Multi-scale Optimal Fusion model for single image dehazing

Dong Zhao, Long Xu, Yihua Yan, Jie Chen, Ling-Yu Duan

Abstract

Image acquisition is usually vulnerable to bad weathers, like haze, fog and smoke. Haze removal, namely dehazing has always been a great challenge in many fields. This paper proposes an efficient and fast dehazing algorithm for addressing transmission map misestimation and oversaturation commonly happening in dehazing. We discover that the transmission map is commonly misestimated around the edges where grayscale change abruptly. These Transmission MisEstimated (TME) edges further result in halo artifacts in patch-wise dehazing. Although pixel-wise method is free from halo artifacts, it has trouble with oversaturation. Therefore, we firstly propose a TME recognition method to distinguish TME and non-TME regions. Secondly, we propose a Multi-scale Optimal Fusion (MOF) model to fuse pixel-wise and patch-wise transmission maps optimally to avoid misestimated transmission region. This MOF is then embedded into patch-wise dehazing to suppress halo artifacts. Furthermore, we provide two post-processing methods to improve robustness and reduce computational complexity of the MOF. Extensive experimental results demonstrate that, the MOF can achieve additional improvement beyond the prototypes of the benchmarks; in addition, the MOF embedded dehazing algorithm outperforms most of the state-of-the-arts in single image dehazing.

Keywords:

Image dehazing, Multi-scale dehazing, Dark channel prior, Multi-scale Optimal Fusion

1. Introduction

Haze is an atmospheric phenomenon caused by the absorption or reflection of light by floating particles in the air. The scene radiance acquired in these kinds of bad weather conditions is dramatically attenuated and contaminated because of atmospheric scattering. The captured images often tend to suffer from faint, e.g. image contrast, content detail and color saturation progressively worsen with the increased scene distance. In practice, the atmosphere is not absolutely free of any particles, so the haze still exists even in good weather conditions. The degradation caused by haze would harm image visual quality perceived on one hand. On the other hand, it would lower the efficiency of following image processing and computer vision tasks, such as feature extraction, recognition and classification [1–3]. Therefore, haze removal or dehazing is necessary before performing following vision tasks in applications, including outdoor surveillance, traffic monitoring, air navigation and remote sensing.

Early approaches often require certain types of additional information. For instance, Polarized filters are used to capture multiple images of the same scene in [4,5], and then different degrees of polarization (DOP) of images are used for haze removal. Nayar and Narasimhan [6] also capture multiple images of the same scene and use the differences of images for estimating the haze properties. Kopf et al. [7] uses existing 3D geographic models of the scene for dehazing. While these methods can enhance the visibility of hazy images, they cannot be applied in applications where additional information or multiple images are not available. Therefore, single image haze removal has been a hot spot of research since it is more applicable.

Like image denoising and restoration [8–13], dehazing is an ill-posed/inverse problem. Various single image dehazing algorithms have been developed to tackle this problem. They can be roughly divided into two classes: prior model based and data driven methods. For easy explaining, we compare the state-of-the-art algorithms in Table A.6 in Appendix A. Prior model based methods are designed to find scene depth by employing certain prior model. The classical priors include dark-channel prior (DCP) [14], Bayesian statistical prior [15, 16], color-lines prior (CLP) [17], color attenuation prior (CAP) [18], difference-structure-preservation prior (DSPP) [19] and color ellipsoid prior (CEP) [20]. A Laplacian-based mechanism [21] and median filter technique [22] were proposed to boost the performance of DCP [14].
Recently, data driven methods were explored. They usually employ machine learning methods to learn transmission map, such as random forest regression in [23], Convolutional Neural Networks (CNNs) in [24,25] and Generative Adversarial Network (GAN) in [26–29]. The challenge of data driven methods is the lack of training data, which compromises the efficiency of learning based methods. Specifically, outdoor scenes are rare in training data since current depth sensor cannot well acquire depths of them.

The comparison between pixel-wise and patch-wise DCPs reveals that they have their respective pros and cons, as illustrated in Fig. 1. On one hand, pixel-wise method is free of halo artifacts. However, it always overestimates haze intensity beyond its actual value, which leads to over-saturated haze removal image, as explained in Fig. 1(d). On the other hand, although the patch-wise dehazing, as shown in Fig. 1(c) and (f), could refrain from over-saturation, it may result in halo artifacts. Further in-depth analysis discovers that transmission map of an image is often misestimated around the edge of an object where image intensity varies abruptly, the region connecting foreground and background where depth changes dramatically. This Transmission MisEstimated (TME) problem would result in halo artifacts during patch-wise dehazing, specifically, inside TME region. While pixel-wise dehazing works well inside TME region. Therefore, fusing pixel-wise transmission map and patch-wise one through the weights indicated by TME would provide us a transmission map which could overcome both halo artifacts and over-saturation. We name this fusing process as Multi-scale Optimal Fusion (MOF) in this work.

The major contributions of this paper are listed as follows. Firstly, TME could provide us the weights for mixing pixel-wise and patch-wise transmission maps, so an approach of recognizing TME is proposed. Secondly, we propose a Multi-scale Optimal Fusion (MOF) model to refine transmission map based on DCP [14]. The proposed model mixes the pixel-wise and patch-wise transmission maps in an optimal manner by using a minimal variance estimation method. Thirdly, we raise a post-processing by employing an exposure enhancing method and a fast Gradient Domain GIF (GD-GIF) [30] to improve the proposed dehazing algorithm more robust and fast. Finally, beyond the DCP dehazing, the MOF model can be easily embedded into other patch-wise dehazing methods for suppressing halo artifacts. In this work, we have tested it in Boundary Constraint Contextual Regularization (BCCR) [31] and Color Attenuation Prior (CAP) [18], and achieved favorable results.

The rest of the paper is organized as follows. Section 2 gives the principle of DCP [14]. In Section 3, the details of the proposed MOF model are presented, including TME region recognition, single-scale transmission map estimation, and multi-scale transmission map estimation. In Section 4, a post-processing is presented for further improving robustness and saving computational complexity of the MOF. In Section 5, the experimental results demonstrate the superior performance of the proposed MOF on a variety of images. Finally, discussions and conclusions are drawn in Sections 6 and 7, respectively.

2. Background

According to Koschmieder’s law, the haze model can be mathematically described as [32]:

\[ I(x) = J(x) \cdot \tau(x) + A(1 - \tau(x)), \]  

(1)

where \( x \) is a spatial location, \( I \) is the observed image, \( J \) is scene radiance, \( A \) is global atmospheric light, and \( \tau(0 \leq \tau(x) \leq 1) \) is the medium transmission, which describes the portion of light that is not scattered and reaches the camera. According to Lambert Beer law [33], for a transparent object, when atmosphere is homogeneous, its transmission map \( \tau(x) \) is related to its depth \( d(x) \) by \( \tau(x) = e^{-\varepsilon d(x)} \) (\( \varepsilon \) represents scattering coefficient of atmosphere). Regarding hazing model of (1), \( J \) is an original image, which is intended to be recovered from its degraded version \( I \). To resolve \( J \) from (1), we have to estimate \( A \), and \( \tau(x) \) from \( I \) beforehand.

To estimate \( A \) and \( \tau(x) \), [14] proposed the DCP based on the observation that, most non-sky patches of the outdoor haze-free images, at least one color channel has some pixels with very low intensity. For an image \( I \), its dark channel is obtained by [14]:

\[ d_I(x) = \min_{y \in I(x)} \min_{c \in \{r, g, b\}} I(x)^c, \]

(2)

where \( I^c \) is a color channel of \( I \) and \( \Omega(x) \) is a local patch centered at \( x \). \( \min_{y \in I(x)} \) is performed on each pixel in the RGB space, and \( \min_{c \in \{r, g, b\}} \) is a minimum filter.

For estimating \( A \), the top 0.1% brightest pixels in the dark channel are picked up at first. Then, among these pixels, the pixels with the highest intensity in the given image \( I \) can represent the global atmospheric light \( A \).

Applying the DCP [14] to both sides of Eq. (1), noting that the dark channel of the haze-free radiance \( J \) tends to be zero, transmission \( t_{\text{pa}} \) can be estimated by [14]:

\[ t_{\text{pa}}(x) = 1 - \alpha \min_{y \in \Omega(x)} \min_{c \in \{r, g, b\}} \frac{I^c(y)}{A^c}. \]

(3)

where \( t_{\text{pa}} \) is the patch-wise transmission map of a local patch \( \Omega(x) \), \( A^c \) represents \( A' \), \( A'' \) or \( A''' \) which is the component of \( A \) at RGB space, and \( \alpha \) is application-based constant used to adaptively keep more haze for the distant objects [14]. As patch size equals to 1, i.e., \( \Omega(x) \) contains only one pixel, patch-wise transmission map \( t_{\text{pa}} \) becomes pixel-wise one, denoted by \( t_{\text{p}} \). In dehazing, transmission map is expected to be constant/smooth for the same object, since more textures would harm the depth of dehazed image. For smoothing transmission map, He et al. [14] further applied the GIF to \( t_{\text{pa}} \) from (3) for smoothing textures, meanwhile preserving image edges as good as possible [34].

After obtaining atmospheric light \( A \) and transmission map \( \tau(x) \) (\( t_{\text{p}} \) or \( t_{\text{pa}} \)), the scene radiance \( J \) is recovered by

\[ J(x) = \frac{I(x) - A}{\tau(x)} + A. \]

(4)

From (4), the success of dehazing mainly comes from a good transmission map \( \tau(x) \). By analyzing the state-of-the-art works, both physical model and deep learning based model fall into the issues of getting good transmission map and atmospheric light, as well as low computational complexity.
3. Multi-scale optimal fusion dehazing model

Since patch-wise and pixel-wise estimation models have their pros and cons, respectively, the straightforward way is to combine their pros and avoid their cons. Fig. 2 describes the process of deriving joint transmission map $t_{mof}$ from patch-wise $t_{pa}$ and pixel-wise $t_p$. A region with large depth variation is illustrated in Fig. 2(a). As discussed above, patch-wise dehazing would result in halo artifacts around this edge with abrupt grayscale change, namely TME region, labeled by dashed curves in Fig. 2(a), so pixel-wise instead of patch-wise dehazing is expected in TME region. For this purpose, the ideal transmission map $t_{mof}$ is computed by

$$t_{mof}(x) = W_{mof}(x) \cdot t_p(x) + W_{mce}(x) \cdot t_{pa}(x),$$

where $t_{pa}$ and $t_p$ represent transmission maps of pixel-wise and patch-wise, $W_{mce}$ and $W_{mof}$ give the weights for combining $t_{pa}$ and $t_p$. Since $0 < t_{mof}(x) < 1$, we have $W_{mce}(x) + W_{mof}(x) = 1$.

To get an ideal transmission map, we should minimize the difference between $t_{mof}(x)$ and $t_p(x)$ inside the TME region, and minimize the difference between $t_{mof}(x)$ and $t_{pa}(x)$ outside the TME region, i.e.

$$W_{mce}(x) \approx 0 \text{ and } W_{mof}(x) \approx 1, \quad x \in TME;$$

$$W_{mce}(x) \approx 1 \text{ and } W_{mof}(x) \approx 0, \quad x \notin TME;$$

indicating that $W_{mce}$ and $W_{mof}$ act as the masks of TME, as illustrated in Fig. 2(b). With (6), we propose our optimal fusion model for transmission map refining using $L_2$ regularization:

$$\min_{t_{mof}} \|W_{mof} \cdot t_{mof} - t_p\|^2_2 + \|W_{mce} \cdot t_{mof} - t_{pa}\|^2_2 + \alpha \|1 - t_{mof}\|^2_2 + R(t_{mof}),$$

where $\alpha$ is a regularization coefficient used to avoid invalid output; the smoothness regularization term $R(t_{mof})$ accounts for the smoothness of transmission map. Since $W_{mce}$ & $W_{mof}$ act as the mask of TME, so we should firstly recognize TME before the following processing.

3.1. Extracting the weight maps $W_{mce}$ & $W_{mof}$

Analyzing how $t_{pa}$ and $t_p$ are computed from (3), TME region can be roughly extracted by following method:

$$D_i = \max ((1 - ad_{pa}) - (1 - ad_{pa}), 0),$$

where $d_{pa}$ and $d_p$ are the dark channels of $I/A$ calculated by pixel-wise and patch-wise estimations, respectively.

However, $D_i$ may contain much textures as illustrated in Fig. 3(a), so it undergoes Gaussian filtering for smoothing textures. The filtered result (denoted by $D_f$) is shown in Fig. 3(b), where it can be observed that some textures are smoothed. Then, for enhancing contrast, $D_f$ further undergoes a tanh function. The enhanced result (denoted by $D_e$) is shown in Fig. 3(c), where the outlines/edges are kept well; meanwhile, textures are smooth. Thus, we finally obtain the $W_{mce} = D_f$ and $W_{mof} = 1 - D_f$, respectively.

3.2. Single-scale optimal fusion model

Let $W_{mce} = D_f$ and $W_{mof} = 1 - D_f$, the single-scale optimal fusion model (7) can be rewritten as:

$$\min_{t_{mof}} \|D_f \cdot t_{mof} - t_{pa}\|^2_2 + \| (1 - D_f) \cdot t_{mof} - t_p\|^2_2 + \alpha \|1 - t_{mof}\|^2_2 + R(t_{mof}).$$

(9)

It is difficult to solve $t_{mof}$ from (9) directly, so a two-step approach is raised to solve (9) approximately: (1) firstly, $t_{mof}$ is solved without the smoothness regularization by

$$t_{mof}(x) = \frac{D_f(x) \cdot t_p(x) + (1 - D_f(x)) \cdot t_{pa}(x) + \lambda_2}{D_f^2(x) + (1 - D_f(x))^2 + \lambda_2};$$

(10)

(2) then the $t_{mof}$ is smoothed by a fast GD-GIF, which will be discussed in Section 4.1.

The transmission map $t_{mof}$ computed from the above two steps are shown in Fig. 1(d) and (e), where the latter is with smoother textures. Applying the proposed single-scale model, the dehazed image is shown in Fig. 3(f) and (g). It can be observed that the haze is removed successfully, which proves the effectiveness of the TME region and achieve almost halo-free output. To explain this observation, we analyze a 1D signal with an abrupt grayscale change for examining transmission map and halo artifact. In Fig. 4(a), the grayscale abruptly changes around $x = 80$. In addition, small grayscale change is along horizontal axis. Fig. 4(b) draws the 1D curve of the region surrounded by dashed curve in Fig. 4(a). Observing Fig. 4(b), transmission map of

![Fig. 3. TME region recognition and refining results. (a)-(c) are $D_i$, $D_f$, and $D_e$ maps, respectively; (d) gives transmission map of the proposed model; (e) gives (d) followed by a GIF; (f) shows the result of the proposed single-scale model; and (g) is the zoom-in image of (f).](image-url)

![Fig. 4. TME and transmission map resulted from single-scale optimal fusion model. (a) gives an input image containing horizontal lines, and an abrupt grayscale change is around line 80. (b) shows different transmission maps and TME region. (c) shows that halos generated from patch-wise model GIF[$t_{pa}$] can be efficiently suppressed by the proposed model $t_{mof}$.](image-url)
the proposed model \( (t_{\text{mo}}) \) is closer to that of pixel-wise one \((t_p)\) in TME region (labeled by dashed lines). In addition, they both coincide with the real edge well around \( x = 80 \). While outside TME region, patch-wise one is expected instead of pixel-wise one, since the latter would result in more complex textures of transmission map, leading oversaturation of dehazing. Observing Fig. 4(c), it is clear that halo artifacts raised by patch-wise model \((GIF(t_p))\) can be successfully suppressed by the proposed model \((t_{\text{mo}})\). Here, “gif” indicates that they are all undergone a GIF for smoothing textures of transmission map.

### 3.3. Multi-scale Optimal Fusion (MOF) model

Single-scale optimal fusion model (9) for transmission map estimating can successfully suppress haloes at TME region by solving (10). However, for complex case, e.g., images of different resolution, or objects of different size in an image, single-scale patch may not be competent. Thus, we extend single-scale model to multi-scale one. To do this, we first estimate the \( t_p \) for different patch size \( r^j \) \((j = 1, 2, \ldots)\). The choice of \( r^j \) depends on image resolution that

\[
 r^j = \lfloor 2 \cdot \log((w \times h)) \rfloor - 1, \quad j = 1, 2, \ldots, \beta. \tag{11}
\]

where \( \lfloor . \rfloor \) is a rounded down function, and \( w \) and \( h \) are the width and high of the input image, respectively. Then, according to (10), different scales of transmission map \( (t_{\text{mo}}^j) \) are calculated. Finally, all these transmission maps are fused together to give a multi-scale transmission map through

\[
t_{\text{mo}} = \sum_j \nu^j \cdot t_{\text{mo}}^j, \quad j = 1, 2, \ldots, \beta, \tag{12}
\]

where \( t_{\text{mo}}^j \) is the fused transmission map, \( \beta \) is the number of scales, \( \nu^j \) is the weight for the \( j \)th scale, conforming to \( \sum_j \nu^j = 1 \). Take two-scale model as an example, for a 600 x 400 image, the patch sizes are \( r^1 = 9 \) and \( r^2 = 18 \) which are very close to the recommendation in literature of DCP [14]. All scales of \( t_{\text{mo}}^j \) take the same \( \lambda \), given in (10), and the same parameters for deriving \( D' \) around the discussion of (8). However, the TME region of small scale/patch is smaller than that of large one, so contrast enhancement of small scale is less than that of large one in fact. To compensate this uneven enhancement, the weight is defined by

\[
 \nu^j = e^{\tau \cdot \beta \cdot (\beta - j + 1)} \cdot \sum_j e^{\tau \cdot \beta \cdot (\beta - j + 1)}, \quad j = 1, 2, \ldots, \beta, \tag{13}
\]

where \( r^j \) is the patch size of the \( j \)th scale, \( \tau \) is the proportional control factor. The larger \( \tau \) indicates more uneven of the weights.

The processing flow of the multi-scale MOF is demonstrated in Fig. 5, where two single-scale fused transmission maps are calculated from two path-wise transmission maps \((t_{\text{mo}}^1, t_{\text{mo}}^2)\) and one pixel-wise transmission map \((t_p)\) firstly. Secondly, these two transmission maps are fused together to give a multi-scale one, namely \( t_{\text{mo}} \), via (12) and (13). Then, \( t_{\text{mo}} \) is followed by a GIF.

Table 1 lists average SSIM [36] and running time of the MOF under four scales \((\beta = 1, 2, 3, 4)\). The statistics of SSIM and running time are calculated over 189 images from O-HAZE database [35] (45 images) and NYU2 Depth database [37] (144 images). From 1, with two scales or more, SSIM is not improved obviously, but running time increases a lot. For a tradeoff between quality efficiency and computational complexity, two scales are used in the proposed model.

### 4. Post-processing

#### 4.1. Fast gradient domain GIF

As mentioned above, for suppressing textures in a transmission map, a fast GD-GIF is further imposed on the output of (12). Comparing to GIF, GD-GIF contains an explicit first-order edge aware constraint, preserving edges better than GIF [30]. GD-GIF [30] is with a little computational overhead. Inspired by [38], we downsample raw transmission map and guidance image firstly, and then upsample them after GD-GIF [30] to reduce its complexity from \( O(N) \) to \( O(N/\tau^2) \).

After getting transmission map \( \tau(I) \) and atmospheric light \( A \) (computed by the method of DCP [14]), the scene radiance is computed by (4).

#### 4.2. Exposure enhancing

The image after haze removal looks so dark that image details cannot be seen clearly. This is because transmission map of the MOF is always less than patch-wise one according to (3). Fig. 6(b) illustrates an example of the difference map of \( t_p - t_{\text{mo}}^j \), where 99.93% pixels are positive, indicating that \( t_p \) are commonly larger than \( t_{\text{mo}}^j \). To address this problem, the dehazed image is further adjusted by an adaptive exposure scaling [23]

\[
\min \| 1 - S \cdot (I_F / J_F) \|_2^2 + \lambda \| S - 1 \|_2^2 + R(S), \tag{14}
\]

where \( S \) is the illumination scaling field, \( I_F \) is the illumination of hazy image \( I \), \( J_F \) is the illumination of scene radiance \( J \), \( \lambda \) is a regularization coefficient, given by the mean of \( J_F \), and \( R(\cdot) \) is a term of smoothness regularization. Following the same way for solving (9), \( S \) is computed from (14) without smoothness regularization term. Then, it is further processed by the fast GD-GIF [30].

For testing the effectiveness of fast GD-GIF [30] and exposure enhancing, a sample image from the O-HAZE database [35] is processed, and the results are demonstrated in Fig. 6. It can be observed that the dehazed image of MOF (Fig. 6(c)) looks quite dark. After exposure enhancing, its visual quality is improved remarkably as shown in Fig. 6(d). Fig. 6(e) and (f) show the MOF further undergoes exposure enhancing and fast GD-GIF [30].

### 5. Experimental results

To evaluate the proposed MOF, three groups of experiments are performed. First, beyond DCP [14], we embed the MOF into other two dehazing methods, Boundary Constraint Contextual Regularization (BCCR) [31] and the Color Attenuation Prior (CAP) [18]. Second, the MOF based DCP [14] is compared with the state-of-the-arts qualitatively, where the benchmarks include He et al. He et al. [14], Tarel [39], Tan et al. [32], Nishino et al. [15], Gibson et al. [40], Meng et al. [31], Fattal et al. [37], Zhu et al. [15], Berman et al. [41, 42] and Bahat et al. [43], as well as the newly developed learning or deep learning based methods, such as Tang et al. (BCCR) [31], Fattal et al. (CL) [17], Zhu et al. (CAP) [18], Berman et al. (Nonlocal) [41, 42] and Bahat et al. [43]. To address this problem, the dehazed image is further adjusted by an adaptive exposure scaling [23]

### Table 1

<table>
<thead>
<tr>
<th>Indexes</th>
<th>( \beta = 1 )</th>
<th>( \beta = 2 )</th>
<th>( \beta = 3 )</th>
<th>( \beta = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.821</td>
<td>0.825</td>
<td>0.826</td>
<td>0.826</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>5.225</td>
<td>7.468</td>
<td>9.553</td>
<td>14.901</td>
</tr>
</tbody>
</table>

*Tested on O-HAZE database [35].

**Note:** The configurations of the benchmarks strictly follow the authors’ recommendations in their papers. The parameters of our method are: \( \lambda = 10^{-6} \) in (10) (all scales of \( t_{\text{mo}} \) take the same \( \lambda \), \( \tau = 0.138 \) in (13), and \( s = 3 \) in the fast GD-GIF.
Fig. 5. Flowchart of the MOF framework using two scales.

Fig. 6. An example of dehazing by using the proposed MOF. (a) is the hazy image. (b) is the difference map of $t_{pa} - t_{mof}$. (c) gives the result of the MOF without exposure enhancing. (d) shows the result of the MOF after exposure enhancing (+E); (e) and (f) gives the results of the proposed MOF cooperated with both the fast GD-GIF (+F) and exposure enhancing (+E), where $s = 2$ and $s = 3$ represent the number of scales in the GD-GIF for downscaling the original transmission map.

Fig. 7. Visual quality comparisons among the pixel-wise, patch-wise and MOF embedded DCP, BCCR and CAP on two images (a) and (b).

5.1. Beyond DCP

As described above, the MOF merges the merits of pixel-wise and patch-wise DCPs. In this part, beyond DCP [14], we prove that the MOF is more applicable, which can be easily embedded into other dehazing methods. For this purpose, the pixel-wise, patch-wise and their combining, i.e., MOF versions of BCCR [31], CAP [18] and DCP [14] are compared with respect to a blind contrast enhancement assessment (BCEA) metric [44]. For the details of BCCR/CAP and BCEA, they are illustrated in Appendices B and C, respectively.

The pixel-wise BCCR (BCCR$_{pi}$) can be realized similar to DCP, i.e. setting the window radius $r_b = 1$. For patch-wise BCCR (BCCR$_{pa}$), we set $r_b = 9$, in our experiments. The corresponding transmission maps of BCCR$_{pi}$ and BCCR$_{pa}$ are noted as $t_{pi}$ and $t_{pa}$. Replacing $t_{pi}$ and $t_{pa}$
Table 2
BCEA measurements on the images in Fig. 7.

<table>
<thead>
<tr>
<th>Image Index</th>
<th>DCP_p</th>
<th>DCP_p0</th>
<th>DCP_p0/m</th>
<th>BCC R_p</th>
<th>BCC R_p0</th>
<th>BCC R_p0/m</th>
<th>CAP_p</th>
<th>CAP_p0</th>
<th>CAP_p0/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>−0.0284</td>
<td>0.0351</td>
<td>0.0833</td>
<td>0.0584</td>
<td>0.0728</td>
<td>0.0592</td>
<td>−0.0504</td>
<td>0.0031</td>
<td>0.0581</td>
</tr>
<tr>
<td>r</td>
<td>1.2152</td>
<td>1.0405</td>
<td>1.3668</td>
<td>1.2757</td>
<td>1.1426</td>
<td>1.3053</td>
<td>1.0596</td>
<td>1.0450</td>
<td>1.1808</td>
</tr>
<tr>
<td>p</td>
<td>0.0110</td>
<td>0.0003</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
<tr>
<td>e</td>
<td>0.8629</td>
<td>0.8264</td>
<td>0.8893</td>
<td>0.4168</td>
<td>0.2580</td>
<td>0.4800</td>
<td>0.0559</td>
<td>0.0989</td>
<td>0.3841</td>
</tr>
<tr>
<td>r</td>
<td>1.5813</td>
<td>1.3216</td>
<td>1.5910</td>
<td>1.4441</td>
<td>1.3044</td>
<td>1.4589</td>
<td>0.7748</td>
<td>0.9527</td>
<td>1.2249</td>
</tr>
<tr>
<td>p</td>
<td>0.0245</td>
<td>0.0001</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 3
Statistics of BCEA indexes for 80 images.

<table>
<thead>
<tr>
<th>Indexes</th>
<th>DCP_p</th>
<th>DCP_p0</th>
<th>DCP_p0/m</th>
<th>BCC R_p</th>
<th>BCC R_p0</th>
<th>BCC R_p0/m</th>
<th>CAP_p</th>
<th>CAP_p0</th>
<th>CAP_p0/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>30</td>
<td>4</td>
<td>46</td>
<td>29</td>
<td>3</td>
<td>48</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>r</td>
<td>8</td>
<td>1</td>
<td>71</td>
<td>1</td>
<td>2</td>
<td>77</td>
<td>3</td>
<td>1</td>
<td>76</td>
</tr>
<tr>
<td>p</td>
<td>8</td>
<td>75</td>
<td>31</td>
<td>19</td>
<td>77</td>
<td>33</td>
<td>17</td>
<td>72</td>
<td>28</td>
</tr>
</tbody>
</table>

By $t_{pa}$ and $t_{pm}$ respectively in (10), we can obtain the MOF based BCCR ($BCCR_{p0/m}$). In original CAP [18], the radius of $\Omega$ is 1, which means that the CAP [18] is a pixel-wise method. It is easy to modify CAP [18] to a patch-wise one by setting the radius of $\Omega$ larger than 1, which is 5 in our testing. We mark the original CAP [18] as $CAP_m$ and the patch-wise one as $CAP_p$. The patch-wise method of $CAP_p$ and $CAP_m$ are denoted as $d_p$ and $d_m$, respectively. Replacing $t_{pa}$ and $t_{pm}$ by $d_p$ and $d_m$ in (10), respectively, the $CAP_m$ can be obtained.

Qualitative comparisons of visual quality are shown in Fig. 7, where the pixel-wise, patch-wise and MOF versions of DCP [14], BCCR [31] and CAP [18] are indicated by $DCP_p$, $BCCR_p$ and $CAP_p$, respectively. From Fig. 7, the results of pixel-wise methods ($DCP_p$, $BCCR_p$ and $CAP_p$) are extremely oversaturated (bounded by a rectangular in yellow). When people watch these scenes, they may feel unnatural and unsatisfied depth sense (highlighted in red rectangular). On the contrary, the dehazed images of patch-wise methods ($DCP_p$, $BCCR_p$ and $CAP_p$) can give us pleasant depth sense, but halo artifacts are generated. By contrast, the MOF based methods ($DCP_{p0/m}$, $BCCR_{p0/m}$ and $CAP_{p0/m}$) can overcome aforementioned flaws of pixel-wise and patch-wise methods, give us pleasant depth sense and free from halo artifacts.

To quantitative comparison, BCEA [44] indexes are computed. In BCEA, three objective indicators, $\epsilon$, $\sigma$, and $\rho$ defined in [44] are employed to quantitatively assess dehazing results. Indicator $\epsilon$ represents the rate of edges newly visible after restoration. Indicator $\sigma$ represents the contrast restoration, which is the mean of normalized gradients of visible edges. Indicator $\rho$ is the occurrence probability of oversaturated pixels which become black or white after restoration. Generally, the larger $\epsilon$, $\sigma$, and the smaller $\rho$, the better the performance.

Table 2 shows the results of pixel-wise methods ($DCP_p$, $BCCR_p$ and $CAP_p$). From Table 2, we can find that the MOF based methods ($DCP_{p0/m}$, $BCCR_{p0/m}$ and $CAP_{p0/m}$) can give us pleasant depth sense, but halo artifacts are generated. By contrast, the MOF based methods ($DCP_{p0/m}$, $BCCR_{p0/m}$ and $CAP_{p0/m}$) can overcome aforementioned flaws of pixel-wise and patch-wise methods, give us pleasant depth sense and free from halo artifacts.

By $t_{pa}$ and $t_{pm}$ respectively in (10), we can obtain the MOF based BCCR ($BCCR_{p0/m}$). In original CAP [18], the radius of $\Omega$ is 1, which means that the CAP [18] is a pixel-wise method. It is easy to modify CAP [18] to a patch-wise one by setting the radius of $\Omega$ larger than 1, which is 5 in our testing. We mark the original CAP [18] as $CAP_m$ and the patch-wise one as $CAP_p$. The patch-wise method of $CAP_p$ and $CAP_m$ are denoted as $d_p$ and $d_m$, respectively. Replacing $t_{pa}$ and $t_{pm}$ by $d_p$ and $d_m$ in (10), respectively, the $CAP_m$ can be obtained.

Qualitative comparisons of visual quality are shown in Fig. 7, where the pixel-wise, patch-wise and MOF versions of DCP [14], BCCR [31] and CAP [18] are indicated by $DCP_p$, $BCCR_p$ and $CAP_p$. From Fig. 7, the results of pixel-wise methods ($DCP_p$, $BCCR_p$ and $CAP_p$) are extremely oversaturated (bounded by a rectangular in yellow). When people watch these scenes, they may feel unnatural and unsatisfied depth sense (highlighted in red rectangular). On the contrary, the dehazed images of patch-wise methods ($DCP_p$, $BCCR_p$ and $CAP_p$) can give us pleasant depth sense, but halo artifacts are generated. By contrast, the MOF based methods ($DCP_{p0/m}$, $BCCR_{p0/m}$ and $CAP_{p0/m}$) can overcome aforementioned flaws of pixel-wise and patch-wise methods, give us pleasant depth sense and free from halo artifacts.

5.2. Evaluations on natural hazy images

Fig. 8 gives the comparisons among the proposed MOF embedded DCP [14] and the state-of-the-art methods. Without confusing, the proposed dehazing method (i.e. $DCP_{p0/m}$) is also labeled by “MOF” in discussion. Tarel et al. [39] method perform well on image Fig. 8(a) but it fails on image Fig. 8(b) highlighted by a rectangular. Nishino et al. [15] and Tan et al. [32] suffers from overly-enhanced visual artifacts, and Gibson et al. [40], Bahat et al. [43] and Chen et al. [45] clean the haze incompletely, leading to the loss of color contrast in the output images. Fattal et al. (CL) [17] and Berman et al. (NLD) [41,42] have the most competitive visual results among all, with plausible details, however, both of them yields over-saturation around branches of the image in Fig. 8(b).

Furthermore, we also compare the MOF with the widely developed deep learning based dehazing methods, including the DNet [24], MSCNN [25], AOD-Net [46], GAN [26] dehazing, and also the learning framework (LF) of Tang et al. [23]. Although the AOD-Net [46] and GAN [28] successfully refrain from both over-saturation and halo artifacts as illustrated respectively in Fig. 9(a) and (c), the AOD-Net [46] loses the sense of depth, e.g. it is difficult to differentiate the tree leaves and the grass in the image given in Fig. 9(a); and the GAN [28] performs slight color-distortion, e.g. the red house in the image shown in Fig. 9(c). LF [23], DNet [24] and MSCNN [25] can achieve non-overestimated results, but they fail to effectively dehaze. GAN [25] performs better than previous learning based methods [23–25], however it produces less adequate dehazing in the distance scenes as red rectangulars shown in Fig. 9(c). By contrast, the proposed MOF recovers richer and more saturated colors, while suppressing most artifacts.

5.3. Evaluations on synthesized hazy images

We also evaluate dehazing methods on the synthesized indoor databases (including NYU2 Depth [37], Middlebury [47], I-HAZE [48] and SOTS-indoor [49]) and outdoor databases (including FRIDA2 [39], HazeRD [50], O-HAZE [35] and SOTS-indoor [49]). Databases of NYU2 Depth [37], Middlebury [47], FRIDA2 [39] and HazeRD [50] provide both clean images and their depths. The synthesized hazy image is generated by imposing the physical model (1) on clean images. I-HAZE [48] and O-HAZE [35] are generate hazy image using real haze produced by a professional haze machine. In this test, the hazy sample images of NYU2 Depth [37] and Middlebury [47] are collected from the work of [46] and [51], respectively.

To evaluate the MOF framework, we adopt four full reference image quality assessment (FRIQA) metrics, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [36], visual signal-to-noise ratio (VSNR) [52] and feature similarity index for color image (FSIMc) [53], comparing with the benchmarks. The PSNR is an most
popular index of image quality, however it was criticized recently due to inconsistent with human visual system (HVS). The SSIM is used to evaluate the structural information preserving ability of the dehazing methods. VSNR is an efficient metric for quantifying the visual fidelity of natural images based on near-threshold and supra-threshold properties of human vision. Higher VSNR means that the method can maintain higher-level properties of vision. In FSIMc, the phase congruency, gradient magnitude and chromatic features are used to represent complementary aspects of image visual quality beyond the SSIM. Higher FSIMc indicates that the restored images are not only more close to the ground truth, but also better perceptual visual quality from human visual system.

The statistics of all FRIQA indexes are given in Tables 4 and 5, where the best performance is highlighted in bold. Visual comparison is provided in Figs. 10 and 11, selecting one sample from indoor and outdoor databases, respectively.

Evaluation on indoor databases. From Table 4, the deep-learning based methods DNet [24] and MSCNN [25] perform better than early dehazing methods BCCR [31], CAP [18] and NLD [41] on indoor databases. For example, DNet [24] gets the highest PSNR score on the first three indoor databases and all indoor samples; MSCNN [25] is the best with respect to FSIMc index. Fig. 10(g) and (h) illustrate that these two methods are able to generate realistic colors then other methods for indoor image dehazing. However, the proposed MOF dehazing can commonly achieve better SSIM and VSNR scores, such as in I-HAZE [48] and the total indoor samples, meaning that this method have more competitive visual results than BCCR [31], CAP [18] and NLD [41].

Additionally, we evaluate all tested methods at different haze levels over SOTS-Indoor [49] database. As illustrated in Fig. 10, although all of the tested methods can restore low or medium level hazy image effectively, only MOF, BCCR [31] and MSCNN [25] can maintain the similar efficiency for high level dehazing. Comparing with MOF, results of BCCR [31] BCCR tends to be little over-saturated. MSCCN [25] is less efficiency for high level hazy image dehazing. The FRIQA indexes listed in Tables 4 and 5 also demonstrate that the proposed MOF method is the best on SOTS-Indoor [49] database with respect to PSNR, VSNR and FSIMc.

Evaluation on outdoor databases. For evaluation on outdoor synthetic images, all tested methods are firstly implemented on FRIDA2 [39] database, where 96 inhomogeneous hazy images of diverse road
Fig. 10. Indoor synthetic image dehazing results. (a) is the hazy image. (b) is the ground truth. (c)-(h) are the results of MOF, BCCR [31], CAP [18], NLD [41], DHNet [24] and MS-CNN [25].

Table 5: Average FRIQA indexes on four outdoor databases.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSIM</td>
<td>0.690</td>
<td>0.678</td>
<td>0.599</td>
<td>0.588</td>
<td>0.609</td>
<td>0.688</td>
</tr>
<tr>
<td>(96 samples)</td>
<td>VSNR</td>
<td>4.205</td>
<td>4.025</td>
<td>5.892</td>
<td>4.527</td>
<td>6.071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.713</td>
<td>0.713</td>
<td>0.713</td>
<td>0.713</td>
<td>0.713</td>
<td>0.713</td>
</tr>
<tr>
<td>(75 samples)</td>
<td>VSNR</td>
<td>7.352</td>
<td>7.352</td>
<td>7.352</td>
<td>7.352</td>
<td>7.352</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.682</td>
<td>0.682</td>
<td>0.682</td>
<td>0.682</td>
<td>0.682</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
</tr>
<tr>
<td>(45 samples)</td>
<td>VSNR</td>
<td>5.814</td>
<td>5.814</td>
<td>5.814</td>
<td>5.814</td>
<td>5.814</td>
<td>5.814</td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.838</td>
<td>0.838</td>
<td>0.838</td>
<td>0.838</td>
<td>0.838</td>
<td>0.838</td>
</tr>
<tr>
<td>SOTS-Outdoor [49]</td>
<td>PSNR</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
</tr>
<tr>
<td>(300 samples)</td>
<td>VSNR</td>
<td>8.915</td>
<td>8.915</td>
<td>8.915</td>
<td>8.915</td>
<td>8.915</td>
<td>8.915</td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
<td>0.921</td>
</tr>
<tr>
<td>FRIDA2 ⋃ HazeRD</td>
<td>PSNR</td>
<td>15.890</td>
<td>15.890</td>
<td>15.890</td>
<td>15.890</td>
<td>15.890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>VSNR</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
</tr>
<tr>
<td>O-HAZE ⋃ SOTS-Outdoor</td>
<td>PSNR</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td>15.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
<td>0.736</td>
</tr>
<tr>
<td>(466 samples)</td>
<td>VSNR</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
<td>8.931</td>
</tr>
<tr>
<td></td>
<td>FSIMc</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
<td>0.876</td>
</tr>
</tbody>
</table>

Fig. 11. Outdoor synthetic images with different haze images. Each haze scene contains five different haze levels. Fig. 11 shows one example with low, medium and high haze levels. We can find that, the MOF achieve better dehazing with visually appealing results at all levels. CAP [18], DNet [24] and MSCNN [25] can recover better dehazed results at low haze level; however, their performances drop at medium and high haze levels. BCCR [31] and NLD [41] have some color distortions in sky.

We add another two outdoor databases, O-HAZE [35] and SOTS-Outdoor [49] in this revision. Finally, we count the average FRIQA indexes of each dehazing method tested on total samples randomly selected from the four outdoor databases, as listed in the last row of Table 5. From Table 5, the proposed MOF gets the best performance on all the indexes, since it can handle well halo artifact and over-saturation simultaneously.

5.4. Computational complexity

We select 692 images with different resolutions from databases of [37,39,47,49,50] to compare the runtimes of various methods.
We executed the published MATLAB codes of BCCR [31], CAP [18], DNet [24] and MSCNN [25]. The DCP [14] is implemented by using the conventional GIF to refine the transmission. The experiments are all implemented on a Windows10 PC with an Inter(R) Core(TM) i7-8700 CPU @ 3.20 GHz processor with 16GB RAM. Experimental results are illustrated in Fig. 12, in which the horizontal axis represents the pixel number of the image and the vertical axis is the running time. In our method, the fast GD-GIF is performed twice. It is with the complexity of $O(N/8)$ (downsampling scale $s = 3$ as discussed in Section 4.1), which is significantly lower than $O(N)$ of the conventional GIF. From Fig. 12, CAP ranks the first with respect to time complexity since it also employs a fast GIF similar to the fast GD-GIF in our method. Benefiting from the fast GD-GIF, our method ranks the second among all benchmarks, better than BCCR [31], DNet [24] and MSCNN [25] although additional exposure enhancing is required.

6. Discussions

The identification of TME region and multi-scale transmission map mainly contributes the superior performance of the proposed MOF. In Section 5, we have shown that the proposed MOF can restore the outdoor hazy images effectively, quantitatively evaluated by BCEA [44] and FRIQA indexes. All the experimental results are provided by the two-scale MOF. The scales, which are the patch sizes, are adaptively calculated by the empirical equation (11). More in-depth theoretical analysis of the number of scales and corresponding patch size would be investigated in our near future research.

However, our dehazing method suffers from the problem of ambiguity between image color and depth [31]. It stated that a haze-free pixel in light color may be wrong recognized as a haze pixel [31]. For example, we may likely confuse the sky color, white object with hazes.
Therefore, without sufficient priors, these pixels are difficult to be reliably recognized as hazed or not. This ambiguity may overestimate the transmission map, resulting in excessive enhancement of the objects in an image. All of the tested algorithms [15,18,31], including the proposed MOF suffer from this ambiguity, as illustrated in Figs. 8(c) and 9(c).

Most existing prior-based methods [54–56] address this ambiguity by introducing additional sky region detection processing. However, accurate sky region detection is complex. Another strategy is to do dehazing in super-pixel domain [57,58], since sky region can be better detected in super-pixel domain. We will extend the MOF framework into super-pixel domain in our future work.

7. Conclusions

In this paper, an optimal multi-scale fusion model, namely MOF, was proposed to overcome the flaws of the classical DCP dehazing. Comparing the pixel-wise and patch-wise transmission maps, it can be found that the halo artifacts are prevailing around regions of abrupt grayscale change, i.e. TME regions, so we first presented a TME recognition method. Then, the TME map was used as the weights to fuse transmission maps of different sizes of DCPs to produce a better one. Such a fusion not only overcomes oversaturation but also suppress halo artifacts. After that, two post-processing methods were proposed to further improve the robustness and computational complexity of the MOF. The extensive experiments of three categories were performed on a variety of haze images, demonstrated that the MOF can provide more perceptually plausible results than most of the existing algorithms. There are still some remaining issues worth carrying out in our near future work, including more in-depth analysis of model parameters and the ambiguity between image color and depth mentioned above.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grants 61572461, 6166114605, U1611461, 11433006, 11790301 and 11790305, the PKU-NTU Joint Research Institute (JRI) Sponsored by a donation from Ng Teng Fong Charitable Foundation, and the CAS “100-Talents” (Dr. Xu Long).

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Appendix A

See Table A.6.

Appendix B

B.1. BCCR [31] dehazing

BCCR [31] method extends the idea of patch-wise DCP [14] in determining the raw transmission map by introducing its lower bound (considering that $J$ is always bounded, i.e. $C_0 < J(x) < C_1(x \in \Omega)$, where $C_0$ and $C_1$ are two constant vectors that are relevant to the given image; then the extrapolation of $J$ is located in a radiance cube bound). Since this contextual assumption often fails to image patches with abrupt depth jumps leading to significant halo artifacts in the dehazing results, a weighting function $W_{\text{bcr}}(x, y)$ on the constraints is introduced: $W_{\text{bcr}}(x, y) = \frac{1}{2} \sum_{i=1}^{k} d_i(x, y) - t(x, y) \approx 0$ which plays a “switch” role of the constraint between $x$ and $y$. When $W_{\text{bcr}}(x, y) = 0$, the corresponding contextual constraint of $t$ between $x$ and $y$ will be canceled. Noting that the role of this $W_{\text{bcr}}$ are very similar to our weighting function $W$. Differently, $W_{\text{bcr}}$ is obtained by the luminance difference of neighboring pixels [66], and its solution is calculated by employing a L1-norm based contextual regularization.

B.2. CAP dehazing

The CAP dehazing recovers the depth by creating a linear model under a color attenuation prior, i.e. the depth of the scene is positively correlated with the difference between lightness $\nu$ and situation $s$: $d(x) \sim \theta_0 + \theta_1 \cdot \nu(x) + \theta_2 \cdot s(x) + \epsilon(x) = N(\theta_0 + \theta_1 \cdot \nu(x) + \theta_2 \cdot s(x), \delta^2)$. In this linear model, the parameters $\theta_0$, $\theta_1$, $\theta_2$ and $\delta$ can be learned from a supervised learning method using 120 million scene points, so that the raw $d(x)$ can be obtained. However, this model may fail in some particular situations. For instance, the white objects in an image are usually with high values of the brightness and low values of the saturation. Therefore, the proposed model tends to regard the objects with white color as distant objects. Unfortunately, this misclassification will result in inaccurate estimation of the depth in some cases. To overcome this problem, the authors process the raw depth map by: $d_i(x) = \min_{y \in \Omega, x}(d(y))$, where $r_c$ is the radius of the window $\Omega$. However, it is obvious that the block artifacts still appear in the image. To refine the depth map, we use the fast guided image filtering [34,38] to smooth the image.

Appendix C

C.1. Definition of the visible edges

In the LIP model [67], the contrast between two pixels $x$ and $y$ of an image $f$ is given by:

$$C_{x,y}(f) = \max(f(x), f(y)) \ominus \min(f(x), f(y)),$$

where $\ominus$ denotes LIP subtraction. Then, the contrast associated to a border $F$ which separates two adjacent regions follows:

$$c_F(f) = \frac{1}{|F|} \ominus \bigoplus_{i \in F} C_{x,y}(f),$$

where $\ominus$ and $\oplus$ denote LIP multiplication and addition. The $F$ can be calculated by certain segmentation method. $\nu$ is the set of couple pixels in adjacent area, and $|\cdot|$ represents the number of elements in a set.

In the BCEA [44], $F$ is calculated by the Köhler’s segmentation method [68]; that is $F(s)$ is the set of all couples $(x, y)$ separated by $s$, in which the $s$ respects the condition of $\min(f(x), f(y)) \leq s \leq \max(f(x), f(y))$ and $y \in V_x(s)$ (four connected domain). For every couple belonging to $F(S)$, the contrast $C_{x,y}(s)$ is computed [44,69]:

$$C_{x,y}(s) = \min \frac{|x - f(x)|}{\max(s, f(x)), \max(s, f(y))}.$$

The mean contrast associated to $F(s)$ is then computed [44,69]:

$$C(s) = \frac{1}{|F(s)|} \sum_{(x,y) \in F(s)} C_{x,y}(s).$$

The best threshold $s_0$ verifies the following condition [44,69]:

$$s_0 = \arg \max_{s \in [0,255]} C(s).$$

It is the threshold which delivers the best mean contrast along the associated border $F(s_0)$. The $F(s_0)$ is the definition of visible edge in the BCEA.

C.2. $\epsilon$, $r$ and $p$ indexes in BCEA

Let $n_j$ and $n_i$ denote respectively the cardinal numbers of the set of visible edges in the original image $I(\in \Omega)$ and the dehazed image $J(\in \Omega)$, then the index $e$ is defined as [44]:

$$e = \frac{n_j - n_i}{n_i}.$$  

The index $r$ is defined as [44]:

$$r = \exp \left( \frac{1}{n_j} \sum_{i \in \Omega} \log r_x \right),$$

where $r_x$ is the gradient ratio between $I$ and $J$ at pixel $x$. 

262
Table A6
Comparisons of state-of-the-art dehazing methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Method description</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tan (08) [32]</td>
<td>This model is developed based on: (1) haze-free image has more contrast than</td>
<td>Improve the contrast of the hazy image</td>
<td>Oversaturated; halo artifact</td>
</tr>
<tr>
<td></td>
<td>hazy one, and (2) airtight $A(I-x(I))$ across image tends to be smooth. Thus, a</td>
<td></td>
<td>generates at depth discontinued region</td>
</tr>
<tr>
<td></td>
<td>optimization function is formulated to seek minimum contrast with smoothness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>constraint of airtight, where the contrast is defined as the number of edges in a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>image patch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fattal (08) [59]</td>
<td>$(1)$ is rewritten as $I(x) = (\hat{x}, R : \hat{x}, A(I-x(I))$ (i.e., $J(x) = I(x) \cdot R$, where $R$ describes the relative intensity of the reflected light, and $\hat{x}$ represents the shading. $C(I, I) = 0 (C$ represents covariance operation)</td>
<td>Usually produce impressive results when the color information is enough</td>
<td>Fails at dense-haze image for the lack of color information; the estimation of $A$ is passive</td>
</tr>
<tr>
<td>He et al. (09)</td>
<td>This state-of-the-art method uses dark channel prior (DCP) to estimate $I(x)$;</td>
<td>Result is natural without over-saturation</td>
<td>High computational complexity;</td>
</tr>
<tr>
<td></td>
<td>uses soft matting [60] or GIF [34] to refine $\hat{x}(I)$</td>
<td></td>
<td>halo artifact</td>
</tr>
<tr>
<td>Tarel–Hautiere (09) [59]</td>
<td>Referring to Fattal [59], the haze model is expressed as $I(x) = I(x) \cdot R : \hat{x}, A(I-x(I))$, denoted as $\hat{W}(x)$, and then computing $V$ from $\max(\min(</td>
<td>W(I) = R(x)\cdot\hat{x}, W(x), 0)$, where $p \in (0, 1)$; $R = \frac{1}{</td>
<td>\hat{W}(x)</td>
</tr>
<tr>
<td>Nishino–Kratz (09) [15]</td>
<td>A haze model $I(x) = (C(x) + D(x))$ is taken, where $I(x) = I(x) \cdot C(x)$, $C$ and $D$ are the functions of scene albedo and depth respectively; with the assumption that $C$ and $D$ are independent, referred to Factorial MRF: $\mu(C, D) \propto (I(C, D) \cdot C(x) \cdot D(D))$ to derive $C$ and $D$, and recurring to EM algorithm to decouple $C$ and $D$ during iteration</td>
<td>Efficient for heavy hazy image dehazing</td>
<td>May cause over-saturation and unnatural distortion</td>
</tr>
<tr>
<td>Ancuti–Ancuti (10) [61]</td>
<td>An hazy image is processed into a white balance image $W$ and a contrast-enhancement image $E$, three weight maps are derived from luminance, chromatic and saliency channels to fuse these two images, specifically in a multi-scale fusion manner, by applying Laplacian and Gaussian operators to $W$, $E$ and weight maps, respectively</td>
<td>Pixel-wise operation reduces the amount of artifacts</td>
<td>Limited to homogeneous hazy images</td>
</tr>
<tr>
<td>Meng et al. (13) [31]</td>
<td>This method extends the idea of DCP [14] in determining the raw transmission map $t$ by introducing its lower bound (considering that $J$ is always bounded, i.e. $C_0 &lt; J(x) &lt; C_1$), where $C_0$ and $C_1$ are two constant vectors that are relevant to the given image; then the extrapolation of $J$ is located in a radianse cube bounded; after that, a $L_1$-norm based contextual regularization is introduced to suppress halo artifacts with smoothness constraint of $t$</td>
<td>Visually pleasing results with faithful color and finer image details and structures</td>
<td>Color distortion occurs at too dark and bright areas</td>
</tr>
<tr>
<td>Fattal (14) [17]</td>
<td>Following the same model of Fattal [59], however different ways to solve $l$ and $t$; the authors find that a color-line (i.e. the values of $I$ in a patch are distributed along a 1D line in RGB space), which is parameterized by the pixel shading $\iota$, whose orientation coincides with the patch albedo $R$, is intersected with another line formed by $A(l-\iota)$ given $A(I-\iota)$ can be obtained recurring to the intersection</td>
<td>Less errors in both transmission estimation and final dehazing at different haze levels</td>
<td>Fails in sky-region dehazing; compromised by strong noise</td>
</tr>
<tr>
<td>Tang et al. (14) [23]</td>
<td>Four handcraft haze-relevant features, multi-scale dark channel, multi-scale local max contrast, hue disparity and multi-scale local max saturation are calculated; Random Forest Regression (RFR) is employed to learn a direct mapping between transmission map $t$ and the four features</td>
<td>Various haze-relevant multi-scale features are beneficial for accurate $t$</td>
<td>Overload of calculating multi-scale features for each pixel; unsatisfactory effect for heavy haze</td>
</tr>
<tr>
<td>Zhu et al. (15) [18]</td>
<td>Recovers the depth by creating a linear model under a color attenuation prior, i.e. the depth of the scene is positively correlated with the difference between lightness $t$ and saturation $s$; the parameters of this linear model is learned from a supervised learning method</td>
<td>Free from over-saturation; details can be moderately enhanced as well</td>
<td>Prone to under-estimating the $t$ in heavy hazy region</td>
</tr>
<tr>
<td>Berman et al. (16) [41]</td>
<td>A nonlocal approach, transmission map is estimated across the entire image instead of local patch: (1) $I(x)$ is first converted to spherical coordinates $I(x) = (\hat{x}, \phi(x), \psi(x))$, haze-line $H$ is clustered by ${</td>
<td>x, \phi(x), \psi(x)}$; (2) for each cluster $H$ estimate maximum radius $\sigma_{max}$; (3) the $t(x)$ are estimated by $t(x) = \hat{x}/f_{max}$</td>
<td>Work well on many real-world images</td>
</tr>
<tr>
<td>Cai et al. (16) [24]</td>
<td>An end-to-end CNN system, named DehazeNet, consists of four layers: (1) feature extraction layer: four haze-relevant features are extracted using a nonlinear function, Maxout unit [62]; (2) multi-scale mapping layer: generate multi-scale features; (3) local extremum layer: instead of patch-wise max-pooling, a pixel-wise extremum operator is applied to keep original resolution and overcome local sensitivity and noise of features; (4) nonlinear regression uses the ReLU activation function, resulting in $\alpha \in [0, 1]$</td>
<td>Can learn four most important haze-relevant features</td>
<td>Rely on training database; unsatisfactory in heavy haze regions</td>
</tr>
<tr>
<td>Ren et al. (16) [25]</td>
<td>A multi-scale CNN: (1) a coarse-scale network predicts a holistic transmission map $t$ of an image; (2) a fine-scale network further refines the global $t$ locally; (3) the coarse-scale output is fed to the fine-scale network at the second layer</td>
<td>More features are learned compared to Cai et al. [24]</td>
<td>Rely on training database; unsatisfactory in heavy haze regions</td>
</tr>
</tbody>
</table>
Let \( n_b \) denote the number of black or white pixels in \( J \). The index \( p \) is defined as [44]:

\[
p = \frac{n_b}{M \times N},
\]

where \( M \) and \( N \) are the width and height of the image.

References


