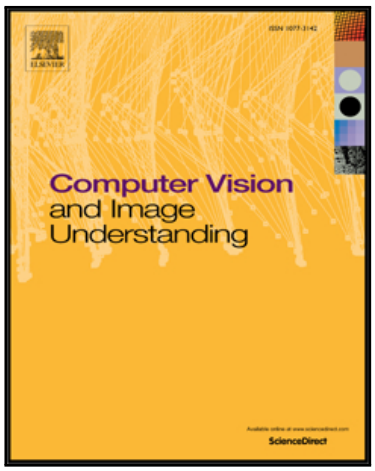


Haze Visibility Enhancement: A Survey and Quantitative Benchmarking

Yu Li, Shaodi You, Michael S. Brown, Robby T. Tan

PII: S1077-3142(17)30159-5
DOI: 10.1016/j.cviu.2017.09.003
Reference: YCVIU 2616

To appear in: *Computer Vision and Image Understanding*

Received date: 5 August 2016
Revised date: 27 July 2017
Accepted date: 17 September 2017

Please cite this article as: Yu Li, Shaodi You, Michael S. Brown, Robby T. Tan, Haze Visibility Enhancement: A Survey and Quantitative Benchmarking, *Computer Vision and Image Understanding* (2017), doi: 10.1016/j.cviu.2017.09.003

Haze Visibility Enhancement: A Survey and Quantitative Benchmarking

Yu Li[a], Shaodi You[b], Michael S. Brown[c], Robby T. Tan[d]

[a]*Advanced Digital Sciences Center, Singapore*
[b]*Data61-CSIRO, Australia and Australian National University, Australia*
[c]*York University, Canada*
[d]*Yale-NUS College and National University of Singapore*

Abstract

This paper provides a comprehensive survey of methods dealing with visibility enhancement of images taken in hazy or foggy scenes. The survey begins with discussing the optical models of atmospheric scattering media and image formation. This is followed by a survey of existing methods, which are categorized into: multiple image methods, polarizing filter-based methods, methods with known depth, and single-image methods. We also provide a benchmark of a number of well-known single-image methods, based on a recent dataset provided by Fattal [1] and our newly generated scattering media dataset that contains ground truth images for quantitative evaluation. To our knowledge, this is the first benchmark using numerical metrics to evaluate dehazing techniques. This benchmark allows us to objectively compare the results of existing methods and to better identify the strengths and limitations of each method.

Keywords: Scattering media, visibility enhancement, dehazing, defogging

1. Introduction

Fog and haze are two of the most common real-world phenomena caused by atmospheric particles. Images captured in foggy and hazy scenes suffer from noticeable degradation of visibility and significant reduction of contrast, as shown in Figure 1. To visually recover scenes from haze or fog can be critical for image processing and computer vision algorithms. Haze-free photographs with clear visual content are what consumers desired when shooting target objects or landscapes; hence, cameras or image-editing softwares that can recover scenes from haze or fog are useful for consumer markets. In addition, many computer vision systems, particularly those for outdoor scenes (e.g., surveillance, intelligent vehicle systems, remote sensing systems), assume clear scenes under good weather. This is because the underlying algorithms, such as object detection, tracking, segmentation, optical flow, obstruction detection, stereo vision are designed with such an assumption. However, mist, fog, and haze are natural phenomena that are inevitable and thus have to be resolved.

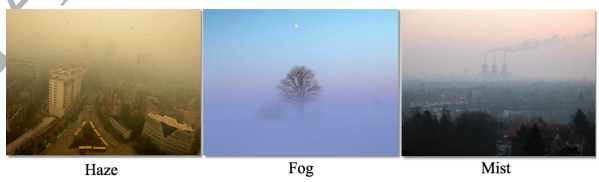


Figure 1: Several examples of images showing the visual phenomena of atmospheric particles. Most of them exhibit significant visibility degradation.

Therefore, addressing this problem is of practical importance.

The degradation in hazy and foggy images can be physically attributed to floating particles in the atmosphere that absorb and scatter light in the environment [2]. This scattering and absorption reduce the direct transmission from the scene to the camera and add another layer of the scattered light, known as airlight [3]. The attenuated direct transmission causes the intensity from the scene to be weaker, while the airlight causes the appearance of the scene to be washed out.

In the past two decades, there has been significant progress in methods that use images taken in hazy scenes. Early work by Cozman and Krotkov [4] and Nayar and Narasimhan [5, 6] uses atmospheric cues to estimate depth. Since then, a number of methods

Email addresses: yul@illinois.edu (Yu Li), shaodi.you@anu.edu.au (Shaodi You), mbrown@eecs.yorku.ca (Michael S. Brown), robby.tan@nus.edu.sg (Robby T. Tan)

have been introduced to explicitly enhance visibility, which can be categorized into ([7, 8]): multi-image-based methods (e.g., [9, 10, 11, 12]), polarizing filter-based methods (e.g., [13, 14]), methods using known depth or geometrical information (e.g., [15, 16, 17, 18]), and single-image methods (e.g., [8, 19, 20, 7, 21, 22, 23, 24, 1, 25, 26]).

Chronologically, Oakley and Satherley's study [15], published in 1998, was the pioneer in proposing a method dealing with poor visibility conditions. The method, however, requires known geometrical information. In 2000, Narasimhan and Nayar's [9] introduced a method that uses multiple images to solve the ill-posedness nature of the problem. It assumes the images are taken under different atmospheric conditions – that is when taking the input images, we need to wait for some time until the fog or haze density levels change, which is impractical for many applications. Subsequently in 2001, polarizing filter-based methods were proposed ([13]). This approach can resolve the problem, since we do not need to wait for atmospheric conditions to change when taking the input images. However, it assumes that the scene is static when the filter is rotated, which still poses problems for real-time applications. More importantly, these two approaches cannot process a single input image. To address this problem, an approach that uses a single input image with additional depth constraints was introduced in 2007 [17]. Motivated by these problems, in 2008, two methods based on a single input image without known geometrical information were proposed [8, 19]. From that point forward, many single input approaches were proposed to address this problem.

In this paper, one of the contributions is to provide a detailed survey on dehazing methods. Our survey provides a holistic view of most of the existing methods. After starting with a brief introduction of the atmospheric scattering optics in Section 2, in Section 3 we provide the survey, where a particular emphasis is placed on the last category of single-image methods, reflecting the recent progress in the field. As part of this survey, we also provide a quantitative benchmarking of a number of the single-image methods. Obtaining quantitative results is challenging as it is difficult to capture ground truth examples for where the same scene has been imaged with and without scattering particles. The work by Fattal [1] synthesizes a dataset by using natural images which associate depth maps that can be used to simulate the spatially varying attenuation in haze and fog images. We have generated an additional dataset using a physically based rendering to simulate environments with scattered particles. Section 4 provides the results of the different methods using both Fattal's dataset [1] and our newly generated benchmark dataset. Our paper is concluded in Section 5 with a discussion on the current state of image dehazing methods and the findings from the benchmark results. In particular, we discuss current limitations with existing methods and possible avenues for research for future methods.

Table 1: Weather condition and the particle type, size, and density [27].

Weather	Particle type	Particle radius (μm)	Density (cm^{-3})
Clean air	Molecule	10^{-4}	10^{19}
Haze	Aerosol	$10^{-2} - 1$	$10 - 10^{3}$
Fog	Water droplets	$1 - 10$	$10 - 100$

2. Atmospheric Scattering Model

Haze is a common atmospheric phenomenon resulting from air pollution, such as dust, smoke, and other dry particles that obscure the clarity of the sky. Sources for haze particles include farming, traffic, industry, and wildfire. As listed in Table 1, the particle size varies from 10^{-2} to $1\mu m$ and the density varies from 10 to 10^3 per cm^3. The particles cause visibility degradation and also color shift. Depending on the view-angle with respect to the sun and the types of the particles, haze may appear brownish or yellowish [28].

Unlike haze, fog or mist is caused by water droplets and/or ice crystals suspended in the air close to the earth's surface [29]. As listed in Table 1, the particle size varies from 1 to $10\mu m$ and the density varies from 10 to 100 per cm^3. Generally, fog particles do not have their own color, and thus their color appearance depends mostly on the surrounding light colors.

In this section, we review the derivation of the optical model for haze or fog, which is known as Koschmieder's law [2]. Discussing the derivation is necessary to understand the physics behind the model. The discussion is based on Narasimhan and Nayar [9] and McCartney [3].

2.1. Optical Modeling

As illustrated in Figure 2(a), when a ray of light hits a particle, the particle will scatter the light to all directions, with magnitudes depending on the particle's size, shape, and incident light wavelengths. Since the directions of scattered rays are moving away from the particle, they are known as outbound rays or out-scattering

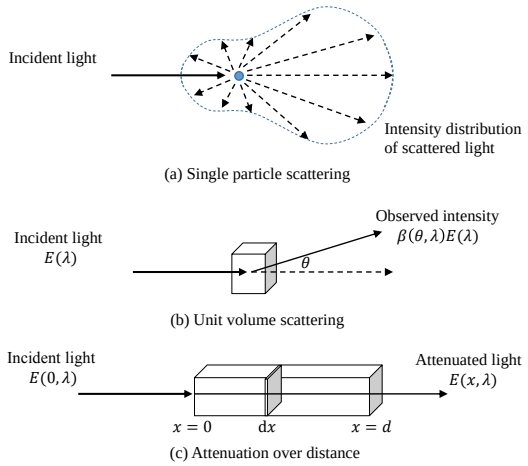


Figure 2: (a) Single particle scattering; (b) unit volume scattering; and (c) light attenuation over distance raised by scattering.

rays. The rays arriving from all directions that hit a particle are referred to as inbound rays or in-scattering rays. As well exploited by Minnaert [30], for a given particle type and incident light wavelength, the outbound light intensity can be modeled as a function between the angle of inbound and outbound light. In this paper, we are more interested in the statistical properties over a large number of particles. Thus, considering the particle density (Table 1), and that each particle can be considered as an independent particle, we can have the statistical relationship between inbound light intensity E and outbound light intensity I [3]):

$$I(\theta,\lambda) = \beta_{p,\mathbf{x}}(\theta,\lambda)E(\lambda), \tag{1}$$

where $\beta_{p,\mathbf{x}}(\theta,\lambda)$ is called the *angular scattering coefficient*. The subindices of β are defined with p indicating its dependency on particle type and density, and $\mathbf{x}$ indicates the dependency spatially. By integrating Eq. (1) over all spherical directions, we obtain the *total scattering coefficient*:

$$I(\lambda) = \beta_{p,\mathbf{x}}(\lambda)E(\lambda). \tag{2}$$

Direct Transmission If we assume a particle medium consists of a small chunk with thickness $\mathrm{d}x$, and a parallel light ray passes through every sheet, as illustrated in Figure 2(c), then the change in irradiance at location x is expressed as:

$$\frac{\mathrm{d}E(x,\lambda)}{E(x,\lambda)} = -\beta_{p,x}(\lambda)d. \tag{3}$$

Integrating this equation between $x = 0$ and $x = d$[1] gives us: $E(d,\lambda) = E_0(\lambda)e^{-\beta(\lambda)dx}$, where E_0 is the irradiance. This formula is known as the Beer-Lambert law.

For non-parallel rays of light, which occur more commonly for outdoor light, factoring in the inverse square law the equation becomes:

$$E(d,\lambda) = \frac{I_0(\lambda)e^{-\beta(\lambda)d}}{d^2}, \tag{4}$$

where I_0 is the intensity of the source, assumed to be a point [3]. Moreover, as mentioned in [6], for overcast sky illumination, the last equation can be written as:

$$E(d,\lambda) = \frac{gL_\infty(\lambda)\rho(\lambda)e^{-\beta(\lambda)d}}{d^2}, \tag{5}$$

where L_∞ is the light intensity at the horizon, ρ is the reflectance of a scene point, and g is the camera gain (assuming the light has been captured by a camera).

Airlight As illustrated in Figure 3(c), besides light from a source (or reflected by objects) that passes through the medium and is transmitted towards the camera, there is environmental illumination in the atmosphere scattered by the same particles also towards the camera. The environmental illumination can be generated by direct sunlight, diffuse skylight, light reflected from the ground, and so on. This type of scattered light captured in the observer's cone of vision is called airlight [3].

Denote the light source as $I(x,\lambda)$. Following the unit volume scattering equation (Eq. (2) and Eq. (3)), we have:

$$\mathrm{d}I(x,\lambda) = \mathrm{d}Vk\beta_{p,\mathbf{x}}(\lambda), \tag{6}$$

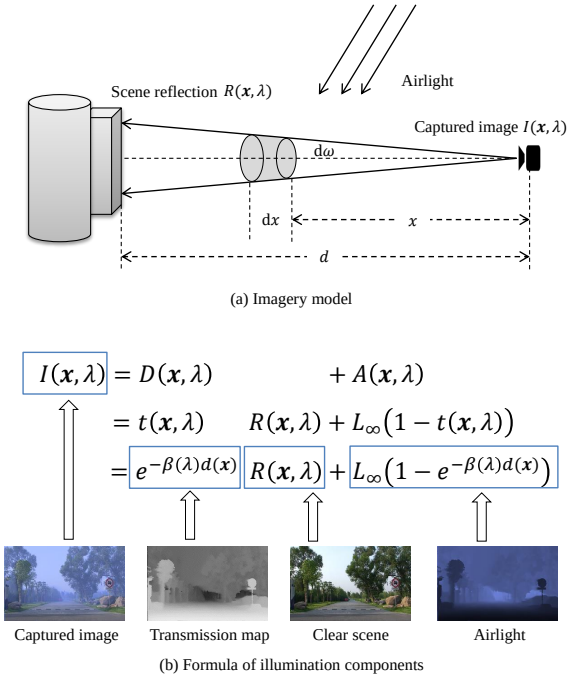
where $\mathrm{d}V = \mathrm{d}\omega x^2$ is a unit volume in the perspective cone. $k\beta_{p,\mathbf{x}}(\lambda)$ is the total scattering coefficient. k is a constant representing the environmental illumination along the camera's line of sight. As with the mechanism for direct transmission in Eq. (4), this light source $\mathrm{d}I$ passes through a small chunk of particles, and the outgoing light is expressed as:

$$\mathrm{d}E(x,\lambda) = \frac{\mathrm{d}I(x,\lambda)e^{-\beta(\lambda)x}}{x^2}, \tag{7}$$

where x^2 is due to the inverse square law of non-parallel rays of light. Therefore, the total radiance at distance d from the camera can be obtained by integrating $\mathrm{d}L = \frac{\mathrm{d}E}{\mathrm{d}\omega}$:

$$L(d,\lambda) = \int_{x=0}^{x=d}\frac{\mathrm{d}E}{\mathrm{d}\omega} = \int_{x=0}^{x=d}\frac{\mathrm{d}I(x,\lambda)e^{-\beta(\lambda)x}}{\mathrm{d}\omega x^2}. \tag{8}$$

[1] We use d for differential and italic d for depth.



(a) Imagery model

(b) Formula of illumination components

Figure 3: Visibility degradation problem in computer vision and computational imaging. (a) Imagery model: with the existence of atmospheric scattering media, light captured by a perspective camera has two components: one is the scene reflection attenuated by the scattering media (direct transmission); the other is the airlight (sunlight, diffused skylight and diffused ground light) scattered by media. (b) Formula and visual example of illumination components. Images are from [23].

Then, based on Eq. (6) and assuming the particles are uniform across the scene (i.e., $\beta_{p,\mathbf{x}}(\lambda) = \beta(\lambda)$), we can express:

$$L(d,\lambda) = k\beta(\lambda)\int_{x=0}^{x=d} e^{-\beta(\lambda)x}\mathrm{d}x \quad (9)$$

$$= k\left(1 - e^{-\beta(\lambda)d}\right). \quad (10)$$

By definition k is the environmental illumination, which in the case of outdoor foggy scenes, is the skylight (L_∞), and thus:

$$L(d,\lambda) = L_\infty\left(1 - e^{-\beta(\lambda)d}\right). \quad (11)$$

This equation is the model of airlight.

Image Formation As illustrated in Figure 3(b), by combining the direct transmission (Eq. (5)) and airlight (Eq. (11)) and assuming that the incoming light intensity to a camera is linearly proportional to the camera's pixel values, the scattered light in the atmosphere captured by the camera can be modeled as:

$$\mathbf{I}(\mathbf{x}) = \mathbf{L}\boldsymbol{\rho}(\mathbf{x})e^{-\beta d(\mathbf{x})} + \mathbf{L}_\infty(1 - e^{-\beta d(\mathbf{x})}). \quad (12)$$

The first term is the direct transmission, and the second term is the airlight. The model is known as Koschmieder's law [2]. The term $\mathbf{I}$ is the image intensity as an RGB color vector,[2] while $\mathbf{x}$ is the 2D image spatial location. The term $\mathbf{L}_\infty$ is the atmospheric light that is assumed to be globally constant and independent from location $\mathbf{x}$. The term $\mathbf{L}$ represents the atmospheric light, the camera gain, and the squared distance, $\mathbf{L} = \mathbf{L}_\infty g/d^2$. The term $\boldsymbol{\rho}$ is the reflectance of an object, β is the atmospheric attenuation coefficient, and d is the distance between an object and the camera. The term β is assumed to be independent from wavelengths, which is a common assumption as we are dealing with particles whose size is larger compared with the wavelength of light, such as, fog, haze, and aerosol [3]. Moreover, β is independent from the spatial image location for homogeneous distribution of atmospheric particles.

In this paper, we denote scene reflection as:

$$\mathbf{R}(\mathbf{x}) = \mathbf{L}\boldsymbol{\rho}(\mathbf{x}). \quad (13)$$

The estimation of Eq. 13 terms is the ultimate goal of dehazing or visibility enhancement, since these terms represent the scene that has not been affected by medium-sized scattered particles. The term $\mathbf{A}(\mathbf{x})$ represents the airlight, and can be denoted as:

$$\mathbf{A}(\mathbf{x}) = \mathbf{L}_\infty(1 - e^{-\beta d(\mathbf{x})}). \quad (14)$$

The function $t(\mathbf{x})$ represents the transmission, as $t(\mathbf{x}) = e^{-\beta d(\mathbf{x})}$. Hence, the scattering model in Eq. (12) can be written as:

$$\mathbf{I}(\mathbf{x}) = \mathbf{D}(\mathbf{x}) + \mathbf{A}(\mathbf{x}), \quad (15)$$

where $\mathbf{D}(\mathbf{x}) = \mathbf{R}(\mathbf{x})t(\mathbf{x})$, the direct transmission.

The above scattering model assumes the images are three channel RGB images. For gray images, we can write a similar formula by transforming the color vectors to scalar variables:

$$I(\mathbf{x}) = D(\mathbf{x}) + A(\mathbf{x}), \quad (16)$$

where $D(\mathbf{x}) = t(\mathbf{x})R(\mathbf{x})$, with $R(\mathbf{x}) = L\rho(\mathbf{x})$, and, $A(\mathbf{x}) = L_\infty(1 - e^{-\beta d(\mathbf{x})})$.

3. Survey on Dehazing Methods

The general goal of dehazing is to recover the clear scene reflection $\mathbf{R}$ (and transmission t, atmosphere light

[2] That is to say we have three sets of equations for wavelength λ at red, green and blue channel separately. The bold fonts indicate this color vector.

color $\mathbf{L}_\infty$) from input $\mathbf{I}$. It is an ill-posed problem as it requires one to infer many unknown parameters from only one equation. In order to make the problem plausible to solve, other information is required. Many early methods propose to use multiple images (e.g., [9]) or use information from other modalities (e.g., depth [18]) to dehaze the images. Compared with dehazing with multiple images as input, single-image dehazing is more challenging. A milestone in single-image dehazing was made with the concurrent publications of Tan [8] and Fattal [19] that propose methods that can automatically dehaze a single image without additional information, such as known geometrical information. These two methods are based on their observations of the characteristics of the hazy and clean images. These characteristics are used as image priors to solve the dehazing problem. Following this trend, different haze-related priors (including the well-received dark channel prior [20]) were proposed and single-image dehazing became the dominant research topic in the field. Recently, a number of methods attempted to use learning frameworks [24, 25] to solve the single-image haze removal problem and demonstrated good results.

As listed in Table 2, we group these methods into four categories according to the inputs([7]): (1) multi-image-based dehazing, (2) polarizing filter-based dehazing, (3) dehazing using known depth, and (4) single-image dehazing. The multi-image category contains all methods that use more than one input image. The polarization-filter category contains all methods that utilize polarizing filters in their methods. While it uses multiple images, the images in this category carry different information from that of the raw multi-image category. Images obtained through a polarizing filter with different polarizing angles have different degrees of polarization. The third category focuses on methods that use a single image and additional geometrical information as their inputs. The fourth category includes methods using a single input image without any additional information. Since it has received the greatest attention recently in the computer vision community, the discussion on this category makes up the largest portion of our survey.

Early Work in Depth Estimation. **Cozman-Krotkov 1997** [4] is one of the earliest methods to analyze images of scenes captured in scattering media. The goal in this work is to extract scene depth by exploiting the presence of the atmospheric scattering effects. This work inspired **Nayar-Narasimhan 1999** [5], who proposed a few methods to estimate depth from hazy scenes. Unlike [4], however, this work does not assume that the haze-free image is provided. While these two methods [4, 5] are pioneers in dealing with atmospheric particles, they are not dehazing methods.

3.1. Multiple Images

Narasimhan-Nayar 2000 [9] extends the analysis of the dichromatic scattering model of [5], which is described as:

$$\mathbf{I}(\mathbf{x}) = p(\mathbf{x})\hat{\mathbf{D}}(\mathbf{x}) + q(\mathbf{x})\hat{\mathbf{A}}(\mathbf{x}), \tag{17}$$

where $\hat{\mathbf{D}}$ and $\hat{\mathbf{A}}$ are the chromaticity values of the direct transmission and the airlight. The terms p and q are the magnitude of the direct transmission and the airlight, respectively. The paper calls the equation the *dichromatic scattering model*, where the word 'dichromatic' is borrowed from [31] due to the similarity of the models.

The method uses multiple images of the same scene taken in different haze density. It works by supposing there are two images taken from the same scene, which share the same color of atmospheric light but have different direct transmission colors. From this, two planes can be formed in the RGB space that intersect each other. In their work [9] utilizes the intersection to estimate the atmospheric light chromaticity, $\hat{\mathbf{A}}$, which is similar to Tominaga and Wandell's method [32] for estimating a light color from specular reflection. The assumption that the images of the same scene have different colors of direct transmission, however, might produce inaccurate estimation since, in many cases, the colors of the direct transmission of the same scene are similar.

The method then introduces the concept of iso-depth, which is the ratio of the direct transmission magnitudes under two different weather conditions. Referring to Eq. (17), and applying it to two images, we have:

$$\frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} = \frac{L_{\infty 2}}{L_{\infty 1}} e^{-(\beta_2-\beta_1)d(\mathbf{x})}, \tag{18}$$

where p is the magnitude of the direct transmission. From this equation, we can infer that if two pairs of pixels have the same ratio, then they must have the same depth: $\frac{p_2(\mathbf{x}_i)}{p_1(\mathbf{x}_i)} = \frac{p_2(\mathbf{x}_j)}{p_1(\mathbf{x}_j)}$. To calculate these ratios, the method provides a solution by utilizing the analysis of the planes formed in the RGB space by the scattering dichromatic model in Eq. (17).

Having obtained the ratios for all pixels, the method proceeds with the estimation of the scene structure, which is calculated by:

$$(\beta_2 - \beta_1)d(\mathbf{x}) = \log\left(\frac{L_{\infty 2}}{L_{\infty 1}}\right) - \log\left(\frac{p_2(\mathbf{x})}{p_1(\mathbf{x})}\right). \tag{19}$$

Table 2: An overview of existing works on vision through atmospheric scattering media.

Method	Category	Known parameters (input)	Estimation (output)	Key idea
Nayar – Narasimham 2000	Multi-images	Two RGB images **I**(**x**) with different weather conditions β_1, β_2	t(**x**), d(**x**)	Iso – depth: comparing different β; color decomposition
Nayar – Narasimham 2003a	Multi-images	Two grayscale or RGB images **I**(**x**) with different weather conditions β_1, β_2	t(**x**), d(**x**), **A**(**x**) and scene reflection **R**(**x**)	Iso – depth
Caraffa-Tarel 2012	Multi-images	Stereo images	d(**x**), **R**(**x**)	Depth from scattering; depth from stereo; spatial smoothness
Li et al. 2015	Multi-images	Monocular video	t(**x**), d(**x**), **R**(**x**)	Depth from monocular video; depth from scattering; photoconsistency
Schechner et al. 2001	Polarizing filter	Two images with different polarization under same weather condition Image with sky region presented	**A**(**x**), t(**x**), d(**x**), **R**(**x**)	Assuming direct transmission **D**(**x**) has insignificant polarization
Schartz et al. 2006	Polarizing filter	Two images with different polarization under same weather condition Image with sky region presented	**A**(**x**), t(**x**), d(**x**), **R**(**x**)	Direct transmission **D**(**x**) has insignificant polarization; **A**(**x**) and **D** (**x**) are statistically independent
Oakley – Satherley 1998	Known depth	Single grayscale image I(**x**) Depth d(**x**)	Atmospheric light: $\boldsymbol{L}_\infty$ Scattering coefficient: β R(**x**)	Mean square optimization; color of the scene is uniform
Nayar – Narasimham 2003b Method 1	Known depth	Single RGB image **I**(**x**) User specified less hazed and more hazed regions	**R**(**x**)	Dichromatic model
Nayar – Narasimham 2003b Method 2	Known depth	Single RGB image **I**(**x**) User specified vanishing point, min depth and max depth	**R**(**x**)	Dichromatic model
Hautière et al. 2007	Known depth	Single image **I**(**x**) Scene of flat ground	$\boldsymbol{L}_\infty$, **R**(**x**)	Depth from calibrated camera
Kopf et al. 2008	Known depth	Single image **I**(**x**) Known 3D model	t(**x**), **R**(**x**)	Transmission estimation using averaged texture from same depth
Tan 2008	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$, t(**x**), **R**(**x**)	Brightest value assumption for atmospheric light L_∞ estimation; maximal contrast assumption for scene reflection R(x) estimation
Fattal 2008	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$, t(**x**), **R**(**x**)	Shading and transmission are locally and statistically uncorrelated
He et al. 2009	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$, t(**x**), **R**(**x**)	Dark channel: outdoor objects in clear weather have at least one color channel that is significantly dark
Tarel – Hautière 2009	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$, t(**x**), **R**(**x**)	Maximal contrast assumption; normalized air light is upper-bounded
Kratz – Nishino 2009	Single image	Single RGB image **I**(**x**)	t(**x**), **R**(**x**)	Scene reflection **R**(**x**) and airlight **A**(**x**) are statistically independent; layer separation
Ancuti-Ancuti 2010	Single image	Single RGB image **I**(**x**)	**A**(**x**), **R**(**x**)	Gray-world color constancy; global contrast enhancement
Meng et al. 2013	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$**, t (x), R(x)**	Dark channel for transmission t(**x**)
Tang et al. 2014	Single image	Single RGB image **I**(**x**)	t (**x**), **R**(**x**)	Learning for transmission t(**x**)
Fattal 2014	Single image	Single RGB image **I**(**x**)	$\boldsymbol{L}_\infty$**, t (x), R(x)**	Color line: small image patch has uniform color and depth but different shading
Cai et al. 2016	Single image	Single RGB image **I**(**x**)	t (**x**), **R**(**x**)	Learning of t(**x**) in CNN framework
Berman et al. 2016	Single image	Single RGB image **I**(**x**)	t (**x**), **R**(**x**)	Non-local haze line; finite color approximation

To be able to estimate the depth, the last equation requires the knowledge of the values of $L_{\infty 1}$ and $L_{\infty 2}$, which are obtained by solving the equation:

$$c(\mathbf{x}) = L_{\infty 2} - \frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} L_{\infty 1}, \tag{20}$$

where c is the magnitude of a vector indicating the distance between the origin of $\mathbf{I}_1$ to the origin of $\mathbf{I}_2$ in the direction of the airlight chromaticity in RGB space, while $\frac{p_2(\mathbf{x})}{p_1(\mathbf{x})}$ is the ratio, which had been computed.

For the true scene color restoration, employing the estimated atmospheric light, the method computes the airlight magnitude of Eq. (17) using:

$$q(\mathbf{x}_i) = L_\infty \left(1 - e^{-\beta d(\mathbf{x}_i)}\right), \tag{21}$$

where:

$$\beta d(\mathbf{x}_i) = \beta d(\mathbf{x}_j) \left(\frac{d(\mathbf{x}_i)}{d(\mathbf{x}_j)}\right), \tag{22}$$

and $\frac{d(\mathbf{x}_i)}{d(\mathbf{x}_j)}$ is computable using Eq. (19). $\beta d(\mathbf{x}_j)$ is a chosen reference point. This is obtained by assuming there is at least one pixel in the image for which the true value of the direct transmission, $\mathbf{D}$, is known (e.g., a black object), since, in this case $\mathbf{I}(\mathbf{x}) = \mathbf{A}(\mathbf{x})$, and $\beta d(\mathbf{x})$ can be directly computed. The method also proposes how to find such a pixel automatically. Note that knowing the value of $q(\mathbf{x}_i)$ in Eq. (21) enables us to dehaze the images straightforwardly.

Narasimhan-Nayar 2003 In a subsequent publication, Narasimhan and Nayar [10] introduce a technique that works for gray or colored images: contrast restoration of iso-depth regions, atmospheric light estimation, and contrast restoration.

In the contrast restoration of iso-depth regions, the method forms an equation that assumes the depth segmentation is provided (e.g., manually by the user) and the atmospheric light is known:

$$\rho(\mathbf{x}_i) = 1 - \left(\sum_j 1 - \sum_j \rho(\mathbf{x}_j)\right) \frac{L_\infty - I(\mathbf{x}_i)}{\sum_j (L_\infty - I(\mathbf{x}_j))}, \tag{23}$$

where the sums are over the same depth regions. As can be seen in the equation, $\rho(\mathbf{x}_i)$ can be estimated up to a linear factor $\sum_j \rho(\mathbf{x}_j)$. By setting $\rho^{min} = 0$ and $\rho^{max} = 1$ and adjusting the value $\sum_j \rho(\mathbf{x}_j)$, the contrast of regions with the same depth can be restored.

To estimate the atmospheric lights, the method utilizes two gray images of the same scene yet different atmospheric lights. Based on the scattering model in Eq. (12), scene reflectance ρ is eliminated. The two equations representing the two images can be transformed into:

$$I_2(\mathbf{x}) = \left[\frac{L_{\infty 2}}{L_{\infty 1}} e^{-(\beta_2 - \beta_1) d(\mathbf{x})}\right] I_1(\mathbf{x}) + \left[L_{\infty 2}\left(1 - e^{-(\beta_2 - \beta_1) d(\mathbf{x})}\right)\right],$$

where indices 1 and 2 indicate image 1 and 2, respectively. From the equation, a two-dimensional space can be formed, where I_1 is the x-axis, and I_2 is the y-axis. In the space, a few pixels will form a line, if those pixels represent objects that have the same depth d yet different reflectance ρ. As a result, if we have different depths, then there will be a few different lines in the space, which intersect at $(L_{\infty 1}, L_{\infty 2})$. The lines representing pixels with the same depth can be detected using the Hough transform. Finally, to restore contrast or to dehaze, the same method as in [9] is used.

Caraffa-Tarel 2012 [11] and later [33] introduce a dehazing method using stereo cameras. The idea is that both airlight and disparity from stereo can indicate the scene depths. Hence, the goal is to jointly estimate the depth and enhance visibility in the stereo images. To achieve this, the authors proposed a cost function for the data term that is a linear combination of the two main log-likelihoods from stereo and fog stereo:

$$E_{data} = \sum_{\mathbf{x}} \alpha E_{data}^{stereo}(\mathbf{x}) + (1 - \alpha) E_{data}^{fog\_stereo}(\mathbf{x}), \tag{24}$$

where $0 \leq \alpha \leq 1$ is the weighting factor, and

$$E_{data}^{stereo}(\mathbf{x}) = \rho\Big(I_L(x, y) - I_R\left(x - \delta(x, y), y\right)\Big), \tag{25}$$

is the standard data term in stereo estimation to measure the intensity constancy between the left-right pair. L, R indicate the left and right views, δ is the stereo disparity, and ρ is a robust function to handle noise and occlusions. The use of E_{data}^{stereo} helps stereo estimation at short distances regardless of whether the clean left image I_{0L} is correctly estimated.

The proposed $E_{data}^{fog\_stereo}$ is composed of two parts:

$$\begin{aligned} &E_{data}^{fog\_stereo}(\mathbf{x}) \\ &= \rho\left(I_{0L}(x, y) e^{-\beta \frac{b}{\delta(x,y)}} + L_\infty (1 - e^{-\beta \frac{b}{\delta(x,y))}}) - I_L(x, y)\right) \\ &+ \rho\left(I_{0L}(x, y) e^{-\beta \frac{b}{\delta(x,y)}} + L_\infty (1 - e^{-\beta \frac{b}{\delta(x,y))}}) - I_R\left(x - \delta(x, y), y\right)\right), \end{aligned} \tag{26}$$

where b relates to stereo calibration parameters. The first part enforces the consistency with the imaging model and the second part is the stereo photometric consistency term that takes into account the haze effect.

Aside from the data terms, the method utilizes prior terms, which are basically the spatial smoothness term

for the estimated disparity δ and the estimated clean left image I_{0L}. The optimization to estimate the two variables δ and I_{0L} is done in a two-step fashion that in each time only one of the variables is optimized, with the other one fixed and then alternate. After a few iterations, it will converge with the solution of δ and I_{0L}.

Li et al. 2015 [12] jointly estimates scene depth and enhances visibility in a foggy video, which, unlike Caraffa-Tarel's method [11], uses a monocular video. Following the work of Zhang et al. [34], it estimates the camera parameters and the initial depth of the scene, which is erroneous particularly for dense fog regions due to the photoconsistency problem in the data term. Similar to [11], Li et al.'s method [12] introduces a photoconsistency data term that involves effect of fog:

$$E_p(d_n) = \frac{1}{|\mathcal{N}(n)|} \sum_{n' \in \mathcal{N}(n)} \sum_{\mathbf{x}} \|\hat{I}_{n'}(\mathbf{x}) - I_{n'}(l_{n \to n'}(\mathbf{x}, d_n(\mathbf{x})))\|,$$

where $l_{n \to t'}(\mathbf{x}, d_n(\mathbf{x})$ projects the pixel x with inverse depth $d_n(\mathbf{x})$ in frame n to frame n'. The intensity, $\hat{I}_{n'}(\mathbf{x}) = (I_n(\mathbf{x}) - L_\infty)\frac{\pi_{n \to n'}(\mathbf{x}, t_n(\mathbf{x}))}{t_n(\mathbf{x})} + L_\infty$, is a synthetic intensity value obtained from the transmission, t_n, which is computable by knowing d_n (note that, in the paper, the scattering coefficient β and the atmospheric light, L_∞, are estimated separately). The projection function $\pi_{n \to n'}(\mathbf{x}, t_n(\mathbf{x}))$ computes the corresponding transmission in the n'-th frame for the pixel $\mathbf{x}$ in the n-th frame with transmission $t_n(\mathbf{x})$. The denominator $\mathcal{N}(t)$ represents the neighboring frames of frame n and $|\mathcal{N}(n)|$ is the number of neighboring frames. By having $\beta(\mathbf{x})$ estimated separately, $t_n(\mathbf{x})$ depends only on $d_n(\mathbf{x})$, and thus d_n is the only unknown in the last equation. The whole idea in the photoconsistency term here is to generate a synthetic intensity value of each pixel from known depth, d, atmospheric light, L_∞, and the particle scattering coefficient, β. Note that the paper assumes β and L_∞ are uniform across the video sequence. Therefore, if those three values are correctly estimated, the generated synthetic intensity values must be correct.

Aside from the photoconsistency term, the method also uses Laplacian smoothing as the transmission smoothness prior. Together with the geometric coherent term and disparity smoothness term, the problem is formulated in a Markov Random Field (MRF) for dense image labeling. After a few iterations, the outcomes are estimated depth maps and defogged images.

3.2. Polarizing Filter

Schechner et al.2001 addresses the issue appearing in the work of Narasimhan and Nayar [9], where it requires at least two images of the same scene taken under different particle densities (i.e., we have to wait until the fog density changes considerably). Unlike [9], Schecher et al.'s [13] uses multiple images captured using polarizing filters, which does not require the fog density to change.

The main assumption employed in this polarized-based method is that the direct transmission has insignificant polarization, and thus the polarization of the airlight dominates the observed light. Based on this, the maximum intensity occurs when airlight passes through the filter. This can be obtained when:

$$I^{\max}(\mathbf{x}) = D(\mathbf{x})/2 + A^{\max}(\mathbf{x}), \tag{27}$$

where D and A are the direct transmission and the airlight, respectively. The minimum intensity (i.e., when the filter can block the airlight at its best) is when:

$$I^{\min}(\mathbf{x}) = D(\mathbf{x})/2 + A^{\min}(\mathbf{x}). \tag{28}$$

Adding up the two states of the polarization, we obtain: $I(\mathbf{x}) = I^{\max}(\mathbf{x}) + I^{\min}(\mathbf{x})$. Based on this, the method estimates the atmospheric light from a sky region and computes its degree of polarization:

$$P = \frac{L_\infty^{\max} - L_\infty^{\min}}{L_\infty^{\min} + L_\infty^{\max}}, \tag{29}$$

and then estimates the airlight for every pixel:

$$A(\mathbf{x}) = \frac{I^{\max}(\mathbf{x}) - I^{\min}(\mathbf{x})}{P}. \tag{30}$$

Based on the airlight, the method computes the transmission: $e^{-\beta d(\mathbf{x})} = 1 - \frac{A(\mathbf{x})}{L_\infty}$, and finally obtains the dehazing result $R(\mathbf{x}) = [I(\mathbf{x}) - A(\mathbf{x})]\, e^{\beta d(\mathbf{x})}$. To obtain the maximum and the minimum intensity values, the filter needs to be rotated either automatically or manually.

Shwartz et al.2006 [14] uses the same setup proposed by Schechner et al.'s [13] but removes the assumption that sky regions are present in the input image. Instead, this method estimates the color of the airlight and of the direct transmission by applying independent component analysis (ICA):

$$\begin{bmatrix} A \\ D \end{bmatrix} = \mathbb{W} \begin{bmatrix} I^{max} \\ I^{min} \end{bmatrix} \tag{31}$$

$$\mathbb{W} = \begin{bmatrix} 1/P & -1/P \\ (P-1)/P & (P+1)/P \end{bmatrix}. \tag{32}$$

In this case, the challenge lies in estimating $\mathbb{W}$ given $[I^{max}, I^{min}]^T$ to produce D and A accurately.

The method claims that while the airlight and direct transmission are in fact statistically dependent there are transformations that can relax this dependency. The method therefore transforms the input data using a wavelet transformation and solves the ICA problem by using an optimization method in the wavelet domain. Aside from P, the method also needs to estimate L_∞, which is done by labeling certain regions manually to have two pixels that have the same values of the direct transmission yet different values of the airlight.

3.3. Known Depth

Oakley-Satherley 1998 [15] is one of the early methods dealing with visibility enhancement in a single foggy image. The enhancement is done in two stages: parameter estimation followed by contrast enhancement. The basic idea of the parameter estimation is to employ the sum of squares method to minimize an error function, between the image intensity and some parameters of the physical model, by assuming the reflectance of the scene can be approximated by a single value representing the mean of the scene reflectance. With these assumptions, the minimization is done to estimate three global parameters: the atmospheric light (L_∞), the mean reflectance of the whole scene $\bar{\rho}$, and the scattering coefficient, β:

$$Err = \sum_{\mathbf{x}}^{M} \left(I(\mathbf{x}) - L_\infty \left(1 + (\bar{\rho} - 1) e^{-\beta d(\mathbf{x})} \right) \right)^2 . \quad (33)$$

The last equation assumes that $L = L_\infty$. Having estimated the three global parameters by minimizing function Err, the airlight is then computed using:

$$A(\mathbf{x}) = L_\infty (1 - e^{-\beta d(\mathbf{x})}). \quad (34)$$

Consequently, the end result is obtained by computing:

$$R(\mathbf{x}) = \left(L_{max} \left(\frac{I(\mathbf{x}) - A(\mathbf{x})}{L_\infty} e^{\beta d(\mathbf{x})} \right) \right)^{\frac{1}{2.2}} , \quad (35)$$

where L_{max} is a constant depending on the maximum gray level of the image display device, and the power 2.2^{-1} is the gamma correction.

The main drawbacks of this method are the assumption that the depth of the scene is known, and the mean reflectance for the whole image is used in the minimization and in computing the airlight. The latter is acceptable if the color of the scene is somehow uniform, which is not the case for general scenes. Tan and Oakley's [35] extended the work of Oakley and Satherley [15] to handle color images by taking into account a colored scattering coefficient and colored atmospheric light.

Narasimhan-Nayar 2003 [16] proposed several methods based on a single input image that requires some user interaction. The first method requires the user to select a region with less haze and a region with more haze of the same reflection as the first one's. From these the two inputs, the approach estimates the dichromatic plane and dehaze pixels that have the same color as the region with less haze. This method assumes the pixels represent scene points that have the same reflection. The second method asks the user to indicate the vanishing point and to input the maximum and minimum distance from the camera. This information is used to interpolate the distance to estimate the clear scene in between. The interpolation is a rough approximation, since depth can be layered and not continuous. To resolve layered scenes, the third method is introduced, which requires depth segmentation that can be done through satellite orthographic photos of buildings.

Hautière et al.2007 [17] proposes a framework for restoring the contrast of images taken in a vehicle. It first computes the scattering coefficient β and obtains the airlight intensity L_∞ from a calibrated camera using the method presented in [36]. Basically the estimation is based on the relationship of the distance d with each line, y in the image, where the assumption of a flat road:

$$d = \frac{a}{y - y_h}, \quad \text{if } v > v_h, \quad (36)$$

where $a = \frac{H\alpha}{\cos^2 \theta}$. The term H is the height of the camera, y is the y-axis of the image coordinates, θ is the angle between the optical axis of the camera and the horizon line. y_h is the horizon line. The term $\alpha = f/w$, with f as the focal length and w as the length of a pixel.

Once the parameter β and L_∞ are estimated, the remaining issue to restore the scene contrast is to estimate the depth d at each pixel. To relax the flat world assumption in Eq. (36) in handling the vertical objects like trees, vehicles, houses, or any objects in the scene, the method in [17] employs depth heuristics. It proposes a rule to detect the sky region and vanishing point. Then it clips large distances using a fixed parameter c to reduce modeling error:

$$d_1 = \begin{cases} \dfrac{a}{y - y_h} & \text{if } y - y_h > c; \\ \dfrac{a}{c - y_h} & \text{if } 0 < y - y_h \leq c. \end{cases} \quad (37)$$

Another depth heuristic in [16] is used to model the depth of objects not belonging to the road surface:

$$d_2 = \frac{\kappa}{y - y_h} \text{ or } \frac{\kappa}{\sqrt{(y - y_h)^2 + (x - x_h)^2}}, \quad (38)$$

where $\kappa \geq c$. The first heuristic is used to model vertical planes like buildings and the second heuristic is used for modeling cylindrical scenes like rural roads. The two parameters c and κ are obtained in an optimization process with a proposed image quality attribute. The final depth excluding the sky region is estimated as

$$d = \min(d_1, d_2). \tag{39}$$

The method [17] also demonstrated three in-vehicle applications like road scene enhancement using this framework.

Kopf et al.2008 [18] attempt to overcome the dehazing problem by utilizing the information provided by an exact 3D model of the input scene and the corresponding model textures (obtained from Landsat data). The main task is to estimate the transmission, $\exp(-\beta d(\mathbf{x}))$, and the atmospheric light, $\mathbf{L}_\infty$.

Since it has the 3D model of the scene, it can collect the average model texture intensity of certain depths ($\hat{I}_h(\mathbf{x})$) from the Landsat data and the corresponding average haze intensity ($\hat{I}_m(\mathbf{x})$) of the same depths from the input image. The two average intensity values can be used to estimate the transmission assuming $\mathbf{L}_\infty$ is known:

$$t(\mathbf{x}) = \frac{\hat{I}_h - L_\infty}{C\hat{I}_m - L_\infty}, \tag{40}$$

where C is a global correction vector and $C\hat{I}_m$ attempts to substitute R, the scene reflectance without the influence of haze. In this method, C is computed from:

$$C = \frac{F_h}{\mathrm{lum}(F_h)} / \frac{F_m}{\mathrm{lum}(F_m)}, \tag{41}$$

where F_h is the average of $I_h(\mathbf{x})$ with $z < z_F$ with $z_F = 1600\,m$, and F_m is the average of the model texture. The function $\mathrm{lum}(c)$ is the luminance of a color c.

The method suggests that $\mathbf{L}_\infty$ is estimated by collecting the average background intensity for pixels whose depth is more than a certain distance ($> 5000m$) from both the input image and the model texture image.

3.4. Single-Image Methods

Tan 2008 [8] is based on two basic observations: first, images on a clear day have more contrast than images in bad weather; second, the airlight whose variation mainly depends on the depth, tends to be smooth. Given an input image, the method estimates the atmospheric light, $\mathbf{L}_\infty$ from the brightest pixels in the input image, and normalizes the color of the input image, from $\mathbf{I}$ to $\tilde{\mathbf{I}}$ by dividing $\mathbf{I}$ by the chromaticity of $\mathbf{L}_\infty$, element-wise. The chromaticity of $\mathbf{L}_\infty$ is the same as $\hat{\mathbf{A}}$ in Eq. (17). By doing this, the airlight $\mathbf{A}$, can be transformed from color vectors into scalars, A. Hence, the visibility enhancement problem can be solved if we know the scalar value of the airlight, A, for every pixel:

$$e^{\beta d(\mathbf{x})} = \frac{\sum_c \mathbf{L}_{2c}}{A(\mathbf{x}) \sum_c \mathbf{L}_{2c}}, \tag{42}$$

$$\tilde{\mathbf{R}}(\mathbf{x}) = \left(\tilde{\mathbf{I}}(\mathbf{x}) - A(\mathbf{x}) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) e^{\beta d(\mathbf{x})}, \tag{43}$$

where c represents the index of RGB channels, and $\tilde{\mathbf{R}}$ is the light normalized color of the scene reflection, $\mathbf{R}$. The values of A range from 0 to $\sum_c \mathbf{L}_{2c}$. The key idea of the method is to find a value of $A(\mathbf{x})$ from that range that maximizes the local contrast of $\tilde{\mathbf{R}}(x)$. The local contrast is defined as:

$$\mathrm{Contrast}(\tilde{\mathbf{R}}(\mathbf{x})) = \sum_{x,c}^{S} |\nabla \tilde{\mathbf{R}}_c(\mathbf{x})|, \tag{44}$$

where S is a local window whose size is empirically set to 5×5. It was found that the correlation between the airlight and the contrast is convex.

The problem can be cast into an MRF framework and optimized using graphcuts to estimate the values of the airlight across the input image. The method works for both color and gray images and was shown able to handle relatively thick fog. One of the drawbacks of the method is the appearance of halos around depth discontinuity due to the local window-based operation. Another drawback is that when the input regions have no textures, the quantity of local contrast will be constant even when the airlight value changes. Prior to the 2008 publication, Tan et al. [37] introduced a fast single dehazing method that uses a color constancy method [38] to estimate the color of the atmospheric light, and utilizes the Y channel of the YIQ color space as an approximation to dehaze.

Fattal 2008 [19] is based on the idea that the shading and transmission functions are locally and statistically uncorrelated. From this, the work derives the shading and transmission functions from Eq. (12):

$$l^{-1}(\mathbf{x}) = \frac{1 - I_A(\mathbf{x})/\|\mathbf{L}_\infty\|}{} + \frac{\eta}{\|\mathbf{L}_\infty\|}, \tag{45}$$

$$t(\mathbf{x}) = 1 - I_A(\mathbf{x}) - \frac{\eta I_{R'}(\mathbf{x})}{\|\mathbf{L}_\infty\|}, \tag{46}$$

where $l(\mathbf{x})$ is the shading function and $t(\mathbf{x})$ is the trans-

mission function. The I_A and $I_{R'}$ are defined as:

$$I_A(\mathbf{x}) = \frac{\langle \mathbf{I}(\mathbf{x}), \mathbf{L}_\infty \rangle}{||\mathbf{L}_\infty||}, \quad (47)$$

$$I_{R'}(\mathbf{x}) = \sqrt{||\mathbf{I}_x||^2 - I_A^2(\mathbf{x})}. \quad (48)$$

Assuming $\mathbf{L}_\infty$ can be obtained from the sky regions, η is estimated by assuming the shading and the transmission functions are statistically uncorrelated over a certain region Ω. This implies that $C_\Omega(l^{-1}, t) = 0$, where function C_Ω is the sample covariance. Hence, η can be defined based on $C_\Omega(l^{-1}, t) = 0$:

$$\eta(\mathbf{x}) = \frac{C_\Omega\left(I_A(\mathbf{x}), h(\mathbf{x})\right)}{C_\Omega\left(I_{R'}(\mathbf{x}), h(\mathbf{x})\right)}, \quad (49)$$

where $h(\mathbf{x}) = (||\mathbf{L}_\infty|| - I_A(\mathbf{x}))/I_{R'}(\mathbf{x})$. Obtaining the values of $t(\mathbf{x})$ and $\mathbf{L}_\infty$ will eventually solve the estimation of the scene reflection, $\mathbf{R}(\mathbf{x})$.

The success of the method relies on whether the statistical decomposition of shading and transmission can be optimum, and whether they are truly independent. Moreover, while it works for haze, the approach was not tried on foggy scenes.

He et al. 2009. The work in [20, 39] observed an interesting phenomenon of outdoor natural scenes with clear visibility. They found that most outdoor objects in clear weather have at least one color channel that is significantly dark. They argue that this is because natural outdoor images are colorful (i.e.,the brightness varies significantly in different color channels) and full of shadows. Hence, they define a dark channel as:

$$J^{dark} = \min_{y \in \Omega(\mathbf{x})} \left(\min_{c \in \{R,G,B\}} R^c(y) \right). \quad (50)$$

Because of the observation that, $J^{dark} \rightarrow 0$, He et al. [20] refer to this as the *dark channel prior*.

The dark channel prior is used to estimate the transmission as follows. Based on Eq. (12), we can express:

$$\frac{I^c(\mathbf{x})}{L^c_\infty} = t(\mathbf{x}) \frac{R^c(\mathbf{x})}{L^c_\infty} + 1 - t(\mathbf{x}). \quad (51)$$

Assuming that we work on a local patch $\Omega(\mathbf{x})$ and denote the patch's transmission as $\tilde{t}(\mathbf{x})$, the overall objective function can be expressed as:

$$\min_{y \in \Omega(\mathbf{x})} \left(\min_c \frac{I^c(\mathbf{x})}{L^c_\infty} \right) = \tilde{t}(\mathbf{x}) \min_{y \in \Omega(\mathbf{x})} \left(\min_c \frac{R^c(\mathbf{x})}{L^c_\infty} \right) + 1 - \tilde{t}(\mathbf{x}),$$

and consequently, due to the dark channel prior:

$$\tilde{t}(\mathbf{x}) = 1 - \min_{y \in \Omega(\mathbf{x})} \left(\min_c \frac{I^c(\mathbf{x})}{L^c_\infty} \right), \quad (52)$$

where $\mathbf{L}_\infty$ is obtained by picking the top 0.1 % brightest pixels in the dark channel. Finally, to have a smooth and robust estimation of $t(\mathbf{x})$ that can avoid the halo effects due to the use of patches, the method employs the matting Laplacian in [40]. One can interpret the dark channel prior as the maximum possible value of the airlight in a local patch, following [8], since the maximum possible value of the airlight is the minimum over the color components.

Tarel-Hautière 2009 noticed that one drawback of the previous methods [8] [19] [20] [39] is the computation time. These methods cannot be applied for real-time applications, where the depths of the input scenes change from frame to frame. Tarel and Hautière [7] introduce a fast visibility restoration method whose complexity is linear to the number of image pixels. Inspired by the contrast enhancement [8], they observed that the value of the normalized airlight, $A(\mathbf{x})$ (where the illumination color is now pure white), is always less than $W(\mathbf{x})$, where $W(\mathbf{x}) = \min_c(\tilde{I}^c(\mathbf{x}))$. Note that, $\tilde{I}^c$ is the pixel intensity value of color channel c after the light normalization. Since it takes time to find the optimal value of $A(\mathbf{x})$, the idea of estimating $A(\mathbf{x})$ rapidly is based on bounds of the possible airlight values [41]:

$$M(\mathbf{x}) = \text{median}_{\Omega(\mathbf{x})}(W)(\mathbf{x}), \quad (53)$$

$$S(\mathbf{x}) = M(\mathbf{x}) - \text{median}_{\Omega(\mathbf{x})}(|W - M|)(\mathbf{x}), \quad (54)$$

$$A(\mathbf{x}) = \max\left(\min(pS(\mathbf{x}), W(\mathbf{x}), 0)\right), \quad (55)$$

where $\Omega(\mathbf{x})$ is a patch centered at x, and p is a constant value, chosen empirically. The last equation means $0 \leq A(\mathbf{x}) \leq W(\mathbf{x})$. The method develops a special filter named the *median of median along lines* to help produce a smooth airlight estimation, $A(\mathbf{x})$. Following this approach, the work in [41] adds a planar scene assumption to make it dedicated to tackling the road scene cases.

Kratz-Nishino 2009 [42] and later [43] offer a new perspective on the dehazing problem. This work poses the problem in the framework of a factorial MRF [44], which consists of a single observation field (the input hazy image), and two separated hidden fields (the albedo and the depth fields). Thus, the idea of the method is to estimate the depth and albedo by assuming that the two are statistically independent. First, it transforms the model in Eq. (12) to:

$$\log\left(1 - \frac{I^c(\mathbf{x})}{L^c_\infty}\right) = \log(1 - \rho^c(\mathbf{x})) - \tilde{d}(\mathbf{x}), \quad (56)$$

$$\tilde{I}^c(\mathbf{x}) = C^c(\mathbf{x}) + D(\mathbf{x}), \quad (57)$$

where c is the index of the color channel, $C^c(\mathbf{x}) = \log(1 - \rho^c(\mathbf{x}))$, and $D(\mathbf{x}) = -\tilde{d}(\mathbf{x})$, and $\tilde{d}(\mathbf{x}) = \beta d(\mathbf{x})$.

Hence, in terms of the factorial MRF, $\tilde{I}^c$ is the observed field, and C^c and D are the two separated hidden fields. Each node in the MRF will connect to the corresponding node in the observed field and to its neighboring nodes within the same field. The goal is then to estimate the value of C^c for all color channels and the depth, D. The objective function consists of the likelihood and the priors C^c and D. The prior of C^c is based on the exponential power distribution of the chromaticity gradients (from natural images), while the prior of D is manually selected from a few different models, depending on the input scene (e.g., either cityscape or terrain). To solve the decomposition problem, the method utilizes an EM algorithm that decouples the estimation of the two hidden fields. In each step, graphcuts are used to optimize the values, resulting in a high computational cost. To make the iteration more efficient good initializations are required. The initialization for the depth is:

$$D_{init}(\mathbf{x}) = \max_{c\in R,G,B}(\tilde{I}^c(\mathbf{x})), \tag{58}$$

which means the upper bound on the depth value at each pixel is assumed to be corresponding to the maximum of observed RGB color values and the maximum value can be used as the initial estimate of the depth layer [43]. In the Bayesian direction, a different method in [45] is later proposed with a novel MRF model and planar constraint. This approach is able to produce better results, especially on road images.

Ancuti-Ancuti 2010. The methods in [21] [22] propose an approach based on image fusion. The idea is to blend information from two images derived from the input image: a white-balanced image, $\mathbf{I}_1$, by using the gray-world color constancy method [46], and a global contrast enhanced image, $\mathbf{I}_2$, which is calculated by $\mathbf{I}_2(\mathbf{x}) = \gamma(\mathbf{I}(\mathbf{x}) - \bar{\mathbf{I}})$, where $\bar{\mathbf{I}}$ is the average intensity of the whole input image and γ is a weighting factor. From both $\mathbf{I}_1$ and $\mathbf{I}_2$, the weights in terms of the luminance, chromaticity, and saliency are calculated. Based on the weights, the output of the dehazing algorithm is

$$\tilde{w^1}(\mathbf{x})\mathbf{I}_1 + \tilde{w^2}(\mathbf{x})\mathbf{I}_2, \tag{59}$$

where $\tilde{w}^k$ is the normalized weights and the index k is either 1 or 2, such that $w^k(\mathbf{x}) = w_l^k w_c^k w_s^k$ and $\tilde{w}^k = w^k / \sum_{k=1}^{2} w^k$. The subscripts l, c, s represent luminance, chromaticity, and saliency, respectively. The three weights' definitions are as follows:

$$w_l^k(\mathbf{x}) = \sqrt{\frac{1}{3}\sum_{c\in R,G,B}\left(I_c^k(\mathbf{x}) - L^k(\mathbf{x})\right)^2}, \tag{60}$$

$$w_c^k(\mathbf{x}) = \exp\left(-\frac{\left(S^k(\mathbf{x}) - S_{max}^k\right)^2}{2\sigma^2}\right), \tag{61}$$

$$w_s^k(\mathbf{x}) = \|I_\omega^k(\mathbf{x}) - I_\mu^k\|, \tag{62}$$

where $L^k(\mathbf{x})$ is the average of the intensity in the three color channels. The term S is the saturation value (e.g., the saturation in the HSI color space). The term σ is set 0.3 as default. The term S_{max} is a constant, where for the HSI color space, it would be 1. The term I_μ^k is the arithmetic mean pixel value of the input, and I_ω^k is the blurred input image. The method produces good results; however, the reasoning behind using the two images ($\mathbf{I}_1$ and $\mathbf{I}_2$) and the three weights is not fully explained and needs further investigation. The fusion approach was also applied to underwater vision [47].

Meng et al. 2013 [23] extends the idea of the dark channel prior [20] in determining the initial values of transmission, $t(\mathbf{x})$, by introducing its lower bound. According to Eq. (12), $t(\mathbf{x}) = (A^c - I^c(\mathbf{x}))/(A^c - R^c(\mathbf{x}))$. As a result, the lower bound of the transmission, denoted as $t_b(\mathbf{x})$, can be defined as:

$$t_b(\mathbf{x}) = \frac{A^c - I^c(\mathbf{x})}{A^c - C_0^c}, \tag{63}$$

where C_0^c is a small scalar value. Since C_0^c is smaller than or equal to $R^c(\mathbf{x})$, then $t_b(\mathbf{x}) \le t(\mathbf{x})$. To anticipate a wrong estimation of $\mathbf{A}$, such as when the value of A^c is smaller than I^c, the second definition of $t_b(\mathbf{x})$ is expressed as:

$$t_b(\mathbf{x}) = \frac{A^c - I^c(\mathbf{x})}{A^c - C_1^c}, \tag{64}$$

where C_1^c is a scalar value, larger than the possible values of A^c and I^c. Combining the two, we obtain:

$$t_b(\mathbf{x}) = \min\left(\max_{c\in R,G,B}\left(\frac{A^c - I^c(\mathbf{x})}{A^c - C_0^c}, \frac{A^c - I^c(\mathbf{x})}{A^c - C_1^c}\right), 1\right).$$

Assuming the transmission is constant for a local patch, the estimated transmission becomes $\tilde{t}(\mathbf{x}) = \min_{y\in\Omega_x}\max_{z\in\Omega_y} t_b(z)$. The method employs a L_1-based regularization formulation to obtain a more robust and smooth transmission map.

Tang et al. 2014 [24], unlike the previous methods, introduces a learning-based method to estimate the transmission. The method gathers multiscale features, such as dark channel [39], local maximum contrast [8],

hue disparity, and local maximum saturation, and uses the random forest regressor [48] to learn the correlation between the features and the transmission $t(\mathbf{x})$. The features related to the transmission are defined as follows:

$$
\begin{aligned}
F_D(\mathbf{x}) &= \min_{y\in\Omega(\mathbf{x})} \min_{c\in R,G,B} \frac{I^c(y)}{A^c},\\
F_C(\mathbf{x}) &= \max_{y\in\Omega(\mathbf{x})} \sqrt{\frac{1}{3|\Omega(y)|}\sum_{z\in\Omega(y)} \|I(y)-\mathbf{I}(z)\|^2},\\
F_H(\mathbf{x}) &= |\mathrm{H}(I_{si}(\mathbf{x})) - \mathrm{H}(I(\mathbf{x}))|,\\
F_S(\mathbf{x}) &= \max_{y\in\Omega(\mathbf{x})} \left(1 - \frac{\min_c I^c(y)}{\max_c I^c(y)}\right),
\end{aligned}
\tag{65}
$$

where $I_{si} = \max[I^c(\mathbf{x}), 1 - I^c(\mathbf{x})]$. For the learning process, synthetic patches are generated from given haze-free patches, fixed white atmospheric light, and random transmission values, where the haze-free images are taken from the Internet. The paper claims that the most significant feature is the dark channel feature; however, other features also play important roles, particularly when the color of an object is the same as that of the atmospheric light.

Fattal 2014 [1] introduces another approach based on color lines. This method assumes that small image patches (e.g., 7×7) have a uniformly colored surface and the same depth, yet different shading. Hence, the model in Eq. (12) can be written as:

$$
\mathbf{I}(\mathbf{x}) = l(\mathbf{x})\hat{\mathbf{R}} + (1-t)\mathbf{L}_\infty, \tag{66}
$$

where $l(\mathbf{x})$ is the shading, and $\mathbf{R}(\mathbf{x}) = l(\mathbf{x})\hat{\mathbf{R}}$. Since the equation is a linear equation, in the RGB space the pixels of a patch will form a straight line (unless when the assumptions are violated–e.g., when patches contain color or depth boundaries). This line will intersect with another line formed by $(1-t)\mathbf{L}_\infty$. Since $\mathbf{L}_\infty$ is assumed to be known, then by having the intersection, $(1-t)$ can be obtained. To obtain $t(\mathbf{x})$ for the entire image, the method has to scan the pixels, extract patches, and find the intersections. Some patches might not give correct intersections; however, if the majority of patches do, then the estimation can be correct. Patches containing object color identical to the atmospheric light color will not give any intersection, as the lines will be parallel. A Gaussian Markov random field (GMRF) is used to do the interpolation.

Sulami et al.'s method [49] uses the same idea and assumptions of the local color lines to estimate the atmospheric light, $\mathbf{L}_\infty$, automatically. First, it estimates the color of the atmospheric light by using a few patches, a minimum of two patches of different scene reflections. It assumes the two patches provide two different straight lines in the RGB space, and the atmospheric light's vector which starts from the origin must intersect with the two straight lines. Second, knowing the normalized color vector, it tries to estimate the magnitude of the atmospheric light. The idea is to dehaze the image using the estimated normalized light vector, and then to minimize the distance between the estimated shading and the estimated transmission for the top 1% brightness value found at each transmission level.

Cai et al. 2016 [25] proposes a learning-based framework similar to [24] that trains a regressor to predict the transmission value $t(\mathbf{x})$ at each pixel (16×16) from its surrounding patch. Unlike [24], which used a handcrafted features, Cai et al. [25] applied a convolutional neural network (CNN) framework with special network design. The network, termed DehazeNet is conceptually formed by four sequential operations (feature extraction, multi-scale mapping, local extremum, and nonlinear regression), which consist of 3 convolution layers, a max-pooling, a maxout unit, and a bilateral rectified linear unit (BReLU, a nonlinear activation function extended from standard ReLU [50]). The training set used is similar to that in [24]–namely, they gathered haze-free patches from Internet to generate hazy patches using the hazy imaging model with random transmissions t and assuming white atmosphere light color ($\mathbf{L}_\infty = [1\,1\,1]^\top$). Once all the weights in the network are obtained from the training, the transmission estimation for a new hazy image patch is simply forward propagation using the network. To handle the block artifact caused by the patch-based estimation, guided filtering [51] is used to refine the transmission map before recovering the scene.

Berman et al. 2016 [26] proposes an algorithm based on a new, non-local prior. This is a departure from existing methods (e.g., [8, 20, 23, 1, 24, 25]) that use patch-based transmission estimation. The algorithm by [26] relies on the assumption that colors of a haze-free image are well approximated by a few hundred distinct colors, that form tight clusters in RGB space and pixels in a cluster are often non-local (spread in the whole image). The presence of haze will elongate the shape of each cluster to a line in color space as the pixels may be affected by different transmission coefficients due to their different distances to the camera. The line, termed haze-line, is informative in estimating the transmission factors. In their algorithm, they first proposed a clustering method to group the pixels and each cluster becomes a haze-line. Then the maximum radius of each cluster is calculated and used to estimate the transmission. A final regulation step is performed to enforce the smoothness of the transmission map.

Table 3: Single-image dehazing methods we compared. The programming language use is denoted as: M for matlab, P for python, C for C/C++. The average runtime is tested on images of resolution 720×480 using a desktop with Xeon E5 3.5GHz CPU and 16GB RAM. Source of the results is denoted as: (No symbol) is code from the authors, (*) is our implementation, (†) is result images that are directly provided by the authors.

Methods	Pub. venue	Code	Runtime(s)
Ancuti 13 [22]	TIP 2013	M*	3.0
Tan 08 [8]	CVPR 2008	C	3.3
Fattal 08 [19]	ToG 2008	M†	141.1
He 09 [20]	CVPR 2009	M*	20
Tarel 09 [7]	ICCV 2009	M	12.8
Kratz 09 [42]	ICCV 2009	P	124.2
Meng 13 [23]	ICCV 2013	M	1.0
Fattal 14 [1]	ToG 2014	C†	1.9
Berman 16 [26]	CVPR 2016	M	1.8
Tang 14 [24]	CVPR 2014	M*	10.4
Cai 16 [25]	TIP 2016	M*	1.7

4. Quantitative Benchmarking

In this section, we benchmark several well-known visibility enhancement methods. Our focus is on recent single-image-based methods. Compared with other approaches, single-image-based approaches are more practical and thus have more potential applications. By benchmarking the methods in this approach, we consider it will be beneficial, since one can know the comparisons of the methods quantitatively.

To compare all methods quantitatively we need to test on a dataset with ground truth. Ideally, similar to what Narasimhan et al. [52] did, the dataset should be created from real atmospheric scenes taken over a long period of time to have all possible atmospheric conditions ranging from light mist to dense fog with various backgrounds of scenes. While it may be possible, it is not trivial, since it has to be done at certain times and locations where fog and haze are present frequently. In addition, the illumination in the scene should keep fixed which means clouds and sunlight distribution should be about the same. Unfortunately, these conditions rarely met. Moreover, it is challenging to have a pixel-wise ground truth of a scene without the effect of particles even on a clear day, particularly for distant objects, as significant amounts of atmospheric particles are always present. These reasons motivated us to use synthesized data. We first performed dehazing evaluations on a recent dataset provided by Fattal [1]. In addition, we created a new dataset using a physics-based rendering technique for the evaluation. In the following sections, we will describe the details of the dataset and present the results of different dehazing methods on these datasets. There are earlier synthetic haze/fog image datasets introduced by Tarel et al. in 2010 [53] and 2012 [41], named FRIDA and FRIDA2 (Foggy Road Image DAtabase). This was the first time a synthetic data of scenes with and without haze was used for quantitative evaluation (MAD) of single image defogging methods. However, the FRIDA and FRIDA2 datasets are dedicated to road scenes where most scene components are simple planes. As a result, these datatsets are not used in this paper.

We compare 11 dehazing methods in total, including most representative dehazing methods published in major venues, as listed in Table 3. We use the codes from the authors if the source codes are available. We implement [22, 20, 24, 25] by strictly following the pipeline and parameter settings described in the paper. For [19] and [1], we directly use the results provided along the dataset [1]. Following the convention in the dehazing papers, we simply use the first author's name with the publication year (e.g., **Tan 08**) to indicate each method.

We mainly categorize the methods into three groups: a heuristic method [22] that doesn't use the haze model Eq. (12), model-based methods that use priors [8, 7, 19, 20, 42, 23, 1, 26], and model-based methods that use learning schemes [24, 25]. Due to different programming languages the runtimes are not comparable and are listed just for reference.

4.1. Evaluation on Fattal's Dataset [1]

Fattal's dataset [1][3] has 11 haze images generated using real images with known depth maps. Assuming a spatially constant scattering coefficient β, the transmission map can be generated by applying the direct attenuation model, and the synthesized haze image can be generated using the haze model Eq. (12). One example of the synthesized images is shown in Figure 4.

There are generally three major steps in dehazing: (1) estimation of the atmospheric light, (2) the estimation of the transmission (or the airlight), and (3) the final image enhancement that imposes a smooth constraint of the neighboring transmission. A study of the atmospheric light color estimation in dehazing can be found in [49]. In our benchmarking, our focus is on evaluating the transmission map estimation and final dehazing results. We therefore directly use ground truth atmospheric light color provided in the dataset for all dehazing methods.

[3] `http://www.cs.huji.ac.il/~raananf/projects/dehaze_cl/results/index_comp.html`
We excluded the *Doll* scene due to invalid link on the page.

Table 4: The mean absolute difference of transmission estimation results on Fattal's dataset [1]. The three smallest values are highlighted.

Methods	Church	Couch	Flower1	Flower2	Lawn1	Lawn2	Mansion	Moebius	Reindeer	Road1	Road2
Tan 08 [8]	0.167	0.367	0.216	0.294	0.275	0.281	0.316	0.219	0.372	0.257	0.186
Fattal 08 [19]	0.377	0.090	0.089	0.075	0.317	0.323	0.147	0.111	0.070	0.319	0.347
Kratz 09 [42]	0.147	0.096	0.245	0.275	0.089	0.093	0.146	0.239	0.142	0.120	0.118
He 09 [20]	0.052	0.063	0.164	0.181	0.105	0.103	0.061	0.208	0.115	0.092	0.079
Meng 13 [23]	0.113	0.096	0.261	0.268	0.140	0.131	0.118	0.228	0.128	0.114	0.096
Tang 14 [24]	0.141	0.074	0.044	0.055	0.118	0.127	0.096	0.070	0.097	0.143	0.158
Fattal 14 [1]	0.038	0.090	0.047	0.042	0.078	0.064	0.043	0.145	0.066	0.069	0.060
Cai 16 [25]	0.061	0.114	0.112	0.126	0.097	0.102	0.072	0.096	0.095	0.092	0.088
Berman 16 [26]	0.047	0.051	0.061	0.115	0.032	0.041	0.080	0.153	0.089	0.058	0.062

Table 5: The mean signed difference of transmission estimation results on Fattal's dataset [1].

Methods	Church	Couch	Flower1	Flower2	Lawn1	Lawn2	Mansion	Moebius	Reindeer	Road1	Road2
Tan 08 [8]	0.013	-0.339	-0.117	-0.268	-00.083	-0.089	-0.301	-0.160	-0.358	-0.148	-0.117
Fattal 08 [19]	0.376	0.088	0.088	0.071	0.317	0.323	0.143	0.073	0.063	0.312	0.327
Kratz 09 [42]	-0.006	0.010	-0.220	-0.267	0.003	-0.013	-0.114	-0.236	-0.083	-0.030	0.067
He 09 [20]	-0.035	-0.045	-0.162	-0.180	-0.091	-0.086	-0.041	-0.208	-0.105	-0.054	-0.047
Meng 13 [23]	-0.112	-0.003	-0.259	-0.266	-0.139	-0.130	-0.101	-0.223	-0.086	-0.109	-0.089
Tang 14 [24]	0.133	0.054	-0.008	-0.046	0.059	0.067	0.089	-0.051	0.013	0.094	0.123
Fattal 14 [1]	-0.019	0.086	-0.021	-0.019	0.063	0.045	0.002	-0.105	0.006	0.005	-0.015
Cai 16 [25]	-0.002	0.086	-0.096	-0.118	0.012	0.017	-0.028	-0.070	0.044	0.001	0.023
Berman 16 [26]	0.009	-0.014	-0.051	-0.115	-0.008	-0.013	-0.076	-0.152	-0.059	-0.041	-0.021

Table 6: The mean absolute difference of final dehazing results On fattal's dataset [1]. The three smallest values are highlighted.

Methods	Church	Couch	Flower1	Flower2	Lawn1	Lawn2	Mansion	Moebius	Reindeer	Road1	Road2
Tan 08 [8]	0.109	0.139	0.098	0.134	0.146	0.146	0.154	0.131	0.150	0.111	0.139
Fattal 08 [19]	0.158	0.055	0.028	0.022	0.116	0.123	0.071	0.039	0.034	0.135	0.165
Kratz 09 [42]	0.099	0.060	0.155	0.161	0.055	0.059	0.085	0.155	0.083	0.073	0.088
He 09 [20]	0.036	0.038	0.078	0.080	0.056	0.057	0.034	0.121	0.061	0.051	0.052
Tarel 09 [7]	0.173	0.112	0.130	0.120	0.146	0.161	0.113	0.143	0.179	0.148	0.176
Ancuti 13 [22]	0.188	0.078	0.276	0.219	0.128	0.144	0.109	0.189	0.145	0.135	0.142
Meng 13 [23]	0.052	0.060	0.114	0.106	0.055	0.055	0.048	0.096	0.065	0.052	0.054
Tang 14 [24]	0.087	0.048	0.017	0.019	0.072	0.078	0.053	0.031	0.053	0.088	0.106
Fattal 14 [1]	0.025	0.053	0.019	0.015	0.035	0.033	0.022	0.076	0.034	0.033	0.038
Cai 16 [25]	0.042	0.069	0.045	0.049	0.061	0.0652	0.040	0.043	0.053	0.057	0.065
Berman 16 [26]	0.032	0.031	0.022	0.045	0.026	0.031	0.049	0.081	0.045	0.040	0.042

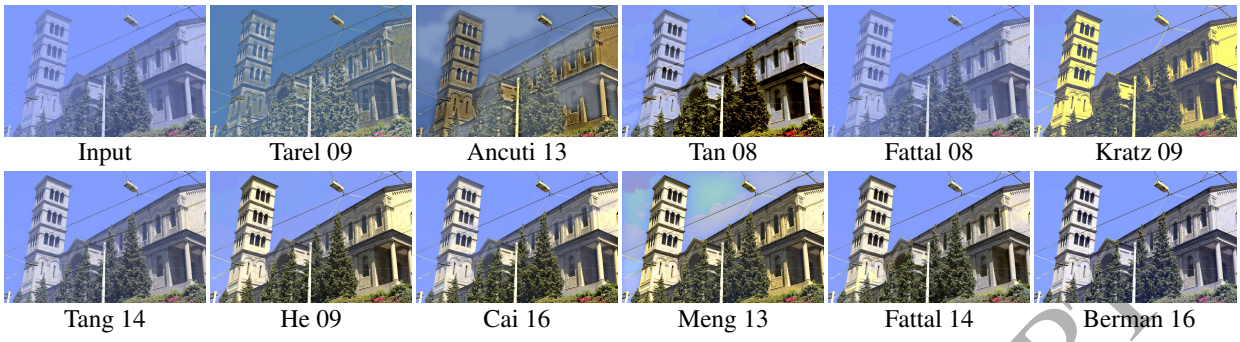


Figure 4: Final haze removal results on the *church* case.

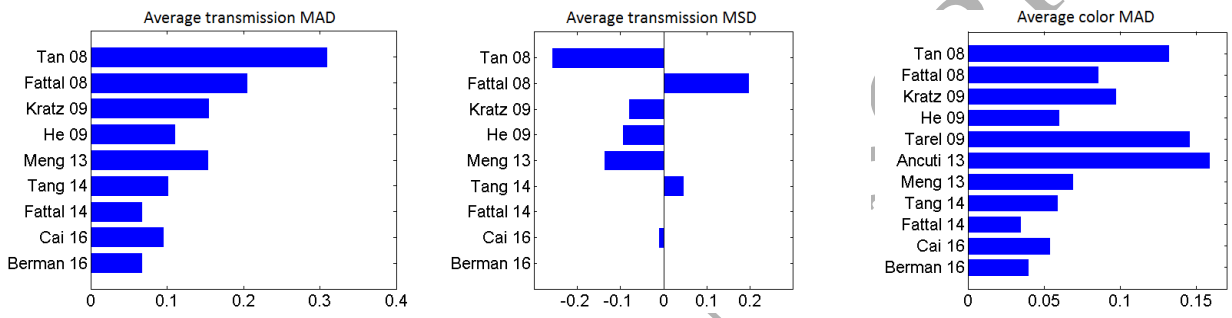


Figure 5: The average performance of different dehazing methods on Fattal's dataset [1].

Transmission Map Evaluation Table 4 lists the mean absolute difference (MAD) of the estimated transmissions (excluding sky regions) of each method to the ground truth transmission. Note that two methods, **Tarel 09** [7] and **Ancuti 13** [22], are not included, as **Tarel 09** [7] directly estimated airlight A in Eq. (14) and **Ancuti 13** [22] does not require the transmission estimation. The three smallest errors for each image are highlighted. We can see no single method can be outstanding for all cases. The recent methods **Fattal 14** [1] and **Berman 16** [26] can obtain more accurate estimation of the transmission for most cases. The early work of **Tan 08** [8] gives less precise estimation. Another early work, **Fattal 08** [19], is not stable and it obtains accurate estimation in a few cases (e.g., *flower2, reindeer*) while it obtains the largest error in some other cases (e.g., *church, road1*).

We plot the average MAD over all 11 cases in Figure 5. It is noticed that in general, the latest methods perform better in the transmission estimation. The methods of **Fattal 14** [1] and **Berman 16** [26] rank at the top, while the two learning-based methods, **Tang 14** [24] and **Cai 16** [25], are in the second place. However, we noticed in our experiments that the learning-based methods heavily rely on the white balance step with correct atmospheric light color. Once there are small errors in atmospheric light color estimation, their performance drops quickly. This indicates the learned models are actually overfilled to the case of white balanced haze images as in the training process it always assumes pure white atmosphere light color. **He 09** [20]'s results also are at a decent rank place. This demonstrates that dark channel prior is an effective prior in the transmission estimation.

We further test the mean signed difference (MSD) on the transmission estimation results (excluding sky regions) as MSD $= \frac{1}{N}\sum_i(\tilde{t}_i - t_i)$, where i is the pixel index, N is the total number of pixels, $\tilde{t}$ is the estimated transmission, and t is the ground truth transmission. By doing so, we can test whether a method overestimates (positive signed difference) or underestimates (negative signed difference) the transmission, which cannot be revealed using the previous MAD metrics. The MSDs are listed in Table 5 and the average MSDs are plotted in Figure 5. It is observed that **Tan 08** [8] mostly underestimates the transmission and as a result it obtains oversaturated dehaze results. **Fattal 08** [19], on the other hand, likely overestimates the transmission, leading to a results with haze still presented in the output. The two methods **He 09** [20] and **Meng 13** [23] also slightly

underestimate the transmission due to the fact they essentially predict the lower bound of transmission.

Dehazing Results Evaluation We evaluate the dehazing results. The mean absolute difference (MAD) of each method (excluding sky regions) to the ground truth clean image is listed in Table 6 and the dehazing results on the *church* case are shown in Figure 4. In Table 6, the three smallest errors for each image are highlighted. Again, no one method can be outstanding for all cases. It is observed that non-model-based method **Ancuti 13** [22] obtains the largest error in the recovery. The visual qualities of their results are also rather inferior compared with other methods (as can be seen in Figure 4). This shows that the image contrast enhancement operation without the haze image model Eq. (12) cannot achieve satisfactory results. Among the rest of the model-based methods, the latest methods, **Meng 13** [23], **Tang 14** [24], **Fattal 14** [1], **Cai 16** [25], and **Berman 16** [26], and also **He 09** [20] generally perform better than early dehazing methods **Tan 08** [8], **Fattal 08** [19], **Tarel 09** [7], and **Kratz 09** [42].

Fattal 14 [1] and **Berman 16** [26] are *the best two methods* that can provide dehazing results that are the closest to the ground truth. This quantitative ranking corresponds well to the overall visual quality for the example shown in Figure 4.

Evaluation with Various Haze Levels

Additionally, we test the performance of each method for different haze levels. In Fattal's dataset [1], he provides a subset of images (*lawn1, mansion, reindeer, road1*) that are synthesized with three different haze levels by controlling the scattering coefficient β. As β increases, denser haze effects will appear. We measure the transmission estimation error and final dehazing error using the mean absolute difference, and the average results over all scenes are plotted in Figure 6.

It is clearly observed that **Fattal 14** [1] stably stands out in achieving fewer errors in both transmission estimation and final dehazing at different haze levels. **Fattal 08** [19] works well only at low haze levels and the performance drops at medium and high haze levels. Looking at the transmission results, we can see **Tan 08** [8]'s, **He 09** [20]'s, and **Meng 13** [23]'s estimation becomes more accurate when haze level increases. This demonstrates that the priors of these three methods are correlated with haze so that these priors can tell more information with more haze. The difference is that **He 09** [20], and **Meng 13** [23] can achieve much smaller transmission errors than **Tan 08** [8], showing the superiority of dark channel prior [20] and boundary constraint [23] against the local contrast [8] for this task. This can be explained by the fact that with heavier haze, the contribution of the airlight $A(x)$ increases, making these types of inputs well-suited to the the dark channel prior and boundary constraint assumptions.

Berman 16 [26] can achieve the least transmission estimation error at medium haze levels but the error increases at both low and heavy haze levels. This may reveal one limitation of **Berman 16** [26] that the haze-lines formed from non-local pixels work well only at certain haze levels. In near clean (low haze level) or heavily hazy scenarios, the haze-lines found may not be reliable. The two learning methods, **Tang 14** [24] and **Cai 16** [25], predict the transmission decently well. For the final dehaze results, most methods obtain large errors in heavy haze except **He 09** [20] and **Fattal 14** [1].

4.2. Evaluation on Our Dataset

Unlike Fattal's dataset, which is generated using images with the haze image model Eq. (12), we generate our dataset using a physically based rendering technique (PBRT) that uses the Monte Carlo ray tracing in a volumetric scattering medium [54]. We render five sets of different scenes under different haze levels of different types – namely, *swamp, house, building, island, villa*. Our scenes are created using freely available 3D models. All five scenes contain large depth variation from a few meters to about $2,000$ meters. We assume a uniform haze density in the space and use homogeneous volumes in our rendering. For each of the five scenes, we render six images. The first one is rendered with no participating media and is considered as the ground truth. The remaining five images are rendered with increasing haze level—namely by evenly increasing the absorption coefficient σ_a and the scattering coefficient σ_s. Figure 7 shows two sets of our generated synthetic data (*building, island*). As can be seen, the visibility of the scene, especially further away objects, decreases when the haze level increases. The whole dataset will be available via a project website.

We have evaluated 9 methods on our dataset (**Fattal 08** [19]'s and **Fattal 14** [1]'s results are not available on our dataset). As the test images in our dataset are rendered with the Monte-Carlo sampling-based ray tracing algorithm, we cannot obtain the transmission map explicitly. Therefore, we quantify the visibility enhancement outputs by comparing them with their respective ground truths. The quantitative measurement is done by using the structural similarity index (SSIM) [55]. While MAD directly measures the closeness of the pixel value to the ground truth, SSIM is more consistent with human visual perception, especially in the cases of dehazing for denser haze levels (haze level beyond 3 in our dataset). SSIM is a popular choice to compute the

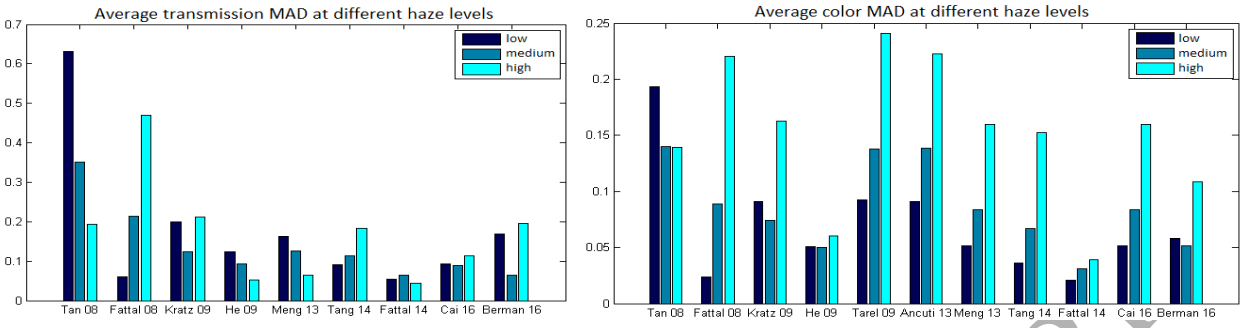


Figure 6: Comparisons of the results for different haze levels.

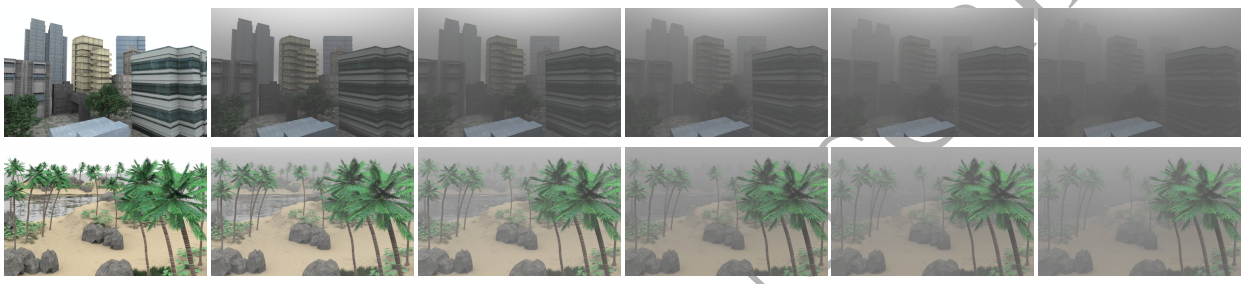


Figure 7: Samples of our synthetic data with increasing haze levels.

structure similarity of two images in image restoration. Unlike MAD, a higher value in SSIM indicates a better match as it is a similarity measurement.

Figure 8 shows the performance of each method in terms of SSIM. It is observed that again the latest methods **Tang 14** [24], **Cai 16** [25], and **Berman 16** [26] generally performed better than others. **He 09** [20] also performs very well, especially in heavier haze levels. This is consistent with our experiment in Section 4.1.

4.3. Qualitative Results on Real Images

We also list three qualitative examples of the dehazing results on real hazy images by different methods in Figure 9 (more visual comparisons can be found in the previous dehazing paper–e.g., [1, 26]). The visual comparison here confirms our findings in the previous benchmarking that **Fattal 14** [1] and **Berman 16** [26] are *the best two methods* that can consistently provide excellent dehazing results. Some early methods, like **Kratz 09** [42], **Tarel 09** [7], and **Ancuti 13** [22] exhibit noticeable limitations in the dehazing results (e.g., oversaturation, boundary artifacts, color shift). **He 09** [20] and **Meng 13** [23] also perform well and obtain similar results as they essentially both predict the lower bound of the transmission. The learning-based methods **Tang 14** [24] and **Cai 16** [25] produce appealing results but tend to leave a noticeable amount of haze in the image.

5. Summary and Discussion

Summary This paper has provided a thorough survey of major methods of visibility enhancement in hazy/foggy scenes. Various modalities, such as multiple images, known approximated depth, stereo, and polarizing filters, have been introduced to tackle the problem. Special emphasis was placed on single-image methods where significant image cues have been explored to enhance visibility, such as local contrast [8], shading-transmission decomposition [19], dark channel prior [20], and line intersection [1]. The tenet of all the methods is to use scene cues to estimate light transmission and to unveil scene reflection based on the estimated transmission. Furthermore, there are two principal properties of the transmission estimation: the estimation of the atmospheric light (both its color and intensity) and the smoothness constraint of the transmission.

We have also conducted the first quantitative benchmark for most representative single-image dehazing methods. Our primary finding from the benchmark is that recent works [1],[26] generally perform better in the dehazing. Machine learning based methods [24, 25] can also get decent results, but their performance is likely to

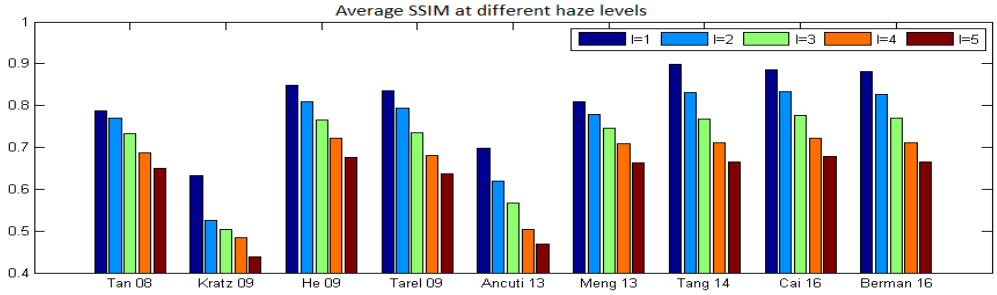


Figure 8: The performance of each method on our dataset on 5 haze levels (l=1,2,3,4,5, low to high) in terms of SSIM.

be affected by the white balancing step. Therefore we still recommend the prior-based methods [1],[26] over the learning-based methods [24, 25] in practical use for robustness. We also found that the popular dark channel prior [20] is an effective prior in dehazing, especially for denser haze levels.

For the dataset used in the benchmark, we picked a dataset from Fattal [1] and also our newly introduced synthetic dataset, which provides ground truth images and haze images with different haze levels. We hope the community can benefit from our dataset by being able to assess new methods more objectively.

Discussion When fog is considerably thick, the problem of visibility enhancement becomes harder. This is because scene reflection is "buried" further underneath the airlight ($\mathbf{A}$) and transmission (t). Considering the scattering model in Eq. (12), when the scattering coefficient β is large–that is, in a thick fog scene–the transmission ($t = e^{-\beta d}$) is small. Consequently, the airlight ($\mathbf{A} = (1-t)\mathbf{L}_\infty$) is dominated by the atmospheric light, $\mathbf{L}_\infty$, and thus the veiling component takes up a greater portion in the image intensity. Also, since the transmission is small, the contribution of scene reflection in the image intensity becomes reduced significantly, due to the multiplication of $\mathbf{R}$ with a fractionally small value of t. The combined airlight and transmission components hide the underlying scene reflection information in the image intensities.

Based on this, some questions might arise: how do we know whether the information of scene reflection is too minuscule to be recovered? How thick is the fog that we cannot extract the scene reflection any longer? Answering such questions is important theoretically, since then we can know the limit of visibility enhancement in bad weather. Furthermore, in thick foggy scenes, due to absorption and scattering to directions other than the line of sight, image blur will be present more prominently and it is not modeled in the current model.

Another issue to note is the application of various onboard camera photo-finishing routines, such as tone-mapping and color manipulation. Although many methods do not explicitly mention the assumption of linearity between the flux of incoming light and the pixel intensity values, based on the scattering model (Eq. (12)), there is an assumption that the image is acting as a linear light-measuring device. While for the purpose of visibility enhancement this might not be an issue, for physically correct scene reflection recovery, the non-linearity of real camera outputs can be a significant issue that needs to be carefully considered.

One interesting finding in our benchmark is that the performance of the recent learning based-methods [24, 25] is just comparable to the prior-based methods like [20, 23, 26]. These learning-based methods essentially learn low-level haze related features. Whether we can do visibility recovery better using the latest machine learning techniques, like deep neural networks, is a possible direction to explore. By doing so, the model can incorporate the semantic information in the dehazing process, which may help reduce some ambiguities in dehaze (e.g., whether a white patch is on a white wall nearby or an object far away in haze). However, as mentioned in Section 1, the haze effect may impair the performance of the vision system that targets high-level tasks. Therefore how to jointly remove haze and estimate high-level information needs careful thought.

Our synthetic dataset is still limited in size. Modelling and rendering a large set of data using physics-based rendering takes a great deal of time and effort. However, continued efforts in producing a larger dataset would be of continued benefit for future work.

Acknowledgment

This study is supported by an Nvidia GPU Grant and a Canadian NSERC Discovery grant. R. T. Tans work in this research is supported by the National Research Foundation, Prime Ministers Office, Singapore under its

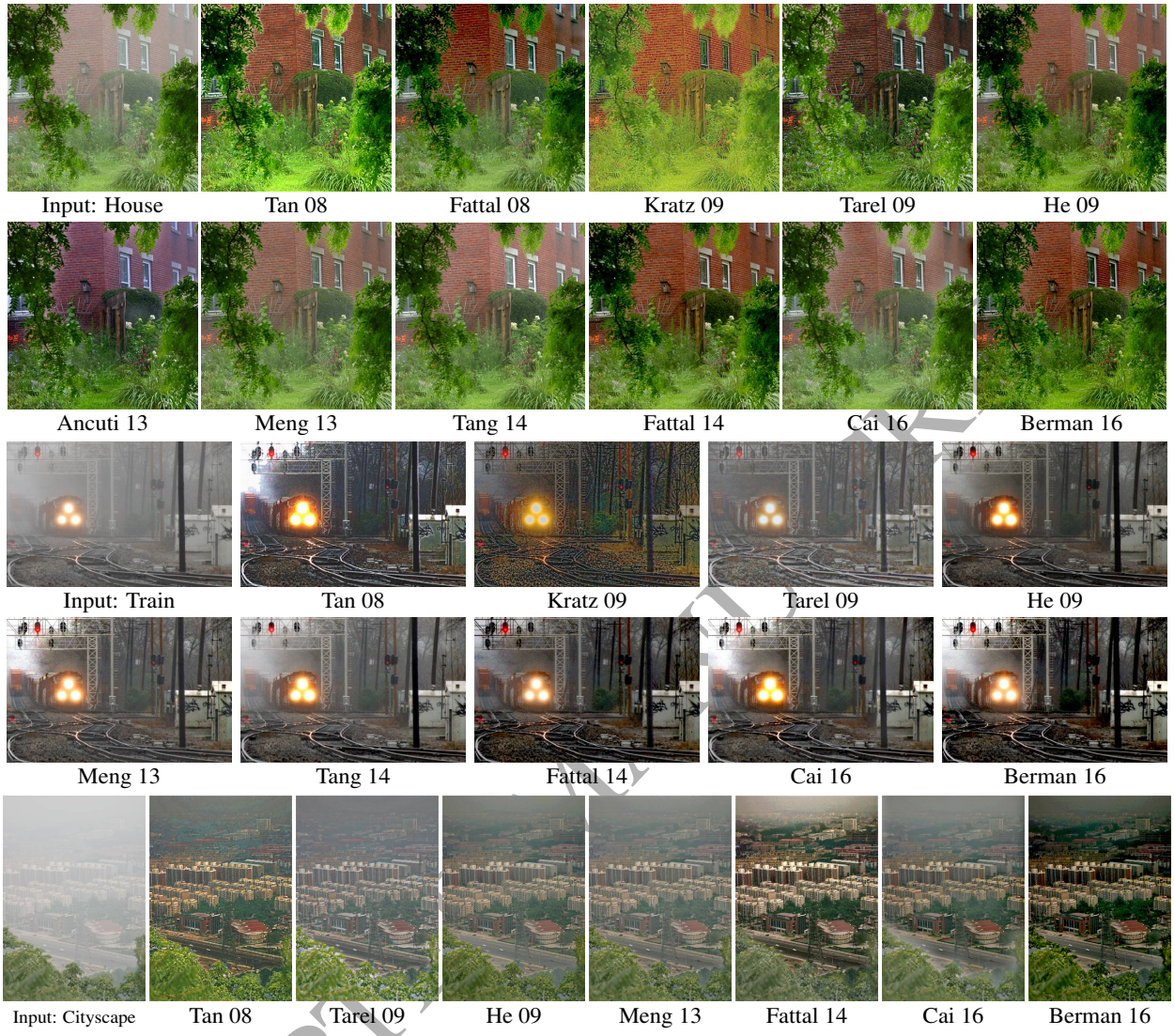


Figure 9: Example comparisons on real images.

International Research Centre in Singapore Funding Initiative.

References

References

[1] R. Fattal, Dehazing using color-lines, ACM Trans. Graph. 34 (1) (2014) 13:1–13:14.

[2] H. Koschmieder, Theorie der horizontalen Sichtweite: Kontrast und Sichtweite, Keim & Nemnich, 1925.

[3] E. J. McCartney, Optics of the atmosphere: scattering by molecules and particles, New York, John Wiley and Sons, Inc., 1976. 421 p. 1.

[4] F. Cozman, E. Krotkov, Depth from scattering, in: IEEE Conf. Computer Vision and Pattern Recognition, 1997.

[5] S. K. Nayar, S. G. Narasimhan, Vision in bad weather, in: IEEE Int'l Conf. Computer Vision, 1999.

[6] S. G. Narasimhan, S. K. Nayar, Vision and the atmosphere, Int'l J. Computer Vision 48 (3) (2002) 233–254.

[7] J.-P. Tarel, N. Hautière, Fast visibility restoration from a single color or gray level image, in: IEEE Int'l Conf. Computer Vision, 2009.

[8] R. T. Tan, Visibility in bad weather from a single image, in: IEEE Conf. Computer Vision and Pattern Recognition, 2008.

[9] S. G. Narasimhan, S. K. Nayar, Chromatic framework for vision in bad weather, in: IEEE Conf. Computer Vision and Pattern Recognition, 2000.

[10] S. G. Narasimhan, S. K. Nayar, Contrast restoration of weather degraded images, IEEE Trans. Pattern Analysis and Machine Intelligence 25 (6) (2003) 713–724.

[11] L. Caraffa, J.-P. Tarel, Stereo reconstruction and contrast restoration in daytime fog, in: Asian Conf. Computer Vision,

2012.

[12] Z. Li, P. Tan, R. T. Tan, S. Z. Zhou, L.-F. Cheong, Simultaneous video defogging and stereo reconstruction, in: IEEE Conf. Computer Vision and Pattern Recognition, 2015.

[13] Y. Y. Schechner, S. G. Narasimhan, S. K. Nayar, Instant dehazing of images using polarization, in: IEEE Conf. Computer Vision and Pattern Recognition, 2001.

[14] S. Shwartz, E. Namer, Y. Y. Schechner, Blind haze separation, in: IEEE Conf. Computer Vision and Pattern Recognition, 2006.

[15] J. P. Oakley, B. L. Satherley, Improving image quality in poor visibility conditions using a physical model for contrast degradation, IEEE Trans. Image Processing 7 (2) (1998) 167–179.

[16] S. G. Narasimhan, S. K. Nayar, Interactive (de) weathering of an image using physical models, in: IEEE Workshop on Color and Photometric Methods in Computer Vision, 2003.

[17] N. Hautière, J.-P. Tarel, D. Aubert, Towards fog-free in-vehicle vision systems through contrast restoration, in: IEEE Conf. Computer Vision and Pattern Recognition, 2007.

[18] J. Kopf, B. Neubert, B. Chen, M. F. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, Deep photo: Model-based photograph enhancement and viewing, ACM Trans. Graphics 27 (5) (2008) 116:1–116:10.

[19] R. Fattal, Single image dehazing, ACM Trans. Graphics 27 (3) (2008) 72.

[20] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, in: IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[21] C. O. Ancuti, C. Ancuti, P. Bekaert, Effective single image dehazing by fusion, in: IEEE Int'l Conf. Image Processing, 2010.

[22] C. O. Ancuti, C. Ancuti, Single image dehazing by multi-scale fusion, IEEE Trans. Image Processing 22 (8) (2013) 3271–3282.

[23] G. Meng, Y. Wang, J. Duan, S. Xiang, C. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in: IEEE Int'l Conf. Computer Vision, 2013.

[24] K. Tang, J. Yang, J. Wang, Investigating haze-relevant features in a learning framework for image dehazing, in: IEEE Conf. Computer Vision and Pattern Recognition, 2014.

[25] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, IEEE Trans. Image Processing 25 (11) (2016) 5187–5198.

[26] D. Berman, T. Treibitz, S. Avidan, Non-local image dehazing, in: IEEE Conf. Computer Vision and Pattern Recognition, 2016.

[27] G. M. Hidy, M. Kerker, Aerosols and Atmospheric Chemistry: The Kendall Award Symposium Honoring Milton Kerker, at the Proceedings of the American Chemical Society, Los Angeles, California, March 28-April 2, 1971, Academic Press, 1972.

[28] M. O. Codes, International codes–wmo no. 306, Geneva–Switzerland: World Meteorological.

[29] C. D. Ahrens, Meteorology today: an introduction to weather, climate, and the environment, West Publishing Company New York, 1991.

[30] M. G. J. Minnaert, The Nature of Light and Colour in the Open Air: Transl.[By] HM Krener-Priest, Rev.[By] KE Brian Jay, Dover, 1954.

[31] S. A. Shafer, Using color to separate reflection components, Color Research & Application 10 (4) (1985) 210–218.

[32] S. Tominaga, B. A. Wandell, Standard surface-reflectance model and illuminant estimation, J. Opt. Soc. Am. A 6 (4) (1989) 576–584.

[33] L. Caraffa, J.-P. Tarel, Combining stereo and atmospheric veil depth cues for 3d reconstruction, IPSJ Transactions on Computer Vision and Applications 6 (2014) 1–11.

[34] G. Zhang, J. Jia, T.-T. Wong, H. Bao, Consistent depth maps recovery from a video sequence, IEEE Trans. Pattern Analysis and Machine Intelligence 31 (6) (2009) 974–988.

[35] K. Tan, J. P. Oakley, Enhancement of color images in poor visibility conditions, in: IEEE Int'l Conf. Image Processing, 2000.

[36] N. Hautière, J.-P. Tarel, J. Lavenant, D. Aubert, Automatic fog detection and estimation of visibility distance through use of an onboard camera, Machine Vision and Applications 17 (1) (2006) 8–20.

[37] R. T. Tan, N. Pettersson, L. Petersson, Visibility enhancement for roads with foggy or hazy scenes, in: IEEE Intelligent Vehicles Symposium, 2007.

[38] R. T. Tan, K. Nishino, K. Ikeuchi, Color constancy through inverse-intensity chromaticity space, J. Opt. Soc. Am. A 21 (3) (2004) 321–334.

[39] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Analysis and Machine Intelligence 33 (12) (2011) 2341–2353.

[40] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Analysis and Machine Intelligence 30 (2) (2008) 228–242.

[41] J.-P. Tarel, N. Hautière, L. Caraffa, A. Cord, H. Halmaoui, D. Gruyer, Vision enhancement in homogeneous and heterogeneous fog, IEEE Intelligent Transportation Systems Magazine 4 (2) (2012) 6–20.

[42] L. Kratz, K. Nishino, Factorizing scene albedo and depth from a single foggy image, in: IEEE Int'l Conf. Computer Vision, 2009.

[43] K. Nishino, L. Kratz, S. Lombardi, Bayesian defogging, Int'l J. Computer Vision 98 (3) (2012) 263–278.

[44] J. Kim, R. Zabih, Factorial markov random fields, in: European Conf. Computer Vision, 2002.

[45] L. Caraffa, J.-P. Tarel, Markov random field model for single image defogging, in: IEEE Intelligent Vehicles Symposium, 2013.

[46] G. Buchsbaum, A spatial processor model for object colour perception, Journal of the Franklin Institute 310 (1) (1980) 1–26.

[47] C. Ancuti, C. O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: IEEE Conf. Computer Vision and Pattern Recognition, 2012.

[48] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

[49] M. Sulami, I. Geltzer, R. Fattal, M. Werman, Automatic recovery of the atmospheric light in hazy images, in: IEEE Int'l Conf. Computational Photography, 2014.

[50] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Int'l Conf. Machine Learning, 2010.

[51] K. He, J. Sun, X. Tang, Guided image filtering, in: European Conf. Computer Vision, 2010.

[52] S. G. Narasimhan, C. Wang, S. K. Nayar, All the images of an outdoor scene, in: European Conf. Computer Vision, 2002.

[53] J.-P. Tarel, N. Hautière, A. Cord, D. Gruyer, H. Halmaoui, Improved visibility of road scene images under heterogeneous fog, in: IEEE Intelligent Vehicles Symposium, 2010.

[54] M. Pharr, G. Humphreys, Physically based rendering: From theory to implementation, Morgan Kaufmann, 2010.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Processing 13 (4) (2004) 600–612.