Dehazed Image Quality Evaluation: From Partial Discrepancy to Blind Perception

Wei Zhou, *Member, IEEE*, Ruizeng Zhang, Leida Li, *Member, IEEE*, Guanghui Yue, Jianwei Gong, Huiyan Chen, and Hantao Liu, *Senior Member, IEEE*

***Abstract*—Nowadays, vision oriented intelligent vehicle systems such as autonomous driving or transportation assistance can be optimized by enhancing the visual visibility of images acquired in bad weather conditions. The presence of haze in such visual scenes is a critical threat. Image dehazing aims to restore spatial details from hazy images. There have emerged a number of image dehazing algorithms, designed to increase the visibility of those hazy images. However, much less work has been focused on evaluating the visual quality of dehazed images. In this paper, we propose a Reduced-Reference dehazed image quality evaluation approach based on Partial Discrepancy (RRPD) and then extend it to a No-Reference quality assessment metric with Blind Perception (NRBP). Specifically, inspired by the hierarchical characteristics of the human perceiving dehazed images, we introduce three groups of features: luminance discrimination, color appearance, and overall naturalness. In the proposed RRPD, the combined distance between a set of sender and receiver features is adopted to quantify the perceptually dehazed image quality. By integrating global and local channels from dehazed images, the RRPD is converted to NRBP which does not rely on any information from the references. Extensive experiment results on both synthetic and real dehazed image quality databases demonstrate that our proposed methods outperform state-of-the-art full-reference, reduced-reference, and no-reference quality assessment models. Furthermore, we show that the proposed dehazed image quality evaluation methods can be effectively applied to tune parameters for image dehazing algorithms and have the potential to be deployed in real transportation systems.**

***Index Terms*—Image dehazing, quality evaluation, reduced-reference, blind/no-reference, partial discrepancy, human visual perception.**

I. Introduction

THE visibility of images is vital not only for delivered consumers, but also for intelligent driving and vehicle systems [1]–[5]. We usually use visible light imaging equipment to acquire visual scenes. In such cases, when the outdoor environment is captured, image visibility would be inevitably degraded by possible bad weather conditions due to the scattering or absorption of light by atmospheric particles [6]. Among them, haze is one of the representative atmospheric phenomena. To reduce the influence of visibility damage caused by hazy effects, many image dehazing algorithms [7]–[9] have been proposed and achieved great success in the field of vision-related systems.

W. Zhou and H. Liu are with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, United Kingdom (email: zhouw26@cardiff.ac.uk; liuh35@cardiff.ac.uk).

R. Zhang, J. Gong and H. Chen are with the School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China (email: gongjianwei@bit.edu.cn).

L. Li is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China. (email: ldli@xidian.edu.cn).

G. Yue is with the School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China. (email: yueguanghui@szu.edu.cn).

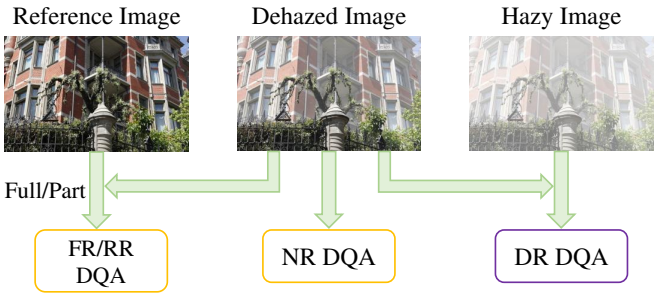


Fig. 1: Illustration of objective DQA categories, where FR and RR DQA methods use full and part reference information, respectively. The DR DQA models rely on hazy images, while NR DQA directly evaluates the perceptual quality from dehazed images.

With the development of image dehazing, dehazed images with a variety of appearances are generated by different image dehazing models. One nature question that arises here is how to effectively evaluate and compare the performance of various image dehazing approaches? Because human subjects are the ultimate receivers of dehazed images, subjective quality evaluation [10] is the most accurate and reliable way to measure the visual quality of dehazed images. Moreover, based on subjective testing, several benchmarking dehazed image quality databases [11]–[13] have been built, including regular and aerial scenarios. Nevertheless, subjective experiments are often time and labor-consuming. The objective quality evaluation [14] is an alternative to subjective ones, in which particularly objective metrics are designed to automatically estimate the perceptually dehazed image quality.

As shown in Fig. 1, similar to traditional image quality assessment (IQA), for objective dehazed quality assessment (DQA), there are generally four categories regarding to existing DQA metrics. These consist of full-reference (FR), reduced-reference (RR), degraded-reference (DR), and no-reference (NR) DQA models. Among them, the FR DQA supposes the pristine reference image is totally available, while the RR DQA only relies on part of the corresponding reference information. Since the hazy image is the degraded counterpart of the original reference image, based on this, the DR DQA methods can be developed. When both haze-free and hazy

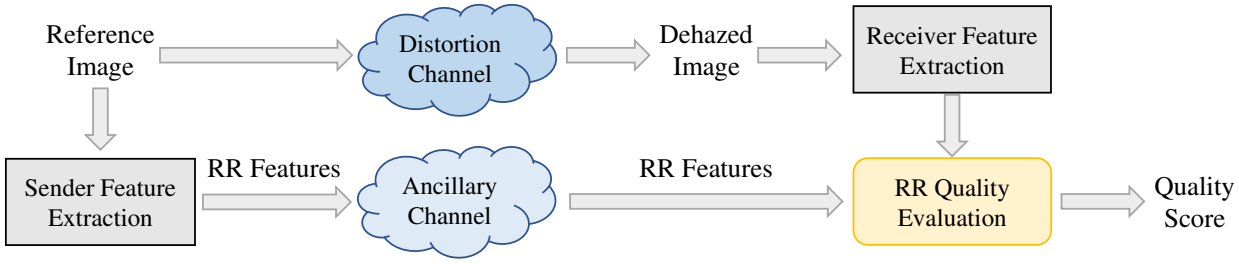


Fig. 2: Deployment of RR quality evaluation systems.

images are unavailable, we may resort to the NR DQA models which directly predict the visual quality from dehazed images.

Although lots of objective quality evaluation methods have shown effectiveness in estimating human perceived quality for conventional distortions such as noise [15], JPEG compression [16], blur [17], etc, the DQA models specifically designed for dehazed images are still in their infancy. As for FR and DR DQA situations, either original haze-free or degraded hazy images can be fully accessible, which is beneficial to the algorithm design of dehazed quality evaluation. However, in practical real-world applications, we may not obtain full information of them. In other words, it becomes more challenging for RR and NR DQA tasks that we focus on. Consequently, in this paper, we first propose a novel RR DQA method and then extend it to a more practical NR DQA model that does not rely on any other information except for the dehazed images. To the best of our knowledge, only [18] developed an RR image quality assessment called perceptual discrepancy learning (PDL) that adapted to dehazed images. We are one of the first to propose a specific RR DQA method and have such an extension to the NR manner. Besides, our proposed quality evaluation models can be applied to tune parameters for image dehazing algorithms.

In our proposed frameworks, to evaluate the perceptual quality of dehazed images produced by various image dehazing algorithms, we consider the cognitive mechanisms of the human visual system (HVS) [19]. To be specific, when subjects perceive dehazed images, luminance discrimination and color appearance [20], [21] are observed from low-level dimensions. Then, the high-level overall naturalness [22], [23] is also taken into account. This process is more likely to be reflected by the hierarchical properties of the human brain [24]. Additionally, the HVS tends to aggregate the visual quality impression from both global and local channels [25]. The codes are publicly available at https://github.com/weizhou-geek/Dehazed-Image-Quality-Evaluation.

Around the above-mentioned aspects, this paper makes the following main contributions:

- We propose RR and NR dehazed quality evaluation methods via partial discrepancy and blind perception, respectively.
- Motivated by the hierarchical properties of dehazed image perceiving process, we propose quality-aware features, including luminance discrimination, color appearance, and overall naturalness.
- Based on the characteristics of the HVS, the global and local aggregation is taken into consideration. Extensive experiments on both synthetic and authentic dehazed quality databases as well as real transportation scenes verify that our proposed metrics have superior performance. In addition, we optimize potential parameter-based image dehazing algorithms by applying our proposed dehazed image quality evaluation methods.

The rest of this paper is organized as follows. The related work of our paper is presented in Section II. Section III introduces the proposed Reduced-Reference dehazed image quality evaluation approach based on Partial Discrepancy (RRPD). In Section IV, we then propose the No-Reference quality assessment metric with Blind Perception (NRBP), which combines both global and local channels. We present experimental results and analysis in Section V, and finally conclude the work in Section VI.

II. Related Work

Objective IQA designs computational models to predict the human perception-based image quality, which is also known as mean opinion score (MOS). The simplest FR IQA model is the peak signal-to-noise ratio (PSNR) which compares the pixel difference between reference and distorted images. However, such a signal fidelity model does not consider the characteristics of the HVS. Therefore, the PSNR cannot estimate the accurate visual quality as human observers perceive. To overcome the drawbacks of signal fidelity, Wang et al. proposed the well-known structural similarity (SSIM) index [26] based on the visual perception of image structures. Besides, several variants of similarity measurement were also developed, such as the multiscale SSIM (MS-SSIM) index [27], the information content weighted SSIM (IW-SSIM) measure [28], the feature similarity (FSIM) index [29], the gradient similarity (GSM) model [30], the gradient magnitude similarity deviation (GMSD) [31], the perceptual similarity (PSIM) measure [32], and the superpixel-based similarity (SPSIM) index [33], etc. Other FR IQA methods called information fidelity criterion (IFC) [34] and visual information fidelity (VIF) [35] evaluate visual quality from the aspect of image information.

Apart from the FR IQA approaches, many NR IQA methods have been proposed during the past decades. According to a spatial natural scene statistic (NSS) model, the classical

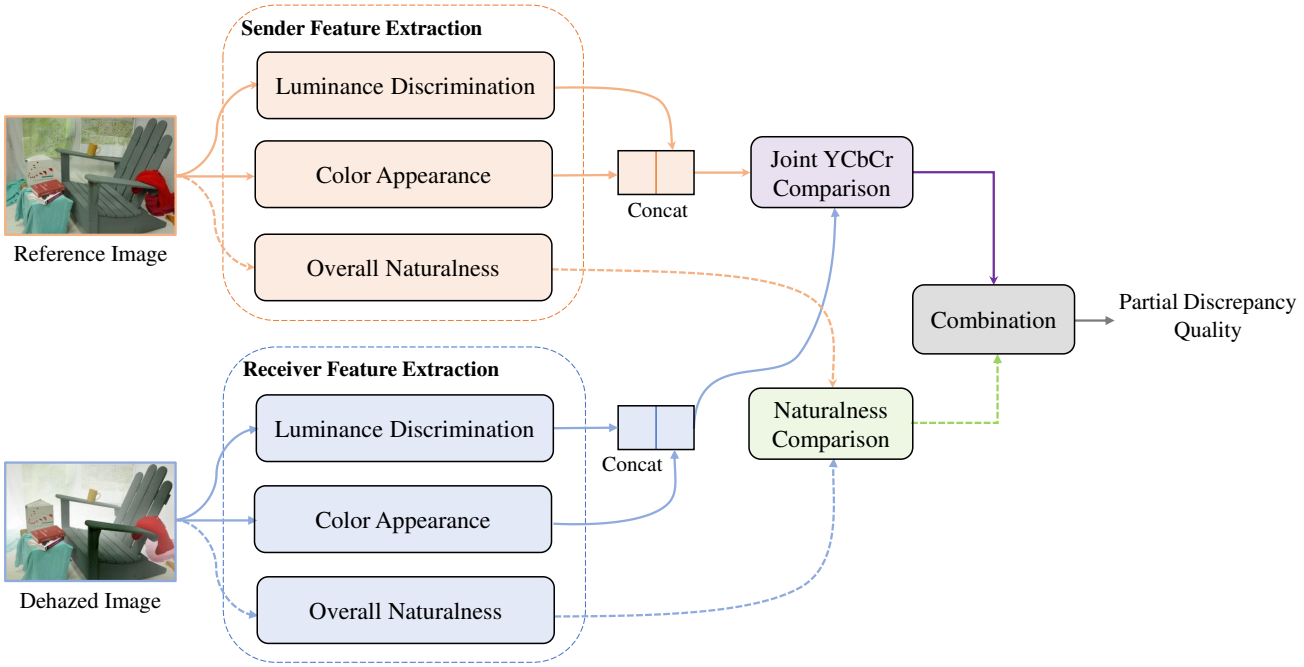


Fig. 3: Framework of the proposed RR quality evaluation method, where luminance discrimination, color appearance, and overall naturalness are employed for feature extraction.

blind/referenceless image spatial quality evaluator (BRISQUE) was built [36]. Other NSS models were also presented, e.g., the blind image quality index (BIQI) [37], the natural image quality evaluator (NIQE) [38], the blind image integrity notator using DCT statistics (BLIINDS-II) [22], the distortion identification-based image verity, and the integrity evaluation (DIIVINE) index [23], etc. In [39], a highly efficient NR IQA model named local pattern statistics index (LPSI) was presented to evaluate the image quality. The multi-task end-to-end optimized deep neural network (MEON) was proposed in [40], which consists of a distortion identification network and a quality prediction network.

The RR IQA is a kind of objective way between FR IQA and NR IQA. The development of RR IQA falls behind and few models have been proposed for this category. For example, the reduced-reference image quality assessment method (RRIQA) [41] was proposed according to an image statistic model in the wavelet transform domain. Afterward, based on the internal generative mechanism and visual saliency detection, the RQMSH [42] and SIRR [43] were proposed to predict the distorted image quality with side information from reference data.

These traditionally generic IQA models introduced above are mainly designed for the quality assessment of natural images. However, the properties of natural images differ from that of dehazed images, even the artifacts generated by various image dehazing algorithms are unlike the traditional distortions such as common Gaussian white noise in natural images. Thus, it is necessary to devise specially dehazed quality evaluation methods. In the literature, representative FR DQA methods include the DEHAZEfr [11] and FRFSIM [13]. Based on visibility and distortion measurement, the NR DQA metric named VDA-DQA [44] was developed by complex contourlet transform. Another NR DQA model was proposed in the HSI color space [45]. Moreover, pairwise comparison-based rank learning was employed to form the NR image restoration quality algorithm [46] which is also suitable for the image dehazing scenario. Due to the possible hazy images, one can also propose DR DQA models, e.g., the objective dehazing quality index (DHQI) [14].

For the quality metrics of RR DQA, only [18] was proposed for image restoration and adapted to image dehazing. Considering that none of the previous attempts specifically focus on the RR DQA problem, in this work, we aim to fill in this blank. Specifically, we propose a partial discrepancy based RR DQA model and then extend it to the NR DQA method with blind perception. Furthermore, our dehazed quality evaluation metrics can also be applied to optimize the parameter selection for the potential image dehazing algorithm.

III. Proposed RR Quality Evaluation Method

In Fig. 2, we give the deployment of RR quality evaluation systems. According to this deployment, there is a feature extraction process at the sender side and a feature extraction followed by a quality evaluation process at the receiver side, for the reference and dehazed images, respectively. The dehazed image is generated through the distortion channel. Moreover, the RR features obtained from sender feature extraction usually have a much lower data rate than the original image data and are transmitted to the receiver by an ancillary channel.

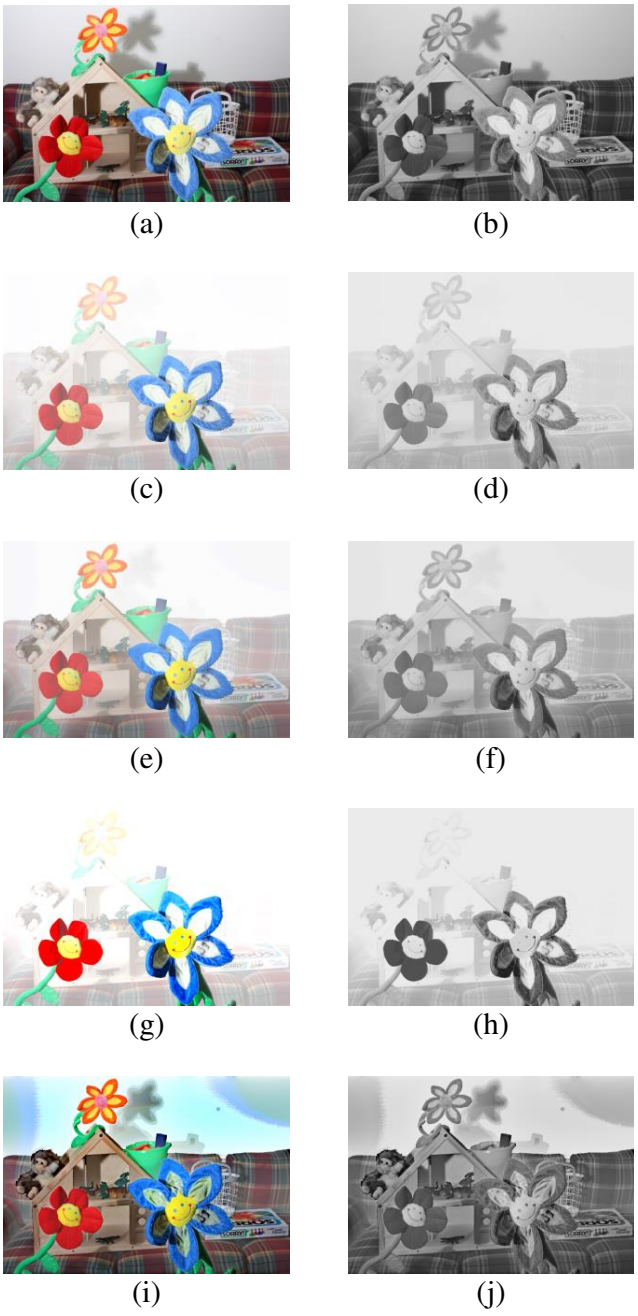


Fig. 4: Luminance quality degradation from image dehazing. (a) A haze-free image; (c) The corresponding hazy image; (e) Dehazed image with haze removal degradation; (g) Dehazed image with structural recovery degradation; (i) Dehazed image with enhancement effects degradation; (b), (d), (f), (h) and (j) are the luminance maps for (a), (c), (e), (g) and (i) respectively.

The framework of our proposed RR quality evaluation method for dehazed images, i.e., RRPD, is shown in Fig. 3. In the proposed RR DQA method, motivated by the hierarchically dehazed image perceiving process, we first develop three groups of quality-aware features, containing luminance discrimination, color appearance, and overall naturalness. Then, the feature comparison and combination are used for partial discrepancy quality estimation. Since real-time performance is very important for automotive applications, we test the time complexity for our proposed framework is 1.32s. Note that for a real-time video scene, our quality metric can be tested every once in a while since the haze level generally could not have huge changes in successively neighboring frames. Therefore, we can use the proposed method in real-world transportation scenarios as an assistant way to monitor image quality, which will benefit automotive applications. In the following subsections, we will introduce the technical details of RRPD.

A. Luminance Discrimination

When human perceives the dehazed images generated by different image dehazing algorithms, luminance information significantly influences perceptual image quality [21]. Fig. 4 shows some examples of the typical luminance quality degradation from image dehazing, where the corresponding luminance maps of reference, hazy, and dehazed images are presented. Our observation is that the quality degradation in the aspect of luminance information mainly comes from haze removal, structural recovery, and enhancement effects. Therefore, we propose the luminance discrimination features described as follows.

1) Haze Removal: The haze in the image causes visibility loss. A lot of mainstream image dehazing approaches have been proposed to remove the haze as much as possible, which can help to increase visibility. Under the circumstances, more haze removal can reflect better visual quality. Here we employ several statistics in the luminance component to quantify the haze removal degrees. Let R denote the luminance map of reference image. The resolution is $M \times N$. To reveal the haze removal, we first compute the mean and standard deviation of luminance map by:

$$\mathrm{MEA} = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} R(m,n), \tag{1}$$

$$\mathrm{STD} = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} [R(m,n) - \overline{R}]^2}, \tag{2}$$

where $\overline{R}$ is the mean operation of luminance map, which is the same as Eqn (1). Then, the median and mode values of luminance map can be obtained by:

$$\mathrm{MED} = \begin{cases} \widehat{R}\left[\frac{M \times N + 1}{2}\right], & \text{if } (M \times N) \text{ is odd} \\ \frac{\widehat{R}\left[\frac{M \times N}{2}\right] + \widehat{R}\left[\frac{M \times N}{2} + 1\right]}{2}, & \text{if } (M \times N) \text{ is even} \end{cases}, \tag{3}$$

$$\mathrm{MOD} = \{R_i \mid P(R_i) \geq P(R_j), i \neq j\}, \tag{4}$$

where $\widehat{R}$ is the ordered list of the luminance map. $P(\cdot)$ indicates the probability.

Except for the mean, standard deviation, median, and mode statistics, the global entropy that determines the 'surprise' of a specific image is also taken into consideration as:

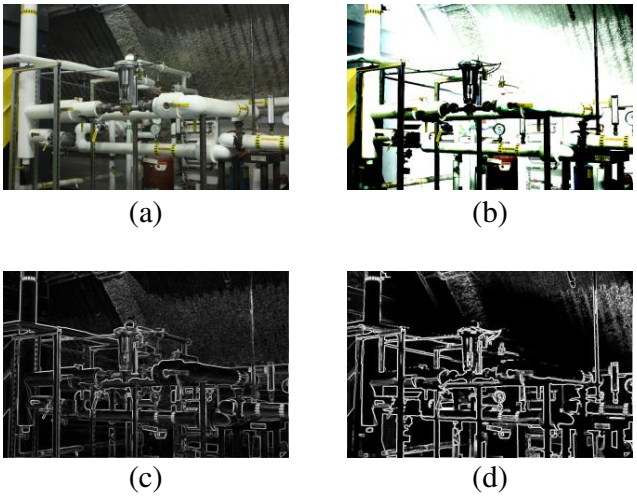


Fig. 5: Comparison of CSF weighted gradient maps. (a) A reference haze-free image; (b) Test dehazed image; (c),(d) The corresponding CSF weighted gradient map of (a) and (b).

$$\mathrm{ENT} = -\sum_{k=0}^{K} P_k \log\left(P_k\right), \tag{5}$$

where K is the maximum pixel value in the whole luminance map. Besides, P_k represents the probability of pixel value equaling to k. With all the extracted luminance statistics, we can exploit them to form the quantification of haze removal degradation.

2) Structural Recovery: Since structures are important for image quality evaluation, many IQA models based on image structures have been proposed [26], [47]–[49]. One main goal of image dehazing algorithms is also to recover the structures. From Fig. 4 (g-h), we find that part of the structural information is missing in the dehazed images, which leads to quality degradation. Thus, according to the characteristics of the HVS, i.e., the contrast sensitivity function (CSF) [50], we first filter the luminance map and then compute the gradient map weighting by CSF to reflect image structures.

Specifically, the CSF can measure the HVS sensitivity to various frequencies of visual stimulus. Based on CSF, human eyes have different abilities to discriminate target brightness at different spatial frequencies. Here we employ a modified CSF function [51] given by:

$$H(f,\varphi) = H_1(f,\varphi)H_2(f,\varphi), \tag{6}$$

where f denotes the radial spatial frequency with cycles per degree of visual angle (c/deg). Moreover, $\varphi \in [-\pi,\pi]$ represents the angular frequency. $H_1(f,\varphi)$ and $H_2(f,\varphi)$ are the frequency response from a circularly symmetric Gaussian filter and the frequency response of a CSF model originally proposed in [52] and adjusted in [53]. These can be computed as follows:

$$H_1(f,\varphi) = e^{-2\pi^2\alpha^2 f^2}, \tag{7}$$

$$H_2(f,\varphi) = \begin{cases} 2.6\left(0.0192+\eta f_\varphi\right)e^{-\eta f_\varphi}, f \geq f_{\text{peak}} \\ 0.981, \text{ otherwise} \end{cases}, \tag{8}$$

where $\alpha = 0.5$ is adopted to control the filter cutoff. Inspired by [54], we set $\eta = 0.114$. f_φ is used to represent direction-based correction of f to reduce the contrast sensitivity along the diagonal, which can be calculated by:

$$f_\varphi = \frac{f}{0.15\cos(4\varphi)+0.85}. \tag{9}$$

We exploit the modified CSF function $H(f,\varphi)$ to filter the luminance map. As shown in Fig. 5, we then obtain the CSF weighted gradient map, which can reflect the quality variation. Finally, the mean, standard deviation, median, mode, and entropy statistics of CSF weighted gradient map are adopted to measure the structural recovery degradation.

3) Enhancement Effects: By using different image dehazing algorithms, the resulting dehazed images may be over-enhanced, as demonstrated by Fig. 4 (i-j). In order to measure the contrast variation caused by image dehazing, we use the mean of local variance and normalized local variance [14] which can determine the enhancement effects.

Specifically, the local variance of luminance map R is computed by:

$$\sigma(x,y) = \sqrt{\sum_{u,v} w(u,v)[R(x+u,y+v)-\mu(x,y)]^2}, \tag{10}$$

where x, y indicate the pixel indexes in the spatial domain. w is a Gaussian weighting window. Moreover, μ represents the local mean, which can be calculated as:

$$\mu(x,y) = \sum_{u,v} w(u,v)R(x+u,y+v). \tag{11}$$

Considering that the local variance is sensitive to the local mean, we also derive the normalized local variance as follows:

$$\gamma(x,y) = \frac{\sigma(x,y)}{\mu(x,y)+C}, \tag{12}$$

where C represents a small positive constant used for avoiding instability. After obtaining the local variance and normalized local variance of luminance map, the mean values of them are employed to quantify the quality degradation from enhancement effects.

B. Color Appearance

Apart from luminance information, color cues are also useful to help the HVS perceive visual signals. There exist many works that have investigated the effects of luminance and chrominance on perceptual image quality [55]–[58]. In these works, chrominance has been proven to achieve a promising gain for the image quality evaluation task. Therefore, we first convert the original image I into YCbCr components by:

$$I \rightarrow \{R, Cb, Cr\}, \tag{13}$$

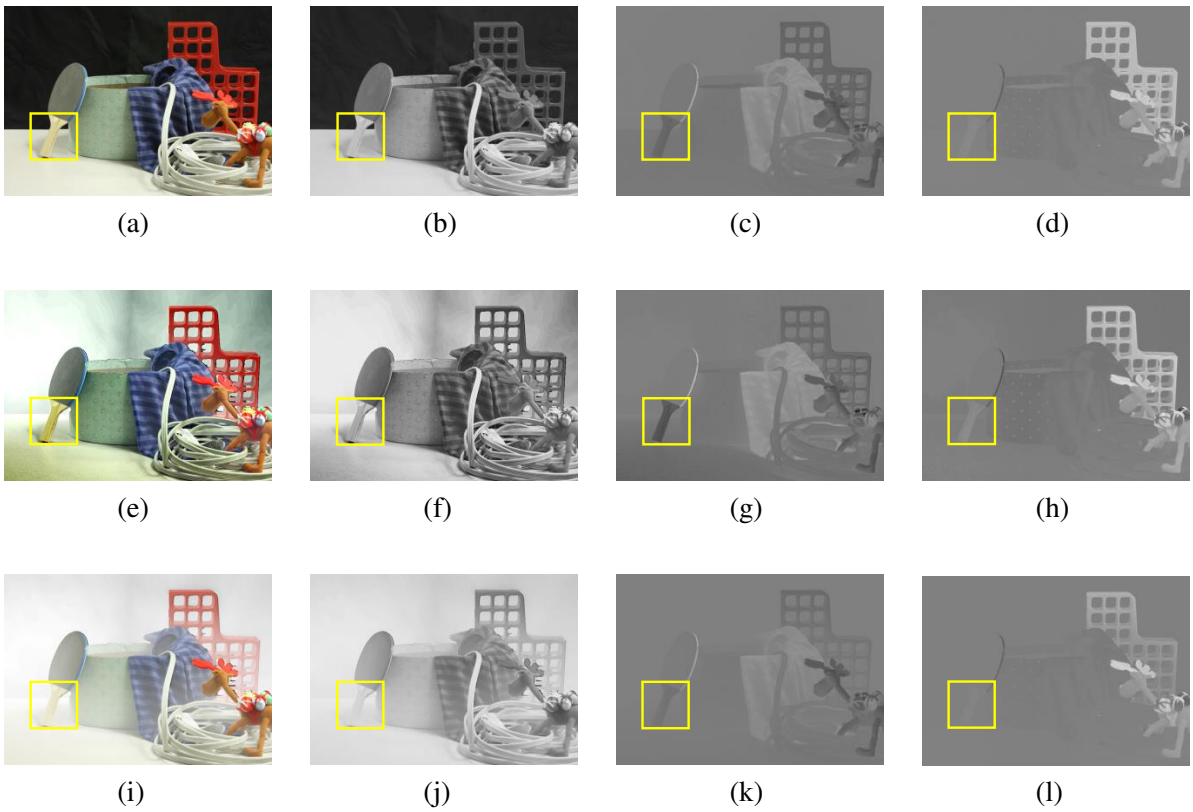


Fig. 6: Color decomposition results. (a) A reference haze-free image; (e) Dehazed image produced by Berman16 [9]; (i) Dehazed image generated by Lai15 [8]; (b-d) The corresponding YCbCr components of (a); (f-h) The corresponding YCbCr components of (e); (j-l) The corresponding YCbCr components of (i).

where R is the reference luminance map. Cb and Cr are two chrominance maps.

We show the color decomposition results in Fig. 6, which include two image dehazing algorithms, i.e., Berman16 [9] and Lai15 [8]. From this figure, it can be seen the color differences in the reference and dehazed ones, indicating that color information can be applied to measure the dehazed image quality. Thus, we then compute the mean, standard deviation, median, mode, and entropy statistics of chrominance to form the color appearance features.

C. Overall Naturalness

Since the HVS works in a hierarchical manner, we utilize the higher level overall naturalness features of the YCbCr components to estimate the quality degradation from a natural sense. To be specific, we first realize the decorrelation of input image, e.g., luminance map R with divisive normalization transform [59] as below:

$$\tilde{R}(x,y) = \frac{R(x,y) - \mu(x,y)}{\sigma(x,y) + C}, \tag{14}$$

where the local variance σ and local mean μ are obtained from Eqn (12) and Eqn (13), respectively. C is the same constant as Eqn (14). Apart from the decorrelation, we also apply the whitening filter to make the distribution more Gaussian-like.

In Fig. 7, we show the statistical distribution for YCbCr components. We see that the behaviors of curves are different for various image dehazing algorithms. Therefore, we then use the generalized Gaussian distribution (GGD) and more general asymmetric generalized Gaussian distribution (AGGD) [60] to fit the statistical distribution. The AGGD with zero mean value mode is given as:

$$g\left(\tau;\lambda,\rho_l,\rho_r\right) = \begin{cases} \frac{\lambda}{(\varsigma_l+\varsigma_r)\Gamma\left(\frac{1}{\lambda}\right)} e^{-\left(\frac{-\tau}{\varsigma_l}\right)^{\lambda}}, \tau < 0 \\ \frac{\lambda}{(\varsigma_l+\varsigma_r)\Gamma\left(\frac{1}{\lambda}\right)} e^{-\left(\frac{-\tau}{\varsigma_r}\right)^{\lambda}}, \text{ otherwise} \end{cases}, \tag{15}$$

where

$$\varsigma_l = \rho_l \sqrt{\frac{\Gamma\left(\frac{1}{\lambda}\right)}{\Gamma\left(\frac{3}{\lambda}\right)}}, \tag{16}$$

$$\varsigma_r = \rho_r \sqrt{\frac{\Gamma\left(\frac{1}{\lambda}\right)}{\Gamma\left(\frac{3}{\lambda}\right)}}, \tag{17}$$

and λ controls the distribution shape. Γ represents the Gamma function. ρ_l and ρ_r are the scale parameters of left and right

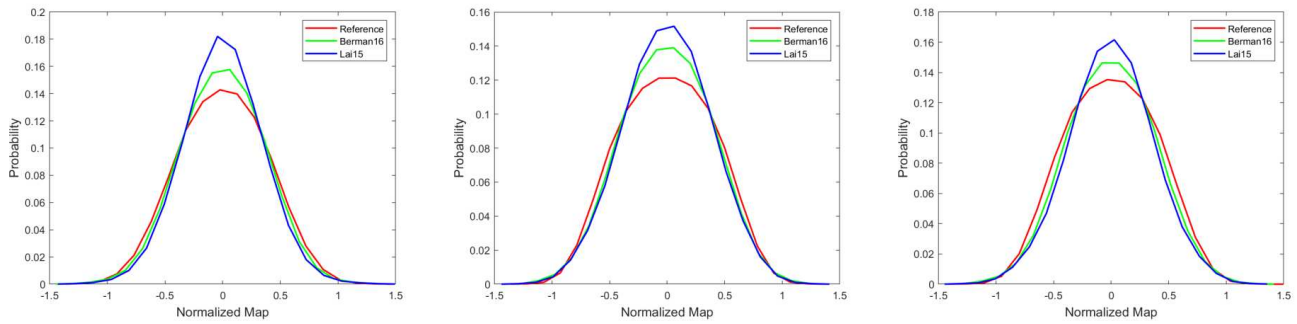


Fig. 7: Statistical distribution of different dehazed images from Fig. 6 after normalization. (a) Statistical distribution for Luminance Y component; (b) Statistical distribution for chrominance Cb component; (c) Statistical distribution for chrominance Cr component.

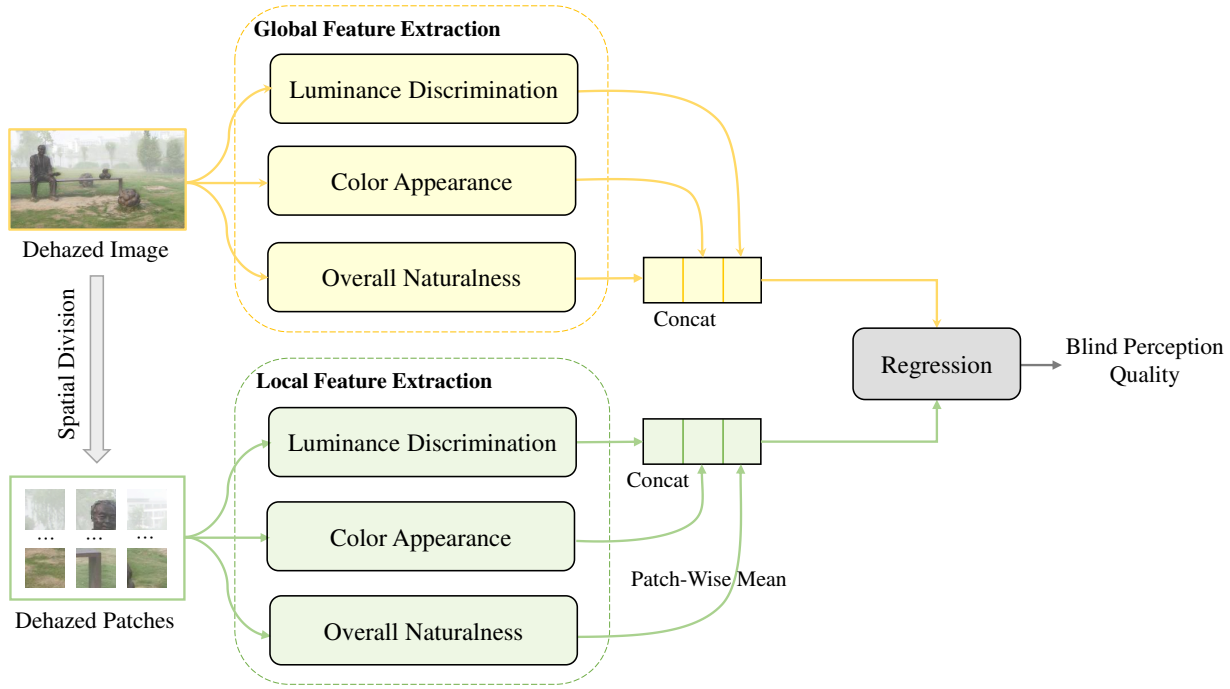


Fig. 8: Framework of the proposed NR quality evaluation method, where global and local channels are used for feature extraction.

sides. The AGGD becomes GGD when $\rho_l = \rho_r$. Moreover, the skewness and kurtosis of GGD are used as complementary features. The shape and scale parameters are estimated by the moment-matching based approach [61]. In addition, the parameters $(\delta, \lambda, \rho_l^2, \rho_r^2)$ of the best AGGD fit are achieved in which δ calculated by:

$$\delta = (\rho_r - \rho_l)\frac{\Gamma\left(\frac{2}{\lambda}\right)}{\Gamma\left(\frac{1}{\lambda}\right)}. \tag{18}$$

With YCbCr components, we adopt two scales, containing the original image scale and a reduced resolution scale proposed in [36] to perform as the overall naturalness features.

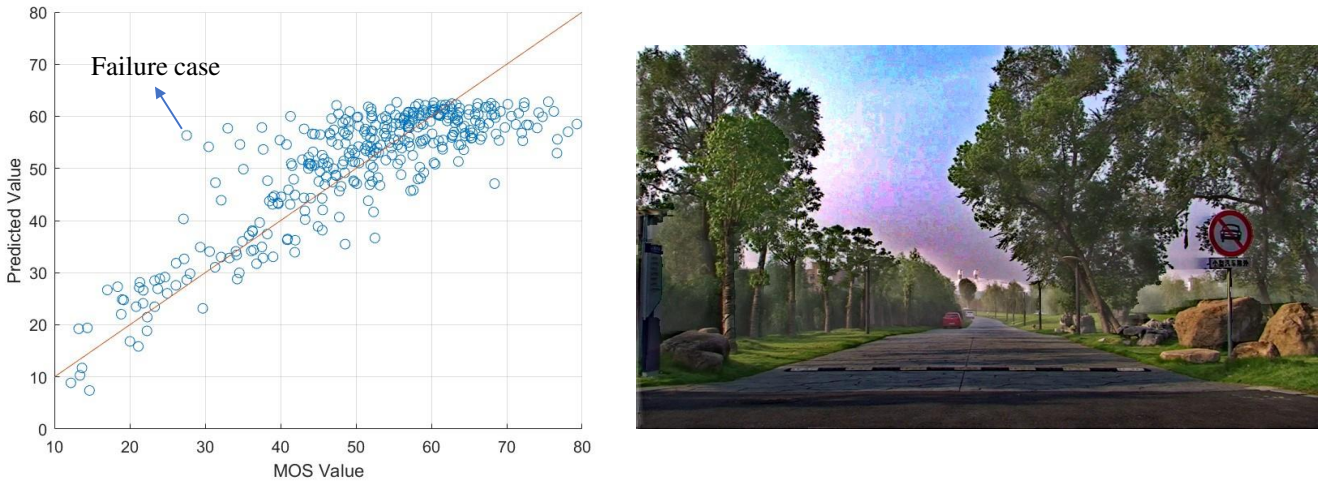
D. Partial Discrepancy Quality Estimation

In analogy to the feature extraction of reference images, identical operations can be applied to the corresponding dehazed images. Suppose that we have the luminance discrimination, color appearance, and overall naturalness features from reference and dehazed images. These features are denoted by LD/LD$'$, CA/CA$'$, and ON/ON$'$, for haze-free and dehazed images, respectively.

We first concatenate LD and CA as well as LD$'$ and CA$'$, leading to LDCA and LD$'$CA$'$. Then, the joint YCbCr comparison and naturalness comparison are computed, which can be combined as the final partial discrepancy quality. Thus, we obtain the proposed RR quality estimation as follows:

TABLE I: PERFORMANCE COMPARISONS OF VARIOUS REDUCED-REFERENCE QUALITY EVALUATION METHODS ON TWO DEHAZED SUBJECTIVE DATABASES.

Databases	SHRQR				SHRQA			
Methods	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
RRIQA [41]	0.3540	0.2487	0.4378	12.4543	0.6624	0.4741	0.6694	12.0025
RQMSH [42]	0.2793	0.1883	0.2846	13.2794	0.6403	0.4565	0.6566	12.1855
SIRR [43]	0.4662	0.3222	0.5053	11.9535	0.5997	0.4153	0.6090	12.8148
PDL [18]	0.0825	0.0552	0.0854	13.802	0.2061	0.1424	0.2001	15.8303
RRPD (without CSF)	0.6562	0.4857	0.7126	9.7185	0.7241	0.5358	0.6998	11.5414
Proposed RRPD	**0.7450**	**0.5633**	**0.8355**	**7.6115**	**0.7608**	**0.5715**	**0.7271**	**11.0921**



(a) (b)

Fig. 9: Example of failure case. (a) Scatter plot of predicted value versus MOS value; (b) A failure case from (a).

$$Q_{\mathrm{RRPD}} = \mathcal{M}(\left|\mathrm{LDCA} - \mathrm{LDCA}'\right|) \times \mathcal{M}(\left|\mathrm{ON} - \mathrm{ON}'\right|), \tag{19}$$

where $\mathrm{LDCA}/\mathrm{LDCA}'$ and ON/ON' are 22-dimensional and 120-dimensional vectors, respectively. $\mathcal{M}$ denotes the mean operation.

IV. PROPOSED NR QUALITY EVALUATION METHOD

In practical applications, the part of reference information may not be always available. Therefore, we extend the proposed RR quality evaluation method to a blind/no-reference dehazed image quality evaluation model called NRBP. The framework of the proposed NR quality evaluation method is shown in Fig. 8, where the only input is the dehazed image, and both global and local channels are used for feature extraction.

A. Local and Global Channels

Inspired by the characteristics of the HVS, to exploit more information from dehazed images, global and local channels are involved in the proposed NRBP framework. Specifically, we conduct a spatial division for the input dehazed image. Here the patch size is set to 32×32. More experimental results can be found in Section V-E.

Similar to the RRPD, we have luminance discrimination, color appearance, and overall naturalness features for both channels. Then, we concatenate the extracted features separately. It should be noted that for local feature extraction, before the concatenation, the patch-wise mean is computed for all patch features of each dehazed image. As for aerial images, the local channel would lead to marginal performance improvement, which is validated in the experimental part (i.e., Section V-C). Therefore, we test the regular images to select the patch size.

B. Blind Perception Quality Estimation

Finally, with the concatenated features from global and local channels, we utilize the support vector regression (SVR) [62] for the blind perception quality generation as:

$$Q_{\mathrm{NRBP}} = \mathcal{F}(\mathrm{LD}', \mathrm{CA}', \mathrm{ON}', {\mathrm{LD}_{\mathrm{local}}}', {\mathrm{CA}_{\mathrm{local}}}', {\mathrm{ON}_{\mathrm{local}}}'), \tag{20}$$

where $\mathcal{F}$ is the SVR operation, and the common parameters are adopted for mainstream image quality evaluation works [63]–[65].

V. EXPERIMENTS AND ANALYSIS

In this section, we will present the experimental results and analysis of our proposed RR DQA and NR DQA models.

TABLE II: PERFORMANCE COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS ON SHRQR DATABASE.

Types	Methods	SRCC	KRCC	PLCC	RMSE
FR IQA	PSNR	0.5972	0.4231	0.6523	10.4996
	SSIM [26]	0.5627	0.3991	0.6225	10.8412
	MS-SSIM [27]	0.5836	0.4160	0.6275	10.7855
	IW-SSIM [28]	0.5657	0.4031	0.6246	10.8175
	FSIM [29]	0.6256	0.4615	0.7418	9.2889
	IFC [34]	0.5549	0.4068	0.7354	9.3873
	VIF [35]	0.6287	0.4705	0.7609	8.9885
	GSM [30]	0.6029	0.4364	0.6946	9.9654
	GMSD [31]	0.6157	0.4531	0.7364	9.3722
	PSIM [32]	0.6238	0.4580	0.7580	9.0350
	SPSIM [33]	0.6553	0.4914	0.7454	9.2348
NR IQA	BIQI [37]	0.0277	0.0235	0.2444	13.4323
	BRISQUE [36]	0.4196	0.2964	0.5767	11.3165
	NIQE [38]	0.4029	0.2843	0.5920	11.1639
	BLIINDS-II [22]	0.3400	0.2353	0.5374	11.6821
	DIIVINE [23]	0.3712	0.2535	0.5016	11.9840
	LPSI [39]	0.3363	0.2355	0.5718	11.3643
	MEON [40]	0.2220	0.1445	0.3111	13.1652
FR DQA	DEHAZEfr [11]	0.8292	0.6430	0.8675	6.8912
	FRFSIM [13]	0.5862	0.4264	0.7467	9.2141
DR DQA	DHQI [14]	0.4240	0.3000	0.6739	10.2338
NR DQA	VDA-DQA [44]	0.4846	0.3421	0.6160	10.9119
	Proposed NRBP	**0.8556**	**0.6823**	**0.8992**	**5.9771**

(a) (b) (c) (d)

Fig. 10: Examples of aerial images. (a) Desert; (b) Forest; (c) Mountain; (d) Viaduct.

First, we provide a brief overview of the adopted databases and criteria in our experiments. Then, the comparisons between the proposed metrics and state-of-the-art methods are conducted. Moreover, for NRBP, we also report the ablation study of individual channels as well as each proposed feature and parameter test with different local patch sizes. Finally, we demonstrate that our proposed metrics perform well for authentic dehazed images as well as real transportation scenes and can also serve as a useful tool to be applied for image dehazing.

A. Databases and Criteria

To verify the proposed RR and NR dehazed quality evaluation methods, we carry out experiments and analysis on two synthetic databases (i.e., SHRQR and SHRQA [11]) and three authentic dehazed image quality databases including IVCDehazing [10], DHQ [14] and exBeDDE [12]. Moreover, we also test the performance on real transportation scenes.

The SHRQR database contains 45 original haze-free images and 360 dehazed images generated by eight typical image dehazing algorithms. Each haze-free image is related to a corresponding hazy image.

The SHRQA database is composed of 30 pristine haze-free images and the corresponding 30 hazy images. Moreover, there exist 240 dehazed images produced by eight image dehazing algorithms which are the same as that of the SHRQR database. Note that the image content in this database is aerial.

The IVCDehazing database has 25 real hazy images and 200 dehazed images which are produced by 8 existing dehazing algorithms. Each dehazed image is associted with a MOS value.

The DHQ database is so far the largest one, which includes 1,750 dehazed images. By 7 representative dehazing algorithms, these dehazed images are generated by real hazy images with different haze densities.

The exBeDDE database is a dehazed image quality database based on real hazy images. There exist 1,670 dehazed images created by 10 typical image dehazing approaches.

For comparison criteria, we adopt four commonly-used metrics as follows:

TABLE III: PERFORMANCE COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS ON SHRQA DATABASE.

Types	Methods	SRCC	KRCC	PLCC	RMSE
FR IQA	PSNR	0.8246	0.6397	0.8040	9.6080
	SSIM [26]	0.8207	0.6267	0.8166	9.3252
	MS-SSIM [27]	0.7895	0.5893	0.7815	10.0811
	IW-SSIM [28]	0.7949	0.5954	0.7841	10.0276
	FSIM [29]	0.7424	0.5471	0.7348	10.9583
	IFC [34]	0.5630	0.3871	0.6140	12.7522
	VIF [35]	0.7048	0.5384	0.7651	10.4044
	GSM [30]	0.7832	0.5892	0.7719	10.2715
	GMSD [31]	0.7103	0.5243	0.6984	11.5633
	PSIM [32]	0.7593	0.5755	0.7338	10.9760
	SPSIM [33]	0.8408	0.6418	0.8348	8.8960
NR IQA	BIQI [37]	0.3710	0.2442	0.3869	14.8984
	BRISQUE [36]	0.1527	0.1023	0.3216	15.2986
	NIQE [38]	0.3634	0.2470	0.4342	14.5547
	BLIINDS-II [22]	0.0895	0.0611	0.3305	15.2491
	DIIVINE [23]	0.2903	0.1915	0.3130	15.3449
	LPSI [39]	0.4170	0.2902	0.4780	14.1918
	MEON [40]	0.0339	0.0246	0.1268	16.0283
FR DQA	DEHAZEfr [11]	0.8615	0.6685	0.8554	8.3692
	DEHAZEfr+ [11]	0.9028	0.7219	0.9017	6.9855
	FRFSIM [13]	0.8065	0.6245	0.7844	10.0216
DR DQA	DHQI [14]	0.5675	0.4341	0.5726	13.2458
NR DQA	VDA-DQA [44]	0.6662	0.4816	0.6728	11.9529
	Proposed NRBP	**0.9158**	**0.7606**	**0.9170**	**6.3036**

TABLE IV: ABLATION TEST OF INDIVIDUAL CHANNELS ON TWO DEHAZED SUBJECTIVE DATABASES.

Databases	SHRQR				SHRQA			
Methods	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
NRBP (only local)	0.8024	0.6252	0.8719	6.7006	0.9078	0.7447	0.9098	6.6050
NRBP (only global)	0.8193	0.6401	0.8765	6.6110	0.9113	0.7500	0.9140	6.4002
NRBP (local + global)	**0.8556**	**0.6823**	**0.8992**	**5.9771**	**0.9158**	**0.7606**	**0.9170**	**6.3036**

TABLE V: PERFORMANCE RESULTS OF EACH PROPOSED FEATURE (LOCAL + GLOBAL) ON SHRQR DATABASE.

Methods	SRCC	KRCC	PLCC	RMSE
Luminance Discrimination	0.6717	0.5055	0.8026	8.1954
Color Appearance	0.5002	0.3592	0.7141	9.5582
Overall Naturalness	0.8008	0.6236	0.8640	6.8815
Proposed NRBP	**0.8556**	**0.6823**	**0.8992**	**5.9771**

TABLE VI: PERFORMANCE RESULTS OF EACH PROPOSED FEATURE (LOCAL + GLOBAL) ON SHRQA DATABASE.

Methods	SRCC	KRCC	PLCC	RMSE
Luminance Discrimination	0.8214	0.6454	0.8220	9.1890
Color Appearance	0.6151	0.4504	0.6546	12.0325
Overall Naturalness	0.7327	0.5461	0.7495	10.5121
Proposed NRBP	**0.9158**	**0.7606**	**0.9170**	**6.3036**

1. Spearman Rank-order Correlation Coefficient (SRCC) is defined by:

$$SRCC = 1 - \frac{6\sum_{t=1}^{T} {d_t}^2}{T(T^2-1)}, \tag{21}$$

where T denotes the total image numbers of the used database. d_t is the rank difference between the t-th image's subjective and objective evaluations.

2. Kendall Rank Correlation Coefficient (KRCC) can be computed as:

$$KRCC = \frac{2(F_c - F_d)}{T(T-1)}, \tag{22}$$

where F_c and F_d are the numbers of concordant and discordant pairs on the database, respectively.

3. Pearson Linear Correlation Coefficient (PLCC) is calculated as follows:

Fig. 11: Image dehazing by using different parameters. (a) A reference haze-free image; (b) The corresponding hazy image; (c) Dehazed image with default parameters, i.e., $\alpha = 0.5$, $\lambda = 1.0$ and $Q_{RRPD} = 38.8803$; (d) Dehazed image with $\alpha = 0.05$, $\lambda = 1.0$ and $Q_{RRPD} = 54.1494$; (e) Dehazed image with $\alpha = 5$, $\lambda = 1.0$ and $Q_{RRPD} = 34.8126$; (f) Dehazed image with $\alpha = 50$, $\lambda = 1.0$ and $Q_{RRPD} = 34.1356$; (g) Dehazed image with $\alpha = 0.5$, $\lambda = 0.1$ and $Q_{RRPD} = 26.9847$; (h) Dehazed image with $\alpha = 0.5$, $\lambda = 0.01$ and $Q_{RRPD} = 32.0869$; (i) Dehazed image with $\alpha = 0.5$, $\lambda = 10$ and $Q_{RRPD} = 81.1083$.

TABLE VII: PERFORMANCE VARIATION OF DIFFERENT PATCH SIZES FOR PROPOSED NRBP (LOCAL + GLOBAL) ON SHRQR DATABASE.

Patch Sizes	SRCC	KRCC	PLCC	RMSE
16×16	0.8482	0.6729	0.8971	6.0858
32×32	**0.8556**	**0.6823**	**0.8992**	**5.9771**
64×64	0.8527	0.6808	0.8981	6.0351
128×128	0.8506	0.6776	0.8930	6.1905

$$PLCC = \frac{\sum_{t=1}^{T}(s_t - \overline{s})(o_t - \overline{o})}{\sqrt{\sum_{t=1}^{T}(s_t - \overline{s})\sum_{t=1}^{T}(o_t - \overline{o})}}, \tag{23}$$

where s_t and o_t represent the t-th subjective and mapped objective quality scores. $\overline{s}$ and $\overline{o}$ are the corresponding mean values of s_i and o_i, respectively.

4. Root Mean Squared Error (RMSE) can be estimated by:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(s_t - o_t)^2}{T}}. \tag{24}$$

The SRCC and KRCC are applied to test prediction monotonicity and the ordinal association between two measured quantities. Meanwhile, the PLCC and RMSE are used to measure prediction accuracy. Note that higher correlation coefficients and lower errors mean better performance.

Additionally, before computing the PLCC and RMSE values of different objective quality evaluation approaches, a five-parameter logistic nonlinear fitting function [66] is employed to map the predicted quality scores into a common scale as:

$$\epsilon(\mathrm{q}) = \beta_1\left(\frac{1}{2} - \frac{1}{1 + \mathrm{e}^{(\beta_2(\mathrm{q}-\beta_3))}}\right) + \beta_4\mathrm{q} + \beta_5, \tag{25}$$

where $(\beta_1 ... \beta_5)$ represent five parameters to be fitted. q and $\epsilon(\mathrm{q})$ are the raw objective score generated by quality

TABLE VIII: PERFORMANCE COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS ON IVCDEHAZING DATABASE.

Types	Methods	SRCC	KRCC	PLCC	RMSE
NR IQA	BIQI [37]	0.1472	0.0956	0.2669	1.4254
	BRISQUE [36]	0.1568	0.1057	0.2036	1.4481
	NIQE [38]	0.1599	0.1057	0.3166	1.4030
	BLIINDS-II [22]	0.0193	0.0106	0.1984	1.4497
	DIIVINE [23]	0.1126	0.0727	0.2057	1.4474
	LPSI [39]	0.0381	0.0271	0.1021	1.4713
	MEON [40]	0.0501	0.0299	0.1435	1.4638
NR DQA	VDA-DQA [44]	0.1375	0.0931	0.3221	1.4002
	Proposed NRBP	**0.6587**	**0.4834**	**0.7161**	**1.0156**

TABLE IX: PERFORMANCE COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS ON DHQ DATABASE.

Types	Methods	SRCC	KRCC	PLCC	RMSE
NR IQA	BIQI [37]	0.2566	0.1725	0.2874	12.5828
	BRISQUE [36]	0.3034	0.2024	0.4757	11.5557
	NIQE [38]	0.4306	0.3020	0.5090	11.3082
	BLIINDS-II [22]	0.3145	0.2146	0.4839	11.4966
	DIIVINE [23]	0.2743	0.1858	0.3843	12.1283
	LPSI [39]	0.3500	0.2387	0.5233	11.1947
	MEON [40]	0.2790	0.1907	0.3111	12.4853
NR DQA	VDA-DQA [44]	0.6220	0.4514	0.6598	9.8721
	Proposed NRBP	**0.8483**	**0.6652**	**0.8608**	**6.6786**

TABLE X: PERFORMANCE COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS ON EXBEDDE DATABASE.

Types	Methods	SRCC	KRCC	PLCC	RMSE
NR IQA	BIQI [37]	0.3497	0.2395	0.3950	0.2454
	BRISQUE [36]	0.3688	0.2522	0.4966	0.2319
	NIQE [38]	0.1422	0.0879	0.2971	0.2551
	BLIINDS-II [22]	0.3486	0.2405	0.3443	0.2508
	DIIVINE [23]	0.4695	0.3321	0.5119	0.2295
	LPSI [39]	0.1212	0.0918	0.2417	0.2592
	MEON [40]	0.0227	0.0150	0.3604	0.2492
NR DQA	VDA-DQA [44]	0.4216	0.2861	0.4811	0.2342
	Proposed NRBP	**0.9115**	**0.7403**	**0.9354**	**0.0944**

assessment models and the regressed objective score after the nonlinear mapping, respectively.

B. Comparisons with Other Quality Evaluation Metrics

We compare the proposed dehazed image quality evaluation methods with existing state-of-the-art ones. For the proposed RRPD, the other compared RR quality evaluation algorithms include PDL [18], RRIQA [41], RQMSH [42], and SIRR [43]. Besides, due to the lack of RR DQA metrics for dehazed images, we also test the RRPD without CSF. The performance results are reported in TABLE I. From this table, we find that our specifically designed RR DQA model outperforms the others. Note that PDL is the only RR DQA method developed for image restoration, which involves dehazed images. Our proposed metric performs significantly better than it, further demonstrating the superiority of the proposed metric. This is because we aim to focus on dehazed images and relevant characteristics rather than the generic image restoration case. Furthermore, removing CSF would reduce the performance of our model, which validates the effectiveness of the proposed CSF weighting. This is mainly attributed to the powerful perceptual properties of CSF, performing closer to the human visual perception in the dehazed image quality evaluation task.

In Fig. 9, we show an example of the failure case of our metric on the SHRQR database. It can be seen that the corresponding dehazed image in (b) suffers from complex artifacts, especially in the sky. To overcome such failure case, more advanced frameworks can be proposed to consider more visual cues and perceptual features.

As for the proposed NRBP, the traditional image quality evaluation methods for comparisons are 11 FR IQA and 8 NR IQA models. Except for these models, existing FR DQA, DR DQA and NR DQA approaches are also taken into consideration. The compared results on SHRQR and SHRQA databases are provided in TABLE II and TABLE III, respectively. As can be seen from these two tables, our proposed NRBP delivers the best performance on both databases. More importantly, the proposed metric even shows superior performance compared with DEHAZEfr [11] and FRFSIM [13] which are two FR DQA models. It should be noted that the DEHAZEfr+ is the improved DEHAZEfr measure for aerial images. Additionally, we notice that the performance numbers of the SHRQA database are generally better than that of the SHRQR database.

(a) (b)

Fig. 12: Illustration of real hazy images in transportation scenes. (a) Out-vehicle view; (b) In-vehicle view.

TABLE XI: COMPARISONS OF DIFFERENT QUALITY EVALUATION METHODS FOR TRANSPORTATION SCENES.

Types	Methods	SRCC	KRCC	PLCC	RMSE
NR IQA	BIQI [37]	0.2265	0.1000	0.6587	1.1909
	BRISQUE [36]	0.1735	0.1167	0.5057	1.3654
	NIQE [38]	0.0382	0.0333	0.7258	1.0888
	BLIINDS-II [22]	0.0923	0.0649	0.2899	1.5147
	DIIVINE [23]	0.1059	0.0167	0.2251	1.5421
	LPSI [39]	0.3147	0.2000	0.4857	1.3835
	MEON [40]	0.1294	0.1000	0.3722	1.4690
NR DQA	VDA-DQA [44]	0.0765	0.1000	0.6888	1.1473
	Proposed NRBP	**0.5794**	**0.4500**	**0.7285**	**1.0846**

This proves that dehazed quality evaluation for regular images is more challenging.

C. Ablation Study of Individual Channels

Since the proposed NRBP architecture has global and local channels, the ablation study is adopted to test the results of individual channels. Specifically, we test three configurations, including the NRBP with only local channel, the NRBP with only global channel, and the NRBP with both local and global channels.

TABLE IV provides the ablation results which are also verified by the significant t-test. We observe that the global channel performs better than the local channel and the combination of the two can boost the final results to some extent. However, the performance improvements for the SHRQR database are obviously larger than that for the SHRQA database. In Fig. 10, we show some examples of aerial images. An interesting trend is that aerial images are inclined to have similar textures. That is, the self-similarity of local patches for these images is relatively higher, compared to that of the regular images. This may be the main reason for that the proposed NRBP with only global channel obtains top-performing results on SHRQA database.

D. Validity of Each Proposed Feature

It is interesting to verify each proposed feature category regarding to luminance discrimination, color appearance, and overall naturalness.

In TABLE V and TABLE VI, we provide the performance results of individual proposed features on SHRQR and SHRQA databases, respectively. It can be seen that the proposed combination (i.e., NRBP) has the best performance, which demonstrates the superiority of our proposed metric. As for SHRQR and SHRQA databases, the overall naturalness and luminance discrimination outperform other individual proposed features separately. This may be due to the differential characteristics of regular and aerial images.

E. Test with Different Local Patch Sizes

Due to the relatively marginal improvement of adding local channel for aerial images, we test different local patch sizes on the SHRQR database.

As illustrated in TABLE VII, several patch sizes are adopted in the experiments, including 16×16, 32×32, 64×64, and 128×128. From the results, we find that the 32×32 leads to the best performance. In addition, with the increase and decrease around 32×32, the performance would drop. This is also consistent with other visual quality evaluation methods

[67]–[69]. Therefore, we choose the patch size equaling to 32×32 in our designed NRBP model.

F. Application to Image Dehazing

A good dehazed quality indicator can not only predict the perceptually visual quality of dehazed images, but also effectively optimize the existing image dehazing algorithms. Here we also validate our proposed dehazed quality evaluation method by applying it to parameter selection of a classical image dehazing approach [7].

We give image dehazing results in Fig. 11. Since the proposed RRPD does not need the training process, here we show the objective quality scores obtained by RRPD. In this figure, higher Q_{RRPD} indicates worse visual quality. The image dehazing algorithm has two parameters, including α and λ. Among the generated results, (c) utilizes the original parameters proposed in [7]. We observe that the proposed NRBP can efficiently select better parameters. That is, (g) with $\alpha = 0.5$, $\lambda = 0.1$ delivers the best perceptual quality.

G. Evaluation on Authentic Dehazed Images

In order to further verify the proposed NR quality evaluation method, we adopt three authentic dehazed image quality databases, i.e., IVCDehazing, DHQ, and exBeDDE.

We show the test results in Table VIII, Table IX, and Table X. Since the pristine images are unavailable, here we compare the proposed NRBP with state-of-the-art NR IQA and NR DQA models. We find that existing traditional NR IQA metrics generally fail to predict the dehazed image quality. In contrast, our proposed NRBP significantly outperforms the NR IQA models. Additionally, the proposed NRBP can also exceed the VDA-DQA [44] which is the specific NR DQA method designed for dehazed images.

H. Evaluation on Real Transportation Scenes

Apart from the synthetic and authentic dehazed image quality databases, we test our methods on real automotive imagery to further verify the performance. Specifically, we select real hazy images in transportation scenes on so far the most challenging IVCDehazing database as shown in Fig. 12. Both the out-vehicle and in-vehicle views are considered in this figure.

We exploit the largest DHD database to train our model and then test it on the transportation scenes. The performance comparisons of different quality evaluation methods are reported in TABLE XI. Again, our proposed NRBP shows the best performance when compared to state-of-the-art algorithms. This further demonstrates the superiority and potential of the proposed method for intelligent vehicle systems.

VI. Conclusion

In this paper, we have presented a new RR dehazed quality evaluation method to predict the visual quality of dehazed images with only part of the original reference information. In the proposed framework, to involve the hierarchical property of the human perception, we propose quality-aware features from the aspects of luminance discrimination, color appearance, and overall naturalness. Then, motivated by the characteristics of the HVS, we extend it to a blind/NR quality evaluation model by integrating both global and local channels. Finally, experimental results on both synthetic and authentic dehazed image quality databases as well as real transportation scenes demonstrate the superiority of our proposed quality evaluation methods. Additionally, our proposed quality metric can also be applied to optimize the existing image dehazing algorithm by tuning its parameters.

In the future, we plan to explore more potential applications of the proposed dehazed quality evaluation methods. For example, using the metric as a loss function to train a more robust neural network for image dehazing. Besides, developing quality evaluation methods for video dehazing and its optimization is also a promising research direction.

References

[1] K. Wang, L. Pu, J. Zhang, and J. Lu, "Gated adversarial network based environmental enhancement method for driving safety under adverse weather conditions," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1934–1943, 2022.

[2] Y. Liu, M. Wang, P. Lasang, and Q. Sun, "Importance biased traffic scene segmentation in diverse weather conditions," *IEEE Transactions on Intelligent Vehicles*, 2023.

[3] K. C. Dey, A. Mishra, and M. Chowdhury, "Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1107–1119, 2014.

[4] M. Negru, S. Nedevschi, and R. I. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2257–2268, 2015.

[5] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6–20, 2012.

[6] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[7] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 617–624.

[8] Y.-H. Lai, Y.-L. Chen, C.-J. Chiou, and C.-T. Hsu, "Single-image dehazing via optimal transmission map under scene priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 1–14, 2015.

[9] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1674–1682.

[10] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *IEEE International Conference on Image Processing*. IEEE, 2015, pp. 3600–3604.

[11] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, and W. Zhang, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2319–2333, 2019.

[12] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Transactions on Image Processing*, vol. 29, pp. 6947–6962, 2020.

[13] W. Liu, F. Zhou, T. Lu, J. Duan, and G. Qiu, "Image defogging quality assessment: Real-world database and method," *IEEE Transactions on Image Processing*, vol. 30, pp. 176–190, 2020.

[14] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2879–2892, 2018.

[15] J. Shen, Q. Li, and G. Erlebacher, "Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2089–2098, 2011.

[16] S. A. Golestaneh and D. M. Chandler, "No-reference quality assessment of JPEG images via a quality relevance map," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 155–158, 2013.

[17] L. Li, Y. Zhou, K. Gu, Y. Yang, and Y. Fang, "Blind realistic blur assessment based on discrepancy learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3859–3869, 2019.

[18] L. Li, B. Hu, Y. Huang, and H. Zhu, "Reduced-reference perceptual discrepancy learning for image restoration quality assessment," in *Artificial Intelligence: First CAAI International Conference, CICAI 2021, Hangzhou, China, June 5–6, 2021, Proceedings, Part I 1*. Springer, 2021, pp. 359–370.

[19] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment and human visual system," in *Visual Communications and Image Processing 2010*, vol. 7744. SPIE, 2010, pp. 316–325.

[20] D. M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *International Scholarly Research Notices*, vol. 2013, 2013.

[21] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[22] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[23] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[24] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. T. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in Neuroinformatics*, vol. 3, p. 571, 2009.

[25] W. Zhou, J. Xu, Q. Jiang, and Z. Chen, "No-reference quality assessment for 360-degree images by analysis of multifrequency information and local-global naturalness," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1778–1791, 2021.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[27] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2. IEEE, 2003, pp. 1398–1402.

[28] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2010.

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

[30] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, 2011.

[31] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2013.

[32] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 3903–3912, 2017.

[33] W. Sun, Q. Liao, J.-H. Xue, and F. Zhou, "SPSIM: A superpixel-based similarity index for full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4232–4244, 2018.

[34] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[35] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

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

[37] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[38] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[39] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *IEEE International Conference on Image Processing*, 2015, pp. 339–343.

[40] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.

[41] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human Vision and Electronic Imaging X*, vol. 5666. SPIE, 2005, pp. 149–159.

[42] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 1–14, 2016.

[43] X. Min, K. Gu, G. Zhai, M. Hu, and X. Yang, "Saliency-induced reduced-reference quality index for natural scene and screen content images," *Signal Processing*, vol. 145, pp. 127–136, 2018.

[44] T. Guan, C. Li, K. Gu, H. Liu, Y. Zheng, and X.-J. Wu, "Visibility and distortion measurement for no-reference dehazed image quality assessment via complex contourlet transform," *IEEE Transactions on Multimedia*, 2022.

[45] B. Ji, Y. Ji, H. Gao, X. Hu, and F. Ding, "No-reference image quality assessment for dehazed images," *Journal of Electronic Imaging*, vol. 31, no. 1, pp. 013 013–013 013, 2022.

[46] B. Hu, L. Li, H. Liu, W. Lin, and J. Qian, "Pairwise-comparison-based rank learning for benchmarking image restoration algorithms," *IEEE Transactions on Multimedia*, vol. 21, no. 8, pp. 2042–2056, 2019.

[47] X. Min, G. Zhai, K. Gu, Y. Fang, X. Yang, X. Wu, J. Zhou, and X. Liu, "Blind quality assessment of compressed images via pseudo structural similarity," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2016, pp. 1–6.

[48] K. Gu, J. Qiao, X. Min, G. Yue, W. Lin, and D. Thalmann, "Evaluating quality of screen content images via structural variation analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 10, pp. 2689–2701, 2017.

[49] W. Zhou, Z. Wang, and Z. Chen, "Image super-resolution quality assessment: Structural fidelity versus statistical naturalness," in *International Conference on Quality of Multimedia Experience*. IEEE, 2021, pp. 61–64.

[50] J. G. Robson, "Spatial and temporal contrast-sensitivity functions of the visual system," *Josa*, vol. 56, no. 8, pp. 1141–1142, 1966.

[51] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5293–5303, 2016.

[52] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.

[53] S. Daly, "Subroutine for the generation of a two dimensional human visual contrast sensitivity function," Eastman Kodak, Rochester, NY, USA, Tech. Rep. 233203Y, 1987.

[54] T. Mitsa and K. L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, 1993, pp. 301–304.

[55] D. Lee and K. N. Plataniotis, "Towards a full-reference quality assessment for color images using directional statistics," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3950–3965, 2015.

[56] D. Temel and G. AlRegib, "CSV: Image quality assessment based on color, structure, and visual system," *Signal Processing: Image Communication*, vol. 48, pp. 92–103, 2016.

[57] D. Lee and K. N. Plataniotis, "Toward a no-reference image quality assessment using statistics of perceptual color descriptors," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3875–3889, 2016.

[58] Z. Wang, L. Shen, Z. Wang, Y. Lin, and Y. Jin, "Generation-based joint luminance-chrominance learning for underwater image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[59] S. Lyu, "Dependency reduction with divisive normalization: Justification and effectiveness," *Neural Computation*, vol. 23, no. 11, pp. 2942–2973, 2011.

[60] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *IEEE International Conference on Image Processing*. IEEE, 2009, pp. 2281–2284.

[61] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, 1995.

[62] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
[63] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010.
[64] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 721–734, 2017.
[65] W. Zhou, L. Shi, Z. Chen, and J. Zhang, "Tensor oriented no-reference light field image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4070–4084, 2020.
[66] A. M. Rohaly, P. J. Corriveau, J. M. Libert, A. A. Webster, V. Baroncini, J. Beerends, J.-L. Blin, L. Contin, T. Hamada, D. Harrison *et al.*, "Video quality experts group: Current results and future directions," in *Visual Communications and Image Processing*, vol. 4067. SPIE, 2000, pp. 742–753.
[67] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.
[68] W. Zhou, Q. Jiang, Y. Wang, Z. Chen, and W. Li, "Blind quality assessment for image superresolution using deep two-stream convolutional networks," *Information Sciences*, vol. 528, pp. 205–218, 2020.
[69] Y. Zhang, H. Liu, Y. Yang, X. Fan, S. Kwong, and C. J. Kuo, "Deep learning based just noticeable difference and perceptual quality prediction models for compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1197–1212, 2021.

BIOGRAPHY

Wei Zhou is an Assistant Professor at Cardiff University, United Kingdom. Dr Zhou was a Postdoctoral Fellow at University of Waterloo, Canada. Wei received the Ph.D. degree from the University of Science and Technology of China in 2021, joint with the University of Waterloo from 2019 to 2021.

Dr Zhou was a visiting scholar at National Institute of Informatics, Japan, a research assistant with Intel, and a research intern at Microsoft Research and Alibaba Cloud. Wei is now an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems. Wei's research interests span multimedia computing, perceptual image processing, and computational vision.

Ruizeng Zhang received the B.E. degree from Shandong University, Jinan, China, in 2016. He is currently working toward the Ph.D. degree with the School of Mechanical Engineering, Beijing Institute of Technology. He was a Visiting Scholar with the Mechatronic Vehicle Systems Lab, University of Waterloo from 2021 to 2023. His main research interest are motion control and planning of autonomous vehicle system.

Leida Li received the B.E. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. From 2009 to 2019, he worked at China University of Mining and Technology, as an Assistant Professor, an Associate Professor, and a Professor. Currently, he is a Professor with the School of Artificial Intelligence, Xidian University. His research interests include multimedia quality assessment, computational aesthetics and affective computing. He served as the Area Chair for IJCAI 2019–2020 and ICME 2023, the Session Chair for ICMR 2019 and PCM 2015, and a TPC for AAAI 2019, ACM MM 2019–2020, ACM MM-Asia 2019, ACII 2019, and PCM 2016. He is now an Associate Editor of the Journal of Visual Communication and Image Representation.

Guanghui Yue received the B.S. degree in communication engineering from Tianjin University in 2014, and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2019. He was a joint Ph.D. student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from September 2017 to January 2019.

He is currently an Associate Professor with the School of Biomedical Engineering, Health Science Center, Shenzhen University. His research interests include bioelectrical signal processing, multimedia quality assessment, 3D image visual discomfort prediction, pattern recognition, and medical image analysis.

Jianwei Gong received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2002. From 2011 to 2012, he was a Visiting Scientist with the Robotic Mobility Group, Massachusetts Institute of Technology, Cambridge, MA, USA.

He is currently a Professor and the Director of the Intelligent Vehicle Research Centre, School of Mechanical Engineering. His research interests include intelligent vehicle environment perception and understanding, decision making, path/motion planning, and control.

Huiyan Chen received the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2004. Since 1981, he has been with the Beijing Institute of Technology, where he is currently a Professor with the Intelligent Vehicle Research Center. His research interests include intelligent vehicle and vehicle automatic transmission technologies.

Hantao Liu is an Associate Professor with the School of Computer Science and Informatics, Cardiff University, United Kingdom. He received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is a founder member of the Delft Image Quality Lab.

His research interests include visual media quality assessment, visual attention modelling and applications, visual scene understanding, medical image perception and human-machine interaction. He is serving as Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology and IEEE Signal Processing Letters.