Single Image Dehazing and Denoising: A Fast Variational Approach[*]

Faming Fang[†], Fang Li[‡], and Tieyong Zeng[§]

Abstract. In this paper, we propose a new fast variational approach to dehaze and denoise simultaneously. The proposed method first estimates a transmission map using a windows adaptive method based on the celebrated dark channel prior. This transmission map can significantly reduce the edge artifact in the resulting image and enhance the estimation precision. The transmission map is then converted to a depth map, with which the new variational model can be built to seek the final haze- and noise-free image. The existence and uniqueness of a minimizer of the proposed variational model is further discussed. A numerical procedure based on the Chambolle–Pock algorithm is given, and the convergence of the algorithm is ensured. Extensive experimental results on real scenes demonstrate that our method can restore vivid and contrastive haze- and noise-free images effectively.

1. Introduction. Due to the presence of the atmosphere, the light reflected from a scenario is always scattered before it reaches the camera lens, and the light collected by any camera lens is usually blended with the airlight [26, 41]. This leads to inevitable image degradation such as increasing noise, reduction of intensity contrast, and loss of color fidelity. This kind of degradation is particularly serious when the weather conditions are poor, i.e., when aerosols such as haze, fog, rain, dust, or fumes are present. For instance, as a common weather phenomenon, fog may produce an albedo effect, which leads to ambiguity and noise [20]. These phenomena, to some extent, have adverse effects on comprehension and extraction of contents from the images. Therefore, effective haze removal (or dehazing) and denoising methods are urgently needed in real applications.

Indeed, the image dehazing and denoising of natural scene images have attracted much attention in imaging science recently. The advantages of such operations are clear. First, the haze- and noise-free images are visually more vivid and appealing; second, the haze- and noise-free images are more suitable for many important applications such as image segmentation, feature extraction, and image fusion. However, as usually the haze depends strongly on unknown depth information, the image dehazing problem is a very challenging task. The problem will be more ill-posed if the input data is only a single image [24]. In the litera-

[†]Department of Computer Science, East China Normal University, Shanghai, China (fmfang@cs.ecnu.cn).

[‡]Department of Mathematics, East China Normal University, Shanghai, China (fli@math.ecnu.edu.cn).

[§]Corresponding author. Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. (zeng@hkbu.edu.hk).

ture, various approaches have been proposed to handle the image dehazing problem. We can roughly classify these methods into two categories: multiple image dehazing and single image dehazing. Among the first kind of approaches, the polarization-based methods can remove haze effectively through multiple images of the same scene with different degrees of polarization [38, 39]. In [30, 32, 34, 38, 39], the dehazed image is obtained from several images of the same scene and angle under different weather conditions. In depth-based methods, an approximate 3D geometrical model of the input scene is needed for dehazing [25, 33]. Since both multiple images and a 3D geometrical model are difficult to acquire in real applications, the above methods are somewhat impractical.

Single image dehazing methods have thus been more attractive. As the input data has only one image, these methods usually need rather strong prior information or assumptions. In [20], relying on the assumption that the transmission and surface shading are locally uncorrelated, the albedo of the scene and the medium transmission are inferred, and impressive dehazed results are thus obtained. However, this method may fail in heavily hazy images or in the case when the assumption is seriously violated. In [41], observing when the airlight (see discussion after (2.2) for definition) in natural images tends to be smooth and the haze-free image has more contrast than the original hazy image, Tan proposes to remove the haze by maximizing the local contrast of the restored image. This approach can produce impressive results, though the physical meaning of the above observations is somewhat vague. In [24], He, Sun, and Tang present a novel prior: the dark channel prior. A powerful single image haze removal method is then proposed based on this prior. Using this method, they can directly estimate the thickness of the haze and recover a high-quality haze-free image. However, the CPU time and memory consumption of this method are rather high, and the resulting image may contain some noise especially for a heavy haze scene.

Moreover, a common disadvantage of the above single image dehazing methods is that they all ignore the presence of noise. As noise is somewhat inevitable in natural images, the dehazed image produced by the above approaches might contain some heavy noise. This phenomenon can be easily observed in [24]. To overcome this drawback, a possible solution is that we can further denoise after obtaining the dehazed image.

A promising denoised result should contain the most reliable estimation of the original image given a degraded input. Since denoising may introduce blurring or artifacts, it has remained a challenge for researchers. In recent decades, extensive denoising methods have been developed, such as [6, 16, 18, 36, 43]. Among them, a noteworthy method is the Rudin–Osher–Fatemi (ROF) denoising model [36], in which a total variation (TV) prior is used as a global regularization term. The TV and its adaptive term, such as weighted TV or vectorial TV (VTV), have been proved to be effective for edge preservation, and they have been applied in many image processing tasks, such as image denoising [4], image restoration [7, 14, 15, 17], deblurring [12], filtering [29], and so on. Note that the minimization on the TV related models can be handled by fast algorithms, such as the Chambolle projection method [11], the Chambolle–Pock algorithm [10], the augmented Lagrangian method [40], and the split Bregman iteration [21].

However, as denoising may causally lead to the blurring of some useful details, the naive approach is that first dehazing and then denoising should be improved. To preserve more details in images, we propose considering a simultaneous procedure for dehazing and denosing.

Indeed, in this paper, inspired by He, Sun, and Tang [24], we first present an improved approach named the windows adaptive method to estimate the transmission map from the hazy image. This improved method can overcome the edge artifact, and then the process of soft matting in [24], which is very time-consuming, can be avoided. By using this transmission map in the data fidelity term, we then propose a new variational model with weighted vectorial total variation (WVTV). By minimizing the energy efficiently, a haze- and noise-free image is then recovered simultaneously from the original single hazy image.

The main contributions will be clear:

- we present a windows adaptive method to estimate the transmission map;
- we propose a new energy model for dehazing and denoising simultaneously;
- we discuss the existence and uniqueness of the minimizer of the proposed energy functional;
- to the best of our knowledge, the framework of the weighted vectorial total variation introduced here is somewhat new and could be applied elsewhere;
- we describe the numerical procedure for our energy based on the Chambolle–Pock algorithm as well as its convergence.

The rest of the paper is organized as follows. In section 2, we briefly introduce the background and related works for the dehazing problem. In section 3, a new windows adaptive method is presented, and we introduce our convex variational model for simultaneously dehazing and denoising. The basic mathematical properties related to the proposed model are then addressed in section 4. Moreover, section 5 describes the numerical scheme and its convergence property. In section 6, we show the results after applying our approach on a set of natural images and illustrate the superior performance of our approach. Finally, some concluding remarks are given in section 7.

2. Background and related works. In computer vision and computer graphics, the formation of a hazy image can naturally be described as follows [20, 24, 30, 31, 33]:

$$\mathbf{I}(x) = t(x)\mathbf{J}(x) + (1 - t(x))\mathbf{A}, \tag{2.1}$$

where $\mathbf{I} : \Omega \to [0, \ 255]^3$ is the observed hazy image, Ω is a connected bounded open subset of $\mathbb{R}^2$ with compact Lipschitz boundary, $x \in \Omega$ is the 2D spatial location in an image, and $\mathbf{A}$ is the atmospheric light, which is assumed to be globally constant. Moreover, $\mathbf{J} : \Omega \to [0, \ 255]^3$ is the surface radiance, and $t : \Omega \to [0, \ 1]$ is the transmission describing the portion of the light that is not distorted. When the atmosphere is homogenous, the transmission $t(x)$ can be described as follows [20, 23]:

$$t(x) = e^{-\eta d(x)}, \tag{2.2}$$

where η is the medium extinction coefficient (typically, we can set it as 1) and $d(x)$ is the depth of scene, which is always nonnegative since $t \le 1$.

In (2.1), the terms $t(x)\mathbf{J}(x)$ and $(1 - t(x))\mathbf{A}$ are called the direct attenuation and airlight, respectively [26, 41]. Since only the hazy image $\mathbf{I}$ is given, we should recover $\mathbf{J}$, $t(x)$, and $\mathbf{A}$ from $\mathbf{I}$. Note that the estimation of $\mathbf{A}$ itself is not so easy. Most of the existing methods, such as [20, 41], set the highest intensity of the input image as $\mathbf{A}$. In this paper, similarly

to [24], we handle this issue with the help of the so-called dark channel prior; for details, see discussion after (2.5).

Indeed, the dark channel prior presented by He, Sun, and Tang [24] is a kind of statistics on haze-free outdoor images. This prior is based on the following observation: in most of the nonsky patches of the haze-free outdoor image, at least one color channel has very low intensity. For an image $\mathbf{J}$, formally, the dark channel $\mathbf{J}_{dark}$ is defined as

$$(2.3) \qquad \mathbf{J}_{dark}(x) = \min_{c \in \{R,G,B\}} \left(\min_{y \in \omega(x)} \mathbf{J}_c(y) \right),$$

where $c \in \{R, G, B\}$ is a color channel, $\mathbf{J}_c$ is a single band image of $\mathbf{J}$, and $\omega(x)$ is a local patch centered at x. Then the dark channel prior can be formulated as

$$(2.4) \qquad \mathbf{J}_{dark}(x) \approx 0.$$

Assuming that the transmission t is a constant in the local patch $\omega(x)$, and taking the min operation in the local patch and three color channels on the haze image $\mathbf{I}$, a rough version of t (denoted t_0) can be obtained by (2.1) and (2.4):

$$(2.5) \qquad t_0(x) = 1 - \nu \min_{c \in \{R,G,B\}} \left(\min_{y \in \omega(x)} \left(\frac{\mathbf{I}_c(y)}{\mathbf{A}_c} \right) \right),$$

where $\nu \in (0, 1]$ is used to keep some depth information of the natural image and is commonly set as $\nu = 0.95$ [24]. Additionally, among the pixels which are the brightest 0.1% in $\mathbf{J}_{dark}$, the pixels with the highest intensity in $\mathbf{I}$ are picked as the atmospheric light $\mathbf{A}$.

After obtaining t_0, a dehazed image can be recovered according to (2.1); i.e.,

$$(2.6) \qquad \mathbf{J_0}(x) = \frac{\mathbf{I}(x) - \mathbf{A}}{\max(t_0(x), \epsilon_0)} + \mathbf{A},$$

where $\epsilon_0 > 0$ is a constant to avoid division by zero (typically, we can set ϵ_0 as 0.1).

As shown in [24], $\mathbf{J_0}$ may contain some edge artifacts around edges (see the roof in Figure 1(b) below, for example) because the transmission t_0 is not so accurate. In order to overcome this defect, the authors of [24] use the soft matting method [28] to refine t_0. However, the procedure of solving a linear equation $MX = b$ is needed for their soft matting method, which leads to high computational complexity. Here M is a very large matting Laplacian matrix with size $N_1 N_2 \times N_1 N_2$ given an image with $N_1 \times N_2$ pixels. Therefore, this method is not suitable for real-time or nearly real-time dehazing. Besides, as mentioned above, the result of He, Sun, and Tang may contain some unpleasant noise especially in the heavy haze scene.

3. The proposed method. In this section, we first present a new method for estimating the transmission map and then construct an energy functional to recover the final haze- and noise-free image.

3.1. Windows adaptive method. As mentioned above, the dehazed result obtained by (2.5) and (2.6) contains some edge artifacts, especially on the boundary of an object. This is due to the limitation that the transmission t is assumed to be a constant in the local patch $\omega(x)$. However, this is not always true. Indeed, when $\omega(x)$ contains a sharp edge, the constant assumption of t is clearly wrong. This inevitably leads to undesired results.

To avoid the edge artifact, now we present a new method for choosing the local patch. We term this method the windows adaptive method.

For a $y \in \omega(x)$ we define a distance to measure the intensity difference between x and y,

$$(3.1) \qquad\qquad \varrho_c(x, y) := |\mathbf{I}_c(x) - \mathbf{I}_c(y)|.$$

The top $r\%$ pairs of pixel and channel which have minimal distance among all bands ($y \in \omega(x)$, $c \in \{R, G, B\}$) are selected (usually, we set $r = 40$). Defining this selected set as $\omega_c(x)$, equation (2.5) can be changed accordingly:

$$(3.2) \qquad\qquad t_0(x) = 1 - \nu \min_{(y,c)\in\omega_c(x)} \left(\frac{\mathbf{I}_c(y)}{\mathbf{A}_c}\right),$$

where again $\nu = 0.95$ and $\mathbf{A} = (\mathbf{A}_R, \mathbf{A}_G, \mathbf{A}_B)$ is obtained along the lines of the method of He, Sun, and Tang (see [24] or the discussion after (2.5) for details).

After obtaining t_0, we can get a new $\mathbf{J}_0$ using (2.6). The experiment done in many images shows that this adaptive method can avoid edge artifacts, as shown in Figure 1(d) below. However, observing those dehazed images, we can also find that they are both dark in intensity and low in contrast, and they contain noise. So we need to further refine the results to get more impressive outcomes.

3.2. The proposed model. Obviously, (2.1) can be rewritten as

$$(3.3) \qquad\qquad \mathbf{A}_c - \mathbf{I}_c = t(\mathbf{A}_c - \mathbf{J}_c).$$

In order to avoid product form, we convert (3.3) into the logarithmic domain, i.e.,

$$(3.4) \qquad\qquad \log(\mathbf{A}_c - \mathbf{I}_c) = \log t + \log(\mathbf{A}_c - \mathbf{J}_c).$$

As shown in (2.2), t is an exponential function of the depth map d, so we have

$$d = -\frac{1}{\eta} \log t,$$

where again η is a positive constant and, for simplicity, we set it to 1 for all our numerical tests.

We further set $f_c = \frac{1}{\eta} \log(\mathbf{A}_c - \mathbf{I}_c)$ and $g_c = \frac{1}{\eta} \log(\mathbf{A}_c - \mathbf{J}_c)$; then (3.4) is equal to

$$(3.5) \qquad\qquad g_c = f_c + d.$$

Let $\boldsymbol{g} = \{g_R, g_G, g_B\}$, $\boldsymbol{f} = \{f_R, f_G, f_B\}$, and $\boldsymbol{d} = \{d, d, d\}$. The above equation can be rewritten as

$$(3.6) \qquad\qquad \boldsymbol{g} = \boldsymbol{f} + \boldsymbol{d}.$$

Our coming variational model will be based on the following assumptions:

- Since $\mathbf{J}$ is the surface radiance at the intersection point of the scene and the real-world ray corresponding to the pixel x, we can reasonably say that $\mathbf{g}$ is piecewise constant. Thus a WVTV, $\int_\Omega h(x)|D\mathbf{g}(x)|$, can be set as the regularization term of $\mathbf{g}$ [3, 4, 5, 17, 27, 37, 42], where $h(x)$ is a positive continuous real-valued function related to the haze thickness (see (3.9) below for a particular choice) and

$$(3.7) \qquad \int_\Omega h|D\mathbf{g}| := \sup_{\boldsymbol{\varphi} \in C_0^1(\Omega, \mathbb{R}^{3\times 2})} \left\{ \int_\Omega \langle \mathbf{g}, \nabla \cdot \boldsymbol{\varphi} \rangle dx : \ |\boldsymbol{\varphi}(x)| \le h(x) \ \forall x \in \Omega \right\}.$$

Here, $D\mathbf{g}$ is the distributional gradient of $\mathbf{g}$, $\boldsymbol{\varphi} := (\varphi_R, \varphi_G, \varphi_B) : \Omega \to \mathbb{R}^{3\times 2}$ is a vector-valued function, $\nabla\cdot$ is the divergence operator such that $\nabla\cdot\boldsymbol{\varphi} = -(\nabla^T\varphi_R, \nabla^T\varphi_G, \nabla^T\varphi_B) : \Omega \to \mathbb{R}^3$, ∇ is the gradient in the classical sense, the product $\langle \cdot \rangle$ is the Euclidean scalar product which implies that $\langle \mathbf{g}, \nabla\cdot\boldsymbol{\varphi} \rangle = -\sum_{c\in\{R,G,B\}}\langle g_c, \nabla^T\varphi_c \rangle$, and the norm $|\cdot|$ can be defined in different ways, such as $|\cdot|_{L^\infty}, |\cdot|_{L^2}$. In this paper, we use $|\cdot|_{L^2}$ since it is the most standard definition of the VTV norm [4]; that is,

$$|\boldsymbol{\varphi}| = \sqrt{\sum_{c\in\{R,G,B\}} \langle \varphi_c, \varphi_c \rangle}.$$

- The depth map is piecewise regular. Therefore, the regularization term is used by $\int_\Omega |Dd|$.

Overall, we propose to minimize the following energy functional for recovering the haze- and noise-free image simultaneously:

$$(3.8) \qquad E(\mathbf{g}, d) := \int_\Omega h|D\mathbf{g}| + \lambda \int_\Omega |Dd| + \frac{1}{2}\int_\Omega |\mathbf{g} - \mathbf{f} - \mathbf{d}|^2 dx + \frac{\gamma}{2}\int_\Omega |d - d_0|^2 dx,$$

where $d_0 = -\frac{1}{\eta}\log t_0$ (see (3.2)) and λ, γ are positive parameters with larger values indicating more important corresponding terms. In the above formula, the last two terms are the data fidelity terms. Since the WVTV can denoise effectively while keeping the boundary [3, 5, 27, 42], and there exists a regularization term on d, the minimization of our energy also can denoise effectively.

Now we give some details on the selection of the function h. Note that, as mentioned previously, the hazy image usually contains some noise. Moreover, we also observe that the noise is not uniform. Usually, it is rather serious in scenes with dense haze and is rather light in scenes of mist. As we aim to maintain as many details as possible while denoising, we should set a large $h(x)$ in the haze-dense area and set a small $h(x)$ in the mist scene. Generally speaking, the depth map d_0 can roughly describe the thickness of haze. In this paper, we assume that $h(x)$ is an increasing function of $d_0(x)$, i.e.,

$$(3.9) \qquad\qquad\qquad h(x) := \frac{1}{1 + ke^{-5d_0(x)}},$$

where k is a nonnegative parameter that determines the depth level of the image and the constant 5 is empirically fixed.

Since the space of functionals of bounded variation $BV(\Omega)$ contains the set of piecewise smooth functions, it is a suitable space for images and has been frequently used in many image processing problems [1, 22, 36]. Therefore, we choose $BV(\Omega)$ as our search space for d. In detail, the space $BV(\Omega)$ is defined as follows:

$$BV(\Omega) = \left\{ d \in L^1(\Omega); \ \int_\Omega |Dd| < +\infty \right\}.$$

Further, for $\boldsymbol{g}$, we can use the vector-valued space $\mathbf{BV}(\Omega)$, which is defined as follows:

$$\mathbf{BV}(\Omega, \mathbb{R}^3) = \left\{ \boldsymbol{g} \in \mathbf{L}^1(\Omega, \mathbb{R}^3); \ \int_\Omega |D\boldsymbol{g}| < +\infty \right\}.$$

Then the total space Λ of our energy functional can be defined as

$$(3.10) \qquad \Lambda := \{(\boldsymbol{g}, d) | (\boldsymbol{g}, d) \in \mathbf{BV}(\Omega, \mathbb{R}^3) \times BV(\Omega)\}.$$

After the functional space has been chosen, the minimizing problem of (3.8) can be cast as the following regular form:

$$(3.11) \qquad \inf_{(\boldsymbol{g},d)\in\Lambda} E(\boldsymbol{g}, d).$$

4. Existence and uniqueness of the minimizer. The goal of this section is to prove the existence and uniqueness of the solution for (3.11) based on fundamental theories in optimization. For this, we need the following two lemmas.

Lemma 4.1 (lower semicontinuity). *If the sequences $\{\boldsymbol{g}^j\} \subset \mathbf{BV}(\Omega, \mathbb{R}^3)$ and $\boldsymbol{g}^* \in \mathbf{BV}(\Omega, \mathbb{R}^3)$ satisfy that $\boldsymbol{g}^j \underset{\mathbf{L}^1(\Omega,\mathbb{R}^3)}{\to} \boldsymbol{g}^*$, i.e.,*

$$\lim_{j\to+\infty} \int_\Omega |\boldsymbol{g}^j - \boldsymbol{g}^*| dx = 0,$$

then

$$\int_\Omega h|D\boldsymbol{g}^*| \le \liminf_{j\to+\infty} \int_\Omega h|D\boldsymbol{g}^j|.$$

Proof. Define

$$S := \{\boldsymbol{\varphi} \in C_0^1(\Omega); |\boldsymbol{\varphi}(x)| \le h(x) \ \forall x \in \Omega\}.$$

For fixed $\boldsymbol{\varphi} \in S$, $\nabla \cdot \boldsymbol{\varphi}$ is bounded on Ω. By (3.7), readily we have

$$\int_\Omega \langle \boldsymbol{g}^*, \nabla \cdot \boldsymbol{\varphi} \rangle dx = \lim_{j\to+\infty} \int_\Omega \langle \boldsymbol{g}^j, \nabla \cdot \boldsymbol{\varphi} \rangle dx \le \liminf_{j\to+\infty} \sup_{\phi\in S} \left\{ \int_\Omega \langle \boldsymbol{g}^j, \nabla \cdot \boldsymbol{\phi} \rangle dx \right\} = \liminf_{j\to+\infty} \int_\Omega h|D\boldsymbol{g}^j|.$$

Taking the sup of $\boldsymbol{\varphi}$ over the set S, we obtain

$$\int_\Omega h|D\boldsymbol{g}^*| \le \liminf_{j\to+\infty} \int_\Omega h|D\boldsymbol{g}^j|. \qquad \blacksquare$$

Lemma 4.2. *If $g \in \mathbf{BV}(\Omega, \mathbb{R}^3)$, and $h_1, h_2 : \Omega \to \mathbb{R}$ are both continuous real-valued functions such that $h_1(x) \leq h_2(x)$ for all $x \in \Omega$, then*

$$\int_\Omega h_1 |Dg| \leq \int_\Omega h_2 |Dg|.$$

Proof. This follows directly from (3.7) since $|\varphi(x)| \leq h_1(x)$ implies $|\varphi(x)| \leq h_2(x)$. ∎

Now we are ready to prove the existence and uniqueness of the solution for (3.11).

Theorem 4.3. *Let Ω be a connected bounded open subset of $\mathbb{R}^2$ with compact Lipschitz boundary, and let $\lambda, \gamma > 0$. Assume that $f \in \mathbf{L}^2(\Omega, \mathbb{R}^3)$ and that $d_0 \in L^2(\Omega)$ is nonnegative; then the minimization problem (3.11) admits a unique solution $(g^*, d^*) \in \Lambda$.*

Proof. (a) *Existence.* First, the energy functional is nonnegative, so there exists the infimum. Further, when g and d are constants, the energy is bounded. Therefore, the energy (3.11) is well defined.

Let $\{(g^j, d^j)\} \in \Lambda$ be a minimizing sequence of problem (3.11). Then there exists a constant M (which could change line by line) such that

$$E(g^j, d^j) = \int_\Omega h|Dg^j| + \lambda \int_\Omega |Dd^j| + \frac{1}{2} \int_\Omega |g^j - f - d^j|^2 dx + \frac{\gamma}{2} \int_\Omega |d^j - d_0|^2 dx \leq M.$$

As $d_0 \in L^2(\Omega)$, we have $\int_\Omega |d_0|^2 dx \leq M$. Observing that $\int_\Omega |d^j - d_0|^2 dx \leq M$, we have

$$\int_\Omega |d^j|^2 dx = \int_\Omega |(d^j - d_0) + d_0|^2 dx \leq 2 \int_\Omega |d^j - d_0|^2 dx + 2 \int_\Omega |d_0|^2 dx \leq M.$$

Then the boundedness of $||d^j||_1$ is automatically obtained by Holder's inequality:

$$||d^j||_1 \leq |\Omega|^{\frac{1}{2}} ||d^j||_2 \leq M.$$

Since $\int_\Omega |Dd^j| \leq M$, we can deduce that $\{d^j\}$ is uniformly bounded in $BV(\Omega)$. Thus, there exist a subsequence (also denoted by $\{d^j\}$) and $d^* \in BV(\Omega)$ such that

$$(4.1) \qquad d^j \underset{L^1(\Omega)}{\to} d^*, \quad d^j \underset{L^2(\Omega)}{\rightharpoonup} d^*, \quad \text{and} \quad d^j \underset{BV-\mathrm{w}^*}{\rightharpoonup} d^*,$$

where $d^j \underset{L^1(\Omega)}{\to} d^*$ means that d^j converges strongly to d^* in $L^1(\Omega)$, $d^j \underset{L^2(\Omega)}{\rightharpoonup} d^*$ means that d^j converges weakly to d^* in $L^2(\Omega)$, and $d^j \underset{BV-\mathrm{w}^*}{\rightharpoonup} d^*$ means that d^j converges to d^* for the weak* topology of $BV(\Omega)$.

Meanwhile, from (3.9), we take $\delta_0 = \frac{1}{1 + ke^{-5 \inf_x d_0(x)}}$ and then $\delta_0 \in (0, 1)$. As a consequence, we have $h \geq \delta_0 \geq \frac{1}{1+k}$. By Lemma 4.2, $\{g^j\}$, where the index j is the same as (4.1), meets

$$(4.2) \qquad \delta_0 \int_\Omega |Dg^j| \leq \int_\Omega h|Dg^j| \leq M$$

and

$$\frac{1}{2} \int_\Omega |g^j - f - d^j|^2 dx \leq M.$$

Because $\boldsymbol{f}$ is given in $\mathbf{L^2}(\Omega, \mathbb{R}^3)$, the L^2 norm of $\boldsymbol{g}^j$ is uniformly bounded. Indeed,

$$\|\boldsymbol{g}^j\|_2 = \|(\boldsymbol{g}^j - \boldsymbol{f} - \boldsymbol{d}^j) + \boldsymbol{f} + \boldsymbol{d}^j\|_2 \leq \|\boldsymbol{g}^j - \boldsymbol{f} - \boldsymbol{d}^j\|_2 + \|\boldsymbol{f}\|_2 + \|\boldsymbol{d}^j\|_2 \leq M.$$

Then we obtain the boundedness of $||\boldsymbol{g}^j||_1$ by

$$||\boldsymbol{g}^j||_1 \leq |\Omega|^{\frac{1}{2}} ||\boldsymbol{g}^j||_2 \leq M.$$

Combining this with (4.2), we can conclude that $\{\boldsymbol{g}^j\}$ is bounded in $\mathbf{BV}(\Omega, \mathbb{R}^3)$. Hence, there exist a subsequence (also denoted by $\{\boldsymbol{g}^j\}$) and $\boldsymbol{g}^* \in \mathbf{BV}(\Omega, \mathbb{R}^3)$ such that

$$(4.3) \qquad \boldsymbol{g}^j \underset{\mathbf{L^1(\Omega)}}{\longrightarrow} \boldsymbol{g}^*, \quad \boldsymbol{g}^j \underset{\mathbf{L^2(\Omega)}}{\rightharpoonup} \boldsymbol{g}^*, \quad \text{and} \quad \boldsymbol{g}^j \underset{\mathbf{BV-w^*}}{\rightharpoonup} \boldsymbol{g}^*.$$

Therefore, up to a subsequence, $\{(\boldsymbol{g}^j, d^j)\}$ satisfies (4.1) and (4.3).

Because of the weak lower semicontinuity (w.l.s.c.) of the L^2 and BV norms, we have

$$\liminf_{j \to +\infty} \lambda \int_\Omega |Dd^j| + \frac{\gamma}{2} \int_\Omega |d^j - d_0|^2 dx \geq \lambda \int_\Omega |Dd^*| + \frac{\gamma}{2} \int_\Omega |d^* - d_0|^2 dx.$$

Since $\boldsymbol{g}^j - \boldsymbol{d}^j \rightharpoonup \boldsymbol{g}^* - \boldsymbol{d}^*$, we deduce that

$$\liminf_{j \to +\infty} \frac{1}{2} \int_\Omega |\boldsymbol{g}^j - \boldsymbol{f} - \boldsymbol{d}^j|^2 dx \geq \frac{1}{2} \int_\Omega |\boldsymbol{g}^* - \boldsymbol{f} - \boldsymbol{d}^*|^2 dx.$$

Finally, due to the lower semicontinuity of WVTV (see Lemma 4.1), we get

$$\liminf_{j \to +\infty} \int_\Omega h|D\boldsymbol{g}^j| \geq \int_\Omega h|D\boldsymbol{g}^*|.$$

Therefore, we can conclude that

$$\min_{(\boldsymbol{g},d) \in \Lambda} E(\boldsymbol{g}, d) = \liminf_{j \to +\infty} E(\boldsymbol{g}^j, d^j) \geq E(\boldsymbol{g}^*, d^*).$$

Hence, $(\boldsymbol{g}^*, d^*)$ is a minimal point of $E(\boldsymbol{g}, d)$.

(b) *Uniqueness.* All the terms in (3.8) are convex, so E is convex. Moreover, the Hessian matrix of $E_0(\boldsymbol{g}, d) := \frac{1}{2} \int_\Omega |\boldsymbol{g} - \boldsymbol{f} - \boldsymbol{d}|^2 dx + \frac{\gamma}{2} \int_\Omega |d - d_0|^2 dx$ is

$$H = \begin{bmatrix} 1 & -3 \\ -1 & 3 + \gamma \end{bmatrix}.$$

As $\gamma > 0$, H is positive definite, and E_0 and E are strictly convex.

Therefore, we can conclude that the minimization problem (3.11) admits a unique solution. ∎

5. Numerical schemes. In this section, the numerical procedure of the proposed energy (3.11) will be implemented. Note that (3.11) can be solved by many efficient methods, such as the augmented Lagrangian method [40] and split Bregman iteration [21]. In this paper, we adopt the multiple-splitting Chambolle–Pock algorithm [10, 13] due to its simplicity and the control of the dual gap.

As shown in [10], the Chambolle–Pock algorithm is a first-order primal-dual algorithm for convex optimization problems with known saddle-point structure. It is a promising method which can solve our nondifferentiable convex optimization problem without inner loop and can reduce the time and space overhead significantly. It has been proved that the Chambolle–Pock algorithm can converge to a saddle-point with rate $O(1/N)$ in finite dimensions. The wide applications in many image processing problems, such as image segmentation, image denoising, and image inpainting, have also demonstrated the superiority of the Chambolle–Pock algorithm [10].

First, we give some notations. In the discrete setting, let $\mathbf{X}$ denote a real vector space $\mathbf{R}^{N_1 \times N_2}$, $g_c \in \mathbf{X}$ denote the discrete image with size $N_1 \times N_2$, $\langle \cdot \rangle$ denote the Euclidean inner product, and I denote the identity matrix. Since $D\mathbf{g}, Dd$ are usually treated as $\nabla \mathbf{g}, \nabla d$ in difference schemes [11] (recall that D and ∇ denote the distributional gradient and classical gradient, respectively), we use $\nabla \mathbf{g}, \nabla d$ instead of $D\mathbf{g}, Dd$ in the following numerical schemes. The discrete gradients $\nabla \mathbf{g}, \nabla d$ are defined as

$$\nabla \mathbf{g} = (\nabla g_R, \nabla g_G, \nabla g_B) = \begin{pmatrix} \nabla_x g_R & \nabla_x g_G & \nabla_x g_B \\ \nabla_y g_R & \nabla_y g_G & \nabla_y g_B \end{pmatrix}, \quad \nabla d = \begin{pmatrix} \nabla_x d \\ \nabla_y d \end{pmatrix},$$

where ∇_x and ∇_y are the discrete derivative operators in the x- and y-directions, respectively. In our numerical experiments, ∇_x and ∇_y are obtained by using finite difference approximations to the derivatives with symmetric boundary conditions in the respective coordinate directions. For example, we define

$$(\nabla_x d)_{i,j} = \begin{cases} d_{i+1,j} - d_{i,j}, & i = 1, 2, \ldots, n-1, \\ 0, & i = n. \end{cases}$$

In addition, $\|\nabla \mathbf{g}\|_1$ and $\|\nabla d\|_1$ denote the discrete total variations of $\mathbf{g}$ and d; i.e.,

$$\|\nabla \mathbf{g}\|_1 = \sum_{i,j} \sqrt{\|\nabla g_R\|_1^2 + \|\nabla g_G\|_1^2 + \|\nabla g_B\|_1^2} = \sum_{i,j} \sqrt{\sum_{c \in \{R,G,B\}} \left\{ (\nabla_x g_c)_{i,j}^2 + (\nabla_y g_c)_{i,j}^2 \right\}},$$

$$\|\nabla d\|_1 = \sum_{i,j} \sqrt{(\nabla_x d)_{i,j}^2 + (\nabla_y d)_{i,j}^2}.$$

Overall, the discrete version of (3.11) is

$$(5.1) \qquad \min_{\mathbf{g},d} E_d(\mathbf{g}, d) := \|h\nabla \mathbf{g}\|_1 + \lambda \|\nabla d\|_1 + \frac{1}{2}\|\mathbf{g} - \mathbf{f} - \mathbf{d}\|^2 + \frac{\gamma}{2}\|d - d_0\|^2,$$

and we can rewrite it as the following equivalent constrained problem:

$$(5.2) \quad \min_{\mathbf{g},d,\mathbf{w_1},w_2} \|h\mathbf{w_1}\|_1 + \lambda \|w_2\|_1 + \frac{1}{2}\|\mathbf{g} - \mathbf{f} - \mathbf{d}\|^2 + \frac{\gamma}{2}\|d - d_0\|^2, \quad \text{s.t. } \mathbf{w_1} = \nabla \mathbf{g}, \ w_2 = \nabla d.$$

To solve this problem, similarly to [13], let us convert it to the primal-dual problem using the Lagrange multipliers method,

$$\max_{\boldsymbol{v_1}, v_2} \min_{\boldsymbol{g}, d, \boldsymbol{w_1}, w_2} L(\boldsymbol{g}, d, \boldsymbol{w_1}, w_2, \boldsymbol{v_1}, v_2) = \|h\boldsymbol{w_1}\|_1 + \lambda\|w_2\|_1 + \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{f} - \boldsymbol{d}\|^2 + \frac{\gamma}{2}\|d - d_0\|^2$$

$$(5.3) \qquad\qquad\qquad\qquad + \langle \boldsymbol{v_1}, \boldsymbol{w_1} - \nabla\boldsymbol{g}\rangle + \langle v_2, w_2 - \nabla d\rangle.$$

Let $\boldsymbol{y} = [\boldsymbol{v_1}, v_2]^T$ and $\boldsymbol{x} = [\boldsymbol{g}, d, \boldsymbol{w_1}, w_2]^T$. Denote

$$K := \begin{bmatrix} -\nabla^T & 0 & I & 0 \\ 0 & -\nabla^T & 0 & I \end{bmatrix}$$

as a continuous linear operator with the induced norm

$$(5.4) \qquad\qquad \|K\| = \max\{\|K\boldsymbol{x}\| : \ \boldsymbol{x} \in \mathbf{X}, \ \|\boldsymbol{x}\| \leq 1\}.$$

The min-max problem (5.3) can be written as

$$(5.5) \qquad\qquad \max_{\boldsymbol{y}} \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{y}) := \langle K\boldsymbol{x}, \boldsymbol{y}\rangle + G(\boldsymbol{x}) - F^*(\boldsymbol{y}),$$

where $G(\boldsymbol{x}) = \|h\boldsymbol{w_1}\|_1 + \lambda\|w_2\|_1 + \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{f} - \boldsymbol{d}\|^2 + \frac{\gamma}{2}\|d - d_0\|^2$ and $F^*(\boldsymbol{y}) = 0$. Note that readily we have

$$F(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

and

$$G^*(-K^T\boldsymbol{y}) = \frac{1}{2}\|\nabla^T\boldsymbol{v_1}\|^2 + \langle \boldsymbol{f}, \nabla^T\boldsymbol{v_1}\rangle + \frac{1}{2\gamma}\|\sum \nabla^T\boldsymbol{v_{1c}} + \nabla^T v_2\|^2 + \langle d_0, \sum \nabla^T\boldsymbol{v_{1c}} + \nabla^T v_2\rangle.$$

Similarly to [10, 13], the primal problem related to the above min-max problem is

$$\min_{\boldsymbol{x}} F(K\boldsymbol{x}) + G(\boldsymbol{x})$$

or

$$\min_{K\boldsymbol{x}=0} G(\boldsymbol{x}).$$

This is exactly our previous problem (5.2).

Proposition 5.1. *The norm of K (see paragraph before (5.4)) is bounded by $\|K\| \leq 3$.*

Proof. Here, we note that $K\boldsymbol{x} = [(\boldsymbol{w_1} - \nabla\boldsymbol{g})^T, \ (w_2 - \nabla d)^T]^T$. Since $\|\nabla\|^2 \leq 8$ (see [11], which can be readily extended to the vectorial case), we deduce that

$$\|K\boldsymbol{x}\|^2 = \|\boldsymbol{w_1} - \nabla\boldsymbol{g}\|^2 + \|w_2 - \nabla d\|^2$$
$$\leq (1 + \tfrac{1}{s})\|\boldsymbol{w_1}\|^2 + (1 + s)\|\nabla\boldsymbol{g}\|^2 + (1 + \tfrac{1}{t})\|w_2\|^2 + (1 + t)\|\nabla d\|^2$$
$$= (1 + \tfrac{1}{s})\|\boldsymbol{w_1}\|^2 + (1 + s)\left(\|\nabla g_R\|^2 + \|\nabla g_G\|^2 + \|\nabla g_B\|^2\right) + (1 + \tfrac{1}{t})\|w_2\|^2 + (1 + t)\|\nabla d\|^2$$
$$\leq (1 + \tfrac{1}{s})\|\boldsymbol{w_1}\|^2 + (1 + s)\left(8\|g_R\|^2 + 8\|g_G\|^2 + 8\|g_B\|^2\right) + (1 + \tfrac{1}{t})\|w_2\|^2 + (1 + t)\|\nabla d\|^2$$
$$\leq (1 + \tfrac{1}{s})\|\boldsymbol{w_1}\|^2 + 8(1 + s)\|\boldsymbol{g}\|^2 + (1 + \tfrac{1}{t})\|w_2\|^2 + 8(1 + t)\|d\|^2,$$

where s, t can be any positive constants.

In particular, we can choose the ones such that the coefficients of the last inequality are the same: $1 + \frac{1}{s} = 8(1 + s) = 1 + \frac{1}{t} = 8(1 + t)$. This leads to $t = s = \frac{1}{8}$. Moreover,

$$\|K\boldsymbol{x}\|^2 \le 9(\|\boldsymbol{w_1}\|^2 + \|\boldsymbol{g}\|^2 + \|w_2\|^2 + \|d\|^2) \le 9\|\boldsymbol{x}\|^2.$$

According to (5.4), we have $\|K\| \le 3$. ∎

Since both $G(\cdot)$ and $F^*(\cdot)$ are proper, convex, and lower semicontinuous functionals, (5.5) is exactly the same problem in [10], and thus we can use the Chambolle–Pock algorithm to solve our model.

The regular Chambolle–Pock algorithm of (5.5) is as follows:

$$(5.6) \qquad \begin{cases} \boldsymbol{y}^{n+1} = \arg\max_{\boldsymbol{y}} L(\bar{\boldsymbol{x}}, \boldsymbol{y}) - \frac{1}{2\tau}\|\boldsymbol{y} - \boldsymbol{y}^n\|^2, \\[2mm] \boldsymbol{x}^{n+1} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{y}^{n+1}) + \frac{1}{2\sigma}\|\boldsymbol{x} - \boldsymbol{x}^n\|^2, \\[2mm] \bar{\boldsymbol{x}} = 2\boldsymbol{x}^{n+1} - \boldsymbol{x}^n, \end{cases}$$

(5.7)

(5.8)

where σ and τ are two positive parameters.

In detail, (5.6) can be rewritten as

$$(\boldsymbol{v_1}^{n+1}, v_2{}^{n+1}) = \arg\max_{\boldsymbol{v_1}, v_2} \langle \boldsymbol{v_1}, \bar{\boldsymbol{w}}_1 - \nabla \bar{g}\rangle + \langle v_2, \bar{w}_2 - \nabla \bar{d}\rangle - \frac{1}{2\tau}(\|\boldsymbol{v_1} - \boldsymbol{v_1}^n\|^2 + \|v_2 - v_2{}^n\|^2).$$

Obviously, the above formula is equal to

$$\begin{cases} \boldsymbol{v_1}^{n+1} = \arg\max_{\boldsymbol{v_1}} \langle \boldsymbol{v_1}, \bar{\boldsymbol{w}}_1 - \nabla \bar{g}\rangle - \frac{1}{2\tau}\|\boldsymbol{v_1} - \boldsymbol{v_1}^n\|^2, \\[2mm] v_2{}^{n+1} = \arg\max_{v_2} \langle v_2, \bar{w}_2 - \nabla \bar{d}\rangle - \frac{1}{2\tau}\|v_2 - v_2{}^n\|^2. \end{cases}$$

Then we can obtain $\boldsymbol{v_1}^{n+1}$ and $v_2{}^{n+1}$ by

$$(5.9) \qquad \begin{cases} \boldsymbol{v_1}^{n+1} = \tau(\bar{\boldsymbol{w}}_1 - \nabla \bar{g}) + \boldsymbol{v_1}^n, \\[2mm] v_2{}^{n+1} = \tau(\bar{w}_2 - \nabla \bar{d}) + v_2{}^n. \end{cases}$$

(5.10)

The subproblem (5.7) can be rewritten as

$$\begin{aligned} (\boldsymbol{g}^{n+1}, d^{n+1}, \boldsymbol{w_1}^{n+1}, w_2{}^{n+1}) = \arg\min_{\boldsymbol{g}, \boldsymbol{w_1}, d, w_2} &\|h\boldsymbol{w_1}\|_1 + \lambda\|w_2\|_1 + \tfrac{1}{2}\|\boldsymbol{g} - \boldsymbol{f} - d\|^2 \\ &+ \tfrac{\gamma}{2}\|d - d_0\|^2 + \langle \boldsymbol{v_1}^{n+1}, \boldsymbol{w_1} - \nabla \boldsymbol{g}\rangle + \langle v_2{}^{n+1}, w_2 - \nabla d\rangle \\ &+ \tfrac{1}{2\sigma}(\|\boldsymbol{g} - \boldsymbol{g}^n\|^2 + \|\boldsymbol{w_1} - \boldsymbol{w_1}^n\|^2 + \|d - d^n\|^2 + \|w_2 - w_2{}^n\|^2). \end{aligned}$$

It is equivalent to the following form:

$$(5.11) \qquad \begin{cases} (\boldsymbol{g}^{n+1}, d^{n+1}) = \arg\min_{\boldsymbol{g}, d} \tfrac{1}{2}\|\boldsymbol{g} - \boldsymbol{f} - d\|^2 + \tfrac{\gamma}{2}\|d - d_0\|^2 - \langle \boldsymbol{v_1}^{n+1}, \nabla \boldsymbol{g}\rangle \\ \qquad\qquad + \langle v_2{}^{n+1}, -\nabla d\rangle + \tfrac{1}{2\sigma}(\|\boldsymbol{g} - \boldsymbol{g}^n\|^2 + \|d - d^n\|^2), \\[2mm] \boldsymbol{w_1}^{n+1} = \arg\min_{\boldsymbol{w_1}} \|h\boldsymbol{w_1}\|_1 + \langle \boldsymbol{v_1}^{n+1}, \boldsymbol{w_1}\rangle + \tfrac{1}{2\sigma}\|\boldsymbol{w_1} - \boldsymbol{w_1}^n\|^2, \\[2mm] w_2{}^{n+1} = \arg\min_{w_2} \lambda\|w_2\|_1 + \langle v_2{}^{n+1}, w_2\rangle + \tfrac{1}{2\sigma}\|w_2 - w_2{}^n\|^2. \end{cases}$$

(5.12)

(5.13)

For (5.11), we have the following proposition.

Proposition 5.2. *The step* (5.11) *has a closed-form solution.*

Proof. We first obtain the first-order variations with respect to $\boldsymbol{g}$ and d:

$$\begin{cases} (\boldsymbol{g}^{n+1} - \boldsymbol{f} - \boldsymbol{d}^{n+1}) + \nabla^T \boldsymbol{v}_1^{n+1} + \frac{1}{\sigma}(\boldsymbol{g}^{n+1} - \boldsymbol{g}^n) = 0, \\ -\sum_c (g_c^{n+1} - f_c - d^{n+1}) + \gamma(d^{n+1} - d_0) + \nabla^T v_2^{n+1} + \frac{1}{\sigma}(d^{n+1} - d^n) = 0. \end{cases}$$

Then the closed-form solution of $\boldsymbol{g}$ and d can be obtained from the above linear system:

$$\begin{cases} d^{n+1} = \dfrac{(1 + \frac{1}{\sigma})B + A}{(1 + \frac{1}{\sigma})(3 + \gamma + \frac{1}{\sigma}) - 3}, \\ \boldsymbol{g}^{n+1} = \dfrac{\boldsymbol{f} + d^{n+1} - \nabla^T \boldsymbol{v}_1^{n+1} + \frac{1}{\sigma}\boldsymbol{g}^n}{1 + \frac{1}{\sigma}}, \end{cases}$$

where $A = \sum_c f_c + \frac{1}{\sigma}\sum_c g_c^n - \sum_c \nabla^T \boldsymbol{v}_{1c}^{n+1}$, $B = -\sum_c f_c + \gamma d_0 + \frac{1}{\sigma}d^n - \nabla^T v_2^{n+1}$, and $\boldsymbol{v}_1 = \{v_{1_R}, v_{1_G}, v_{1_B}\}$. ∎

The solutions of (5.12) and (5.13) can be directly solved by

$$(5.14) \qquad \begin{cases} \boldsymbol{w}_1^{n+1} = \dfrac{\boldsymbol{w}_1^n - \sigma \boldsymbol{v}_1^{n+1}}{|\boldsymbol{w}_1^n - \sigma \boldsymbol{v}_1^{n+1}|} \max(|\boldsymbol{w}_1^n - \sigma \boldsymbol{v}_1^{n+1}| - \sigma h, 0), \\[2mm] (5.15) \qquad w_2^{n+1} = \dfrac{w_2^n - \sigma v_2^{n+1}}{|w_2^n - \sigma v_2^{n+1}|} \max(|w_2^n - \sigma v_2^{n+1}| - \sigma \lambda, 0). \end{cases}$$

Additionally, the formula (5.8) is given as follows:

$$\begin{cases} (5.16) \qquad \bar{d} = 2d^{n+1} - d^n, \\ (5.17) \qquad \bar{\boldsymbol{g}} = 2\boldsymbol{g}^{n+1} - \boldsymbol{g}^n, \\ (5.18) \qquad \bar{\boldsymbol{w}}_1 = 2\boldsymbol{w}_1^{n+1} - \boldsymbol{w}_1^n, \\ (5.19) \qquad \bar{w}_2 = 2w_2^{n+1} - w_2^n. \end{cases}$$

For the convergence assessment, we propose using the primal-dual gap [10] as our stopping criterion. Classically, the primal-dual gap for (5.5) is defined as

$$(5.20) \qquad \mathcal{G}(\boldsymbol{x}, \boldsymbol{y}) := F(K\boldsymbol{x}) + G(\boldsymbol{x}) + F^*(\boldsymbol{y}) + G^*(-K^T \boldsymbol{y}).$$

Given a prescribed positive parameter ρ (very small, say 10^{-4}) and the initial energy $E_d(\boldsymbol{g}^0, d^0)$, where $\boldsymbol{g}^0 = \boldsymbol{f}$ and $d^0 = d_0$ are initial values of $\boldsymbol{g}$ and d, when the primal-dual gap is less than $\rho \cdot E_d(\boldsymbol{g}^0, d^0)$ our algorithm can be considered to reach a steady state.

Therefore, taking all of the above into account, the overall computation procedure of the proposed method is as presented in Algorithm 1.

The convergence of the proposed algorithm can be guaranteed by the following theorem.

Theorem 5.3. *Denote $L = \|K\|$, and choose the parameters σ, τ such that $\sigma \tau L^2 < 1$. Let $(\boldsymbol{x}^n, \boldsymbol{y}^n) = (\boldsymbol{g}^n, d^n, \boldsymbol{w}_1^n, w_2^n, \boldsymbol{v}_1^n, v_2^n)$ be the sequence derived from Algorithm 1. Then the following hold:*

Algorithm 1: The computational procedure of the Chambolle–Pock algorithm for the proposed variational model (5.1).

- Initialize: $\bar{d} = d^0 = d_0$, $\bar{g} = g^0 = f + d^0$, $v_1^0 = 0$, $v_2{}^0 = 0$, $\bar{w}_1 = w_1^0 = \nabla g^0$, $\bar{w}_2 = w_2{}^0 = \nabla d^0$

- Repeat until $\mathcal{G}(x^n, y^n) < \rho \cdot E_d(g^0, d^0)$:

$$
\begin{cases}
v_1^{n+1} = \tau(\bar{w}_1 - \nabla \bar{g}) + v_1^n, \\
v_2{}^{n+1} = \tau(\bar{w}_2 - \nabla \bar{d}) + v_2{}^n, \\
d^{n+1} = \frac{(1+\frac{1}{\sigma})B + A}{(1+\frac{1}{\sigma})(3+\gamma+\frac{1}{\sigma}) - 3}, \\
\boldsymbol{d}^{n+1} = (d^{n+1}, d^{n+1}, d^{n+1}), \\
g^{n+1} = \frac{f + d^{n+1} - \nabla^T v_1^{n+1} + \frac{1}{\sigma} g^n}{1 + \frac{1}{\sigma}}, \\
w_1^{n+1} = \frac{w_1^n - \sigma v_1^{n+1}}{|w_1^n - \sigma v_1^{n+1}|} \max(|w_1^n - \sigma v_1^{n+1}| - \sigma h, 0), \\
w_2{}^{n+1} = \frac{w_2{}^n - \sigma v_2{}^{n+1}}{|w_2{}^n - \sigma v_2{}^{n+1}|} \max(|w_2{}^n - \sigma v_2{}^{n+1}| - \sigma \lambda, 0), \\
\bar{d} = 2d^{n+1} - d^n, \\
\bar{g} = 2g^{n+1} - g^n, \\
\bar{w}_1 = 2w_1^{n+1} - w_1^n, \\
\bar{w}_2 = 2w_2{}^{n+1} - w_2{}^n.
\end{cases}
$$

(a) *there exists a saddle-point* (x^*, y^*) *such that*

$$x^n \to x^*, \qquad y^n \to x^*;$$

(b) *the* (g^*, d^*) *is exactly the solution of* (5.1).

Proof. One can easily prove (a) following the proof of Theorem 1 in [10]. Further, as shown in Proposition 3.1 in [19], (a) is equivalent to (b). Thus Theorem 5.3 is proved. ∎

6. Experimental results and analysis. In this section, in order to examine the effectiveness of the proposed method, we describe and analyze the experimental results on some natural hazy images. Note that all the following experiments are implemented in MATLAB R2012b on an Intel 3.33 GHz PC with 8 GB RAM. In what follows, we set the parameters of the depth map (2.2) to $\eta = 1$; the parameters of energy (3.8) to $k = 50$, $\lambda = 0.01$, and $\gamma = 0.1$; and the stopping criterion of Algorithm 1 to $\rho = 10^{-4}$. Further, according to Theorem 5.3, we need $\sigma\tau L^2 < 1$ to ensure convergence. Since $L = \|K\| \leq 3$, we need $\sigma\tau < 1/9$. In our experiments, we set $\sigma = 1$ and $\tau = 0.11$.

6.1. The procedure of the proposed method. We now display the overall procedure of the proposed method. In Figure 1, (a) is the original input, (b) is a haze-free image using the dark channel prior presented by [24], (c) is the estimated transmission map using our windows adaptive method (see section 3.1), and (d) is a recovered haze-free image using (c) and formula (2.6). Clearly, (d) has no edge artifact, while there are some edge artifacts on the boundary of some objects in (b). That is to say, the proposed windows adaptive method outperforms the method of estimating the transmission from [24]. However, (d) is dark in intensity, low in contrast, and contains noise. To improve these aspects, we further transform
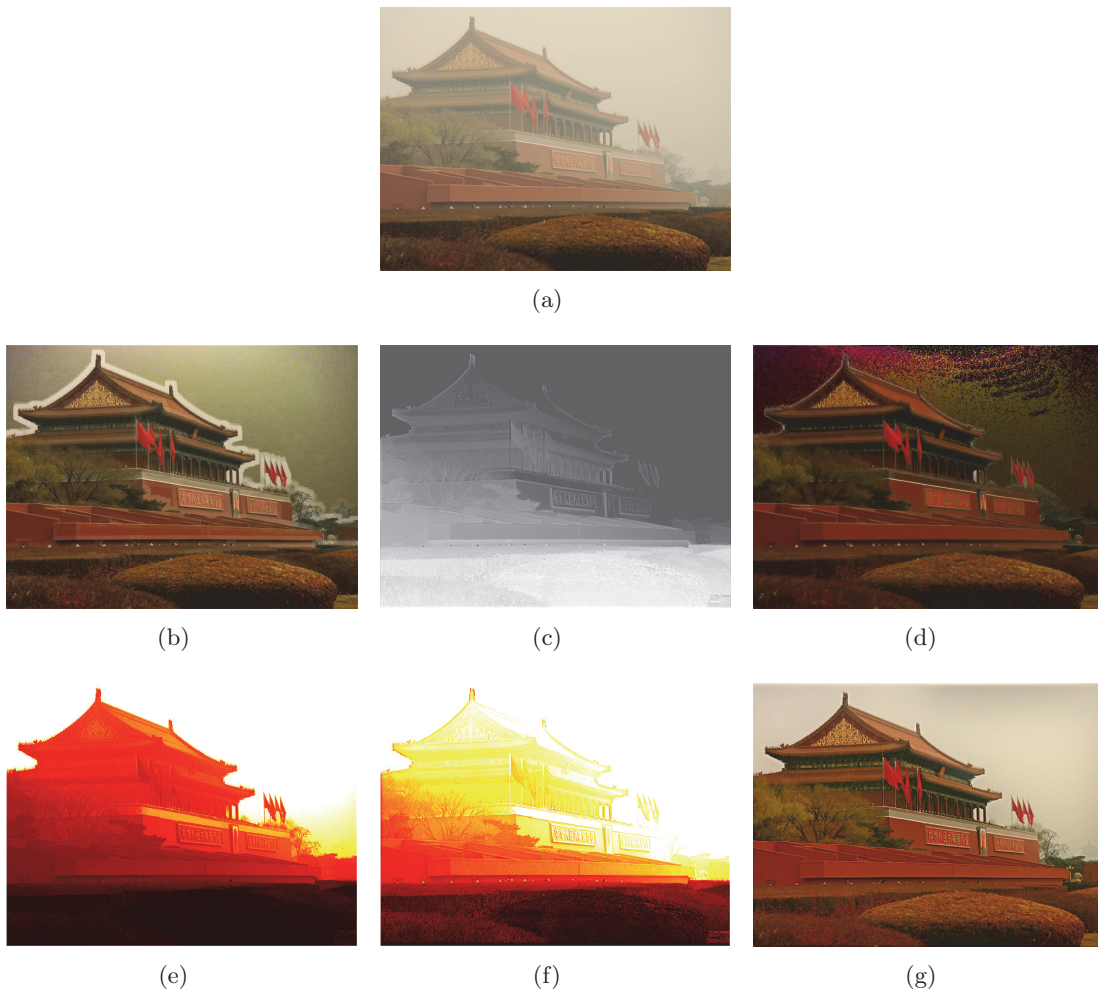
Figure 1. *The procedure of the proposed method:* (a) *The original input image (600×450),* (b) *the recovered haze-free image using a dark channel prior,* (c) *the estimated transmission map of* (a) *using the windows adaptive method,* (d) *the recovered haze-free image using* (b), (e) *the depth map corresponding to* (b) *(shown in false colors, similarly hereinafter),* (f) *the refined depth map using the proposed method,* (g) *recovered haze-free and noise-free image obtained using the proposed method.*

(c) to the depth map (see (e)), and then obtain our final result using the energy (3.11). Note that (f) and (g) are the final depth map and recovered haze-free and noise-free image, respectively. Obviously, (g) is bright-colored, vivid, and contains little noise.

The convergence curve for Figure 1 is presented in Figure 2. The convergence curve is actually the sequence of the primal-dual gap derived from Algorithm 1. We can see from the figure that our algorithm converges rapidly.

Here note that $E_d(\boldsymbol{g}^0, d^0) = 1.239 \times 10^5$. The stopping criterion is given by $\mathcal{G}(\boldsymbol{x}^n, \boldsymbol{y}^n) < 12.39$; see condition (5.20). In this case, the number of iterations is $n = 68$.

6.2. Comparison study. In this part, we will present some comparison studies from the vision observation and numerical efficiency aspects using six images.
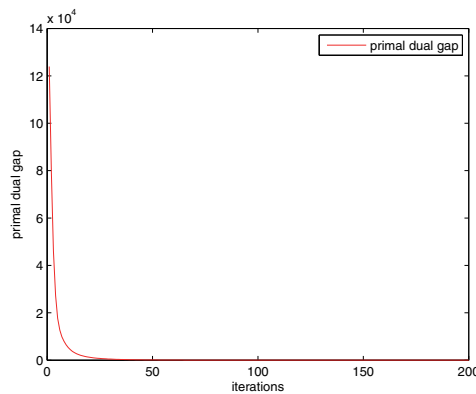
Figure 2. *The convergence curve (i.e., the sequence of the primal-dual gap, derived from Algorithm 1) of Figure 1.*

Comparison 1: First, we compare our method with a dehazing model as well as a dehaze+denoise procedure, where we select the work of He, Sun, and Tang [24] as the representative of the dehazing model, and the dehaze+denoise procedure is the procedure in which we first dehaze the image using that model and then denoise it using the ROF model [36] (note that the regularizing parameter of the ROF model is set as 0.1 in our experiment). The results are shown in Figures 3–5.

In Figure 3, we first generate a hazy and noisy image by adding white Gaussian noise of standard deviation 0.1 to a hazy image (as shown in (a)), then use this synthesized image to obtain the images (b)–(d) by the dehazing model, the dehaze+denoise procedure, and the proposed model, respectively. To see the difference of these results more clearly, we show close-ups of an example area of Figure 3 (in the red box) in Figure 4.

As can be observed from Figures 3 and 4, although the result of the dehazing model is haze-free and informative, it contains some noise, which is particularly serious in the distant view. The result of the dehaze+denoise procedure contains little noise, but some details of the whole image are lost. Nevertheless, the proposed method has the advantages of both the dehazing model and the dehaze+denoise procedure. Indeed, our result contains more detail and less noise. For instance, the outline of mountains in the upper right corner is either noisy or blurry in Figures 4(b)–(c), while it is much more clear in Figure 4(d). The reason is that our model contains the WVTV term which can denoise effectively while keeping the edges. Therefore, our result outperforms the dehazing model and the dehaze+denoise procedure. Analogously, Figure 5, which shows a real hazy image and its outcomes, also demonstrates the superiority of our method.

Comparison 2: Next, we compare our method with those of Tan [41], Fattal [20], and He, Sun, and Tang [24]. These results are shown in the Figures 6–10. In each of these figures, (a) shows a hazy original input, (b)–(e) are the results of Tan's, Fattal's, He et al.'s, and our methods, respectively. Observing Figures 6 and 7, we can see that all these methods can dehaze effectively, and our results are comparable to others.

However, in Figure 8, the results show that Tan's method can dehaze effectively, while the resulting image is distorted in color and contains serious noise. Fattal's method can dehaze
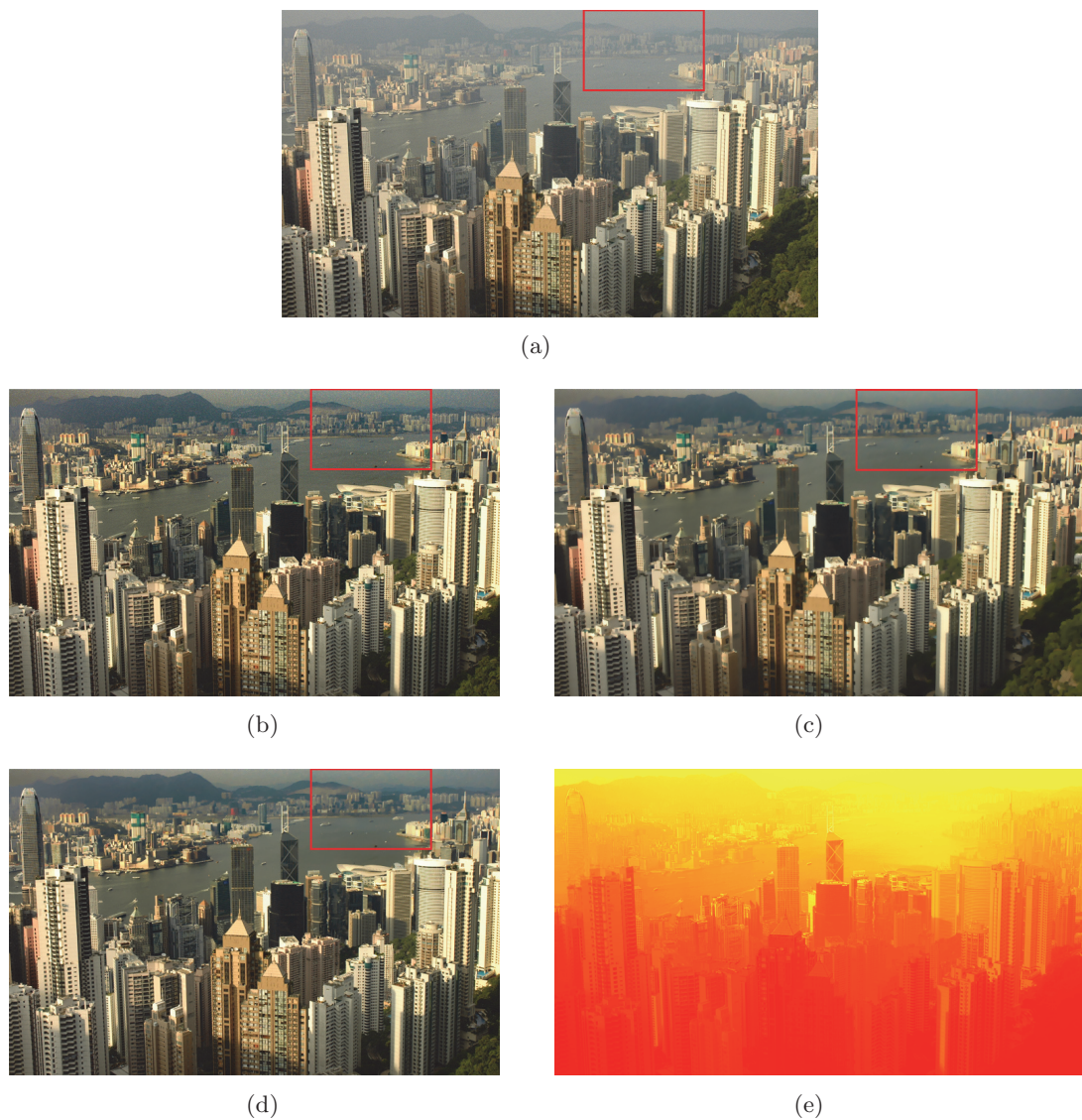
(a)

(b)

(c)

(d)

(e)

Figure 3. *The synthesized noisy image and processed results:* (a) *The synthesized noisy image (white Gaussian noise of standard deviation* 0.1, 457 × 800*); the results of* (b) *the dehazing model,* (c) *dehaze+denoise, and* (d) *the proposed method;* (e) *the depth map using the proposed method.*

with little oversaturated color and little noise, but it cannot remove the dense haze. He et al.'s method can dehaze without sacrificing the fidelity of the color, yet its output also has a lot of noise in the area with a distant view. In contrast, the result from the proposed method not only preserves color well, but also provides a high quality dehazed image with little noise. These differences are more apparent in Figure 9, in which close-up views of the white box areas in Figure 8 are displayed. For example, the roof edges in Figure 9(b)–(d) are unsharp due to the existence of noise or haze, while the boundary is clearer in Figure 9(e). Note that the block effects in (b) and (d) are not due to JPEG decompression, but due to the dehazing
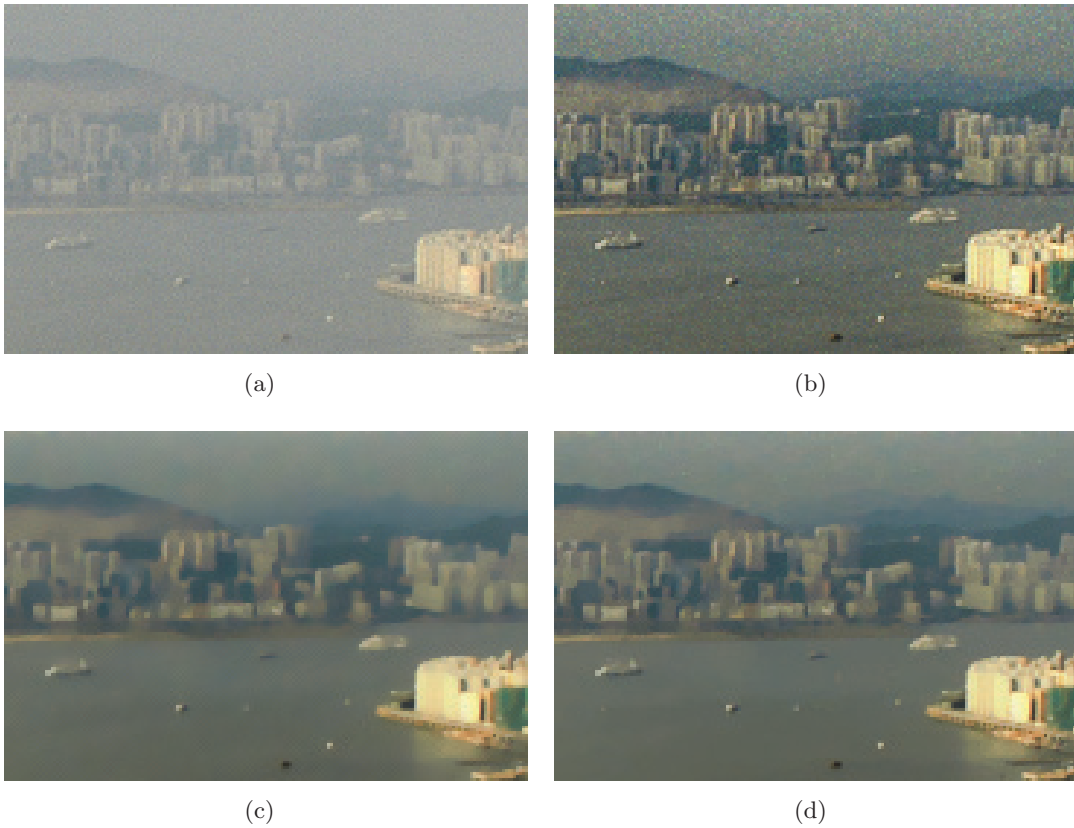
Figure 4. *Close-ups from the boxed area of Figure* 3 *(*120 × 180*): (a) The synthesized hazy image; the results of* (b) *the dehazing model,* (c) *dehaze+denoise, and* (d) *the proposed method.*

Table 1
Comparison of the proposed method with that of He, Sun, and Tang's work [24] *in time and memory usage.*

Figure	Image size	Time (s)		Total memory (Mb)		Peak memory (Mb)	
		He et al.	Proposed	He et al.	Proposed	He et al.	Proposed
Figure 3	457 × 800	44.101	**4.229**	1426.856	**680.080**	463.300	**17.176**
Figure 5	432 × 432	23.442	**2.810**	725.148	**281.800**	235.940	**8.764**
Figure 6	600 × 400	29.532	**3.387**	1117.536	**381.476**	303.704	**11.272**
Figure 7	1000 × 327	38.254	**3.744**	1350.516	**582.160**	413.880	**15.368**
Figure 8	600 × 400	29.755	**3.277**	1116.032	**390.784**	303.700	**11.272**
Figure 10	1024 × 768	94.376	**7.006**	2984.200	**1452.360**	998.892	**36.948**

methods. Hence, our method outperforms others to some extent. The obvious differences in the red and white boxes of each panel in Figure 10 also demonstrate the advantages of our method.

Efficiency comparison: To evaluate computational efficiency, the proposed method is further compared with the work of He, Sun, and Tang [23]. Precisely, we calculate time cost and memory usage for processing each image in Figures 3–10. The comparative results are shown in Table 1. It should be noted that the best value of each measure is marked in bold.
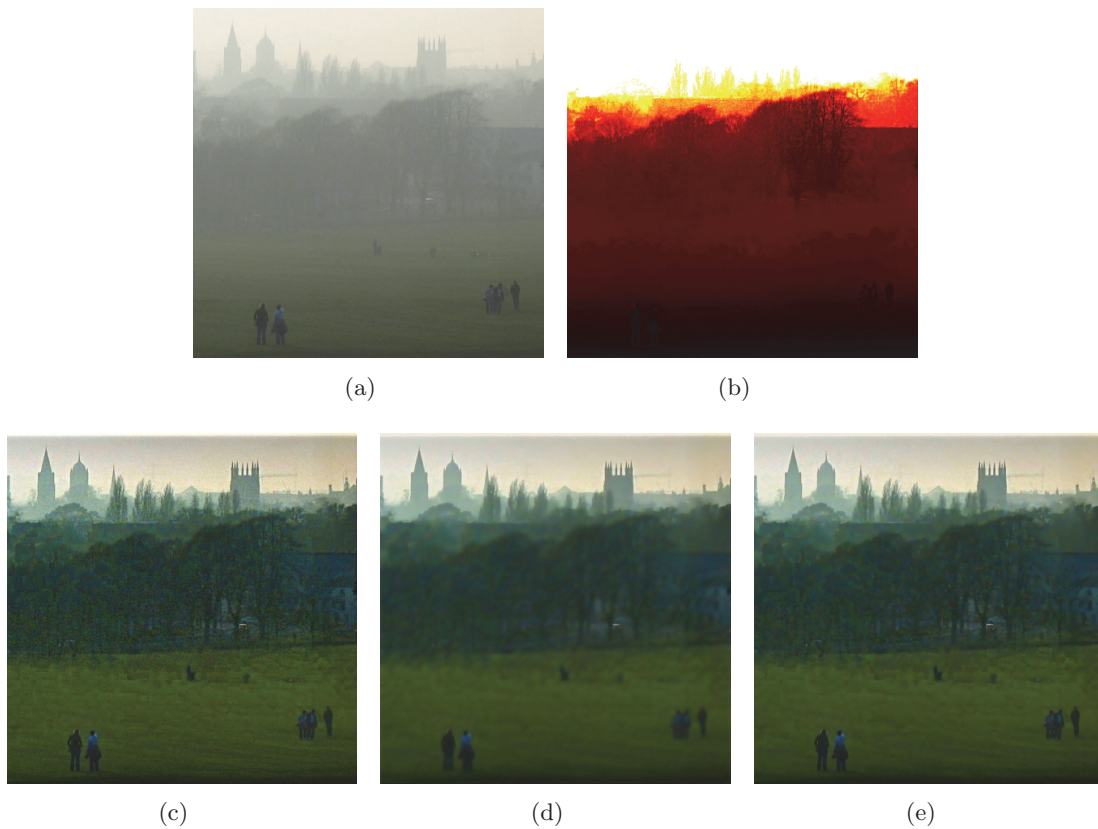
Figure 5. *Comparison with other methods:* (a) *The original input hazy image (*432×432*);* (b) *the depth map using the proposed method; the results of* (c) *the dehazing model,* (d) *dehaze+denoise, and* (e) *the proposed method.*

From Table 1, we find that the time cost of our method is about 10% that of the model of He, Sun, and Tang. That is, our method is much more efficient. The main reason is that the proposed method has no complex term, and we have applied the Chambolle–Pock algorithm, which has remarkable advantages in time overhead and RAM usage, to implement our energy, while there is a soft matting step which involves a large time-consuming matting Laplacian matrix in He, Sun, and Tang's model.

As for memory usage, we use two variables, the total memory and the peak memory, to measure efficiency. The total memory is the total amount of the available memory allocated to a process, and the peak memory is the maximal physical memory that the process really used. In Table 1, we can see that the total and peak memories used by our method are 50% and 4%, respectively, of those of He, Sun, and Tang's model. That is, our method highly outperforms their model in memory usage. As has been stated, a large Laplacian equation which is memory-consuming needs to be solved in He, Sun, and Tang's model. Since our method has no such memory-consuming term, the statistics in Table 1 verify the aforementioned standpoints.

6.3. The influence of parameters. As shown in (3.8), there are three parameters, i.e., k, λ, and γ, in our functional energy, where k is a nonnegative parameter that determines

Figure 6. *Comparison with other methods:* (a) *The original input hazy image (600 × 400),* (b) *Tan's result,* (c) *Fattal's result,* (d) *He et al.'s result,* (e) *the proposed result,* (f) *the depth map using the proposed method.*

the depth level of the image (see (3.9)), and λ, γ are two positive parameters that adjust the importance of the two corresponding energy terms. We now analyze the influence of these parameters.

We first study the effect of parameter k. In Figure 11, we display the results when the values of k are varied. We can see that, as k increases, the result contains more detailed information, while its noise increases simultaneously. Meanwhile, when $k = 10$ and 50, the results can not only retain the details but also reduce the noise. This phenomenon is also reflected in Figure 12, in which the unidimensional examples of function h with different k are
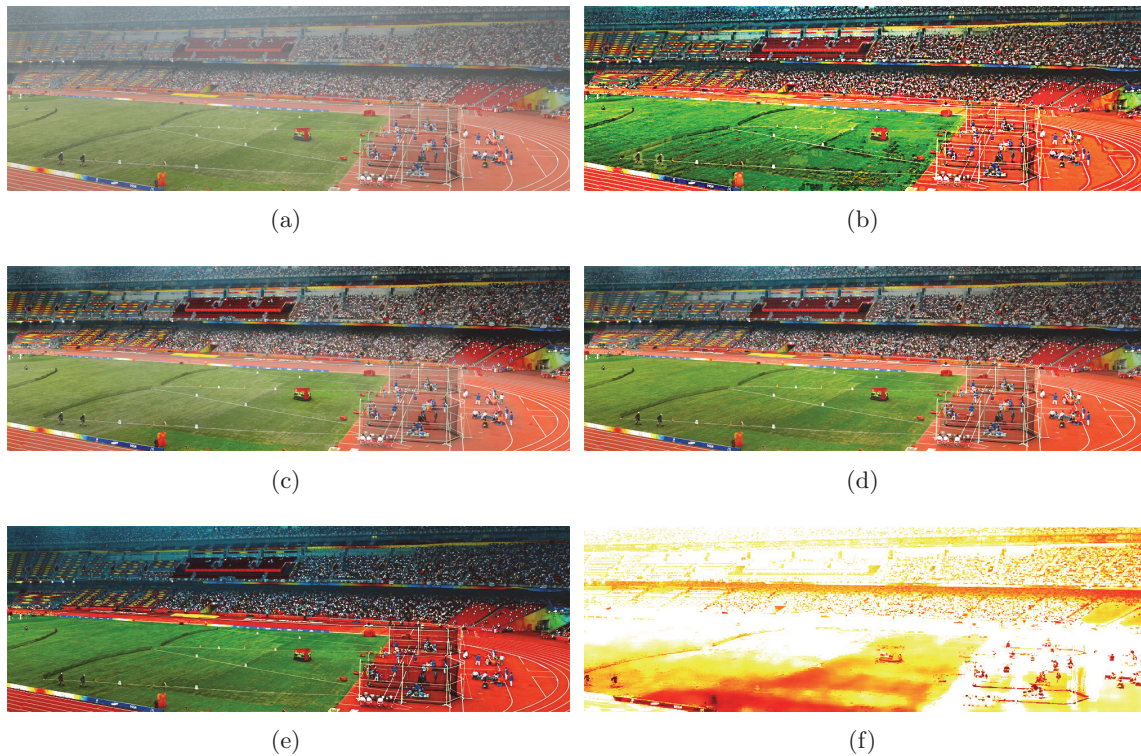
Figure 7. *Comparison with the methods of Tan, Fattal, and He, Sun, and Tang:* (a) *The original input hazy image (*1000×327*),* (b) *Tan's result,* (c) *Fattal's result,* (d) *He et al.'s result,* (e) *the proposed result,* (f) *the depth map of the proposed method.*

presented. In this figure, when k is too small, e.g., $k = 0$ or 0.001, all values of h approach 1. That is, the denoising procedure influenced by the depth map d_0 is very limited, and this leads to the lack of details for the whole image. In contrast, when k is too large, e.g., $k = 1000$, a lot of values of h approach 0, which causally leads to the noisy result. When $k = 10$ or 50, h is generally balanced. This leads to the removal of the noise when d_0 is large while keeping the details when d_0 is small. Since the dehazed image does not have much noise at close range (i.e., d_0 is small), the whole result has little noise and much detail. Therefore, based on the above analysis, $k \in [10, 50]$ is suitable. Further, as shown in Figure 11, the results change very little as k varies, so our method is rather stable with k.

In the next experiment, we test the quality of the result for different λ and γ. A lot of experiments have shown that, for an image with little noise, the difference is not obvious when λ and γ change. To make the comparison more obvious, we add white Gaussian noise of standard deviation 0.1 to Figure 1(a) and then produce results with different λ and γ. Here we fix k to be 50. The results are shown in Figure 13. We can see from Figure 13 that the results have less noise and become blurrier when λ increases and γ decreases. Thus we can set a bigger λ and smaller γ for images with much noise and set a smaller λ and bigger γ for images with little noise. It should be noted, since very small γ will lead to an unstable result, that we can't set γ to be too small. In our experiment, we find that $\gamma \in [0.005, 5]$ is suitable.
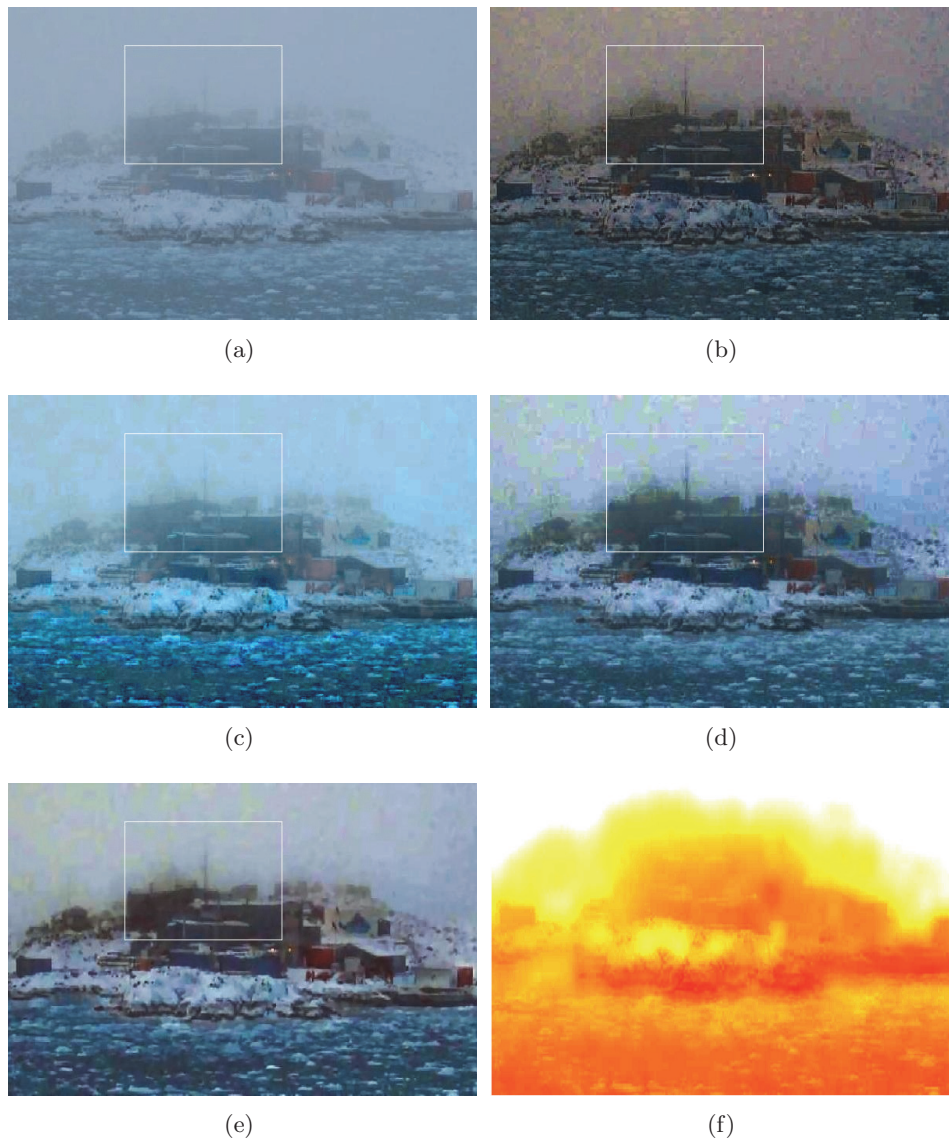
(a)   (b)

(c)   (d)

(e)   (f)

Figure 8. *Comparison with other methods:* (a) *The original input hazy image (600 × 400),* (b) *Tan's result,* (c) *Fattal's result,* (d) *He et al.'s result,* (e) *the proposed result,* (f) *the depth map using the proposed method.*

Meanwhile, we can also find that the images in Figure 13 change very little, which indicates that the proposed method is rather robust to λ and γ.

7. Conclusions. In this paper, we have introduced a new fast dehazing and denoising method based on the variational approach. The proposed method first estimates a transmission map using the windows adaptive method, and then converts the transmission map to a depth map. Using this depth map, we construct an energy functional to seek the final haze- and noise-free image. The existence and uniqueness of the solution of the proposed variational model are then discussed. To minimize the objective function efficiently, we adopt
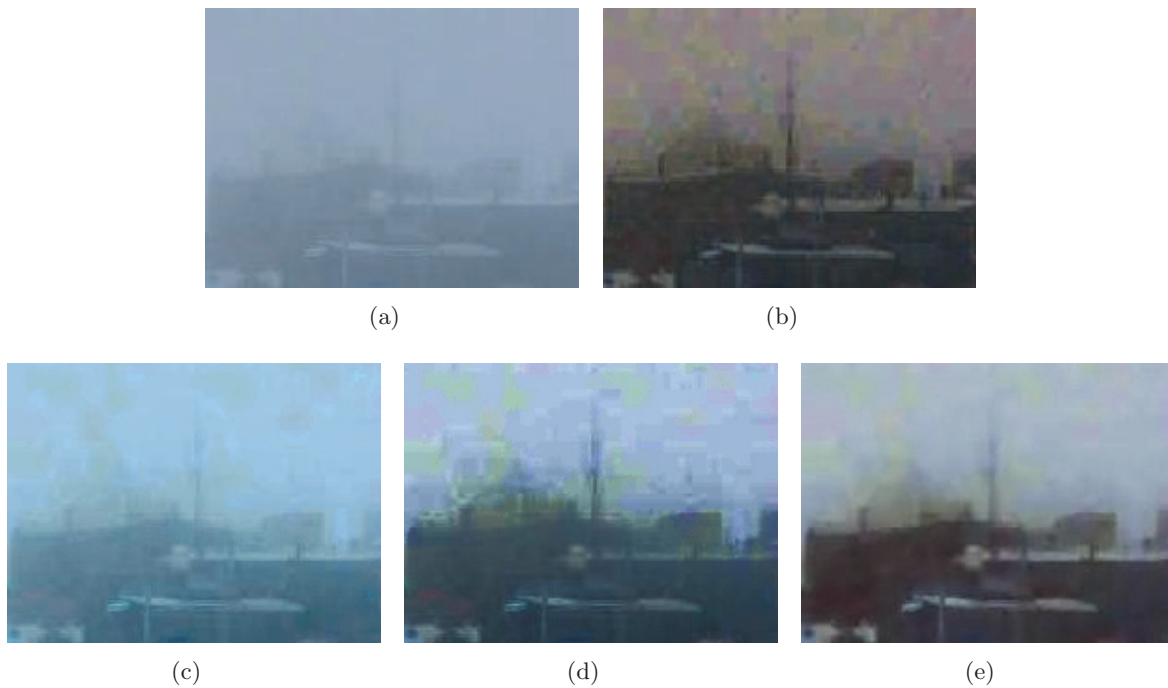
Figure 9. *Close-ups from the boxed area of Figure* 8 *(*200 × 150*):* (a) *The input hazy image,* (b) *Tan's result,* (c) *Fattal's result,* (d) *He et al.'s result,* (e) *the proposed result.*

the multiple-splitting Chambolle–Pock algorithm, which solves our problem without an inner loop. In order to prove its effectiveness, we tested the proposed method on many real scenes and compared its performance with that of several existing methods. The results demonstrate that our method is generally better than the compared methods especially in scenes with noise. Furthermore, the comparison of time and memory usage with those of He, Sun, and Tang's method also shows the efficiency of the proposed method. Therefore, we can conclude that the proposed method can remove haze and noise efficiently and effectively.

Simultaneously dehazing and denoising is still a challenging problem, and there are many open questions that should be studied. Our method is effective, but it also can be improved. First, since our windows adaptive method is based on the dark channel prior, it also can't deal with sky regions very reasonably. Till now, there are few suitable models handling sky regions more naturally. The building of models to handle the sky regions will be the aim of our further research. Second, our results are acceptable but some artifacts are still visible (see Figure 6(e), for example). There may be some difficulty in the choice of the parameter r used for the windows adaptive method. A constant r may not be suitable for all images. It is interesting to study some adaptive methods for the choice of r. Furthermore, our method may fail when the haze image model (2.1) is physically invalid. We will extend our research based on some more advanced models [35]. Additionally, as our method is a general framework, it can be extended in many possible ways, such as the application of total generalized variation [2] or tight-frame [7, 8, 9]. We leave this as further works.

(a)   (b)

(c)   (d)

(e)   (f)

Figure 10. *Comparison with other methods:* (a) *The original input hazy image (*1024×768*),* (b) *Tan's result,* (c) *Fattal's result,* (d) *He et al.'s result,* (e) *the proposed result,* (f) *the depth map of the proposed method.*
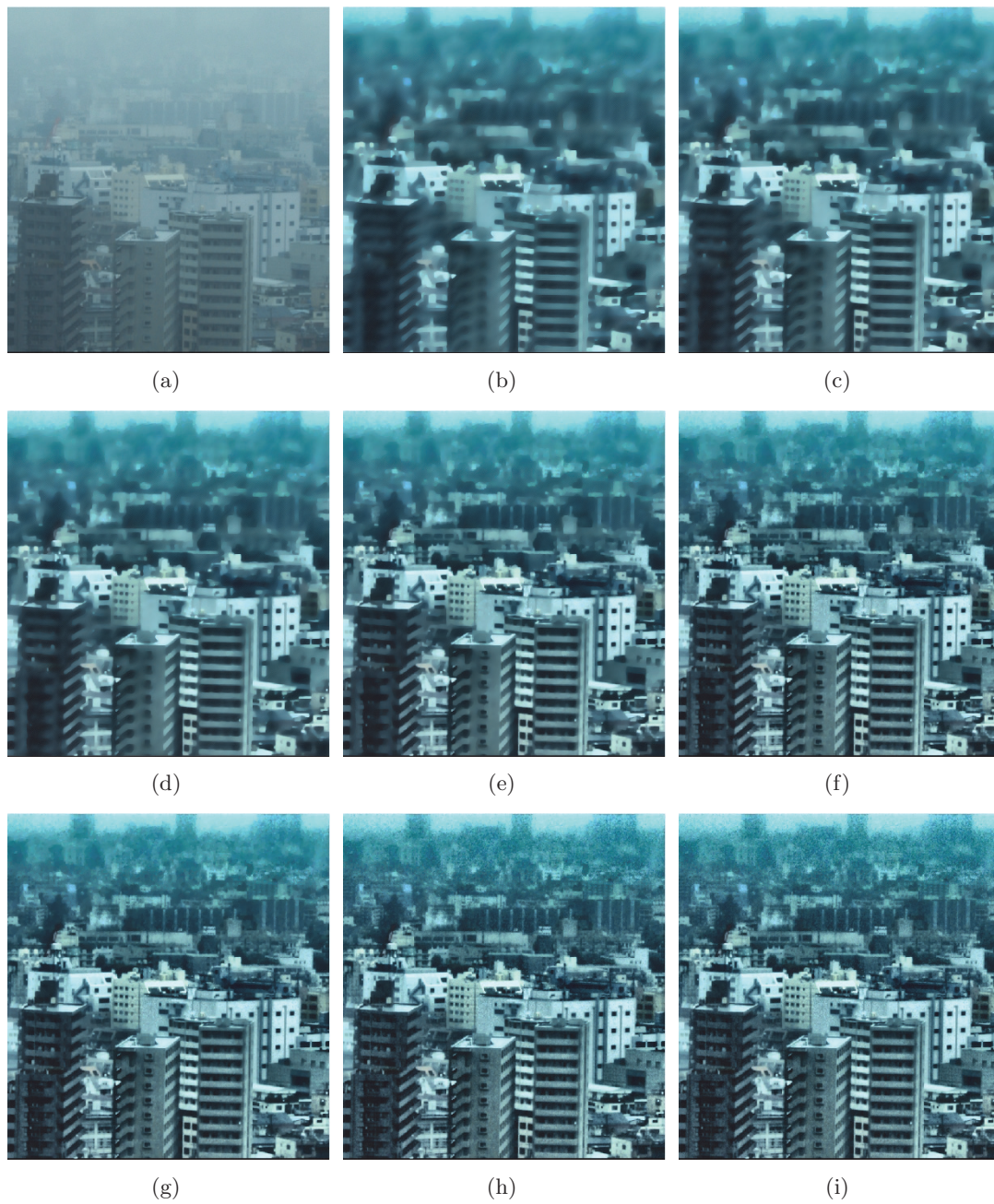
Figure 11. *The results with different k:* (a) *The original input hazy image (400 × 429); the result with* (b) $k = 0$, (c) $k = 0.001$, (d) $k = 1$, (e) $k = 10$, (f) $k = 50$, (g) $k = 100$, (h) $k = 500$, (i) $k = 1000$.
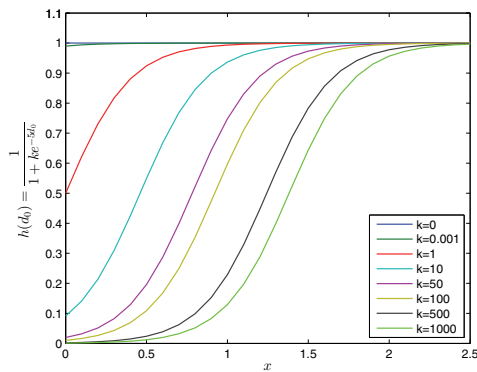
Figure 12. *The unidimensional examples of function h with different values of parameter k.*
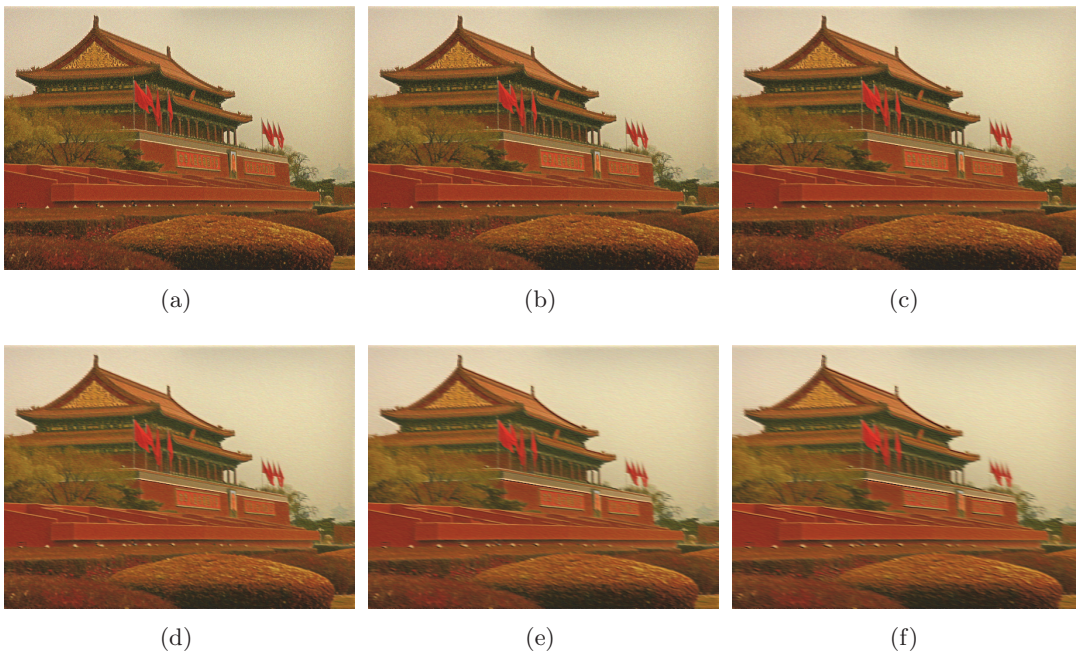

(a)   (b)   (c)

(d)   (e)   (f)

Figure 13. *The results with fixed $\gamma = 1$ and λ set to be (a) 10^{-3}, (b) 10^{-2}, (c) 10^{-1}; the results with fixed $\lambda = 1$ and γ set to be (d) 1, (e) 10^{-1}, (f) 10^{-2}.*

REFERENCES

[1] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, London, 2000.

[2] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526.

[3] X. BRESSON AND T. CHAN, *Active contours based on Chambolle's mean curvature motion*, in Proceedings of IEEE ICIP 2007, IEEE Press, Piscataway, NJ, 2007, pp. I33–I36.

[4] X. BRESSON AND T. CHAN, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Probl. Imaging, 2 (2008), pp. 455–484.

[5] X. BRESSON, S. ESEDOGLU, P. VANDERGHEYNST, J. THIRAN, AND S. OSHER, *Fast global minimization of the active contour/snake model*, J. Math. Imaging Vision, 28 (2007), pp. 151–167.

[6] A. BUADES, B. COLL, AND J. M. MOREL, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530.

[7] J. CAI, B. DONG, S. OSHER, AND Z. SHEN, *Image restoration: Total variation, wavelet frames, and beyond*, J. Amer. Math. Soc., 4 (2012), pp. 1033–1089.

[8] J. CAI, H. JI, C. LIU, AND Z. SHEN, *Framelet-based blind motion deblurring from a single image*, IEEE Trans. Image Process., 21 (2012), pp. 562–572.

[9] A. CHAI AND Z. SHEN, *Deconvolution: A wavelet frame approach*, Numer. Math., 106 (2007), pp. 529–587.

[10] A. CHAMBOLLE AND A. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[11] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[12] R. CHAN, Y. DONG, AND M. HINTERMÜLLER, *An efficient two-phase l^1-TV method for restoring blurred image with impulse noise*, IEEE Trans. Image Process., 19 (2010), pp. 1731–1739.

[13] R. CHAN, H. YANG, AND T. ZENG, *A two-stage image segmentation method for blurry images with Poisson or multiplicative gamma noise*, SIAM J. Imaging Sci., 7 (2014), pp. 98–127.

[14] S. CHAN, R. KHOSHABEH, K. GIBSON, P. GILL, AND T. NGUYEN, *An augmented Lagrangian method for total variation video restoration*, IEEE Trans. Image Process., 20 (2011), pp. 3097–3111.

[15] T. CHAN, S. ESEDOGLU, F. PARK, AND A. YIP, *Total variation image restoration: Overview and recent developments*, in Mathematical Models of Computer Vision, Springer-Verlag, Berlin, 2005, pp. 17–31.

[16] Q. CHANG, X. TAI, AND L. XING, *A compound algorithm of denoising using second-order and fourth-order partial differential equations*, Numer. Math. Theory Methods Appl., 2 (2009), pp. 353–376.

[17] Y. CHEN AND T. WUNDERLI, *Adaptive total variation for image restoration in BV space*, J. Math. Anal. Appl., 272 (2002), pp. 117–137.

[18] Y. DONG, R. CHAN, AND S. XU, *A detection statistic for random-valued impulse noise*, IEEE Trans. Image Process., 16 (2007), pp. 1112–1120.

[19] I. EKELAND AND R. TÉMAM, *Duality in convex optimization*, in Convex Analysis and Variational Problems, Classics in Appl. Math. 28, SIAM, Philadelphia, 1999, Chapter 3, pp. 46–72.

[20] R. FATTAL, *Single image dehazing*, in ACM SIGGRAPH08, New York, 2008, pp. 1–9.

[21] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[22] Y. GOUSSEAU AND J.-M. MOREL, *Are natural images of bounded variation?*, SIAM J. Math. Anal., 33 (2001), pp. 634–648.

[23] K. HE, J. SUN, AND X. TANG, *Single image haze removal using dark channel prior*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1956–1963.

[24] K. HE, J. SUN, AND X. TANG, *Single image haze removal using dark channel prior*, IEEE Trans. Pattern Anal. Mach. Intell., 33 (2011), pp. 2341–2353.

[25] J. KOPF, B. NEUBERT, B. CHEN, M. COHEN, D. COHEN-OR, O. DEUSSEN, M. UYTTENDAELE, AND D. LISCHINSKI, *Deepphoto: Model-based photograph enhancement and viewing*, ACM Trans. Graphics, 27 (2008), pp. 233–254.

[26] H. KOSCHMIEDER, *Theorie der horizontalen Sichtweite*, Beiträge zur Physik der freien Atmosphäre, Keim & Nemnich, Frankfurt, Germany, 1924.

[27] S. LEUNG AND S. OSHER, *Global minimization of the active contour model with TV-inpainting and two-phase denoising*, in Variational, Geometric, and Level Sets Methods in Computer Vision: Third International Workshop, Lecture Notes in Comput. Sci. 3752, Springer-Verlag, New York, Berlin, 2005, pp. 149–160.

[28] A. LEVIN, D. LISCHINSKI, AND Y. WEISS, *A closed form solution to natural image matting*, IEEE Trans. Pattern Anal. Mach. Intell., 30 (2008), pp. 228–242.

[29] C. LOUCHET AND L. MOISAN, *Total variation as a local filter*, SIAM J. Imaging Sci., 4 (2011), pp. 651–694.

[30] S. NARASIMHAN AND S. NAYAR, *Chromatic framework for vision in bad weather*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, Piscataway, NJ, 2000, pp. 598–605.

[31] S. NARASIMHAN AND S. NAYAR, *Vision and the atmosphere*, Int. J. Comput. Vision, 48 (2002), pp. 233–254.

[32] S. NARASIMHAN AND S. NAYAR, *Contrast restoration of weather degraded images*, IEEE Trans. Pattern Anal. Mach. Intell., 25 (2003), pp. 713–724.

[33] S. NARASIMHAN AND S. NAYAR, *Interactive deweathering of an image using physical models*, in Proceedings of the IEEE Workshop on Color and Photometric Methods in Computer Vision, held in conjunction with ICCV 2003, 2003; available online at http://www.ri.cmu.edu/publication_view.html?pub_id=5039.

[34] S. NAYAR AND S. NARASIMHAN, *Vision in bad weather*, in Proceedings of the International Conference on Computer Vision, Vol. 2, IEEE Computer Society Press, Piscataway, NJ, 1999, pp. 820–827.

[35] A. PREETHAM, P. SHIRLEY, AND B. SMITS, *A practical analytic model for daylight*, in Proceedings of the 26th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99), ACM, New York, 1999, pp. 91–100.

[36] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[37] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, in Proceedings of the Eleventh Annual International Conference of the Center for Nonlinear Studies on Experimental Mathematics: Computational Issues in Nonlinear Science, Los Alamos, NM, 1992, pp. 259–268.

[38] Y. SCHECHNER, S. NARASIMHAN, AND S. NAYAR, *Instant dehazing of images using polarization*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, IEEE Press, Piscataway, NJ, 2001, pp. I325–I332.

[39] S. SHWARTZ, E. NAMER, AND Y. SCHECHNER, *Blind haze separation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE Press, Piscataway, NJ, 2006, pp. 1984–1991.

[40] X. TAI AND C. WU, *Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model*, in Scale Space and Variational Methods in Computer Vision (SSVM2009), Lecture Notes in Comput. Sci. 5567, Springer, New York, 2009, pp. 502–513.

[41] R. TAN, *Visibility in bad weather from a single image*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, Piscataway, NJ, 2008, pp. 1–8.

[42] M. WERLBERGER, T. POCK, M. UNGER, AND H. BISCHOF, *A variational model for interactive shape prior segmentation and real-time tracking*, in Scale Space and Variational Methods in Computer Vision (SSVM2009), Lecture Notes in Comput. Sci. 5567, Springer, New York, 2009, pp. 200–211.

[43] A. C. YAU, X. TAI, AND M. K. NG, *Compression and denoising using l_0-norm*, Comput. Optim. Appl., 50 (2011), pp. 425–444.