<table>
<thead>
<tr>
<th>Metadata of the chapter that will be visualized online</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Book Title</strong></td>
</tr>
<tr>
<td><strong>Copyright Year</strong></td>
</tr>
<tr>
<td><strong>Copyright Holder</strong></td>
</tr>
<tr>
<td><strong>Chapter Title</strong></td>
</tr>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>Family Name</td>
</tr>
<tr>
<td>Particle</td>
</tr>
<tr>
<td>Division</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>Family Name</td>
</tr>
<tr>
<td>Particle</td>
</tr>
<tr>
<td>Division</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>Family Name</td>
</tr>
<tr>
<td>Particle</td>
</tr>
<tr>
<td>Division</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td><strong>Author</strong></td>
</tr>
<tr>
<td>Family Name</td>
</tr>
<tr>
<td>Particle</td>
</tr>
<tr>
<td>Division</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Email</td>
</tr>
</tbody>
</table>
11 Image Quality Assessment: A Multiscale Geometric Analysis-Based Framework and Examples

Xinbo Gao¹ · Wen Lu² · Dacheng Tao³ · Xuelong Li⁴
¹Video and Image Processing System Lab, School of Electronic Engineering, Xidian University, China
xbgao@ieee.org
²Video and Image Processing System Lab, School of Electronic Engineering, Xidian University, China
luwen@mail.xidian.edu.cn
³School of Computer Engineering, Nanyang Technological University, Singapore
dacheng.tao@gmail.com
⁴State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, Shaanxi, China
xuelong_li@ieee.org
xuelong_li@opt.ac.cn
xuelongli@gmail.com

1 Introduction ................................................................. 2

2 Multiscale Geometric Analysis .................................................. 7

3 Multiscale Geometric Analysis for Image Quality Assessment ...................... 9

4 Performance Evaluation .......................................................... 15

5 Conclusion ........................................................................ 21

© Springer-Verlag Berlin Heidelberg 2010
Abstract

This chapter is about objective image quality assessment (IQA), which has been recognized as an effective and efficient way to predict the visual quality of distorted images. Basically, IQA has three different dependent degrees on original images, namely, full-reference (FR), no-reference (NR), and reduced-reference (RR). To begin with, the fundamentals of IQA are introduced and a broad treatment of the current state-of-the-art IQA are also given. A novel framework for IQA is focused upon to mimic the human visual system (HVS) by incorporating the merits from multiscale geometric analysis (MGA), contrast sensitivity function (CSF), and Weber’s law of just noticeable difference (JND). Thorough empirical studies were carried out upon the laboratory for image and video engineering (LIVE) database against subjective mean opinion score (MOS) and demonstrate that the proposed framework has good consistency with subjective perception values and the objective assessment results well reflect the visual quality of images.

1 Introduction

With the rapid development of information technology, digital image as a media for representing and communicating has witnessed a tremendous growth. A huge number of processing methods have been proposed to treat images for different purposes. The performance of these methods highly depends on the quality of images after processing. Therefore, how to evaluate the image quality has become a burning question in recent years. The problems of image quality assessment (IQA) (Wang and Bovik 2006) occur in many applications, such as image compression, enhancement, communication, storage and watermarking, etc. In the process of image compression, lossy compression techniques may introduce artificial block, blurring and ringing effects, which can lead to image degradation. In poor transmission channels, transmission errors or data dropping may happen, which can lead to imperfect quality and distortion of the received video data. The past 5 years have demonstrated and witnessed the Tremendous and imminent demands for IQA methods have occurred in recent years, at least in the following three ways: (1) They can be used to monitor image quality for quality control systems; (2) They can be employed to benchmark image processing systems and algorithms; (3) They can also be embedded into image processing systems to optimize algorithms and parameter settings.

Existing IQA metrics can be categorized into subjective (Rec. ITU-R) and objective methods (Wang and Bovik 2006). The former is based on quality, which is assessed by human observers; the latter depends on quantified parameters, which are obtained from the model to measure the image quality.

Because human observers are the ultimate receivers of the visual information contained in an image, a subjective method whose results are directly given by human observers is probably the best way to assess the quality of an image. The subjective method is one in which the observers are asked to evaluate the picture quality of sequences using a continuous grading scale and to give one score for each sequence. A number of different subjective methods are represented by ITU-R Recommendation BT.500 (Rec. ITU-R). The subjective quality measurement has been used for many years as the most reliable form of quality measurement. However, subjective experiment requires human viewers working over a long period and repeated experiments are needed for many image objects. Furthermore, it is expensive, time...
consuming, and cannot be easily and routinely performed for many scenarios, for example, real-time systems. Moreover, there currently is no precise mathematical model for subjective assessment. As a consequence, there is a need for an objective quality metric that accurately matches the subjective quality and can be easily implemented in various image systems, leading to the emergence of objective IQA.

Objective IQA makes use of the variation of several original or distorted image characteristics, which is caused by degradation to represent the variation of image perceptual quality. Many objective quality metrics for predicting image distortions have been investigated. According to the availability of original image information, there is a general agreement (Wang and Bovik 2006) that objective quality metrics can be classified into three categories: full-reference (FR), no-reference (NR), and reduced-reference (RR).

To evaluate the quality of a distorted image, FR metrics, which have access to both whole original and reconstructed information, provide the most precise evaluation results compared with NR and RR. Generally, FR metrics can be divided into two steps: one tends to construct the errors between original and distorted images, then produce distortion maps; the other provides the global IQA by pooling the errors. Conventional FR IQA methods (Avcibas et al. 2002; Eskicioglu and Fisher 1995) calculate pixel-wise distances, for example, peak signal-to-noise ratio (PSNR) and mean square error (MSE), between the distorted image and the corresponding reference. These measurements are attractive because of their simplicity and good performance when images with the same type of degradation are compared. However, when images with different types of degradations are compared, the results may not be consistent well with that of subjective IQA. Besides, images with various types of degradations but the same value could have very different subjective qualities. This is mainly because they are based on pixel-to-pixel difference calculations, which disregard image content and the characteristics of human visual perception.

Recently, a great deal of efforts has been made to develop visual models (Mannos and Sakrison 1974) that take advantage of the known characteristics of human visual system (HVS). The aim of HVS-based IQA is to evaluate how strong the error signal is perceived by the HVS, according to the characteristics of human visual error sensitivity. A number of IQA methods have been proposed to evaluate the perceptual quality. Two famous models, the Daly Visible Differences Predictor (VDP) and Sarnoff Visual Discrimination Model (VDM), were proposed by Daly (1993) and Lubin (1995), respectively. The Daly VDP receives distorted and corresponding reference as input and produces difference map as output, which predicts the probability of detection for dissimilarities throughout the whole images. If two images vary substantially, the probability of prediction will be one, and as the differences aggravate, the probability does not increase further. The Sarnoff VDM was designed for physiological plausibility as well as for speed and simplicity. While the Daly VDP is an example of a frequency domain visual model, the Sarnoff VDM operates in the spatial domain. In Karunasekera and Kingsbury (1995), these two excellent models have been comparatively evaluated. The Watson (1993) metric uses the DCT-based perceptual error measurement. The quantified errors for every coefficient in every block are scaled by the corresponding visual sensitivities of every DCT basis function in each block. Miyahara et al. (1998) reported a new methodology for the determination of an objective metric. This methodology is applied to obtain a picture quality scale (PQS) for the coding of achromatic images over the full range of image quality presented by the MOS. Damera-Venkata et al. (2000) modeled the degraded image as an reference image polluted by linear frequency distortion and additive noise contamination. Since the psycho-visual effects of frequency distortion and noise
contamination are independent, they decouple these two sources of degradation and measure their effect on the HVS. Instead of computing a residual image, they compute a model restored image by applying the restoration algorithm on original image, using the same parameters as those used while restoring the degraded image. A number of limitations of HVS-based IQA methods were discussed in Wang and Bovik (2006), because they must rely on several strong assumptions and generalizations.

Different from traditional HVS-based error-sensitivity IQA, structural similarity-based IQA (Wang and Bovik 2006) is based on the following philosophy: The main function of the HVS is to extract structural information from the viewing field, and the HVS is highly adapted for this purpose. Therefore, measurements of structural information loss can provide a good approximation to image perceived distortion. Wang et al. introduces a complementary framework for IQA based on the degradation of structural information and develops a structural similarity (SSIM) index to demonstrate its promise through a set of intuitive examples. Extensive experimental results have demonstrated that SSIM achieves better performance compared with the traditional methods. Sheikh et al. proposed to quantify the loss of image information in the distortion process and explore the relationship between image information and visual quality. They modeled natural image in the wavelet domain using Gaussian scale mixtures (GSM) (Wang and Bovik 2006) and employed natural scene statistics (NSS) model for IQA (Wang and Bovik 2006). The visual information fidelity (VIF) method was derived from the combination of a statistical model for natural scenes, an image distortion model, and a HVS model in information-theoretic setting. The experimental results demonstrate that it outperformed traditional IQA methods by a sizable margin. Sheikh et al. also furnished a statistical evaluation of recent FR algorithms (Tao et al. 2007b).

Since FR metrics need full information of images on the comparison between corresponding regions of the original image and the degraded image, this requirement makes the metrics less than optimal for some applications, which require images and videos to be broadcasted or transmitted through the data network. For these applications, FR metrics might be impossible or too expensive to allocate the extra bandwidth that is required to send information of images. Therefore, researchers are seeking a rational IQA method that could work without any prior information. NR IQA emerges in demand, which does not require any information of the original image, and only makes a distortion analysis of the distorted image to assess its quality. It addresses a fundamental distinction between fidelity and quality, that is, HVS usually does not need any reference to determine the subjective quality of a target image. This kind of metric is most promising in the context of a video broadcast scenario, since the original images or videos are not accessible to terminal users in practice.

NR IQA is a relatively new research direction with promising future. However, NR is also a difficult task, and most NR metrics are designed for one or a set of predefined specific distortion types and unlikely to generalize for image with other types of distortions, for example, blocking effect in JPEG, ringing and blurring artifact in JPEG2000. In practical applications, they are useful only when the distortion type of the distorted images are fixed and known. so there is a big gap between NR metrics and real scenarios. Wang and Bovik (2006) proposed a computational and memory efficient quality model for JPEG images. Li (2002) proposed to appraise the image quality by three objective measures: edge-sharpness level, random noise level, and structural noise level, which jointly provide a heuristic approach to characterizing most important aspects of visual quality. Sheikh et al. (2006) use NSS models to blindly measure the quality of image compressed by JPEG2000 scheme. They claim that
natural scenes contain nonlinear dependencies that are disturbed by the compression process, and this disturbance can be quantified and related to human perceptual quality.

FR metrics may not be applicable because of the absence of original images. On the other hand, NR or “blind” IQA is such an extremely difficult task that it is impossible to apply a universal NR metric to practical applications. Although there is an urgent demand for NR QA methods that are applicable to a wide variety of distortions, unfortunately no such method has been proposed and extensively tested. RR IQA provides an alternative solution that compromises the FR and NR methods. It only makes use of partial information from the original images to evaluate the perceptual quality of the distorted image. In general, certain features or physical measures are extracted from the original image and then transmitted to the receiver as extra information for evaluating the quality of the image or video.

RR metrics may be less accurate for evaluating image quality than the FR metrics, but they are less complicated and make real-time implementations more affordable. RR metrics are very useful for monitoring quality on transmission network. In such contexts, the features are transmitted with the coded sequence if they correspond to a reasonable overhead. They also can be used to track image quality degradations and control the streaming resources in real-time visual communication systems. Recently, the VQEG (2000a) has included RR image and video quality assessment as one of its directions for future development. RR methods are useful in a number of applications. Masry et al. (2006) presented a computationally efficient video distortion metric that can operate in FR and RR model as required. This metric is based on a model of the HVS and implemented using the wavelet transform and separable filters. The visual model is parameterized using a set of video frames and the associated quality scores. Wolf and Pinson (2005) presented a new RR video quality monitoring system under low bandwidth. This system utilizes less than 10 kbits/s of reference information from source video stream.

In the RR model, the key issue is to determine which features have the best ability to capture distortion between the original image and the distorted image. The best features are the ones that are able to produce the highest correlation between objective quality scores and subjective ones. In order to select the best feature to design a reduced description IQA method, Lu et al. (Gao et al. 2008a; Lu et al. 2008; Li et al. 2009) utilized some transforms to extract features for representing image based on HVS. Wang and Bovik (2006) proposed an RR IQA method called 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. They use the Kullback–Leibler distance to represent this error, so that only a relatively small number of RR features are needed for the evaluation of image quality.

In recent years, neural computing has emerged as a practical technology, with successful applications in many fields. The majority of these applications are concerned with problems in image processing. They can address most of the various steps that are involved in the image processing chain: from the preprocessing to the image understanding level. For example, in the process of image construction, Wang and Wahl trained a Hopfield artificial neural network (ANN) for the reconstruction of 2D images from pixel data obtained from projections (Wang and Wahl 1997). In image enhancement, Chandresakaran et al. (1996) used a novel
feed-forward architecture to classify an input window as containing an edge or not. The weights of the network were set manually instead of being obtained from training. In the phase of image understanding, it is considered as part of artificial intelligence or perception, which involves recognition, classification, and relational matching. So it seems that neural networks can be used to an advantage in certain problems of image understanding, especially in feature extraction. Feature extraction can be seen as a special kind of data reduction, of which the goal is to find a subset of informative variables based on image data. In image processing, an effective approach to extract feature can be employed to represent image sparsely. A well-known feature-extraction ANN is Oja's neural implementation of a one-dimensional principal component analysis (PCA) (Oja 1982). To aim at IQA, feature extraction is probably the most important stage, and effective features can well reflect the quality of digital images and vice versa. So the neural network for feature extraction is used to evaluate image quality in RR mode. In Bishop (1995) and Lampinen and Oja (1998), Callet et al. (2006) use a convolutional neural network (CNN) that allows a continuous time scoring of the video to complete the QA in MPEG-2 video. Objective features are extracted on a frame-by-frame basis on both the reference and the distorted sequences. They are derived from a perceptual-based representation and integrated along the temporal axis using a time-delay neural network (TDNN). By realizing a nonlinear mapping between nonsensory features extracted from the video frames and subjective scores, the experimental results demonstrated that TDNN can be useful to assess the perceived quality of video sequences. Gastaldo and Zunino (2004) use a circular back propagation (CBP) feed-forward network to process objective features extracted from JPEG images and to return the associated quality scores. Gastaldo et al. (2002) feed the CBP network, estimating the corresponding perceived quality; objective features are continuously extracted from compressed video streams on a frame-by-frame basis. The resulting adaptive modeling of subjective perception supports a real-time system for monitoring displayed video quality.

Although the aforementioned RR metrics achieve a good solution for IQA problems, there is still much room to further improve the performance of RR IQA, because the existing methods fail to consider the statistical correlations of transformed 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; for example, lines and curves, and wavelet coefficients are dense for smooth image edge contours.

To target the aforementioned problems, to further improve the performance of RR IQA, and to broaden RR IQA related applications, a novel HVS driven framework is proposed. The new framework is consistent with HVS: MGA decomposes images for feature extraction to mimic the multichannel structure of HVS, CSF reweights 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 (Mallet 1989), contourlet (Do and Vetterli 2005), wavelet-based contourlet transform (WBCT) (Eslami and Radha 2004), and hybrid wavelets and directional filter banks (HWD) (Eslami and Radha 2005), and different transforms capture different types of image geometric information. Extensive experiments based on laboratory for image and video engineering (LIVE) database (Sheikh et al. 2003) against subjective MOS (VQEG 2000a) have been conducted to demonstrate the effectiveness of the new framework.
Multiscale Geometric Analysis

Multiscale geometric analysis (MGA) (Romberg 2003) 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, for example, edges, which nominally span a high-dimensional space but contain important features approximately concentrated on lower-dimensional subsets, for example, curves.

In the past decades, scientists devoted themselves to finding a simple way to present images. With the advances in science, MGA has formed a big family with a large number of members. Wavelet is the first one that can analyze image by multiscale and multidirection transforms, and can recover the original image lossless. Due to their good nonlinear approximation (NLA) performance for piecewise smooth function in one dimensions, wavelets have been successfully applied to many image processing tasks, such as low bit-rate compression and de-noising. To utilize wavelets, one can catch point or zero-dimensional discontinuities effectively. However, in two-dimension or other higher dimensions, the wavelet is not able to depict the singulars efficiently and effectively. In essence, wavelets in two-dimension obtained by a tensor-product of one-dimensional wavelets will be good at isolating the discontinuities at edge points, but will not see the smoothness along the contours.

Taking Fig. 1 as an example (Do and Vetterli 2005) for approximating the contour efficiently, the wavelet is limited to using brushes of square shapes along the contour, with different sizes corresponding to the multiresolution of wavelets. As the resolution gets finer, the limitation of the wavelet scheme can be clearly seen since it requires many finer dots to capture the contour. The X-let styles, which are the expected transforms, on the other hand, have more freedom in making brush strokes in different directions and rectangular shapes that follow the contour. As hoped, the expected NLA methods are much more efficient than the wavelet.

In addition, wavelets have only three directions and lack the important feature of directionality; hence, they are not efficient in retaining textures and fine details in these applications. There have been several efforts toward developing geometrical image transforms. According to the above analysis, the expected multiscale geometrical image transforms should contain the following features (Do and Vetterli 2005). Multi-resolution, localization, critical sampling, directionality, and anisotropy.

Fig. 1
The wavelet versus other lets as approximating the 2D contour.
Recently, some more powerful MGA representations were created. In 1998, Candès pioneered a nonadaptive method for high dimension functions representation, named Ridgelet (Candès 1998). And in the same year, Donoho gave the method for constructing the orthonormal ridgelet. The ridgelet can approximate the multivariable functions containing the line-singularity effectively, but for the others with curve-singularity, it performs only as well as wavelet. In 2000, Pennec and Mallat proposed a new system of representation, called bandelet (Pennec and Mallat 2000). It can represent the images based on the edges and track the geometrically regular directions of images. Thus, if one knew the geometrically regular directions of the images, bandelet would lead to optimal sparse representations for images and have great potential in image compression. In order to achieve optimal approximation behavior in a certain sense for 2D piecewise smooth functions in $\mathbb{R}^2$ where the discontinuity curve is a $C^2$ function, Candès and Donoho constructed the curvelet (Candès and Donoho 2000) transform. More specifically, an M-term NLA for such piecewise smooth function using curvelets has $L^2$ square error decaying. An attractive property of the curvelet system is that such correct approximation behavior is simply obtained via thresholding a fixed transform. The key features of the curvelet elements are that they exhibit very high directionality and anisotropy. However, the original construction of the curvelet transform is intended for functions defined in the continuum space; when the critical sampling is desirable, the development of discrete transforms for sampled images remain a challenge.

Based on the key features, namely directionality and anisotropy, which make curvelets an efficient representation for 2D piecewise smooth functions with smooth discontinuity curves, Do and Vetterli proposed a new image transform, contourlet, which is also called pyramidal directional filter bank (PDFB) (Bamberger and Smith 1992). The contourlet first uses the Laplacian pyramid (LP) (Burt and Adelson 1983) to decomposition the image and capture point singularity. Then, it combines all the singular points in every direction to one coefficient by directional filter bank. The contourlet provides a multiscale and directional decomposition for images with a small redundancy factor, and a frame expansion for images with frame elements like contour segments, which is the reason that it is named contourlet. The connection between the developed discrete and continuous-domain constructions is made precise via a new directional multi-resolution analysis, which provides successive refinements at both spatial and directional resolutions. The contourlet transform can be designed to satisfy the anisotropy scaling relation for curves and thus it provides a curvelet-like decomposition for images. Contourlet transform 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 and HWD are designed to optimize the representation of image features without redundancy.

For IQA, one needs 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, the effective representation of visual information is especially important. Natural images are not simple stacks of 1D piecewise smooth scan-lines and points of discontinuity are typically located along smooth curves owing to smooth boundaries of physical objects. As a result of a separable extension from 1D bases, 2D wavelet transform is not good at representing visual information of image. Consequently, image is disposed of with linear features and wavelet transform is not effective. However, MGA transforms can capture the characteristics of image, for example, lines, curves, cuneiform, and the contour of object. As mentioned in Table 1, different transforms of MGA capture different features of an image and complement each
other. Based on mentioned requirements for IQA, it is reasonable to consider a wide range of explicit interactions between multiscale methods and geometry, for example, contourlet (Do and Vetterli 2005), WBCT (Eslami and Radha 2004), and HWD (Eslami and Radha 2005).

3 Multiscale Geometric Analysis for Image Quality Assessment

As discussed in Sect. 2, MGA contains a series of transforms, which can analyze and approximate geometric structure while providing near optimal sparse representations. In the image sparse representation, one 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 chapter, a novel framework for IQA (Gao et al. 2008b) is developed by applying MGA transforms to decompose images and extract effective features. This framework (Wang and Bovik 2006) quantifies the errors between the distorted and the reference images by mimicking the error sensitivity function in and HVS. The objective of this framework is to provide IQA results, which have good consistency with subjective perception values. Figure 2 shows the framework for IQA.

3.1 MGA-Based Feature Extraction

A number of MGA transforms (Do 2001), which are contourlet, WBCT, and HWD, are considered for image decomposition and feature extraction. Moreover, wavelet is utilized as the baseline for comparison. However, there is a wide application of neural computing that can be applied to perform feature extraction for image representation sparsely. Feature extraction (Bishop 1995) is treated as a means for reducing dimensionality of the image data and preserving feature data separability well, for example, feed-forward ANN, self-organizing feature map (SOMs), Hopfield ANN. Some methods of neural computing are selected to extract feature of image data further. In this framework, MGA is utilized to decompose images and then extract features to mimic the multichannel structure of HVS.

3.1.1 Wavelet Transform

Wavelet transform (Mallat 1989) is well known as an approach to analyze signal in both time and frequency domains simultaneously and adaptively. Features are extracted effectively from

<table>
<thead>
<tr>
<th>Metric Type</th>
<th>Metric</th>
<th>Type</th>
<th>ROCC</th>
<th>OR</th>
<th>MAE</th>
<th>RMSE</th>
<th>ROCC</th>
<th>OR</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR FR</td>
<td>JPEG</td>
<td>0.9229</td>
<td>0.8905</td>
<td>0.1886</td>
<td>7.118</td>
<td>9.154</td>
<td>0.933</td>
<td>0.9041</td>
<td>0.0947</td>
<td>6.407</td>
</tr>
<tr>
<td>MSSIM FR</td>
<td>JPEG2000</td>
<td>0.9674</td>
<td>0.9485</td>
<td>0.04</td>
<td>4.771</td>
<td>5.832</td>
<td>0.949</td>
<td>0.9368</td>
<td>0.0651</td>
<td>5.422</td>
</tr>
<tr>
<td>WNISM RR</td>
<td>JPEG</td>
<td>0.9291</td>
<td>0.9069</td>
<td>0.1486</td>
<td>6.0236</td>
<td>8.2446</td>
<td>0.9261</td>
<td>0.9135</td>
<td>0.1183</td>
<td>6.135</td>
</tr>
</tbody>
</table>

The performance of PSNR, MSSIM, and WNISM on the LIVE database

Table 1
signal, especially non-stationary signal, by multiscale operation. In this manuscript, three levels of wavelet transform are applied to decompose the image into nine highpass subbands and a lowpass residual subband. Then all the highpass subbands are selected for feature extraction in the proposed framework. Figure 3 shows decomposition of the image by wavelet transform and a set of selected subbands (marked with white dashed boxes and white numerals) are used for feature extraction in the proposed framework.

### 3.1.2 Contourlet Transform

Contourlet transform (Mallat 1989) can capture the intrinsic geometrical structure that is a key in visual information. Contourlet consists of two major stages: the multiscale decomposition and the direction decomposition. At the first stage, it uses LP (Bishop 1995) to capture the point discontinuities. For the second stage, it uses directional filter banks (DFB) (Burt and Adelson 1983) to link point discontinuities into linear structures. Here, every image is decomposed into three pyramidal levels. Based on the characteristics of DFB for decomposition, half of the directional subbands are selected for the feature extraction. The decomposition of image by contourlet transform and a set of selected subbands (marked...
with white dashed boxes and numerals) are used for feature extraction in the proposed framework as shown in Fig. 4.

3.1.3 WBCT

WBCT (Eslami and Radha 2004) has a construction similar to the contourlet transform, which consists of two filter bank stages. The first stage is subband decomposition by wavelet transform. The second stage of the WBCT is angular decomposition. In this stage, WBCT employs DFB as in the contourlet transform. Here, each image is decomposed into two wavelet levels, and the number of DFB decomposition levels at each highpass of the finest and finer scales is equal to 3. The image is decomposed into 48 high frequency directional subbands and a lowpass residual subband in all. Like contourlet, half of the subbands at each fine scale are selected to extract features of image. The decomposition of image by WBCT transform and a set of selected subbands (marked with numerals and gray blocks) are applied for feature extraction in the proposed framework as demonstrated in Fig. 5.
3.1.4 HWD Transform

Like WBCT, the HWD (Eslami and Radha 2005) employs wavelets as the multiscale decomposition. Then DFB and modified DFB are applied to some of the wavelets subbands. When the HWD (HWD1 and HWD2 respectively) is employed in the proposed framework, each image is decomposed into two wavelet levels, and the number of DFB decomposition levels at each highpass of the finest and finer scales is equal to 3. Thus, the image is decomposed into 36 high frequency directional subbands and a lowpass residual subband in all. Half of the directional subbands at each fine scale (for HWD1 and HWD2, 16 directional subbands in all, respectively) are selected to extract features of image. The HWD decomposition of image and a set of selected subbands (marked with numerals and gray blocks) are applied for feature extraction in the proposed frameworks demonstrated in Fig. 6.
3.2 CSF Masking

MGA is introduced to decompose images and then extract features to mimic the multichannel structure of HVS; that is, HVS (Wandell 1995) works similar to a filter bank (containing filters with various frequencies). CSF (Miloslavski and Ho 1998) measures how sensitive one is to the various frequencies of visual stimuli; that is, one is 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, it can be perceived as a gray image if stripes are very thin; otherwise, these stripes can be distinguished. 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) (Miloslavski and Ho 1998), that is,

\[ H(f) = a(b + cf)^{d(f)} \]

where \( f = f_n \times f_s \) is 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 Bamberger and Smith (1992), a, b, c, and d are 2.6, 0.192, 0.114, and 1.1, respectively. The sampling frequency \( f_s \) is defined as (Nadenau et al. 2003)

\[ f_s = \frac{2v \cdot \tan(0.5^\circ)}{0.0254} \]

where \( v \) is the viewing distance with units of meter and is the resolution power of the display with units of pixels/in. In this framework, \( v \) is 0.8 m (about 2–2.5 times height of the display), the display is 21 in. with the resolution of 1024 x 768, and pixels/in. According to the Nyquist sampling theorem, it 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, one can 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, a series of contourlet coefficients \( c_{kij} \) are obtained, 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 \( k \)th level, and \( j \) represents the coefficient index. By using CSF masking, the coefficient \( c_{kij} \) is scaled to \( x_{kij} = H(f_k) \cdot c_{kij} \).

3.3 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. 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 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

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

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

In JND threshold, some parameters are so delicate that they have to be selected by empirical values. Others may be tunable by trial and error or cross validation process by observing the overall performance of the system. For optimizing JND threshold, some methods of neural computing (Bishop 1995) might be applied to train parameters or variables for autotif threshold. Neural networks provide a whole family of divergent formalisms to optimize corresponding values for improving the overall performance, for example, radial basis function (RBF) network.

### 3.4 Normalization of Sensitivity

By using JND threshold $T$, one 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 $n$th selected subband that are larger than $T$ obtained from Eq. 5. The number of coefficients in the $n$th selected subband is $C(n)$. Therefore, for a given image, one can obtain the normalized histogram with $L$ bins ($L$ subbands are selected) for representation and the $n$th entry is given by

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

### 3.5 Sensitivity Errors Pooling

Based on Eq. 4, the normalized histograms can be obtained for both the reference and the distorted images as $P_R(n)$ and $P_D(n)$, respectively. In this framework, the metrics of the distorted image quality can be defined as

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

where $S = \sum_{n=1}^{L} |P_R(n) - P_D(n)|$ is the city-block distance between $P_R(n)$ and $P_D(n)$, and $Q_0$ is a constant used to control the scale of the distortion measure. In this framework, one can set $Q_0$ as 0.1. The log function is introduced here to reduce the effects of large $S$ and enlarge the effects of small $S$, so that one 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, for example, Euclidean norm. This is also the case for the base 2 for the logarithm. The entire function preserves the monotonic property of $S$. 
4 Performance Evaluation

In this section, the performance of the proposed framework based on different MGA transforms are compared with standard IQA methods, that is, PSNR, WNISM, and MSSIM, based on the following experiments: the consistency experiment, the cross-image and cross-distortion experiment, and the rationality experiment. At the beginning of this section, the image database for evaluation are first briefed.

The LIVE database (Sheikh et al. 2003) is widely used to evaluate the image quality measures, in which 29 high-resolution RGB color images are compressed at a range of quality levels using either JPEG or JPEG2000, producing a total of 175 JPEG images and 169 JPEG2000 images. To remove any nonlinearity due to the subjective rating process and to facilitate comparison of the metrics in a common analysis space, following the procedure given in the video quality experts group (VQEG) (VQEG 2000) test, variance-weighted regression analysis is used in a fitting procedure to provide a nonlinear mapping between the objective and subjective MOS. After the nonlinear mapping, the following three metrics are used as evaluation criteria (VQEG 2000): Metric 1 is the Pearson linear correlation coefficient (CC) between the objective and MOS after the variance-weighted regression analysis, which provides an evaluation of prediction accuracy; Metric 2 is the Spearman rank-order correlation coefficient (ROCC) between the objective and subjective scores, which is considered as a measure of the prediction monotonicity; Metric 3 is the outlier ratio (OR), percentage of the number of predictions outside the range of twice of the standard deviation of the predictions after the nonlinear mapping, which is a measure of the prediction consistency. In addition, the mean absolute error (MAE) and the root mean square error (RMSE) of fitting procedure are calculated after the nonlinear mapping.

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

4.1 The Consistency Experiment

In this subsection, the performance of the proposed IQA framework will be compared with PSNR (Avcibas et al. 2002), WNISM (Wang and Bovik 2006), and the well-known full reference assessment metric, MSSIM (Wang and Bovik 2006). The evaluation results for all IQA methods being compared are given as benchmark in Table 1. Table 2 shows the performance of the proposed IQA framework with different MGA transforms on the LIVE database.

<table>
<thead>
<tr>
<th>Metric</th>
<th>JPEG</th>
<th>JPEG2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>ROCC</td>
</tr>
<tr>
<td>Wavelet</td>
<td>1</td>
<td>0.9587</td>
</tr>
<tr>
<td>Contourlet</td>
<td>3</td>
<td>0.9493</td>
</tr>
<tr>
<td>WBCT</td>
<td>3</td>
<td>0.9728</td>
</tr>
<tr>
<td>HWD1</td>
<td>2</td>
<td>0.9704</td>
</tr>
<tr>
<td>HWD2</td>
<td>3</td>
<td>0.9728</td>
</tr>
</tbody>
</table>
evaluation results of the proposed IQA framework with different MGA transforms, e.g., wavelet, contourlet, WBCT, HWD1, and HWD2. Figures 7 and 8 present the scatter plots of MOS versus the predicted score by objective metrics after the nonlinear mapping.

Table 2 shows that the results of MGA transforms using the proposed framework provide a higher effectiveness to IQA for JPEG and JPEG2000 images, respectively. Comparing

Fig. 7
Scatter plots of mean opinion score (MOS) versus different image quality assessment (IQA) methods for JPEG and JPEG2000 images.
Fig. 8
Trend plots of Lena with different distortions using the proposed framework with contourlet.
Table 2 with Table 1, one can observe that the metric obtained by the proposed framework provides much better performance than WNISM including better prediction accuracy (higher CC), better prediction monotonicity (higher ROC), and better prediction consistency (lower OR, MAE, and RMSE). Particularly when HWD are employed in the proposed framework, better performance will be achieved than that of MSSIM index. The only tuning parameter \( \alpha \) in the proposed framework responses to different distortions.

Since the key stage of IQA is how to represent images effectively and efficiently, it is necessary to investigate different transforms. Wavelet is localized in spatial domain and frequency domain and can extract local information with high efficiency, so it optimally approximates a target function with 1D singularity. However, wavelet cannot achieve the sparse representation of edges even if it captures the point singularity effectively. In order to represent edges in an image sparsely, contourlet analyzes the scale and the orientation respectively and reduces the redundancy by approximating images with line-segment similar basis ultimately. To further reduce the redundancy, 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 multi-orientation for image decomposition. Table 3 shows the proposed IQA framework with HWD works much better than previous standards.

### 4.2 The Rationality Experiment

To verify the rationality of the proposed framework, contourlet and WBCT are chosen to test the Lena image with different distortions, which are blurring (with smoothing window of \( W^a W \)), additive Gaussian noise (mean = 0, variance = \( V \)), impulsive salt–pepper noise (density = \( D \)), and JPEG compression (compression rata = \( R \)).

Figures 9 and 10 (all images are 8 bits/pixel and resized from 512 x 512 to 128 x 128 for visualization) show the relationships between the Lena image with different distortions and the IQA methods prediction trend to the corresponding image.

#### Table 3

The value of different IQA metrics for images in Fig. 11

<table>
<thead>
<tr>
<th>Metric</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>24.8022</td>
<td>24.8013</td>
<td>24.8041</td>
</tr>
<tr>
<td>MSSIM</td>
<td>0.9895</td>
<td>0.9458</td>
<td>0.6709</td>
</tr>
<tr>
<td>Wavelet</td>
<td>1.0000</td>
<td>0.3208</td>
<td>0.2006</td>
</tr>
<tr>
<td>Contourlet</td>
<td>1.0000</td>
<td>0.2960</td>
<td>0.2423</td>
</tr>
<tr>
<td>WBCT</td>
<td>1.0000</td>
<td>0.2345</td>
<td>0.1929</td>
</tr>
<tr>
<td>HWD1</td>
<td>1.0000</td>
<td>0.2740</td>
<td>0.2166</td>
</tr>
<tr>
<td>HWD2</td>
<td>1.0000</td>
<td>0.2704</td>
<td>0.2094</td>
</tr>
</tbody>
</table>
For JPEG compression, it can be found 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 the IQA framework; that is, information in the lowpass subband is not considered 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.

### 4.3 Sensitivity Test Experiment of the Proposed Framework

Currently, MSE and PSNR are the most commonly used objective quality metrics for images. However, they do not correlate well with the perceived quality measurement. \( \textbf{Figure 11} \) shows three degraded “Lena” images with different types of distortion but with the same PSNR; however, the perceived qualities of them are obviously varied. As shown in \( \textbf{Table 3} \), the proposed framework with different MGA transform can distinguish them very well. It is noted that $Q$ is insensitive to the changes of intensity, which means a slight change of gray value at image level will not affect the image quality when $Q$ equals 1, it is consistent with the HVS character that only if the distorted image and the reference image have the same variance of light, the two images have the same visual quality.
5 Conclusion

This chapter has described a RR IQA framework by incorporating merits of MGA and HVS. In this framework, sparse image representation based on MGA is utilized to mimic the multichannel structure of HVS, and then to balance the magnitude of coefficients obtained by MGA mimic nonlinearities of HVS, and JND is utilized to produce a noticeable variation in sensory experience. 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 the conventional standard RR IQA method. Since RR methods require full limited information of the reference image, it could be a serious impediment for many applications. It is essential to develop NR image quality metrics that blindly estimate the quality of images. Recently, tensor-based approaches (Tao et al. 2006, 2007a, b) have been demonstrated to be an effective way for image representation in classification problems, so it is valuable to introduce them for image quality.

Acknowledgments

This research was supported by National Science Foundation of China (60771068, 60702061, 60832005), the Open-End Fund of National Laboratory of Pattern Recognition in China and National Laboratory of Automatic Target Recognition, Shenzhen University, China, and the
Program for Changjiang Scholars and innovative Research Team in University of China (IRT0645). The authors would also like to thank the Nanyang Technological University Nanyang Assistant Professor Start-Up Grant (under project number M58020010).

References


Gao XB, Lu W, Li XL, Tao DC (2008b) Image quality assessment based on multiscale geometric analysis. Accepted to IEEE Trans Image Process


**Author Query Form**

**Handbook of Natural Computing**  
**Chapter No.: 11**

<table>
<thead>
<tr>
<th>Query Refs.</th>
<th>Details Required</th>
<th>Author's response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU1</td>
<td>Please provide correct author name in citation for Ref. 'Rec. ITU-R'.</td>
<td></td>
</tr>
<tr>
<td>AU2</td>
<td>Reference Damera-Venkata does not match with Damera (2000) in ref. list Please check.</td>
<td></td>
</tr>
<tr>
<td>AU3</td>
<td>Reference Sheikh et al. does not match with Sheikh (2003) in ref. list Please check.</td>
<td></td>
</tr>
<tr>
<td>AU4</td>
<td>Please check the sentence “Feature extraction... Hopfield ANN” for sense.</td>
<td></td>
</tr>
<tr>
<td>AU5</td>
<td>Please check the uncited References Chitprasert and Rao (1990), Petersena et al. (2002), and Sheikh et al. (2003).</td>
<td></td>
</tr>
<tr>
<td>AU6</td>
<td>Please check and provide the complete details for Reference Rec. ITU-R.</td>
<td></td>
</tr>
<tr>
<td>AU7</td>
<td>Please check the complete details of Ref. Tao et al. (2007b).</td>
<td></td>
</tr>
<tr>
<td>AU8</td>
<td>This sentence seems to be long and unclear. Please check.</td>
<td></td>
</tr>
</tbody>
</table>