Quality Evaluation of Image Dehazing Methods Using Synthetic Hazy Images

Xiongkuo Min, *Member, IEEE*, Guangtao Zhai, *Member, IEEE*, Ke Gu, Yucheng Zhu, Jiantao Zhou, *Member, IEEE*, Guodong Guo, *Senior Member, IEEE*, Xiaokang Yang, *Fellow, IEEE*, Xinping Guan, *Fellow, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

***Abstract*—To enhance the visibility and usability of images captured in hazy conditions, many image dehazing algorithms (DHAs) have been proposed. With so many image DHAs, there is a need to evaluate and compare these DHAs. Due to the lack of the reference haze-free images, DHAs are generally evaluated qualitatively using real hazy images. But it is possible to perform quantitative evaluation using synthetic hazy images since the reference haze-free images are available and full-reference (FR) image quality assessment (IQA) measures can be utilized. In this paper, we follow this strategy and study DHA evaluation using synthetic hazy images systematically. We first build a synthetic haze removing quality (SHRQ) database. It consists of two subsets: regular and aerial image subsets, which include 360 and 240 dehazed images created from 45 and 30 synthetic hazy images using 8 DHAs, respectively. Since aerial imaging is an important application area of dehazing, we create an aerial image subset specifically. We then carry out subjective quality evaluation study on these two subsets. We observe that taking DHA evaluation as an exact FR IQA process is questionable, and the state-of-the-art FR IQA measures are not effective for DHA evaluation. Thus, we propose a DHA quality evaluation method by integrating some dehazing-relevant features, including image structure recovering, color rendition, and over-enhancement of low-contrast areas. The proposed method works for both types of images, but we further improve it for aerial images by incorporating its specific characteristics. Experimental results on two subsets of the SHRQ database validate the effectiveness of the proposed measures.**

***Index Terms*—Image dehazing, dehazing algorithm evaluation, quality assessment, synthetic haze, regular/aerial image.**

I. Introduction

IMAGES captured from outdoor scenes using visible light imaging devices can suffer from visibility reduction due to the atmospheric scattering caused by atmospheric particles such as haze and cloud [1], [2]. This problem is particularly serious in aerial imaging since the atmosphere condition is uncontrollable, and haze or cloud is frequently observed in such scenario. To get clear and visually pleasuring images under hazy atmosphere conditions and to facilitate further image analysis, many image dehazing algorithms (DHAs) have been proposed [3]–[11]. Specifically for aerial images captured by remote sensing satellites, there are also some algorithms proposed to detect and remove the haze or cloud [12]–[14]. With so many available image DHAs, how to evaluate the perceptual quality of the dehazing and select the best DHA becomes a problem. What's more, the DHAs have not been systematically tested in aerial images yet, while aerial imaging is an important application area of dehazing. Compared with the extensive research of DHAs, the evaluation of DHAs falls behind, and the evaluation of DHAs in aerial images is even less researched. In the literature, DHAs evaluation generally follows two strategies: using real hazy images and using synthetic hazy images. A comparison of these two strategies is illustrated in Fig. 1.

Evaluating DHA using real hazy images is straightforward, and it can be interpreted as a no-reference (NR) image quality assessment (IQA) problem. As illustrated in Fig. 1, we can evaluate DHAs by assessing the perceptual quality of the dehazed images. The most intuitive way is to evaluate the dehazed images qualitatively by human subjects. It is reliable but expensive and time-consuming, and it can not be embedded into any optimization frameworks and practical systems. A better way is to introduce some quantitative evaluators. But this is difficult since image dehazing is a complicate process, and different DHAs can have distinctive effects. Since the primary

Manuscript received August 16, 2018; revised November 3, 2018 and January 20, 2019; accepted February 1, 2019. Date of publication February 27, 2019; date of current version August 23, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61831015, 61521062, and 61527804, in part by the China Postdoctoral Science Foundation under Grant BX20180197, in part by the Macau Science and Technology Development Fund under Grants FDCT/022/2017/A1 and FDCT/077/2018/A2, and in part by the Research Committee at the University of Macau under Grants MYRG2016-00137-FST and MYRG2018-00029-FST. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hantao Liu. *(Corresponding author: Guangtao Zhai.)*

X. Min is with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: minxiongkuo@gmail.com).

G. Zhai, Y. Zhu, X. Yang, and W. Zhang are with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn; zyc420@sjtu.edu.cn; xkyang@sjtu.edu.cn; zhangwenjun@sjtu.edu.cn).

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn).

J. Zhou is with the Department of Computer and Information Science, Faculty of Science and Technology, and also with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 999078, China (e-mail: jtzhou@umac.mo).

G. Guo is with the Institute of Deep Learning and with the National Engineering Laboratory for Deep Learning Technology and Application, Baidu Research, Beijing 100193, China (e-mail: guoguodong01@baidu.com).

X. Guan is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xpguan@sjtu.edu.cn).

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

Digital Object Identifier 10.1109/TMM.2019.2902097

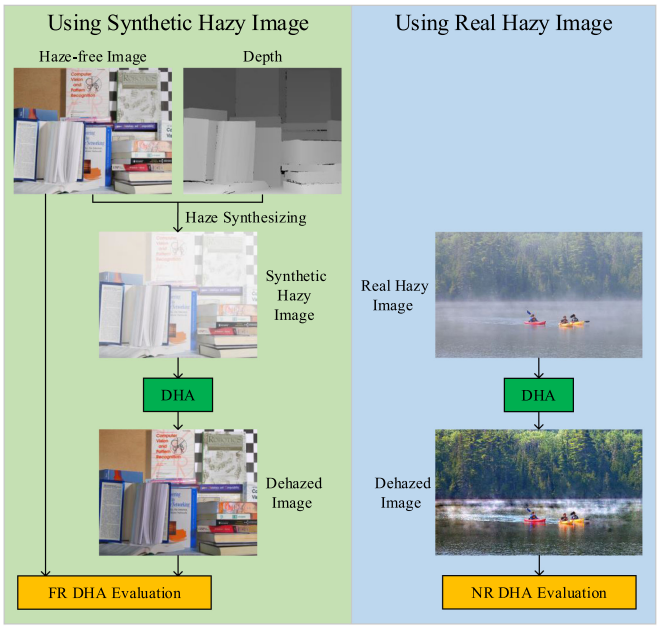


Fig. 1. Comparison of DHA evaluation using synthetic and real hazy images.

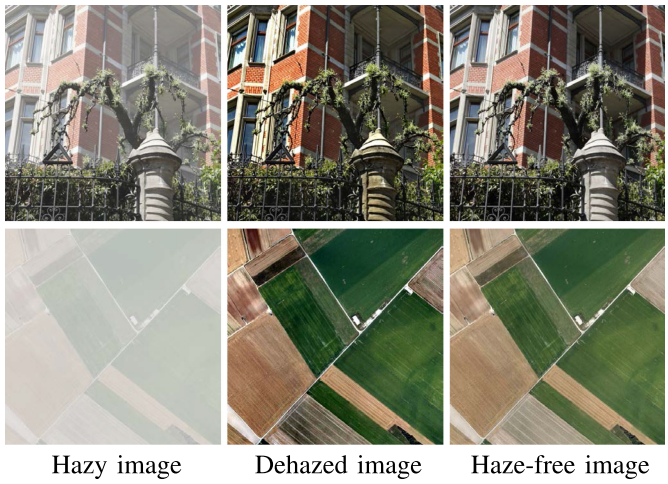


Fig. 2. The dehazed image may not be so close to the haze-free image from an image fidelity perspective, but it still has high perceptual quality.

objective of dehazing is to remove haze and enhance contrast, some evaluation methods have been proposed by accessing the contrast enhancement quality [15]. But these measures show low correlation with the overall perceptual quality [16], since DHAs not only remove haze but also introduce various other effects. For overall quality assessment, we should consider contrast enhancement, image structure recovering, color rendition, over-enhancement, etc [17]. There are also some other measures proposed to assess the quality of enhanced images [18], [19], but these measures are designed to evaluate general image enhancement algorithms, and they are not suitable and reliable for DHA evaluation.

Since performing quantitative DHA evaluation with real hazy images is difficult, some researchers suggest using hazy images synthesized from haze-free images [1], [2], [8]–[11], [20], and conduct quantitative evaluation with the available haze-free images. These methods follow the framework illustrated in the left part of Fig. 1, and generally consist of several key steps:

1) synthesizing hazy images using haze-free images and the corresponding depth;
2) dehazing using the target DHA;
3) assessing the quality of the dehazed images using the haze-free images as the "ground-truth".

Under such strategy, DHA evaluation can be solved via full-reference (FR) IQA between the haze-free and dehazed images. Many papers follow this strategy [1], [8]–[11], [20], and utilize some basic FR IQA measures such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [21] for evaluation. Owing to the availability of the "ground-truth" haze-free images, it is easy to perform quantitative evaluation and comparison. Thus this strategy is becoming more and more widely accepted. Besides DHA evaluation, some learning-based DHAs also follow this strategy since the "ground-truth" is available.

Compared with DHA evaluation in regular images, DHA evaluation in aerial images is even less researched. Similar to the situations in regular images, DHA evaluation in aerial images is conducted using either real or synthetic hazy images [12]–[14]. When using real hazy images, either direct qualitative comparison [13] or some quantitative measures designed for regular images are utilized [12]. When using synthetic hazy images, simple quality measures like PSNR are used [14]. In general, DHA evaluation in aerial images inherits the strategies and measures from the evaluation in regular images. No comprehensive DHAs evaluation is conducted in aerial images yet, and no specific DHA evaluation measures is designed for aerial images.

In this paper, we study DHA evaluation in both regular and aerial images using synthetic haze systematically. We investigate DHA evaluation from the perceptual quality point of view. In another words, human beings are the ultimate receiver of the dehazed images and the goal is to estimate the human perceived dehazing quality. We first construct a synthetic haze removing quality (SHRQ) database, which includes a regular image subset and an aerial image subset. Both subsets include dehazed images created from synthetic hazy images using 8 representative DHAs. Haze-free images and the corresponding depth images are used to synthesize the hazy images using the widely utilized haze model. We then carry out subjective quality evaluation study on the two subsets of the SHRQ database. Results show that the state-of-the-art FR IQA measures are not effective for DHA evaluation. We observe that using the haze-free images as the "ground-truth" and taking DHA evaluation as an exact FR IQA process can be problematic. The FR DHA evaluation is slightly different from tradition FR IQA. Dehazing is an image enhancement process, whose desired enhancement degree is hard to control. If the underlying reference haze-free image is not sharp enough, it is possible for the dehazed image to have a better perceptual quality than the reference image since humans often prefer clear and sharp images. As illustrated in Fig. 2, sometimes the dehazed image may not be so close to the reference image from an image fidelity perspective, but it still has high perceptual quality. Whereas in FR IQA, being closer to the reference image means better perceptual quality.

To solve the low correlation problem of current FR IQA measures in FR DHA evaluation, we propose a quality measure which is a combination of 3 components: image structure recovering, color rendition, and over-enhancement of low-contrast areas. We use the haze-free image as a reference of the original image content, and derive a similarity term to measure the image structure recovering during dehazing. To tackle the problem described in the previous paragraph, we modify the structure features before calculating the similarity. Besides the image structures, we incorporate another two factors including color rendition and over-enhancement, since some DHAs can cause undesirable side effects such as color shift and over-enhancement of the low-contrast areas. The proposed method works for both regular and aerial images, but we further improve it for aerial images by incorporating the specific characteristics of aerial images. The effectiveness of the proposed measure is verified on the SHRQ database. Considering that DHA evaluation using synthetic hazy images is becoming more widely used, the proposed measure can be of great value under such evaluation strategy.

The remainder of this paper is organized as follows. In Section II, we shortly review some related works. Section III describes the construction of the SHRQ database and the subjective user study. Section IV presents the details of the proposed measure. In Section V, we improve the proposed measure for aerial images by incorporating the specific characteristics. Effectiveness of the proposed method is verified in Section IV, where we also give comprehensive analyses of the proposed measure. Section VII concludes this paper.

II. Related Works

In this section, we shortly review the atmospheric scattering model, the state-of-the-art DHAs, and the evaluation of DHAs.

A. Atmospheric Scattering Model

In the areas of computer vision and computer graphics, hazy image formation is widely modeled by the following atmospheric scattering model [22]

$$\mathbf{I}(x) = \mathbf{J}(x)t(x) + \mathbf{A}\big(1 - t(x)\big), \tag{1}$$

where $\mathbf{I}$ is the observed hazy image, $\mathbf{J}$ is the real scene radiance, t is the medium transmission, $\mathbf{A}$ is the global atmospheric light, and x denotes the pixel index. In homogenous atmosphere, the transmission t can be modeled by

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

where β is the scattering coefficient of the atmosphere, and d indicates the distance from the scene point to the camera. The *attenuation* term $\mathbf{J}(x)t(x)$ describes how scene radiance gets attenuated when traversing from a scene point to the camera, whereas the *airlight* term $\mathbf{A}\big(1 - t(x)\big)$ describes how atmosphere reflects environmental illumination to the camera. This atmospheric scattering model is the basis of many DHAs.

B. Dehazing Algorithms (DHAs)

With the demand of dehazing in consumer photography and computational imaging, many DHAs have been proposed in recent years. The goal of dehazing is to estimate $\mathbf{J}$, t, and $\mathbf{A}$ from the observed image $\mathbf{I}$. Fattal [3] refined the image formation model to account for surface shading. Based on that, he introduced a method to remove haze and estimate transmission. Tarel and Hautière [4] took dehazing as a particular filtering problem, and proposed a method based on median filter. As an alternative of the median filter, they proposed an edge and corner preserving filter. He *et al.* [5] introduced a dark channel prior (DCP) for dehazing. The DCP describes a phenomenon that in the non-sky regions at least one color channel has very low intensity at some pixels. Xiao and Gan [6] introduced another method based on guided joint bilateral filter. Meng *et al.* [7] explored the inherent boundary constraint on the transmission, and proposed a regularization method to remove haze. Tang *et al.* [8] utilized random forest to learn a regression model to estimate the transmission from some haze-relevant features. Lai *et al.* [9] derived the optimal transmission map under two scene constraints, i.e., locally consistent scene radiance and context-aware scene transmission. Berman *et al.* [10] introduced a non-local method based on the assumption that an image could be represented via a few hundreds of distinct colors, which formed clusters in RGB space. Cai *et al.* [11] introduced an end-to-end system for dehazing via neural network. There are also some DHAs specifically designed for aerial images. Long *et al.* [13] proposed a remote sensing image DHA based on DCP. Pan *et al.* [12] introduced a method based on the differences of dark channels between remote sensing images and regular images. Xu *et al.* [14] proposed a cloud removal method by learning the sparse representations of the cloudy and cloud-free areas. The readers can refer to the relevant surveys for more DHAs [1], [2].

C. DHA Evaluation

An overview and discussion of DHA evaluation has been given in Section I. The situation is that current DHA evaluation follows the two strategies illustrated in Fig. 1. Using real hazy images is the most desirable way, but it is not easy to conduct quantitative evaluation. Though some quantitative measures are proposed [15], they show low correlation with the overall perceptual quality of the dehazing [16], [17]. The reason is that they only focus on the quality of contrast enhancement, and they omit the sky regions where side-effects occur frequently. Thus qualitative comparison is more recognized under such strategy, but it undergoes all drawbacks of subjective testing. To overcome this situation, using synthetic hazy images is introduced [1], [2], [8]–[11], [20]. It is easy to perform quantitative evaluation under such strategy, and it has been becoming widely recognized. Besides DHA evaluation, some methods use this strategy to learn a mapping from hazy images to haze-free images [8], [11]. For aerial images, the DHA evaluation inherits the methods from DHA evaluation for regular images. In this paper, we first propose a quality measure for reliable DHA evaluation, and then improve it for aerial images by incorporating the specific characteristics.

III. Subjective Evaluation of Dehazing Methods Using Synthetic Hazy Images

Intuitively, current state-of-the-art FR IQA measures may not be effective enough for FR DHA evaluation due to the reasons described in Section I. To verify this and to facilitate the design of FR DHA evaluation method, we construct a synthetic haze removing quality (SHRQ) database and conduct subjective experiments on this database.

A. Database Construction

The SHRQ database consists of two subsets: the regular image subset and aerial image subset. Regular images denote indoor or outdoor images captured in our daily life, while aerial images denote the visible-light images captured by remote sensing satellite.

1) The Regular Image Subset: We collect 45 high quality haze-free regular images and the corresponding depth from [23] and the Middlebury Stereo Datasets [24], [25]. The image resolution varies from 610×555 to 1024×680. We utilize the atmospheric scattering model described in Section II-A to synthesize haze. The real scene radiance $\mathbf{J}$ and the depth d are available. Following the work presented in [11], [20], the scattering coefficient of the atmosphere β is set by default to 1, which indicates moderate and homogenous haze, and the atmospheric light $\mathbf{A}$ is set to 1. Then we can synthesize hazy image $\mathbf{I}$ via Eq. (1) and Eq. (2). The synthesized hazy images are then processed by 8 state-of-the-art DHAs, including Fattal08 [3], Tarel09 [4], He09 [5], Xiao12 [6], Meng13 [7], Lai15 [9], Berman16 [10], and Cai16 [11]. A total of 360 regular dehazed images are generated.

2) The Aerial Image Subset: We collect 30 high quality aerial images from the AID database [26]. All images share the same resolution of 600×600. Considering that aerial images are captured from very high altitude and most scene points share a similar depth, we ignore the influence of depth and synthesize hazy images via Eq. (1) directly. The transmission t is set as a constant randomly selected from range [0.1, 0.7], and the atmospheric light $\mathbf{A}$ is set to a random value in range [0.7, 1]. Similar haze synthesizing protocol has been used in many dehazing studies [2], [8]. The same 8 DHAs are used to generate 240 aerial dehazed images.

The dehazed images, synthesized hazy images, and the reference haze-free images together constitute the SHRQ database. Example reference and the corresponding synthesized hazy images are shown in Fig. 3.

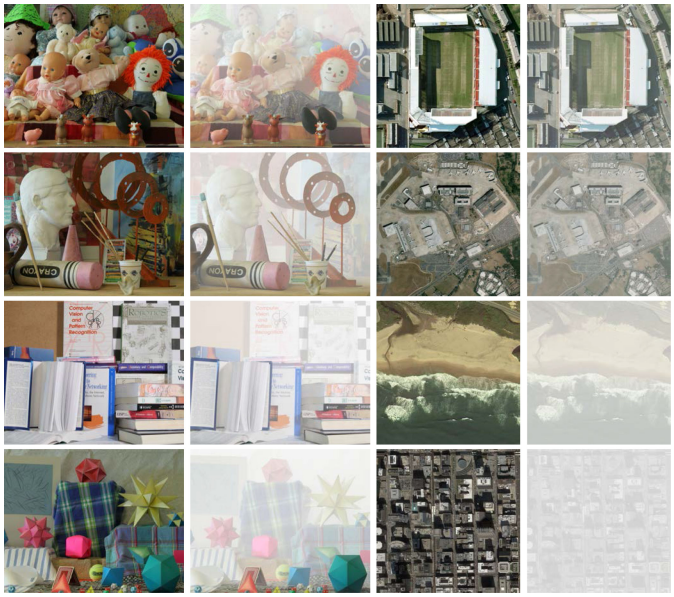


Fig. 3. Example reference and the corresponding synthesized hazy images in the SHRQ database. Left: regular images, right: aerial images.

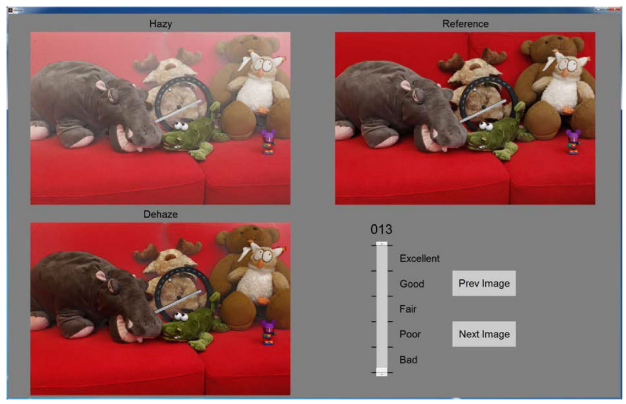


Fig. 4. GUI for subjective rating. The hazy, reference, and dehazed images are shown to the subjects.

B. Subjective Testing

We perform subjective quality assessment experiments with the SHRQ database. Human subjects need to rate the quality of the dehazed images using a five-grade continuous quality scale. Besides the dehazed image, the hazy image and the reference haze-free image are also shown, and subjects are asked to give an overall rating considering both haze removing effect and image content preserving effect. It means that a good DHA should not only remove haze but also preserve the original image content. All test images are shown in a random order with a MATLAB graphical user interface (GUI) on a LED monitor, which is calibrated according to the recommendations of ITU-R BT.500-13 [27]. A screenshot of the rating GUI is shown in Fig. 4. All images are shown at the original resolutions. A total of 38 subjects participate in the subjective experiments. The subjects are seated at a viewing distance of around 3 times the image height in a laboratory environment which has normal indoor illumination levels. The full test is divided into 3 sessions, and each session lasts less than 30 minutes. 17 valid subjects participate in the two sessions of the regular subset, and 18 valid subjects participate in another session of the aerial subset. Table I lists an overview of the test methodology and condition.

C. Data Processing and Analysis

We follow the recommendations given in [27] to exclude outliers and reject subjects. Rating for an image is considered

TABLE I
SUBJECTIVE EXPERIMENT SETUP

Category	Item	Detail
Monitor	Model	EIZO RX440 / LED / 29.8 in
	Resolution	2560×1600
Methodology	Method	Triple-stimulus
	Quality-scale	5-grade continuous
	Presentation order	Random
Test settings	Sessions	3
	Subjects number	35 valid / 3 outliers 24 male / 14 female
	Viewing distance	3 times image height
	Environment	Laboratory

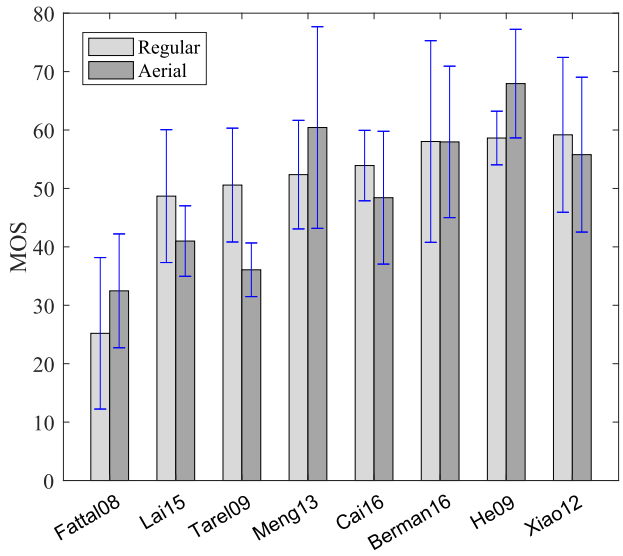


Fig. 5. Mean and std of subjective ratings of all compared DHAs in the regular and aerial subsets of the SHRQ database.

as outlier if it is outside 2 (if normal) or $\sqrt{20}$ (if non-normal) standard deviations (stds) about the mean rating of that image. A subject with more than 5% outlier evaluations are rejected. Four subjects are rejected in our experiments. Mean opinion score (MOS) is calculated for each dehazed image. We first normalize the ratings of each subject: $z_{ij} = \frac{r_{ij} - r_i}{\sigma_i}$, where r_i is the mean rating given by the ith subject, and σ_i is the std. Then the ratings for each image are averaged: $z_j = \frac{1}{N_j} \sum_{i=1}^{N_j} z_{ij}$, where N_j is the number of valid ratings (after outlier removing) for the jth image. At last a liner scaling is applied to derive the final MOS value: $MOS_j = \frac{100(z_j + 3)}{6}$.

To have a compare of the DHAs, we illustrate the mean and std of the subjective ratings of the dehazed images derived from each DHA in Fig. 5. Subjective scores of the regular and aerial subsets are shown separately. The overall performance rankings of all compared DHAs are similar in regular and aerial images. He09, Xiao12, and Berman16 perform the best, while Fattal08, Lai15, and Tarel09 show lower mean ratings, and the rest are in between. But note that the specific relative rankings are quite different in regular and aerial images. It indicates that some DHAs are better at regular images while some other are better at aerial images. The stds of all DHAs are quite large, which means that image content has influence on the effectiveness of the DHAs.

TABLE II
SRCC PERFORMANCE OF FR IQA MEASURES ON THE SHRQ DATABASE

Subset	PSNR	SSIM	MS-SSIM	VIF	MAD
Regular	0.5972	0.5627	0.5836	0.6287	0.5780
Aerial	0.8246	0.8207	0.7895	0.7048	0.6308
Subset	IW-SSIM	GSI	FSIM	IFC	PSIM
Regular	0.5657	0.6029	0.6256	0.5549	0.6238
Aerial	0.7949	0.7832	0.7424	0.5630	0.7593

D. Performance of Current FR IQA Measures

In recent years, some quality evaluators based on structural computational model of human visual system have been proposed [28]–[31]. As discussed in Section I, the similarity between dehazed images and reference haze-free images calculated via image quality evaluators is used as the quantitative evaluator under the strategy of DHA evaluation using synthetic hazy images [1], [8]–[11], [20]. We test if state-of-the-art FR IQA measures are accurate enough for FR DHA evaluation. We select 10 recognized FR IQA measures, including PSNR, SSIM [21], MS-SSIM [32], VIF [33], MAD [34], IW-SSIM [35], GSI [36], FSIM [37], IFC [38] and PSIM [39], and test their performance on the constructed SHRQ database. Table II lists the corresponding performance. Only spearman rank-order correlation coefficient (SRCC) is listed here. More evaluation results using other criteria are given in Section VI. It is observed that FR IQA measures are relatively more effective in aerial images. But all measures are not effective enough, and a more effective measure is needed for DHA evaluation.

IV. THE PROPOSED OBJECTIVE QUALITY MEASURE

To develop a new dehazing quality measure, we first need to analyse the introduced typical artifacts. Fig. 6 illustrates several typical failures of image dehazing, including poor dehazing effect, structural damage, color shift, and over-enhancement of the low-contrast areas. Some DHAs adopt a moderate strategy to avoid side-effects during dehazing, but it may leave too much haze in the image. Structural damage is another source of distortion, and the image content is destroyed during dehazing. Color shift generally occurs when the hazy is dense, and it makes the inferring of the original color difficult. Over-enhancement is generally observed in the low-contrast areas, where some hardly-perceived image details are taken as the image structures and enhanced out. To cope with these distortions, we introduce a quality measure by integrating three components: image structure recovering, color rendition, and over-enhancement. The measure proposed in this section is a general measure, and it works for both regular and aerial images. We will incorporate the specific characteristics of aerial images, and make it more effective for aerial images in the next section. The details of the measure are as follows.

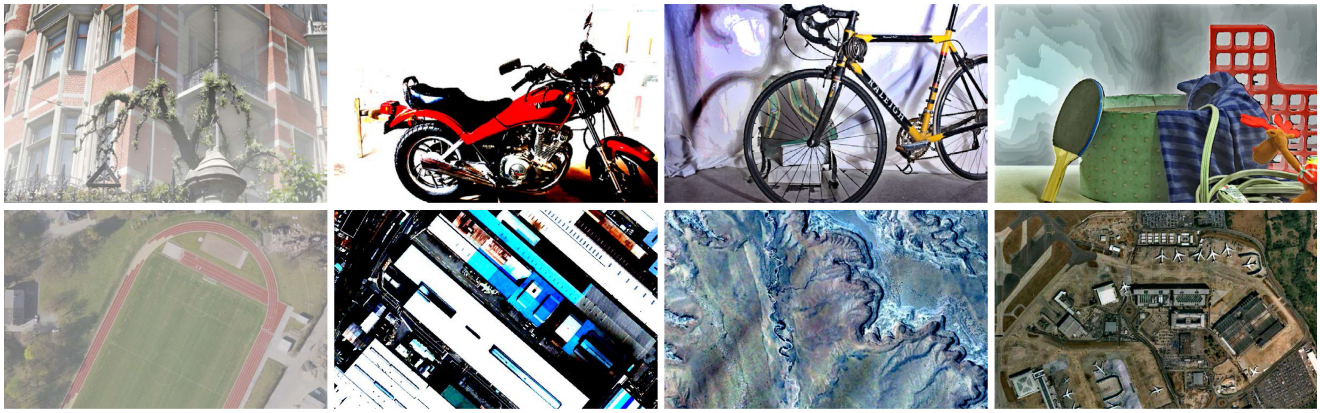


Fig. 6. Several typical failures of image dehazing. All 8 dehazed images has relatively low perceptual quality. Top: regular image, bottom: aerial image.

A. Image Structure Recovering

We deal with the poor dehazing effect and structural damage with a unified structure recovering map, since these two distortions both result in loss of image structures. Image structure is widely used in many IQA measures due to its effectiveness of capturing image degradation [21], [32], [35]–[37], [39], [40]. We extract haze-aware structural features and do some modifications to them to consider the differences between FR DHA evaluation and FR IQA. Then structural similarity is utilized as the core feature. We only consider luminance information when extracting structural features.

1) Haze-Aware Structural Features: The extracted haze-aware structural feature is based on traditional structural features. Given an image $\mathbf{I}$, we first compute the local mean and variance [21], [41]–[43]

$$\boldsymbol{\mu}(i,j) = \sum_{k,l} \mathbf{w}(k,l)\mathbf{I}(i+k, j+l), \tag{3}$$

$$\boldsymbol{\sigma}(i,j) = \sqrt{\sum_{k,l} \mathbf{w}(k,l)\big[\mathbf{I}(i+k,j+l) - \boldsymbol{\mu}(i,j)\big]^2}, \tag{4}$$

where i, j are pixel indexes, $\boldsymbol{\mu}$ is the local mean, and $\mathbf{w}$ is a local Gaussian weighting window. $\boldsymbol{\sigma}$ is a good structural feature which is sensitive to haze, since haze introduces contrast reduction. But local variance $\boldsymbol{\sigma}$ is sensitive to local mean $\boldsymbol{\mu}$. Thus we derive the normalized local variance

$$\boldsymbol{\eta} = \frac{\boldsymbol{\sigma}}{\boldsymbol{\mu} + \epsilon_1}, \tag{5}$$

where ϵ_1 is a positive constant used to avoid instability, and $\boldsymbol{\eta}$ is the desired haze-aware structural feature. This feature was previously used in [44] to estimate the haze density. It has been proved to be a good descriptor of both haze and image structures. Via Eq. (3)–Eq. (5), we can derive the local mean, variance, and normalized variance for the reference haze-free image $\boldsymbol{I}_r$ and the dehazed image $\boldsymbol{I}_d$: $\boldsymbol{\mu}_r, \boldsymbol{\mu}_d, \boldsymbol{\sigma}_r, \boldsymbol{\sigma}_d, \boldsymbol{\eta}_r, \boldsymbol{\eta}_d$.

2) Structural Feature Modification: In FR IQA, structural similarity is a frequently utilized strategy. But as shown in Table II, the traditional way of using structural similarity is not effective enough for FR DHA evaluation. It may be caused by two reasons. The first is that the traditional structural features are not specifically designed for dehazing and they are not sensitive enough to haze. The second is that they have not considered the problem illustrated in Fig. 2, and there are some differences between FR DHA evaluation and traditional FR IQA. In this paper, we first extract haze-aware structural features as described above, and then do some modifications to them to cope with these problems.

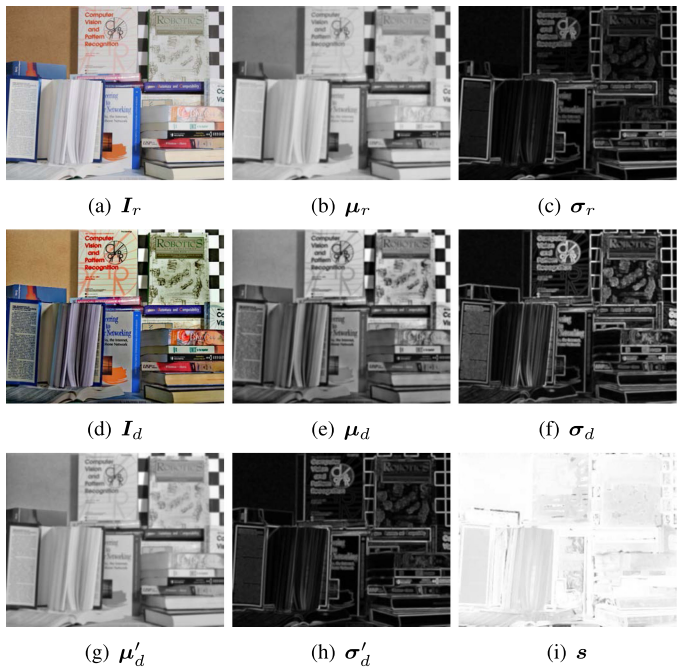


Fig. 7. An illustration of the situation that $\boldsymbol{I}_d$ is more enhanced than $\boldsymbol{I}_r$. The relevant feature maps (linear scaled for better visualization) are also shown.

Fig. 7 illustrates another example which has the same problem shown in Fig. 2. The relevant feature maps are also shown. $\boldsymbol{I}_d$ is not close to $\boldsymbol{I}_r$ from a perspective of image fidelity, but $\boldsymbol{I}_d$ still has high perceptual quality. We observe that this situation generally occurs when $\boldsymbol{I}_d$ is more enhanced than $\boldsymbol{I}_r$. In this case, we generally have $\boldsymbol{\mu}_r > \boldsymbol{\mu}_d$ and $\boldsymbol{\sigma}_r < \boldsymbol{\sigma}_d$, since some DHAs try to depress the luminance and boost the contrast to enhance the image. This phenomenon is easily observed in Fig. 7. If we

compare $\boldsymbol{I}_d$ and $\boldsymbol{I}_r$ directly like traditional FR IQA measures, the similarity will be low. Thus we modify $\boldsymbol{\mu}_d$, $\boldsymbol{\sigma}_d$, and $\boldsymbol{\eta}_d$ to solve this problem.

Traditional FR IQA measures will punish the situation that $\boldsymbol{I}_d$ has lower $\boldsymbol{\mu}$ and higher $\boldsymbol{\sigma}$ than $\boldsymbol{I}_r$. We think that contrast enhancement often decreases $\boldsymbol{\mu}$ and increases $\boldsymbol{\sigma}$, and they do not do much harm to the perceptual quality. Thus we try to weaken this punishment. Specifically, we modify $\boldsymbol{\mu}_d$ via

$$\begin{aligned}\boldsymbol{\mu}'_d &= f_\mu(\boldsymbol{\mu}_d, \boldsymbol{\mu}_r) \\ &= \begin{cases} \boldsymbol{\mu}_r + k \cdot (\boldsymbol{\mu}_d - \boldsymbol{\mu}_r) & \text{if } \boldsymbol{\mu}_d < \boldsymbol{\mu}_r \\ \boldsymbol{\mu}_d & \text{otherwise} \end{cases}, \end{aligned} \tag{6}$$

and modify $\boldsymbol{\sigma}_d$ via

$$\begin{aligned}\boldsymbol{\sigma}'_d &= f_\sigma(\boldsymbol{\sigma}_d, \boldsymbol{\sigma}_r) \\ &= \begin{cases} \boldsymbol{\sigma}_r + k \cdot (\boldsymbol{\sigma}_d - \boldsymbol{\sigma}_r) & \text{if } \boldsymbol{\sigma}_d > \boldsymbol{\sigma}_r \\ \boldsymbol{\sigma}_d & \text{otherwise} \end{cases}, \end{aligned} \tag{7}$$

where $\boldsymbol{\mu}'_d$ and $\boldsymbol{\sigma}'_d$ are the modified local mean and variance of the dehazed image, k is a linear scaling factor which is lower than 1 and it is empirically set. We will test the proposed method's sensitivity to k in Section VI. f_μ and f_σ increase the calculated similarity between $\boldsymbol{I}_d$ and $\boldsymbol{I}_r$ when $\boldsymbol{\mu}_d < \boldsymbol{\mu}_r$ and $\boldsymbol{\sigma}_d > \boldsymbol{\sigma}_r$. After the modification, we can derive the modified normalized local variance of the dehazed image

$$\boldsymbol{\eta}'_d = \frac{\boldsymbol{\sigma}'_d}{\boldsymbol{\mu}'_d + \epsilon_1}. \tag{8}$$

3) Structure Recovering: Then we compute the haze-aware structure recovering map using the similarity function widely used in IQA [21], [32], [35]–[37], [39]

$$\boldsymbol{s} = \frac{2\boldsymbol{\eta}_r \cdot \boldsymbol{\eta}'_d + \epsilon_2}{\boldsymbol{\eta}_r^2 + \boldsymbol{\eta}_d'^2 + \epsilon_2}, \tag{9}$$

where ϵ_2 is a positive constant acting the same stabilization function as ϵ_1. Fig. 7 illustrates examples of the original and modified feature maps and the final structure recovering map. Compared with traditional image fidelity measures, better similarity between $\boldsymbol{I}_r$ and $\boldsymbol{I}_d$ is shown in $\boldsymbol{s}$.

B. Color Rendition

Besides structures, color information is also important cue for quality estimation [37], [39]. Since dehazing can cause color shift as illustrated in Fig. 8, we incorporate a color rendition component into the proposed method. The widely used YIQ color space is utilized to transfer the given image

$$\begin{cases} \boldsymbol{y} = & 0.299\boldsymbol{r} + 0.587\boldsymbol{g} + 0.114\boldsymbol{b} \\ \boldsymbol{i} = & 0.596\boldsymbol{r} - 0.274\boldsymbol{g} - 0.322\boldsymbol{b}, \\ \boldsymbol{q} = & 0.211\boldsymbol{r} + 0.523\boldsymbol{g} + 0.312\boldsymbol{b} \end{cases} \tag{10}$$

where $\boldsymbol{r}$, $\boldsymbol{g}$, $\boldsymbol{b}$ are the RGB components of the input image, $\boldsymbol{y}$ is the transferred luminance information, and $\boldsymbol{i}$, $\boldsymbol{q}$ represent the chrominance information.

Since we have considered luminance information in the structure recovering component, we only consider chrominance information here. For $\boldsymbol{I}_r$ and $\boldsymbol{I}_d$, we apply the same transformation, and derive the chrominance as $\boldsymbol{i}_r$, $\boldsymbol{i}_d$ and $\boldsymbol{q}_r$, $\boldsymbol{q}_d$. Then we calculate the color rendition as

$$\boldsymbol{c} = \boldsymbol{c}_i \cdot \boldsymbol{c}_q = \frac{2\boldsymbol{i}_r \cdot \boldsymbol{i}_d + \epsilon_3}{\boldsymbol{i}_r^2 + \boldsymbol{i}_d^2 + \epsilon_3} \cdot \frac{2\boldsymbol{q}_r \cdot \boldsymbol{q}_d + \epsilon_3}{\boldsymbol{q}_r^2 + \boldsymbol{q}_d^2 + \epsilon_3}, \tag{11}$$

where ϵ_3 is a stabilization constant. Fig. 8 illustrates an example of the calculated color rendition map. It is observed that $\boldsymbol{c}$ captures the color shift well.

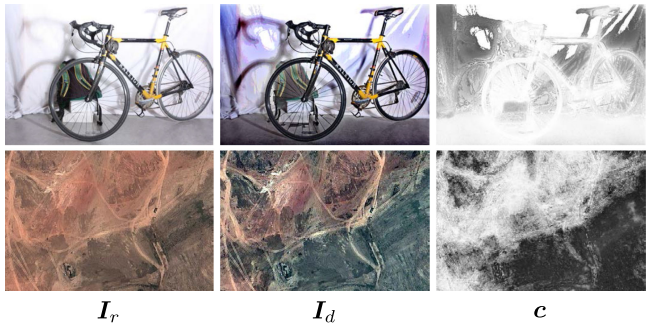


Fig. 8. An illustration of color shift and the calculated color rendition map. Top: regular image, bottom: aerial image.

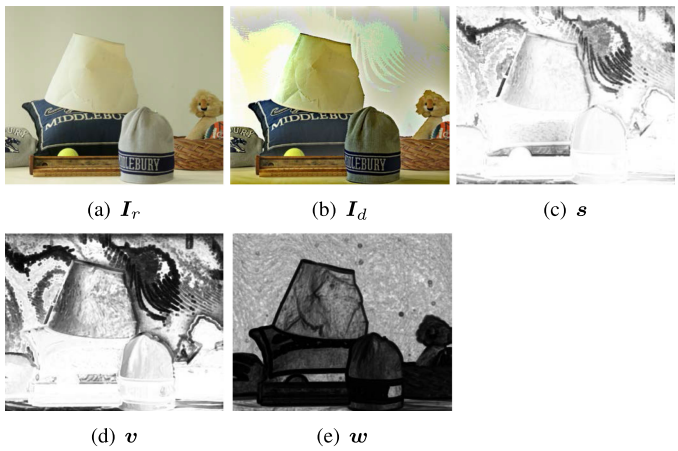


Fig. 9. An illustration of over-enhancement in the low-contrast areas and several related feature maps.

C. Over-Enhancement

As illustrated in Fig. 9, over-enhancement in low-contrast areas is another major distortion. Some hardly-observed image details are enhanced as image structures. Ideally, such kind of over-enhancement should be able to be captured by structure-like descriptors, since it introduces large variance changes. It is observed from Fig. 9(c) that the introduced structure recovering map $\boldsymbol{s}$ does capture such over-enhancement. But compared with its harm to the overall perceptual quality, the describing ability is weak. It results from two reasons: on one hand, over-enhancement in low-contrast areas can do much more harm than in textured areas, and on the other hand, low-contrast areas are usually background areas which occupy a small part of the scene. Thus we introduce an over-enhancement term specifically for such distortions.

We introduce a variance similarity map $\boldsymbol{v}$ to better describe the over-enhancement in low-contrast areas

$$\boldsymbol{v} = \frac{2\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_d + \epsilon_4}{\boldsymbol{\sigma}_r^2 + \boldsymbol{\sigma}_d^2 + \epsilon_4}. \tag{12}$$

As illustrated in Fig. 9(c) and Fig. 9(d), $\boldsymbol{s}$ describes the perceptual quality of textured foreground areas better, whereas $\boldsymbol{v}$ captures the over-enhancement of smooth background areas better. Considering the success of content-based or visual attention-based pooling in IQA [45]–[51], we introduce a content-based weighting map

$$\boldsymbol{w} = \frac{1}{\boldsymbol{\sigma}_r + \epsilon_5}. \tag{13}$$

Fig. 9(e) illustrates an example of $\boldsymbol{w}$. Contrary to traditional content-based weighting maps which highlight the content-rich areas, $\boldsymbol{w}$ highlights the low-contrast background areas where over-enhancement occurs frequently. It is what we desired. Then the overall over-enhancement is dscribed as

$$o = \frac{\sum_{i,j} \boldsymbol{v}(i,j) \cdot \boldsymbol{w}(i,j)}{\sum_{i,j} \boldsymbol{w}(i,j)}, \tag{14}$$

where i, j are pixel indexes.

D. Overall Quality Estimation

Finally the overall quality Q is estimated from image structure recovering map $\boldsymbol{s}$, color rendition map $\boldsymbol{c}$, and over-enhancement o

$$Q = \frac{1}{Z} \sum_{i,j} \boldsymbol{s}(i,j) \cdot \left[\boldsymbol{c}(i,j)\right]^{\lambda} \cdot o, \tag{15}$$

where Z is a normalization factor which represents the total number of pixels, and i, j are pixel indexes. Similar to FSIM$_c$ [37], we introduce a parameter λ to adjust the importance of the color information. λ is empirically set, and we will test the method's sensitivity to it in Section VI.

V. Improved Quality Measure for Aerial Images

The above measure is a general dehazing quality measure which works for various kinds of images, but it has not considered the characteristics of aerial images. In this section, we first analyse the differences between regular image dehazing and aerial image dehazing, and then improve the above measure specifically for aerial images.

A. Differences Between Regular and Aerial Image Dehazing

We have observed two major differences between regular and aerial image dehazing, which involve the two major features utilized by the above measure, i.e., the color rendition and the over-enhancement of low-contrast areas. The color rendition feature is used to incorporate the color shift introduced by DHAs. We single out the color rendition feature from Eq. (15), and describe it as

$$q_c = \frac{1}{Z} \sum_{i,j} \left[\boldsymbol{c}(i,j)\right]^{\lambda}. \tag{16}$$

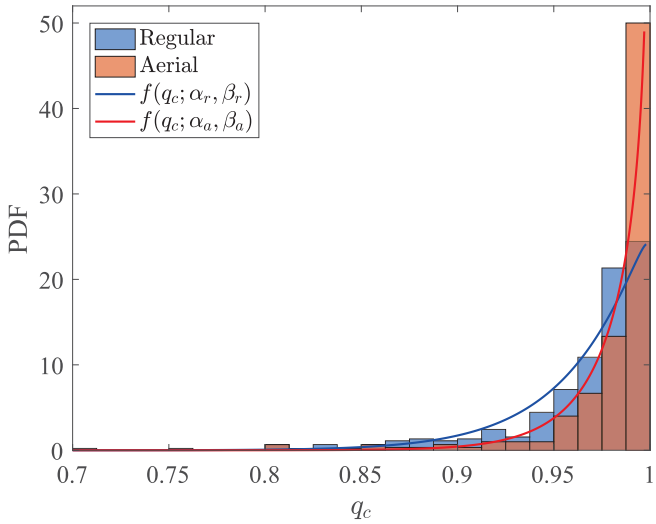


Fig. 10. Regular and aerial images' histograms of the color rendition feature q_c and the beta distribution fittings. q_c of aerial images are more clustered near 1, which indicates weaker quality describing ability. PDF: probability density function.

Then we compute q_c in the regular and aerial subsets of the SHRQ database. The histograms of q_c are illustrated in Fig. 10. It is observed that the histograms generally follow the beta distribution, whose probability density function (PDF) is

$$f(q_c;\alpha,\beta) = \frac{q_c^{\alpha-1}(1-q_c)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, \tag{17}$$

where $\alpha, \beta > 0$ are two shape parameters, $\mathrm{B}(\cdot)$ is the beta function. The fitted curves are also given in Fig. 10. It is observed from both the histograms and fitted cures that q_c of aerial images are more clustered near 1 and more regular images have lower q_c values. It indicates that q_c has weaker quality describing ability in aerial images.

The over-enhancement feature is used to consider the over-enhancement in low-contrast areas introduced by DHAs. Examples of over-enhancement effect are illustrated in Fig. 6. From this figure, we can observe the differences between the over-enhancement in regular and aerial images. The over-enhancement in regular image is generally observed in the low-contrast areas, where some hardly-observed image details are enhanced as image structures, and such kind of over-enhancement does much harm to the perceptual quality. While the over-enhancement in aerial image is wide-spread over the whole image, and such over-enhancement is easily covered by the image details and thus does less harm to the perceptual quality. The reason behind such difference is that the contents of aerial images are large scale earth surface captured from high altitude and these images are generally rich in image details. Whereas the contents of regular images are usually in much smaller scale and often contains large smooth areas with low contrast. To have a better understanding of this, we compute the average local variance of all reference regular and aerial images in the SHRQ database

$$v = \frac{1}{Z} \sum_{i,j} \boldsymbol{\sigma}_r, \tag{18}$$

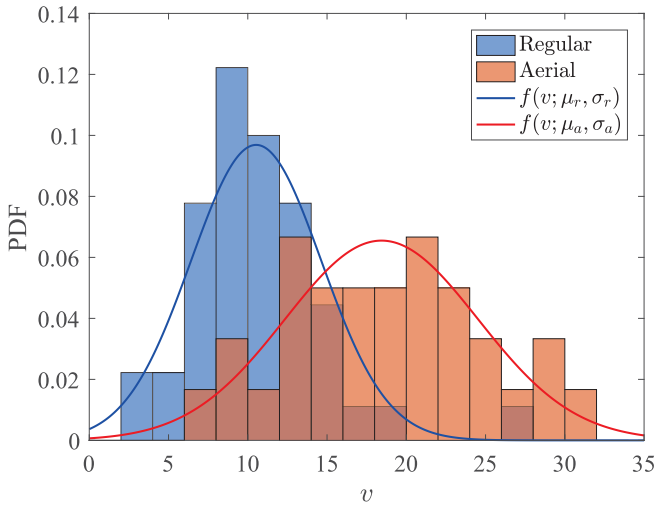


Fig. 11. Regular and aerial images' histograms of the average local variance v and the normal distribution fittings. Aerial images have larger local variances than regular images. PDF: probability density function.

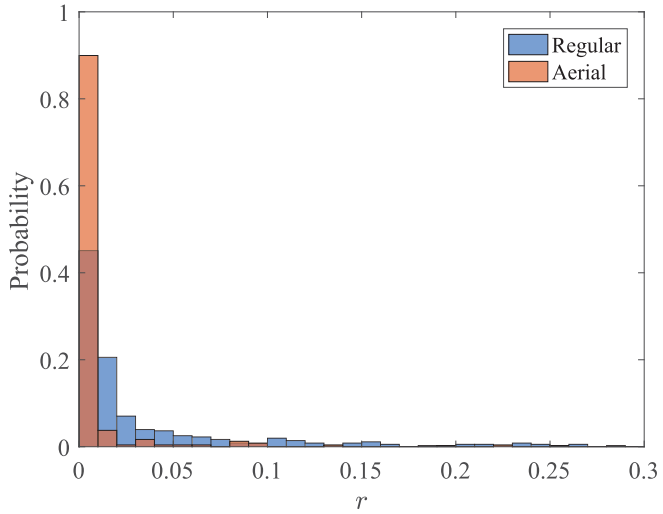


Fig. 12. Regular and aerial images' histograms of the proportion of image areas with low variance similarity and low local variance.

and illustrate their histograms in Fig. 11. It is observed that they generally follow the normal distribution, whose PDF is

$$f(v;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(v-\mu)^2}{2\sigma^2}\right), \tag{19}$$

where μ,σ are the mean and variance parameters. The fitted curves are also given in Fig. 11. It is observed from both the histograms and fitted cures that aerial images have larger local variances than regular images, which means that they have much richer image details.

The over-enhancement feature described by Eq. (14) computes the variance similarity of low-contrast areas. Such low-contrast areas have two characteristics: low variance similarity and low local variance. We collect some statistics of regular and aerial images' proportion of such areas. More specifically, we identify the image areas which have the lowest 30% variance similarity and the lowest 30% local variance at the same time, and calculate the proportion of such areas in the image. Fig. 12 illustrates the histograms of this proportion in regular and aerial images of the SHRQ database. It is observed that more than 90% of aerial images have no more than 1% of such areas. It indicates that over-enhancement is seldom observed in the low-contrast areas in aerial images.

B. Improved Measure for Aerial Images

To consider the differences between regular and aerial image dehazing described above, we improve the quality measure described by Eq. (15) for aerial images. First, considering that q_c has weaker quality describing ability in aerial images, we increase the importance of color information, i.e., the parameter λ, for aerial images. By increasing λ, q_c will spread to lower values and thus have stronger quality describing ability. Then, considering that over-enhancement is seldom observed in low-contrast areas in aerial images while the general over-enhancement effect can be well described by the image structure recovering feature, we remove the over-enhancement term from Eq. (15). Finally, the improved quality measure for aerial images can be described by

$$Q=\frac{1}{Z}\sum_{i,j}\boldsymbol{s}(i,j)\cdot\left[\boldsymbol{c}(i,j)\right]^{\lambda'}, \tag{20}$$

where the parameter λ' is specifically set for aerial images and it is larger than λ. λ' is also empirically set, and we will test the method's sensitivity to it in Section VI.

VI. Experimental Results

A. Experimental Settings

We test the proposed DHA quality measure on both subsets of the constructed SHRQ database. Since the proposed method follows the framework of FR IQA, we compare it with 10 state-of-the-art FR IQA measures, including PSNR, SSIM [21], MS-SSIM [32], VIF [33], MAD [34], IW-SSIM [35], GSI [36], FSIM [37], IFC [38], and PSIM [39]. We use the original implementations of all compared algorithms.

Following the practices in [39], [41], [52]–[54], we first map the predicted scores nonlinearly using a five-parameter logistic function

$$q(s)=\beta_1\left(\frac{1}{2}-\frac{1}{1+e^{\beta_2(s-\beta_3)}}\right)+\beta_4 s+\beta_5, \tag{21}$$

where $\{\beta_i|i=1,2,...,5\}$ are parameters determined via curve fitting, s and $q(s)$ are predicted and mapped quality scores. After mapping the predicted scores, we then evaluate the IQA measures using three criteria: Spearman rank-order correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC) and root-mean-square error (RMSE), which measure the prediction monotonicity, linearity and accuracy, respectively. Higher SRCC, PLCC and lower RMSE indicate better performance.

B. Performance Comparison With FR Measures

Table III summarizes the performance comparison results. We test the original measure on both subsets and test the

TABLE III
PERFORMANCE COMPARISON WITH FR IQA MEASURES ON THE SHRQ DATABASE. TIME: SECONDS/IMAGE

Subset	Criteria	A	B	C	D	E	F	G	H	I	J	K	L
		PSNR	SSIM	MS-SSIM	VIF	MAD	IW-SSIM	GSI	FSIM	IFC	PSIM	Pro.	Pro.+
Regular	SRCC	0.5972	0.5627	0.5836	0.6287	0.5780	0.5657	0.6029	0.6256	0.5549	0.6238	**0.8292**	-
	PLCC	0.6591	0.6225	0.6276	0.7609	0.6950	0.6172	0.6946	0.7419	0.7354	0.7580	**0.8675**	-
	RMSE	10.417	10.841	10.784	8.9885	9.9602	10.900	9.9650	9.2882	9.3873	9.0350	**6.8912**	-
Aerial	SRCC	0.8246	0.8207	0.7895	0.7048	0.6308	0.7949	0.7832	0.7424	0.5630	0.7593	0.8615	**0.9028**
	PLCC	0.8040	0.8166	0.7815	0.7651	0.6382	0.7841	0.7719	0.7348	0.6061	0.7338	0.8583	**0.9017**
	RMSE	9.6080	9.3252	10.081	10.404	12.438	10.028	10.272	10.958	12.852	10.976	8.2912	**6.9855**
	Time	**0.0017**	0.0109	0.0256	0.5985	0.8369	0.2353	0.1105	0.0123	0.5890	0.0394	0.0302	0.0286

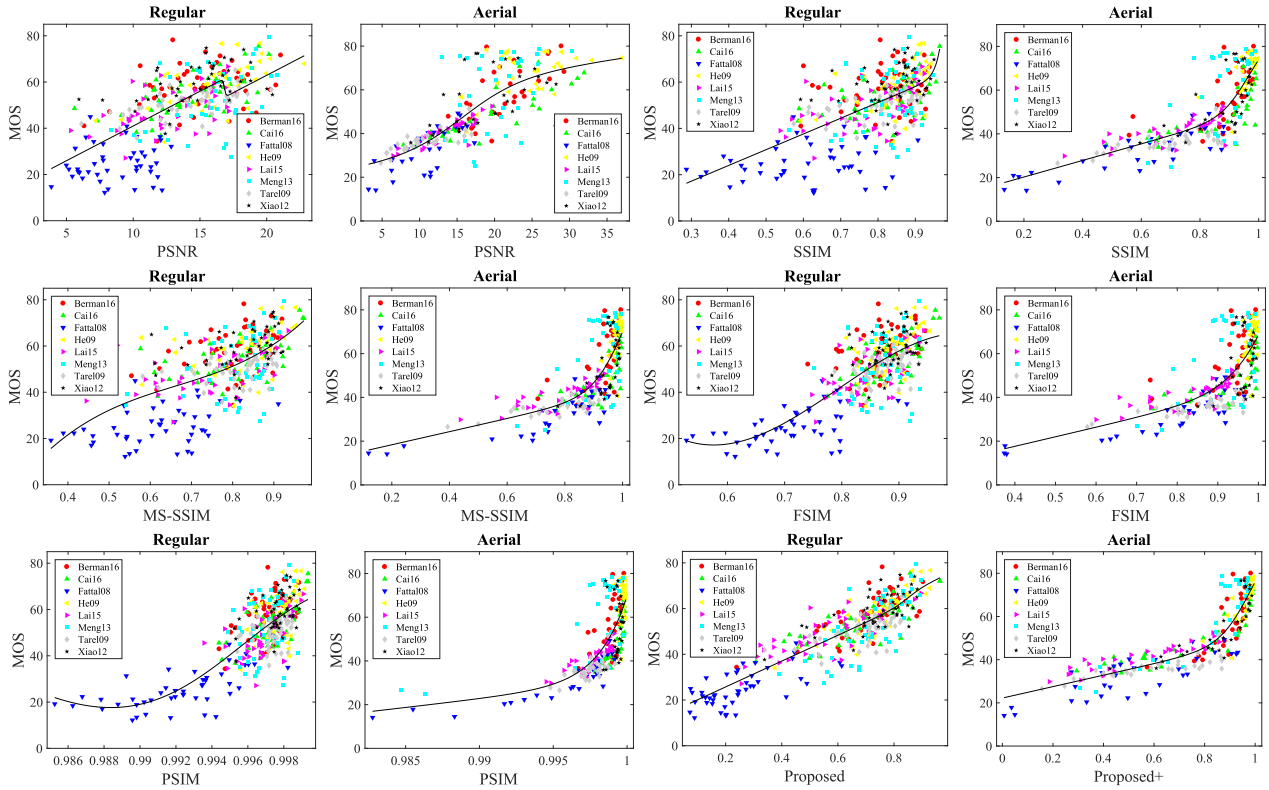


Fig. 13. Scatter plots of the proposed methods and representative FR IQA measures on the regular and aerial subsets of the SHRQ database. The (black) lines are curves fitted with the five-parameter logistic function. Different colors and shapes of the scatter points represent different DHAs.

specifically improved measure on the aerial image subset. It is observed that the proposed methods show the best performance on both subsets, and they lead by a large margin. Traditional FR IQA measures show certain prediction ability, but they are not accurate enough. It confirms the previous analysis that FR DHA evaluation is not an exact FR IQA problem. The haze-free image can be taken as a reference, but it may not be accurate to use it as the "ground truth" for DHA evaluation. The proposed method considers such differences, and extracts some dehazing-relevant features, thus achieves better performance. For aerial images, the original measure is also effective and it outperforms traditional FR IQA measures by a large margin, while the improved measure outperforms the original measure significantly. Fig. 13 illustrates the scatter plots of the proposed methods and representative FR IQA measures on both subsets of the SHRQ database. It is observed that the scatter points of the proposed methods are more clustered near the fitted curves on both subsets.

We conduct statistical significance tests to testify if the performance between any two models are statistically different on both subsets of the SHRQ database. Similar to [55], we test the performance by comparing the variances of residuals between subjective ratings and the nonlinear mapped scores. Higher/lower variance indicates worse/better performance. The utilized F-statistic is based on the ratio of two models' residual variances. The null hypothesis is that the residuals of two quality models are from the same distribution and they are statistically indistinguishable with a 95% confidence. We compare every possible pairs of models, and illustrate the significance test results in Fig. 14. The significant superiorities of the proposed methods are obvious, whereas many competitors are statistically indistinguishable with each other.

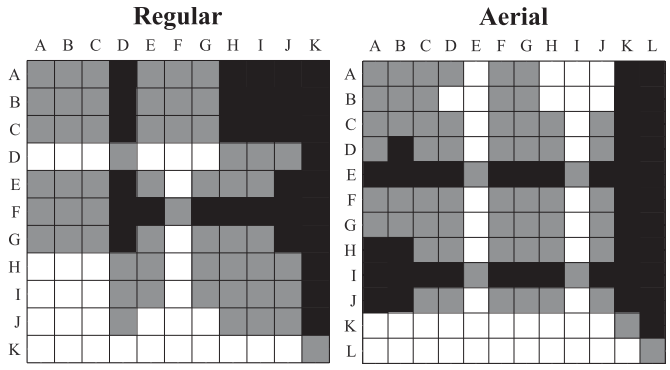


Fig. 14. Statistical significance test results on both subsets of the SHRQ database. A white/black block indicates that the row model is statistically better/worse than the column model. A gray block indicates that the row and column models are statistically indistinguishable. A-L are model indexes given in Table III.

C. Performance Comparison With NR Measures

As illustrated in Fig. 1, DHAs can be evaluated either using synthetic hazy images in a FR style, or using real hazy images in a NR style. If there are some accurate and reliable quality measures, NR evaluation using real hazy images would be the more straightforward and desirable way. But some subjective quality evaluation study [16] shows that current blind IQA models are not suitable for NR DHA evaluation. To further confirm this, we also test some relevant blind quality estimators, including

- Blind IQA measures for dehazed or contrast-enhanced images, including FADE [44], the three evaluators e, r and NS introduced in [15], BIQME [18], Fang15 [56], NIQMC [57], and DHQI [17].
- General-purpose blind IQA measures, e.g., BRISQUE [58], NFERM [29], dipIQ [59], MEON [60], BPRI [61], and BMPRI [62], which are assumed to be able to handle general IQA problems.

We test these measures on both subsets of the SHRQ database, and report their performance in Table IV. It is observed that no measure is effective for dehazed image quality prediction, which agrees with the study conducted in [16].

D. Ablation Experiment

The proposed methods are composed of several key components. We test the contributions of each part by conducting a series of ablation experiments. We test the performance of the proposed methods as shown in Eq. (15) and Eq. (20) under the following circumstances:

Case 1: The complete algorithm;
Case 2: Excluding the structure recovering term $\boldsymbol{s}$;
Case 3: Excluding the color rendition term $\boldsymbol{c}$;
Case 4: Excluding/Including the over-enhancement term o;
Case 5: Excluding the feature modification, i.e., $k = 1$.

The methods shown in Eq. (15) and Eq. (20) are tested on the regular and aerial subsets of the SHRQ database, respectively. The excluded or included are the 4 core parts of the proposed method. In *Case 4*, we test the methods by excluding term o from Eq. (15) or including term o into Eq. (20).

Table V lists the results of the ablation experiments. Any one of the last 4 cases has lower performance than *Case 1*, which means that they all make some contributions to the overall method. *Case 2* shows the lowest performance, which means that the structure recovering term $\boldsymbol{s}$ contributes most to the methods. Moreover, modifying the structure features introduces quite a large improvement, which confirms the previous analysis about the differences between FR IQA and FR DHA evaluation. The color rendition term $\boldsymbol{c}$ contributes little on the regular subset because the dehazed images in which the color information makes a large difference only occupy a small part of the database. While on the aerial subset, the color rendition term contributes much more because we have increased the importance of color information considering the differences between regular and aerial image dehazing as described in Section V. The over-enhancement term o contributes to the overall method in regular images, whereas it makes no contribution in aerial images. It confirms the previous analyses about the differences between regular and aerial image dehazing.

E. Parameter Sensitivity

The proposed method involves few parameters. The most important parameters are the linear scaling factor k in Eq. (6)–Eq. (7) and the color importance parameter λ in Eq. (15) or λ' in Eq. (20). k is parameter of the both measures, while λ and λ' are parameters of the original and improved measures, respectively. They are empirically set to 0.2, 0.1, and 0.35. We test the parameter sensitivity of k, λ, and λ' from 0.1 to 0.2 with a step of 0.05, from 0 to 0.2 with a step of 0.05, and from 0.25 to 0.45 with a step of 0.05, respectively. Both parameters can vary at the same time. The original and improved measures are tested on the regular and aerial subsets, respectively. Table VI lists the SRCC performance of all settings. It can be observed that the performance remains stable within a significantly wide range.

F. Computational Complexity

We test the computational complexity of all compared algorithms, and report the average running time (seconds/image) for 100 pairs of images with a fixed resolution of 512×512 in Table III. The algorithms are tested with MATLAB R2017a operating on a computer with Intel Core i7-7700K CPU @3.60 GHz and 32 GB RAM. The running time includes all feature extraction and quality prediction time. For all competitors, we use the original implementations provided by the authors. It is observed that the proposed methods have considerably low computational complexity. The improved measure performs slightly faster than the original measure since we exclude the over-enhancement feature.

G. Discussion

Considering the difficulty of evaluating DHAs quantitatively using real hazy images, the FR DHA evaluation using synthetic hazy images seems more promising. It is convenient to conduct quantitative evaluation and comparison, thus it has been

TABLE IV
PERFORMANCE COMPARISON WITH NR IQA MEASURES ON THE SHRQ DATABASE

Subset	Criteria	FADE	e	r	NS	BIQME	Fang15	NIQMC	DHQI	BRISQUE	NFERM	dipIQ	MEON	BPRI	BMPRI	Pro.(+)
Regular	SRCC	0.2958	0.2344	0.0200	0.0144	0.2751	0.4539	0.4025	0.4241	0.4196	0.1913	0.0417	0.2220	0.0144	0.2206	**0.8292**
	PLCC	0.2722	0.3026	0.1807	0.4324	0.2505	0.5728	0.5551	0.6621	0.5767	0.4574	0.1699	0.3151	0.1664	0.3721	**0.8675**
	RMSE	13.329	13.203	13.624	12.490	13.411	11.355	11.522	10.380	11.317	12.318	13.651	13.147	13.659	12.857	**6.8912**
Aerial	SRCC	0.6569	0.0980	0.1340	0.4024	0.7018	0.3820	0.6119	0.5675	0.1527	0.3992	0.0707	0.0339	0.2382	0.0895	**0.9028**
	PLCC	0.6743	0.1957	0.2311	0.2283	0.7062	0.6046	0.6325	0.6172	0.3261	0.4782	0.1231	0.2319	0.3709	0.3151	**0.9017**
	RMSE	11.931	15.845	15.720	15.730	11.440	12.870	12.515	12.713	15.274	14.190	16.034	15.717	15.005	15.334	**6.9855**

TABLE V
PERFORMANCE RESULTS OF THE ABLATION EXPERIMENTS

Subset	Criteria	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 5*
Regular	SRCC	0.8292	0.6832	0.8211	0.7545	0.7526
	PLCC	0.8675	0.7839	0.8617	0.8169	0.8337
	RMSE	6.8912	8.6001	7.0295	7.9903	7.6500
Aerial	SRCC	0.9028	0.5462	0.8514	0.8666	0.8391
	PLCC	0.9017	0.5368	0.8650	0.8562	0.8229
	RMSE	6.9855	13.632	8.1075	8.3463	9.1798

TABLE VI
SRCC PERFORMANCE RESULTS OF THE PARAMETER SENSITIVITY TESTS. TOP: THE GENERAL MEASURE ON THE REGULAR SUBSET, BOTTOM: THE IMPROVED MEASURE ON THE AERIAL SUBSET

λ \ k	0.1	0.15	0.2	0.25	0.3
0	0.8210	0.8216	0.8211	0.8204	0.8187
0.05	0.8279	0.8277	0.8273	0.8256	0.8243
0.1	0.8309	0.8304	0.8292	0.8268	0.8252
0.15	0.8301	0.8295	0.8282	0.8262	0.8242
0.2	0.8277	0.8274	0.8260	0.8235	0.8213
λ' \ k	0.1	0.15	0.2	0.25	0.3
0.25	0.9049	0.9051	0.9045	0.9021	0.8995
0.3	0.9067	0.9060	0.9052	0.9025	0.8992
0.35	0.9051	0.9043	0.9028	0.9007	0.8983
0.4	0.9038	0.9025	0.9007	0.8988	0.8966
0.45	0.9018	0.9009	0.8991	0.8978	0.8950

becoming more and more widely accepted. Many recent works have utilized this strategy for DHA evaluation [1], [8]–[11], [20]. This paper follows this FR DHA evaluation strategy. We point out that using these FR measures as the criteria is questionable, and they have low correlation with the subjective evaluation. Moreover, effective quality measures are proposed for better DHA evaluation in regular and aerial images.

VII. CONCLUSION

Evaluating DHAs using hazy images synthesized from reference haze-free images is a widely adopted strategy. In the current literature, FR IQA measures are used as the evaluation criteria. But no work has been conducted to test the effectiveness of FR IQA measures in DHA evaluation. In this paper, we study this strategy systematically. We construct a SHRQ database. It consists of a regular image subset and an aerial image subset, which respectively include 360 and 240 dehazed images created from 45 and 30 synthetic hazy images using 8 representative DHAs. A subjective user study is conducted on the database. We observe that there are some differences between FR DHA evaluation and FR IQA, and state-of-the-art FR IQA measures are not suitable for the objective of FR DHA evaluation.

To tackle these problems, we analyse the typical distortions introduced by dehazing, and propose a FR DHA evaluation method considering the image structure recovering, color rendition, and over-enhancement of low-contrast areas. To consider the differences between FR DHA evaluation and FR IQA, we modify the structural features of the dehazed image. The differences between regular and aerial image dehazing are also analysed, and we improve the method for aerial images by incorporating the specific characteristics of aerial images. Effectiveness of the proposed method is verified on both subsets of the SHRQ database, and all key components contribute to the overall method. Due to the lack of reliable quantitative measures for DHA evaluation using real hazy images, evaluating DHAs using synthetic hazy images is a promising way. The proposed method provides an effective evaluation measure, which is of great value for such strategy.

REFERENCES

[1] Y. Li, S. You, M. S. Brown, and R. T. Tan, "Haze visibility enhancement: A survey and quantitative benchmarking," *Comput. Vis. Image Understanding*, vol. 165, pp. 1–16, 2017.
[2] B. Li *et al.*, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[3] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 72:1–72:9, 2008.
[4] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE Conf. Comput. Vis.*, 2009, pp. 2201–2208.
[5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1956–1963.
[6] C. Xiao and J. Gan, "Fast image dehazing using guided joint bilateral filter," *Vis. Comput.*, vol. 28, no. 6–8, pp. 713–721, 2012.
[7] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Conf. Comput. Vis.*, 2013, pp. 617–624.
[8] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2995–3000.
[9] Y.-H. Lai, Y.-L. Chen, C.-J. Chiou, and C.-T. Hsu, "Single-image dehazing via optimal transmission map under scene priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 1–14, Jan. 2015.
[10] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1674–1682.
[11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[12] X. Pan, F. Xie, Z. Jiang, and J. Yin, "Haze removal for a single remote sensing image based on deformed haze imaging model," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1806–1810, Oct. 2015.

[13] J. Long, Z. Shi, W. Tang, and C. Zhang, "Single remote sensing image dehazing," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 59–63, Jan. 2014.

[14] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, "Cloud removal based on sparse representation via multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2998–3006, May 2016.

[15] N. Hautière, J.-P. Tarel, D. Aubert, and E. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Anal. Stereology*, vol. 27, no. 2, pp. 87–95, 2011.

[16] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 3600–3604.

[17] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[18] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.

[19] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3003–3010.

[20] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2226–2230.

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

[22] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[23] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, p. 13, 2014.

[24] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[25] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[26] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[27] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-13, Jan. 2012.

[28] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.

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

[30] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "Efficient image deblocking based on postfiltering in shifted windows," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 122–126, Jan. 2008.

[31] X. Yang, W. Ling, Z. Lu, E. Ong, and S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, 2005.

[32] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Sig., Syst., Comput.*, 2003, pp. 1398–1402.

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

[34] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 1–21, 2010.

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

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

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

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

[39] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[40] X. Min *et al.*, "Blind quality assessment of compressed images via pseudo structural similarity," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2016, pp. 1–6.

[41] X. Min *et al.*, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.

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

[43] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, Apr. 2013.

[44] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[45] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.

[46] 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 Process.*, vol. 145, pp. 127–136, 2018.

[47] X. Min, G. Zhai, K. Gu, and X. Yang, "Fixation prediction through multimodal analysis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 1, pp. 6:1–6:23, 2017.

[48] X. Min *et al.*, "Visual attention analysis and prediction on human faces," *Inf. Sci.*, vol. 420, pp. 417–430, 2017.

[49] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1266–1278, Jun. 2016.

[50] W. Zhang and H. Liu, "Study of saliency in objective video quality assessment," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1275–1288, Mar. 2017.

[51] Y. Zhu, G. Zhai, and X. Min, "The prediction of head and eye movement for 360 degree images," *Signal Process., Image Commun.*, vol. 69, pp. 15–25, 2018.

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

[53] Y. Zhou *et al.*, "Blind quality index for multiply distorted images using bi-order structure degradation and nonlocal statistics," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3019–3032, Nov. 2018.

[54] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of DIBR-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914–926, Apr. 2018.

[55] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[56] Y. Fang *et al.*, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[57] K. Gu *et al.*, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.

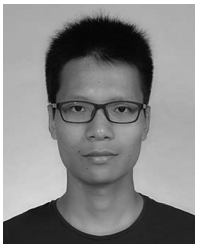
[58] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[59] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[60] K. Ma *et al.*, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[61] X. Min *et al.*, "Blind quality assessment based on pseudo reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.

[62] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 508–517, Jun. 2018.



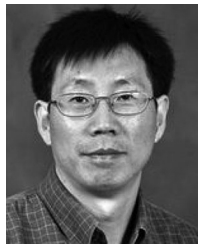
Xiongkuo Min (M'19) received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018. From Jan. 2016 to Jan. 2017, he was a visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently a Postdoctoral Fellow with Shanghai Jiao Tong University. His research interests include visual quality assessment, visual attention modeling, and perceptual signal processing. He was the recipient of the Best Student Paper Award at IEEE ICME 2016.

Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Postdoctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Erlangen, Germany. His research interests include multimedia signal processing and perceptual signal processing. He was the recipient of the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012.

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently the Professor with the Beijing University of Technology, Beijing, China. His research interests include image analysis, environmental perception, quality assessment, and machine learning. Dr. Gu is currently the Associate Editor for the IEEE ACCESS and the *IET Image Processing*, and is the Reviewer for 20 top SCI journals. He was the leading special session organizer in VCIP2016 and ICIP2017, and serves as the Guest Editor in the Digital Signal Processing journal. He is the recipient of the Best Paper Award of the IEEE TRANSACTIONS ON MULTIMEDIA, the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and the excellent Ph.D. thesis award from the Chinese Institute of Electronics in 2016.

Yucheng Zhu received the B.E. degree from the Shanghai Jiao Tong University, Shanghai, China, in 2015. He is currently working toward the Ph.D. degree with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He was the recipient of the Grand Challenge Best Performance Award in ICME 2017 and 2018.

Jiantao Zhou (M'11) received the B.Eng. degree from the Department of Electronic Engineering, Dalian University of Technology, Dalian, China, in 2002, the M.Phil. degree from the Department of Radio Engineering, Southeast University, Nanjing, China, in 2005, and the Ph.D. degree from the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, in 2009. He held various research positions with the University of Illinois at Urbana-Champaign, the Hong Kong University of Science and Technology, and the McMaster University. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China. His research interests include multimedia security and forensics, multimedia signal processing, artificial intelligence, and big data. He holds four granted U.S. patents and two granted Chinese patents. He has coauthored two papers that received the Best Paper Award at the IEEE Pacific-Rim Conference on Multimedia in 2007 and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Guodong Guo (M'07–SM'07) received the B.E. degree in automation from Tsinghua University, Beijing, China, the Ph.D. degree in pattern recognition and intelligent control from the Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in computer science from the University of Wisconsin–Madison, Madison, WI, USA. He is an Associate Professor with the Department of Computer Science and Electrical Engineering, West Virginia University (WVU), Morgantown, WV, USA. In the past, he visited and researched in several places, including Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France, Ritsumeikan University, Kyoto, Japan, Microsoft Research, Beijing, China, and North Carolina Central University, Durham, NC, USA. He has authored a book entitled *Face, Expression, and Iris Recognition Using Learning-Based Approaches* (University of Wisconsin–Madison, 2008), coedited a book entitled *Support Vector Machines Applications* (Springer, 2014), and authored or coauthored about 100 technical papers. His current research interests include computer vision, machine learning, and multimedia. Dr. Guo was a recipient of the North Carolina State Award for Excellence in Innovation in 2008, the Outstanding Researcher at College of Engineering and Mineral Resources (CEMR) and WVU in 2013 and 2014, and the New Researcher of the Year at CEMR and WVU in 2010 and 2011. He was selected as the "People's Hero of the Week" by Broadband and Social Justice Blog under the Minority Media and Telecommunications Council in 2013. Two of his papers were selected as "The Best of FG'13"and "The Best of FG'15", respectively.

Xiaokang Yang (F'19) received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000. He is currently a Distinguished Professor with the School of Electronic Information and Electrical Engineering, and the Deputy Director of the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. From 2000 to 2002, he was a Research Fellow with the Centre for Signal Processing, Nanyang Technological University, Singapore. From 2002 to 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he visited the Institute for Computer Science, University of Freiburg, Freiburg im Breisgau, Germany, as an Alexander von Humboldt Research Fellow. He has authored or coauthored more than 200 refereed papers, and has filed 60 patents. His current research interests include image processing and communication, computer vision, and machine learning. Prof. Yang is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He was a Series Editor of Springer CCIS, and an Editorial Board Member of *Digital Signal Processing*. He is a member of Asia–Pacific Signal and Information Processing Association, the VSPC Technical Committee of the IEEE Circuits and Systems Society, and the MMSP Technical Committee of the IEEE Signal Processing Society. He is also the Chair of the Multimedia Big Data Interest Group of MMTC Technical Committee, IEEE Communication Society.

Xinping Guan (F'18) received the Ph.D. degree in control and systems from the Harbin Institute of Technology, Harbin, China, in 1999. He is currently a Chair Professor with Shanghai Jiao Tong University, Shanghai, China, where he is the Deputy Director of the University Research Management Office, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. From 1998 to 2008, he was a Professor and the Dean of School of Electrical Engineering, Yanshan University, Qinhuangdao, China. He has authored and/or coauthored 4 research monographs, more than 270 papers in the IEEE TRANSACTIONS and other peer-reviewed journals, and numerous conference papers. His current research interests include industrial cyberphysical systems, wireless networking and applications in smart city and smart factory, and underwater sensor networks. As a Principal Investigator, he has finished/been working on many national key projects. He is the Leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China. He is an Executive Committee Member of the Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council. He was the recipient of the First Prize of Natural Science Award from the Ministry of Education of China, in 2006 and 2016, and the Second Prize of the National Natural Science Award of China in 2008. He was the recipient of the "IEEE Transactions on Fuzzy Systems Outstanding Paper Award" in 2008. He is a "National Outstanding Youth" honored by NSF of China, "Changjiang Scholar" by the Ministry of Education of China, and "State-level Scholar" of the "New Century Bai Qianwan Talent Program" of China.

Wenjun Zhang (M'00–SM'04–F'19) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987, and 1989, respectively. After three years working as an Engineer with Philips, Nuremberg, Germany, he went back to his Alma Mater, in 1993 and became a Full Professor in electronic engineering, in 1995. As the Project Leader, he successfully developed the first Chinese HDTV prototype system, in 1998. He was one of the main contributors of the Chinese DTTB Standard (DTMB) issued, in 2006. He holds more than 76 patents and authored/coauthored more than 90 papers in international journals and conferences. He is the Chief Scientist of the Chinese Digital TV Engineering Research Centre, an industry/government consortium in DTV technology research and standardization, and the Director of Cooperative MediaNet Innovation Center, an excellence research cluster affirmed by the Chinese Government. His main research interests include digital video coding and transmission, multimedia semantic analysis, and broadcast/broadband network convergence.