Multimed Tools Appl (2018) 77:15409–15430
DOI 10.1007/s11042-017-5122-y

Color and sharpness assessment of single image dehazing

Jessica El Khoury[1] · Steven Le Moan[2] · Jean-Baptiste Thomas[1] · Alamin Mansouri[1]

Received: 2 February 2017 / Revised: 30 June 2017 / Accepted: 16 August 2017 /
Published online: 4 September 2017
© Springer Science+Business Media, LLC 2017

Abstract Image dehazing is the process of enhancing a color image of a natural scene that contains an undesirable veil of fog for visualization or as a pre-processing step for computer vision systems. In this work, we investigate the performances of eleven state-of-the-art image quality metrics in evaluating dehazed images, and discuss challenges in designing an efficient dehazing evaluation metric. This is done through a composite study based on the agreement between subjective and objective evaluations. Accordingly, we evaluate five state-of-the-art dehazing algorithms. We use two semi-indoor scenes, degraded with several levels of fog. One important aspect of these scenes is that the fog-free images are available and can therefore serve as ground-truth data for dehazing methods evaluation. This study shows that the best working dehazing method depends on the density of fog. There seems to be a clear distinction between what people perceive as good quality in terms of color restoration and in terms of sharpness restoration. Most metrics show limitations in providing proper quality prediction of dehazing. According to the introduction and analysis, a contribution of this work is to point out the flaws in the evaluation and development of dehazing methods. Our observations might be considered when designing efficient methods and metrics dedicated to image dehazing.

✉ Jessica El Khoury
jessica.el-khoury@u-bourgogne.fr

Steven Le Moan
steven.lemoan@hig.no

Jean-Baptiste Thomas
jean-baptiste.thomas@u-bourgogne.fr

Alamin Mansouri
alamin.mansouri@u-bourgogne.fr

1 LE2I Laboratory, Université de Bourgogne Franche-Comté, Dijon, France

2 The Norwegian Colour and Visual Computing Laboratory, NTNU in Gjøvik, Gjøvik, Norway

Keywords Single image dehazing · Color · Sharpness · Image quality assessment · Objective assessment · Psychometric experiment

1 Introduction

1.1 Visibility degradation model

When taking a picture in bad weather, the light emanating from the scene and reaching the camera's sensor is scattered, which forms the so-called *airlight* [23] and results in an undesirable veil that reduces contrast and chroma in the picture [34]. In this context, the image formation can be modeled, according to Koschmieder [23], as the sum of the direct transmission $J(x)t(x)$ and the airlight $A(x) = A_\infty(1 - t(x))$, weighted by a transmission factor $t(x)$:

$$I(x) = J(x)t(x) + A_\infty(1 - t(x)) \tag{1}$$

where x represents the image coordinates. I is the image formed on the camera's sensor. The transmission factor t depends on the scene depth d (distance to the sensor) and on the scattering coefficient β of the haze, such that $t = e^{-\beta.d(x)}$ [23]. Unlike other image degradation, haze is a natural, depth-dependent perturbation that spans over the whole image. The degradation and the loss of information increases with depth, as the amount of fog between the imaged surface and the sensor increases. Hazy and foggy images have also different prevailing colors, which depend on the scattering particles density and on the ambient light.

1.2 Dehazing

The process to recover J from I is known as *dehazing* or *defogging*. Although weather conditions differ in particle types, sizes, concentrations and the resulting veil hue and visibility [36], since the haze model (1) is valid for a wide range of weather conditions [35], the same methods are used indiscriminately for haze, mist, fog and other aerosols. To avoid any ambiguity, we designate this process by *dehazing* in this paper.

Dehazing is used for a variety of applications ranging from simple visualization to e.g. monitored driving [14, 50], outdoor security systems [39], or remote surveillance systems [3].

Single image dehazing methods can either be based on an inversion of (1) [23] or simply on image enhancement precepts such as contrast enhancement [1, 56]. The former category of methods rely on an estimation of the airlight and the depth from image statistics or additional information, whereas methods from the latter category are mostly based on global classical image processing methods (e.g. histogram equalization, PDE, etc.).

1.3 Motivation and contribution

A great attention has been given to dehazing research domain over the last decade or more. This is reflected by a dramatically increasing in the number of publications, which propose various dehazing approaches and/or address quality assessment of dehazed images [28]. The large variety of dehazing methods requires some common tools to be evaluated and compared to each other in a reliable way.

Related evaluations are typically carried out in a qualitative way [29, 48, 49, 62] and only a few studies have considered subjective experiments [6, 27, 31]. However, evaluation results achieved by dehazing metrics are often unreliable and do not show a satisfactory correlation with subjective judgments.

The lack of a fog-free reference image in the assessment of dehazing quality is a considerable problem, which we recently addressed by the first database of not simulated but real foggy scenes with various—controlled—levels of fog [12]. A similar dataset has been recently released [30]. It contains four synchronous and registered RGB/NIR images of indoor hazy scenes, available with the ground-truth haze-free image. However, only one haze level is provided. Thus, using such database, the investigation of the performance of both dehazing methods and quality metrics across fog levels is not possible.

In this study we consider the quality as the ability of a process to eliminate undesirable effects from a degraded image and to render it similar to the original image. This is done while considering color and sharpness restoration. The available fog-free image is considered as reference and it is assumed to have a perfect quality. Accordingly, we analyze the performance of several state-of-the-art dehazing methods, as well as the behavior of various image quality assessment (IQA) metrics in both objective and subjective terms. We single out the biased behavior that the metrics show when the assessed features change. To our knowledge, this is the first report of a user study on dehazing based on image sharpness and color assessment (as opposed to single image quality) over various known levels of real fog. A variety of image quality indices grouped into three main categories, are used and analyzed (see Table 1). We discuss how these indices correlate with human judgment. A contribution of this work is to point out the flaws in the evaluation and development of dehazing methods. Our observations might be considered when designing efficient methods and metrics dedicated to image dehazing.

In Sections 2 and 3 we focus on the categories of the used quality metrics and the selected dehazing methods. We provide in Section 4 a global description of the image database. In Section 5, we describe the components of the subjective experiment. We then evaluate in Section 6 the behavior of dehazing algorithms and image quality metrics. An objective evaluation is also provided based on the prediction of the metrics compared to the actual observer's rating. In Section 7, we discuss our statistical findings before to conclude and to suggest future work.

2 Dehazing quality assessment metrics

The quality evaluation of dehazed images is challenging, since the fog-free reference image is not available. In our recent database, in addition to real images of different fog densities, we provide the fog-free image (cf. Section 4). Hence, in addition to the metrics in reference with the foggy image and the metrics without reference image we can use some metrics in reference with the fog-free image (Table 1).

2.1 Metrics in reference with the foggy image

In this category, the dehazed images are compared to the foggy image. As soon as it shows a superiority in some features such as contrast, it is considered to be of better quality. Such metrics are usually used to evaluate dehazing methods.

Early dehazing-dedicated quality metrics are based on the measurement of structural distortions, which are not dominant in dehazed images [46]. Although the important role that non-structural features such as color degradation play in the quality assessment of dehazed images, its impact was mainly ignored. The most recent indicators take into consideration

Table 1 Image quality assessment metrics

Metric	Reference image
MS-SSIM [53]	Fog-free
VIF [44]	Fog-free
MS-iCID [26]	Fog-free
FSIMc [59]	Fog-free
VSI [61]	Fog-free
NIQE [33]	Without reference
FADE [8]	Without reference
Hautière's r [18]	Foggy
Hautière's e [18]	Foggy
CNC [17]	Foggy
CS [13]	Foggy

this aspect [8, 17], pushed by recent works that reveal the impact of color shift on perceived image quality.

Hautière et al. [18] proposed to compute the rate of new visible edges after dehazing, e, that points out the ability of the dehazing method to restore the edges that were not visible in the foggy image. They proposed also the geometric mean r, which reveals the contrast enhancement as the ratio between the gradient of the visible edges before and after dehazing.

Fang et al. [13] proposed an evaluation metric called Contrast ascension-Structural similarity (CS) that combines the ascension of contrast degree C with the structural similarity S between the hazy and the dehazed images. When C is high, the difference of image quality between the original and the processed images is great. S measures the consistency of edges before and after dehazing. The structure of both images is more similar when S is higher. This means that the dehazing method did not insert a lot of artificial edges, such as blocking artifacts. In this type of metrics, the color evaluation is completely ignored and the contrast, which is supposed to be a good indication when it is improved, may be over-enhanced and leads therefore to undesirable results.

Guo et al. [17] proposed a metric called Contrast-Naturalness-Colorfulness (CNC). $C_{contrast}$ like e, consists in evaluating image contrast by calculating the edges ratio between dehazed and hazy images. N is the image naturalness and $C_{colorfulness}$ is the image colorfulness [21]. The last two components estimate the perceived quality of the color reproductions in terms of naturalness and saturation [58]. Since over-enhancement is one of dehazing problems, like the previous metrics, this one does not judge the image enhancement regarding the preservation of original features.

2.2 Metrics without reference image

Recently, a Fog Aware Density Evaluator (FADE) [8] was proposed to evaluate the performance of dehazing algorithms through the definition of the perceptual fog density of the image. This is a no-reference metric. Since foggy images are characterized by low contrast, faint color, and shifted luminance, this metric evaluates defogged image density through the calculation by means of Mahalanobis distance the similarity with Natural Scene Statistics (NSS) and fog aware statistical features, including sharpness, contrast, image entropy, saturation in HSV color space and colorfulness. These features have been extracted from 500 images of foggy and 500 of fog-free images of different natural images databases such as

LIVE IQA database [45], the Berkeley image segmentation database [32], and copy-right free web sources like Flickr.

In addition, we used also a no-reference model for contrast-distorted images, incorporating NSS, which is called Natural Image Quality Evaluator (NIQE) [33]. This index is not attached to any specific distortion type. It predicts the overall image's naturalness.

2.3 Metrics in reference with the fog-free image

In our experiment, having the fog-free image, we consider some full-reference metrics to evaluate the dehazed images. These metrics have never been used on real foggy images.

Five full-reference objective IQA models were used. We selected Multi-Structural Similarity (MS-SSIM) [53], that predicts image quality considering contrast, structure and luminance.

Visual Information Fidelity (VIF) [44] has been used to quantify the amount of shared information between the dehazed and the reference fog-free images that the brain can extract from both images, such as signal attenuation and the noise added after enhancement. The Multi-scale improved Color-Image-Difference (MS-iCID) [26] was basically proposed to evaluate gamut mapping and tone mapping distortions. It is sensitive to many expected distortions generated by dehazing methods, such as lightness difference L_L, lightness-contrast L_C, lightness-structure L_S, chroma-difference C_L, hue-difference H_L, chroma-contrast C_C and chroma-structure C_S. It has been selected as it provides useful information on how the colors are distorted by fog.

The perceived quality degradation of a given image induces perceptible changes in the visual saliency map. In the Visual Saliency-induced Index (VSI) [61], the detection of contrast and saturation changes, which are strongly altered by fog and fog removal, is boosted by the gradient modulus map and the Scharr gradient operator, respectively. A similar metric, Feature Similarity Index (FSIMc) [59], which considers color similarity in addition, has been also used.

3 Dehazing methods

Single image dehazing methods can be split into two categories: physics-based methods and image enhancement methods. Methods in the former category are based on the visibility degradation model (1). The unknown parameters are first estimated based on prior assumptions followed by the inversion of the model and the restoration of the clear image. Methods in the latter category do not consider the physical causes of image degradation caused by weather conditions, but rely instead on observers preference to improve the quality of the image [60]. Although physics-based methods are supposed to provide more accurate recovery, they fail when the haze model is physically invalid and when the scene objects are similar to the atmospheric light [55]. For this study, we selected five methods that are representative of the state-of-the-art (see Table 2).

3.1 Physics-based methods

He et al. [19] introduced the concept of Dark Channel Prior (DCP). It relies on the assumption that, for a given pixel in a color image of a natural scene, one channel (red, green or blue) is usually very dark (it has a very low intensity). The atmospheric light A_∞ tends to brighten these dark pixels. Thus, it is estimated over the darkest pixels in the scene. A

Table 2 Summary of selected dehazing methods. For each method, the default implementations' parameters proposed by the authors, have been used

Dehazing method	Approach	Physics-based
DCP [38]	Statistics-based assumptions & filtering approach	✓
FAST [49]	Filtering approach	✓
FUSION [1]	White balance and contrast enhancement weights fusion	✓
VAR [16]	Variational approach & histogram equalization	×
CLAHE [56]	Histogram equalization	×

transmission t map is then deduced after subtracting A_∞ from the hazy image I. In order to smooth and reduce noise and artifacts in the transmission map, which is equivalent to the depth map, several approaches have been investigated. For instance, in a second version, the Laplacian matting has been replaced with a guided filter [20], which aims to refine the transmission map with a good behavior near the edges and a fast and non-approximate linear-time [38]. Although it is one of the first proposed methods, it is still considered as very effective and robust since it is based on minimal assumptions and does not consider any pre-processing step. Many other methods are based on it including some variations and improvements, [51, 52, 57]. Some other are used in different application domains like underwater and satellite image enhancement [9, 43, 54].

Unlike DCP, FAST method proposed by Tarel et al. [49] considers a pre-processing and a post processing steps: a white balancing is first performed and a tone mapping is applied after dehazing since corrected images are usually with a higher dynamic range than the original ones. The atmospheric light A_∞ is calculated as a percentage standard deviation of the local mean of the image whiteness. As in DCP, the transmission t map is then deduced from A_∞ and a new filter, which preserves edges and corners as an alternate to the median filter is then applied to reduce noise and artifacts. However, it requires to adjust various parameters, which complicate the task of users.

The fusion technique (FUSION) [1] consists of dividing the hazy image into two inputs by performing white balancing and contrast enhancement. The first one represents a remedy for the first issue one can find in hazy images, which is the color cast resulting from the airlight A. The second one handles the low contrast caused by the transmission factor t, which represents the attenuation resulting from the light scattering. In order to combine back both inputs, as it is modeled in (1) (Direct transmission + Airlight), while maintaining a great visibility, their significant characteristics are filtered through the calculation of three weight maps: luminance, chromatic and saliency. The resulting image is then smoothed by Laplacian and Gaussian pyramid to reduce artifacts introduced due to the weight maps. Although the intricacy of this method is lower than the other physics-based methods, referring to Sahu [42], the fusion based image dehazing leads to good results in a very reduced time since it does not calculate the transmission map.

3.2 Image enhancement

Galdran et al. [15] assumed that the physical model is too simple to be able to model the real situation, and it fails to handle the change of the size of atmospheric particles and the non-uniformity of the illuminant as the brightness of the atmosphere is spatially variant. They proposed, a variational framework (VAR) for image dehazing that deals with spatially-variant features. It performs contrast enhancement on hazy regions of the image throughout an iterative procedure allowing to control the degree of restoration of the visibility in the scene, while considering white world assumption adapted to haze-free images. As the gradient descent achieved as soon as the Mean Square Error between one iteration and the next one falls below 0.02, Fabien et al. [40] assume that there is no guarantee of convergence of the algorithm towards the minimum of the functional. In a further work [16], an additional energy term was added to control the degree of saturation of the processed image in order to avoid over-saturated results. This method is not dedicated to image dehazing. It can be applied for contrast enhancement, no matter the physical reason behind.

Xu [56] presented the contrast limited adaptive histogram equalization CLAHE to restore fog degraded color images. This method aims at enhancing image contrast, no matter how the related physical phenomenon is modelled. First, RGB image is converted into HSI color space. The reason to use HSI is that it represents Hue, Saturation and Intensity in separated axis, so they can be processed independently. Second, the intensity component of the image is processed by CLAHE [41]. Hue and saturation remain unchanged. Finally the processed image is converted back to RGB. We selected this method, since it simply applies a local histogram equalization based on color component.

We limited our evaluation to these five methods since they are representative of various categories and they are usually used for benchmarking.

4 Database

Instead of using synthetic foggy images [2], which fail to represent accurately the physical interaction between light and atmospheric particles [11], we used the CHIC database [7, 12], which is composed of two real semi-indoor scenes with real fog. These were basically created in order to compare dehazed images with the available original scene. The scenes were set up in a closed rectangular room, which is large enough to simulate the effect of the distance and the fog density on the objects radiance (length = 6.35 m, width = 6.29 m, height = 3.20 m, diagonal = 8.93 m) with a large window (length = 5.54 m, height = 1.5 m) that allows a large amount of outdoor light to get in, in a sunny day.

Scene 1 shows a typical indoor view (Fig. 1). It contains objects with different characteristics such as shapes, colors, positions and surface types (glossy or rough). Scene 2 contains objects with bigger geometric shapes (Fig. 2). The wall behind with the white lines represent distinctive elements to investigate the algorithms' handling near edges. The photo session of each scene lasted 20 min. The geometry of closed building and sun path have been taken into account as well as the clear weather forecast to maximize the chances that it remains steady. To create real foggy images, a fog machine (FOGBURST 1500), which emits a dense vapor, was used. A large amount of fog is initially emitted until it is evenly distributed in the room and forms an opaque layer. Fog is then progressively evacuated through the window. For each scene, the fog-free image was captured, then nine images with different densities of fog, from an opaque layer to a very light one were shot. The position of the used camera, Nikon D7100, was kept constant through shoots. Note that the density of fog of each level is not the same for scene 1 and scene 2 (Table 3), nor the distance from the scene to the camera: the distance between the camera and the back color chart is 7 m in scene 1 and 4.25 m in scene 2.

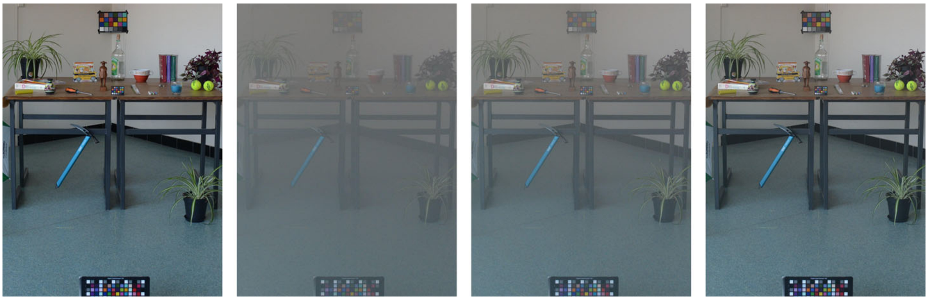


Fig. 1 Scene 1. From left to right: Fog-free; Level A (highest fog level); Level B (medium fog level); Level C (lowest fog level). Image size is 1537 × 2049. These images are successively acquired within 5 min

Considering the dehazing methods performance, among the nine foggy images from the CHIC database, three images of not very high fog densities have been processed by the selected dehazing methods (see Figs. 1, 2 and Table 3). According to Narasimhan et al. [35], the haze model is more suitable for short ranges of distance. In other words, it fails to correctly estimate the light attenuation of scenes covered by a relatively thick veil of fog. Indeed, from a certain level of fog, color image dehazing no longer produces good perceived results, so it was not relevant to select more levels with higher fog densities. Moreover, for intermediate light fog densities, the fidelity of recovered features can be considered. When the fog is basically very light, the dehazed images are usually over-contrasted and therefore more preferred than the original fog-free images (cf. Section 6.3).

Level A, level B and level C denote the highest level, the medium level, and the lowest level among the selected fog levels of scene 1 and scene 2, respectively.

In Table 3, the relative transmittance of fog is calculated with respect to airlight at a given distance over the black patch of original foggy images, as follows:

$$T = \frac{S_{level_n} - S_{airlight}}{S_{fog-free} - S_{airlight}} \tag{2}$$

Fig. 2 Scene 2. From left to right: Fog-free; Level A (highest fog level); Level B (medium fog level); Level C (lowest fog level). Image size is 1537×3073. These images are successively acquired within 5 min

Table 3 Relative transmittance T of fog in original foggy images of scene 1 and scene 2

T	Level A	Level B	Level C
Scene 1	28%	75%	84%
Scene 2	15%	54%	74%

where S_{level_n}, $S_{airlight}$ and $S_{fog-free}$ are the spectral values of green in images of different fog levels, of the airlight image of our database where the scene is completely covered by fog and the fog-free image.

5 Experiment

Objective The goal is to evaluate how metrics correlate with the visual judgments while considering color and sharpness recovery. Accordingly, the selected algorithms are assessed with the best performing metrics. This is mainly done to find out the dehazing method that succeeds to remove the fog while preserving color and sharpness.

In this work, we consider the similarity between the fog-free and the processed images according to the most perceived criteria. The sharpness similarity that points at the same time to the ability to remove fog and the resulting amount of noise and artifacts. The color similarity that points to the ability to estimate accurately the airlight and the ability to find an acceptable compromise between the fog removal and the naturalness of images.

Observers A group of 20 subjects, men and women, has been asked to rank images. The judgments of only 14 observers have been retained. Their age range from 20 to 40. They were students and researchers at Gjøvik University College. Most of them were familiar with images quality features.

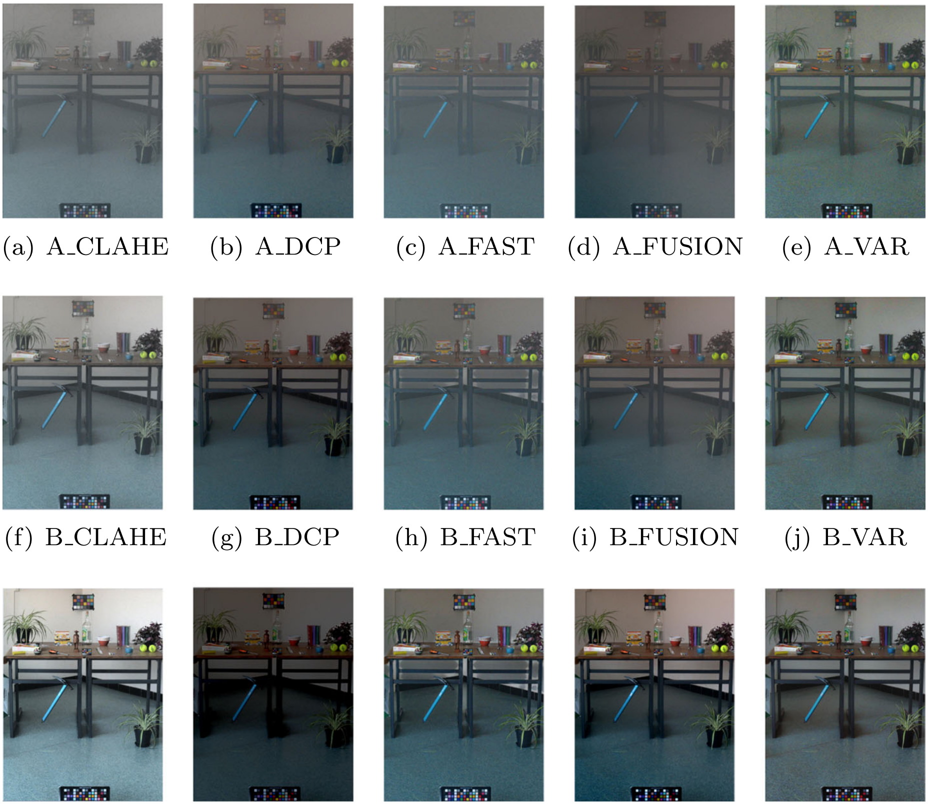
Stimuli The images of fog level A, B and C of scene 1 and scene 2 were used. The five algorithms were applied to all of these images (dehazed images are shown in Figs. 3 and 4). Thus, in the subjective evaluation, for each fog level of a scene, the five dehazed images are compared to the correspondent fog-free image in terms of color and in terms of sharpness similarities.

Thus, the observers examined 36 images for each feature. Indeed, for the same fog level, 6 images (5 dehazed images and the fog-free image) of scene 1 were displayed followed by 6 images of scene 2.

Procedure At first, we had to make clear to participants the nature of the images, the applied processing as well as the objective and the steps of this procedure.

The experiments were run in a dark room. Observers were seated approximately at 70 cm from the display. For each fog density, observers were asked to rank the dehazed images, considering the degree of matching with the fog-free image in terms of color and in terms of sharpness, respectively. For each feature, we displayed 6 sets of 6 images, starting from level A to level C of scene 1 and scene 2, respectively.

This experiment has been done using the well-established comparison process Rank Order Correlation. The images were displayed using *Quickeval* [37], a web-based tool for psychometric image evaluation. Observers were allowed to move the image to a zoomed-in view, and to focus only on a portion of the displayed scene at a time. In this experiment, we



(a) A_CLAHE (b) A_DCP (c) A_FAST (d) A_FUSION (e) A_VAR

(f) B_CLAHE (g) B_DCP (h) B_FAST (i) B_FUSION (j) B_VAR

(k) C_CLAHE (l) C_DCP (m) C_FAST (n) C_FUSION (o) C_VAR

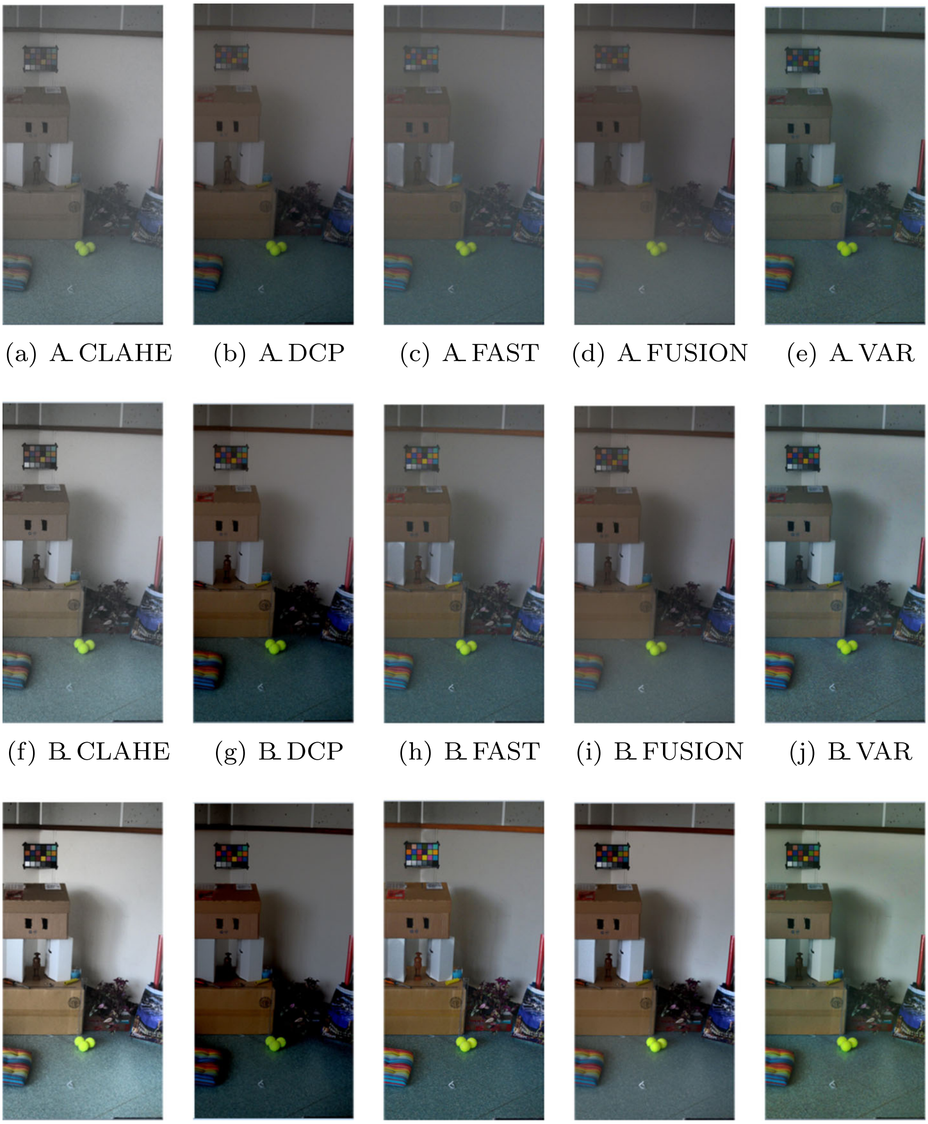
Fig. 3 The foggy images of scene 1 processed by the selected dehazing methods at fog levels A, B and C

did not impose any constraints related to the time an observer could spend on the task, nor to the portion of scene that was to be considered. The overall evaluation time of participants varied from 35 to 45 min.

6 Evaluation

6.1 Subjective evaluation of dehazing algorithms

We first performed a screening of the 20 participants according to [22] and found that 6 of them were not valid. Results from these outlying observers were discarded. Overall, we noted a relatively large inter-observer variability (see error bars on Fig. 5), which we assume can be explained by several factors. First, the complexity of the task: even though we asked the observers to isolate the influence of color and that of sharpness, we believe that the extent to which they were able to do so varied greatly from one observer to another. Secondly, the software that we used allowed observers to focus on a portion of a scene. Although we did not record how each observer browsed the scenes, we argue that there might have been some variations in terms of what they considered as regions of interest. In other words,



(a) A_CLAHE (b) A_DCP (c) A_FAST (d) A_FUSION (e) A_VAR

(f) B_CLAHE (g) B_DCP (h) B_FAST (i) B_FUSION (j) B_VAR

(k) C_CLAHE (l) C_DCP (m) C_FAST (n) C_FUSION (o) C_VAR

Fig. 4 The foggy images of scene 2 processed by the selected dehazing methods at fog levels A, B and C

different observers chose to focus on different parts of the scene, and that behavior cannot be modeled by saliency only.

We computed the average ranks, over all 14 observers, given to each dehazing method in each sequence and obtained the results shown in Fig. 5. Again, the large error bars suggest a large inter-observer variability. However in a few cases, observers tend to agree to a relatively large extent. For example, according to these results, the CLAHE method performs significantly worse than the other four in terms of recovering colors on the second scene, for the fog level A. Observers also agreed that this method performs poorly on the

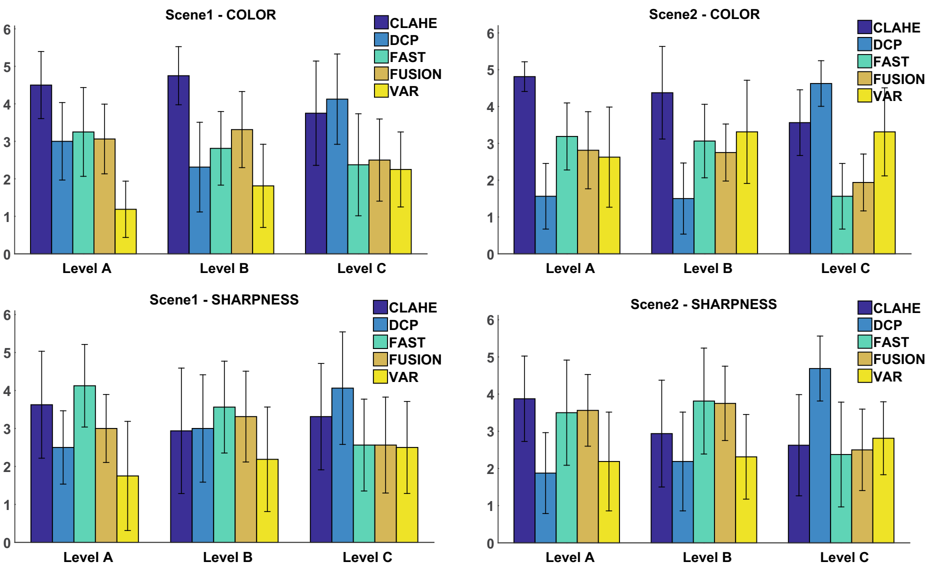


Fig. 5 Average ranks obtained for each image sequence, and for each level of haze. First row: color, second row: sharpness. First column: scene 1, second column: scene 2

first scene for level B and that the DCP method performs badly in terms of both color and sharpness recovery in scene 2 for level C. We can also see that VAR performs best in terms of recovering color on the first scene and for level A.

Overall, we found that inter-observer variability is larger for the sharpness recovery assessment (second row in Fig. 5) than for that of color recovery by 19%. This is consistent with the feedback we obtained from most observers about the higher difficulty of the second part of the experiment.

Note that these results, i.e. large inter-observer variability and lack of consistency across scenes, have also been reported in [31]. Here we demonstrate in addition, and in a different framework where the observers had access to the original image, that there seems to be no consistency either across fog levels and between color and sharpness preservation. Not a single method performs the best not the worst overall.

6.2 Subjective evaluation of image quality metrics

We chose to use the *hit rate* to measure the agreement between subjective and objective data. It corresponds to the proportion of correct predictions made by the metric in a pairwise comparison setup and takes into account both intra- and inter-observer variability. A hit rate h has a value between 0 and 1, although it does have an upper bound in that it is impossible to obtain a perfect agreement with all observers. There is a maximal achievable hit rate $h_{\max}$, also referred to as *majority hit rate* [26], which is defined as the rate that would be obtained should the metric constantly agree with the majority of the observers. Additionally, h also has a pseudo lower bound in that a hit rate of 0.5 is considered as that of chance. Consequently, we also computed the *rescaled* hit rate $\hat{h}$, as follows:

$$\hat{h} = \frac{h - 0.5}{h_{\max} - 0.5} \tag{3}$$

Table 4 Hit rates obtained by each metric: (rescaled rates in brackets)

		COLOR	SHARPNESS
FG	MS-SSIM	0.50 (0.00)	0.60 (0.50)
	VIF	0.48 (−0.08)	0.60 (0.49)
FC	MS-iCID	0.46 (−0.14)	0.52 (0.10)
	FSIMc	0.52 (0.07)	0.62 (0.63)
	VSI	0.58 (0.28)	0.63 (**0.67**)
N	NIQE	0.63 (0.43)	0.58 (0.39)
	FADE	0.63 (**0.45**)	0.55 (0.27)
HG	Hautière's *r*	0.48 (−0.06)	0.59 (0.44)
	Hautière's *e*	0.60 (0.36)	0.57 (0.38)
	CS	0.56 (0.22)	0.55 (0.26)
HC	CNC	0.60 (0.34)	0.57 (0.35)

The table is divided in four parts: metrics in reference with the fog-free greyscale image **(FG)**, metrics in reference with the fog-free color image **(FC)**, no-reference metrics **(N)**, metrics in reference with the foggy greyscale image **(HG)** and metrics in reference with the foggy color image **(HC)**

and which ranges between -1 (worst) and $+1$ (best). Note that negative values can be considered as worse than chance (i.e. worse than a random predictor). The subjective ranking data was therefore transformed in pairwise comparison preferences for each image sequence independently and objective preferences were computed with the benchmark metrics.

Table 4 gives the obtained hit rates for each metric of the benchmark, for each level of haze and overall. Additionally, we isolated individual features of the MS-iCID and CNC metrics in order to have a better insight in terms of which perceptual attributes contribute the most to a good dehazed image quality. Table 5 reports the results obtained. Note that, for the sake of readability, all values were rounded, which explains that some seemingly identical h lead to different $\hat{h}$. Note also that $h_{\max}$ is different in each part of the experiment.

In order to evaluate if the difference between two hit rates is significant, we also performed a two-sample binomial test with Yule's confidence intervals [4]. We assumed that the metrics' predictions of observer choices can be modeled as binomial distributions. The test assesses whether or not two scores are likely to come from the same distribution. If yes, they are not significantly different. We found that, for color recovery assessment, MS-SSIM, VIF, r and MS-iCID perform significantly worse than CS, VSI, e, CNC, NIQE and FADE, while NIQE and FADE perform significantly better than CS and FSIMc. As for sharpness recovery assessment, VSI performs significantly better than r, e, CNC, NIQE, FADE and MS-iCID.

These results suggest that:

– State-of-the-art image quality metrics demonstrate only a limited accuracy on dehazing color and sharpness assessment (the maximal hit rate obtained with our benchmark is $\hat{h} = 0.67$).
– Traditional metrics in reference with the fog-free image perform better at assessing sharpness than color recovery. Figure 6 shows a case of total disagreement between these metrics and observers, implying in particular that chromatic attributes such as chroma and hue are especially important in dehazing quality assessment, compared to more conventional quality assessment. Although we strove to isolate the influence of

Table 5 Hit rates obtained by individual perceptual attributes: COLOR (rescaled rates in brackets)

	Color	Sharpness
MS-iCID: $\mathbf{L}_L$	0.46 (−0.14)	0.56 (0.31)
MS-iCID: $\mathbf{L}_C$	0.54 (0.12)	0.61 (**0.57**)
MS-iCID: $\mathbf{L}_S$	0.52 (0.06)	0.49 (−0.02)
MS-iCID: $\mathbf{C}_L$	0.69 (**0.64**)	0.57 (0.37)
MS-iCID: $\mathbf{H}_L$	0.42 (−0.28)	0.45 (−0.26)
MS-iCID: $\mathbf{C}_C$	0.64 (0.48)	0.51 (0.06)
MS-iCID: $\mathbf{C}_S$	0.41 (−0.33)	0.47 (−0.16)
CNC: contrast	0.60 (0.36)	0.57 (0.38)
CNC: colorfulness	0.57 (0.25)	0.52 (0.12)
CNC: naturalness	0.53 (0.09)	0.52 (0.12)

The MS-iCID attributes are as follows (refer to [25] for details): $\mathbf{L}_L$: lightness-difference, $\mathbf{L}_C$: lightness-contrast, $\mathbf{L}_S$: lightness-structure, $\mathbf{C}_L$: chroma-difference, $\mathbf{H}_L$: hue-difference, $\mathbf{C}_C$: chroma-contrast, and $\mathbf{C}_S$: chroma-structure

color and sharpness in our experiment, a recent study by Le Moan et al. [24] suggested that, given no particular instructions on which kind of distortions to focus on, observers tend to judge achromatic ones such as JPEG artifacts, less severely than chromatic ones such as hue shift. Consequently, the rankings obtained for color recovery assessment may be somewhat more meaningful than that of sharpness for the *overall* quality assessment case, thus emphasizing the need for traditional metrics in reference with the fog-free image to give more weight to chromatic distortions when used for dehazing quality assessment.

– Metrics that are proposed or used for dehazing evaluation (FADE, r, e, CNC and CS) can be outperformed by metrics in reference with the fog-free image. In particular, VSI outperforms them all significantly when it comes to sharpness recovery assessment.
– NIQE and FADE, which are both based on natural scene statistics (NSS), perform particularly well at assessing color recovery assessment. This supports the conclusions obtained in [31]

From the results in Table 5, we observe that the perceptual attributes that correlate the best with the subjective data are the MS-iCID $\mathbf{C}_L$ (chroma-difference) and $\mathbf{C}_C$ (chroma-contrast) terms for color, and $\mathbf{L}_C$ (lightness-contrast) for sharpness. A significance analysis reveals that $\mathbf{C}_L$ performs significantly better than all other attributes for color, whereas for sharpness, $\mathbf{L}_C$ performs significantly better than all other attributes except $\mathbf{C}_L$ and CNC's contrast term. This implies that attributes such as hue and lightness-structure have a relatively small importance in dehazing quality assessment. Even chroma-structure seems to be irrelevant in this context. Finally, it is interesting to note that some of these features perform better individually than together: the MS-iCID $\mathbf{C}_L$ term performs significantly better than the metric itself, but also than FADE and NIQE, thus making $\mathbf{C}_L$ the best metric in our benchmark to estimate color restoration quality in dehazing.

6.3 Objective evaluation of dehazing algorithms

We first calculated the raw scores of all selected metrics for all dehazed images of scene 1 and scene 2 sorted according to fog levels (see Figs. 7, 8 and 9). Overall, metrics that

Fig. 6 Case of total disagreement between metrics in reference with the fog-free image and observers for color recovery quality assessment (scene 1, level B). From left to right: result from CLAHE, original image and result from FAST. While all metrics rate the CLAHE dehazed image as more fidel to the original, all observers agreed that FAST is better in terms of color. Indeed, the FAST image seems more vivid, although it clearly shows artifacts just under the tables and around the plants. When it comes to evaluating the structure however, a majority of observers (9 out of 14) find the CLAHE image more fidel. This suggests that these metrics in our benchmark do not perform well at balancing the importance of chromatic and achromatic attributes for dehazing quality assessment

belong to the same category have monotonic trends across fog levels. Note that the metrics in reference with the fog-free image, which measure the similarity (fidelity) between test and reference images, show that the fidelity of an image's features is better maintained at low fog densities (see Fig. 7). This means that dehazing methods are globally more efficient at these levels.

In order to evaluate methods by metrics in reference with the foggy image, we calculated the score of the fog-free image considered as the optimal one. Amongst these metrics, the scores of r and e which refer to the gain of visibility, are higher at high fog densities. This denotes that a good range of improvement is done at high fog densities, thus, the foggy and the recovered image are less similar. In other words, the less similarity would refer to a good improvement in quality. However, it is not always true. For the same fog level in Fig. 8, the gain of visibility of corrected images is higher than the one of fog-free image. This does not agree with the perceptual judgment. Thus, the highest dissimilarity does not necessarily mean the best quality, but rather underlines the artifacts and the edges due to noise, which is likely to occur after dehazing process. These indices that have been widely used to assess new dehazing methods are inconsistent with the human visual evaluation of color, which was evoked in [17] and proved again in Table 4. Indeed, in such metrics, the apparent problems of over-enhancement and color distortion are overlooked. They mainly focus on the recovered structures that may be an amplified noise and blocking artifacts [31]. The same analysis can be made for CS metric.

Besides the structural distortions, which are mainly considered by the previous metrics of the same category, CNC index claims to correlate the color effect resulting from dehazing with the perceptual judgment. However, according to [47], the colors of natural scenes may vary with environmental conditions such as ambient light. Moreover, the color naturalness based on skin, grass and sky colors do not represent the whole natural colors. Thus, CNC shows the same behavior like the structure evaluating metrics (r, e and CS) without great improvement in color assessment (see Tables 4 and 5).

Considering no-reference metrics, NIQE and FADE, they are both able to quantify the evolution of dehazed image quality. However, FADE seems to be more sensitive to over-saturated colors, which usually occur when the fog is very light (level C). At this level, the

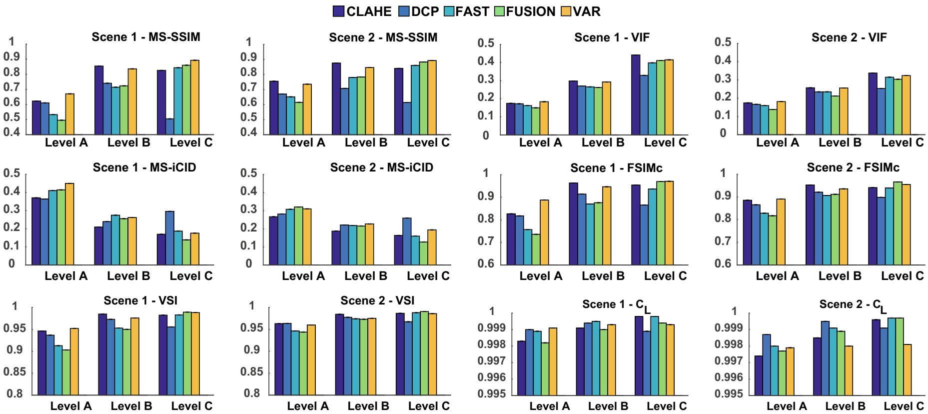


Fig. 7 The scores of the metrics in reference with the fog-free image, which are obtained for each scene at each fog level. These five metrics (MS-SSIM, VIF, MS-iCID, FSIMc and VSI) evaluate the quality of the dehazed image as being the similarity rate of specified features between the foggy and the fog-free images. For light fog, all of these metrics show the high similarity rate compared to higher level of haze. Some of them such as MS-SSIM, FSIMc, VSI and NIQE show that DCP does not perform well in the lowest fog level. The subjective analysis shows that C_L and VSI are the most relevant metrics to evaluate color and sharpness. Both of these metrics' values of different methods are noticeably close to each other

processed images seem to be of higher quality than the original fog-free image. However, higher contrast and saturation, which are mainly considered by FADE, do not necessarily denote a lower fog density [47]. Thus, such metric, which assesses better color than sharpness, seems to underline the preference rather than fidelity (see Table 4). Now, we only consider C_L and VSI that seem to be statistically the most adequate metrics to judge respectively color and sharpness restoration. These two metrics denote a good recovery of color and sharpness at low fog densities (see Fig. 7). At the highest fog level of A, VAR followed by CLAHE and DCP perform the best in sharpness recovery. For color recovery, while VAR and DCP maintain a good performance in addition to FAST, CLAHE seems to be significantly bad. At the medium level of B, all methods show a similar performance in sharpness

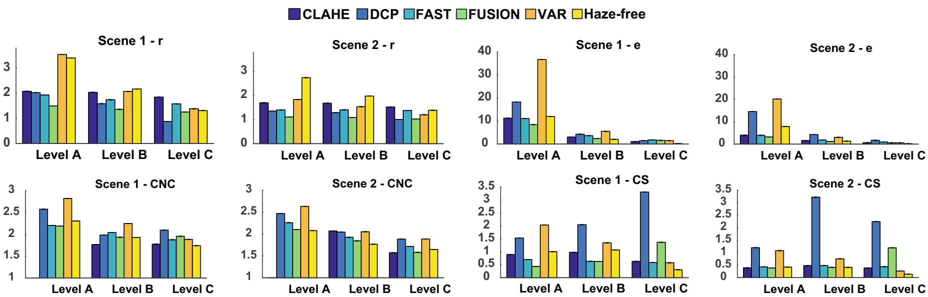


Fig. 8 The scores of the metrics in reference with the foggy image, which are obtained for each scene at each fog level. These four metrics (r, e, CNC, CS) compare the dehazed image to the foggy image. The dissimilarity is small for low levels of fog. The values of metrics on different dehazing methods are noticeably close to each other. e, CS and CNC scores exceed, across levels, the fog-free image scores. This is because appearing artifacts are accounted for edges in the resulting image

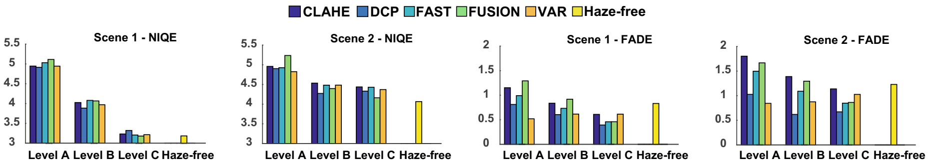


Fig. 9 No-reference NSS-based metrics scores obtained for each scene at each fog level. NIQE is a non haze-dedicated metric. According to it, the image of the highest quality is the fog-free image. Dehazed images, which were basically covered by a light fog, have almost the same quality of the fog-free image. While for FADE, which is a haze-dedicated metric, some methods across the three fog levels beat the fog-free image's score. Indeed, it presents a weakness in evaluating the preservation of image's features comparing to the fog-free image

recovery, with a significant advancement of FAST and FUSION and a significant regress of DCP's performance comparing to others. In case of color, CLAHE remains the bad performing method and VAR seems to be bad in preserving color comparing to other methods. At the lowest fog level of C, all methods are globally efficient in sharpness recovery, except DCP, which fails as well in color recovery with VAR, while the other methods are more suitable.

In all of this, the physics-based dehazing method DCP based on minimum strong assumptions and the variational approach of contrast enhancement VAR are the most suitable for relatively high fog density. On the other hand, FAST and FUSION the physics-based dehazing method which perform white balance and contrast adjustment, perform well for thin layer of fog. Otherwise, the dehazed image remain hazy. CLAHE, which is a histogram equalization method, has a particular behavior. At all considered fog levels CLAHE has the ability to preserve sharpness. However, it widely fails to recover original colors.

7 Discussion

The application of fog and fog removal induces several types of distortions, such as noise, color fade, contrast degradation and artifacts near the edges, which are spatially variant and spatially correlated to imaged objects: the accuracy of colors recovery depends largely on initial objects colors and on their distances to the camera [5]. Thus, an adequate full-scale metric is required.

Although there has lately been a considerable effort to overcome the limitations of existing metrics used for dehazing evaluation, over-saturation and color distortion still can not be well identified [10, 47].

According to the statistical significance, it seems that in color-based metrics, image structure evaluating features are more weighted than color evaluating features. The individual feature C_L, which calculates the chromatic difference in CIELAB color space [26], is significantly better than all other metrics in color preservation judgment. VSI, which was basically detecting perceptual quality distortions as changes in visual saliency map, was not sensitive to contrast change. It was then boosted through the insertion of the Scharr gradient operator similarity map [61].

The clue for mastering the quality assessment of dehazed images is to investigate and quantify the impact an image feature distortion may create in an image category. In particular, color and sharpness, which are differently weighted within metrics and poorly considered in metrics that have been always used to evaluate dehazing.

8 Conclusion and future work

Quality assessment of single image dehazing is a challenging task. According to the target application, quality is differently considered. Although single image dehazing has been an active field for several years, not much work has been dedicated to subjective and objective quality assessment. This is mainly because original fog-free image, which is assumed to have the highest visual quality is not easily available.

Using a new foggy image database that contains scenes degraded with several levels of fog with their fog-free images, we conduct a composite study based on the agreement between subjective and objective dehazing evaluation. We discuss how color and shapness restoration are handled by dehazing methods and assessed by quality metrics.

We single out the clear distinction between what people perceive as good quality in terms of color restoration and in terms of sharpness fidelity across the fog levels. The performance of dehazing methods changes across fog levels. Image enhancement techniques except variational approach, are only suitable for recovery at thin fog. Otherwise, the colors are badly recovered.

Based on our observations, for natural foggy images with no reference image, metrics without reference image would be adequate for dehazing if they were designed to handle adequately the distortions that may be introduced and amplified differently through dehazing processing, while respecting the improvement's boundaries.

In order to support the strong conclusions, some technical points will be addressed in the near future. Further research direction would be the validation of these contributions on other similar indoor scenes with reference images and natural foggy images. A complementary work would focus on the investigation of how regions of interest are defined in the image and how these change between observers and across fog levels.

Acknowledgements The authors thank the Open Food System project as well as the National Research Council of Norway for funding. Open Food System is a research project supported by Vitagora, Cap Digital, Imaginove, Aquimer, Microtechnique and Agrimip, funded by the French State and the Franche-Comté Region as part of The Investments for the Future Programme managed by Bpifrance, www.openfoodsystem.fr.

References

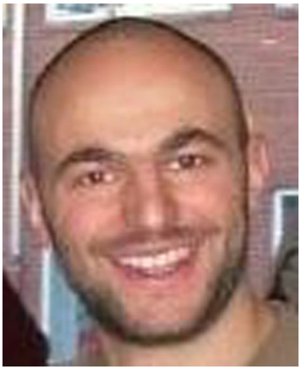
1. Ancuti CO, Ancuti C (2013) Single image dehazing by multi-scale fusion. IEEE Trans Image Process 22(8):3271–3282
2. Ancuti C, Ancuti CO, De Vleeschouwer C (2016) D-hazy: a dataset to evaluate quantitatively dehazing algorithms. In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp 2226–2230
3. Anitharani M (2013) Haze removal of secure remote surveillance system. IOSR J Eng 3:10–17
4. Brown L, Li X (2005) Confidence intervals for two sample binomial distribution. J Stat Plan Inference 130(1):359–375
5. Chandler DM (2013) Seven challenges in image quality assessment: past, present, and future research. ISRN Signal Processing 2013
6. Chen Z, Jiang T, Tian Y (2014) Quality assessment for comparing image enhancement algorithms. In: 2014 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 3003–3010
7. CHIC (Color Hazy Image for Comparison). http://chic.u-bourgogne.fr
8. Choi LK, You J, Bovik AC (2015) Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Trans Image Process 24(11):3888–3901
9. Drews P, do Nascimento E, Moraes F, Botelho S, Campos M (2013) Transmission estimation in underwater single images. In: 2013 IEEE international conference on computer vision workshops (ICCVW). IEEE, pp 825–830

10. El Khoury J, Thomas J-B, Mansouri A (2014) Does dehazing model preserve color information? In: 2014 tenth international conference on signal-image technology and internet-based systems (SITIS). IEEE, pp 606–613
11. El Khoury J, Thomas J-B, Mansouri A (2015) Haze and convergence models: experimental comparison. In: AIC 2015
12. El Khoury J, Thomas J-B, Mansouri A (2016) A color image database for haze model and dehazing methods evaluation. In: International conference on image and signal processing. Springer, pp 109–117
13. Fang S, Yang J, Zhan J, Yuan H, Rao R (2011) Image quality assessment on image haze removal. In: Control and decision conference (CCDC), 2011 Chinese. IEEE, pp 610–614
14. Fleyeh H, Dougherty M (2005) Road and traffic sign detection and recognition. In: Proceedings of the 16th Mini-EURO conference and 10th meeting of EWGT, pp 644–653
15. Galdran A, Vazquez-Corral J, Pardo D, Bertalmío M (2014) A variational framework for single image dehazing. In: Computer vision-ECCV 2014 workshops. Springer, pp 259–270
16. Galdran A, Vazquez-Corral J, Pardo D, Bertalmío M (2015) Enhanced variational image dehazing. Appl Opt 27:25
17. Guo F, Tang J, Cai Z (2014) Objective measurement for image defogging algorithms. J Cent South Univ 21:272–286
18. Hautière N, Tarel J-P, Aubert D, Dumont E et al (2008) Blind contrast enhancement assessment by gradient ratioing at visible edges. Image Anal Stereol J 27(2):87–95
19. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353
20. He K, Sun J, Tang X (2013) Guided image filtering. IEEE Trans Pattern Anal Mach Intell 35(6):1397–1409
21. Huang K-Q, Wang Q, Wu Z-Y (2006) Natural color image enhancement and evaluation algorithm based on human visual system. Comput Vis Image Underst 103(1):52–63
22. ITU-R BT.500-12. Recommendation: methodology for the subjective assessment of the quality of television pictures (1993)
23. Koschmieder H (1925) Theorie der horizontalen Sichtweite: Kontrast und Sichtweite, Keim & Nemnich, Munich
24. Le Moan S, Urban P (2014) Image-difference prediction: from color to spectral. IEEE Trans Image Process 23(5):2058–2068
25. Le Moan S, Preiss J, Urban P (2015) Evaluating the multi-Scale iCID metric. In: Larabi M-C, Triantaphillidou S (eds) Image quality and system performance XII, vol 9396, San Francisco, February. SPIE, pp 9096–38
26. Lissner I, Preiss J, Urban P, Lichtenauer MS, Zolliker P (2013) Image-difference prediction: from grayscale to color. IEEE Trans Image Process 22(2):435–446
27. Liu X, Hardeberg JY (2013) Fog removal algorithms: survey and perceptual evaluation. In: 2013 4th European workshop on visual information processing (EUVIP). IEEE, pp 118–123
28. Liu S, Rahman MA, Wong CY, Lin SCF, Jiang G, Kwok N (2015) Dark channel prior based image dehazing: a review. In: 2015 5th international conference on information science and technology (ICIST). IEEE, pp 345–350
29. Lu H, Li Y, Nakashima S, Serikawa S (2016) Single image dehazing through improved atmospheric light estimation. Multimed Tools Appl 75(24):17081–17096
30. Lüthen J, Wörmann J, Kleinsteuber M, Johannes SA (2017) A rgb/nir data set for evaluating dehazing algorithms. Electron Imaging 2017(12):79–87
31. Ma K, Liu W, Wang Z (2015) Perceptual evaluation of single image dehazing algorithms. In: IEEE international conference on image processing
32. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Eighth IEEE international conference on computer vision, 2001. ICCV 2001. Proceedings, vol 2. IEEE, pp 416–423
33. Mittal A, Soundararajan R, Bovik AC (2013) Making a completely blind image quality analyser. IEEE Signal Process Lett 22:209–212
34. Narasimhan SG, Nayar SK (2000) Chromatic framework for vision in bad weather. In: IEEE conference on computer vision and pattern recognition, 2000. Proceedings, vol 1. IEEE, pp 598–605
35. Narasimhan SG, Nayar SK (2003) Contrast restoration of weather degraded images. IEEE Trans Pattern Anal Mach Intell 25(6):713–724
36. Nayar SK, Narasimhan SG (1999) Chromatic framework for vision in bad weather. In: The proceedings of the seventh IEEE international conference on computer vision, 1999, vol 2. IEEE, pp 820–827

37. Ngo KV, Storvik JJ, Dokkeberg CA, Farup I, Pedersen M (2015) Quickeval: a web application for psychometric scaling experiments. In: IS&T/SPIE electronic imaging. International Society for Optics and Photonics, pp 93960O–93960O
38. Pang J, Au OC, Guo Z (2011) Improved single image dehazing using guided filter. In: Asia-Pacific signal and information processing association annual summit and conference 2011
39. Pettersson N (2013) Gpu-accelerated real-time surveillance de-weathering
40. Pierre F, Migerditichan P (2015) Débrumage variationnel. In: GRETSI
41. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, Romeny BH, Zimmerman JB, Zuiderveld K (1987) Adaptive histogram equalization and its variations. Comput Vis Graph Image Process 39(3):355–368
42. Sahu C, Sahu R (2014) Comparative study on fusion based image dehazing. Int J Adv Res Comput Commun Eng 3:7057–7060
43. Sathya R, Bharathi M, Dhivyasri G (2015) Underwater image enhancement by dark channel prior. In: 2015 2nd international conference on electronics and communication systems (ICECS). IEEE, pp 1119–1123
44. Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE Trans Image Process 15(2):430–444
45. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Process 15(11):3440–3451
46. Shen W, Hao S, Qian J, Li L (2017) Blind quality assessment of dehazed images by analyzing information, contrast, and luminance. J Netw Intell 2(1):139–146
47. Song Y, Luo H, Lu R, Ma J (2017) Dehazed image quality assessment by haze-line theory. J Phys: Conf Ser 844(1):012045
48. Sun W (2013) A new single-image fog removal algorithm based on physical model. Optik-Int J Light Electron Opt 124(21):4770–4775
49. Tarel J-P, Hautière N (2009) Fast visibility restoration from a single color or gray level image. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 2201–2208
50. Tarel J-P, Hautière N, Caraffa L, Cord A, Halmaoui H, Gruyer D (2012) Vision enhancement in homogeneous and heterogeneous fog. IEEE Intell Transp Syst Mag 4(2):6–20
51. Ullah E, Nawaz R, Iqbal J (2013) Single image haze removal using improved dark channel prior. In: 2013 Proceedings of international conference on modelling, identification & control (ICMIC). IEEE, pp 245–248
52. Wang Y, Wu B (2010) Improved single image dehazing using dark channel prior. In: 2010 IEEE international conference on intelligent computing and intelligent systems (ICIS), vol 2. IEEE, pp 789–792
53. Wang Z, Simoncelli EP, Bovik AC (2003) Multiscale structural similarity for image quality assessment. In: Conference record of the thirty-seventh Asilomar conference on signals, systems and computers, 2004, vol 2. IEEE, pp 1398–1402
54. Wang L, Xie W, Pei J (2015) Patch-based dark channel prior dehazing for RS multi-spectral image. Chin J Electron 24(3):573–578
55. Wang W, Yuan X, Wu X, Liu Y (2017) Dehazing for images with large sky region. Neurocomputing 238:365–376
56. Xu Z, Liu X, Ji N (2009) Fog removal from color images using contrast limited adaptive histogram equalization. In: 2nd international congress on image and signal processing, 2009. CISP'09. IEEE, pp 1–5
57. Xu H, Guo J, Liu Q, Ye L (2012) Fast image dehazing using improved dark channel prior. In: 2012 international conference on information science and technology (ICIST). IEEE, pp 663–667
58. Yendrikhovski SN, Blommaert FJJ, de Ridder H (1998) Perceptually optimal color reproduction. In: Photonics West'98 electronic imaging. International Society for Optics and Photonics, pp 274–281
59. Zhang L, Zhang L, Mou X, Zhang D (2011) FSIM: a feature similarity index for image quality assessment. IEEE Trans Image Process 20(8):2378–2386
60. Zhang H, Liu Q, Wu Y, Yang F (2013) Single image dehazing combining physics model based and non-physics model based methods. J Comput Inf Syst 9(4):1623–1631
61. Zhang L, Shen Y, Li H (2014) VSI: a visual saliency-induced index for perceptual image quality assessment. IEEE Trans Image Process 23(10):4270–4281
62. Zhu Q, Mai J, Shao L (2014) Single image dehazing using color attenuation prior. In: Proceedings of the BMVC

Jessica El Khoury received her Master in Image Processing in 2013 from the Université de Bourgogne in France and her Master Telecommunications Engineering from the Université Antonine in Lebanon. She received her PhD in Computer and Imaging Science in 2016 from the Université de Bourgogne Franche-Comté. Her research interests include color and multispectral imaging in bad conditions.

Steven Le Moan received his Master's and Engineer degrees in Signal Processing from the University of Rennes 1, France, in 2009 and his PhD degree in Computer and Imaging Science from the University of Burgundy, France, in 2012. He was then enrolled as a postdoctoral researcher at the University of Technology in Darmstadt, Germany, and at the Norwegian Colour and Visual Computing Laboratory at NTNU in Gjøvik, Norway. He is currently a Lecturer at Massey University in Palmerston North, New Zealand. His research is mainly focused on colour and multispectral image processing.



Jean-Baptiste Thomas received his Bachelor in Applied Physics in 2004 and his Master in Optics, Image and Vision in 2006, both from the Université Jean Monnet in France. He received his PhD from the Université de Bourgogne in 2009. After a stay as researcher at the Gjøvik University College and then at the C2RMF (Centre de Recherche et de Restauration des Musées de France), he is now Maître de Conférences (Associate Professor) at the Université de Bourgogne. His research focuses on color science and on color multi-spectral imaging.

Alamin Mansouri received his PhD from the University of Bourgogne, France in 2005. Afterwards, he was postdoctoral fellow at the ColorLab laboratory in Norway. He is now full professor at the university of Bourgogne Franche-Comté and member of the Laboratory LE2I (Laboratoire d'Electronique, Informatique et Image) where he is leading the thematic project of Physical and Multimodal Imaging. His current research is focused on multispectral imaging (both sides acquisition and processing/analysis) and its interactions with cultural heritage and media technology.