Dense Scattering Layer Removal

Qiong Yan, Li Xu, Jiaya Jia
The Chinese University of Hong Kong
{qyan,xuli,leojia}@cse.cuhk.edu.hk

Abstract

We propose a new model, together with advanced optimization, to separate a thick scattering media layer from a single natural image. It is able to handle challenging underwater scenes and images taken in fog and sandstorm, both of which are with significantly reduced visibility. Our method addresses the critical issue – this is, originally unnoticeable impurities will be greatly magnified after removing the scattering media layer – with transmission-aware optimization. We introduce non-local structure-aware regularization to properly constrain transmission estimation without introducing the halo artifacts. A selective-neighbor criterion is presented to convert the unconventional constrained optimization problem to an unconstrained one where the latter can be efficiently solved.

1. Background

Dense scattering medium is one of the main causes of inconsistency in visual perception and human understanding [2]. Scattering layers often make the originally clear landmarks look distant, which explains why people think they move slower than normally when driving in fog and swimming underwater. The reduced visibility inevitably handicaps visual recognition and understanding. In contrast to its practical importance, previous approaches [16, 17, 3, 4, 18, 11] assume thin scattering layers caused, for example, by haze.

When tackling the more challenging dense-scattering-medium problem where visibility is significantly reduced, we notice inherent issues. On the one hand, the underlying structure is contaminated, making its restoration require spatially neighboring information. On the other hand, ubiquitous camera noise, image artifacts, and physically existing impurities (such as dust) in the media could be greatly amplified and influential in visual restoration. If they are not dealt with properly, erroneous estimation of structures shrouded by dense scattering media could be resulted in. It is noteworthy that intuitively applying denoising [1, 8] or performing regularized inversion [14] is not competent to
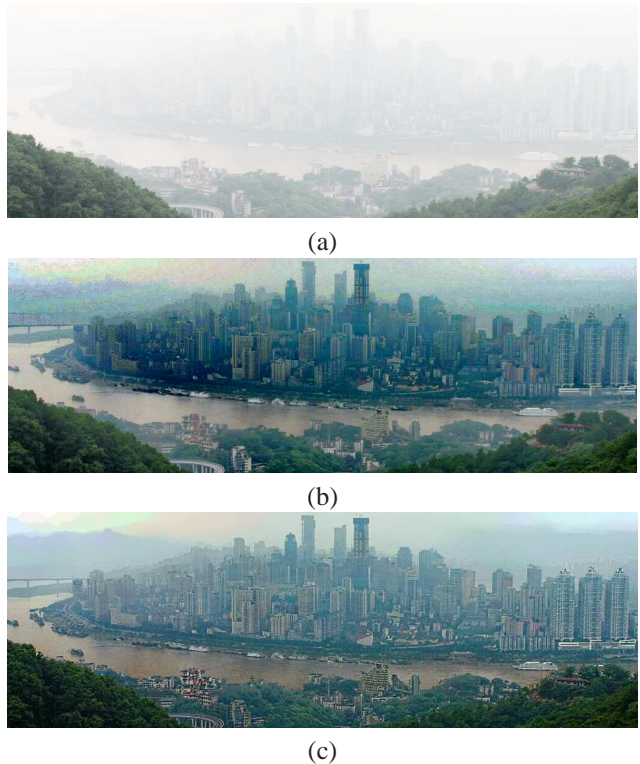

Figure 1. Noise in the input haze image (a) is remarkably amplified with state-of-the-art dehazing method [4], yielding visually unpleasing result (b). Our result shown in (c) contains clearer image structure.

solve this issue, due to the spatially varying properties.

We tackle the vital structure preserving and noise suppressing issues by proposing several novel strategies to properly enhance pictures shot in fog, haze, dust, and even underwater scenes. Both degraded structure recovery and significant noise suppression lead to the use of neighboring information, which motivates the use of non-local total variation strategy to regularize transmission and latent image estimation. It enables us to deal with noisy input, in the meantime preserving sharp discontinuities. The direct involvement of non-local terms, however, results in a complicated constrained optimization problem and is com-

putationally very expensive. We propose a novel *selective-neighbor* criterion to convert it to an unconstrained continuous optimization procedure. By incorporating transmission-aware noise-control terms into the regularized energy function, the proposed method becomes very effective in dense scattering layer removal.

1.1. Related Work

Central to visual restoration from scattering media is transmission estimation. On the hardware side, polarizers were used during picture taking, which help estimate part of the medium transitivity [16] or augments visibility for underwater vision [15]. 3D scene models were used in [6] to guide transmission estimation.

Single-image software solutions are also popular [10, 9, 17, 3, 4]. They are generally based on priors on transmission and scene radiance. Tan [17] developed a method mainly based on the observation that images with enhanced visibility have higher contrast and airlight depends on the distance to the viewer. Fattal [3] regarded transmission and surface shading (reflection) as locally uncorrelated in a hazed image. Independent Component Analysis (ICA) was employed to estimate scene albedo and medium transitivity. A dark-channel prior was proposed in [4] to initialize transmission estimation followed by refinement through soft matting.

These methods produce the results by simply inverting transmission blending with the underlying structures, which can generally magnify image noise and visual artifacts. One example is shown in Fig. 1, where image shown in (b), the result by direct inversion, becomes noisy after haze-removal. This is the major problem when dealing with dense media, where ubiquitous floating impurities can be notably intensified.

Regularized inversion was employed in [14]. Local regularization however by nature cannot handle strong image noise and compression artifacts. Taral et al. [18] used depth dependent median filtering. In dense media, the sizes are hard to determine accurately. The point-wise transmission constraint was recently introduced in [11], accompanying with an Expectation-Maximization (EM) solver. It does not tackle the noise boosting problem. Joshi et al. [5] explicitly pointed out the noise issue and developed a method to remove artifacts by averaging *multiple* images with weights. In comparison, our method is a robust single-image approach taking into account the noise suppression, transmission estimation, and computation efficiency. A unified framework is developed for enhancing pictures taken in challenging underwater environment or in meteorological phenomena.
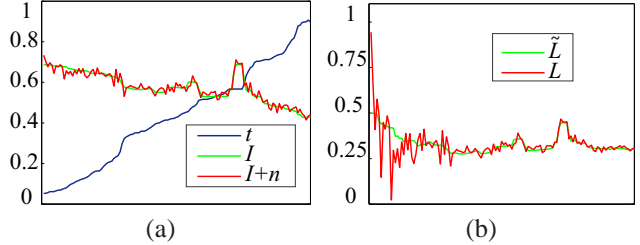

Figure 2. Noise magnification. 1D signals I and $I + n$ in (a) are with very small difference due to noise (x-axis: position; y-axis: value). However for the part with small transmission t (left side of (a)-(b)), the finally computed L in (b) is notably dissimilar to the ground truth $\tilde{L}$ owing to noise amplification.

1.2. The Problem

The model of surface radiance blended with the backscattered light can be simply expressed as

$$I(\mathrm{x}) = t(\mathrm{x})L(\mathrm{x}) + (1 - t(\mathrm{x}))B, \qquad (1)$$

where $L(\mathrm{x})$ denotes the surface radiance that we would have sensed without the scattering medium. x indexes the 2D coordinates. B is the backscattered light color vector determined by ambient illumination, also referred to as airlight or veiling light. $t(\mathrm{x})$ is the transmission component which relates to the scene depth $d(\mathrm{x})$ through $e^{-\eta d(\mathrm{x})}$, η is the attenuation coefficient, determined by scattering property of the medium.

Scattering layer removal requires an estimation of the transmission map $t(\mathrm{x})$, the light vector B, and then more importantly, the restoration the latent image $L(\mathrm{x})$. Based on the estimation of t and B, the latent image L can be recovered as

$$L(\mathrm{x}) = B - \frac{B - I(\mathrm{x})}{t(\mathrm{x})}. \qquad (2)$$

The simple inversion works well for general thin scattering media [16, 3, 4]. It however invokes problem for pixels with small transmission t, which happens when the object is distant or the medium is dense.

We now analyze how the estimation errors Δt would affect the result. In Eq. (2), Δt causes the error of L. We estimate its magnitude as

$$
\begin{aligned}
|\delta_L| &= \left| (B - \frac{B - I}{t + \Delta t}) - (B - \frac{B - I}{t}) \right| \\
&= \left| \frac{B - I}{t} \cdot \frac{\Delta t}{t + \Delta t} \right| \leq \left| \frac{\Delta t}{t + \Delta t} \right|. \quad (3)
\end{aligned}
$$

It indicates that the magnitude $|\delta_L|$ does not increase quickly with Δt. In fact, the larger Δt is, the slower it expands. 10% error in t would result in a similar amount of error in L.

In comparison, considering the inevitable camera noise, even though it is visually unnoticeable within the fog layer, its saliency could be notably raised after dehazing. Denoting the input image noise as n, the resulted noise magnitude $|\delta_n|$ between the finally estimated L using Eq. (2) and the ground truth $\tilde{L}$ without noise is calculated as

$$|\delta_n| = \left| (B - \frac{B - (I + n)}{t}) - (B - \frac{B - I}{t}) \right| = \left| \frac{n}{t} \right|. \tag{4}$$

For pixels with the corresponding t smaller than 0.1, which happens when the object is distant or the medium is dense, the resulted noise will be magnified by more than 10 times. It reveals the fact that effective and precise noise suppression is imperative in these cases and should be regarded as similarly important as transmission estimation. Methods of [3, 4] truncate t to prevent small values, which could leave part of the haze unremoved in dense media.

Fig. 2 illustrates the noise magnification problem for small t. In the plot (a), small perturbation in I results large deviation in L when transmission t is mall, as shown in (b).

2. Approach

Our method consists of two major steps to respectively update the transmission layer and the latent image. We automatically detect the brightest pixels, or allow users to draw scribbles containing a few sample pixels, to determine the backscattering light B.

2.1. Modeling Transmission t

By defining the logarithmic transmission, $D(\mathrm{x}) = \ln t(\mathrm{x}) = -\eta d(\mathrm{x})$, we alter our goal to compute D instead. t can be afterwards calculated using $t(\mathrm{x}) = e^{D(\mathrm{x})}$. The reason to define D is that depth d is exactly negative of D for each pixel. Most natural scene priors, such as the piece-wise spatial smoothness, can be imposed on depth (and correspondingly D), but not on the transmission variables that form different distributions.

Re-arranging terms in Eq. (1) and taking the logarithm yield

$$D(\mathrm{x}) = \ln(|B - I(\mathrm{x})|) - \ln(|B - L(\mathrm{x})|). \tag{5}$$

By further denoting $\bar{i}(\mathrm{x}) = \ln(|B - I(\mathrm{x})|)$, and $\bar{l}(\mathrm{x}) = \ln(|B - L(\mathrm{x})|)$, we express the log likelihood (without normalization) as

$$E_D(D) = \sum_{\mathrm{x}} \sum_{c} |D(\mathrm{x}) - (\bar{i}^c(\mathrm{x}) - \bar{l}^c(\mathrm{x}))|^2. \tag{6}$$

To model the smoothness property of the scene depth and at the same time preserve discontinuities, we resort to a non-local total variation regularizer. Our model considers neighboring pixels as well as non-neighboring ones in local windows, which is expressed as

$$E_S(D) = \sum_{\mathrm{x}} \sum_{\mathrm{y} \in W(\mathrm{x})} w_d(\mathrm{x}, \mathrm{y}) |D(\mathrm{x}) - D(\mathrm{y})|, \tag{7}$$

where $W(\mathrm{x})$ is a patch centered at x. $w_d(\mathrm{x}, \mathrm{y})$ plays the key role in shaping the local support. It is weak when $D(\mathrm{x})$ and $D(\mathrm{y})$ are found not necessarily similar, and is a large weight otherwise.

One may define w_d in Eq. (7) with respect to the structures in the input observation I. That is, if two pixels x and y have similar values in I, $w_d(\mathrm{x}, \mathrm{y})$ needs to be large to encourage strong spatial smoothness. However, we found that this strategy could adversely affect regularization. For instance, a pixel within a small texture pattern makes weights $w_d(\mathrm{x}, \mathrm{y})$ small for most of ys in $W(\mathrm{x})$, which correspondingly results in the lack of strong support in regularization. Worse, small texture-like structures in these cases generally do not correspond to actual depth discontinuity. So, instead of simply basing w_d on I, we propose a guided weight

$$w_d(\mathrm{x}, \mathrm{y}) = g(|S(\mathrm{x}) - S(\mathrm{y})|, \sigma_s), \tag{8}$$

where $g(x, \sigma) = \exp(-x^2/2\sigma^2)$. S is a structure map produced using the texture-structure decomposition method [19] to suppress the excessive details. One example is shown in Fig. 3. The structure map is useful to remove excessive details while still preserving large-scale structures.

Constrained Model There is a lower-bound for pixel-wise transmission, based on the non-negativity of scene luminance. It always holds that

$$t(\mathrm{x}) = \frac{B - I(\mathrm{x})}{B - L(\mathrm{x})} \geq 1 - \frac{I(\mathrm{x})}{B}. \tag{9}$$

The inequality is derived from $L(\mathrm{x}) \geq 0$. Therefore, the lower-bound for transmission given that all color components of B are not zero can be derived as

$$t(\mathrm{x}) \geq \max \left(1 - \min_{c \in \{r,g,b\}} \frac{I^c(\mathrm{x})}{B^c}, 0 \right). \tag{10}$$

This property can inhibit the adverse flattening effect in computing t, which accordingly prevents the halo artifacts in the latent image estimation.

Now, given the terms defined in Eqs. (6) and (7) and the constraint introduced in Eq. (10), the final objective function to estimate D (and correspondingly t) is written as

$$\begin{aligned} \min \quad & E_D + \lambda E_S \\ \text{s. t.} \quad & D(\mathrm{x}) \geq v(\mathrm{x}), \end{aligned} \tag{11}$$

| (a) Input | (b) Structure map S | (c) t estimated without S | (d) t estimated with S | (e) Result with t in (d) |

Figure 3. Structure layer extraction and its importance in transmission estimation compared to defining w_d on the input image I.
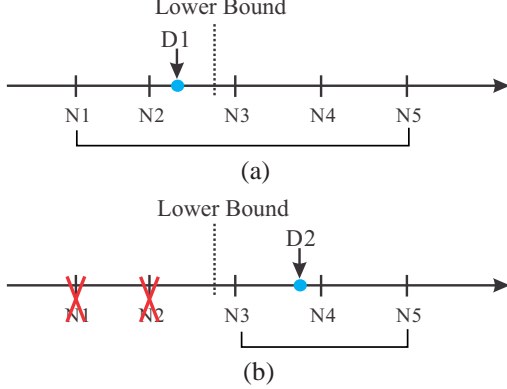

Figure 4. Illustration of neighbor selection. (a) Taking all neighbors N1-N5 in regularization yields result D1, violating the transmission lower-bound condition. (b) Our method explicitly removes N1 and N2 in regularization because they are smaller than the lower bound, and obtain the result D2, naturally satisfying the condition.

where

$$v(\mathrm{x}) = \ln\left(\max\left(1 - \min_c \frac{I^c(\mathrm{x})}{B^c}, \epsilon\right)\right). \quad (12)$$

ϵ is a small positive number to avoid $\ln 0$. v forms a map the same size as D, and encodes the lower-bound of D for all pixels. One example is shown in Fig. 5(b), which specifies the smallest possible value of D for each pixel.

Eq. (11) is a constrained non-linear optimization problem and is difficult to solve due to the non-local regularizer and the pixel-wise constraint. Directly applying the feasible descent direction algorithm results in very slow convergence and obtains only a local minimum even with a small-size image input.

Problem Conversion　We present a new method to convert the constrained optimization problem to an unconstrained one. The benefit is twofold. This conversion simplifies the objective function by embedding the inequality in the new form. Also, it enables transmission estimation by a very simple and efficient procedure. As a result, the originally difficult optimization can be nicely achieved in our method.

Our scheme is to iteratively update transmission for each pixel by computationally trackable relaxation. In each pass, we adaptively select suitable neighboring pixels y in Eq. (7) for regularization, based on their current transmission values, so that Eq. (10) can be explicitly satisfied. Intuitively, if one pixel y in Eq. (7) possibly pulls x out of the required range in Eq. (10), we discard it in regularization. In detail, we consider the following two cases.

1) When $\min_c I^c(\mathrm{x})/B^c < 1$, desirably, $D(\mathrm{x})$ needs to satisfy the condition $D(\mathrm{x}) \geq v(\mathrm{x})$, where $v(\mathrm{x})$ is the "lower bound" defined in Eq. (12) and illustrated in Fig. 4. In regularization, we deliberately set $w_d(\mathrm{x}, \mathrm{y}) = 0$ for all ys, whose current transmission estimates $D(\mathrm{y}) < v(\mathrm{x})$. So these ys will not be used to regularize x, as shown in Fig. 4(b). This naturally guarantees the lower-bound condition.

2) When $\min_c I^c(\mathrm{x})/B^c \geq 1$, the condition $D(\mathrm{x}) \geq \ln(\epsilon)$ can always be satisfied because the current estimates $D(\mathrm{y})$ for all ys also satisfy the lower-bound condition, i.e., $D(\mathrm{y}) \geq \ln(\epsilon)$. Hence, Eq. (7) can be applied directly with all ys.

In accordance to the analysis, we modify the weight definition in Eq. (7) slightly to

$$\tilde{w}_d(\mathrm{x}, \mathrm{y}) = \begin{cases} g(|S(\mathrm{x}) - S(\mathrm{y})|, \sigma_s), & \text{if } D(\mathrm{y}) \geq v(\mathrm{x}); \\ 0, & \text{otherwise.} \end{cases}$$
$$(13)$$

If $D(\mathrm{y})$ is smaller than the possible lower bound of x, to avoid dragging x out of the allowed range during estimation, we break the connection between x and y by setting $\tilde{w}_n(\mathrm{x}, \mathrm{y}) = 0$.

Normalization in each local window after removing the zeroed-out values yields

$$w_n(\mathrm{x}, \mathrm{y}) = \frac{\tilde{w}_d(\mathrm{x}, \mathrm{y})}{\sum_{\mathrm{y} \in W(\mathrm{x})} \tilde{w}_d(\mathrm{x},\mathrm{y})} \quad (14)$$

The modified regularization term is therefore

$$E_{\tilde{S}}(D) = \sum_{\mathrm{x}} \sum_{\mathrm{y} \in W(\mathrm{x})} \tilde{w}_d(\mathrm{x}, \mathrm{y})|D(\mathrm{x}) - D(\mathrm{y})|. \quad (15)$$

Now without loss of generality, by incorporating the *selective-neighbor* scheme in regularization, we prevent

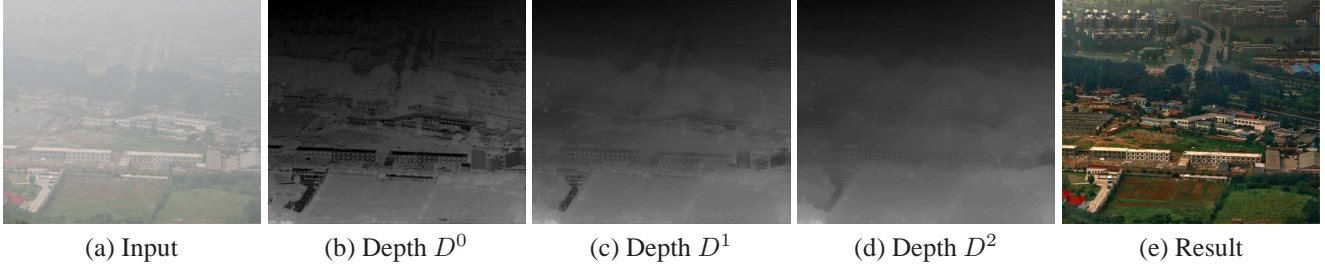| (a) Input | (b) Depth D^0 | (c) Depth D^1 | (d) Depth D^2 | (e) Result |

Figure 5. Depth maps (i.e., negative D maps) estimated in two iterations. Depth map D^0 is initialized with the lower bound measure v.

problematic estimation that violates the transmission lower-bound condition. The final objective function is written as

$$E(D) = E_D(D) + \lambda E_{\tilde{S}}(D), \qquad (16)$$

without explicit hard constraints.

Eq. (16) is not only simple in its representation, but also bears the advantage to find an extremely simple method for optimization based on computationally tractable relaxation to iteratively update D for pixels respectively.

More details about the numerical solver is included in Appendix A.1. Fig. 5 shows how the $-D$ map is improved in iterations.

2.2. Inferring Latent Image L

To compute L given the t estimate, we do not directly solve Eq. (2) since this scheme suffers from noise magnification. Instead, we apply optimization to infer a visually plausible L image.

For robustness, we define our data energy function as

$$E_d(L) = \sum_{\mathrm{x}} t(\mathrm{x})^2 |L(\mathrm{x}) - L_0(\mathrm{x})|^2, . \qquad (17)$$

L_0 is the result intuitively computed using Eq. (2). The data energy term in Eq. (17) suggests that the optimized latent image should be similar to L_0 weighted by $t(\mathrm{x})^2$. When t is large – that is, the object is not distant – we should trust L_0 because noise is not magnified too much according to our analysis.

We also provide a transmission-aware regularization term, which employs smoothness priors to further suppress noise. It is expressed as another non-local total variation:

$$E_s(L) = \sum_{\mathrm{x}} \sum_{\mathrm{y} \in W(\mathrm{x})} \bar{m}(\mathrm{x}, \mathrm{y})|L(\mathrm{x}) - L(\mathrm{y})|. \qquad (18)$$

The weight map $\bar{m}$ is normalized from m that contains two respective constraints to suppress noise. m is defined as

$$m(\mathrm{x}, \mathrm{y}) = g(|t(\mathrm{x}) - t(\mathrm{y})|, \sigma_t)$$
$$\cdot g(\|P(L_0, \mathrm{x}) - P(L_0, \mathrm{y})\|_2, \sigma_L), \qquad (19)$$

where $g(\cdot, \sigma)$ is a Gaussian tradeoff with standard deviation σ. The first term calculates the transmission similarity

between pixels, based on the fact that noise levels are magnified with respect to transmission. The second term actually measures the patch matching fidelity. $P(L_0, \mathrm{x})$ denotes a 7×7 window in L_0 centered at x. $\|P(L_0, \mathrm{x}) - P(L_0, \mathrm{y})\|_2$ uses a windowed L2-norm error measure to robustly estimate the color difference between pixels. Combining the two weight terms, if two pixels are in the same depth layer and have akin neighbors, their similarity is high. This patch-based error measure is much more robust than pixel-wise operations.

The final energy for estimating L is therefore given by

$$E(L) = E_d(L) + \lambda_L E_s(L). \qquad (20)$$

Our solver is similar to that for D estimation. Two or three iterations are enough to produce the results. See Appendix A.2 for more details.

3. Experimental Results

We convert the input image I to a linear color space before applying our method and perform the inverse gamma correction to coarsely curtail the effect of nonlinear color transform from the camera.

3.1. Quantitative evaluation

We quantitatively evaluate our method using images with ground truth depth [13]. We collect 10 images with various color and texture patterns and generate fog contaminated inputs using Eq. (1). Peak Signal to Noise Ratios (PSNRs) are calculated based on our restored images and the ground truth ones. Gaussian noise with $\sigma = 10$ is added to the inputs to simulate noise and dust in the air. We test the performance of our method under different fog thickness settings by varying η in $t = e^{-\eta d(\mathrm{x})}$. Sample images and the statistics are shown in Fig. 6, where (a) is an input with $\eta = 1$; (b) and (c) are our restored transmission and image; (d) shows the average PSNR under different densities of the scattering medium. We compare our method with the one using dark channel prior [4]. To better evaluate the importance of non-local regularizer, we also compare our method with regularized haze removal [14] by replacing non-local regularizer with a local one in the
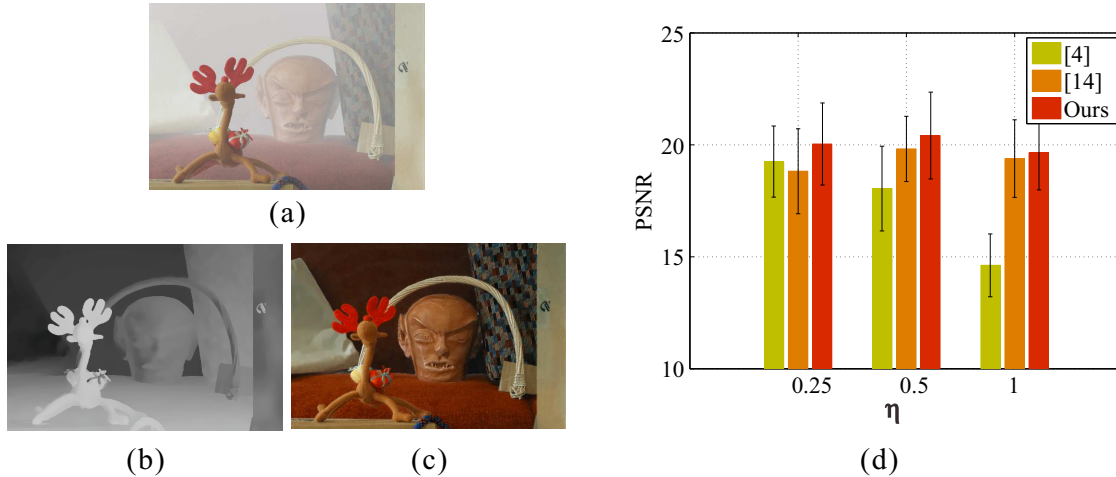
(a)

(b)

(c)

(d)

Figure 6. Quantitative evaluation. (a) is the foggy input. (b) and (c) show our restored transmission and latent images. (d) shows the statistics for different approaches for three levels of fog thickness.
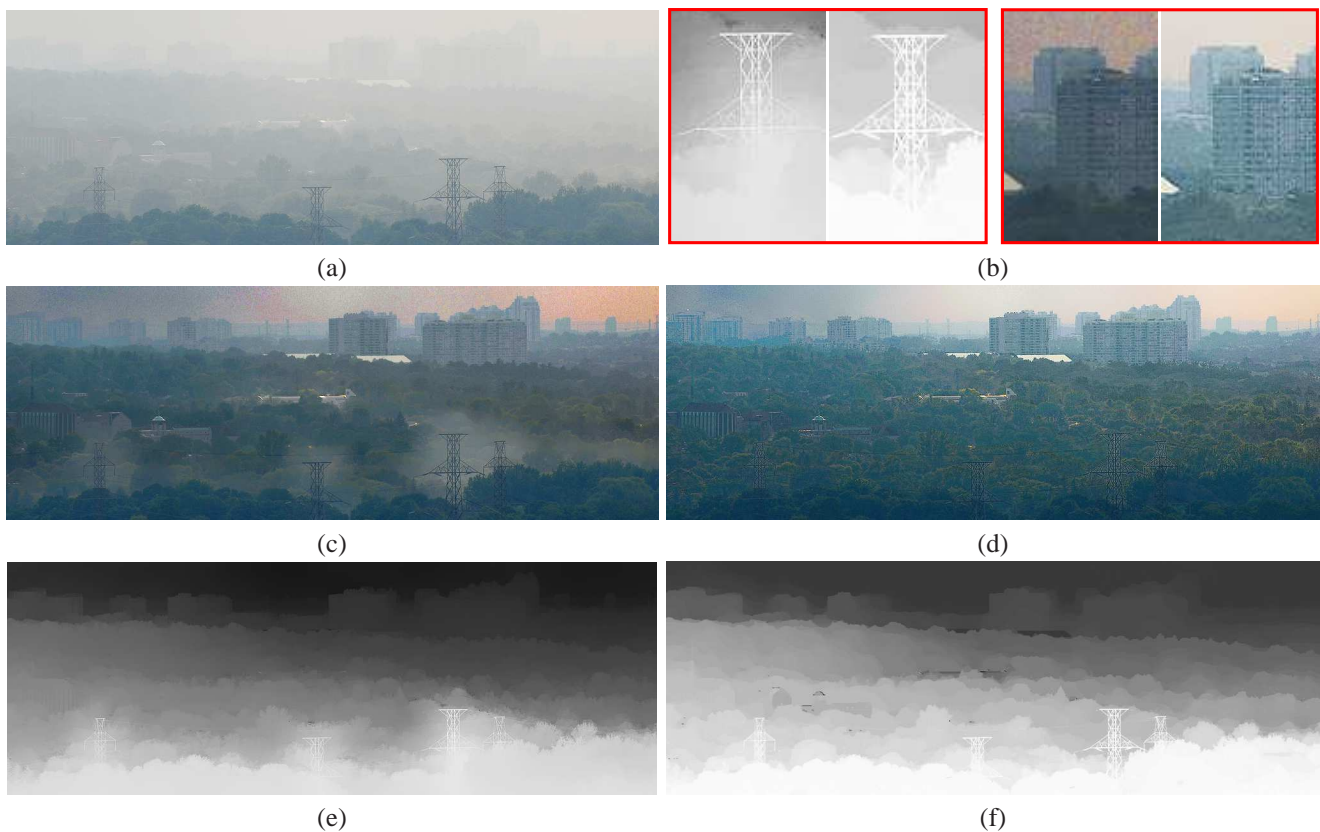

(a)

(b)

(c)

(d)

(e)

(f)

Figure 7. Fog image. (a) Input image. (c) and (e) Image L and map t of [4], which does not consider dense scattering media. (d) and (f) Our results. (b) Close-ups. Our results are on the right.

image restoration step. It can be seen that our method has a clear superiority in preserving structures with the increases of fog thickness, validating the effectiveness of our model in handling dense scattering layers. In what follows, we show more results on natural images.

3.2. Comparison on transmission map

Our non-local total variation smoothness term, working in concert with our point-wise data fidelity, is able to preserve thin structures comparing to the patch-based prior define in [4]. One example is shown in Fig. 7. (e) is

(a) Input

(b) Result of [4]

(c) Denoise (BM3D with $\sigma = 5$) before dehazing [4]

(d) Result by denoising (b) (BM3D with $\sigma = 15$)

(e) Result of [14]

(f) Our result

(g)   (h)   (i)   (j)   (k)   (l)

Figure 8. A dehazing example. Given input (a), our method produces the result (f). It not only removes intensive noise, also retains a great amount of details as well. Close-ups are shown in the last row.

the transmission map of [4], where inaccurate structure boundaries exist. They induce halos to the result shown in (c). Our method, shown in (f), preserves the structural edges. The corresponding restored image is in (d). Close-ups are shown in (b). We note that fine structures are very common in natural scenes. Their restoration is thus vital in scattering media removal.

3.3. Comparison on noise reduction

We exhibit our advantageous ability in handling significant noise. We compare our result with a previous regularized restoration approach [14], as well as strategies applying denoising before and after layer removal. One example is shown in Fig. 8. In the result of [4] (b) unnoticeable

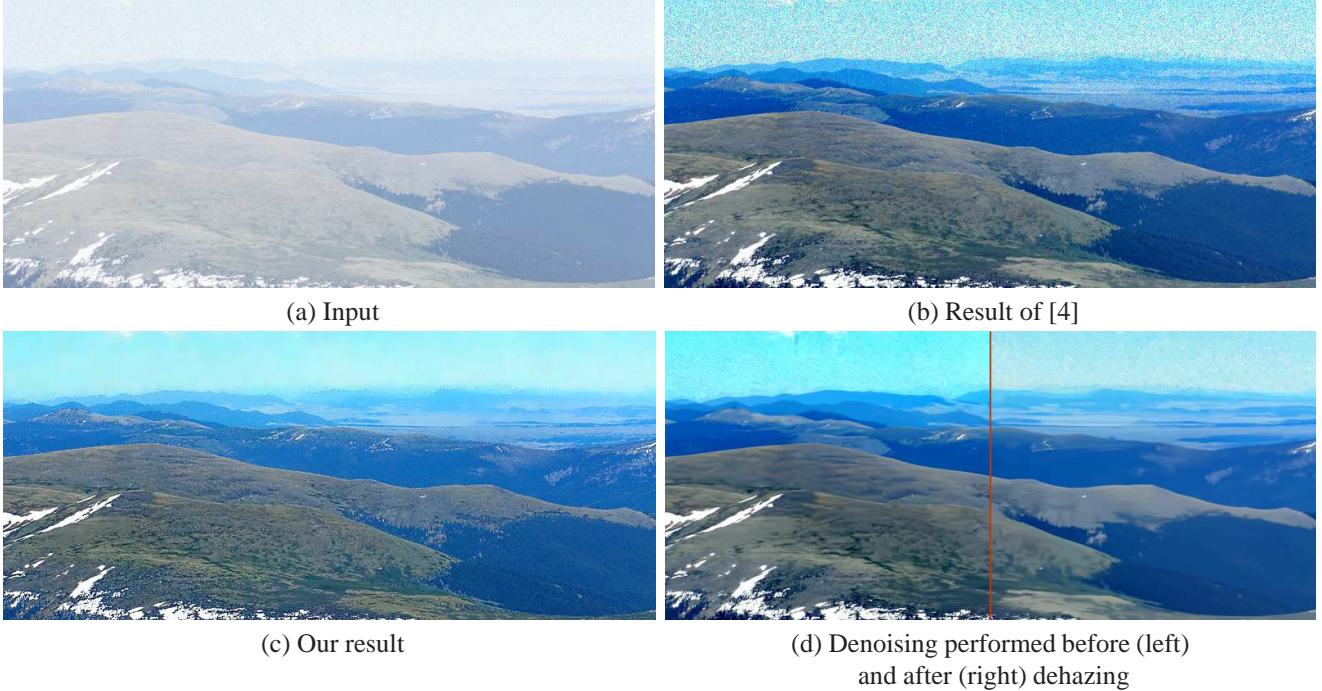|	
(a) Input	(b) Result of [4]
(c) Our result	(d) Denoising performed before (left) and after (right) dehazing

Figure 9. Heavy fog lifting. For the challenging input image (a), the state-of-the-art method produces the result shown in (b). Intuitively applying denoising before and after dehazing also cannot produce reasonable results, as shown in (d). Our result in (c) contains most structural details with noise being suppressed.

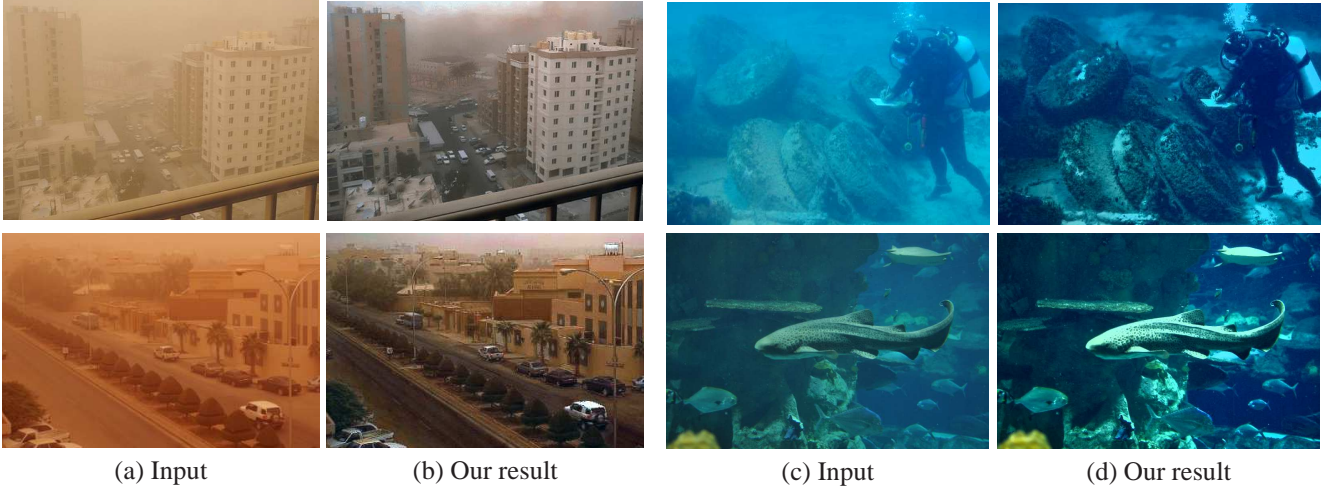|			
(a) Input	(b) Our result	(c) Input	(d) Our result

Figure 10. Underwater and dust storm picture restoration.

noise in the original image is greatly enhanced. The result shown in (c), which performs denoising before dehazing, still contains much noise even with spatial regularization. Denoising after dehazing, on the other hand, can hardly remove intensive noise out of the latent image structure. As shown in (d), even state-of-the-art BM3D denoising method destroys many latent image details, while considerable noise in the sky is left over. (e) is the result of [14] with a local TV regularizer. (f) is our final result. Noise is suppressed while underlaying structures are well preserved. Close-ups in the last row are from (a)-(f) respectively. Another example on noise reduction is shown in Fig. 9, demonstrating the effectiveness of our transmission-aware regularization.

3.4. General scattering medium removal

Our restoration method makes no assumption on the scattering media, making it applicable to fog, dust storm,

and underwater pictures. The example in Fig. 10(a) consists of a dense sand layer and is restored by our method in (b). For underwater environment, however, applying our method to all image channels results inaccurate color, since the red channel is generally much weaker than the other two due to the absorbtion by water. We choose to apply our method to the intensity channel only. An example is shown in Fig. 10(c)-(d). A large amount of structural information is recovered and contrast is greatly enhanced.

4. Concluding Remarks

We have presented a new model for scattering media layer removal from a single image. We introduced the transmission lower-bound condition and provided a very effective optimization framework incorporating several novel terms to solve the challenging noise amplification and depth estimation problems. Our method applied to images taken in fog, sandstorm, and underwater scenes.

A. Numerical Solver

In this section we describe our the numerical solvers (referring to Sections 2.1 and 2.2 in the paper) to compute the depth D and the latent clear image L respectively.

A.1. Depth D Solver

Depth D is computed by minimizing Eq. (16) in the paper, where $E(D)$ is written as

$$E(D) = \sum_{\mathbf{x}} \sum_{c} |D(\mathbf{x}) - (\bar{i}^c(\mathbf{x}) - \bar{l}^c(\mathbf{x}))|^2 +$$
$$\lambda \sum_{\mathbf{x}} \sum_{\mathbf{y} \in W(\mathbf{x})} \tilde{w}_d(\mathbf{x}, \mathbf{y}) |D(\mathbf{x}) - D(\mathbf{y})|, \quad (21)$$

where $\bar{i}^c(\mathbf{x})$ and $\bar{l}^c(\mathbf{x})$ are the observed and latent images in the logarithm domain. $\tilde{w}_d(\mathbf{x}, \mathbf{y})$ is a structure-aware penalty function defined in the paper. Directly minimizing Eq. (21) is difficult as the non-local smoothness term [12] involves a number of neighboring pixels, making gradient descent work poorly. We instead employ an iterative relaxation scheme to compute D.

Initially, we set $D^{(0)} = v$ as the start point. In the τ-th iteration, we minimize the energy function

$$E(D^{(\tau)}) = \sum_{\mathbf{x}} \sum_{c} |D^{(\tau)}(\mathbf{x}) - (\bar{i}^c(\mathbf{x}) - \bar{l}^c(\mathbf{x}))|^2 +$$
$$\lambda \sum_{\mathbf{x}} \sum_{\mathbf{y} \in W(\mathbf{x})} \tilde{w}_d(\mathbf{x}, \mathbf{y}) |D^{(\tau)}(\mathbf{x}) - D^{(\tau-1)}(\mathbf{y})|,$$
$$(22)$$

where the superscript (τ) indexes iterations. In solving $D^{(\tau)}(\mathbf{x})$ for each $\mathbf{x}$, relaxation keeps other values fixed, which results in an efficient estimation process. By further

---

Algorithm 1 Relaxation Procedure to Compute D.

1: **input:** initial depth $D^{(0)} = v$; the maximum number of iterations n
2: $\tau = 1$.
3: Apply median filter to $D^{(\tau-1)}$ according to Eq. (23), and get $D^{(\tau)}$.
4: $\tau = \tau + 1$. If $\tau \leq n$, go to Step 2.
5: **output:** D

---

denoting by $\{D_h^{(\tau-1)}\}_{h=0}^{W-1}$ the sorted $D^{(\tau-1)}$ within the window $W(\mathbf{x})$ in an ascending order and by w_h the correspondingly sorted weight $\tilde{w}_d(\mathbf{x}, \mathbf{y})$, $D^{(\tau)}(\mathbf{x})$ can be written as

$$E(D^{(\tau)}(\mathbf{x})) = \sum_{c} |D^{(\tau)}(\mathbf{x}) - (\bar{i}^c(\mathbf{x}) - \bar{l}^c(\mathbf{x}))|^2 +$$
$$\lambda \sum_{h=0}^{W-1} w_h |D^{(\tau)}(\mathbf{x}) - D_h^{(\tau-1)}|. \quad (23)$$

Eq. (23) has two essential features that enable efficient computation. One is that the data term is strictly convex and differentiable, with its derivative $2 \cdot \sum_c (D^{(\tau)}(\mathbf{x}) - (\bar{i}^c(\mathbf{x}) - \bar{l}^c(\mathbf{x})))$ a bijective function. The other property is that weight w_h is non-negative for all hs by definition. With these properties and also according to the proof in [7], Eq. (23) can be efficiently optimized by computing the median value in

$$M = \left\{ D_0^{(\tau-1)}, \ldots, D_{W-1}^{(\tau-1)}, r_0, \ldots, r_W \right\}, \quad (24)$$

where

$$r_h = \frac{\sum_c (\bar{i}^c(\mathbf{x}) - \bar{l}^c(\mathbf{x}))}{3} + \frac{\lambda}{6} \left(-\sum_{j=0}^{h-1} w_j + \sum_{j=h}^{W-1} w_j \right),$$
$$(25)$$

for $h = 0, \ldots, W$.

Based on this relaxation scheme, we present the overview of our algorithm in Algorithm 1. Typically, λ is set to 15. Two or three passes, i.e., $n \leq 3$, to update all pixels, are enough to yield good results in our experiments.

A.2. Optimizing Latent Image L

The energy function for solving L (in Section 2.2 in the paper) is similar to Eq. (21). We decompose it into channels $c \in \{r, g, b\}$, and write it as

$$E(L^c) = \sum_{\mathbf{x}} t(\mathbf{x})^2 |L^c(\mathbf{x}) - L_0^c(\mathbf{x})|^2 +$$
$$\lambda_L \sum_{\mathbf{x}} \sum_{\mathbf{y} \in W(\mathbf{x})} \bar{m}(\mathbf{x}, \mathbf{y}) |L^c(\mathbf{x}) - L^c(\mathbf{y})|. \quad (26)$$

Algorithm 2 Relaxation Procedure to Compute L.

1: **input:** initial image $L^{(0)} = L_0$; the maximum number of iteration n
2: $\tau = 1$.
3: Apply median filter to $L^{(\tau-1)}$ according to Eq. (27), which yields $L^{(\tau)}$.
4: $\tau = \tau + 1$. If $\tau \leq n$, go to Step 2.
5: **output:** L

By applying similar relaxation, we get the following pixel-wise energy minimization problem

$$E(L^{c(\tau)}(\mathbf{x})) = t(\mathbf{x})^2|L^{c(\tau)}(\mathbf{x}) - L_0^c(\mathbf{x})|^2 +$$
$$\lambda_L \sum_{h=0}^{W-1} m_h|L^{c(\tau)}(\mathbf{x}) - L_h^{c\,(\tau-1)}|. \quad (27)$$

It can also be solved by finding the median value in

$$\left\{ L_0^{c(\tau-1)}, \ldots, L_{W-1}^{c}{}^{(\tau-1)}, z_0, \ldots, z_W \right\}, \quad (28)$$

where

$$z_h = L_0^c(\mathbf{x}) + \frac{\lambda_L}{2t(\mathbf{x})^2} \left(-\sum_{j=0}^{h-1} m_j + \sum_{j=h}^{W-1} m_j \right), \quad (29)$$

for $h = 0, \ldots, W$.

Our algorithm is shown in Algorithm 2. λ_L usually varies from 0.001 to 0.02 to prevent over-smoothing. Iteration number $n = 1 \sim 3$ is enough.

References

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image restoration by sparse 3d transform-domain collaborative filtering. In *SPIE Electronic Imaging*, volume 6812, 2008.

[2] G. Day and P. Schoemaker. Driving through the fog: managing at the edge. *Long Range Planning*, 37(2):127–142, 2004.

[3] R. Fattal. Single image dehazing. *ACM Trans. Graph.*, 27(3), 2008.

[4] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *Computer Vision and Pattern Recognition*, pages 1956–1963, 2009.

[5] N. Joshi and M. F. Cohen. Seeing mt. rainier: Lucky imaging for multi-image denoising. In *International Conference on Computational Photography*, 2010.

[6] J. Kopf, B. Neubert, B. Chen, M. F. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski. Deep photo: model-based photograph enhancement and viewing. *ACM Trans. Graph.*, 27(5):116, 2008.

[7] Y. Li and S. Osher. A new median formula with applications to pde based denoising. *Communications in Mathematical Sciences*, 7(3):741–753, 2009.

[8] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *International Conference on Computer Vision*, pages 2272–2279, 2009.

[9] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(6):713–724, 2003.

[10] S. K. Nayar and S. G. Narasimhan. Vision in bad weather. In *International Conference on Computer Vision*, pages 820–827, 1999.

[11] K. Nishino, L. Kratz, and S. Lombardi. Bayesian defogging. *International Journal of Computer Vision*, pages 1–16, 2011.

[12] G. Peyré, S. Bougleux, and L. D. Cohen. Non-local regularization of inverse problems. In *ECCV (3)*, pages 57–68, 2008.

[13] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *Computer Vision and Pattern Recognition*, 2007.

[14] Y. Y. Schechner and Y. Averbuch. Regularized image recovery in scattering media. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(9):1655–1660, 2007.

[15] Y. Y. Schechner and N. Karpel. Clear underwater vision. In *Computer Vision and Pattern Recognition*, pages 536–543, 2004.

[16] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Polarization-based vision through haze. *Applied Optics*, 42(3), 2003.

[17] R. T. Tan. Visibility in bad weather from a single image. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[18] J.-P. Tarel and N. Hautière. Fast visibility restoration from a single color or gray level image. In *International Conference on Computer Vision*, pages 2201–2208, 2009.

[19] L. Xu, Q. Yan, Y. Xia, and J. Jia. Structure extraction from texture via natural variation measure. *ACM Trans. Graph.*, 31(6), 2012.