\left| \nabla^2 I_r^{\Delta} \big|_{r=0,c} \right| + \left| \nabla^2 I_r^{\Delta} \big|_{r=9,c} \right| \right] + \sum_{r=0}^{7} \left[\left| \nabla^2 I_c^{\Delta} \big|_{r,c=0} \right| + \left| \nabla^2 I_c^{\Delta} \big|_{r,c=9} \right| \right] \right\}. \tag{20}$$

I^{Δ} is the quantized block using quantizer offset Δ. r and c are the pixel indices of the row and column, respectively (for the pixel inside the block, $0 \le r \le 7$ and $0 \le c \le 7$). $\nabla I_r|_{r,c} = I_{r,c} - I_{r-1,c}$ denotes the first order vertical gradient at the pixel (r, c), while $\nabla^2 I_c|_{r,c} = \nabla I_c|_{r,c} - \nabla I_c|_{r,c-1} = I_{r,c} - 2I_{r,c-1} + I_{r,c-2}$ denotes the second order horizontal gradient at the pixel (r, c).

The criterion of gradient smoothness cannot always ensure accurate predictions of the quantizers. Fortunately, even if the predicted quantizer contains some errors, it will only cause a distortion in high-frequency coefficients which will not significantly change the texture pattern in the block. Actually, the quantizers optimization does improve the RD performance (i.e., decrease the RD function) in most cases, despite the quality degradation due to inaccurate quantizers predictions. Sometimes, quantizers predictions can even improve the quality of reconstructed images since it works like a de-block algorithm. After choosing the optimal quantization table and the optimal quantizers, regular JPEG encoding processes are performed.

To implement the proposed perceptual coding scheme, an RDO module should be integrated into the JPEG encoder while a post-processing module into the decoder. By the standard JPEG decoder, the bit stream can be decoded successfully, but not accurately unless the post-processing is performed.

B. Experimental Results

We compared the DCTune codec [19] and the proposed DCTex codec in coding color images for different compression ratios. The coding efficiency is evaluated by perceptual RD curve. Subjective tests were conducted to assess perceptual distortion of coded images. As shown in Fig. 4, six color images from LIVE image databases were coded, with typical resolution of 768 × 512 or 480 × 720. Seventeen subjects (ten males and seven females) rated the image quality with 5-grade score (5 for the best and 1 for the worst) under the condition which strictly conformed to the standard recommendation [44]. Together with the reference image, each coded image was reconstructed and shown in a 19-inch LCD monitor. The coding scheme for the presented image was unknown to all subjects, and the presenting order is randomized for each subject. At the beginning, three training images, which spanned the range of slight, moderate, and heavy compression levels, were shown to stabilize the subjects' opinion. Among all the subjects, one male subject is detected as outlier according to [44]. The average of the 16 remain scores, i.e., mean opinion score (MOS), is used to indicate the perceptual distortion of a coded image. Since our LCD monitor with a pixel pitch of 0.294 mm/pixel was set at its recommended resolution of 1280 × 1024 and the viewing distance was three times the display's height (about 900 mm), we configure the DCTune codec with a pixel per degree of 53.43.

The RD curves are shown in Fig. 5–7, where the vertical axis is the subjective quality and the horizontal axis is the bit rate of coded images. The DCTex Matlab codec was based on the work in [45]. Interested readers can refer to the website: http://ivp.ee.cuhk.edu.hk/projects/demo/piqm/index.html.

We have three major observations based on the results shown in Fig. 5–7. First, the DCTex codec presents the better RD performance than the DCTune codec and the regular JPEG in statistics. Since the DCTune codec is often inferior to JPEG, we compare the bit rate of coded images given MOS of 3.5 between

Fig. 4. Six RGB images used in LIVE image database downscaled for display, including *Lighthouse, Plane, Buildings, Womanhat, Parrot* and *Rapids*.

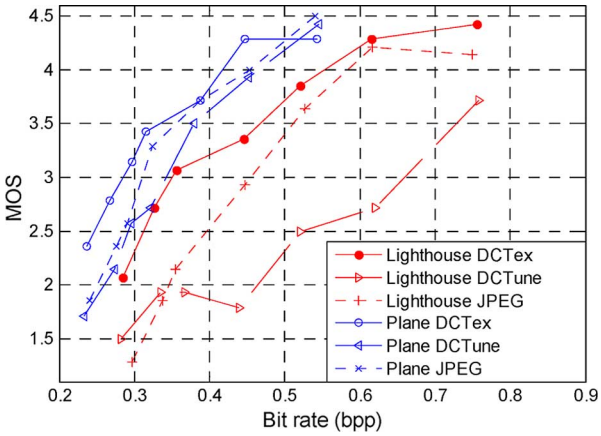


Fig. 5. RD curves of *Lighthouse* and *Plane* coded by JPEG, DCTune, and DCTex codec. The higher MOS means the less distortion.

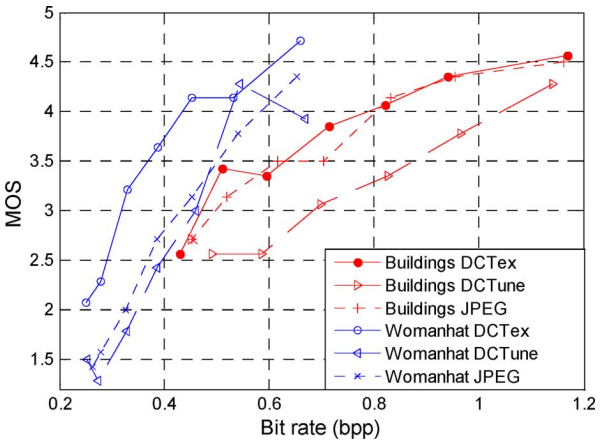


Fig. 6. RD curves of *Buildings* and *Womanhat* coded by JPEG, DCTune, and DCTex codec. The higher MOS means the less distortion.

the DCTex codec and JPEG in Table VI (when no image has exactly the expected MOS, we interpolate the bit rate according to the RD curve). The DCTex codec saves about 5% $\sim$ 26% bits given MOS of 3.5 for most cases, although costs a little more bits (<4%) for *Buildings* and *Rapids*.

Second, the DCTex codec achieves much better RD performance at the middle bit rate, yet the RD performance of the three codec are similar at the high bit rate. The high bit rate case

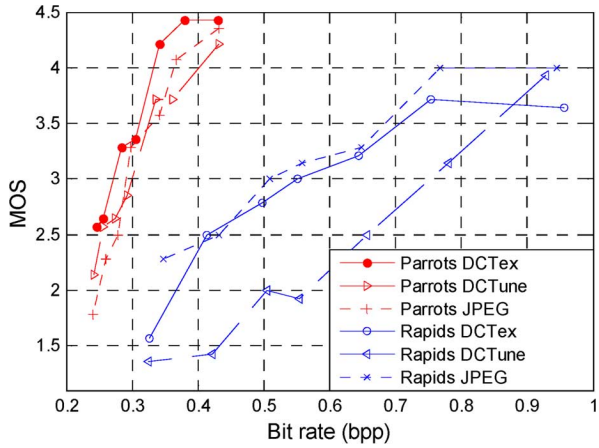


Fig. 7. RD curves of *Parrots* and *Rapids* coded by JPEG, DCTune, and DCTex codec. The higher MOS means the less distortion.

TABLE VI
BIT RATES FOR JPEG AND DCTEX CODEC AT MOS OF 3.5

image	Perceptual coding method	
	JPEG	DCTex
Lighthouse	0.511bpp	0.468bpp (−8.4%)
Plane	0.356bpp	0.333bpp (−6.5%)
Buildings	0.616bpp	0.629bpp (+2.1%)
Womanhat	0.501bpp	0.368bpp (−26.5%)
Parrots	0.330 bpp	0.311 bpp (−5.8%)
Rapids	0.684 bpp	0.706 bpp (+3.2%)

can be regarded as "perceptual lossless" since the corresponding MOS are often above 4 and "saturated".

Third, the DCTex codec outperforms the regular JPEG to a large extent for *Lighthouse* and *Buildings*. These two images both contain the smooth and texture regions yet few obvious regions of interest. When visual attention is not attracted by or focused on a few regions, texture masking effect probability becomes important, and thus the DCTex codec guide coding is more effective than regular JPEG.

C. Complexity Analysis

Although a little more complicated than the standard JPEG, the DCTex encoder is fairly fast. In our implementation above, 1) the DCT is performed only once; 2) the six components of the quantization table and the quantizer offsets are optimized separately; 3) the distortion can be calculated locally and repeated a few times. Suppose J (usually $6+n/2$) coding parameters need be optimized within an average range of R (no more than 17) and the image has a total of n blocks. When optimizing each parameter in the DCText encoder, only the distortion involved corresponding block or subband needs to be computed. It is equivalent to computing the entire image distortion by $O(R)$ times when coding an image. But if MSSIM or VIF guides the RD optimization, optimizing each parameter needs to perform inverse DCT (IDCT) and compute the entire image distortion, and such operations needs to be repeated by $O(R\cdot J\cdot n)$ times for coding an image.

The DCTune encoder appears fast, too, because it does not calculate the distortions in all subbands all the time. Instead, substituted by the ∞th norm for frequency error pooling, the

total distortion is approximated by the maximal one of subband distortions [23]. This assumption makes it reasonable to separately optimize each component of quantization table, but may lead to a suboptimal solution due to inaccurate quality prediction.

The proposed DCTex decoder is fast since the quantizer prediction requires only additions and comparisons. To be specific, it needs to perform IDCT only once during trying all the Δ from -4 to 12. When Δ increases by 1, the image is added by a fixed pattern in the pixel domain which is a constant combination of the last 54 DCT bases.

VI. Conclusion

For many image applications, it is very important to exploit the perceptual distortion in line with perception. In this paper, we propose the DCTex metric, which provides two advantages:

- *Accuracy*. It predicts image quality in a way that highly correlates with the subjectively-rated databases.
- *Simplicity*. It decomposes the distortion into independent blocks and subbands, and facilitates the perceptual rate-distortion optimization.

With the explicit distortion formulation, perceptual image coding is posed as an optimization problem. Currently, the bottleneck of RD optimization comes from lack of an efficient model of bit rate with respect to quantizers, given the simple quality metric. Nevertheless, since the proposed metric is easy to optimize, it may be applied to a broad range of applications.

References

[1] H. R. Sheikh, LIVE Image Quality Assessment Database, Release 2, 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality.

[2] P. Le Callet, Subjective Quality Assessment IRCCyN/IVC Database, 2005. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/.

[3] Y. Horita, MICT Image Quality Evaluation Database. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html.

[4] N. Ponomarenko, Tampere Image Database 2008 version 1.0, 2008. [Online]. Available: http://www.ponomarenko.info/tid2008.htm.

[5] D. M. Chandler and S. S. Hemami, A57 Database. [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[6] D. M. Chandler, CSIQ Image Database, 2010. [Online]. Available: http://vision. okstate.edu/index.php?loc=csiq.

[7] U. Engelke, Wireless Imaging Quality (WIQ) Database, 2010. [Online]. Available: http://www.bth.se/tek/rcg.nsf/pages/wiq-db.

[8] F. Autrusseau, Subjective Quality Assessment of LAR Coded Art Images, 2009. [Online]. Available: http://www.irccyn.ec-nantes.fr/~autrusse/Databases/ LAR.

[9] T. N. Pappas, J. P. Allebach, and D. L. Neuhoff, "Model-based digital halftoning," *IEEE Signal Processing Mag.*, vol. 20, pp. 14–27, 2003.

[10] T. N. Pappas, J. Chen, and D. Depalov, "Perceptually-based techniques for image segmentation and semantic classification," *IEEE Commun. Mag.*, vol. 45, pp. 44–51, 2007.

[11] H. O. Altun, A. Orsdemir, and G. Sharma *et al.*, "Optimal spread spectrum watermark embedding via a multistep feasibility evaluation," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 371–387, Feb. 2009.

[12] F. X. J. Lukas, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. 30, no. 7, pp. 1679–1692, 1982.

[13] W. Lin, D. Li, and X. Ping, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 900–909, 2005.

[14] A. Hyvärinen, *Natural Image Statistics A Probabilistic Approach to Early Computational Vision*. New York: Springer-Verlag, 2009.

[15] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. Int. Conf. Image Processing*, 1994, vol. II, pp. 982–986.

[16] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA: Morgan & Claypool, 2006.

[17] S. Winkler, "Perceptual video quality metrics—A review," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds. Boca Raton, FL: CRC, 2005.

[18] Z. Wang, A. C. Bovik, and H. R. Sheikh *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[19] S. Voloshynovskiy, A. Herrigel, and N. Baumgaertner *et al.*, "A stochastic approach to content adaptive digital image watermarking," in *Proc. Information Hiding*, 2000, vol. 1768, pp. 211–236.

[20] H. H. Y. Tong and A. N. Venetsanopoulos, "A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking," in *Proc. Int. Conf. Image Processing*, 1998, vol. 3, pp. 428–432.

[21] M. A. Georgeson, "Over the limit: Encoding contrast above threshold in human vision," in *Limits of Vision*, J. Kulikowski, Ed. London, U.K.: Erlbaum, 1990, pp. 106–119.

[22] P. J. Bex and K. Langley, "The perception of suprathreshold contrast and fast adaptive filtering," *J. Vision*, vol. 7, no. 12, pp. 1–23, 2007.

[23] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," in *Society for Information Display Digest of Technical Papers, XXIV*, 1993, pp. 946–949.

[24] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[25] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. Chichester, U.K.: Wiley, 2005.

[26] A. Ninassi, O. Le Meur, and P. Le Callet *et al.*, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proc. Int. Conf. Image Processing*, 2007, vol. 1–7, pp. 733–736.

[27] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, 2010.

[28] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, 2010.

[29] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[30] A. B. Watson, DCTune 2.0: Perceptual Optimization of JPEG Images (and Perceptual Error Metric). [Online]. Available: http://vision.arc.nasa.gov/dctune.

[31] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[32] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[33] N. Ponomarenko, F. Silvestri, and K. Egiazarian *et al.*, "On between-coefficient contrast masking of DCT basis functions," in *Proc. Int. Workshop Video Processing and Quality Metrics for Consumer Electronics VPQM*, 2007.

[34] F. Zhang, S. Li, and L. Ma *et al.*, "Limitation and challenges of image quality measurement," in *Proc. SPIE Visual Communications and Image Processing Conf.*, 2010, pp. 1–8.

[35] M. Sendashonga and F. Labeau, "Low complexity image quality assessment using frequency domain transforms," in *Proc. Int. Conf. Image Processing*, 2006, pp. 385–388.

[36] X. B. Gao, W. Lu, and D. C. Tao *et al.*, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[37] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Human Vision, Visual Process., Digit. Display III*, vol. 1666, pp. 365–374, 1992.

[38] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Diego, CA: Elsevier, 2002, p. 216.

[39] VQEG, Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, 2000.

[40] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. 33, no. 3, pp. 551–557, 1985.

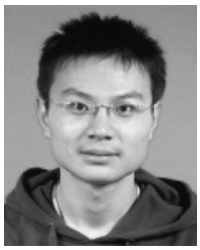
[41] K. N. Ngan, K. S. Leong, and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 11, pp. 1743–1750, Nov. 1989.

[42] K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.*, vol. 359, pp. 381–400, 1985.

[43] Z. Wang, Q. Li, and X. L. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," in *Proc. Int. Conf. Image Processing*, 2007, vol. 1–7, pp. 685–688.
[44] ITU-R-Recommendation-BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002.
[45] W. Lin and L. Dong, "Adaptive down-sampling to improve image compression at low bit rates," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2513–2521, Sep. 2006.

Fan Zhang (M'11) received the B. E. and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2008, respectively, both in electronic and information engineering.

He was a visiting student at the Nanyang Technological University in 2008 and a postdoctor at the Chinese University of Hong Kong from 2009 to 2010. He has been a research engineer in Technicolor (Beijing, China) Company since 2010. His research interests are quality of experience and perceptual watermarking.



Lin Ma (S'09) received the B.S. and M.S. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, both in computer science. He is now pursuing the Ph.D. degree in the Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK).

He was a Research Intern in Microsoft Research Asia from October 2007 to March 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from November 2008 to July 2009. His research interests include image/video quality assessment, super-resolution, restoration, compression.

Mr. Ma received the best paper award in the Pacific-Rim Conference on Multimedia (PCM) 2008.

Songnan Li (S'08) received the B.S. and M.S. degrees from the Department of Computer Science, Harbin Institute of Technology, Harbin, China, in 2004 and 2006, respectively. He is now pursuing the Ph.D. degree in the Department of Electronic Engineering, Chinese University of Hong Kong.

His research interests include visual quality assessment, video de-interlacing, video compression, and algorithm complexity optimization.

King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from the Loughborough University, Loughborough, U.K.

He is currently a Chair Professor at the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a full Professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Perth, Australia. He holds honorary and visiting professorships at numerous universities in China, Australia, and Southeast Asia. He has published extensively including three authored books, five edited volumes, over 300 refereed technical papers, and nine edited special issues in journals. In addition, he holds ten patents in the areas of image/video coding and communications.

Prof. Ngan is an Associate Editor of the *Journal on Visual Communications and Image Representation*, as well as an Area Editor of *EURASIP Journal of Signal Processing: Image Communication*, and served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and *Journal of Applied Signal Processing*. He chaired a number of prestigious international conferences on video signal processing and communications, and served on the advisory and technical committees of numerous professional organizations. He was a general co-chair of the IEEE International Conference on Image Processing (ICIP) held in Hong Kong in September 2010. He is a Fellow of IET (U.K.) and IEAust (Australia), and an IEEE Distinguished Lecturer in 2006–2007.