Signal Processing: Image Communication 55 (2017) 130–145

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Reduced-reference image quality assessment based on distortion families of local perceived sharpness

Yi Zhang[a,*], Thien D. Phan[b], Damon M. Chandler[a]

[a] Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Shizuoka 432-8561 Japan
[b] Department of Information Technology, University IT Gia Dinh, Ho Chi Minh City, Vietnam

ARTICLE INFO

Keywords:
Reduced-reference quality assessment
Image quality
Local sharpness
Distortion family

ABSTRACT

Previous research on reduced reference (RR) image quality assessment (IQA) suggested that appropriate RR features should provide efficient summaries of reference images and be sensitive to a variety of image distortions. The multi-scale local sharpness maps are effective RR features because they can capture smooth, edge, and textured areas of the reference image, and they are affected differently by different distortion types. Motivated by this observation, in this paper, we propose an efficient RR IQA algorithm using local sharpness. Our method, called S4RR, employs four sharpness maps (two FISH maps and two local standard deviation maps) to assess image quality via two main stages. The first stage soft-classifies the distorted image into eight distortion families based on an analysis of the different scatter-plot shapes of the sharpness map values of distorted image vs. reference image. The second st age performs distortion-family-specific quality assessment based on measuring the local sharpness variations between reference and distorted images by using seven types of local statistics and six distance measures. Finally, the soft-classification probabilities computed from the first stage are combined with the distortion-family-specific quality scores to yield a class-weighted average, which serves as the final S4RR quality index. Experiment results tested on various databases show that with less than 5% RR information, the proposed S4RR algorithm achieves better/competitive performance as compared to other state-of-the-art FR/RR IQA algorithms.

1. Introduction

The ability to assess image quality in a manner that is consistent with human subjective ratings of quality is an important task for many image processing systems. Over the last several decades, algorithms for image quality assessment (IQA) have been intensively researched and developed. An IQA algorithm can be classified into three main categories: (1) full-reference (FR), in which the algorithm's inputs are the reference and distorted images; no-reference (NR), in which the algorithm needs only the distorted image; and (3) reduced-reference (RR), in which partial information about the reference image is made available and compared to the input distorted image. IQA research has recently been shifting more towards the NR/RR categories, due to the often impractical requirement of providing the full reference image, particularly in streaming applications. Although NR IQA algorithms have yet to outperform state-of-the-art FR IQA algorithms, modern RR IQA algorithms have begun to approach the performance of FR IQA.

Research on RR IQA has primarily focused on finding appropriate *RR features* that can summarize the reference image while being sensitive to a variety of image distortions. Although it is challenging to determine the type and amount of information to use for RR features, many effective RR algorithms have been developed. These approaches can be roughly classified as those which: (1) use changes in natural-scene-statistics (NSS) [1–3]; (2) calculate differences in transform domain coefficients [4–8]; (3) calculate differences in other features [9–11]; or (4) those which create reduced versions of FR algorithms [12–16].

Despite the difference of the NSS models, image transforms, and the RR features employed, previous approaches to RR IQA have generally been limited in two ways: they are either limited by the number of distortion types on which they can operate (typically up to four common distortion types such as additive white Gaussian noise, Gaussian blur, JPEG, and JPEG2000), or limited in their ability to achieve high quality estimate performance because of the small number of features/scalars used as the reduced information. These distortion-specific RR [9–11] and some other NR algorithms [17,18] have attempted to be generic by combining techniques for analyzing more distortion types. However, they have not been tested on a wide variety of distortion types, such as the numerous distortion types in the TID2008 [19] and TID2013 [20] databases.

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2017.03.020
Received 1 September 2016; Received in revised form 27 March 2017; Accepted 27 March 2017
Available online 02 April 2017

Some other RR algorithms employ from a few to a hundred scalars for the reduced information, and yet can achieve performances only at the level of MSE/PSNR. More recently, Soundararajan et al. [16], and Liu et al. [21] provided two RR algorithms, RRED and SPCRM, respectively. These algorithms have been reported to achieve better performance by increasing the amount of reduced information. When utilizing 2.8% and 3.1% (in terms of scalar numbers) reduced information (for RRED and SPCRM, respectively), these algorithms are nearly as good as current state-of-the-art FR algorithms such as MS-SSIM [22], VIF [23], MAD [24], or FSIM [25] on many large image databases (TID2008 [19], LIVE [26], CSIQ [27]). However, we believe that with appropriate features, RR algorithms can be even better than FR algorithms, particularly considering that humans often do not require full information about the reference image to assess the quality of a distorted image. Therefore, we seek an RR approach that can (1) operate on a wider range of distortion types; (2) be competitive and/or outperform state-of-the-art FR/RR IQA algorithms; and (3) require equal/less amount of reduced information.

In this paper, we present a new RR IQA algorithm that estimates quality based on an analysis of the local sharpness differences between the reference and distorted images. Previous work on image sharpness has focused on developing a sharpness index, a single value that quantifies the overall sharpness of an image (e.g., [28–33] etc.) Some algorithms also generate sharpness maps that quantifies the perceived sharpness of different local areas within an image, and these maps have been used for NR IQA of Gaussian blurred images [32,33] and JPEG2000 compressed images [34]. In our work, the sharpness features are employed for RR IQA of images that contain more general distortion types.

The proposed algorithm, which we called S4RR, employs four sharpness maps for feature extraction: (1) the luminance-based Fast wavelet-based Image SHarpness (FISH) map, (2) the downsampled local lightness distance based (LLD-based) FISH map, (3) the original-scale local standard deviation (LSD) map, and (4) the downsampled-scale LSD map. We argue that changes in these sharpness maps can effectively capture the perceived distortions as long as such changes are quantified in a manner that adapts to different *distortion families*. What makes this approach a good candidate for RR IQA is the fact that sharpness maps can provide efficient summaries of the reference images and are sensitive to a variety of image distortions. This notion is demonstrated in Figs. 1 and 2. As shown in Fig. 1, the reference image and four distorted images are shown in Row (a) with four common distortion types: additive Gaussian noise, Gaussian blur, JPEG compression, and JPEG2000 compression. Rows (b)-(e) show the four sharpness maps for each of the five images, respectively. Fig. 2 shows the luminance-based FISH maps for the four distortion types at three different levels of distortion. Observe that different distortion types and levels produce different sharpness maps. Therefore, it is possible to estimate the quality degradation by measuring the deviations in these distorted sharpness maps from the reference sharpness maps.

Based on the four sharpness maps, S4RR operates via two main stages. The first stage performs distortion identification. Unlike many of the previous NR IQA works (e.g., [17,35,18], etc.) that treat individual distortions as separate classes, we propose in this paper the concept of *distortion families* and each distortion family consists of one or more distortion types that display considerable similarity regarding to certain properties (for example, different types of the noise-corrupted images can be clustered into one distortion family, as they are all distorted by the noise). Consequently, in this stage, 30 classification features are extracted from the two FISH maps (luminance-based and LLD-based) and are fed into a classifier, which estimates the probability that the image is afflicted by one of the multiple distortion families. To be more specific, we build eight distortion families from the 24 distortion types in the TID2013 database [20] based on the scatter plots of the sharpness

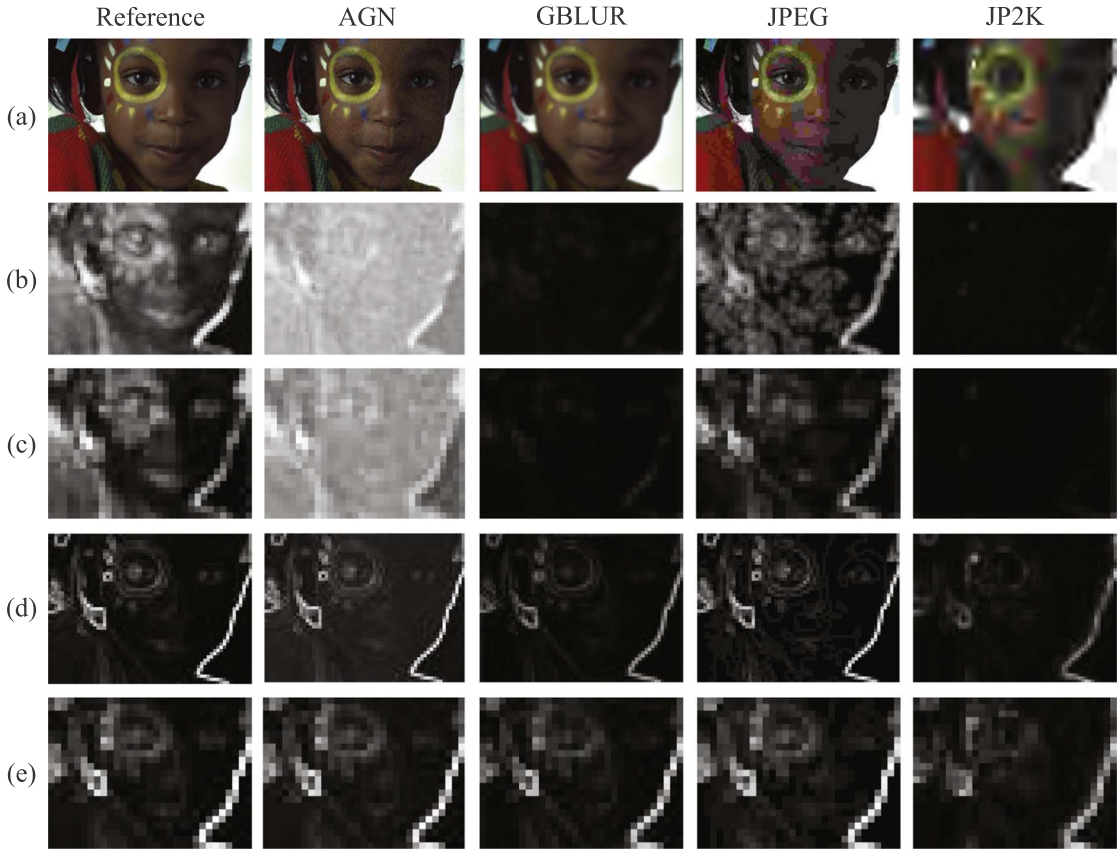

Fig. 1. One reference image ("I15.bmp") and its four distorted versions (row *a*) from the TID2013 database [20] used to demonstrate that sharpness maps are sensitive to different distortion types. Rows *b*, *c*, *d*, and *e* show the four sharpness maps respectively.

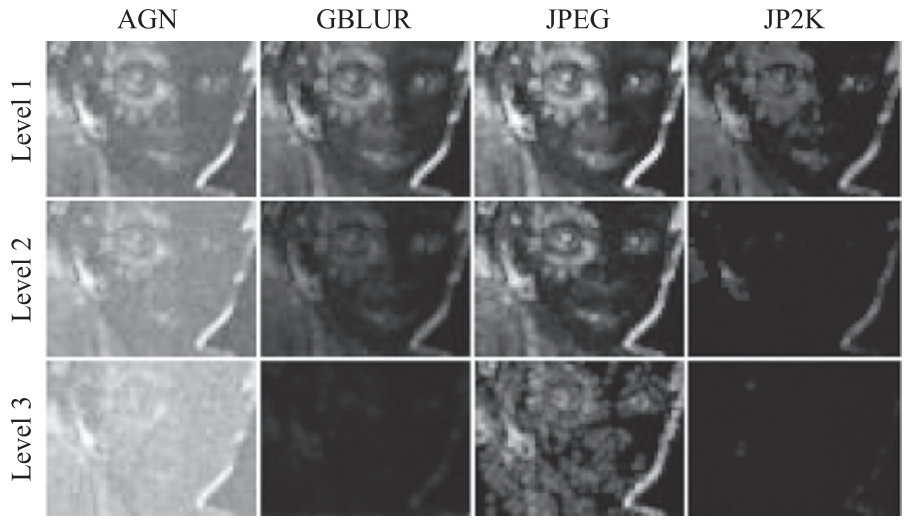

Fig. 2. Luminance-based FISH maps for the 12 distorted versions of the reference image "I15.bmp" used to demonstrated that sharpness maps are sensitive to different distortion levels. Four distortion types (AGN, GBLUR, JPEG, and JP2K) and three distortion levels in the TID2013 database [20] are shown here.

values of distorted maps vs. reference map (see Section 3.2.1 for details). Apparently, this process can greatly reduce the inherent distortion identification errors (due to the imperfectness of the extracted classification features) as compared to the 24 cases, which is also the main reason for the proposed distortion family concept.

The second stage of S4RR performs distortion-family-specific quality assessment (QA). By assuming that distortion types of the same distortion family share common characteristics in sharpness maps, in this stage, a specific regression model for each distortion family is used to map the extracted sharpness-difference-based features to quality estimates. In this work, the three sharpness maps (the luminance-based FISH map and two LSD maps) are employed to extract the RR features. We propose four types of local statistics and six distance measures, and consequently 46 regression features are extracted to quantify both sharpness value changes and sharpness map structure changes (see Section 3.3 for details). Finally, the distortion family probabilities estimated in the first stage and the distortion-family-specific quality scores estimated in the second stage are combined to yield an overall quality score of the image.

The main contributions of this work are as follows. First, we propose an RR IQA algorithm that estimates image quality based on sharpness-map analysis. We demonstrate in our work that sharpness maps can serve as efficient image-quality-related features that are sensitive to different distortion types and levels. Second, to enable our algorithm to work on a wider range of distortion types, various approaches/strategies are proposed for each of the two stages. For the distortion identification stage, we propose the concept of distortion families and classify the 24 distortion types in TID2013 into eight distortion families by analyzing their different scatter-plot shapes. For the distortion-family-specific quality assessment stage, we propose four types of local statistics and six distance measures to quantify the sharpness map differences between reference and distorted images. Third, the two-stage framework and the distortion family concept proposed in our paper might provide a useful way to investigate other features/feature maps under FR/RR conditions, as long as these features/feature maps can efficiently capture the reference information, change differently for different distortion types/families, and be sensitive to different distortion intensities. Compared to the similar two-stage framework employed by most NR IQA algorithms (e.g., [17,35,18]), which treat individual distortion type as one class, the proposed distortion family concept is obviously more suitable to the wider-range-distortion case, because the former one usually generates more distortion identification errors.

This paper is organized as follows. Section 2 reviews existing approaches for RR IQA. In Section 3, we describe details of the proposed S4RR algorithm. In Section 4, we analyze and discuss the performance of S4RR on various image quality databases. General conclusions are presented in Section 5.

2. Previous work

Although it still remains challenging to decide what kind of reduced information and how much of that information should be used for an RR IQA algorithm, a notable amount of research progress has been made [9,1,36,37,4,5,2,6,10,7,3,16,12,8,11,38,15,13,14,39,21]. It is difficult to classify RR IQA algorithms into separate classes. However, there are a few noticeable trends used to calculate the reduced information and to estimate image quality. Some RR IQA methods employ NSS models to extract RR features [1–3]. Some other approaches extract the statistics of images in various transform domains [4–8]. Other works focus on developing RR algorithms which employ more than one strategy for as many distortion types as possible [9–11]. More recently, some researchers attempt to find the relationships between current FR models and RR algorithms [12–16]. In this section, we provide a brief review of current RR methods based on these main trends.

2.1. Methods based on natural-scene statistics

Natural-scene statistics have been used for RR IQA algorithms due to their ability to provide compact descriptions of the reference and distorted images. For example, in [1], Wang et al. proposed an RR IQA algorithm in the wavelet domain by using the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and distorted images. In [2], Li et al. employed a Gaussian-scale-mixture-based statistical model of wavelet coefficients to estimate image quality. In [3], Ma et al. developed an RR algorithm based on a generalized Gaussian distribution model applied to DCT coefficients.

2.2. Methods based on other transform domain differences

Estimating quality based on the differences between reference and distorted grouplet, curvelet, and contourlet transform coefficients is also a common RR approach. For example, in [4], Maalouf et al. calculated information regarding textures and gradients of the images via a grouplet transform, and employed CSF filtering and thresholding to estimate quality. In [5], Gao et al. performed a multiscale geometric

analysis using curvelets, bandlets, wavelets, and contourlets for RR IQA. In [6], Tao et al. employed the city-block distance to measure the differences in "visual sensitive" coefficients in the contourlet domain. In [7], Xue et al. developed an RR IQA algorithm based on Weibull distribution of wavelet coefficients. In [8], Lin et al. used the directional information of the image obtained in the complex wavelet domain to estimate image quality. In [40], Golestaneh et al. proposed an RR IQA algorithm based on the divisive normalization transform (DNT) of locally weighted gradient magnitudes. The RR features are then obtained by computing and averaging the entropy of each DNT subband.

2.3. Methods based on distortion-specific features

Distortion-specific features, such as blocking, blurring, ringing, intensity masking, and lost blocks measures, have proved useful for RR IQA. For example, in [9], Gunawan et al. computed blocking or blurring features based on local harmonic amplitude information computed from edge-detected images. The two separate measurements were then combined to estimate image quality. In [10], Engelke et al. calculated five measures for four artifacts in wireless imaging: blocking, blur, ringing, and intensity masking and lost blocks. A combination of these five measures was shown to be quite effective for RR IQA on the Wireless Imaging Quality database [10]. In [11], Bordevic et al. calculated four RR quality features, structural information, contrast quality, histogram information, and a no-reference JPEG measure, to estimate quality. The proposed RR algorithm employed a neural network statistical estimator to refine these features, and was shown to be effective on three distortion types, Gaussian blur, JPEG, and JPEG2000 in the TID2008 database [19].

2.4. Methods based on current FR algorithms

Another approach to RR IQA has been to develop reduced versions of full-reference IQA algorithms. For example, in [12], Rehman et al. developed the RR-SSIM algorithm, which is a reduced-information version of SSIM [41]. From the same SSIM idea, Bhateja et al. [13,14] measured the structural dissimilarity to estimate quality. Visual information fidelity [23] has also inspired RR algorithms [15,16]. For example, in [16], Soundararajan et al. developed a framework for RR IQA based on information-theoretical measures of differences between the reference and distorted images by using the entropy of wavelet coefficients. When the RR information (the number of scalars required) was around 2.8% of the image size, the algorithm was reported to perform nearly as good as the best-performing FR IQA algorithms.

2.5. Summary of existing methods

In summary, current methods have approached the RR IQA problem by using changes in NSS, differences in transform domain coefficients, differences in distortion-specific features, or by creating reduced versions of FR algorithms. However, these algorithms have been tested only on a limited number of distortion types and have generally yielded predictive performances below that of state-of-the-art FR IQA algorithms. Some RR algorithms employed from a few to a hundred scalars for the reduced information, and could achieve performances only as good as MSE/PSNR. Only Soundararajan et al. [16] and Liu et al. [21] provided RR algorithms (RRED and SPCRM), which are nearly as good as current state-of-the-art FR algorithms when utilizing 2.8% and 3.1% reduced information, respectively.

In the following section, we describe our approach to RR IQA, the S4RR algorithm, which estimates quality based on differences in local image sharpness between the reference and distorted images. As we will demonstrate, with approximately 4.88% reduced information (in terms of the number of bits), S4RR can be competitive with or outperform state-of-the-art FR/RR IQA algorithms.

3. Algorithm

The proposed S4RR algorithm is based on the assumption that distortions affect both the individual sharpness values (which correspond to specific local areas of an image) and the local structure of an image's sharpness map. In this way, it is possible to measure the differences between reference and distorted sharpness maps as proxy estimates of quality degradation. With this idea and motivated by some of the previous NR IQA algorithms (e.g., DIIVINE [17], C-DIIVINE [35], and DESIQUE [18]), S4RR operates via two main stages: (1) distortion identification, and (2) distortion-family-specific quality assessment.

In the first stage, S4RR extracts 15 features from each of the two FISH maps to perform soft classification. To accurately quantify the quality degradation, in the second stage, S4RR extracts in total 46 features from the three sharpness maps (one luminance-based FISH map and two LSD maps) based on four types of local statistics and six distance measures. Both stages employ the support vector machine (SVM) to perform classification and regression. Finally, the soft-classification probabilities and the distortion-family-specific quality estimations are combined to yield a class-weighted average score which serves as the final S4RR quality degradation index.

A block diagram of the S4RR algorithm is shown in Fig. 3. In the following subsections, we discuss details for each stage.

3.1. Reduced-reference information

In this section, we describe the method to collect the reduced-reference information from various sharpness maps. We analyze the amount of RR information in terms of bit number based on sharpness algorithm, block size, and block overlap. The "RR information" refers to the amount of side information about the reference image that is needed for the QA process (e.g., side information sent by the encoder in a streaming QA setting). In our work, this side information is the four sharpness maps from the reference images.

Also, it is important to note that the amount of RR information used in our algorithm depends on the image size. As we have mentioned in Section 1, recent approaches to RR IQA have demonstrated competitive performance with even FR IQA methods by using RR information that also varies with image size (e.g, [16,21]). Although our algorithm and these previous variable RR-information-sized algorithms most often require more side information than previous fixed-sized approaches,

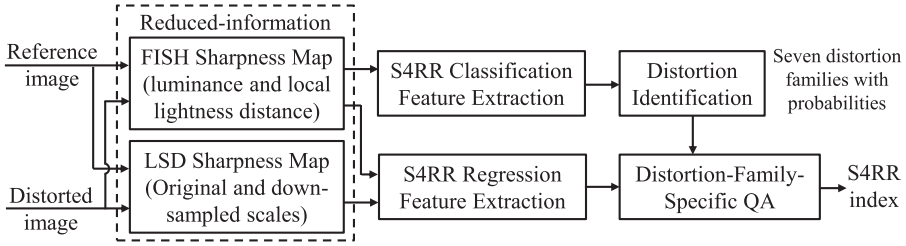

Fig. 3. Block diagram of the S4RR algorithm. Note that in the distortion identification stage, we only consider seven distortion families with probabilities although the 24 distortion types in the TID2013 database [20] are actually clustered into eight distortion families. See text for details.

the amount of RR information is still only a small fraction of the full reference image information.

3.1.1. FISH sharpness map

The FISH sharpness map is a block-based version of the Fast wavelet-based Image SHarpness algorithm [33]. A three-level discrete wavelet transform (DWT) is employed in FISH. The sharpness index for each block of size 16×16 is then given by a combination of the log-energies of blocks of size 8×8, 4×4, and 2×2, respectively, for the subbands from the three DWT levels. FISH employs 50% overlap (i.e., 4, 2, and 1 pixels of overlap for each DWT level) between neighboring blocks to generate the sharpness map. Therefore, the FISH sharpness map is 1/64th the size of the input image.

In this work, we compute FISH sharpness maps of the original image and the image's local lightness distance (LLD) feature (same as that computed in [42]), denoted by "Luminance-based FISH map" and "LLD-based FISH map," respectively. In [42], we demonstrated that LLD can be a useful feature for predicting stereoscopic image quality. Here, we use the LLD FISH map only for its efficacy in distortion identification. Thus, the original LLD FISH map is filtered and downsampled (by 2 in each direction) to further reduce the RR information. Consequently, the amount of reduced information for the two FISH sharpness maps is $1/64 + 1/256 = 1.95\%$ of the original image size.

3.1.2. LSD map

The local standard deviation (LSD) has been employed previously in noise estimation [43], contrast enhancement [44], texture analysis [45], and quality assessment of JPEG-compressed images [46]. Here, we argue that LSD is also a good candidate for local sharpness/quality estimate. The first LSD map is formed by calculating the standard deviation of each block of size 12×12 (with 4 pixels of overlap between neighboring blocks) of the original image. Then, the same procedure is applied to the filtered and downsampled-by-2 version of the image[1] to obtain the second LSD map. Therefore, the amount of reduced information for the two scales is $1/8^2 + 1/2^2/8^2 = 1.95\%$ of the original image size.

3.1.3. Quantization

Based on the aforementioned points, we extract a total of four sharpness maps from the reference image: (1) Luminance-based FISH map, (2) downsampled LLD-based FISH map, (3) original scale LSD map, and (4) downsampled scale LSD map. These four sharpness maps are then quantized to 10 bits for each pixel value. Thus, in terms of the number of bits, the amount of reduced information is $(10/8)\times(1.95\%+1.95\%) = 4.88\%$ of the original image. As we will demonstrate next, features extracted from these sharpness maps can be effective for both distortion identification and quality assessment.

3.2. Sharpness features for classification

As discussed in Section 1 and demonstrated in Figs. 1 and 2, sharpness maps are sensitive to a variety of distortion types and levels. Although different distortion types can exhibit different visual impacts on the images, some distortion types exhibit the same or similar impacts. Thus, it is necessary to determine the distortion families from the distorted image before using the sharpness maps to estimate quality. In this section, we first describe the eight distortion families derived from the TID2013 database [20]. Then, we provide details on extracting various classification features from the reference- and distorted-image sharpness maps to perform soft classification.

3.2.1. Distortion families

The TID2013 database contains 24 individual distortion types: (1) additive Gaussian noise (AGN), (2) additive noise in color components (ACN), (3) spatially correlated noise (SCN), (4) masked noise (MN), (5) high frequency noise (HFN), (6) impulse noise (IN), (7) quantization noise (QN), (8) Gaussian blur (GBLUR), (9) image denoising (ID), (10) JPEG compression (JPEG), (11) JPEG2000 compression (JP2K), (12) JPEG transmission errors (JPG_TE), (13) JPEG2000 transmission errors (J2K_TE), (14) non eccentricity pattern noise (NEPN), (15) local block-wise distortions (LBW), (16) mean shift (MS), (17) contrast change (CTC), (18) change of color saturation (CCS), (19) multiplicative Gaussian noise (MGN), (20) comfort noise (CN), (21) lossy compression of noisy image (LCN), (22) color quantization with dither (CQ), (23) chromatic aberrations (CA), and (24) sparse sampling (SS). To group these 24 distortion types into different distortion families, we use the scatter plots of the sharpness map values of distorted image vs. reference image. Some examples [2] are shown in Fig. 4.

As shown in Fig. 4(a), the scatter plot of the sharpness map values of a reference image versus itself fits perfectly to the line $y=x$. However, when images are distorted, their sharpness maps will be changed, and such changes are reflected in the scatter-plot shapes. Fig. 4(b) shows the luminance-based FISH maps for the 24 distorted versions of the reference image, each of which represents one distortion type in the TID2013 database. Also shown in Fig. 1(b) are the 24 corresponding scatter plots of the distorted sharpness map values vs. the corresponding reference values. Note that for each scatter plot in Fig. 4(b), the horizontal axis represents the reference map values and the vertical axis represents the corresponding distorted map values. From Fig. 4(b), observe that different distortion types display different scatter-plot shapes. However, also observe that some distortion types interestingly share considerable similarities.

To model the different shapes of these scatter plots, we fit each of them by using a three-parameter second-order polynomial curve, which is given by

$$y = \lambda_1 \cdot x^2 + \lambda_2 \cdot x + \lambda_3, \tag{1}$$

where λ_i $(i = 1, 2, 3)$ are the curve parameters. After polynomial fitting, we calculate the sum of the root square error which is used as the forth parameter (denoted by *Err*) for analysis. To be more generic, we obtain these four parameters from all the 24 different natural-scene images in the TID2013 database [20], and each image content is chosen at its most distorted version for clearer observation. Note that scatter plots of both the contrast reduction and contrast enhancement images which belong to the same 17^{th} distortion in the database are fitted (denoted by 17A and 17B, respectively). Consequently, we fit 25 scatter plots for each image content, which results into a total of 2400 values (24 image content $\times$ 25 scatter plots $\times$ 4 parameters). The mean and standard deviation (denoted by μ and σ, respectively) of the four parameters for each distortion type in the TID2013 database are shown in Table 1. Also shown in Table 1 are the number of distorted images (or scatter plots) from which these statistical numbers are calculated. Based on these obtained parameters along with the scatter plots, we classify the 24 distortion types in TID2013 into eight distortion families. The corresponding mappings are shown in Tables 1, 2, and summarized as follows:

- Id 1: The first distortion family, which we called "noise-based distortion," contains seven distortion types, all of which introduce certain levels or types of noise corruptions. Thus, their corresponding scatter plots show significant increases in the images' sharpness values (i.e., the scatter plots are far above the line $y=x$). From Table 1, we observe that this distortion family usually has a small value of λ_2 and a large value of λ_3. Because their scatter plots are

[1] In this paper, we used bicubic interpolation to downsample the images; the output pixel value is a weighted average of pixels in the nearest 4×4 neighborhood.

[2] Here, we use the luminance-based FISH map for example.

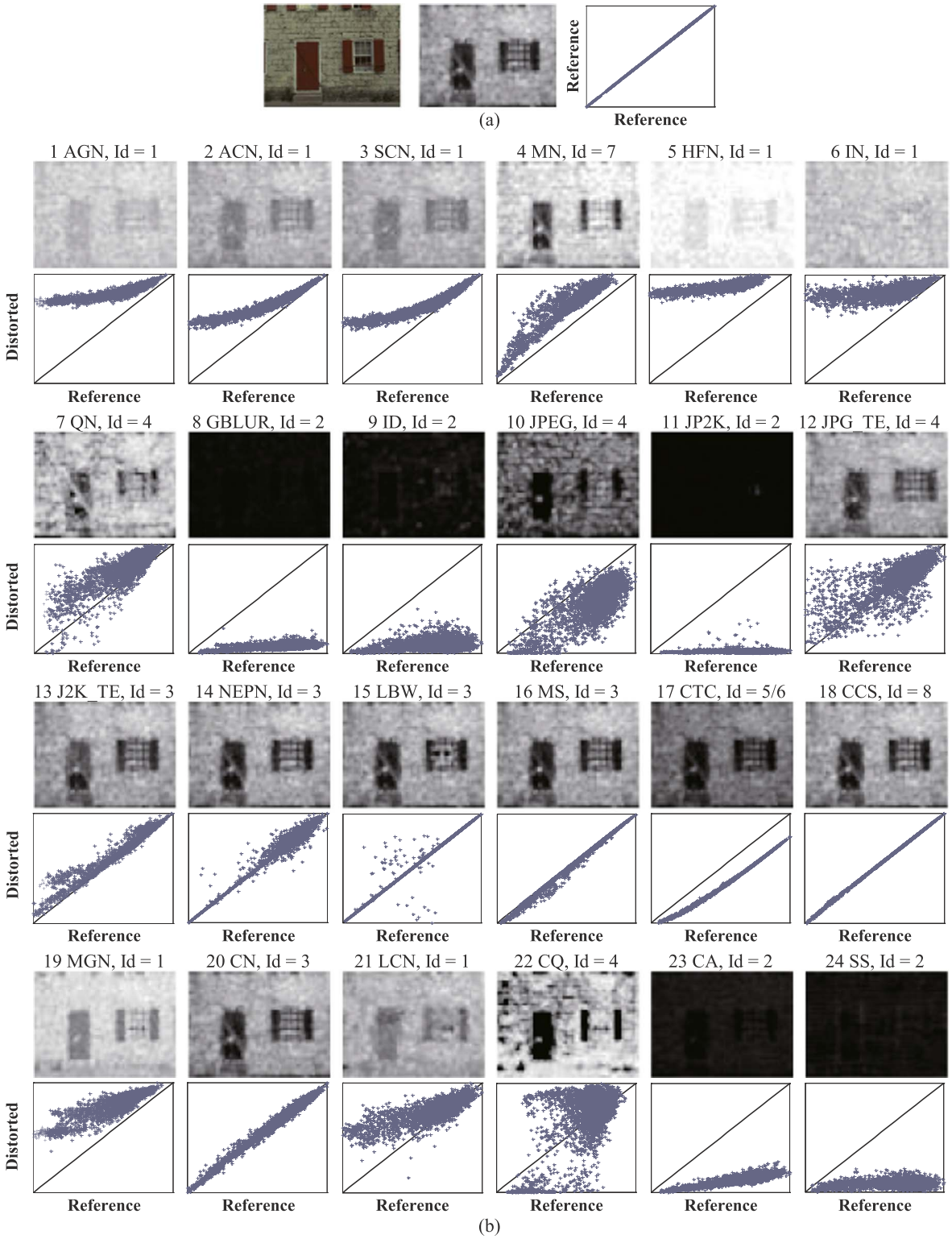

Fig. 4. Scatter plots of the sharpness map values of one reference image (*I01.bmp* in TID2013) versus its 24 distorted versions. Shown in (a) is a reference image, its luminance-based FISH map, and a scatter plot of the map values vs. itself. Shown in (b) are the luminance-based FISH maps and the scatter plots of the map values of 24 distorted images vs. the reference image. Each distorted image contains one of the 24 distortion types (denoted by the number 1, 2,…, 24, respectively). For each scatter plot in (b), the horizontal axis represents the reference map values and the vertical axis represents the corresponding distorted map values. Also included in (b) are the corresponding distortion family number (denoted by "Id") for each distortion type. Notice that different distortions may display different scatter-plot shapes, but scatter plots of some distortion types can share great shape similarity. See text for details.

usually *narrow*, the fitting error (*Err*) is also relatively small.

- Id 2: The second distortion family is called "blur-based distortion" because all five distortion types assigned to this group generally bring some kind of the blur effect to an image. Consequently, the image's sharpness values decrease significantly compared to the reference sharpness values (i.e., the scatter plots are far below the line $y=x$). Again, we observe small values of λ_2 and *Err*, but what is different from the first family is the also small value of λ_3.
- Id 3, 4: The third distortion family produces only a slight effect on the image's local sharpness values, while the fourth distortion family

Table 1
The mean and standard deviation of the four polynomial fitting parameters for the 24 distortion types in the TID2013 database.

Distortion type	# images	Distortion family Id	λ_1		λ_2		λ_3		*Err*	
			μ	σ	μ	σ	μ	σ	μ	σ
1. AGN	24	1	0.31	0.12	0.00	0.14	0.68	0.07	1.40	0.19
2. ACN	24	1	0.41	0.07	0.10	0.10	0.50	0.07	1.24	0.19
3. SCN	24	1	0.56	0.08	−0.06	0.10	0.51	0.07	1.27	0.18
4. MN	24	7	−0.88	0.47	1.69	0.41	0.11	0.09	3.48	1.17
5. HFN	24	1	0.18	0.15	0.04	0.16	0.80	0.06	1.45	0.15
6. IN	24	1	0.42	0.31	−0.15	0.33	0.73	0.09	2.89	0.57
7. QN	24	4	−0.67	0.94	1.68	1.01	−0.05	0.27	8.47	2.67
8. GBLUR	24	2	0.04	0.09	0.06	0.10	0.01	0.02	1.10	0.14
9. ID	24	2	0.25	0.21	−0.08	0.18	0.06	0.04	2.75	0.44
10. JPEG	24	4	0.24	0.59	0.51	0.64	0.02	0.14	6.90	1.27
11. JP2K	24	2	0.18	0.44	−0.13	0.38	0.04	0.08	1.17	1.20
12. JPG_TE	24	4	−0.19	0.43	0.86	0.50	0.15	0.15	5.67	1.96
13. J2K_TE	24	3	0.23	0.38	0.48	0.40	0.27	0.14	4.54	1.90
14. NEPN	24	3	−0.17	0.09	1.13	0.08	−0.01	0.01	2.35	0.46
15. LBW	24	3	−0.07	0.11	1.04	0.09	0.00	0.02	2.17	0.38
16. MS	24	3	0.01	0.23	1.02	0.25	−0.03	0.06	1.26	1.00
17A. CTC	24	5	−0.23	0.34	1.28	0.38	0.00	0.11	1.90	0.93
17B. CTC	24	6	0.24	0.06	0.63	0.07	−0.05	0.02	0.37	0.12
18. CCS	24	8	0.00	0.01	0.99	0.02	0.00	0.00	0.35	0.17
19. MGN	24	1	0.35	0.60	0.04	0.63	0.63	0.16	3.99	0.96
20. CN	24	3	−0.02	0.06	0.96	0.06	0.03	0.02	1.26	0.19
21. LCN	24	1	0.04	0.65	0.29	0.78	0.43	0.23	5.44	1.59
22. CQ	24	4	−0.48	1.39	1.15	1.55	0.24	0.41	10.72	3.60
23. CA	24	2	0.06	0.07	0.24	0.27	−0.01	0.03	0.92	0.38
24. SS	24	2	0.01	0.12	0.06	0.09	0.05	0.05	1.55	0.80

Table 2
Mappings between eight distortion families and 24 distortion types in the TID2013 database.

Id	Distortion families	Distortion types
1	Noise-based distortion	AGN, ACN, SCN, HFN IN, MGN, LCN
2	Blur-based distortion	GBLUR, ID, JP2K, CA, SS
3	Spatially localized distortion	J2K_TE, NEPN, LBWD, CN
4	Spatially correlated broadband distortion	QN, JPEG, JPG_TE, CQ
5	Contrast reduction	CTC
6	Contrast enhancement	CTC
7	Masked Noise	MN
8	Change of color saturation	CCS

gives rise to much broader gains and losses. Therefore, we call these families "spatially localized distortion" and "spatially correlated broadband distortion," respectively. Observe from Fig. 4 that these two groups display either very *narrow* or very *wide* scatter-plot shapes. Specifically, because the third distortion family has a scatter plot that falls almost on the region of the line $y=x$, it has a nearly zero value of λ_1, a large value of λ_2, and small values of λ_3 and *Err*. In comparison, the forth distortion family has a very *wide* scatter-plot shape, and thus it has very random values of λ_1, λ_2, and λ_3 (this can be seen from their corresponding σ values), and because of such randomness, this distortion family gives rise to the largest fitting error (*Err*) among all the distortions considered here.

- Id 5, 6: The fifth and sixth distortion families are derived from a single distortion type in the original TID2013 database [20]. We have these two splits (contrast reduction and contrast enhancement) because each displays different scatter plot positions as compared to the line $y=x$, although both plots are highly correlated (see Fig. 5). Unlike many other distortions, the contrast change does not introduce any noticeable artifacts (e.g., ringing, blurring, and blocking artifacts), because of which the sharpness map structure of the original image is well maintained in its contrast-change versions. Referring to Table 1, we see small values of λ_1, λ_3, and *Err*, and large values of λ_2 for both distortion families.
- Id 7, 8: The last two distortion families, masked noise and change of color saturation, are exactly the same as the original distortion types in TID2013. Note from Fig. 4 that both types display different scatter-plot shapes from the others. Specifically, for the masked-noise-corrupted image, usually less noise exists in its smooth areas, while the edge or textured areas often have more noise corruption (due to the different noise-masking capabilities). Thus, we observe from Fig. 4 that the scatter plot of such distortion displays a middle-bulged curve shape, which indicates that there are more sharpness increases on the image's textured areas and less on the smooth/edge areas. Apparently, this significantly differs from the first noise-based distortion family although they all have additive noise. As for the polynomial fitting, we see larger values of λ_2 and *Err*, and a smaller value of λ_3, as compared to Id 1. Also, the color-saturation-change images display scatter plots that almost coincide with the line $y=x$; hence, the polynomial fitting for this distortion produces nearly zero values of λ_1, λ_3, and *Err*, and λ_2 is close to 1. Note that changes of color saturation do not introduce any visible distortions when images are viewed in gray scales. Thus, we do not consider this distortion type (CCS) in our algorithm.

Despite the aforementioned eight distortion families, we do acknowledge that for some distortion types (e.g., LCN in Id 1, J2K_TE and LBW in Id 3), it is difficult to group them into any one of those distortion families, because their scatter-plot shapes and fitting parameters lie near the boundaries of different distortion types/families. This is possible when different distortion types share certain properties or when images are relatively less distorted, both of which can make it difficult or even impossible for a strict definition (hard classification) of different distortion families. However, because we use a soft classification to achieve the probability-weighted quality estimate (see Eq. (24)), the proposed distortion families can thus be non-strictly defined for some special cases, as long as such definitions or classifications do not significantly affect the algorithm performance.

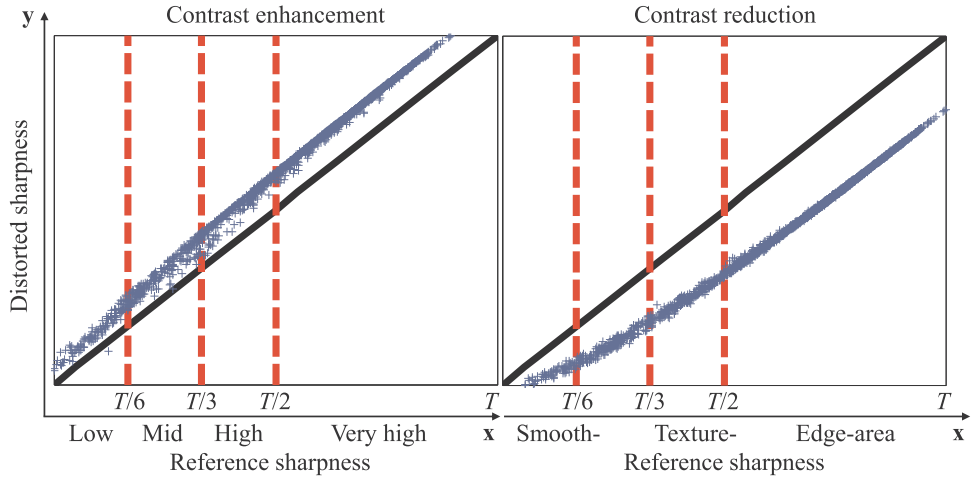

Fig. 5. A demonstration of different scatter plot positions as compared to the line $y=x$ for contrast enhancement and contrast reduction images. These figures also serve as exemplary demos for classification feature extraction.

3.2.2. Classification feature extraction

With these eight distortion families, we next extract classification features for each of them based on the two FISH sharpness maps (luminance-based and LLD-based). As shown in Fig. 5, the reference-sharpness axis (x-axis) is divided into four intervals from left to right: low-valued, mid-valued, high-valued, and very-high-valued, which roughly correspond to the smooth areas, textured areas, and edge areas of an image. These four intervals, combined with the $y=x$ line, give rise to eight regions on the scatter plot, which will be used for classification feature extraction.

Let S^{ref} and S^{dst} denote the FISH sharpness maps (luminance-based or LLD-based) of the reference and distorted images, respectively. Then the four regions on reference sharpness map (denoted by L^{ref}, M^{ref}, H^{ref}, and V^{ref}, respectively) and four regions on distorted sharpness map (denoted by L^{dst}, M^{dst}, H^{dst}, and V^{dst}, respectively) are given by

$$L^{ref/dst}(l) = S^{ref/dst}(l),\ \forall\ l | S^{ref}(l) < T/6 \tag{2}$$

$$M^{ref/dst}(m) = S^{ref/dst}(m),\ \forall\ m | T/6 \le S^{ref}(m) < T/3 \tag{3}$$

$$H^{ref/dst}(h) = S^{ref/dst}(h),\ \forall\ h | T/3 \le S^{ref}(h) < T/2 \tag{4}$$

$$V^{ref/dst}(v) = S^{ref/dst}(v),\ \forall\ v | T/2 \le S^{ref}(v) \tag{5}$$

where l, m, h, and v denote the indices of S^{ref} corresponding to low, mid, high, and very high values, and T denotes the maximum value of a sharpness map.

Based on the reference and distorted sharpness maps, and the eight regions on scatter plot, we extract 15 sharpness features ($f_1 - f_{15}$) from each FISH map (luminance-based FISH map or LLD-based FISH map). Detailed information is given in Table 3, where l, m, h, and v have the same meaning as in Eqs. (2)–(5). Therefore, a total of 30 features are extracted from the two FISH maps, which are then fed into the SVM model for soft classification. We demonstrate in Section 4 that these proposed features are highly effective at distortion identification.

3.3. Sharpness features for regression

In this section, we describe details of how to extract the regression features from the three sharpness maps (one luminance-based FISH map and two LSD maps) to perform the distortion-family-specific QA task (note that we do not use the downsampled LLD-based FISH map for regression feature extraction). The main goal of this stage is to give a quality estimate for each distortion family based on the differences between the reference and distorted sharpness maps. As mentioned previously, distortions will cause changes to both the individual sharpness values and the local structure of an image's sharpness map. Thus, the key point of the second stage in S4RR is to extract efficient features that can describe such a change. To this end, we propose four types of local statistics and six distance measures to quantify both the sharpness value changes and sharpness map structure changes between the reference and distorted sharpness maps. We provide details in the following subsections.

3.3.1. Local statistics

The local statistics aim at modeling the relationship between neighboring pixel-pairs on a sharpness map, which to some extent represent the map structure. Applying each local statistic to the original sharpness map will produce another rescaled map, which we call sharpness feature map.[3] Thus, four types of sharpness feature maps can be generated from each of the three original sharpness maps.

Let $\mathbf{S}$ denote one of the three original sharpness maps of an image, and its elements are denoted by S_i. For each 2×2 block (without overlap), we compute the local maximum and local standard deviation as

$$S_{f_1} = \max\{S_1, S_2, S_3, S_4\}, \tag{6}$$

$$S_{f_2} = \sqrt{\frac{1}{3}\sum_{i=1}^{4}(S_i - \overline{S})^2}, \tag{7}$$

where S_i ($i = 1, 2, 3, 4$) denotes one of the four sharpness values within each 2×2 block, and $\overline{S}$ denotes their average value.

In addition, to measure the sharpness change within each block, we propose another local statistic called top-variation. Specifically, given a 2×2 block, we first compute the six-element variation vector (denoted by $\mathbf{v}$) via

$$\mathbf{v} = [|S_1 - S_2|, |S_1 - S_3|, |S_1 - S_4|, |S_2 - S_3|, |S_2 - S_4|, |S_3 - S_4|].$$

The top-variation-based statistic is then given by

$$S_{f_3} = \sqrt{(v_1^2 + v_2^2)/2}, \tag{8}$$

where v_1 and v_2 denote the two largest elements of $\mathbf{v}$. Note that the above-mentioned three local statistics (S_{f_1} - S_{f_3}) are computed for 2×2 blocks. To extract more about the image's structure information, we also consider the 3×3 block size. In this paper, we use the local sharpness distance computed upon 3×3 blocks to measure the sharpness change between one image area and its surrounding areas. Specifically, the local sharpness distance is computed via

[3] Here, we use the term "sharpness feature map" to distinguish from the term "sharpness map" that is originally computed based on FISH [33] and LSD from the image.

Table 3
All 15 classification features, their formulas and purposes.

Feature	Formulas	Purpose/Good for:
Linear correlation between two maps	$f_1 = corr(\mathrm{S}^{ref}, \mathrm{S}^{dst})$	A high value of f_1 suggests a high quality image or a global-contrast-changing image.
Linear correlation between the reference and error maps	$f_2 = corr(\mathrm{S}^{ref}, (\mathrm{S}^{dst} - \mathrm{S}^{ref}))$	A high value of f_1 and f_2 suggests a contrast enhanced image.
Maximum of the error map	$f_3 = max(\mathrm{S}^{dst} - \mathrm{S}^{ref})$	Appears sharpened (noise-added) if $f_3 + f_4 > 0$. Appears blurred if $f_3 + f_4 < 0$.
Minimum of the error map	$f_4 = min(\mathrm{S}^{dst} - \mathrm{S}^{ref})$	
Average of the error map	$f_5 = mean(\mathrm{S}^{dst} - \mathrm{S}^{ref})$	Appears sharpened if $f_5 > 0$ and appears blurred if $f_5 < 0$.
Average of the increment in S^{ref} map	$f_6 = mean(\mathrm{S}^{dst}(s) - \mathrm{S}^{ref}(s))$ for all $\mathrm{S}^{dst}(s) > \mathrm{S}^{ref}(s)$	(1) $f_6 \approx 0$ and $f_7 \ll 0$: Blurred family (e.g., GBLUR, JP2K, etc.) (2) $f_6 \gg 0$ and $f_7 \approx 0$: Noise family (e.g., AGN, ACN, SCN, etc.) (3) $f_6 \gg 0$ and $f_7 \ll 0$: Two-sided distorted (e.g., JPG_TE, CQ, etc.)
Average of the decrement in S^{ref} map	$f_7 = mean(\mathrm{S}^{dst}(s) - \mathrm{S}^{ref}(s))$ for all $\mathrm{S}^{dst}(s) < \mathrm{S}^{ref}(s)$	
Average of the increment in L^{ref} map	$f_8 = mean(\mathrm{L}^{dst}(l) - \mathrm{L}^{ref}(l))$ for all $\mathrm{L}^{dst}(l) > \mathrm{L}^{ref}(l)$	A high value suggests a noise-added image.
Average of the decrement in L^{ref} map	$f_9 = mean(\mathrm{L}^{dst}(l) - \mathrm{L}^{ref}(l))$ for all $\mathrm{L}^{dst}(l) < \mathrm{L}^{ref}(l)$	A low value suggests pixel-value quantization in the smooth area.
Average of the increment in M^{ref} map	$f_{10} = mean(\mathrm{M}^{dst}(m) - \mathrm{M}^{ref}(m))$ for all $\mathrm{M}^{dst}(m) > \mathrm{M}^{ref}(m)$	Similar to f_8
Average of the decrement in M^{ref} map	$f_{11} = mean(\mathrm{M}^{dst}(m) - \mathrm{M}^{ref}(m))$ for all $\mathrm{M}^{dst}(m) < \mathrm{M}^{ref}(m)$	Similar to f_9
Average of the increment in H^{ref} map	$f_{12} = mean(\mathrm{H}^{dst}(h) - \mathrm{H}^{ref}(h))$ for all $\mathrm{H}^{dst}(h) > \mathrm{H}^{ref}(h)$	Similar to f_8
Average of the decrement in H^{ref} map	$f_{13} = mean(\mathrm{H}^{dst}(h) - \mathrm{H}^{ref}(h))$ for all $\mathrm{H}^{dst}(h) < \mathrm{H}^{ref}(h)$	Similar to f_9
Average of the increment in V^{ref} map	$f_{14} = mean(\mathrm{V}^{dst}(v) - \mathrm{V}^{ref}(v))$ for all $\mathrm{V}^{dst}(v) > \mathrm{V}^{ref}(v)$	Similar to f_8
Average of the decrement in V^{ref} map	$f_{15} = mean(\mathrm{V}^{dst}(v) - \mathrm{V}^{ref}(v))$ for all $\mathrm{V}^{dst}(v) < \mathrm{V}^{ref}(v)$	Similar to f_9

$$\mathrm{S}_{f_4} = |\mathrm{S}_i - \overline{\mathrm{S}}_B| \tag{9}$$

where $\overline{\mathrm{S}}_B$ denotes the average sharpness value of a 3×3 block centered around S_i.

Although all of these local statistics are employed to model the sharpness change between the reference and distorted images, different local statistics may focus on capturing slightly different distortions and/or image descriptors. Specifically, S_{f_1} mainly captures the individual sharpness value change. This statistic is more sensitive to the Gaussian blur distortion which decreases the sharpness values across all image areas, because the local maximum of a sharpness map (especially in an edge area) is often the first to be degraded by blurring. In comparison, S_{f_2} and S_{f_4} mainly capture the sharpness map structure change. Both statistics are more sensitive to distortions such as JPEG and JPEG2000 compressions, and the various noise corruptions, because all these distortions will add additional artifacts (e.g., spots, aliasing, ringing, and blocking artifacts) that ultimately change the image's sharpness map structure. The top-variation statistic (S_{f_3}) is motivated by the work in [32] that the total variation of an image region in the spatial domain can be effective local sharpness measurement especially when the contrast is taken into account. Therefore, we employ the top-variation statistic built upon the LSD map to mainly capture the contrast change distortions. Despite the different emphasis of different local statistics in representing distortions, we do find that for most distortion types, these four local statistics weigh similarly in predicting image quality, and they are all required for better QA performance when all distortion types in all testing databases are considered.

We refer interested readers to the online supplement (http://vision.eng.shizuoka.ac.jp/s4rr/) for an additional discussion and demonstration.

3.3.2. Distance measure

With the three original sharpness maps and four types of the regenerated sharpness feature maps (denoted by $\mathbf{S}_{f_i}$, $i = 1, 2, \ldots, 4$), the other crucial issue is to measure the distance between a distorted sharpness/sharpness feature map and the corresponding reference sharpness/sharpness feature map (also referred as a "sharpness map pair"). To solve this problem, we employ three difference maps and two pooling strategies, which ultimately give rise to six distance measures.

Let $\mathbf{S}^{ref}$ and $\mathbf{S}^{dst}$ denote the sharpness (feature) maps extracted from the reference and distorted images, respectively. The three difference maps (denoted by $\mathbf{D}_i$, $i = 1, 2, 3$) are given by

$$\mathbf{D}_1 = |\mathbf{S}^{ref} - \mathbf{S}^{dst}| \tag{10}$$

$$\mathbf{D}_2 = \log\left(\frac{1 + 2\mathbf{S}^{ref}\mathbf{S}^{dst}}{1 + (\mathbf{S}^{ref})^2 + (\mathbf{S}^{dst})^2}\right) \tag{11}$$

$$\mathbf{D}_3 = \frac{(2\mu_{\mathbf{S}^{ref}}\mu_{\mathbf{S}^{dst}})(2\sigma_{\mathbf{S}^{ref}\mathbf{S}^{dst}})}{\left(\mu^2_{\mathbf{S}^{ref}} + \mu^2_{\mathbf{S}^{dst}} + C\right)\left(\sigma^2_{\mathbf{S}^{ref}} + \sigma^2_{\mathbf{S}^{dst}} + C\right)} \tag{12}$$

where μ and σ are the mean and standard deviation calculated within each 3×3 window; and $C=0.001$ is a small constant that prevents division by zero. Note that $\mathbf{D}_3$ is computed the same way as is done in SSIM [41]; the only differences are the sliding window size and constant value.

In order to collapse each of these difference maps into scalars which represent the map distance, two pooling strategies are used: (1) top 1% value pooling, and (2) average pooling. Specifically, for the $\mathbf{D}_1$ map, two scalars are computed by

$$f_{\mathbf{D}_1}(1) = \sqrt{\frac{1}{T_1}\sum_{i=1}^{T_1}[\mathbf{D}_1(i)]^2}, \tag{13}$$

$$f_{\mathbf{D}_1}(2) = \sqrt{\frac{1}{N_1}\sum_{i=1}^{N_1}[\mathbf{D}_1(i)]^2}, \tag{14}$$

where N_1 denotes the total number of pixel values in the $\mathbf{D}_1$ map; T_1 denotes the number of values in the set of the 1% largest values of the $\mathbf{D}_1$ map; and where $\mathbf{D}_1(i)$, $i = 1, 2, \ldots, T_1/N_1$ denotes a value in the $\mathbf{D}_1$ map. Similarly, for the $\mathbf{D}_2$ and $\mathbf{D}_3$ maps, the four scalars are given by

$$f_{\mathbf{D}_2}(1) = \frac{1}{T_2}\sum_{i=1}^{T_2}\mathbf{D}_2(i), \tag{15}$$

$$f_{\mathbf{D}_2}(2) = \frac{1}{N_2}\sum_{i=1}^{N_2}\mathbf{D}_2(i), \tag{16}$$

Table 4
Summary of sharpness (feature) maps used and the corresponding extracted features.

Feature maps	Distance measure	Feature Id
FISH, $\mathbf{S}_{f_4}$ of FISH	$f_{\mathbf{D}_1}(1)$, $f_{\mathbf{D}_2}(1)$, $f_{\mathbf{D}_3}(2)$	f_1 - f_6
LSD, $\mathbf{S}_{f_1}$ - $\mathbf{S}_{f_4}$ of LSD	$f_{\mathbf{D}_1}(1)$, $f_{\mathbf{D}_1}(2)$, $f_{\mathbf{D}_2}(1)$, $f_{\mathbf{D}_2}(2)$	f_7 - f_{46}

$$f_{\mathbf{D}_3}(1) = \frac{1}{T_3}\sum_{i=1}^{T_3}\mathbf{D}_3(i), \tag{17}$$

$$f_{\mathbf{D}_3}(2) = \frac{1}{N_3}\sum_{i=1}^{N_3}\mathbf{D}_3(i), \tag{18}$$

where N_2, N_3, $\mathbf{D}_2(i)$, and $\mathbf{D}_3(i)$ have the same meanings as N_1 and $\mathbf{D}_1(i)$ in Eqs. (13) and (14). T_2 and T_3 denote the number of pixel values which receive the 1% smallest value indices of the $\mathbf{D}_2$ and $\mathbf{D}_3$ maps, respectively. Thus, for each sharpness map pair, six scalars which serve as regression features, are computed to represent the map distance.

3.3.3. Regression feature extraction

The regression features are mainly extracted by applying the six distance measures to various sharpness/sharpness feature maps. However, we found that these regenerated sharpness feature maps are not always effective, which means that some feature maps regenerated from the three original sharpness maps are useless or redundant, and can even degrade the overall performance. Also, different maps may require different distance measures to achieve the best quality prediction results. A brief summary of the maps used and the features extracted is shown in Table 4.

Specifically, for the FISH sharpness map, three distance measures [$f_{\mathbf{D}_1}(1)$, $f_{\mathbf{D}_2}(1)$, and $f_{\mathbf{D}_3}(2)$] are applied to both the original luminance-based FISH map and its regenerated local sharpness distance map (i.e., $\mathbf{S}_{f_4}$), which yield six features (2 maps × 3 distance measures). For the LSD sharpness map, four sharpness feature maps are first generated. Then, we apply the four distance measures [$f_{\mathbf{D}_1}(1)$, $f_{\mathbf{D}_1}(2)$, $f_{\mathbf{D}_2}(1)$, and $f_{\mathbf{D}_2}(2)$] to the original LSD map, and its four regenerated sharpness feature maps (i.e., $\mathbf{S}_{f_1}$ - $\mathbf{S}_{f_4}$), which yield another 40 features (2 scales × 5 maps × 4 distance measures).

In summary, we extract in total 46 regression features from the three original sharpness maps (one luminance-based FISH map and two LSD maps), and these features will be used to perform the distortion-family-specific QA task, which is the second stage of S4RR.

3.4. Quality estimation

Given the 30 classification features and 46 regression features, the final stage of S4RR is to employ the probability-based weighted-summation strategy to estimate image quality. Specifically, given a test image, S4RR first employs a support vector classification (SVC) model to measure the probability that the distortion in the image falls into one of the seven distortion families.

Let $\mathbf{p}$ denote the k-dimensional vector of probabilities ($k=7$ in this work), and its elements are denoted by $p(i)$ ($i = 1, 2, \ldots, k$). For any classification feature vector $\mathbf{x}$ and its corresponding label y, the goal is to estimate $p(i) = P(y = i|\mathbf{x})$. Based on the "one-against-one" approach [47] for multi-class classification, the pairwise class probabilities (i.e., $r_{ij} \approx P(y = i|y = i \text{ or } j, \mathbf{x})$) are estimated by using method proposed in [48]. After collecting all r_{ij} values, the second approach of [49] is employed to obtain $p(i)$ by solving the following optimization problem:

$$\min_{p}\ \frac{1}{2}\sum_{i=1}^{k}\sum_{j:j\neq i}(r_{ji}p(i) - r_{ij}p(j))^2$$
$$\text{subject to } p(i) \geq 0,\ \forall\, i,\ \sum_{i=1}^{k} p(i) = 1. \tag{19}$$

More details for solving Eq. (19) are referred to [49].

After obtaining these probability values, we next select the top two highest probabilities (denoted by p_m, $m=1,2$) and their corresponding distortion-family-specific support vector regression (SVR) models for quality estimation. In this work, we employ ϵ-SVR [50] to perform the regression task, as this prediction approach has been proved efficient and widely used in many applications.

Specifically, given a set of training data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$ where $\mathbf{x}_l \in \mathbb{R}^d$ and $y \in \mathbb{R}$, ϵ-SVR aims to find a function, which takes the form

$$f(\mathbf{x}) = \langle w, \phi(\mathbf{x})\rangle + b, \tag{20}$$

such that $f(\mathbf{x})$ has at most ϵ deviation from the desired target y for all the training data, and at the same time is as flat as possible. In Eq. (20), $\phi(\mathbf{x}) \in \mathbb{R}^{d'}$ ($d' > d$) is a non-linear projection, which maps the data from the original d-dimensional domain to a higher d'-dimensional feature space; $w \in \mathbb{R}^d$ is the weight vector, and $b \in \mathbb{R}$ is the bias. This can be done by solving the following optimization problem:

$$\min_{w,b,\xi,\xi^*}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \langle w, \phi(\mathbf{x}_i)\rangle - b \leq \epsilon + \xi_i \\ \langle w, \phi(\mathbf{x}_i)\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i,\ \xi_i^* \geq 0 \end{cases} \tag{21}$$

where ξ_i and ξ_i^* are the slack variables introduced to allow for some small errors; $C > 0$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated. By using Lagrange multipliers and a quadratic programming solver, the final solution of Eq. (21) is given by

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\cdot K(\mathbf{x}_i, \mathbf{x}) + b, \tag{22}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x})$ is a kernel function; α_i and α_i^* are Lagrange multipliers, which are both zeros if $\mathbf{x}_i$ does not contribute to the error function. More details for training ϵ-SVR models are referred to [51,52].

Finally, the trained SVR model will map the 46-regression-feature vector (denoted by $\mathbf{z}$) to the associated quality score, which is given by

$$q_m = \sum_{i=1}^{l_m}(\alpha_{m_i} - \alpha_{m_i}^*)\cdot K_m(\mathbf{x}_{m_i}, \mathbf{z}) + b_m, \tag{23}$$

where $m=1,2$ represents the two SVR models corresponding to the two highest probabilities. The final quality, denoted by S4RR, is then given by

$$\text{S4RR} = \frac{\sum_{m=1}^{2} p_m q_m}{\sum_{m=1}^{2} p_m}. \tag{24}$$

Note that only the top two highest probabilities and the corresponding SVR models are selected for QA; we have found that generally fewer or more than two probabilities/SVR models gives rise to more QA errors. Also note that the SVR requires training on TID2013 [20] using the mean opinion scores (MOS). Thus, larger values of S4RR denote predictions of better image quality. As we will demonstrate next, training the seven distortion families on TID2013 using these 30 classification features and 46 regression features can yield excellent prediction performance on a variety of distorted images from other databases.

4. Results

In this section, we analyze S4RR's ability to predict image quality by using various image quality databases. We also compare the performance of S4RR to other FR, RR, and NR IQA algorithms.

4.1. Training

We trained our models on the TID2013 [20] database, which contains 25 reference images and 3000 distorted images spanning 24 distortion types. As mentioned in Section 3.2, we do not train on CCS, because changes of color saturation do not introduce any visible distortions when images are viewed in grayscale. Thus, the total number of training images was 2875, which cover seven distortion families. We use the LIBSVM package [51] to implement the training. To improve the predictive performance, optimal radial basis function kernel parameters were used for both the classification and the regression.

4.2. Testing

To assess the predictive performance of S4RR, five databases of subjective image quality were used: (1) the LIVE database [26], (2) the CSIQ database [27], (3) the TID2008 database [19], (4) the IVC database [53], and (5) the Toyama [54] database.

The LIVE database contains 29 reference images and 779 distorted images that span five distortion categories: JPEG compression, JPEG2000 compression, Gaussian blur, additive Gaussian noise, and fast fading (FF). The CSIQ database consists of 30 original images distorted using six different types of distortions at four to five different levels: JPEG compression, JPEG2000 compression, Gaussian blur, additive pink Gaussian noise (APGN), additive white Gaussian noise (AWGN), and global contrast decrements. The TID2008 database consists of 25 reference images and 1700 distorted images over 17 distortion types. The IVC database contains 10 original images and 185 distorted versions that cover four distortion types: JPEG, JPEG2000 compression, LAR coding, and Blurring. The Toyama database contains 14 original images and 168 distorted versions that cover two distortion types: JPEG and JPEG2000 compression. The subjective ratings in the LIVE and CSIQ databases are provided in the term of differential mean opinion scores (DMOS); ratings in the TID2008, IVC, and Toyama databases are presented as mean opinion scores (MOS).

We compared S4RR with various FR, RR, and NR quality assessment methods for which code is publicly available. The four FR methods were SSIM [41], multiscale structure similarity (MS-SSIM) [22], visual information fidelity (VIF) [23], and most apparent distortion (MAD) [24]. The five RR methods were wavelet-domain natural image statistics model (WNISM) [55], hybrid wavelets and directional filter banks (HWD) [5], entropy of DNT coefficients of local weighted gradient (REDLOG) [40], reduced reference entropic differencing (RRED) [16], and regularity of phase congruency metric (SPCRM) [21]. The three NR methods were the spatial-spectral entropy-based quality index (SSEQ) [56], the derivative statistics-based quality evaluator (DESIQUE) [18], and the convolutional-neural-network-(CNN-) based IQA framework proposed in [57].

Before evaluating the performance of a particular quality assessment method on a particular database, we applied a logistic transform to bring the prediction values on the same scales as the MOS values. The logistic transform recommended by Video Quality Experts Group [58] is a four-parameter sigmoid given by

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp(-\frac{x-\tau_3}{|\tau_4|})} + \tau_2, \tag{25}$$

where x denotes the raw predicted score, and where τ_1, τ_2, τ_3, and τ_4 are free parameters selected to provide the best fit of the predicted scores to the MOS/DMOS values. Three criteria were used to measure the prediction monotonicity and prediction accuracy of each algorithm: (1) the Spearman Rank-Order Correlation Coefficient (SROCC), (2) the Pearson Linear Correlation Coefficient (CC), and (3) the Root Mean Square Error (RMSE) after non-linear regression. Note that the logistic transform in Eq. (25) will affect only CC and RMSE, not SROCC.

Table 5
Median/mean SROCC, CC, and RMSE values across 1000 train-test combinations on the TID2013 database.

		Median			Mean		
Metric	# bits	SROCC	CC	RMSE	SROCC	CC	RMSE
SSIM	8×N	0.695	0.713	0.881	0.695	0.713	0.881
MS-SSIM	8×N	0.861	0.868	0.624	0.862	0.868	0.623
VIF	8×N	0.740	0.799	0.756	0.741	0.800	0.753
MAD	8×N	0.854	0.862	0.636	0.854	0.863	0.635
WNISM	162	0.541	0.598	1.006	0.546	0.607	0.997
HWD	17×32	0.503	0.558	1.042	0.507	0.560	1.041
REDLOG	6×32	0.748	0.774	0.797	0.748	0.774	0.796
RRED	32N/36	0.834	0.854	0.654	0.835	0.855	0.651
SPCRM	32N* / 32	0.829	0.860	0.641	0.830	0.861	0.640
S4RR	10N/25.6	**0.887**	**0.901**	**0.545**	**0.884**	**0.899**	**0.549**

4.3. Cross-validation test on TID2013

In this section, we perform a cross-validation test on TID2013 by splitting the database into two non-overlapping sets based on the reference images content. A 20% subset of the database was used for training and a remaining 80% subset for testing. We compared with the four FR and five RR IQA methods in terms of median/mean SROCC, CC and RMSE values computed over 1000 trials, and the results are shown in Table 5. Also included is the amount of side information that each algorithm has to extract from the reference image, where N denotes the reference image size, and where N^* denotes the normalized image size according to [21]. Note that in order to provide fair comparison, all algorithms were tested on the same 80% subsets of images in TID2013. Also note that the three NR methods were not chosen to perform this cross-validation test because they were initially designed to be trained on the LIVE database [26].

In order to evaluate statistical significance, we also performed a one-sided t-test with a 95% confidence level between SROCC values generated by these algorithms across the 1000 train-test trials. The results are shown in Table 6, in which "1", "0", "−1" indicate that the mean correlation of the row (algorithm) is statistically superior, equivalent, or inferior to the mean correlation of the column (algorithm).

To demonstrate that the S4RR classification features can be used for different distortion family identification, Table 7 shows the median and mean classification accuracy of the classifier in the first stage of S4RR for seven distortion families in TID2013, as well as across all distortions (note that the numbers 1, 2,…,7 in Table 7 denote the seven distortion families). To show the performance consistency of each of the algorithms considered here, Fig. 6 shows a plot the mean and standard deviation of SROCC values across these 1000 trials.

According to the cross-validation test results, S4RR performs statistically the best among all FR/RR IQA algorithms considered here. In regards to the classification accuracy, S4RR still performs quite well, with an average of more than 90% of the images classified correctly. Note that in our cross-validation test, only 20% of the database images are used for training, and the remaining 80% for testing. As we have oberved, increasing the number of trained images in TID2013 will lead to significant performance improvement of S4RR. However, it is important to note that this cross-validation test is not necessarily the best way to gauge the performance of any given algorithm because the amounts of distortions (i.e., the severity levels) are similar in both training and testing sets. Thus, in the following sections, we show the results of these FR and RR IQA algorithms on other databases.

4.4. Performance on other databases

In this section, we evaluate the performance of S4RR for predicting the qualities of images on the LIVE, CSIQ, TID2008, IVC, and Toyama

Table 6
Results of the one-sided *t*-test performed between SROCC values generated by different measures. "1", "0", "−1" indicates that the algorithm in the row is statistically superior, equivalent, or inferior to the algorithm in the column.

	SSIM	MS-SSIM	VIF	MAD	WNISM	HWD	REDLOG	RRED	SPCRM	S4RR
SSIM	0	−1	−1	−1	1	1	−1	−1	−1	−1
MS-SSIM	1	0	1	1	1	1	1	1	1	−1
VIF	1	−1	0	−1	1	1	−1	−1	−1	−1
MAD	1	−1	1	0	1	1	1	1	1	−1
WNISM	−1	−1	−1	−1	0	1	−1	−1	−1	−1
HWD	−1	−1	−1	−1	−1	0	−1	−1	−1	−1
REDLOG	1	−1	1	−1	1	1	0	−1	−1	−1
RRED	1	−1	1	−1	1	1	1	0	1	−1
SPCRM	1	−1	1	−1	1	1	1	−1	0	−1
S4RR	1	1	1	1	1	1	1	1	1	0

Table 7
Mean and median classification accuracy across 1000 train-test trials.

Classification accuracy (%)	1	2	3	4	5	6	7	All
Mean	93.0	95.5	90.8	85.2	94.6	73.9	84.7	91.1
Median	93.1	95.6	91.0	85.5	95.0	77.5	86.0	91.2

databases. In this case, we trained S4RR on the entire TID2013 database, which contains 2875 distorted images (the CCS distorted images were not considered) with 23 distortion types. As mentioned before, these 23 distortion types were clustered into seven distortion families. Thus, our trained S4RR algorithm contains one SVC model and seven SVR models. For a fair comparison, the three NR IQA algorithms were also trained on these same 2875 distorted images in TID2013.

4.4.1. Contributions of individual sharpness features

To analyze the contributions of each of these sharpness (feature) maps towards the overall performance, we created nine abridged version of S4RR in which each version used only part of the regression features in the second stage of S4RR. Specifically, for the "FISH" and "S_{f_4} of FISH" versions, each employed three features for regression; for the "LSD" and "S_{f_i} of LSD" ($i = 1, \ldots, 4$) versions, each employed four features for regression. The other two versions, denoted by "FISH all" and "LSD all", respectively, indicate that all the six FISH-map-related features and all the 40 LSD-map-related features are used in the regression stage. Each of these versions was trained on the 2875 distorted images in the TID2013 database via the same 30 classification features and the same SVM parameters. The testing was performed on the aforementioned five databases, and the results of analysis are shown in Table 8.

As observed in Table 8, the best performance is achieved when all the 46 regression features are used. In fact, the 40 LSD-map-related features can already give rise to high SROCC and CC values on the five testing databases. Thus, the purpose of adding six more FISH-map-related features is to further enhance the algorithm performance. Also, as we have found in our experiment, extracting the same six features from the downsampled LLD-based FISH map will conversely decrease the algorithm performance, especially for the IVC and Toyama databases. Hence, in practice, we only use the luminance-based FISH map for regression feature extraction.

4.4.2. Overall performance

The overall testing results in terms of SROCC, CC, and RMSE on the entire set of images from the LIVE, CSIQ, TID2008, IVC, and Toyama databases are shown in Table 9, in which italicized entries denote FR algorithms, and entries marked by "*" denote the NR algorithms. The results of the best-performing FR algorithm in each case are italicized and bolded, and the results of the best-performing RR algorithm are bolded. Note that among the five RR algorithms, three of them (WNISM, HWD, and REDLOG) extract a much smaller amount of side information from the reference image, and thus their performances are not quite competitive. In comparison, RRED and SPCRM extract more side information proportional to the image size, and thus these two algorithms can achieve good results on the five testing databases.

From Table 9, it is clear that compared with other FR/RR/NR IQA methods, S4RR performs quite well in predicting quality. Although the side information (number of bits) extracted by S4RR is nearly equivalent to RRED and SPCRM (assuming that both algorithms use 32-bit

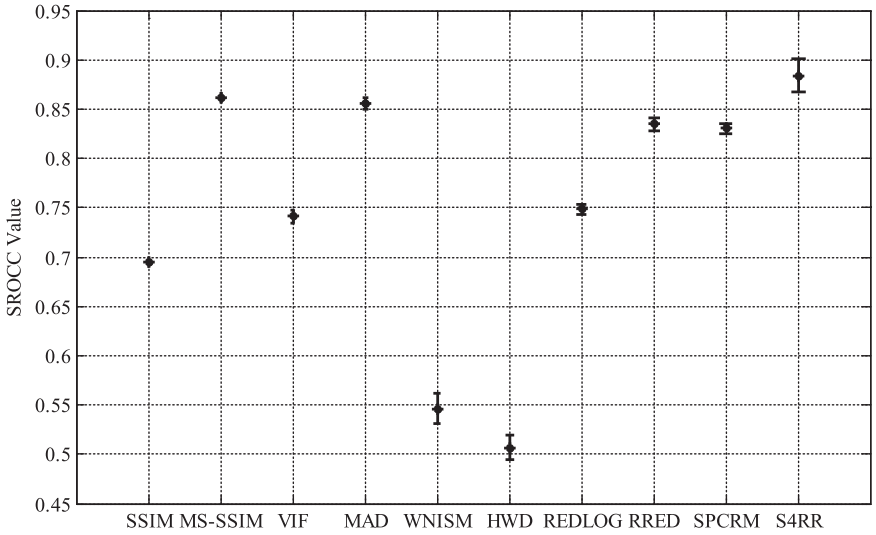

Fig. 6. Mean SROCC and standard error bars for various algorithms across the 1000 train-test trials on the TID2013 database.

Table 8
SROCC and CC for abridged versions of S4RR using only part of the feature maps in the distortion-family-specific QA stage. For reference, the SROCC and CC values of the full S4RR algorithm are also included (denoted by "ALL"). See text for details.

Feature Map	LIVE		CSIQ		TID2008		IVC		Toyama	
	SROCC	CC	SROCC	CC	SROCC	CC	SROCC	CC	SROCC	CC
FISH	0.900	0.890	0.802	0.806	0.800	0.820	0.786	0.778	0.816	0.813
S_{f_4} of FISH	0.850	0.838	0.744	0.761	0.784	0.799	0.683	0.688	0.737	0.764
FISH all	0.910	0.900	0.811	0.813	0.814	0.828	0.817	0.816	0.896	0.909
LSD	0.921	0.916	0.909	0.903	0.895	0.882	0.864	0.872	0.872	0.871
S_{f_1} of LSD	0.883	0.876	0.880	0.872	0.867	0.852	0.804	0.804	0.750	0.744
S_{f_2} of LSD	0.894	0.869	0.911	0.904	0.893	0.882	0.904	0.909	0.907	0.904
S_{f_3} of LSD	0.911	0.903	0.913	0.906	0.894	0.882	0.902	0.907	0.920	0.918
S_{f_4} of LSD	0.926	0.908	0.936	0.944	0.928	0.924	0.911	0.918	0.894	0.896
LSD all	0.946	0.939	0.949	0.954	0.945	0.946	0.916	0.924	0.919	0.922
ALL	0.954	0.951	0.955	0.959	0.953	0.953	0.926	0.932	0.930	0.932

Table 9
Overall performances of S4RR and other FR/RR/NR algorithms on the LIVE, CSIQ, TID2008, IVC, and Toyama databases. Italicized entries denote FR algorithms. Entries marked by "*" denote the NR algorithms. Results of the best-performing FR algorithm are italicized and bolded, and results of the best-performing RR algorithm are bolded.

		SSIM	*MS-SSIM*	*VIF*	*MAD*	CNN*	DESIQUE*	SSEQ*	WNISM	HWD	REDLOG	RRED	SPCRM	S4RR
SROCC	LIVE	0.910	0.945	0.963	**0.967**	0.730	0.740	0.796	0.747	0.748	0.932	0.943	0.944	**0.954**
	CSIQ	0.837	0.914	0.919	**0.947**	0.654	0.619	0.575	0.705	0.700	0.848	0.918	0.941	**0.955**
	TID2008	0.626	**0.853**	0.750	0.834	0.659	0.965	0.916	0.495	0.440	0.684	0.824	0.834	**0.953**
	IVC	0.779	0.884	0.896	**0.915**	0.669	0.511	0.566	0.401	0.459	0.786	0.899	0.923	**0.926**
	Toyama	0.786	0.886	0.909	**0.936**	0.777	0.628	0.414	0.645	0.796	0.825	0.823	0.883	**0.930**
	Average	0.750	0.889	0.849	**0.897**	0.679	0.799	0.771	0.599	0.583	0.786	0.875	0.889	**0.951**
CC	LIVE	0.901	0.934	0.959	**0.967**	0.708	0.738	0.785	0.736	0.742	0.923	0.934	0.941	**0.951**
	CSIQ	0.815	0.897	0.925	**0.950**	0.736	0.649	0.667	0.693	0.667	0.842	0.908	0.944	**0.959**
	TID2008	0.639	**0.839**	0.805	0.831	0.701	0.965	0.914	0.559	0.449	0.725	0.825	0.852	**0.953**
	IVC	0.792	0.893	0.903	**0.921**	0.672	0.539	0.593	0.520	0.504	0.794	0.905	0.929	**0.932**
	Toyama	0.796	0.893	0.914	**0.941**	0.775	0.621	0.410	0.654	0.815	0.823	0.827	0.888	**0.932**
	Average	0.750	0.878	0.876	**0.897**	0.713	0.806	0.790	0.630	0.581	0.802	0.871	0.898	**0.952**
RMSE	LIVE	11.850	9.782	7.733	**6.942**	19.295	18.447	16.920	18.489	18.324	10.484	9.784	9.245	**8.447**
	CSIQ	0.152	0.116	0.100	**0.082**	0.178	0.200	0.196	0.189	0.196	0.141	0.110	0.086	**0.074**
	TID2008	1.032	**0.730**	0.795	0.747	0.957	0.354	0.545	1.113	1.199	0.924	0.758	0.703	**0.407**
	IVC	0.743	0.549	0.524	**0.475**	0.902	1.026	0.981	1.041	1.052	0.741	0.519	0.450	**0.441**
	Toyama	0.757	0.564	0.507	**0.425**	0.791	0.981	1.142	0.947	0.725	0.712	0.704	0.575	**0.453**
	Average	3.078	2.477	2.067	**1.868**	4.627	4.191	3.962	4.546	4.543	2.736	2.493	2.339	**2.027**

quantizations for their quality-map values), the proposed algorithm can achieve better performance on all the five testing databases considered here. Specially, on the CSIQ, TID2008, and IVC databases, S4RR performs even better than the best FR IQA algorithms. Observe that SSEQ and DESIQUE can also achieve competitively high performances on TID2008, but their performances on other databases are relatively weak. These SROCC/CC values demonstrate that the proposed *distortion family* concept and the sharpness-based features are effective at the QA task across various databases.

The last rows of the SROCC, CC, and RMSE results in Table 9 show the average values, where the averages were weighted by the actual number of distorted images tested in each database. Also shown in Fig. 7 are scatter plots of logistic-transformed S4RR quality predictions versus subjective ratings (MOS or DMOS) on different databases. Despite the presence of some outliers, the plots are generally heteroscedastic. In summary, when looking at the overall performance across databases, S4RR has a better average performance than other RR IQA methods.

4.4.3. Performance on individual distortion types

In order to demonstrate that S4RR can achieve fairly good quality evaluation across different distortion types, we also report the performance of S4RR on individual distortion types from the five previously mentioned databases. The same logistic transform was used and the performance for each distortion type was evaluated by extracting the corresponding transformed scores previously computed when all the database images were considered. Table 10 shows the results in terms of SROCC values. Also included are results of the four FR, three NR, and five RR IQA algorithms for comparison. Italicized entries denote the FR IQA algorithms. Entries marked by "*" denote the NR IQA algorithms. Bold entries denote the best-performing FR/RR IQA algorithm for each distortion type on each database.

From the table, we see that S4RR provides fairly good or better predictions on most distortion types in comparison to the other five RR IQA algorithms. Specifically, on the LIVE and CSIQ databases, the three RR algorithms (RRED, SPCRM, and S4RR) seem to have equal performances; there are only minor variations when comparing their SROCC values on each distortion type. On the TID2008 database, S4RR and the two NR models (SSEQ and DESIQUE) inevitably demonstrate the best performances among all FR/RR/NR IQA algorithms considered. This is due to the fact that all these three algorithms were trained on TID2013, which shares distortion similarities with TID2008 (although the two databases were collected in two separate experiments with two separate subject pools), and thus these testing results are

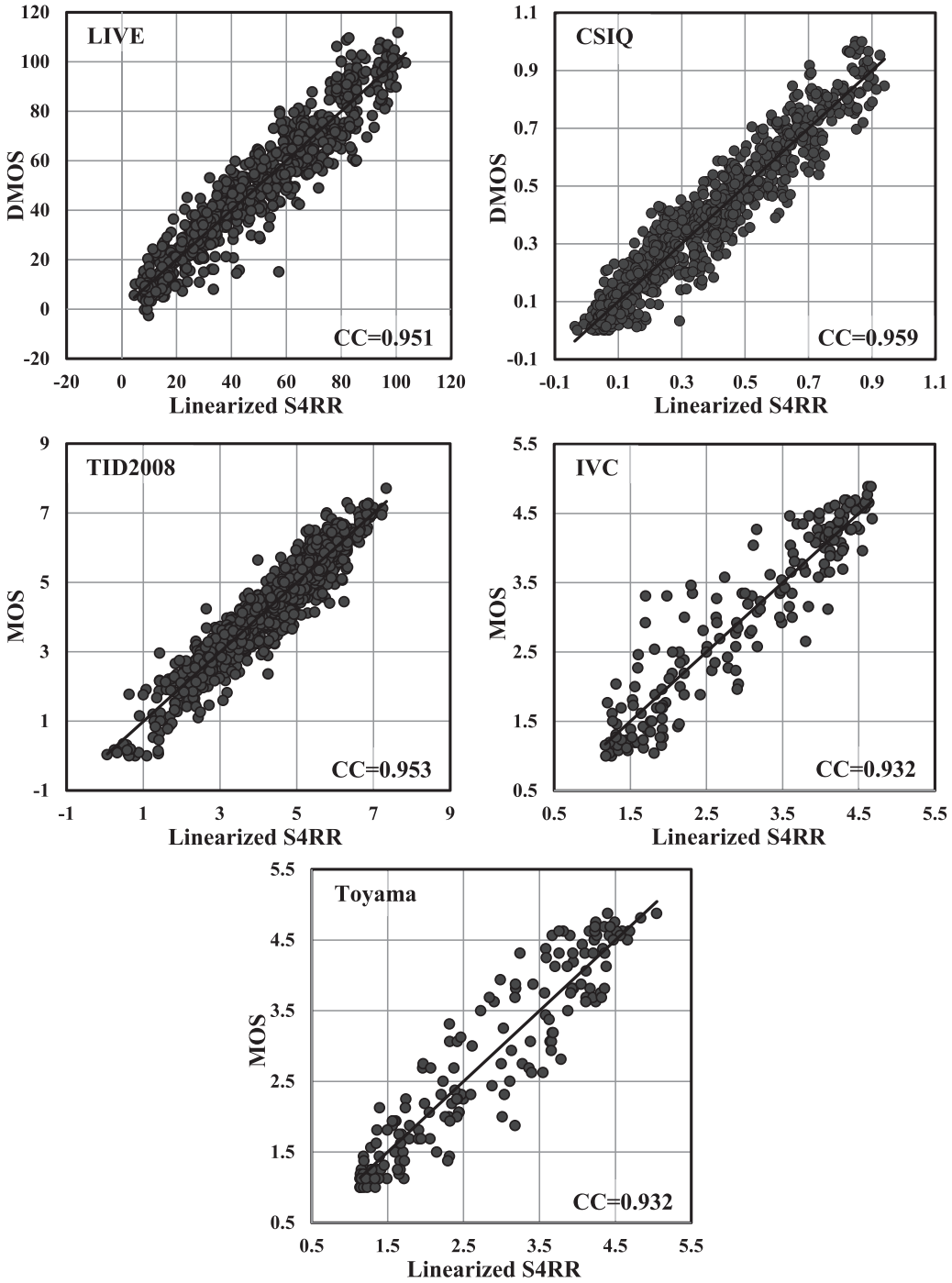

Fig. 7. Scatter plots of objective scores predicted by S4RR algorithm after logistic transform versus subjective scores on different image databases. Note that the x axis across all five figures represents the predicted value transformed via Eq. (25); the y axis represents the true differential mean opinion score (DMOS) value for the LIVE and CSIQ databases, true MOS value for the TID2008, IVC, and Toyama databases.

rather expected. In comparison, S4RR has some difficulties in predicting the qualities of images in the IVC and Toyama databases, which differ significantly from TID2013. Despite that, S4RR still performs fairly well on the four distortion types considered, and we believe that such a slight dip in performance is acceptable.

From Table 10, we also observe that although some distortion types [e.g., the fast fading (denoted by "FF") in LIVE, the additive pink Gaussian noise distortion (denoted by "APGN") in CSIQ, and the LAR coding distortion in IVC] are not "seen" in the training database, yet competitive performance on these distortions can still be achieved. This important finding indicates that our algorithm can possibly generalize to new distortions as long as this new distortion displays the same/similar quality degradation properties as those in the training database. We acknowledge that the classification process employed in our work has the potential of not being able to generalize to new distortion types. Indeed, this is a potential weakness that exists for all the machine-learning-based IQA approaches. Nonetheless, our soft-classification-driven approach has the potential to be robust to new distortion types, as long as those distortions share similar perceptual impacts as those in our existing distortion families.

In summary, when looking at the performance on individual distortion types, S4RR is still the best overall choice to estimate image quality.

Table 10
SROCC of S4RR and other FR/RR/NR quality assessment algorithms on different types of distortion on the LIVE, CSIQ, TID2008, IVC, and Toyama databases. Italicized entries denote FR algorithms. Entries marked by "*" denote the NR algorithms. Results of the best-performing FR algorithm are italicized and bolded, and results of the best-performing RR algorithm are bolded.

		SSIM	*MS-SSIM*	*VIF*	*MAD*	CNN*	DESIQUE*	SSEQ*	WNISM	HWD	REDLOG	RRED	SPCRM	S4RR
LIVE	JP2K	0.935	0.965	**0.968**	**0.968**	0.864	0.807	0.838	0.933	0.945	0.940	0.958	0.957	**0.969**
	JPEG	0.945	0.979	**0.984**	0.976	0.883	0.847	0.813	0.920	0.890	0.950	**0.976**	0.975	0.975
	AGN	0.963	0.973	**0.985**	0.984	0.471	0.775	0.762	0.870	0.870	0.939	**0.978**	0.973	0.965
	GBLUR	0.894	0.958	**0.972**	0.947	0.933	0.663	0.832	0.915	0.882	0.906	**0.968**	0.959	0.944
	FF	0.941	0.932	**0.965**	0.957	0.832	0.514	0.713	0.923	0.926	**0.950**	**0.943**	0.928	**0.943**
CSIQ	AGN	0.925	0.947	**0.957**	0.954	0.717	0.730	0.718	0.759	0.825	0.865	0.935	0.919	**0.941**
	JPEG	0.922	0.962	**0.970**	0.961	0.866	0.753	0.690	0.786	0.897	0.936	0.952	0.956	**0.959**
	JP2K	0.921	0.969	0.967	**0.975**	0.876	0.831	0.752	0.901	0.913	0.957	0.963	0.958	**0.970**
	FN	0.894	0.933	0.951	**0.957**	0.441	0.271	0.232	0.729	0.786	0.850	**0.936**	0.923	0.935
	GBLUR	0.924	0.972	**0.975**	0.968	0.880	0.711	0.747	0.907	0.920	0.948	0.963	**0.969**	0.958
	Contrast	0.740	**0.952**	0.936	0.921	0.061	0.254	0.172	0.830	0.928	**0.938**	**0.938**	0.924	0.913
TID2008	AGN	0.797	0.810	**0.880**	0.839	0.749	0.949	0.927	0.590	0.577	0.829	0.820	0.826	**0.939**
	ACN	0.811	0.806	**0.879**	0.825	0.756	0.924	0.910	0.568	0.588	0.796	0.850	0.818	**0.909**
	SCN	0.827	0.819	**0.870**	0.868	0.685	0.957	0.957	0.586	0.572	0.737	0.842	0.830	**0.955**
	MN	0.812	0.815	**0.870**	0.734	0.595	0.938	0.927	0.593	0.725	0.583	0.833	0.778	**0.926**
	HFN	0.843	0.868	**0.907**	0.886	0.795	0.968	0.965	0.677	0.693	0.914	0.909	0.890	**0.943**
	IN	0.746	0.687	**0.833**	0.065	0.609	0.944	0.936	0.567	0.532	0.467	0.741	0.677	**0.834**
	QN	0.803	**0.854**	0.796	0.816	0.735	0.953	0.937	0.631	0.477	0.685	0.831	0.856	**0.908**
	GBLUR	0.939	**0.961**	0.955	0.920	0.874	0.980	0.943	0.866	0.899	0.861	**0.957**	0.920	0.952
	ID	0.927	**0.957**	0.919	0.943	0.860	0.986	0.949	0.884	0.859	0.909	0.949	0.947	**0.973**
	JPEG	0.899	**0.935**	0.917	0.927	0.891	0.970	0.967	0.795	0.819	0.905	0.933	0.917	**0.949**
	JP2K	0.887	**0.973**	0.971	0.971	0.900	0.985	0.951	0.870	0.930	0.947	0.968	0.959	**0.981**
	JPEG_TE	0.819	**0.874**	0.858	0.866	0.666	0.956	0.919	0.801	0.846	0.752	0.870	0.884	**0.889**
	JP2K_TE	0.840	**0.853**	0.851	0.839	0.634	0.962	0.826	0.593	0.644	0.775	0.742	0.848	**0.948**
	NEPN	0.695	0.733	0.761	**0.829**	0.509	0.915	0.868	0.452	0.719	0.473	0.713	0.819	**0.867**
	LBW	**0.885**	0.761	0.832	0.797	0.135	0.698	0.655	0.583	0.742	0.745	0.824	**0.849**	0.785
	MS	0.718	**0.736**	0.514	0.516	0.114	0.769	0.458	0.285	0.480	0.615	0.538	**0.725**	0.518
	Contrast	0.484	0.640	**0.819**	0.272	0.078	0.931	0.878	0.651	0.696	0.629	0.542	0.036	**0.946**
IVC	JP2K	0.850	0.930	0.936	**0.940**	0.714	0.679	0.820	0.799	0.858	0.906	**0.968**	0.965	0.922
	JPEG	0.719	0.847	0.892	**0.923**	0.772	0.362	0.476	0.440	0.455	0.822	0.897	0.917	**0.924**
	LAR coding	0.712	0.875	0.888	**0.928**	0.737	0.431	0.319	0.709	0.467	0.645	0.862	0.885	**0.897**
	BLUR	0.869	0.944	**0.973**	0.964	0.783	0.616	0.839	0.852	0.901	0.919	**0.955**	0.941	0.945
Toyama	JP2K	0.914	0.945	**0.956**	0.955	0.683	0.747	0.370	0.911	0.903	0.888	0.921	0.921	**0.951**
	JPEG	0.626	0.835	0.907	**0.916**	0.868	0.500	0.455	0.872	0.818	0.825	0.771	**0.924**	0.913

5. Conclusion

This paper presented an RR IQA algorithm, called S4RR, which operates based on analyzing image sharpness maps. The proposed algorithm operates via two main stages. In the first stage, two FISH maps (luminance-based FISH map and downsampled LLD-based FISH map) are employed to extract 30 classification features, which are used to estimate the probabilities that the image is afflicted by one of the multiple distortion families. In the second stage, three sharpness maps (one luminance-based FISH map and two LSD maps) are employed to extract 46 regression features based on seven types of local statistics and six distance measures. These regression features, which represent the sharpness map differences between the reference and distorted images, are then fed into the corresponding SVR models to compute the distortion-family-specific quality scores. Finally, the probabilities from the first stage and the associated quality scores from the second stage are combined to yield an overall image quality prediction. Experimental results on various databases demonstrated that S4RR can achieve competitive or better performance than many other state-of-the-art FR/RR IQA algorithms.

Future work could build upon S4RR in several ways. One potential extension could be better incorporating the HVS modeling into the distortion-family specific QA stage. Contrast sensitivity models and all masking models could potentially supplement our existing QA stage. The main difficulty of such an approach, however, is the need for perceptual models that could operate primarily using the distorted image. Another way to build upon S4RR would be to use a multistage framework, in which multiple alternative measures are used besides sharpness. For example, specialized measures such as texture-based features or color-based features can possibly provide improved quality estimates depending on the distortion type. Thus, our current work is limited to sharpness, and we envision future work employing an array of other perceptually relevant measures.

Acknowledgments

The authors would like to thank the reviewers and associate editor for their helpful comments on this work. We are also thankful for the financial support provided by the Suzuki Foundation (Japan) and by the hardware support from Nvidia Corporation (USA).

References

[1] Z. Wang, E. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in: Electronic Imaging 2005, International Society for Optics and Photonics, pp. 149–159.
[2] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Top. Signal Process. 3 (2009) 202–211.
[3] F.Z. Lin Ma, Songnan Li, K.N. Ngan, Reduced-reference image quality assessment using reorganized DCT-based image representation, IEEE Trans. Multimed. 13 (2011) 824–829.
[4] A. Maalouf, M.-C. Larabi, C. Fernandez-Maloigne, A grouplet-based reduced reference image quality assessment, in: International Workshop on Quality of Multimedia Experience, IEEE, pp. 59–63.
[5] X. Gao, W. Lu, D. Tao, X. Li, Image quality assessment based on multiscale geometric analysis, IEEE Trans. Image Process. 18 (2009) 1409–1423.

[6] D. Tao, X. Li, W. Lu, X. Gao, Reduced-reference IQA in contourlet domain, IEEE Trans. Syst., Man, Cybern., Part B: Cybern. 39 (2009) 1623–1627.
[7] W. Xue, X. Mou, Reduced reference image quality assessment based on weibull statistics, in: Proceedings of the 2nd International Workshop on Quality of Multimedia Experience, IEEE, pp. 1–6.
[8] Z. Lin, J. Tao, Z. Zheng, Reduced-reference image quality assessment based on average directional information, in: IEEE Proceedings of the 11th International Conference on Signal Processing, volume 2, IEEE, pp. 787–791.
[9] I. P. Gunawan, M. Ghanbari, Reduced-reference picture quality estimation by using local harmonic amplitude information, in: London Communications Symposium, volume 2003.
[10] U. Engelke, M. Kusuma, H.-J. Zepernick, H.-J. Zepernick, Reduced-reference metric design for objective perceptual quality assessment in wireless imaging, Signal Process.: Image Commun. 24 (2009) 525–547.
[11] D. Bordevic, D. Kukolj, M. Pokric, I. Ostojic, Image quality assessment using reduced-reference nonlinear model, in: 2013 IEEE Proceedings of the 11th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, pp. 167–170.
[12] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (2012) 3378–3389.
[13] V. Bhateja, A. Srivastava, A. Kalsi, A reduced reference distortion estimation measure for color images, in: 2014 International Conference on Signal Processing and Integrated Networks, IEEE, pp. 699–704.
[14] V. Bhateja, A. Kalsi, A. Srivastava, Reduced reference IQA based on structural dissimilarity, in: 2014 International Conference on Signal Processing and Integrated Networks, IEEE, pp. 63–68.
[15] J. Wu, W. Lin, G. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimed. 15 (2013) 1700–1705.
[16] R. Soundararajan, A.C. Bovik, RRED indices: reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2012) 517–526.
[17] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (2011) 3350–3364.
[18] Y. Zhang, D.M. Chandler, No-reference image quality assessment based on log-derivative statistics of natural scenes, J. Electron. Imaging 22 (2013) 043025.
[19] N. Ponomarenko, M. Carli, F. Battisti, TID2008 - a database for evaluation of full-reference visual quality assessment metrics, Adv. Mod. Radioelectron. 10 (2009) 30–45.
[20] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, Color image database TID2013: Peculiarities and preliminary results, in: 2013 Proceedings of the 4th European Workshop on Visual Information Processing, IEEE, pp. 106–111.
[21] D. Liu, Y. Xu, Y. Quan, P. Le Callet, Reduced reference image quality assessment using regularity of phase congruency, Signal Process.: Image Commun. 29 (2014) 844–855.
[22] Z. Wang, E. Simoncelli, A. Bovik, Multiscale structural similarity for image quality assessment, in: Conference Record Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, volume 2, pp. 1398 – 1402, vol. 2.
[23] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2006) 430–444.
[24] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (2010) 011006.
[25] L. Zhang, D. Zhang, X. Mou, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (2011) 2378–2386.
[26] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, LIVE image quality assessment database release 2, 2005. ⟨http://live.ece.utexas.edu/research/quality/⟩.
[27] E.C. Larson, D. Chandler, Categorical image quality CSIQ database, Online, ⟨http://vision.okstate.edu/csiq⟩(2010).
[28] J. Caviedes, S. Gurbuz, No-reference sharpness metric based on local edge kurtosis, in: Proceedings of International Conference on Image Processing, volume 3, pp. III-53 - III-56.
[29] J. Dijk, M. Van Ginkel, R.J. Van Asselt, L.J. Van Vliet, P.W. Verbeek, A new sharpness measure based on gaussian lines and edges, in: International Conference on Computer Analysis of Images and Patterns, Springer, pp. 149–156.
[30] C. Vu, D.M. Chandler, S3: A spectral and spatial sharpness measure, in: 2009 Proceedings of the First International Conference on Advances in Multimedia.
[31] R. Hassen, Z. Wang, M. Salama, No-reference image sharpness assessment based on local phase coherence measurement, in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 2434–2437.
[32] C. Vu, T. Phan, D. Chandler, S3: a spectral and spatial measure of local perceived sharpness in natural images, IEEE Trans. Image Process. 21 (2012) 934–945.
[33] P.V. Vu, D.M. Chandler, A fast wavelet-based algorithm for global and local image sharpness estimation, IEEE Signal Process. Lett. 19 (2012) 423–426.
[34] P. Vu, D. Chandler, A no-reference quality assessment algorithm for JPEG2000-compressed images based on local sharpness, in: IS & T/SPIE Electronic Imaging, International Society for Optics and Photonics, pp. 865302–865302.
[35] Y. Zhang, A.K. Moorthy, D.M. Chandler, A.C. Bovik, C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes, Signal Process.: Image Commun. 29 (2014) 725–747.
[36] M. Carnec, P. Le Callet, D. Barba, Visual features for image quality assessment with reduced reference, in: IEEE International Conference on Image Processing, volume 1, IEEE, pp. I-421.
[37] M. Carnec, P. Le Callet, D. Barba, Objective quality assessment of color images based on a generic perceptual reduced reference, Signal Process.: Image Commun. 23 (2008) 239–256.
[38] M. Zhang, X. Mou, H. Fujita, L. Zhang, X. Zhou, W. Xue, Local binary pattern statistics feature for reduced reference image quality assessment, in: IS & T/SPIE Electronic Imaging, International Society for Optics and Photonics, pp. 86600L–86600L.
[39] J. Farah, M.-R. Hojeij, J. Chrabieh, F. Dufaux, Full-reference and reduced-reference quality metrics based on SIFT, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 161–165.
[40] S.A. Golestaneh, L.J. Karam, Reduced-reference quality assessment based on the entropy of dnt coefficients of locally weighted gradients, in: 2015 IEEE International Conference on Image Processing, IEEE, pp. 4117–4120.
[41] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.
[42] Y. Zhang, D.M. Chandler, 3D-MAD: a full reference stereoscopic image quality estimator based on binocular lightness and contrast perception, IEEE Trans. Image Process. 24 (2015) 3810–3825.
[43] L. Beaurepaire, K. Chehdi, B. Vozel, Identification of the nature of noise and estimation of its statistical parameters by analysis of local histograms, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 4, IEEE, pp. 2805–2808.
[44] D.-C. Chang, W.-R. Wu, Image contrast enhancement based on a histogram transformation of local standard deviation, IEEE Trans. Med. Imaging 17 (1998) 518–531.
[45] R. Wang, A. Hanson, E. Riseman, Texture analysis based on local standard deviation of intensity, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 482–488.
[46] A. Gore, H. Kansal, S. Gupta, Local standard deviation based image quality metrics for JPEG compressed images, Indones. J. Electr. Eng. Comput. Sci. 12 (2014) 7280–7286.
[47] S. Knerr, L. Personnaz, G. Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network, in: Neurocomputing, Springer, 1990, pp. 41–50.
[48] H.-T. Lin, C.-J. Lin, R.C. Weng, A note on platt's probabilistic outputs for support vector machines, Mach. Learn. 68 (2007) 267–276.
[49] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[50] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[51] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27.
[52] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[53] P. Le Callet, F. Autrusseau, Subjective quality assessment irccyn/ivc database, 2005. ⟨http://www.irccyn.ec-nantes.fr/ivcdb/⟩.
[54] Y.K.Z.M.P.S. Yuukou Horita, Keiji Shibata, Subjective quality assessment Toyama database, 2008. ⟨http://mict.eng.u-toyama.ac.jp/mict/⟩.
[55] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.-H. Yang, A.C. Bovik, Quality-aware images, IEEE Trans. Image Process. 15 (2006) 1680–1689.
[56] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Signal Process.: Image Commun. 29 (2014) 856–863.
[57] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733–1740.
[58] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii (2003). ⟨http://www.vqeg.org⟩.