Image Quality Assessment Based on Multiscale Geometric Analysis

Xinbo Gao, *Senior Member, IEEE*, Wen Lu, Dacheng Tao, *Member, IEEE*, and Xuelong Li, *Senior Member, IEEE*

***Abstract*—Reduced-reference (RR) image quality assessment (IQA) has been recognized as an effective and efficient way to predict the visual quality of distorted images. The current standard is the wavelet-domain natural image statistics model (WNISM), which applies the Kullback–Leibler divergence between the marginal distributions of wavelet coefficients of the reference and distorted images to measure the image distortion. However, WNISM fails to consider the statistical correlations of wavelet coefficients in different subbands and the visual response characteristics of the mammalian cortical simple cells. In addition, wavelet transforms are optimal greedy approximations to extract singularity structures, so they fail to explicitly extract the image geometric information, e.g., lines and curves. Finally, wavelet coefficients are dense for smooth image edge contours. In this paper, to target the aforementioned problems in IQA, we develop a novel framework for IQA to mimic the human visual system (HVS) by incorporating the merits from multiscale geometric analysis (MGA), contrast sensitivity function (CSF), and the Weber's law of just noticeable difference (JND). In the proposed framework, MGA is utilized to decompose images and then extract features to mimic the multichannel structure of HVS. Additionally, MGA offers a series of transforms including wavelet, curvelet, bandelet, contourlet, wavelet-based contourlet transform (WBCT), and hybrid wavelets and directional filter banks (HWD), and different transforms capture different types of image geometric information. CSF is applied to weight coefficients obtained by MGA to simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS. JND is finally introduced to produce a noticeable variation in sensory experience. Thorough empirical studies are carried out upon the LIVE database against subjective mean opinion score (MOS) and demonstrate that 1) the proposed framework has good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images, 2) different transforms in MGA under the new framework perform better than the standard WNISM and some of them even perform better than the standard full-reference IQA model, i.e., the mean structural similarity index, and 3) HWD performs best among all transforms in MGA under the framework.**

***Index Terms*—Contrast sensitivity function (CSF), human visual system (HVS), image quality assessment (IQA), just noticeable difference (JND), multiscale geometric analysis (MGA), reduced-reference (RR).**

I. Introduction

THE objective of *image quality assessment* (IQA) [38] is to provide computational models to measure the perceptual quality of an image. In recent years, a large number of methods have been designed to evaluate the quality of an image, which may be distorted during acquisition, transmission, compression, restoration, and processing (e.g., watermark embedding). The past five years have demonstrated and witnessed the tremendous and imminent demands of IQA methods in various applications, including evaluating and optimizing image processing algorithms and systems.

Existing IQA methods can be categorized into subjective [21] and objective methods [38]. Results of a subjective method are directly given by human observers, so it is probably the best way to assess the quality of an image. This is because human observers are the ultimate receivers of the visual information contained in an image. However, subjective IQA methods are expensive and time consuming, so they cannot be easily and routinely performed for many scenarios, e.g., real time systems. The latter depends on the quantified parameters which are obtained from metrics to measure the image quality. Metrics are usually obtained from either reference or distorted images to reflect a number of image characteristics. Evaluation results obtained from a good objective IQA method should be statistically consistent with subjective methods. According to the availability of a reference image, there is a general agreement [38] that objective quality metrics can be divided into three categories: *full-reference* (FR), *no-reference* (NR), and *reduced-reference* (RR) methods.

To evaluate the quality of a distorted image, FR methods [29] usually provide the most precise evaluation results in comparing with NR and RR. Conventional FR IQA methods calculate pixel-wise distances, e.g., *peak signal-to-noise ratio* (PSNR) and *mean square error* (MSE), between a distorted image and the corresponding reference, but they have not been in agreement with perceived quality measurement widely [1], [12]. Recently, a number of efforts [9], [33] have been proposed for IQA in developing metrics [8], [18] to reflect image visual intrinsic properties according to well known characteristics of the *human visual system* (HVS). An alternative complementary framework, i.e., *structural similarity* (SSIM), for IQA is introduced by Wang *et al.* [38] based on the degradation of structural information. Extensive performance evaluations have

Manuscript received June 20, 2008; revised February 10, 2009. First published May 12, 2009; current version published June 12, 2009. This work was supported in part by: the National Science Foundation of China (60771068, 60702061, 60832005); the Open-End Fund of National Laboratory of Pattern Recognition in China and the National Laboratory of Automatic Target Recognition, Shenzhen University, China; the Program for Changjiang Scholars and innovative Research Team in University of China (IRT0645); and the Nanyang Technological University Start-Up Grant (M58020010). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Laurent Younes.

X. Gao and W. Lu are with the School of Electronic Engineering, Xidian University, Xi'an 710071, Shaanxi Province, China (e-mail: xbgao@mail.xidian.edu.cn; luwen@mail.xidian.edu.cn).

D. Tao is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dctao@ntu.edu.sg).

X. Li is with the School of Computer Science and Information Systems, Birkbeck College, University of London, London WC1E 7HX, U.K. (e-mail: xuelong@dcs.bbk.ac.uk).

Digital Object Identifier 10.1109/TIP.2009.2018014

shown SSIM achieves the state-of-the-art performance. However, SSIM only considers the local correlation in an image, so detected features are not enough for precise IQA. In addition, all FR methods demand the corresponding original or perfect image as reference which is always not available in practical scenarios.

To evaluate the quality of a distorted image in real time systems, NR methods [20], [38], which output evaluation results without the corresponding reference, have been designed. However, all these methods rely on strong hypotheses, i.e., they are designed for one or a set of predefined distortions, so there is a big gap between NR methods and real scenarios. Wang *et al.* [38] proposes a computational and memory efficient quality assessment model for JPEG images to deal with the blocking-artifact. Sheikh *et al.* [38] uses *natural scene statistics* (NSS) to measure the quality of a JPEG2000 image. They claimed that natural scenes contain nonlinear dependencies that are disturbed by the compression process, and this disturbance can be quantified and related to human perceptions quality. Therefore, they perform well only when distortions are known and modeled precisely.

To provide a compromise between FR and NR, RR methods [7], [16], [19], which become popular in recent years, have been designed for IQA by employing partial information of the corresponding reference. Based on results in natural image statistics, Wang *et al.* [38] proposed the *wavelet-domain natural image statistic metric* (WNISM), which achieves promising performance for image visual perception quality evaluation. The underlying factor in WNISM is the marginal distribution of wavelet coefficients of a natural image conforms to the generalized Gaussian distribution. Based on this fact, WNISM measures the quality of a distorted image by the fitting error between the wavelet coefficients of the distorted image and the Gaussian distribution of the reference.

Although WNISM has been recognized as the standard method for RR IQA, it fails to consider the statistical correlations of wavelet coefficients in different subbands and the visual response characteristics of the mammalian cortical simple cells. Moreover, wavelet transforms cannot explicitly extract the image geometric information, e.g., lines and curves, and wavelet coefficients are dense for smooth image edge contours. In summary, features, which serve the key role in metric-based IQA (including both FR and RR), utilized in WNISM are not effective enough. Therefore, there is still a big room to further improve the performance of RR IQA.

In this paper, to target the aforementioned problems in WNISM, to further improve the performance of RR IQA, and to broaden RR IQA related applications, a novel HVS driven framework is proposed for IQA. This framework is constructed by pooling *multiscale geometric analysis* (MGA) [10], [27], *contrast sensitivity function* (CSF) [35], and the Weber's law of *just noticeable difference* (JND) [35] together. The new framework is consistent with HVS: MGA decomposes images for feature extraction to mimic the multichannel structure of HVS, CSF re-weights MGA decomposed coefficients to mimic the nonlinearities inherent in HVS, and JND produces a noticeable variation in sensory experience. This framework contains a number of different ways for IQA because MGA offers a series of transforms including wavelet [23], curvelet [6], bandelet [26], contourlet [11], *wavelet-based contourlet transform* (WBCT) [13], and *hybrid wavelets and directional filter banks* (HWD) [14], and different transforms capture different types of image geometric information. Extensive experiments based on LIVE database [28] against subjective *mean opinion score* (MOS) [21] have been conducted to demonstrate the effectiveness of the new framework.

The remainder of the paper is organized as follows. In Section II, an overview of MGA is provided to better understand the advantages of the proposed framework. Section III presents the proposed HVS driven framework. Experimental results are presented in Section IV, and Section V concludes.

II. Multiscale Geometric Analysis

Wavelet transform [23] have been successfully applied in a wide variety of signal processing tasks, e.g., speech signal compression and voice-based person identification, because it is an optimal greedy approximation to extract singularity structure for 1-D piecewise smooth signals. Although the 2-D extension, i.e., 2-D wavelet transform, can also be applied to image processing relevant applications, e.g., compression, de-noising, restoration, segmentation, and structure detection, it can only deal with the singularity problem of point. To our knowledge, image contains rich and varied information, e.g., texture and edges, and wavelet transform is not effective in dealing with directional information. Therefore, it is essential to develop a directional representation framework for precisely detecting orientations of singularities like edges in a 2-D image while providing near optimal sparse representations.

Multiscale geometric analysis (MGA) [26] is such a framework for optimally representing high-dimensional function. It is developed, enhanced, formed and perfected in signal processing, computer vision, machine learning, and statistics. MGA can detect, organize, represent, and manipulate data, e.g., edges, which nominally span a high-dimensional space but contain important features approximately concentrated on lower dimensional subsets, e.g., curves. MGA contains a large number of tools and gets wavelet transform involved as a special case. MGA tools range from Jones' traveling salesman theorem for multiscale approximation of data in dyadic cubes, to beamlet analysis for multiscale radon transformation, and to curvelet analysis for special space-frequency tilings. Therefore, MGA can be utilized to a large variety of applications, e.g., medical imaging, object categorization, and image compression. However, it is still not clear in how to apply MGA transforms for IQA, although they are plausibly for this application. This is because MGA can analyze and approximate geometric structure while providing optimally sparse representations.

For IQA, we need to find MGA transforms, which perform excellently for reference image reconstruction, have perfect perception of orientation, are computationally tractable, and are sparse and effective for image representation. Among all requirements for IQA, effective representation of visual information is especially important. Natural images are not simple stacks of 1-D piecewise smooth scan-lines and points of dis-

TABLE I
MAIN FEATURES CAPTURED BY DIFFERENT MGA TRANSFORMS

Transform	Main feature captured by MGA methods
Wavelet	Point
Curvelet	Continues closed curve on smooth plane C^2
Bandelet	Continues closed curve on smooth plane $C^{\alpha}(\alpha > 2)$
Contourlet	Area with subsection smooth contour
WBCT	Area with smooth contour
HWD	Area with smooth contour with angle

continuity are typically located along smooth curves owing to smooth boundaries of physical objects. As a result of a separable extension from 1-D bases, 2-D wavelet transform is not good at representing visual information of image. Consequently, when we dispose of image with linear features, wavelet transform is not effective. However, MGA transforms can capture the characteristics of image, e.g., lines, curves, cuneiforms and the contour of object. As mentioned in Table I, different transforms of MGA capture different features of an image, and complement to each other. Based on mentioned requirements for IQA, it is reasonable to consider a wide range of explicit interactions between multiscale methods and geometry, e.g., curvelet [6], bandelet [26], contourlet [11], *wavelet-based contourlet transform* (WBCT) [13], and *hybrid wavelets and directional filter banks* (HWD) [14].

Curvelet transform [6] combines multiscale ridgelets with a spatial bandpass filtering operation. It captures continues closed curves on a smooth plane C^2. Bandelet transform [26] tracks regular geometric orientation of adaptive images and it captures continues closed curves on a smooth plane $C^{\alpha}(\alpha > 2)$. Contourlet transform [11] aims to achieve an optimal approximation rate of piecewise smooth functions with discontinuities along twice continuously differentiable curves. Therefore, it captures areas with subsection smooth contours. WBCT [13] and HWD [14] are designed to optimize the representation of image features without redundancy.

III. MULTISCALE GEOMETRIC ANALYSIS-BASED IMAGE QUALITY ASSESSMENT

As discussed in Section II, MGA contains a series of transforms, which can analyze and approximate geometric structure while providing near optimal sparse representations. The image sparse representation means we can represent the image by a small number of components, so little visual changes of the image will affect these components significantly. Therefore, sparse representations can be well utilized for IQA. In this paper, we develop a novel framework for IQA by applying MGA transforms to decompose images and extract effective features. This framework quantifies the errors between the distorted and the reference images by mimicking the error sensitivity function [38] in the *human visual system* (HVS). The objective of this framework is providing IQA results for distorted images, which have good consistency with subjective perception values. This framework incorporates merits from three components, i.e., MGA, *contrast sensitivity function* (CSF), and the Weber's law of *just noticeable difference* (JND), to model the process of image perception.

Fig. 1 shows the framework for IQA and it works with the following stages: 1) an MGA transform, e.g., curvelet, bandelet, contourlet, WBCT, and HWD, is utilized to decompose both the reference image at the sender side and the distorted image at the receiver side, 2) CSF masking is utilized to balance subbands coefficients in different scales obtained by the MGA transform. With this stage, we can simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS, 3) JND produces a noticeable variation in sensory experience, 4) a histogram is constructed for image representation, each bin of the histogram corresponds to the amount of visual sensitive coefficients of a selected subband, and finally the normalization step is applied to the histogram, and 5) the IQA result is the transformed city-block distance between the normalized histograms of the reference and distorted images. In this section, these five stages are detailed in the following subsections.

A. MGA-Based Feature Extraction

In this paper, a number of MGA transforms, which are curvelet, bandelet, contourlet, WBCT and HWD, are considered for image decomposition and feature extraction. Moreover, wavelet is utilized as the baseline for comparison. In this framework, MGA is utilized to decompose images and then extract features to mimic the multichannel structure of HVS.

- **Wavelet transform**: It [23] is popular to analyze signals in both time and frequency domains simultaneously and adaptively. By using multiscale operation, it extracts effective features to represent signals, especially for nonstationary signals (as mentioned in Table I, it extracts singularity structure of an image). In our experiments, three-level wavelet transform with "Daubechies" filters, is applied to decompose the image into nine highpass subbands and a lowpass residual subband. Coefficients in all nine highpass subbands are preserved for further processing, e.g., CSF masking and JND. As shown in Fig. 2, selected subbands are marked with white dashed boxes and numerals.
- **Curvelet transform:** It [6] is a higher dimensional extension of wavelet transform. It captures curved singularities with very few coefficients in a nonadaptive manner and remains coherent waveforms under the action of the wave equation in a smooth medium. Therefore, it represents images at different scales and different orientations (as mentioned in Table I, it captures continues closed curves on a smooth plane C^2). Theoretically, it is a multiscale pyramid transform and each scale captures different orientations in different positions. In fine scales, it captures needle-shaped elements. In the proposed framework, coefficients in all subbands but the low frequency subband are selected for further processing.
- **Bandelet transform**: It [26] takes the advantage of the geometrical regularity of image structure, represents sharp image transitions, e.g., edges, especially, decomposes images along multiscale vectors that are elongated in the direction of a geometric flow. In the new framework, coeffi-

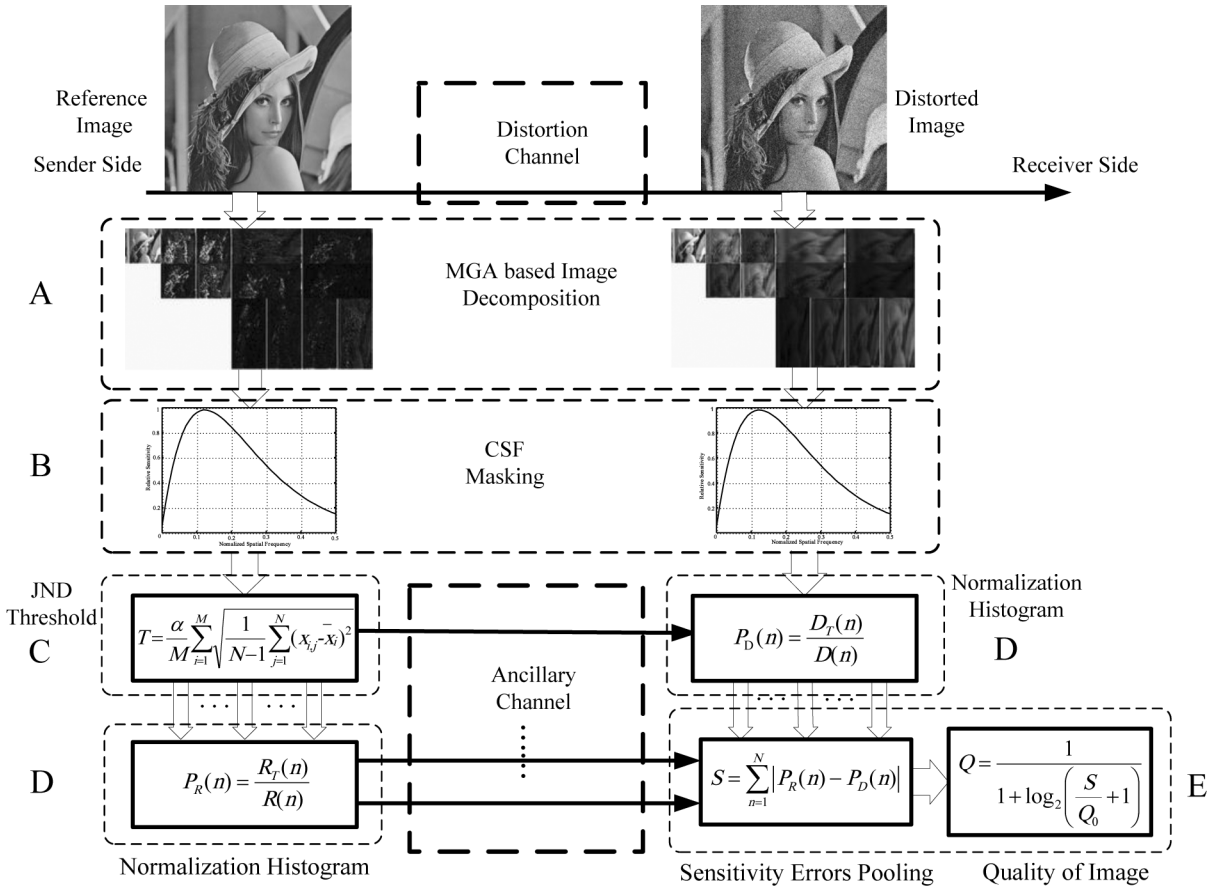


Fig. 1. Multiscale geometric analysis-based image quality assessment framework. (Sender side is applied to extract the normalized histogram of the reference image. Receiver side is applied to extract the normalized histogram of the distorted image. Ancillary channel is applied to transmit the extracted normalized histogram.)

cients in all highpass subbands are selected for further processing. As shown in Fig. 3, selected subbands are marked with white dashed boxes and numerals.

- **Contourlet transform**: It [11] is constructed via filter banks and can be viewed as an extension of wavelets with directionality. For implementation, it utilizes Laplacian pyramid [2] to capture point discontinuities and directional filter banks [3] to link point discontinuities into linear structures. Based on these two steps, it captures the intrinsic geometrical structure of an image. In the proposed framework, the image is decomposed into three pyramidal levels. Based on the characteristics of directional filter banks ("9–7" biorthogonal filters [3], [5]) for decomposition, coefficients in half of the directional subbands are selected for further processing. As shown in Fig. 4, contourlet transform is utilized to decompose an image and a set of selected subbands are marked with white dashed boxes and numerals.
- **WBCT**: It [13] is an enhanced version of contourlet transform based on two stages of filter banks that are nonredundant and perfect for reconstruction. It decomposes images both radially and angularly to an arbitrary extent and obeys the anisotropy scaling law. Therefore, it approximates natural images containing contours and oscillatory patterns efficiently. As mentioned in Table I, it captures areas with smooth contours. In this proposed framework, the image is decomposed into two levels based on wavelet transform "Daubechies" filters and the number of levels for directional filter banks ("9–7" biorthogonal filters) decomposition is three at each highpass of the finest and finer scale. Based on these filters, an image is decomposed into 48 high frequency directional subbands and a lowpass residual subband. Coefficients in half of the subbands at each fine scale are selected for further processing. As shown in Fig. 5, WBCT is utilized to decompose an image and a set of selected subbands are marked with white dashed boxes and numerals.
- **HWD Transform**: It [14] is similar to WBCT and employs wavelet transform for multiscale decomposition. Directional filter banks in HWD1 and modified directional filter banks in HWD2 are applied to some subbands obtained by wavelet transform. In this proposed framework, image is decomposed into two levels based on wavelet transform "Daubechies" filters and the number of levels for directional filter banks ("9–7" biorthogonal filters) decomposition is three at each highpass of the finest and finer scale. Based on these filters, an image is decomposed into 32 high frequency directional subbands and a lowpass residual subband. Coefficients in half of subbands at each fine scale are selected for further processing. As

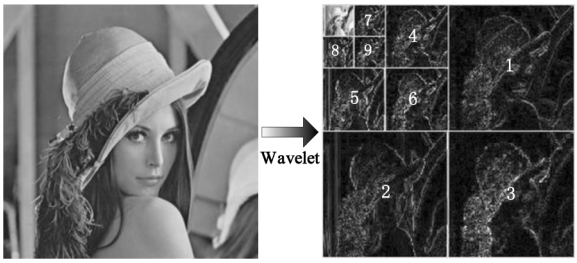


Fig. 2. Wavelet transform-based image decomposition.

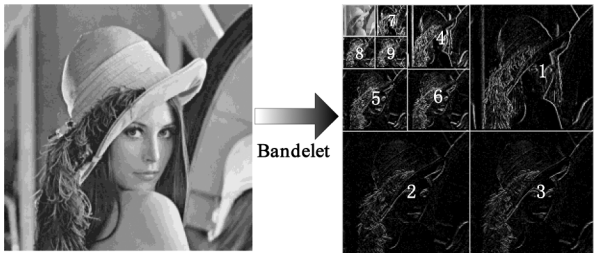


Fig. 3. Bandelet transform-based image decomposition.

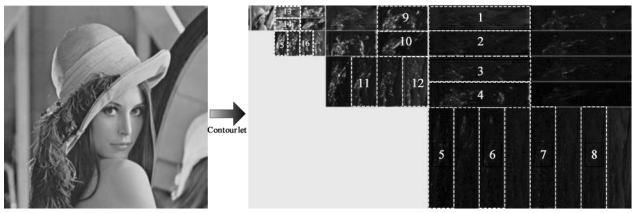


Fig. 4. Contourlet transform-based image decomposition.

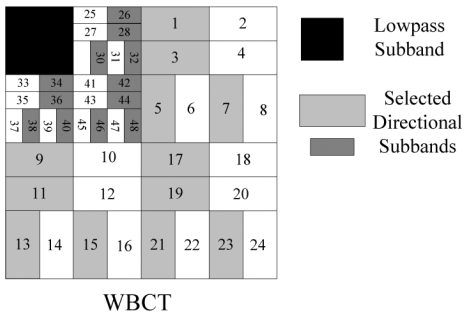


Fig. 5. WBCT-based image decomposition.

shown in Fig. 6, HWD transform is utilized to decompose an image and a set of selected subbands are marked with white dashed boxes and numerals.

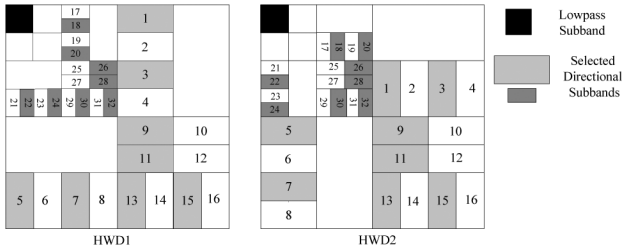


Fig. 6. HWD transform-based image decomposition.

B. CSF Masking

MGA is introduced to decompose images and then extract features to mimic the multichannel structure of HVS, i.e., HVS [36] works similar to a filter bank (containing filters with various frequencies). CSF [24] measures how sensitive we are to the various frequencies of visual stimuli, i.e., we are unable to recognize a stimuli pattern if its frequency of visual stimuli is too high. For example, given an image consisting of horizontal black and white stripes, we will perceive it as a gray image if stripes are very thin; otherwise, we can distinguish these stripes. Because coefficients in different frequency subbands have different perceptual importance, it is essential to balance the MGA decomposed coefficients via a weighting scheme, CSF masking. In this framework, the CSF masking coefficients are obtained by *modulation transfer function* (MTF) [24], i.e.,

$$H(f) = a(b + cf)\exp(-cf)^d \tag{1}$$

where $f = f_n \cdot f_s$, the center frequency of the band, is the radial frequency in cycles/degree of the visual angle subtended, f_n is the normalized spatial frequency with units of cycles/pixels, and f_s is the sampling frequency with units of pixels/degree. According to [4], a, b, c, and d are 2.6, 0.192, 0.114, and 1.1, respectively.

The sampling frequency f_s is defined as

$$f_s = \frac{2 \cdot v \cdot \tan(0.5^\circ) \cdot r}{0.0254} \tag{2}$$

where v is the viewing distance with units of meter and r is the resolution power of the display with units of pixels/inch. In this framework, v is 0.8 meter (about 2–2.5 times height of the display), the display is 21 inches with the resolution of 1024×768, and $r = \sqrt{1024^2 + 768^2}/21 = 61$ pixels/inch. According to the *Nyquist sampling theorem*, f changes from 0 to $f_s/2$, so f_n changes from 0 to 0.5. Because MGA is utilized to decompose an image into three scales from coarse to fine, we have three normalized spatial frequencies, $f_{n1} = 3/32, f_{n2} = 3/16, f_{n3} = 3/8$. Weighting factors are identical for coefficients in an identical scale.

For example, if contourlet transform is utilized to decompose an image, we obtain a series of contourlet coefficients $c_{i,j}^k$, where k denotes the level index (the scale sequence number) of contourlet transform, i stands for the serial number of directional subband index at the kth level, and j represents the coefficient index. By using CSF masking, the coefficient $c_{i,j}^k$ is scaled to $x_{i,j}^k = H(f_k) \cdot c_{i,j}^k$.

C. JND Threshold

Because HVS is sensitive to coefficients with the larger magnitude, it is valuable to preserve visually sensitive coefficients. JND, a research result in psychophysics, is a suitable way for this function. It measures the minimum amount, by which stimulus intensity must be changed to produce a noticeable variation in the sensory experience. In our framework, MGA is introduced to decompose an image and highpass subbands contain the primary contours and textures information of the image,

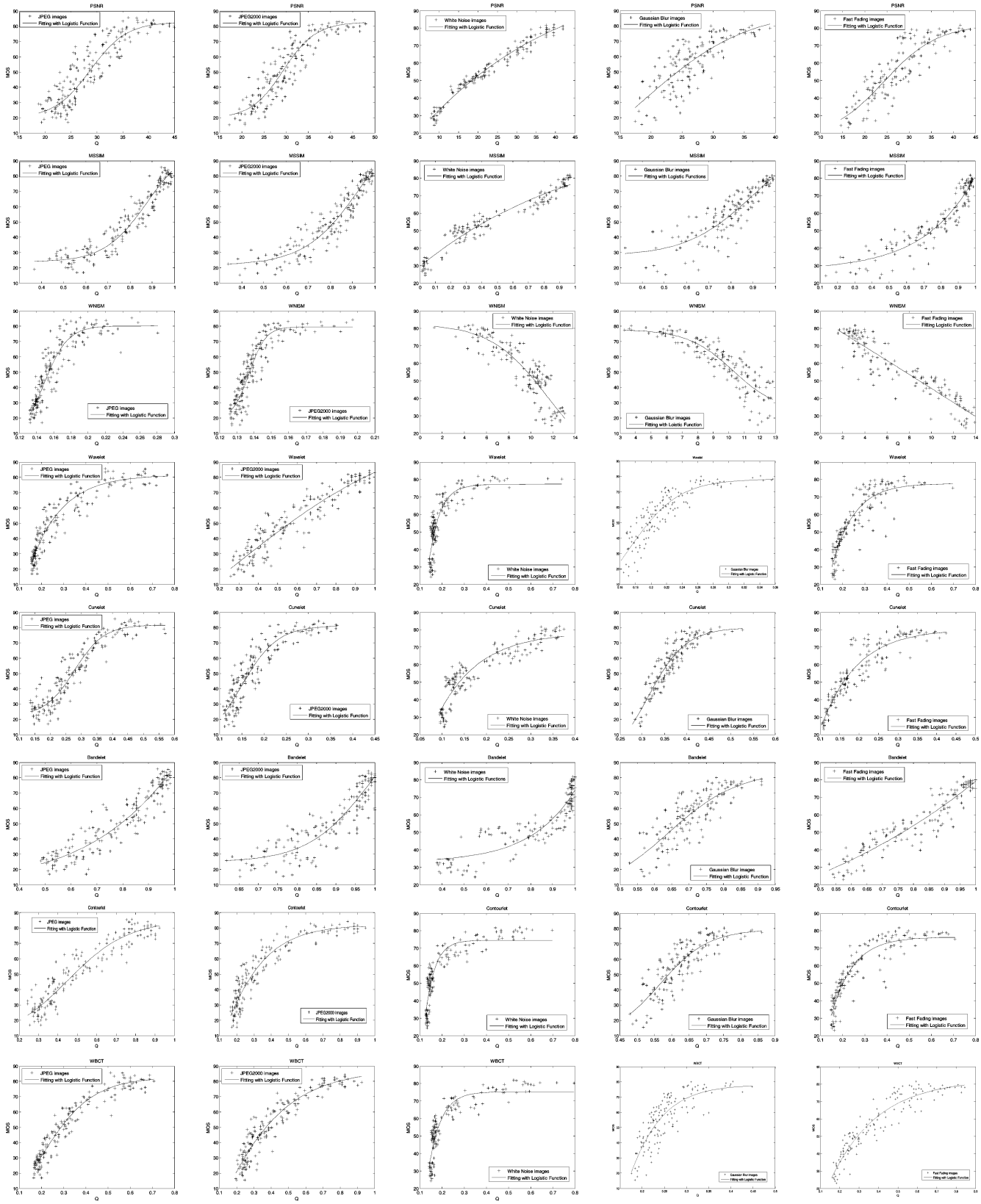


Fig. 7. Scatter plots of MOS versus kinds of IQA methods, i.e., PSNR, MSSIM, WNISM, Wavelet, Curvelet, Bandelet, Contourlet, WBCT, HWD1, and HWD2, for JPEG, JPEG2000, WN, Gblur, and FF distorted images.

CSF masking makes coefficients have similar perceptual importance in different frequency subbands, and then JND is calculated to obtain a threshold to remove visually insensitive coefficients. The amount of visual sensitive coefficients reflects

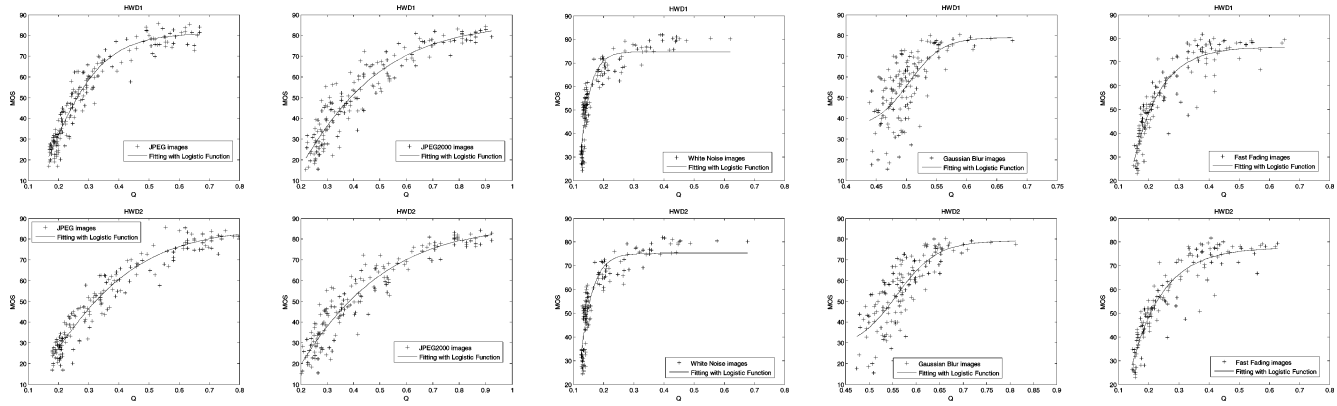


Fig. 7. (*Continued*) Scatter plots of MOS versus kinds of IQA methods, i.e., PSNR, MSSIM, WNISM, Wavelet, Curvelet, Bandelet, Contourlet, WBCT, HWD1, and HWD2, for JPEG, JPEG2000, WN, Gblur, and FF distorted images.

TABLE II
PERFORMANCE OF PSNR, MSSIM, AND WNISM ON THE LIVE DATABASE

		JPEG					JPEG2000				
Metric	Type	CC	ROCC	OR	MAE	RMSE	CC	ROCC	OR	MAE	RMSE
PSNR	FR	0.9229	0.8905	0.1886	7.1180	9.1540	0.9330	0.9041	0.0947	6.407	8.3130
MSSIM	FR	0.9674	0.9485	0.0400	4.7710	5.8320	0.9490	0.9368	0.0650	5.422	6.7090
WNISM	RR	0.9291	0.9069	0.1486	6.0236	8.2446	0.9261	0.9135	0.1183	6.135	7.9127
		WN					Gblur				
Metric	Type	CC	ROCC	OR	MAE	RMSE	CC	ROCC	OR	MAE	RMSE
PSNR	FR	0.9858	0.9853	--	2.1639	2.6797	0.7834	0.7816	--	9.7723	9.7723
MSSIM	FR	0.9695	0.9629	--	3.2566	3.9163	0.8740	0.8942	--	5.7595	7.6391
WNISM	RR	0.8897	0.8703	--	5.8079	7.2910	0.8873	0.9147	--	5.5555	7.2512
		FF									
Metric	Type	CC	ROCC	OR	MAE	RMSE					
PSNR	FR	0.8895	0.8903	--	7.5158	7.5158					
MSSIM	FR	0.9428	0.9411	--	4.2968	5.4846					
WNISM	RR	0.9230	0.9229	--	4.8330	6.3277					

the visual quality of the reconstructed images. The lower the JND threshold is, the more coefficients are utilized for image reconstruction and the better visual quality of the reconstructed image is. Therefore, the normalized histogram reflects the visual quality of an image. Here, the JND threshold is defined as

$$T = \frac{\alpha}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{N_i - 1}\sum_{j=1}^{N_i}(x_{i,j} - \bar{x}_i)^2} \tag{3}$$

where $x_{i,j}$ is the jth coefficient of the ith subband in the finest scale and $\bar{x}_i$ is the mean value of the ith subband coefficients, M is the amount of selected subbands in the finest scale, N_i is the amount of coefficients of ith subband, and α is a tuning parameter corresponding to different types of distortion.

D. Normalized Histogram for Image Representation

By using JND threshold T, we can count the number of visually sensitive coefficients in the n th selected subband and define the value as $C_T(n)$, which mean the number of coefficients in the nth selected subband which are larger than T obtained from (5). The number of coefficients in the nth selected subband is $C(n)$. Therefore, for a given image, we can obtain the normalized histogram with L bins (L subbands are selected) for representation and the nth entry is given by

$$P(n) = \frac{C_T(n)}{C(n)}. \tag{4}$$

E. Sensitivity Errors Pooling

Based on (4), we can obtain the normalized histograms for both the reference and the distorted images as $P_R(n)$ and $P_{\mathrm{D}}(n)$, respectively. In this framework, we define the metrics of the distorted image quality as

$$Q = \frac{1}{1 + \log_2\left(\frac{S}{Q_0} + 1\right)} \tag{5}$$

where $S = \sum_{n=1}^{L}|P_R(n) - P_D(n)|$ is the city-block distance between $P_R(n)$ and $P_{\mathrm{D}}(n)$, and Q_0 is a constant used to control the scale of the distortion measure. In this framework, we set Q_0 as 0.1. The log function is introduced to reduce the effects of large S and enlarge the effects of small S, so that we can analyze a large scope of S conveniently. There is no particular reason to choose the city-block distance, which can be replaced by others, e.g., Euclidean norm. So does for the base 2 for the logarithm. The entire function preserves the monotonic property of S.

TABLE III
PERFORMANCE OF THE PROPOSED IQA FRAMEWORK WITH DIFFERENT MGA TRANSFORMS ON THE LIVE DATABASE

	JPEG						JPEG2000					
Metric	α	CC	ROCC	OR	MAE	RMSE	α	CC	ROCC	OR	MAE	RMSE
Wavelet	1	0.9587	0.9391	0.0914	5.1012	6.5812	6	0.9487	0.9333	0.0473	5.3486	6.8453
Curvelet	3	0.9519	0.9288	0.1029	5.4161	6.8502	1	0.9362	0.9170	0.0888	5.9377	7.5636
Bandelet	2	0.9605	0.9382	0.0571	4.8923	6.3294	5	0.9496	0.9330	0.0710	5.1833	6.7656
Contourlet	3	0.9493	0.9309	0.1029	5.3177	6.6941	2	0.9451	0.9273	0.0710	5.5818	6.9616
WBCT	3	0.9728	0.9527	0.0457	4.1162	5.3750	6	0.9565	0.9390	0.0414	4.9891	6.4718
HWD1	2	0.9704	0.9526	0.0514	4.3305	5.5646	5	0.9540	0.9333	0.0493	5.2857	6.7972
HWD2	3	0.9728	0.9543	0.0400	4.1135	5.3206	5	0.9540	0.9362	0.0473	5.1357	6.6312
	WN						Gblur					
Metric	α	CC	ROCC	OR	MAE	RMSE	α	CC	ROCC	OR	MAE	RMSE
Wavelet	6	0.9115	0.9251	--	5.6576	6.5671	1	0.9295	0.9324	--	4.5961	5.7977
Curvelet	5	0.9582	0.9585	--	3.6607	4.5698	3	0.9189	0.9131	--	4.7284	6.2007
Bandelet	6	0.9057	0.9156	--	5.6885	6.7699	5	0.8290	0.8313	--	6.8386	8.7941
Contourlet	5	0.9155	0.9261	--	5.4488	6.4224	2	0.8848	0.9004	--	5.5172	7.3275
WBCT	6	0.9142	0.9261	--	5.5343	6.4703	1	0.8581	0.8527	--	6.6345	8.0747
HWD1	4	0.9218	0.9365	--	5.2141	6.1894	6	0.6798	0.6964	--	8.942	11.5315
HWD2	4	0.9193	0.9321	--	5.3137	6.2854	4	0.8049	0.8282	--	7.0912	9.3295
	FF											
Metric	α	CC	ROCC	OR	MAE	RMSE						
Wavelet	3	0.9129	0.9137	--	5.0224	6.7130						
Curvelet	1	0.9372	0.9378	--	4.3326	5.7380						
Bandelet	5	0.9379	0.9412	--	4.5427	5.7079						
Contourlet	2	0.9263	0.9376	--	4.6163	6.1959						
WBCT	8	0.9479	0.9504	--	4.0593	5.2404						
HWD1	2	0.9215	0.9168	--	4.7260	6.3881						
HWD2	7	0.9336	0.9386	--	4.4261	5.8946						

IV. PERFORMANCE EVALUATION

In this section, we compare the performance of the proposed framework based on different MGA transforms with standard IQA methods, i.e., PSNR, WNISM, and the *mean structural similarity index* (MSSIM), based on the following experiments: the consistency experiment, cross-image and cross-distortion experiment, sensitivity experiment, and the rationality experiment. At the beginning of this section, we first brief the image database for evaluation.

The *laboratory for image & video engineering* (LIVE) database [28] has been recognized as the standard database for IQA measures performance evaluation. This database contains 29 high-resolution 24 bits/pixel RGB color images and 175 corresponding JPEG and 169 JPEG2000 compressed images, as well as 145 *white noisy* (WN), 145 *Gaussian blurred* (Gblur), and 145 *fast-fading* (FF) Rayleigh channel noisy images at a range of quality levels.

To remove nonlinearity, which is raised by the subjective rating process, and to facilitate empirical comparison of different IQA methods fairly, the *video quality experts group* (VQEG) [34] suggested first applying the variance-weighted regression analysis to provide a nonlinear mapping between the objective and subjective *mean opinion score* (MOS), and then using three metrics as evaluation criteria [34]. The first metric is the Pearson linear *correlation coefficient* (CC) between the objective and MOS after the variance-weighted regression analysis. This metric provides an evaluation of prediction accuracy. The second metric is the *Spearman rank-order correlation coefficient* (ROCC) between the objective and subjective scores. This metric measures the prediction monotonicity. The third metric is the *outlier ratio* (OR), which is the percentage of the number of predictions outside the range of twice of the standard deviation of the predictions after the nonlinear mapping. It is a measure of the prediction consistency. In addition, we compute the *mean absolute error* (MAE), and the *root mean square error* (RMSE) of fitting procedure after the nonlinear mapping.

In the following parts, we compare the performance of different IQA methods based on the aforementioned image database and metrics.

A. Consistency Experiment

In this subsection, we compare the performance of the proposed IQA framework with PSNR, WNISM, and the well known full reference assessment metric, MSSIM. The evaluation results for all IQA methods being compared are given as benchmark in Table II. Table III shows the evaluation results of the proposed IQA framework with different MGA transforms, e.g., wavelet, curvelet, bandelet, contourlet, WBCT, HWD1 and HWD2. The OR results of WN, Gblur and FF are not available because the corresponding subjective scores are not provided in [28] either. Fig. 7 presents the scatter plots of MOS versus the predicted score by objective metrics after the nonlinear mapping.

For JPEG and JPEG2000: Table III demonstrates that the results of all these MGA transforms using the proposed framework yield a higher effectiveness to IQA. Comparing Table III with Table II, it is noticed that the metric obtained by the proposed framework gives much better performance than WNISM in terms of better prediction accuracy (higher CC), better prediction monotonicity (higher ROCC) and better

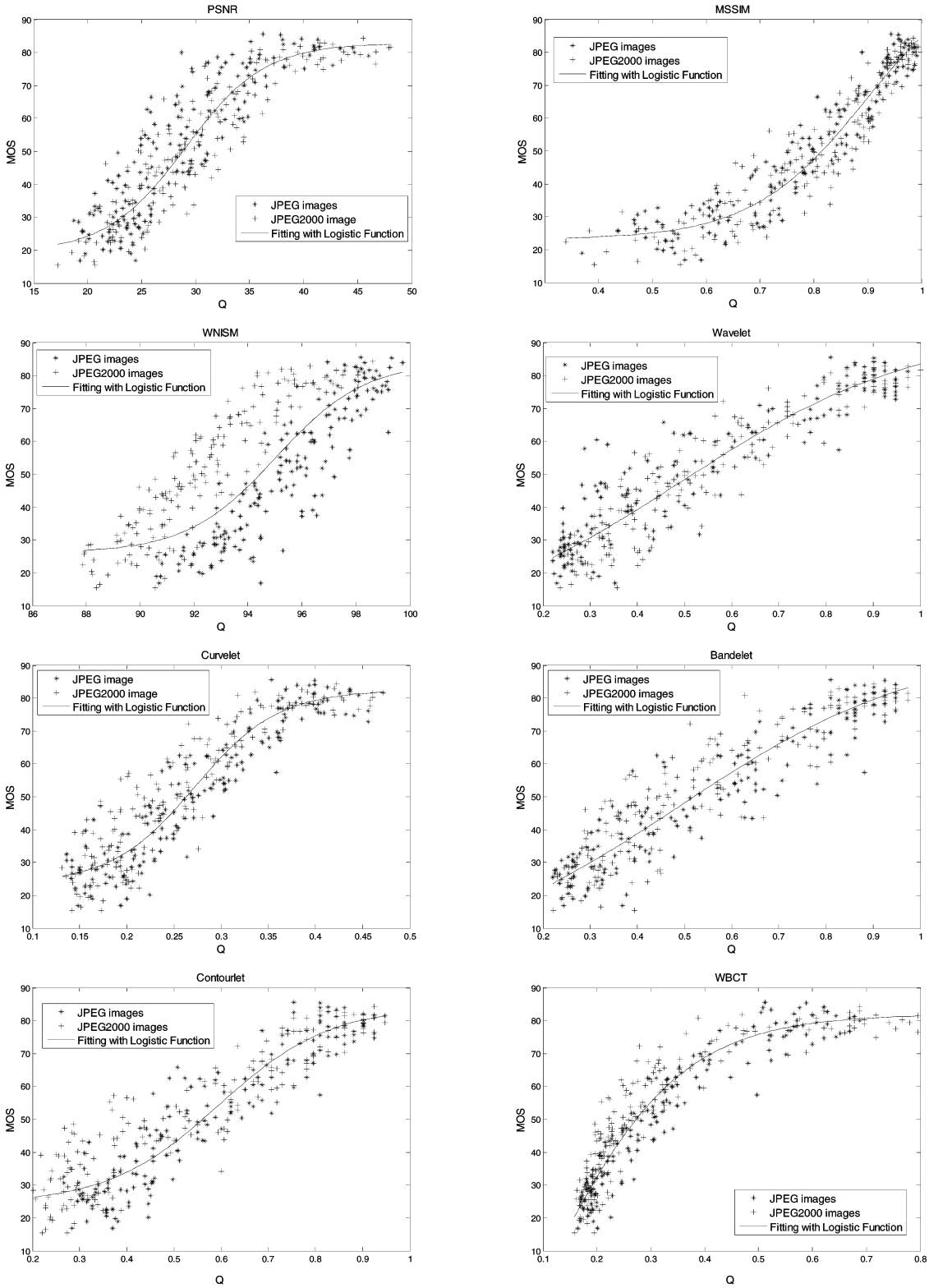


Fig. 8. Scatter plots of MOS versus different metrics using database of JPEG and JPEG2000 images.

prediction consistency (lower OR, MAE, and RMSE). Especially, when WBCT and HWD are employed in the proposed framework, better performance is achieved than that of MSSIM index. The only tuning parameter α in the proposed framework responses to different distortions.

For WN distortion: PSNR outperforms other methods, because PSNR directly measures the difference between the original image and the corresponding WN distorted image. The performance of the proposed framework is comparable to that of MSSIM and better than that of WNISM.

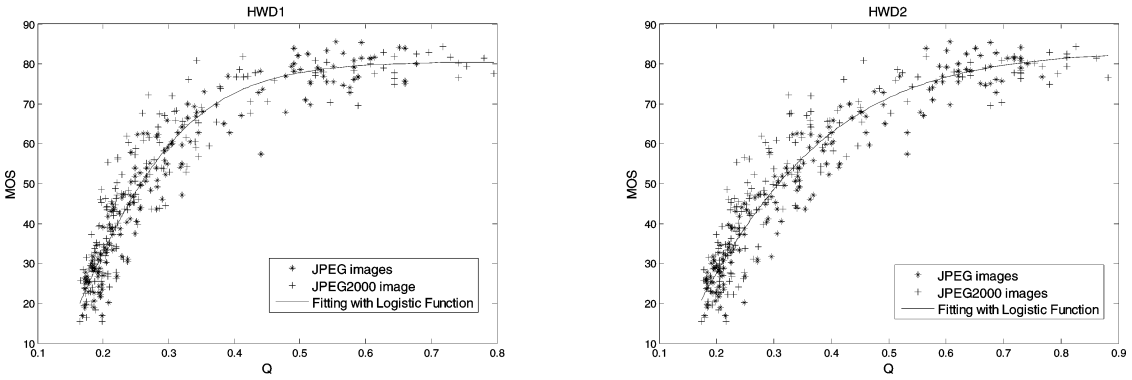


Fig. 8. (*Continued*) Scatter plots of MOS versus different metrics using database of JPEG and JPEG2000 images.

For Gblur Distortion: Gaussian blur is blurring an image by using a Gaussian function and, thus details in an image (corresponding to the high frequency part of the image) are removed. Because Gaussian function is isotropical, transforms of wavelet, curvelet, and contourlet using the proposed framework perform well in comparing with other methods.

For FF Distortion: Fast-fading is the distortion that a signal experiences over a propagation media. FF produces block artifacts and blurs the original image. Experimental results in Tables II and III, as well as Fig. 7, show that the proposed framework outperforms WNISM and is comparable with MSSIM.

It is universally agreed in the community that the key stage of IQA is how to represent images effectively and efficiently, so it is necessary to investigate kinds of transforms. Wavelet is localized in spatial and frequency domain and can extract local information with high efficiency, so it optimally approximates a target function with 1-D singularity. However, wavelet cannot achieve the sparse representation of edges even it captures the point singularity effectively. In order to represent edges in an image sparsely, curvelet transform has been proposed. The basis function of curvelet can be expressed by rectangle, and it is nearly optimal to fit the image with curve singularity. However, because curvelet is defined in the continuous domain, there is a huge redundancy when it is transformed to discrete domain. Bandelet can trace the geometric regular direction adaptively by taking edge information into account. However, its application is limited because it needs prior knowledge of the geometric features in images. Inheriting the anisotropy of each scale from curvelet, contourlet analyzes the scale and orientation respectively and reduces the redundancy by approximating images with line-segment similar basis ultimately. To further reduce the redundancy, WBCT and HWD are proposed to represent images sparsely and precisely. Redundancy is harmful for IQA. If an image is represented by a large number of redundant components, little visual changes will affect the quality of the image slightly. Both transforms are nonredundant, multiscale and multiorientation for image decomposition. Table III shows the proposed IQA framework with WBCT and HWD works much better than previous standards for JPEG and JPEG2000 images, wavelet transform is optimal to Gblur distortion, curvelet transform and PSNR perform best on WN distortion, and WBCT outperforms the other methods for FF distortion.

TABLE IV
PERFORMANCE OF PSNR, MSSIM, AND WNISM ON THE LIVE DATABASE

Metric	Type	CC	ROCC	OR	MAE	RMSE
PSNR	FR	0.9273	0.8957	0.1366	6.846	8.878
MSSIM	FR	0.9594	0.9429	0.0552	5.2211	6.5495
WNISM	RR	0.7968	0.7651	0.3140	10.614	13.096

TABLE V
PERFORMANCE OF THE PROPOSED FRAMEWORK USING MGA TRANSFORMS

Transform	α	CC	ROCC	OR	MAE	RMSE
Wavelet	4	0.9426	0.9113	0.098	5.82	7.696
Curvelet	3	0.9083	0.9104	0.1192	6.0884	7.8426
Bandelet	5	0.9531	0.9342	0.0669	5.2014	6.7907
Contourlet	5	0.9353	0.8999	0.107	6.2535	8.1106
WBCT	3	0.9599	0.9376	0.061	5.0068	6.6216
HWD1	2	0.96	0.9397	0.061	4.9836	6.5327
HWD2	3	0.9624	0.9418	0.055	4.8445	6.3657

B. Cross Image and Cross-Distortion Experiment

Many IQA methods have been shown to be consistent when applied to distorted images created from the same reference image by using the same type of distortions. However, the effectiveness of these methods degrades significantly when applied to a set of images originating from different reference images including a variety of different types of distortions. To further evaluate the effectiveness of the proposed framework, apart from previous experiments, we conducted cross-image and cross-distortion tests, which are critical in evaluating the effectiveness of an IQA method.

In this experiment, we compare the cross-distortion and cross-image performances of different quality assessment methods on the LIVE database. Scatter plots of MOS versus different metrics on the database of JPEG and JPEG2000 images are shown in Fig. 8. The evaluation results for all methods are given in Table IV, and the evaluation results for the proposed framework with different MGA transforms are given in Table V.

According to Tables IV and V, the proposed framework with different MGA transforms performs better than the standard RR IQA method, e.g., WNISM. The proposed IQA framework based on WBCT and HWD performs even better than the standard FR IQA method, e.g., MSSIM.

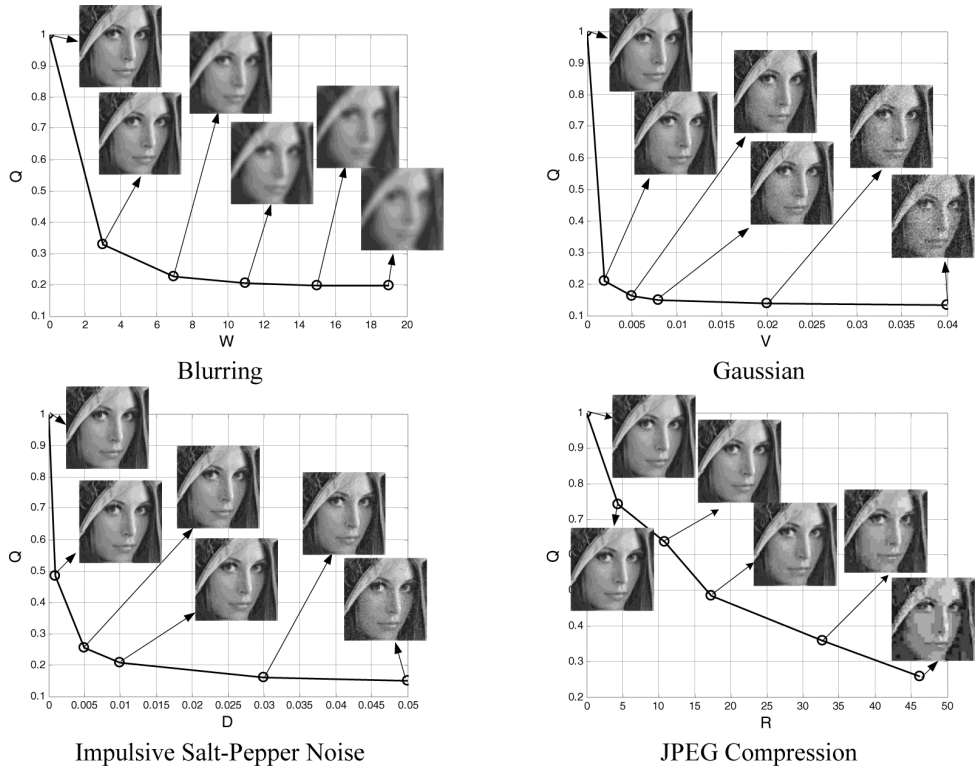


Fig.9. Trend plots of Lena with different types of distortion using contourlet of the proposed framework (all of the images are 8bits/pixel; cropped from 512 × 512 to 128 × 128 for visibility).

C. Rationality Experiment

To verify the rationality of the proposed framework, we choose contourlet and WBCT to test the Lena image with different distortions, which are blurring (with smoothing window of $W \times W$), additive Gaussian noise $(\text{mean} = 0, \text{variance} = V)$, impulsive Salt-Pepper noise $(\text{density} = D)$, and JPEG compression $(\text{compression rata} = R)$.

Figs. 9 and 10 (all of the images are 8 bits/pixel; cropped from 512×512 to 128×128 for visibility) show the Lena image with different types of distortions and the metrics prediction trend to the corresponding image, respectively. It is found that the proposed framework prediction trend to image drop with the increasing intensity of different types of image distortions. It is consistent well with the tendency of the decreasing image quality in fact. So the results demonstrate the rationality of the proposed framework.

For JPEG compression, we find that Q for IQA drops with the increasing intensity of different distortions, which is consistent with the tendency of the decreasing image quality. That is the proposed IQA framework works well for the JPEG distortion. The coding scheme of JPEG (JPEG2000) is based on the discrete cosine transform (discrete wavelet transform). In JPEG (JPEG2000) coding stage, the lowpass subband is compressed in a high compression rate and highpass subbands are compressed in a low compression rate to achieve a good compression rate for the whole image while preserving the visual quality of the image. This procedure is similar to our IQA framework, i.e., we do not consider information in the lowpass subband because most information in the lowpass subband is preserved; and the output value Q is obtained from highpass subbands only because low compression rates are utilized on them and the quality of the reconstructed image is strongly relevant to highpass subbands (the defined JND threshold). Therefore, the proposed scheme adapts well for JPEG and JPEG2000 distortions.

For blurring, additive Gaussian noise distortion and impulsive salt-pepper noise, Q for IQA drops sharply initially and then slowly because MGA transforms cannot explicitly extract effective information from images with these distortions. However, based on these figures, Q can still reflect the quality of distorted images with blurring, additive Gaussian noise distortion and impulsive salt-pepper noise, although the performance is not good.

D. Sensitivity Test Experiment of the Proposed Framework

Currently, PSNR is one of the most popular objective quality metrics for images. However, it does not correlate well with the perceived quality measurement as shown in Table VI. Fig. 11 shows three degraded "Lena" images with different distortions but with almost identical PSNR. Another popular objective quality metric, MSSIM is also reported in Table VI, which shows MSSIM cannot perform well for the contrast stretching distortion, although the perceived quality of the contrast stretching distorted image is worse than that of mean shift transformed image. As shown in Fig. 11, the proposed framework with different MGA transforms can distinguish both the contrast stretching and JPEG compression distortions well. It is worth emphasizing that Q is insensitive to the changes of intensity, i.e., changes of grey values at image level cannot affect the image quality significantly when Q is 1. This is

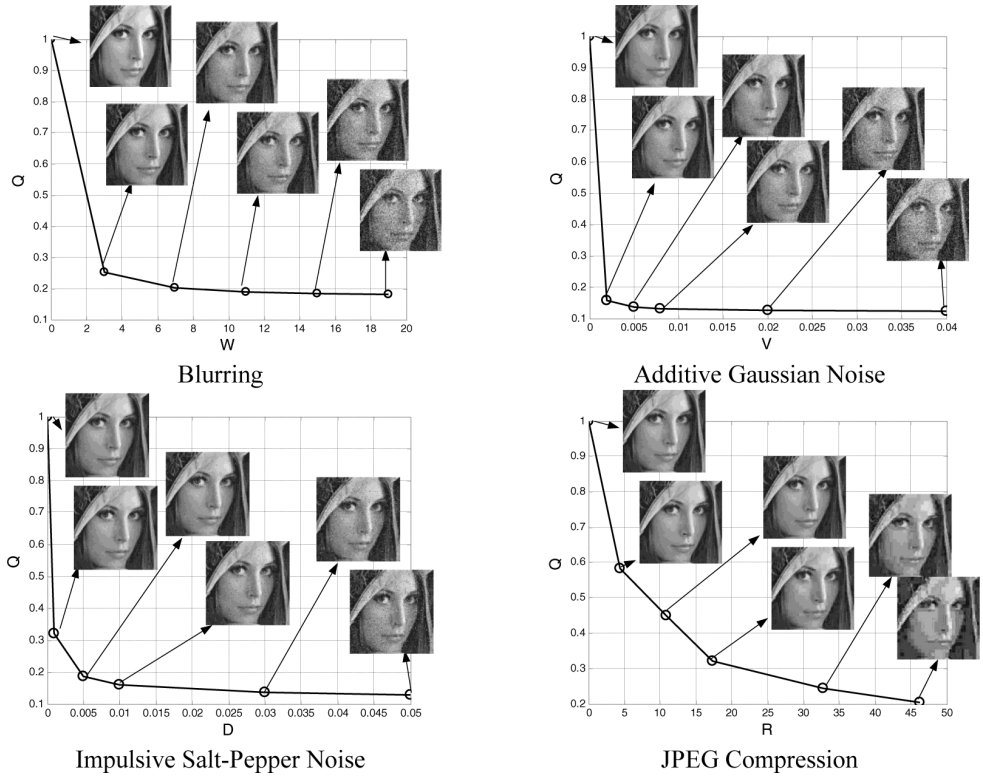


Fig. 10. Trend plots of Lena with different types of distortion using WBCT of the proposed framework.

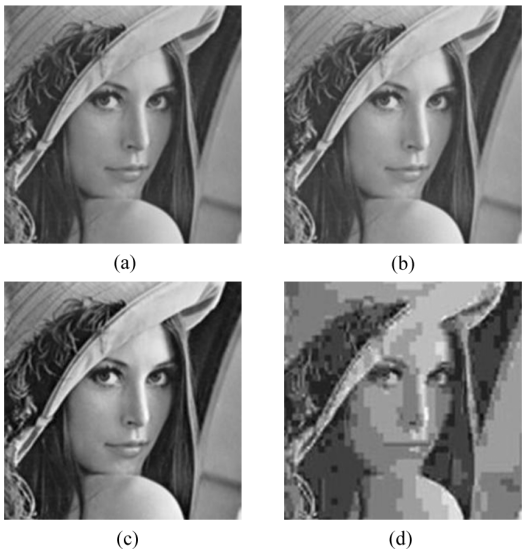


Fig. 11. Lena image with the same PSNR but different perceived quality. (a) The reference image; (b) mean shift; (c) contrast stretching; (d) JPEG compression.

TABLE VI
VALUE OF DIFFERENT METRICS FOR IMAGES IN FIG. 11

Metric	(b)	(c)	(d)
PSNR	24.8022	24.8013	24.8041
MSSIM	0.9895	0.9458	0.6709
Wavelet	1.0000	0.3208	0.2006
Curvelet	1.0000	0.1510	0.1369
Bandelet	1.0000	0.2947	0.2125
Contourlet	1.0000	0.2960	0.2423
WBCT	1.0000	0.2345	0.1929
HWD1	1.0000	0.2740	0.2166
HWD2	1.0000	0.2704	0.2094

because the lowpass section of the image is not considered in the proposed IQA framework.

E. Low Data Rate

This section shows variations of the performance measures (i.e., CC, ROCC, MAE, and RMSE) with respect to the bit rate of the ancillary channel, over which the RR information is sent. Fig. 12 shows that the contourlet, WBCT, HWD1, HWD2, wavelet, bandelet, and curvelet transforms using the proposed framework for RR IQA have low data rates for representing features of the reference image with different distortions. The number of features, sent over the ancillary channel for RR IQA, is mainly proportional to the selected subbands. For example, if the contourlet transform is utilized to decompose an image into three levels or sixteen highpass subbands associated with one lowpass subband, as shown in Fig. 4, we need 16 features to represent the 16 highpass subbands for RR IQA under the proposed framework. With the increasing number of features, we can usually obtain a better IQA performance measure (e.g., higher CC) for different distortions. However, the IQA performance measures (e.g., CC) will not increase with the increasing number of features for the WN distortion, because images will be damaged by this distortion.

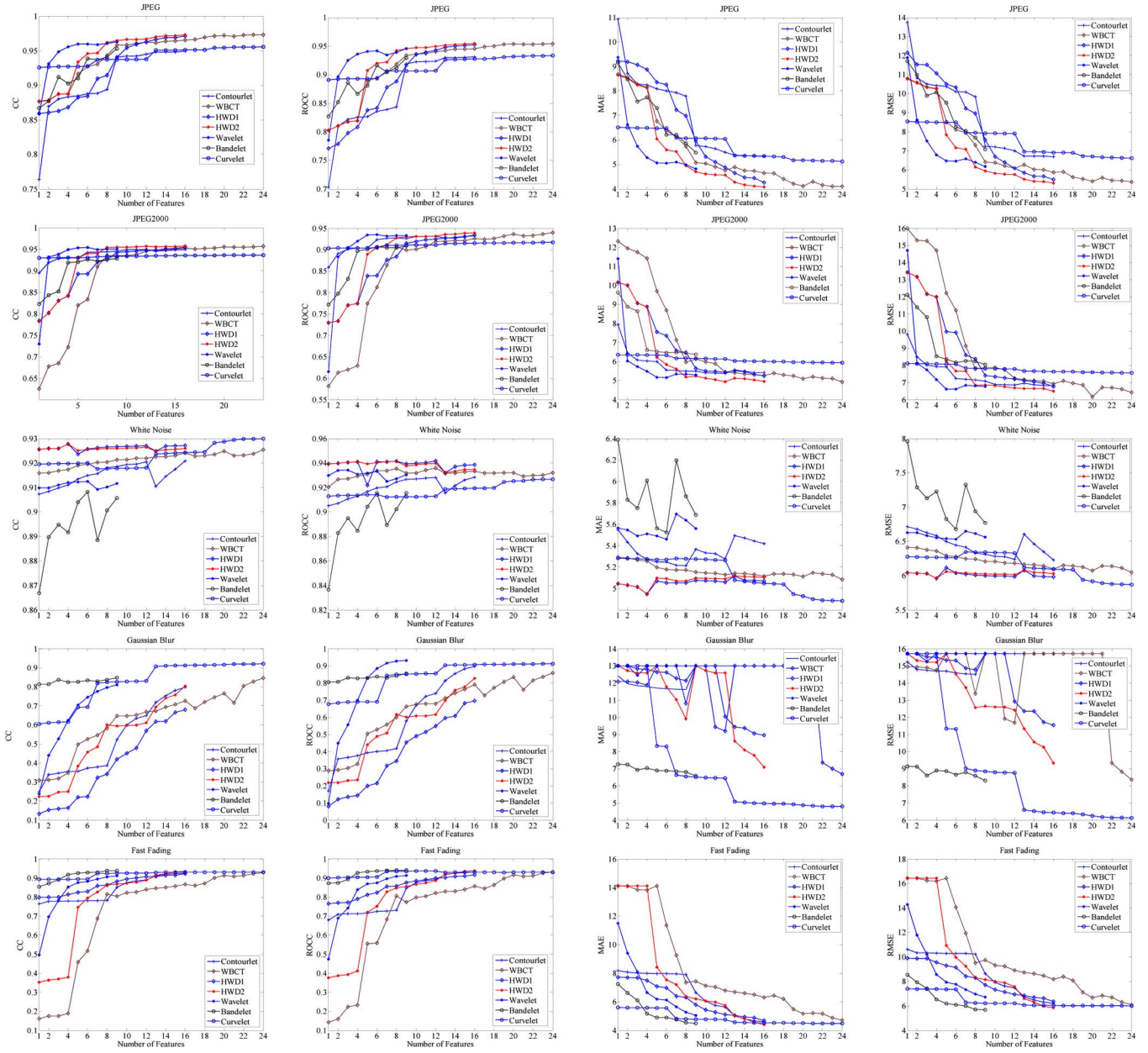


Fig. 12. Variations of different performance measures versus the bit rate of the ancillary channel.

F. Summary

In this paper, we present a novel IQA framework by incorporating merits from MGA, CSF, and JND. A large number of experimental results demonstrate this framework outperforms conventional IQA standards, i.e., PSNR, MSSIM, and WNISM, for JPEG and JPEG2000 images as well as WN, Gblur, and FF distorted images. Advantages of this framework are as follows.

1) **General purpose:** A number of different transforms, e.g., wavelet, curvelet, bandelet, contourlet, WBCT and HWD, for image decomposition can be applied in the proposed framework for IQA. All these transforms can work well for different image distortions, and WBCT and HWD perform much better than the others, especially for JPEG and JPEG2000 images.
2) **Sound effectiveness:** The objective assessment of the proposed framework performs consistently with the subjective perception, and can evaluate visual quality of JPEG and JPEG2000 image effectively. Particularly, by applying WBCT and HWD for image decomposition, the framework has better effectiveness not only than the reduced-reference method, WNSIM, but also than the full reference method as MSSIM, in terms of CC, ROCC, OR, MAE, and RMSE. In addition, different transforms have different advantages in different distortions. For example, wavelet transform is optimal to Gblur distortion, curvelet transform and PSNR perform best on WN distortion, and WBCT outperforms the other methods for FF distortion.
3) **Low data rate:** The proposed framework has a relatively low data rate for representing features of the reference image, i.e., a relatively low bits are used to represent the

image features (e.g., for wavelet-based decomposition, there are only 10 features/image utilized for representation; contourlet-based decomposition corresponds to 17 features/image; WBCT-based decomposition corresponds to 25 features/image; and HWD-based decomposition corresponds to 17 features/image).

4) **Flexible to different distortions:** The proposed framework contains only one free parameter α, which corresponds to different distortions, for calculating the JND threshold. In an identical distortion, α is identical to different images. Empirical studies show that by tuning α to different distortions, we can achieve better performances for IQA. In addition, if we set α as a constant for different distortions, the IQA performance is still acceptable. In our experiments, we tune α from $\{1, 2, 3, \cdots, 9\}$ and choose the one corresponding to the best performance as shown in experiments.

V. Conclusion

In this paper, a reduced-reference image quality assessment framework is proposed by incorporating merits of *multiscale geometry analysis* (MGA), *contrast sensitivity function* (CSF), and the Weber's law of *just noticeable difference* (JND). In comparing with existing image quality assessment approaches, the proposed one has strong links with the *human visual system* (HVS): sparse image representation is utilized to mimic the multichannel structure of HVS, CSF is utilized to balance magnitude of coefficients obtained by MGA to mimic nonlinearities of HVS, and JND is utilized to produce a noticeable variation in sensory experience. In this framework, images are represented by normalized histograms, which correspond to visually sensitive coefficients. The quality of a distorted image is measured by comparing the normalized histogram of the distorted image and that of the reference image. Thorough empirical studies show that the novel framework with a suitable image decomposition method performs better than conventional standard reduced-reference image quality assessment method, i.e., *wavelet-domain natural image statistics model* (WNISM), and even better than the standard full-reference IQA model, i.e., the *mean structural similarity index* (MSSIM).

Although the proposed framework has good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images, there are still some issues deserve to be further investigated in the future. Since reference methods require full information of the reference image or limited information of the reference image, which could be a serious impediment for many applications. It is essential to develop *no-reference* (NR) image quality metrics that blindly estimate the quality of images. Recently, tensor-based approaches [30]–[32] have been demonstrated to be an effective way for image representation in classification problems, so it is valuable to introduce them for image quality. With the explosive growth of video records on Internet and in personal databases as well as the rapid progress of the computer power, *video quality assessment* (VQA) also becomes one of the most active research fields. In the future, we will further the proposed MGA-based framework to VQA by integrating the merits of the spatial temporal information [37] and the human perception [36].

Acknowledgment

The authors would like to thank the Associate Editor Prof. L. Younes and the anonymous reviewers for their constructive comments on this paper.

References

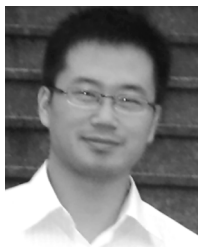
[1] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–213, 2002.
[2] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
[3] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, Apr. 1992.
[4] B. Chitprasert and K. R. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1040–1044, Jul. 1990.
[5] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commmun. Pure Appl. Math.*, vol. 45, no. 5, pp. 485–560, 1992.
[6] E. J. Candès and D. L. Donoho, "Curvelets-a surprising effective nonadaptive representation for objects with edges," in *Curves and Surfaces*. Nashville, TN: Vanderbilt Univ. Press, 2000, pp. 105–120.
[7] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316–1327, May 2006.
[8] S. Daly, , A. B. Watson, Ed., "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993, pp. 179–206.
[9] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 636–650, Apr. 2000.
[10] M. N. Do, "Directional Multiresolution Image Representations," Ph.D. dissertation, École Polytechnique Fédéral de Lausanne, France, 2001.
[11] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[12] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.
[13] R. Eslami and H. Radha, "Wavelet-based contourlet transform and its application to image coding," in *Proc. IEEE Int. Conf. Image Processing*, Piscateway, NJ, 2004, pp. 3189–3192.
[14] R. Eslami and H. Radha, "New image transforms using hybrid wavelets and directional filter banks: Analysis and design," in *Proc. IEEE Int. Conf. Image Processing*, Genova, Italy, 2005, pp. 11–14.
[15] Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment Phase II VQEG, Aug. 2003 [Online]. Available: http://www.vqeg.org/
[16] X. Gao, W. Lu, X. Li, and D. Tao, "Wavelet-based contourlet in quality evaluation of digital images," *Neurocomputing*, vol. 72, no. 12, pp. 378–385, 2008.
[17] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image Process.*, vol. 4, no. 6, pp. 713–724, Jun. 1995.
[18] J. Lubin, "A visual discrimination mode for image system design and evaluation," in *Visual Models for Target Detection and Recognition*, E. Peli, Ed. Singapore: World Scientific, 1995, pp. 207–220.
[19] W. Lu, X. Gao, D. Tao, and X. Li, "A wavelet-based Image Quality Assessment Method," *Int. J. Wavelets, Multires., Inf., Process.*, vol. 6, no. 4, pp. 541–551, 2008.
[20] X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Processing*, New York, 2002, vol. 1, pp. 449–452.
[21] Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R Rec. BT. 500-11.
[22] J. L. Mannos and D. J. Sakrison, "The effect of visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 525–536, Feb. 1974.

[23] S. Mallat, "A theory for multiresolution decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[24] M. Miloslavski and Y.-S. Ho, "Zerotree wavelet image coding based on the human visual system model," in *Proc. IEEE Asia-Pacific Conf. Circuits and Systems*, 1998, pp. 57–60.
[25] M. J. Nadenau, J. Reichel, and M. Kunt, "Wavelet-based color image compression: Exploiting the contrast sensitivity," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 58–70, Jan. 2003.
[26] E. Le Pennec and S. G. Mallat, "Image compression with geometrical wavelets," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, 2000, pp. 661–664.
[27] J. K. Romberg, "Multiscale Geometric Image Processing," Ph.D. dissertation, Rice Univ., Houston, TX, 2003.
[28] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, (2003) LIVE Image Quality Assessment Database [Online]. Available: http://live.ece.utexas.edu/research/quality
[29] 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.
[30] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[31] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Human carrying status in visual surveillance," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2006, pp. 1670–1677.
[32] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Supervised tensor learning," in *Knowledge and Information Systems*. New York: Springer, 2007.
[33] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, 1994, vol. 2, pp. 982–986.
[34] Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment [Online]. Available: http://www.vqeg.org/ 2000, VQEG
[35] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," *SPIE Human Vision, Visual Processing, and Digital Display IV*, 1993.
[36] B. A. Wandell, *Foundations of Vision*, 1st ed. Sunderland, MA: Sinauer, 1995.
[37] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," presented at the SPIE Int. Symp. Voice, Video, and Data Communication, Boston, MA, Sep. 1999.
[38] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York: Morgan & Claypool, 2006.

Xinbo Gao (M'02–SM'07) received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, China, in 1994, 1997, and 1999, respectively.

He joined the Department of Electric Engineering, Xidian University, as a Lecturer in 1999 and is currently Professor, Director of the VIPS Lab, and Director of International Office, Xidian University. His research interests are computational intelligence, machine learning, information processing and analysis, pattern recognition, and artificial intelligence. In these areas, he has published four books and around 100 technical articles in refereed journals and proceedings including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS, etc.

Dr. Gao is a member of the IEEE Xi'an Section Executive Committee and the Membership Development Committee Chair; a senior member of IET, and Vice Chairman of IET Xi'an Network; a senior member of Chinese Institute of Electronics (CIE), senior member of China Computer Federation (CCF), and an Executive member of China Society of Image and Graphics (CSIG). He has won various awards/honors including the New Century Excellent Talents in University of China (2004), the Award of the Title Pacemaker of Excellent Young Teacher of Shaanxi Province (2005), and the Young Teacher Award of High School by the Fok Ying Tung Education Foundation (2006). He is on the editorial boards of journals including *EURASIP Signal Processing* (Elsevier), *Journal of Test and Measurement Technology, Journal of Data Acquisition and Processing*, etc. He has served as general chair/co-chair or program committee chair/co-chair or PC member for around 30 major international conferences.



Wen Lu received the B.S. and M.S. degrees in electrical engineering from Xidian University, China, in 2002 and 2006, respectively. He is currently pursuing Ph.D. degree in pattern recognition and intelligent systems at Xidian University.

His research interests include image and video understanding, visual quality assessment, and computational vision.

Dacheng Tao (M'07) received the B.Eng. degree from the University of Science and Technology of China (USTC), the M.Phil. degree from the Chinese University of Hong Kong (CUHK), and the Ph.D. degree from the University of London, London, U.K.

Currently, he holds a Nanyang titled academic position with the School of Computer Engineering, Nanyang Technological University. He is a Visiting Professor at Xi Dian University, a Guest Professor at Wu Han University, and a Visiting Research Fellow at the University of London. His research is mainly on applying statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and visual surveillance. He has published around 100 scientific papers including IEEE TPAMI, TKDE, TIP, CVPR, ECCV, ICDM, ACM Multimedia, KDD etc.,

Dr. Tao has received best paper runner up awards. Previously, he gained several Meritorious Awards from the International Interdisciplinary Contest in Modeling, which is the highest level mathematical modeling contest in the world, organized by COMAP. He is an Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *Neurocomputing* (Elsevier), and the Official Journal of the International Association for Statistical Computing—*Computational Statistics & Data Analysis* (Elsevier). He is an editorial board member of *Advances in Multimedia* (Hindawi). He has authored/edited six books and eight special issues, including CVIU, PR, PRL, SP, and *Neurocomputing*. He has (co-)chaired for special sessions, invited sessions, workshops, and conferences. He has served with more than 50 major international conferences including CVPR, ICCV, ECCV, ICDM, KDD, and *Multimedia*, and more than 15 top international journals including TPAMI, TKDE, TOIS, TIP, TCSVT, TMM, TIFS, TSMC-B, *Computer Vision and Image Understanding* (CVIU), and *Information Science*.

Xuelong Li (M'02–SM'07) is the Reader in Cognitive Computing at the University of London, London, U.K., a Visiting Professor at Tianjin University, and a Guest Professor at the University of Science and Technology of China. His research focuses on cognitive computing, multimedia signal processing and management, and pattern recognition. His research activities are sponsored in part by EPSRC, the British Council, and the Royal Society. He has authored over 150 publications in over 90 journals (nearly 40 IEEE transactions).

His papers have received Best Paper Awards and finalists. He is an author of a Springer monograph, an editor of three scientific books, and a guest editor of ten special issues. He is an associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS. He was/is also an editor of ten other international journals. He has served as a chair for conferences nearly 30 times and as a conference program committee member more than 120 times. He is an academic committee member of the China Society of Image and Graphics. He chairs the Technical Committee (TC) on Cognitive Computing of IEEE SMCS. He is also a member three other SMCS TCs, a Chapters Coordinator of IEEE SMCS, a member of the IEEE CIS Continuing Education Sub-Committee, and an elected member of IEEE SPS TC on Machine Learning for Signal Processing.