# Fine-Grained Quality Assessment for Compressed Images

Xinfeng Zhang, *Member, IEEE*, Weisi Lin, *Fellow, IEEE*, Shiqi Wang, *Member, IEEE*, Jiaying Liu, *Senior Member, IEEE*, Siwei Ma, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— Image quality assessment (IQA) has attracted more and more attention due to the urgent demand in image services. The perceptual-based image compression is one of the most prominent applications that require IQA metrics to be highly correlated with human vision. To explore IQA algorithms that are more consistent with human vision, several calibrated databases have been constructed. However, the distorted images in the existing databases are usually generated by corrupting the pristine images with various distortions in coarse levels, such that the IQA algorithms validated on them may be inefficient to optimize the perceptual-based image compression with fine-grained quality differences. In this paper, we construct a large-scale image database which can be used for fine-grained quality assessment of compressed images. In the proposed database, reference images are compressed at constant bitrate levels by JPEG encoders with different optimization methods. To distinguish subtle differences, the pair-wise comparison method is utilized to rank them in subjective experiments. We select 100 reference images for the proposed database, and each image is compressed into three target bitrates by four different JPEG optimization methods, such that 1200 distorted images are generated in total. Sixteen well-known IQA algorithms are evaluated and analyzed on the proposed database. With the devised fine-grained IQA database, we expect to further promote image quality assessment by shifting it from a coarse-grained stage to a fine-grained stage. The database is available at: https://sites.google.com/site/zhangxinf07/fg-iqa.**

***Index Terms*— Image quality assessment, peceptual image compression, image database, subjective assessment, fine-grained distortion levels.**

## I. Introduction

IMAGE quality assessment (IQA) aims to measure the perceived visual signal quality according to its statistical characteristics and human perceptual mechanism, which is widely required in numerous image processing applications. IQA plays a vital role in guiding many visual processing algorithms and systems, as well as their implementation, optimization and verification [1]–[4]. In particular, image compression is one of the most representative applications of IQA, which can be utilized in the rate-distortion optimization process to obtain compressed images with better visual quality at the same bit-rate level [5]–[10]. The traditional image compression methods mainly utilize the signal-fidelity based quality metrics, which are less correlated with human perceptual quality, *e.g.,* MAE (mean absolute error), MSE (mean square error), SNR (signal-to-noise ratio), PSNR (peak SNR) and their relatives. Although these metrics possess many favorable properties, *e.g.,* clear physical meaning and high efficiency for calculation, they severely hinder the compression performance improvement in further reducing the visual redundancies in images due to their poor consistency with human visual perception.

To obtain more consistent measures with human visual perception, many perceptual quality metrics have been proposed during the recent years. According to the availability of a reference image, these methods can be divided into three categories, *i.e.,* full reference (FR) ones where the pristine reference image is available, reduced reference (RR) ones where partial information of the reference image is available and no reference (NR) ones where the reference image is unavailable. For image compression problem, the reference images are available at the encoder side such that the FR-IQA algorithms are applicable. By contrast, for image restoration or enhancement problems, where the reference image is in absence, robust NR-IQA algorithms are required.

The well-known Structural SIMilarity (SSIM) index [11] measures the patch similarity between the reference and distorted images instead of the pixel-level distortion calculation. It is based on the assumption that the human visual system (HVS) tends to perceive the local structures and achieves more consistent results with subjective quality assessment on popular databases. Furthermore, its variants, *e.g.,* Multi-Scale SSIM (MS-SSIM) [12], Feature similarity index (FSIM) [13], Information Weighted SSIM (IW-SSIM) [14], further improve the quality assessment performance by measuring local structure distortions in different spaces. Considering that the gradients are sensitive to distortions, the gradient magnitude similarity is utilized in IQA algorithms *e.g.,* Gradient Magnitude Similarity (GSM) [15] and Gradient Magnitude Similarity Deviation (GSMD) [16], which achieve comparable

Manuscript received December 7, 2017; revised May 29, 2018 and August 7, 2018; accepted September 17, 2018. Date of publication October 8, 2018; date of current version November 2, 2018. This work was supported in part by Peking University, in part by the National Natural Science Foundation under Grant 61632001, in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800, and in part by the Top-Notch Young Talents Program of China. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kalpana Seshadrinathan. *(Corresponding author: Siwei Ma.)*

X. Zhang and W. Lin are with the Rapid-Rich Object Search Lab, Nanyang Technological University, Singapore 639798 (e-mail: xfzhang@ntu.edu.sg; wslin@ntu.edu.sg).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

J. Liu is with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: liujiaying@pku.edu.cn).

S. Ma and W. Gao are with the School of Electronics Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing 100871, China (e-mail: swma@pku.edu.cn; wgao@pku.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2018.2874283

performance with SSIM and its variants with fewer computation complexities. The Visual Information Fidelity (VIF) [17] and Information Fidelity Criterion (IFC) [18] are another kind of FR-IQA methods introducing the natural scene statistics (NSS) into image fidelity measurement, which can well quantify the loss of the information that could be extracted by the brain.

Besides the FR-IQA, there are also many NR-IQA and RR-IQA algorithms which mainly utilize the NSS and human visual features. Mittal *et al.* [19] proposed blind/referenceless image spatial quality evaluator (BRISQUE) algorithm to assess general distorted images by utilizing a NSS model of locally normalized luminance coefficients. Zhai *et al.* [20] and Gu *et al.* [21] introduced the free-energy principle from human brain theory to model the perception and understanding of an image as an active inference process, and designed HVS-inspired features to qualify image quality. Ye *et al.* proposed a Codebook Representation for No-Reference Image Assessment (CORNIA) algorithm, which utilizes the learned features from the raw-image-patches instead of hand-crafted features to improve the IQA generalization. There are also many other NR-IQA algorithms designed for different kinds of distortions, *e.g.,* the JPEG and JPEG2000 compression distortions [22], [23], the deblocked images [24] and so on.

These well designed IQA algorithms have achieved more consistent results with human visual perception compared with signal-fidelity based IQA algorithms on many calibrated public databases, *e.g.,* LIVE [25], [26] and TID2008 [27]. Although the correlation coefficients between some IQA algorithms and subjective scores are even up to more than 0.9 on existing databases, they are still obviously inconsistent with subjective results for some cases especially in distinguishing the subtle quality difference in practice. This problem is more prominent in perceptual-based image compression application, the target of which is to achieve the most visually pleasing quality at given bitrates. Meanwhile, the perceptual image qualities generated from different coding parameters, *e.g.,* prediction modes and partition modes etc [28], are quite close. Moreover, given a target bitrate, there may not be significant quality differences when applying different rate control algorithms to predict quantization parameter (QP) values.

To the best of our knowledge, most of the existing IQA databases usually contain limited distortion levels (5-6 levels) covering the whole quality range from "Bad" to "Excellent", which make the images in adjacent distortion levels obviously different and easy to rank. In the following of the paper, we use the terms "coarse-grained" and "fine-grained" to describe the obvious and subtle quality differences between two images. More specifically, the images with "fine-grained" quality difference correspond to the compressed ones generated using different optimization methods at the same or approached bitrate, while the images with "coarse-grained" quality differences correspond to the compressed ones generated using the same codec at obvious different bitrates in this paper. Therefore, these databases with coarse-grained distortion variations for the same image may not be able to provide sufficient information to further improve the performance of IQA algorithms in evaluating fine-grained quality differences. We think that this may be one of the reasons why the improvement of image compression is marginal when applying the state-of-the-art IQA algorithms to rate-distortion optimization process. Another weakness for the existing IQA databases is that they only contain a few reference images with limited visual content, *e.g.,* about 20-30 reference images. Although the Waterloo Exploration Database [29] is the latest one with large-scale reference images which contains 4,744 pristine natural images and 94,880 distorted images, it is impossible to perform subjective experiments on them. As such, the authors have to rank them using several most-trusted FR-IQA measures and determine the pairwise preference when the prediction score is relative large, which is also a coarse-grained quality database.

In this paper, we push the IQA research beyond its current scope, and promote the IQA in the new challenges of the fine-grained quality assessment task by constructing a large-scale IQA database with fine-grained distortion differences. In the new database, we carefully select 100 reference images with different real-world contents, and compress them into 3 target bitrates. At each target bitrate, each reference image is compressed by JPEG encoders with four optimized quantization schemes, and this configuration satisfies the practical perceptual-based image compression scenario that aims at achieving best visual quality at the same bitrate. Therefore, each reference image corresponds to 4 distorted images at the given bitrate, and there are 1200 distorted images in total for all three target bitrates. Since the quality differences are marginal among every 4 distorted images corresponding to the same reference image, they are difficult to be distinguished from single stimulus subjective experiments. To provide the faithful rank on the quality of these images, the pair-wise comparison subjective experiments are conducted to rank every 4 distorted images corresponding to the same reference image, which lead to 1800 comparisons for all the distorted images. We invited 30 subjects in the subjective experiments and up to 54,000 comparisons are conducted. Finally, we analyze 16 state-of-the-art IQA algorithms on the proposed database, and show that there is still a large room to improve the IQA in the prediction of the fine-grained quality preference.

This database is constructed to provide benchmark for compressed image quality assessment, and also benefit for perceptual-based image compression. This is because that, the existing compressed image quality databases with coarse-grained quality differences are inefficient to evaluate IQA methods on images with fine-grained quality differences. However, in perceptual-based image compression problem, for each coding block there are many coding modes to select according to their rate-distortion costs. These distortion differences generated by different coding modes are fine-grained. Therefore, the proposed database can help researchers in image compression community to select the best IQA method to do the perceptual based image optimization.

The remainder of this paper is organized as follows. Section II reviews the related works for IQA databases and perceptual-based image compression. Section III introduces the proposed fine-grained IQA database construction including the data generation and subjective experiments. Section IV

TABLE I
SUMMARY OF IQA DATABASES WITH COMPRESSION DISTORTIONS

| Database | No. of Ref. Images | No. of Dist. Images | No. of Dist. Types | No. of JPEG Compression Levels/Compressed images | Subjects No. for each image |
|---|---|---|---|---|---|
| LIVE | 29 | 779 | 5 | 8/233 | 20-29 subjects for each image using single stimulus |
| TID2008 | 25 | 1700 | 17 | 4/100 | 838 subjects performs 256428 pair-wise comparisons |
| TID2013 | 25 | 3000 | 24 | 5/125 | 971 subjects performs 524340 pair-wise comparisons |
| CSIQ | 30 | 866 | 6 | 5/150 | 25 subjects perform multi-stimulus subjective experiments |
| CID2013 | N/A | 480 | 12-14 | N/A | 26-30 subjects perform multi-stimulus subjective experiments |
| SD-IVL [30] | 20 | 380 | 2 | 9/180 | 26-30 subjects perform multi-stimulus subjective experiments |
| MICT [31] | 14 | 196 | 2 | 7/98 | 16 subjects for each image using single stimulus |
| IVC [32] | 10 | 195 | 4 | 5/75 | 15 subjects for each image using multi-stimulus |
| MCL-JCI [33] | 50 | 5000 | 1 | 100/5000 | 30 subjects for each image using pair-wise comparisons |
| FG-IQA | 100 | 1200 | 1 | 3/1200 | 30 subjects performs 54,000 pair-wise comparisons |

presents the detailed analysis of the popular IQA algorithms on the proposed fine-grained IQA database. Finally, we conclude this paper in Section V.

## II. RELATED WORK

### A. IQA Databases and Challenges

Several IQA databases have been released in recent years and widely utilized in evaluating the image quality assessment algorithms [25]–[27], [33]–[36]. Sheikh *et al.* [25], [26] created the well-known LIVE database which contains 29 reference images and 779 distorted images with five distortion types: JPEG2000 compression, JPEG compression, white noise, Gaussian blur and transmission errors. The perceptual quality of these images corresponds to the Mean Opinion Score (MOS), which is obtained from subjective experiments. The LIVE database adopts the single stimulus categorical rating method [37], where the subjects were asked to provide their perceptual quality on each image in the five categories "Bad", "Poor", "Fair", "Good" and "Excellent", which are converted to the values 1, 2, 3, 4 and 5 when calculating the MOS. Around 20-29 human observers rated each image. For each distortion type, the perceptual quality of these distorted images roughly covered the entire quality range. However, each kind of distortion types only has very limited distortion levels, which make them relatively easy to distinguish by subjects. The average bitrate increase between consecutive JPEG compression distortion levels used in LIVE is more than 60% for the same image, which is not practical in perceptual-based image compression optimization among different coding modes.

The TID2008 [27] is another large-scale database containing 1700 distorted images, which is generated from 25 reference images with 17 types of distortions for each reference image, including JPEG and JPEG2000 compression, Gaussian noise, Gaussian blur, etc. There are 4 different levels for each distortion type. The MOS is obtained from 838 subjective experiments carried out by 838 observers using pair-wise comparisons with forced-choice method, which can obtain more stable subjective quality rank by reducing the insecurity of the observers for the quality distribution in the whole database. Ponomarenko *et al.* further extended the TID2008 to TID2013 with more distorted images by generating 24 types of distortions for each image. However, the number of distortion levels is maintained to be 5 in TID2013 for each distortion type [34]. In particular, for JPEG compressed images, the average bitrate increase between consecutive JPEG compression distortion levels is around 77% for each image, which makes the compression distortions in different levels easy to distinguish.

The CSIQ [35] is another popular image quality assessment database, which uses totally different reference images from those in LIVE, TID2008 and TID2013, where most of the reference images are from Kodak test images. The CSIQ database consists of 30 reference images including five categories, *i.e.,* animals, landscapes, people, plants and urban. Each reference image is processed by adding one of the six distortions (JPEG compression, JPEG2000 compression, contrast decrements, additive pink Gaussian noise, additive white Gaussian noise and Gaussian blurring) with four or five levels. In total, there are 866 distorted images, which are evaluated by multi-stimulus subjective experiments to collect 5,000 subjective ratings from 25 different observers, and finally the differential mean opinion scores (DMOS) are obtained.

The CID2013 [38] is a more complex database for NR-IQA and the distorted images are contaminated by many concurrent distortion types, such as real photographic images captured by different digital cameras. There are 480 images captured with 79 different cameras in CID2013, and each image is evaluated by 26-30 subjects using a hybrid absolute category rating-pair comparison. The MCL-JCI dataset [33] is a special database which focuses on the JPEG compressed images and measures the just-noticeable difference (JND) points. It consists of 50 reference images with resolution $1920 \times 1080$ and 100 JPEG compressed images for each reference image with the quality factor (QF) ranging from 1 to 100. Each individual set of compressed images was evaluated by 30 subjects in a controlled environment. The application of the JND to compressed image quality assessment was also discussed in [39].

There are also many other databases developed in the literature, *e.g.,* in [30], [31], [36], and [40]. However, all of them utilized coarse-grained distortion levels, and they are still in small scale which may be not enough to evaluate IQA algorithms. The summary of the database information is

TABLE II

THE PERFORMANCE COMPARISON FOR DIFFERENT IQA METHODS ON JPEG COMPRESSED IMAGES IN LIVE DATABASE

| | $bpp \approx 0.3$ | | | All JPEG images | | |
|---|---|---|---|---|---|---|
| IQA | SROCC | KRCC | PLCC | SROCC | KRCC | PLCC |
| PSNR | 0.4036 | 0.2750 | 0.5699 | 0.8410 | 0.6360 | 0.8595 |
| PSNRHVS | 0.3183 | 0.1899 | 0.4963 | 0.9035 | 0.7199 | 0.9455 |
| VSI | 0.3787 | 0.2532 | 0.3308 | 0.9089 | 0.7308 | 0.9444 |
| SSIM | 0.4299 | 0.2897 | 0.4567 | 0.8999 | 0.7112 | 0.9249 |
| MS-SSIM | 0.4302 | 0.3369 | 0.3642 | 0.9126 | 0.7460 | 0.9425 |
| IWSSIM | 0.3774 | 0.3172 | 0.2834 | 0.9084 | 0.7468 | 0.9420 |
| FSIM | 0.4112 | 0.3312 | 0.3586 | 0.9137 | 0.7500 | 0.9457 |
| RFSIM | 0.3479 | 0.2356 | 0.4720 | 0.8931 | 0.7083 | 0.9258 |
| SRSIM | 0.3943 | 0.3080 | 0.2915 | 0.9116 | 0.7450 | 0.9461 |
| UQI | 0.1836 | 0.1463 | 0.0736 | 0.8247 | 0.6179 | 0.8473 |
| MAD | 0.3571 | 0.2736 | 0.4276 | 0.9061 | 0.7394 | 0.9346 |
| GSM | 0.3675 | 0.2665 | 0.3081 | 0.9102 | 0.7414 | 0.9483 |
| GSMD | 0.3662 | 0.2609 | 0.3877 | 0.9080 | 0.7333 | 0.9448 |
| IFC | 0.2014 | 0.1709 | 0.0901 | 0.8640 | 0.6701 | 0.9032 |
| VIF | 0.4163 | 0.3467 | 0.4278 | 0.9138 | 0.7579 | 0.9437 |

illustrated in Table I. To show the challenges of IQA in fine-grained quality difference, we carry out a sanity check that selects the JPEG compressed images with the bitrates about 0.3bpp in LIVE database to calculate the three correlation coefficients, *i.e.,* SROCC, KRCC and PLCC, between MOS and the scores from different IQA algorithms. The results are shown in Table II. We can see that the correlation coefficients for the distorted images with approximate bitrates are much lower than those calculated by involving all the JPEG compressed images with different levels in LIVE database. In particular, all the three correlation coefficients are lower than 0.57 for JPEG images with the bitrate approximating 0.3bpp. By contrast, most of these correlation coefficients are larger than 0.9 for all the compressed images with coarse-grained quality levels. These results reveal two limitations of the existing IQA databases:

1) The MOS values of the existing databases constructed from single stimulus subjective experiments may not be accurate enough for fine-grained quality image assessment.
2) The existing IQA algorithms have not been sufficiently evaluated on fine-grained quality prediction.

### *B. IQA in Visual Data Compression*

Although image coding standards have normalized the bitstream, various coding parameters or modes determined according to different IQA metrics will lead to obviously distinct compression performance. In JPEG, one of the optional coding parameters is the customized quantization table, and the default table is determined empirically based on human perception [41]. For example, the quantization table of luminance component at quality factor (QF) equal to 50 is shown in Fig.1(a), which is scaled to generate quantization tables for other quality factors. Besides the JPEG default quantization table, the open source and well optimized JPEG codec, *libjpeg* [42], adopted another 8 quantization tables, and one of them is an optimized quantization table based on MS-SSIM as shown in Fig.1(b).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 11 | 10 | 16 | 24 | 40 | 51 | 61 |
| 12 | 12 | 14 | 19 | 26 | 58 | 60 | 55 |
| 14 | 13 | 16 | 24 | 40 | 57 | 69 | 56 |
| 14 | 17 | 22 | 29 | 51 | 87 | 80 | 62 |
| 18 | 22 | 37 | 56 | 68 | 109 | 103 | 77 |
| 24 | 36 | 55 | 64 | 81 | 104 | 113 | 92 |
| 49 | 64 | 78 | 87 | 103 | 121 | 120 | 101 |
| 72 | 92 | 95 | 98 | 112 | 100 | 103 | 99 |

(a)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 17 | 20 | 21 | 30 | 34 | 56 | 63 |
| 18 | 20 | 20 | 26 | 28 | 51 | 61 | 55 |
| 19 | 20 | 21 | 26 | 33 | 58 | 69 | 55 |
| 26 | 26 | 26 | 30 | 46 | 87 | 86 | 66 |
| 31 | 33 | 36 | 40 | 46 | 96 | 100 | 73 |
| 40 | 35 | 46 | 62 | 81 | 100 | 111 | 91 |
| 46 | 66 | 76 | 86 | 102 | 121 | 120 | 101 |
| 68 | 90 | 90 | 96 | 113 | 102 | 105 | 103 |

(b)

(c)

(d)

Fig. 1. Examples of quantization table and the corresponding compressed JPEG images. (a) JPEG default quantization table at quality factor equal to 50; (b) Optimized quantization table with the optimization goal of MS-SSIM; (c) JPEG image using default quantization table at QF = 10, 0.234 bbp, PSNR = 30.45, SSIM = 0.819, MS-SSIM = 0.946; (d) JPEG image using MS-SSIM optimized quantization table, 0.226 bpp, PSNR = 30.49, SSIM = 0.818, MS-SSIM = 0.953.

In [43] and [44], the researchers proposed the image dependent quantization table optimization based on the signal-fidelity based metric, MSE, which achieved significant bit-rate saving at the same quality measured by PSNR. However, these optimization strategies cannot ensure the same visual quality improvement due to the poor correlation between the perceptual quality and PSNR. In [5], [7], and [45], the researchers introduced SSIM and its variants into image and video coding to optimize the rate distortion process, but the performance improvement is not so satisfying yet. Channappayya *et al.* [5] derived the upper and lower bounds on the average SSIM index as a function of quantization rate for different source distributions *e.g.*, uniform, Gaussian and Laplacian distributions, for the first time. Wang *et al.* [7] utilized SSIM as the quality metric in rate distortion optimization instead of the MSE and achieved about 5%-10% bit-rate saving compared with original H.264/AVC. Ou *et al.* [45], applied SSIM to perceptual rate control problem achieving 0.008 SSIM gain (corresponding to 14% bitrate saving). From these work, we can see that the quality improvements are still small.

In essence, regarding the perceptual-based image compression, although various encoding optimization strategies can improve the image quality at the same bit-rate level, the quality fluctuations are usually limited within a small range. However, most of the traditional IQA databases only contain coarse-grained compression distortion levels, and they cannot well evaluate IQA algorithms on the fine-grained quality prediction for image compression problem. For example, the JPEG

Fig. 2. Sample reference images in the FG-IQA database.

images in Fig.1(c) and 1(d) are compressed at the similar bitrates using the scaled quantization tables in Fig.1(a) and 1(b), respectively. Although the image in Fig.1(d) shows fewer blocking artifacts, it has a lower SSIM value but higher PSNR and MS-SSIM values compared to the image shown in Fig.1(c). These different IQA algorithms show opposite quality rankings on the fine-grained distortion levels, which motivates us to revisit the existing IQA algorithms and investigate their appropriateness in distinguishing fine-grained distortions. In view of this, we construct the first fine-grained IQA database (FG-IQA) in large scale for compressed images to advance the development of the fine-grained IQA and perceptual-based image compression.

## III. The Fine-Grained Database Construction for Image Quality Assessment

### A. Data Preparation and Processing

The scale and content diversity are two important factors for a database to better explore the visual quality problem and evaluate the existing IQA algorithms [46]–[50]. Different from the previous database which only contains 20-30 reference images, we take advantage of plentiful image content in the Waterloo Exploration Database and carefully select 100 reference images from them, which contain men, women, buildings, indoor/outdoor scenes, cars, airplanes, statue, food, animals and plants, etc. Some examples of these reference images are shown in Fig.2. The resolutions of these images range from $400 \times 400$ to $723 \times 480$.

In the proposed FG-IQA database, we focus on the block-based compression distortions, where JPEG compression is utilized. We utilized the JPEG codec developed by the Independent JPEG Group [51] with four categories of quantization tables generated according to different principles. The first category is the default table of JPEG standard (denoted as $T^{(J)}$) and one example is shown in Fig.1(a) at QF equal to 50, and for the other compression ratio, the corresponding quantization tables are derived based on the following equations,

$$s = (QF < 50)\ ?\ (5000/QF)\ :\ (200 - 2QF), \quad (1)$$

$$T_{QF}^{(J)} = (s * T_{50}^{(J)} + 50)/100. \quad (2)$$

TABLE III
The Bitrate Distribution for Three Scenarios

| | $T^{(J)}$ | $T^{(U)}$ | $T^{(P)}$ | $T^{(M)}$ | STD of bitrate |
|---|---|---|---|---|---|
| Bitrate $b_1$ | 1.9615 | 1.9614 | 1.9615 | 1.9721 | 0.0047 |
| Bitrate $b_2$ | 2.3756 | 2.3754 | 2.3757 | 2.3823 | 0.0037 |
| Bitrate $b_3$ | 2.6734 | 2.6716 | 2.6735 | 2.6816 | 0.0053 |

The second category of quantization table denoted as $T^{(U)}$ is a uniform matrix, which is also widely utilized in compression. The third category of quantization table denoted as $T^{(P)}$ is the derived for individual images based on the rate-distortion optimization according to PSNR [43], and the corresponding source code can be downloaded from website.[1] The last category of quantization table is an optimized one according to MS-SSIM denoted as $T^{(M)}$, which is implemented in the open source JPEG codec, *libjpeg* [42].

For each reference image, we first apply the JPEG codec with default quantization table to compress reference images into three target quality levels with QF equal to 10, 30 and 50, corresponding to low, middle and high bit-rate scenarios, respectively. The corresponding three bitrates are denoted as $b_1$, $b_2$ and $b_3$. For each bitrate, we apply the other three quantization table derivation methods by exhaustively searching all possible quantization tables in their own quantization space. As such, the corresponding optimal ones with the closest bitrate for each compressed image using the JPEG default quantization table can be obtained. For example, one reference image is firstly compressed at bitrate $b_1$ using JPEG default quantization table. Then, for the quantization table $T^{(U)}$, we compress the reference image using the uniform quantization table with the elements from 1 to 255 and compare its bitrate with $b_1$. The compressed image using $T^{(U)}$ is selected when it has the minimum bitrate deviation from $b_1$. Similar process is applied to other categories of quantization table derivation methods. For all three bitrate scenarios, we can generate all the JPEG compressed images with fine-grained distortion levels at each target bitrate.

After the process, the average bitrate deviations from the target ones are 0.007%, 0.002% and 0.54% at $b_1$, 0.01%, 0.0007% and 0.28% at $b_2$, 0.07%, 0.003% and 0.31% at $b_3$ for $T^{(U)}$, $T^{(P)}$ and $T^{(M)}$ on average respectively. The bitrate for most distorted images using non-default quantization table is within 0.5% deviation from the target ones generated by JPEG default quantization table. The bitrate distribution of the proposed FG-IQA database is illustrated in Table III. Here, the STD of bitrate means the average of bitrate standard deviation for every four distorted images at each bitrate case. We can see that each reference image is compressed by different quantization tables into very approximate bitrate, which can well simulate the practical circumstances in perceptual-based image coding. The overall bitrate is relative large because the chroma components consume too many bits due to the all 1 quantization table, and the actual bitrates for luminance component at 0.26bpp, 0.64bpp and 0.92bpp.

[1] http://pages.cs.wisc.edu/~ratnakar/rdopt.html

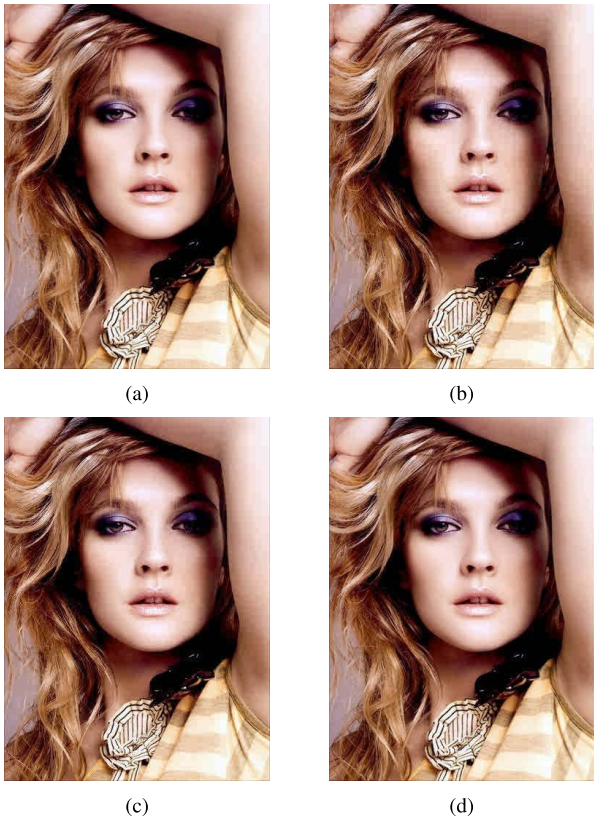


Fig. 3. Examples of distorted images compressed by JPEG at bitrate $b_1$, using quantization tables $T^{(J)}$, $T^{(U)}$, $T^{(P)}$ and $T^{(M)}$ for (a)∼(d), respective. (a) PSNR = 31.2154, SSIM = 0.9085, MS-SSIM = 0.9890, VIF = 0.6122, bitrate = 1.5891 bpp, (b) PSNR = 31.8654, SSIM = 0.9039, MS-SSIM = 0.9809, VIF = 0.4950, bitrate = 1.5904 bpp, (c) PSNR = 31.5465, SSIM = 0.9041, MS-SSIM = 0.9834, VIF = 0.5053, bitrate = 1.5878 bpp and (d) PSNR = 31.5892, SSIM = 0.9109, MS-SSIM = 0.9887, VIF = 0.5981, bitrate = 1.5895 bpp.

Since HVS is more sensitive to the luminance and most IQA algorithms are only applied on luminance component, we apply different quantization tables to luminance component. For the quantization table of the chroma components, all elements are set as 1, and this avoids the influence of the compression distortions in chroma components. In addition, although the distortions only generated in luminance components, but the distortions can be spread to all the R, G, B components due to the conversion of luminance values, Y = 16+0.2568R+0.5041 G+0.0979B. Therefore, we adopt the color images for the database, which are more popular than gray images. Fig.3 shows some images at bitrate $b_2$ with the four quantization methods. We can see that although these images are almost with the same bitrate, they show different perceptual qualities, where the subjective quality of Fig.3(a) with fewer blocking artifacts is obviously better than that of Fig.3(b). As such, there are quality differences but it is difficult to distinguish them using the single stimulus categorical rating method in subjective experiment as that in most of the IQA database.

### B. Subjective Experiments

To accurately distinguish the qualities of these images with fine-grained difference, we take the ordering by

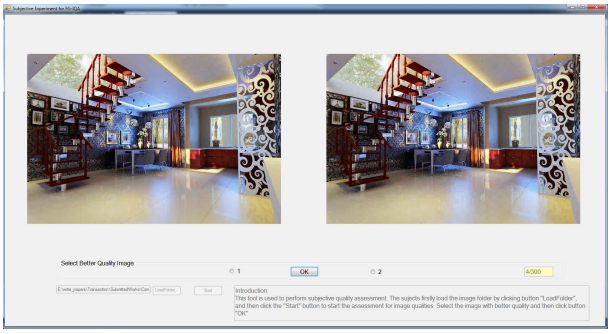


Fig. 4. Screen-shot of the software used in our subjective experiments. Each pair images corresponds to the same reference image compressed by JPEG into approximate bitrate using differen optimization methods.

force-choice pairwise comparison method [37] in our subjective experiment, which is more reliable and stable than individual single-stimulus evaluations by reducing the variations among researchers due to no reference for quality standard. In the subjective experiments, we invite 30 undergraduate and graduate students as subjects. The distorted images in each bitrate case are divided into two batches equally. Then, each batch contains 200 distorted images generated from 50 reference images. Then, the subjective experiments can be performed batch by batch, and it needs about 40-50 minutes to finish all the judgements in one batch. In the experiment, one subject can finish the evaluation of several batches at different time. This strategy can further reduce the instability due to fatigue.

To ensure the accuracy, we compare all the combinations for the distorted images corresponding to the same reference image instead of using the balanced incomplete block designs to reduce the trials [52]. In each subjective experiment, the subjects are shown a pair of distorted images corresponding to the same reference image compressed by JPEG with different optimization methods. The subjects are asked to view the two distorted images with a specified viewing distance (around 2-2.5 screen heights) and are always forced to choose one image with better quality even if they are difficult to distinguish them (*i.e.,* a forced-choice design). There are 6 comparisons for 4 compressed images corresponding to the same reference image, *i.e.,*

$$\Omega = \{(I_{T^{(J)}}, I_{T^{(U)}}), (I_{T^{(J)}}, I_{T^{(P)}}), (I_{T^{(J)}}, I_{T^{(M)}}), (I_{T^{(U)}}, I_{T^{(P)}}), (I_{T^{(U)}}, I_{T^{(M)}}), (I_{T^{(M)}}, I_{T^{(P)}})\}, \quad (3)$$

where $I_{T^{(*)}}$ represents the compressed image $I$ using the quantization table $T^{(*)}$ at the same target bitrate. Therefore, each batch needs to perform 300 comparison trials, the order of which is randomly generated. The GUI of the developed subjective software is shown in Fig.4, which is run on Windows OS PC with screen resolution 1920 × 1080 placed in laboratory with normal indoor lighting.

### C. FG-IQA Database Summary

After the subjective experiments, we further process the subjective results and organize the database to be convenient for evaluate. In our subjective experiments, there are 600 comparisons for 400 distorted images at each bitrate scenario, in total

TABLE IV

THE PREFERENCE NUMBER OF PAIRWISE COMPARISONS AT DIFFERENT BITRATE SCENARIOS, AND THE STATISTICAL DISTRIBUTION OF THE IMAGES WITH BETTER QUALITY

| bit-rate | preference probability | preference number | $T^{(J)}$ vs. $T^{(U)}$ | $T^{(J)}$ vs. $T^{(P)}$ | $T^{(J)}$ vs. $T^{(M)}$ | $T^{(U)}$ vs. $T^{(P)}$ | $T^{(U)}$ vs. $T^{(M)}$ | $T^{(M)}$ vs. $T^{(P)}$ |
|---|---|---|---|---|---|---|---|---|
| b1 | >90% | 276 | (24,0) | (4,13) | (0,52) | (1,61) | (0,89) | (32,0) |
| | 80%-90% | 119 | (24,0) | (2,14) | (0,18) | (1,26) | (0,7) | (27,0) |
| | 70%-80% | 99 | (26,0) | (7,16) | (0,16) | (1,6) | (0,3) | (23,1) |
| | 60%-70% | 61 | (13,3) | (10,10) | (0,11) | (0,2) | (0,1) | (10,1) |
| | 50%-60% | 45 | (6,4) | (13,11) | (1,2) | (0,2) | (0,0) | (3,3) |
| | Overall | 600 | (93,7) | (36,64) | (1,99) | (3,97) | (0,100) | (95,5) |
| b2 | >90% | 134 | (31,0) | (10,0) | (0,0) | (0,26) | (0,53) | (14,0) |
| | 80%-90% | 150 | (35,0) | (25,0) | (1,5) | (0,30) | (0,22) | (32,0) |
| | 70%-80% | 116 | (15,0) | (26,0) | (2,14) | (0,21) | (0,15) | (23,0) |
| | 60%-70% | 117 | (12,1) | (21,2) | (16,25) | (1,12) | (1,5) | (20,1) |
| | 50%-60% | 83 | (1,5) | (10,6) | (11,26) | (3,7) | (2,2) | (4,6) |
| | Overall | 600 | (94,6) | (92,8) | (30,70) | (4,96) | (3,97) | (93,7) |
| b3 | >90% | 97 | (31,0) | (22,0) | (0,0) | (0,2) | (0,24) | (18,0) |
| | 80%-90% | 92 | (21,0) | (15,0) | (2,0) | (1,14) | (1,21) | (17,0) |
| | 70%-80% | 137 | (19,0) | (31,0) | (9,0) | (1,30) | (0,22) | (25,0) |
| | 60%-70% | 154 | (19,0) | (21,2) | (35,6) | (2,25) | (1,19) | (22,2) |
| | 50%-60% | 120 | (6,4) | (2,7) | (25,23) | (7,18) | (6,6) | (14,2) |
| | Overall | 600 | (96,4) | (91,9) | (71,29) | (11,89) | (8,92) | (96,4) |

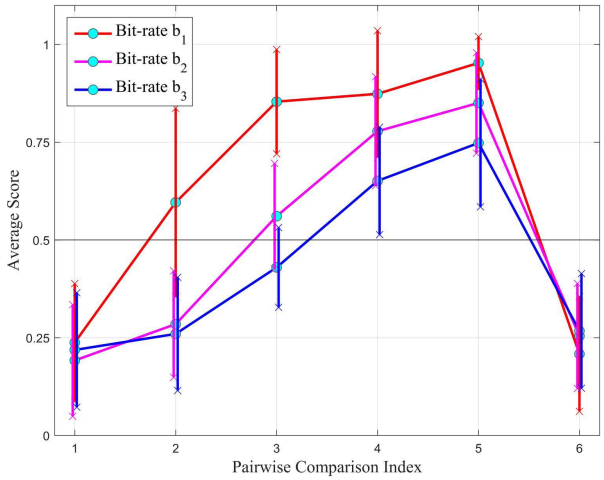



Fig. 5. The score distribution of the FG-IQA database at different bitrate scenarios.

1800 comparisons for the three bitrate scenarios. For each comparison, 30 subjects are required to provide force choice for better quality image, which leads to 54,000 subjective experiments in total.

The final statistical results of the subjective experiments for the three bitrate scenarios are illustrated in Table IV and Fig.5. In Table IV, the preference probability reflects the reliability of pair-wise comparisons, which is calculated based on the percentage of subjects for the image with better quality. For example, the preference probability >90% means that for one comparison, more than 90% of the subjects do the same selection. The "preference number" column indicates the selection amount corresponding to the preference probability. The $4^{th} \sim 9^{th}$ columns show the image distribution with better quality between the every comparison. For example, the element (24,0) in the $2^{th}$ row and $4^{th}$ means that there are 24 images with better quality compressed by JPEG using quantization table $T^{(J)}$ and 0 images compressed by JPEG using $T^{(U)}$, and the preference probability of these comparison results are beyond 90%.

From Table IV, we can see that for $b_1$ case, there are 276, 119, 99, 61 and 45 comparisons with their preference probability more than 90%, 80%, 70% and 60% respectively, while for $b_2$, the corresponding numbers are 134, 150, 116, 117 and 83, and for $b_3$ the corresponding numbers are 97, 92, 137, 154 and 120. These results show that although these images are compressed into close bitrates with fine-grained quality difference, subjects can still perceive the quality variations with high probability. In addition, at low bitrate, the fine-grained quality difference is easier to be perceived compared with that at high bitrate.

To show the image distribution with better quality directly, we assign a score for each comparison in Eq.(3), *i.e.,* 1 is assigned when the first image in a pair has better quality, while 0 is assigned when the second image has better quality. Fig.5 shows the image distribution with better quality, where the horizontal axis represents the index of the six combinations in Eq.3, and the average scores of the vertical axis are calculated as,

$$s_i = (0 * p_{i,0} + 1 * p_{i,1}), \tag{4}$$

$$\mu = \frac{1}{100} \sum_{i=1}^{100} s_i, \tag{5}$$

where $p_{i,0}$ and $p_{i,1}$ are the subject percentages for the $i^{th}$ comparison pair with scores 0 and 1 respectively at a given bitrate. The error bars in Fig.5 represent the standard deviation

of the scores in Eq.4, which is calculated as,

$$\sigma = \sqrt{\frac{1}{100}\sum_{i=1}^{100}(s_i - \mu)^2}. \quad (6)$$

From the results, we can see that the perceptual quality of the compressed images using JPEG default quantization table and MS-SSIM based optimized quantization table is obviously better than those using uniform quantization table and PSNR based optimized quantization table respectively for all three bitrate scenarios. At low bitrate $b_1$, the quality of the compressed images using PSNR and MS-SSIM based optimized quantization tables are better than that of compressed images with JPEG default quantization table and uniform quantization table respectively. However, at middle and high bitrate scenarios, $b_2$ and $b_3$, the perceptual quality of compressed images using default quantization table is superior to that of compressed images using PSNR based optimized quantization table, and even better than that using MS-SSIM based optimized quantization table at high bitrate scenario $b_3$.

Moreover, we can see that the existing quantization table optimization techniques both targeting for PSNR and MS-SSIM still cannot achieve the optimal perceptual quality. One of the important reasons may be that the existing image quality assessment methods can only predict the image perceptual quality in coarse levels, but they are not well correlated with HVS in fine-grained level. The performances of these optimized image compression methods targeting for different IQA metrics are unstable and inconsistent. Therefore, more efficient IQA algorithms and databases are demanded to further advance the perceptual-based image processing, especially for the image compression application.

## IV. Benchmark Analysis and Experimental Results

### A. Evaluation Methods for IQA Models

One of the main objectives of the FG-IQA database is to evaluate the performance of different IQA models, *i.e.,* how well an objective metric agrees with subjective preferences of subjects. In the IQA field, the Kendall rank correlation coefficient (KRCC) [53], Spearman rank-order correlation coefficient (SROCC) [54] and Pearson linear correlation coefficient (PLCC) are three widely used measures to evaluate the IQA metrics. Herein, the KRCC is calculated as,

$$KRCC = \frac{n_c - n_d}{0.5n(n-1)}, \quad (7)$$

where $n$ is the length of the ranking ($n = 4$ for our database), $n_c$ is the number of concordant pairs and $n_d$ is the number of discordant pairs over all pairs of entries in the ranking. The SROCC measures the monotonic relationship between two vectors, *e.g.,* MOS and the objective scores calculated from IQA algorithms, and PLCC measures their linear correlation. Herein, we take the Bradley-Terry model [55] to derive the MOS from pairwise comparison results. The three correlation coefficients are in the range of $[-1, 1]$, and the larger coefficients correspond to more consistency between two vectors. More specially, the correlation coefficients are equal to 1 in case of the perfect agreement and $-1$ indicates the case of perfect disagreement. In the case of correlation coefficients equalling to 0, the rankings are considered to be independent. In particular, since for each image content there are only 4 compressed versions, the amount is not enough to perform the nonlinear regression with four or five parameters [26], [56] before calculating PLCC as the common ways in the IQA field. Therefore, we directly calculated the PLCC without performing the regression process in this paper.

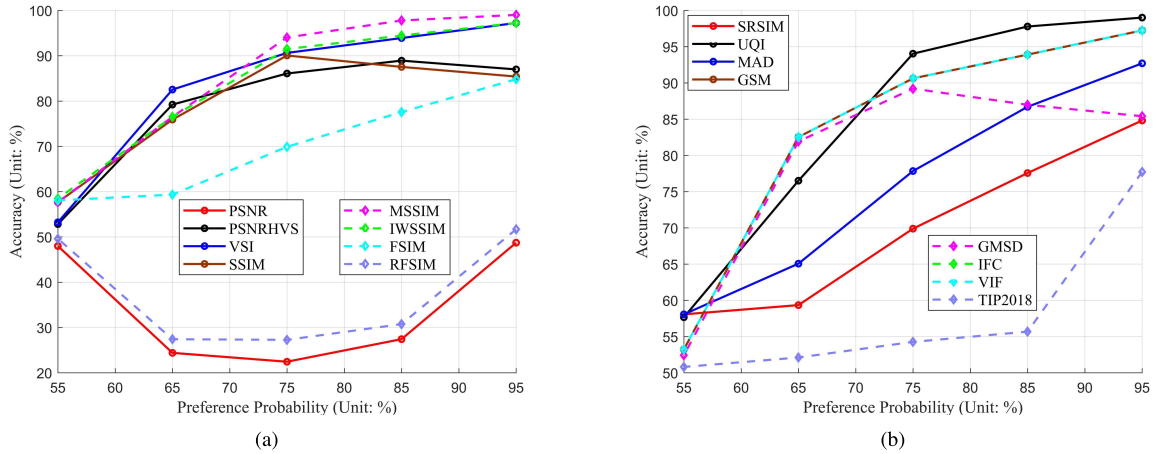
### B. Benchmark Analyses and Experimental Results

To analyze the efficiency of IQA algorithms, we apply 15 state-of-the-art full reference image quality assessment methods on the proposed FG-IQA database, to investigate their performance and demonstrate the new challenges in fine-grained image quality assessment problem. The FR-IQA algorithms include PSNR, PSNR-HVS (PSNR with HVS properties) [57], VSI [58], SSIM [11], IW-SSIM [14], MS-SSIM [12], FSIM [13], RFSIM [59], SR-SIM [60], UQI [61], GSM [15], GSMD [16], VIF [17], IFC [18], MAD (Most Apparent Distortion) [35], and one CNN based IQA method, WaDIQaM-FR [62]. The implementations of all algorithms are obtained from the authors or public websites. Here, the CNN model for WaDIQaM-FR is trained on TID2008 images.

First, we evaluate the pairwise preference consistency using the classic correlation coefficients KRCC, SROCC and PLCC, as shown in Table V. The KRCC, SROCC and PLCC are the average values for the distorted images of the same reference image, and the top 2 correlation coefficient values are highlighted. We can see that the PSNR is poorly correlated with human perceptual quality, and even contrary to subjective results on average for different bitrate scenarios. Combining the HVS features, the PSNR-HVS and VSI achieve more consistent results with the subjective results. More interestingly, the SSIM and its variations also show diverse results on the FG-IQA database. Although the SSIM achieves good correlation with human perceptual quality on existing coarse-grained databases, it is poorly correlated with human perceptual quality in fine-grained quality assessment, especially at high bitrate coding scenario. However, the variations of SSIM achieve much better correlated results with subjective results, especially MS-SSIM and IWSSIM, which introduce additional HVS features to improve the performance of SSIM. The MS-SSIM takes advantage of the multi-scale SSIM to capture the contrast sensitivity characteristics of HVS, *i.e.,* the contrast sensitivity decreases along both high- and low-frequency directions. The IWSSIM exploits the information content to weight local SSIM values in the pooling stage, which also utilizes the same HVS features by calculating the weights in Laplacian pyramid decomposition images with five scales. The other methods also achieve comparable performance but these correlation coefficients are obviously lower than those on the existing databases with coarse-grained distortion levels. Specifically, the performance of WaDIQaM-FR is also poor on the proposed database, and this may because the training data with coarse-grained quality levels is different from the proposed database with fine-grained quality levels.

TABLE V
THE KRCC, SROCC AND PLCC FOR DIFFERENT IQA ALGORITHMS AT DIFFERENT BITRATE SCENARIOS

| IQA methods | KRCC | | | | SROCC | | | | PLCC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $b_1$ | $b_2$ | $b_3$ | Average | $b_1$ | $b_2$ | $b_3$ | Average | $b_1$ | $b_2$ | $b_3$ | Average |
| PSNR | 0.3733 | -0.5567 | -0.7433 | -0.3089 | 0.4521 | -0.6943 | -0.8373 | -0.3598 | 0.6054 | -0.6502 | -0.8592 | -0.3013 |
| PSNR-HVS | 0.5133 | 0.6800 | **0.7833** | 0.6589 | 0.5929 | 0.7963 | **0.8593** | 0.7495 | 0.6052 | 0.8369 | **0.8674** | 0.7698 |
| VSI | 0.6800 | 0.6833 | 0.6200 | 0.6611 | 0.7608 | 0.7755 | 0.6769 | 0.7377 | 0.6677 | 0.8390 | 0.6570 | 0.7213 |
| SSIM | 0.5567 | 0.2133 | -0.2767 | 0.1644 | 0.6461 | 0.2095 | -0.3917 | 0.1546 | 0.6752 | 0.1858 | -0.4635 | 0.1325 |
| MS-SSIM | **0.8333** | **0.8133** | 0.6967 | **0.7811** | **0.8896** | **0.8795** | 0.7758 | **0.8483** | **0.8106** | **0.8715** | 0.8184 | **0.8335** |
| IWSSIM | 0.8033 | **0.8200** | 0.6367 | **0.7533** | 0.8744 | **0.8755** | 0.7378 | **0.8292** | 0.7953 | **0.8717** | 0.8078 | **0.8250** |
| FSIM | 0.6567 | 0.6700 | 0.3733 | 0.5667 | 0.7317 | 0.7715 | 0.3746 | 0.6259 | 0.6664 | 0.7557 | 0.3327 | 0.5849 |
| RFSIM | 0.6933 | 0.3167 | 0.1200 | 0.3767 | 0.7900 | 0.3037 | 0.1102 | 0.4013 | 0.7425 | 0.3211 | 0.1467 | 0.4034 |
| SRSIM | 0.6233 | 0.4667 | 0.3933 | 0.4944 | 0.7177 | 0.5232 | 0.3547 | 0.5318 | 0.6546 | 0.5800 | 0.3056 | 0.5134 |
| UQI | 0.6867 | 0.6800 | 0.5833 | 0.6500 | 0.7612 | 0.7143 | 0.6374 | 0.7043 | 0.7528 | 0.7161 | 0.6455 | 0.7048 |
| MAD | **0.8700** | 0.7167 | 0.5300 | 0.7056 | **0.9296** | 0.8135 | 0.6048 | 0.7826 | **0.9156** | 0.8278 | 0.5974 | 0.7802 |
| GSM | 0.6867 | 0.7400 | 0.7300 | 0.7189 | 0.7748 | 0.8215 | 0.8133 | 0.8032 | 0.6928 | 0.8574 | 0.8000 | 0.7834 |
| GSMD | 0.6633 | 0.6800 | **0.7833** | 0.7089 | 0.7435 | 0.7943 | **0.8593** | 0.7991 | 0.6645 | 0.8531 | 0.8420 | 0.7866 |
| IFC | 0.7200 | 0.6867 | 0.7733 | 0.7267 | 0.8069 | 0.8003 | 0.8533 | 0.8202 | 0.7204 | 0.8631 | **0.8544** | 0.8127 |
| VIF | 0.7167 | 0.6767 | 0.7733 | 0.7222 | 0.8009 | 0.7943 | 0.8493 | 0.8148 | 0.6994 | 0.8596 | 0.8463 | 0.8018 |
| WaDIQaM-FR | 0.4300 | 0.4667 | 0.4033 | 0.4333 | 0.4789 | 0.5743 | 0.4672 | 0.5068 | 0.5406 | 0.5951 | 0.4994 | 0.5450 |




(a) (b)

Fig. 6. Preference consistency ratios of different IQA algorithms on the proposed FG-IQA database at different preference probability ranges.

For the three correlation coefficients, these IQA methods shows similar characteristics. As a whole, MS-SSIM and IW-SSIM achieve top 2 performance for most cases, and the GSMD and IFC achieve better results at high bitrate scenario while MAD performs better at low bitrate case.

In addition, we analyze the relationship among different IQA algorithms by calculating their KRCC in Table VII and Table VIII to reveal their reliability in FG-IQA. From Table VII, we can see that these IQA algorithms show poor correlation, and there are even some negative values between IQA algorithms, *e.g.,* PSNR&PSNR-HVS, PSNR&IWSSIM, PSNR&UQI, PSNR-HVS&MAD and SSIM&IFC etc. Moreover, we also calculate the KRCC for the objective quality of every combination of these IQA algorithms for images at all bitrate scenarios. The distortion levels are distinct among different bitrate scenarios, and there are both fine-grained and coarse-grained distortion levels in the database combined by all the images. From Table VIII, we can see that all the IQA algorithms illustrate positive correlation, and even high correlation (up to 0.86 is observed) for many IQA algorithms. These results further prove that although existing IQA algorithms have achieved great success in predicting the human perceptual quality for coarse-grained quality differences, they

TABLE VI
THE KRCC FOR DIFFERENT IQA ALGORITHMS ON DIFFERENT TYPES OF IMAGES

| Image Types | IQA methods | $b_1$ | $b_2$ | $b_3$ | Average |
|---|---|---|---|---|---|
| Portrait Images | MS-SSIM | 0.8889 | 0.8889 | 0.7778 | 0.8519 |
| | IWSSIM | 0.8000 | 0.9111 | 0.6222 | 0.7778 |
| | IFC | 0.7111 | 0.7333 | 0.8222 | 0.7556 |
| | VIF | 0.7111 | 0.7111 | 0.8222 | 0.7481 |
| Building Images | MS-SSIM | 0.8222 | 0.9111 | 0.6889 | 0.8074 |
| | IWSSIM | 0.8000 | 0.9333 | 0.6889 | 0.8074 |
| | IFC | 0.7556 | 0.7333 | 0.8889 | 0.7926 |
| | VIF | 0.7333 | 0.7333 | 0.8889 | 0.7852 |

TABLE VII

THE KRCC OF THE OBJECTIVE QUALITY BETWEEN DIFFERENT IQA ALGORITHMS FOR JPEG IMAGES USING $T^{(J)}$ AT BITRATE $b_2$

| | PSNR | PSNR-HVS | VSI | SSIM | MS-SSIM | IWSSIM | FSIM | RFSIM | SRSIM | UQI | MAD | GSM | GSMD | IFC | VIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | 1.000 | -0.173 | 0.253 | 0.424 | 0.179 | -0.054 | 0.385 | 0.295 | 0.181 | -0.311 | -0.083 | 0.066 | 0.420 | -0.570 | 0.157 |
| PSNR-HVS | -0.173 | 1.000 | -0.026 | 0.264 | 0.431 | 0.399 | -0.009 | -0.104 | -0.024 | 0.006 | -0.211 | 0.020 | 0.142 | 0.215 | 0.289 |
| VSI | 0.253 | -0.026 | 1.000 | 0.285 | 0.212 | 0.098 | 0.272 | 0.676 | 0.797 | -0.089 | -0.029 | 0.720 | 0.379 | -0.147 | 0.102 |
| SSIM | 0.424 | 0.264 | 0.285 | 1.000 | 0.676 | 0.389 | 0.485 | 0.237 | 0.240 | -0.122 | -0.183 | 0.170 | 0.640 | -0.257 | 0.438 |
| MS-SSIM | 0.179 | 0.431 | 0.212 | 0.676 | 1.000 | 0.651 | 0.371 | 0.163 | 0.203 | 0.087 | -0.128 | 0.181 | 0.559 | 0.024 | 0.618 |
| IWSSIM | -0.054 | 0.399 | 0.098 | 0.389 | 0.651 | 1.000 | 0.173 | 0.062 | 0.137 | 0.245 | -0.114 | 0.155 | 0.257 | 0.322 | 0.666 |
| FSIM | 0.385 | -0.009 | 0.272 | 0.485 | 0.371 | 0.173 | 1.000 | 0.213 | 0.244 | 0.086 | 0.203 | 0.204 | 0.526 | -0.214 | 0.221 |
| RFSIM | 0.295 | -0.104 | 0.676 | 0.237 | 0.163 | 0.062 | 0.213 | 1.000 | 0.657 | -0.099 | -0.087 | 0.584 | 0.260 | -0.170 | 0.113 |
| SRSIM | 0.181 | -0.024 | 0.797 | 0.240 | 0.203 | 0.137 | 0.244 | 0.657 | 1.000 | -0.002 | 0.013 | 0.737 | 0.348 | -0.025 | 0.124 |
| UQI | -0.311 | 0.006 | -0.089 | -0.122 | 0.087 | 0.245 | 0.086 | -0.099 | -0.002 | 1.000 | 0.349 | 0.141 | -0.050 | 0.523 | 0.133 |
| MAD | -0.083 | -0.211 | -0.029 | -0.183 | -0.128 | -0.114 | 0.203 | -0.087 | 0.013 | 0.349 | 1.000 | 0.054 | 0.035 | 0.096 | -0.168 |
| GSM | 0.066 | 0.020 | 0.720 | 0.170 | 0.181 | 0.155 | 0.204 | 0.584 | 0.737 | 0.141 | 0.054 | 1.000 | 0.289 | 0.065 | 0.116 |
| GSMD | 0.420 | 0.142 | 0.379 | 0.640 | 0.559 | 0.257 | 0.526 | 0.260 | 0.348 | -0.050 | 0.035 | 0.289 | 1.000 | -0.232 | 0.309 |
| IFC | -0.570 | 0.215 | -0.147 | -0.257 | 0.024 | 0.322 | -0.214 | -0.170 | -0.025 | 0.523 | 0.096 | 0.065 | -0.232 | 1.000 | 0.160 |
| VIF | 0.157 | 0.289 | 0.102 | 0.438 | 0.618 | 0.666 | 0.221 | 0.113 | 0.124 | 0.133 | -0.168 | 0.116 | 0.309 | 0.160 | 1.000 |

TABLE VIII

THE KRCC OF THE OBJECTIVE QUALITY BETWEEN DIFFERENT IQA ALGORITHMS FOR JPEG IMAGES USING $T^{(J)}$ AT ALL THE THREE BITRATES

| | PSNR | PSNR-HVS | VSI | SSIM | MS-SSIM | IWSSIM | FSIM | RFSIM | SRSIM | UQI | MAD | GSM | GSMD | IFC | VIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | 1.000 | 0.438 | 0.451 | 0.623 | 0.503 | 0.416 | 0.566 | 0.468 | 0.369 | 0.186 | 0.377 | 0.403 | 0.569 | 0.159 | 0.490 |
| PSNR-HVS | 0.438 | 1.000 | 0.456 | 0.616 | 0.715 | 0.717 | 0.655 | 0.365 | 0.419 | 0.471 | 0.529 | 0.504 | 0.697 | 0.596 | 0.752 |
| VSI | 0.451 | 0.456 | 1.000 | 0.492 | 0.497 | 0.455 | 0.523 | 0.745 | 0.859 | 0.287 | 0.396 | 0.837 | 0.566 | 0.335 | 0.476 |
| SSIM | 0.623 | 0.616 | 0.492 | 1.000 | 0.781 | 0.659 | 0.698 | 0.442 | 0.435 | 0.355 | 0.420 | 0.491 | 0.727 | 0.365 | 0.668 |
| MS-SSIM | 0.503 | 0.715 | 0.497 | 0.781 | 1.000 | 0.852 | 0.747 | 0.420 | 0.467 | 0.509 | 0.536 | 0.543 | 0.805 | 0.563 | 0.831 |
| IWSSIM | 0.416 | 0.717 | 0.455 | 0.659 | 0.852 | 1.000 | 0.701 | 0.378 | 0.448 | 0.578 | 0.567 | 0.535 | 0.735 | 0.687 | 0.860 |
| FSIM | 0.566 | 0.655 | 0.523 | 0.698 | 0.747 | 0.701 | 1.000 | 0.436 | 0.493 | 0.518 | 0.670 | 0.562 | 0.819 | 0.499 | 0.727 |
| RFSIM | 0.468 | 0.365 | 0.745 | 0.442 | 0.420 | 0.378 | 0.436 | 1.000 | 0.726 | 0.224 | 0.313 | 0.684 | 0.453 | 0.254 | 0.409 |
| SRSIM | 0.369 | 0.419 | 0.859 | 0.435 | 0.467 | 0.448 | 0.493 | 0.726 | 1.000 | 0.306 | 0.388 | 0.826 | 0.537 | 0.358 | 0.459 |
| UQI | 0.186 | 0.471 | 0.287 | 0.355 | 0.509 | 0.578 | 0.518 | 0.224 | 0.306 | 1.000 | 0.619 | 0.416 | 0.478 | 0.712 | 0.534 |
| MAD | 0.377 | 0.529 | 0.396 | 0.420 | 0.536 | 0.567 | 0.670 | 0.313 | 0.388 | 0.619 | 1.000 | 0.481 | 0.627 | 0.580 | 0.559 |
| GSM | 0.403 | 0.504 | 0.837 | 0.491 | 0.543 | 0.535 | 0.562 | 0.684 | 0.826 | 0.416 | 0.481 | 1.000 | 0.597 | 0.460 | 0.539 |
| GSMD | 0.569 | 0.697 | 0.566 | 0.727 | 0.805 | 0.735 | 0.819 | 0.453 | 0.537 | 0.478 | 0.627 | 0.597 | 1.000 | 0.505 | 0.762 |
| IFC | 0.159 | 0.596 | 0.335 | 0.365 | 0.563 | 0.687 | 0.499 | 0.254 | 0.358 | 0.712 | 0.580 | 0.460 | 0.505 | 1.000 | 0.632 |
| VIF | 0.490 | 0.752 | 0.476 | 0.668 | 0.831 | 0.860 | 0.727 | 0.409 | 0.459 | 0.534 | 0.559 | 0.539 | 0.762 | 0.632 | 1.000 |

are still inefficient in predicting human perceptual quality in fine-grained quality differences. Moreover, it is a more challenging problem for fine-grained image quality assessment.

Based on the research in [46]–[50], image content also has influence on subjective quality estimation. We select 15 portrait images and building images respectively from the proposed database, where the portrait images with simple structures compared with that of building images. We find that the average preference probability for the portrait images is 81%, while for building images it is 79%, which show that the quality assessment is easier for subjects on images with simple structures. We further explore the IQA performance on these images using MS-SSIM, IWSSIM, IFC and VIF, which perform better than others on the proposed database. Table VI shows their KRCC values on portrait and building images respectively. We can see that MS-SSIM performs much better on portrait images than that on building images since it applies low-pass filters to images to estimate quality with multi-scale images, which may filter out some high-frequency information in building images with complex structures. However, the information loss based IQA methods, IWSSIM, IFC and VIF, can well capture the complex structures especially in middle and high frequency, and achieve better performance on building images. Therefore, the quality assessment is also closely related with the image content, which is an important clue for the further IQA research.

To visualize the performance of IQA algorithms, the pairwise preference consistency ratios in terms of different subjective preference probability are illustrated in Fig.6(a) and Fig.6(b). Herein the pairwise preference consistency ratio is defined as [29],

$$P = \frac{M_c}{M}, \tag{8}$$

where $M$ is the amount of the image pairs, $M_c$ is the number of concordant pairs of an IQA model, *i.e.,* the accuracy ratio of the IQA model predicting the correct preference. In Fig.6(a), the GSM and GSMD achieve the same accuracy ratio, and the IFC and VIF are also the same accuracy on average. From the results, we can see that VSI, MS-SSIM, IWSSIM, FSIM, SRSIM, UQI, MAD, IFC and VIF achieve more consistent results with subjective results, where the accurate ratio increases along with the preference probability. However, PSNR, PSNR-HVS, SSIM, RFSIM, GSM and GSMD show accuracy degradation in high preference probability intervals to some extent, especially SSIM, GSM and GSMD showing obvious accuracy decrease when preference probability is beyond 80%. The PSNR and RFSIM metrics show more interesting results, as they almost have the opposite judgements against subjective results. Even in the highest preference probability interval, they only achieve 50% accuracy ratio, which almost approximates to random results. These results prove that some existing IQA models perform poorly in distinguishing the fine-grained distortion levels, which are feasible to determine by human visual system. Therefore, these metrics may not be suitable for perceptual-based image compression because the distortion differences between various coding modes are usually marginal. Moreover, the fine-grained image quality assessment is demanded and should be evaluated on the FG-IQA databases.

## V. Conclusion

In this paper, we have proposed a new and large scale IQA database with fine-grained distortion levels, targeting to advance the development for both the fine-grained quality assessment and perceptual-based image compression. The reliable subjective experiment with pair-wise comparison is utilized to rank the qualities of these distorted images. We also provided in-depth analyses for state-of-the-art IQA algorithms on the proposed FG-IQA database, and showed new challenges for fine-grained IQA. The FG-IQA database is made publicly available to facilitate future IQA research. The proposed database can be utilized to evaluate IQA algorithms and provides more accurate and comprehensive evaluation jointly with existing coarse-grained IQA databases. In addition, the latent factors influencing perceptual quality for images produced by the block-based image compression will also be investigated in our future work.

## Acknowledgment

The authors would like to thank Dingquan Li and Bin Deng of Peking University for their help in providing results of the CNN based image quality assessment method. They also would like to thank the three anonymous reviewers for their helpful comments and suggestions.

This work was carried out at the Rapid-Rich Object Search (ROSE) Laboratory, Nanyang Technological University, Singapore. The ROSE Laboratory was supported by the National Research Foundation, Singapore, under its Interactive Digital Media Strategic Research Programme.

## References

[1] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[2] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 1–14, Jan. 2018.

[3] H. Wang, X. Zhang, C. Yang, and C.-C. J. Kuo, "Analysis and prediction of JND-based video quality model," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2018, pp. 278–282.

[4] Y. Zhang, W. Lin, X. Zhang, Y. Fang, and L. Li, "Aspect ratio similarity (ARS) for image retargeting quality assessment," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1080–1084.

[5] S. S. Channappayya, A. C. Bovik, and R. W. Heath, Jr., "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.

[6] Z. Chen, W. Lin, and K. N. Ngan, "Perceptual video coding: Challenges and approaches," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2010, pp. 784–789.

[7] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[8] X. Zhang *et al.*, "Low-rank-based nonlocal adaptive loop filter for high-efficiency video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2177–2188, Oct. 2017.

[9] X. F. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao, "Just-noticeable difference-based perceptual optimization for JPEG compression," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 96–100, Jan. 2017.

[10] S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, "Nonlocal in-loop filter: The way toward next-generation video coding?" *IEEE MultiMedia*, vol. 23, no. 2, pp. 16–26, Apr./Jun. 2016.

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

[12] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[13] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

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

[15] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[16] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[17] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

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

[19] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[20] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[21] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.

[22] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2002, p. 1.

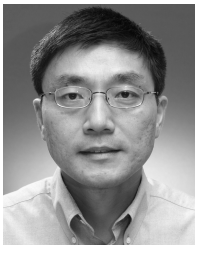
[23] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 1918–1927, Nov. 2005.

[24] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.

[25] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2*. Accessed: Oct. 2018. [Online]. Available: http://live.ece.utexas.edu/research/quality

[26] 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.
[27] N. Ponomarenko *et al.*, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[28] T. K. Tan *et al.*, "Video quality evaluation methodology and verification testing of HEVC compression performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 76–90, Jan. 2016.
[29] K. Ma *et al.*, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, Feb. 2017.
[30] S. Corchs, F. Gasparini, and R. Schettini, "No reference image quality classification for JPEG-distorted images," *Digit. Signal Process.*, vol. 30, pp. 86–100, Jul. 2014.
[31] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
[32] A. Ninassi, F. Autrusseau, and P. Le Callet, "Pseudo no reference image quality metric using perceptual data hiding," *Proc. SPIE*, vol. 6057, p. 60570G, Feb. 2006.
[33] L. Jin *et al.*, "Open access statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis," *Electron. Imag.*, vol. 2016, no. 13, pp. 1–9, 2016.
[34] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.
[35] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, 2010.
[36] F. De Simone, L. Goldmann, V. Baroncini, and T. Ebrahimi, "Subjective evaluation of JPEG XR image compression," *Proc. SPIE*, vol. 7443, p. 74430L, Sep. 2009.
[37] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Comput. Graph. Forum*, vol. 31, no. 8, pp. 2478–2491, 2012.
[38] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "CID2013: A database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.
[39] J. Y. Lin *et al.*, "Experimental design and analysis of JND test on coded image/video," *Proc. SPIE*, vol. 9599, p. 95990Z, Sep. 2015.
[40] Y. Horita. (2008). *MICT Image Quality Evaluation Database*. [Online]. Available: http://r0k.us/graphics/kodak/
[41] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. 18–34, Feb. 1992.
[42] (2016). *Libjpeg*. [Online]. Available: https://github.com/thorfdbg/libjpeg
[43] V. Ratnakar and M. Livny, "An efficient algorithm for optimizing DCT quantization," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 267–270, Feb. 2000.
[44] E. H. Yang and L. Wang, "Joint optimization of run-length coding, Huffman coding, and quantization table with complete baseline JPEG decoder compatibility," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 63–74, Jan. 2009.
[45] T.-S. Ou, Y.-H. Huang, and H. H. Chen, "SSIM-based perceptual rate control for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 682–691, May 2011.
[46] E. Allen, S. Triantaphillidou, and R. Jacobson, "Image quality comparison between JPEG and JPEG2000. I. Psychophysical investigation," *J. Imag. Sci. Technol.*, vol. 51, no. 3, pp. 248–258, 2007.
[47] S. Triantaphillidou, E. Allen, and R. Jacobson, "Image quality comparison between JPEG and JPEG2000. II. Scene dependency, scene analysis, and classification," *J. Imag. Sci. Technol.*, vol. 51, no. 3, pp. 259–270, 2007.
[48] E. Allen, S. Triantaphillidou, and R. Jacobson, "Perceptibility and acceptability of JPEG 2000 compressed images of various scene types," *Proc. SPIE*, vol. 9016, p. 90160W, Feb. 2014.
[49] S. E. Corchs, G. Ciocca, E. Bricolo, and F. Gasparini, "Predicting complexity perception of real world images," *PLoS ONE*, vol. 11, no. 6, p. e0157986, 2016.
[50] S. Corchs, G. Ciocca, and F. Gasparini, "Human perception of image complexity: Real scenes versus texture patches," *J. Alzheimer's Disease*, vol. 53, no. s1, 2016.
[51] (2016). *Independent JPEG Group*. [Online]. Available: http://www.ijg.org/
[52] H. Gulliksen and L. R. Tucker, "A general procedure for obtaining paired comparisons from multiple rank orders," *Psychometrika*, vol. 26, no. 2, pp. 173–183, Jun. 1961. [Online]. Available: https://doi.org/10.1007/BF02289713
[53] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, nos. 1–2, pp. 81–93, Jun. 1938.
[54] W. W. Daniel, *Applied Nonparametric Statistics*. Boston, MA, USA: Houghton Mifflin, 1978.
[55] D. R. Hunter, "MM algorithms for generalized Bradley–Terry models," *Ann. Statist.*, vol. 32, no. 1, pp. 384–406, 2004.
[56] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4286–4297, Sep. 2016.
[57] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *Proc. 2nd Int. Workshop Video Process. Quality Metrics*, vol. 4, 2006, pp. 1–4.
[58] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.
[59] L. Zhang, L. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 321–324.
[60] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. 19th IEEE Int. Conf. Image Process. (ICIP)*, Sep./Oct. 2012, pp. 1473–1476.
[61] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[62] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.

**Xinfeng Zhang** (M'16) received the B.S. degree in computer science from the Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. From 2014 to 2017, he was a Research Fellow with the Rapid-Rich Object Search Lab, Nanyang Technological University, Singapore. He is currently a Post-Doctoral Fellow with the School of Electrical Engineering System, University of Southern California, Los Angeles, CA, USA. His research interests include image and video processing, image and video compression.



**Weisi Lin** (M'92–SM'98–F'16) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from Kings College, London University, London, U.K., in 1992.

He was involved in teaching and research with Zhongshan University, Shantou University, Shantou, China, Bath University, Bath, U.K., the National University of Singapore, the Institute of Microelectronics, Singapore, and the Institute for Infocomm Research, Singapore. He was the Laboratory Head of Visual Processing and the Acting Department Manager of Media Processing, Institute for Infocomm Research. He is currently a Full Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.

**Shiqi Wang** received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from Peking University in 2014. From 2014 to 2016, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From 2016 to 2017, he was a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has authored over 40 technical proposals to ISO/MPEG, ITU-T, and AVS standards. His research interests include image/video compression, analysis and quality assessment.

**Jiaying Liu** (S'08–M'10–SM'17) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, in 2005, and the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, in 2010. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored over 100 technical articles in refereed journals and proceedings, and holds 28 granted patents. Her current research interests include image/video processing, compression, and computer vision.

She was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. She was a Visiting Researcher at Microsoft Research Asia in 2015 supported by the Star Track for Young Faculties. She has served as a TC Member at the IEEE CAS-MSA/EOT and APSIPA IVM and an APSIPA Distinguished Lecturer from 2016 to 2017. She is a CCF/IEEE Senior Member.

**SIWEI MA** received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He held a post-doctoral position with the University of Southern California, Los Angeles, CA, USA, from 2005 to 2007. He joined the Institute of Digital Media, School of Electronics Engineering and Computer Science, Peking University, Beijing, where he is currently a Professor. He has authored over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

**WEN GAO** (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently a Professor of computer science with the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing, China. Before joining Peking University, he was a Professor of computer science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is a member of the China Engineering Academy. He has published extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interfaces, and bioinformatics. He has been the chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE International Conference on Multimedia and Expo and ACM Multimedia, and served on the Advisory and Technical Committees of numerous professional organizations. He served or serves on the Editorial Board of several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, the *Journal of Visual Communication*, and *Image Representation*.