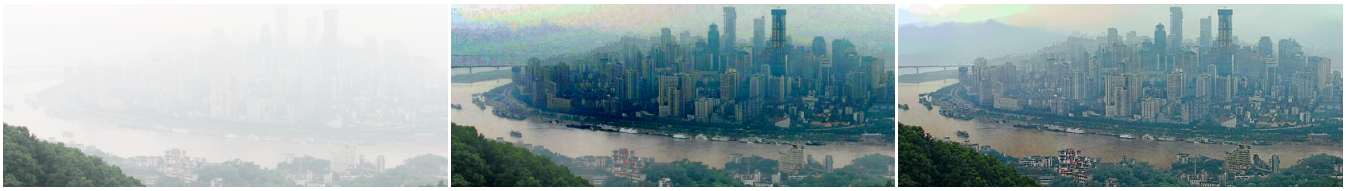
Dense Scattering Layer Removal

Qiong Yan Li Xu Jiaya Jia*

Department of Computer Science and Engineering
The Chinese University of Hong Kong



(a) (b) (c)

Figure 1: *Noise in the input image (a) is remarkably amplified with state-of-the-art dehazing methods [He et al. 2009], yielding result (b). Our result shown in (c) contains more image structures.*

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.

Keywords: haze removal, noise reduction

1 Background

Dense scattering medium is one of the main causes of inconsistency in visual perception and human understanding [Day and Schoemaker 2004]. 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 [Schechner et al. 2003; Tan 2008; Fattal 2008; He et al. 2009; Tarel and Hautière 2009] assume thin scattering layers caused, for example, by haze.

When tackling the 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, 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 [Dabov et al. 2008] or performing regularized inversion [Schechner and Averbuch 2007] is not competent to 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 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 structures. The direct involvement of non-local terms, however, results in a complicated constrained optimization problem. 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 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 to estimate part of the medium transitivity [Schechner et al. 2003] or augments visibility for underwater vision [Schechner and Karpel 2004]. 3D scene models were used in [Kopf et al. 2008] to guide transmission estimation.

Single-image software solutions are also popular [Nayar and Narasimhan 1999; Narasimhan and Nayar 2003; Tan 2008; Fattal 2008; He et al. 2009]. They are generally based on priors on transmission and scene radiance. Tan [2008] 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 [2008] 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 [He et al. 2009] to initialize transmission estimation followed by refinement through soft matting.

These methods simply invert transmission blending with the underlying structures, which can generally magnify image noise and visual artifacts. One example is shown in Fig. 1. The result (b) by direct inversion, becomes noisy after haze-removal. This is the

*e-mail: {qyan, xuli, leojia}@cse.cuhk.edu.hk

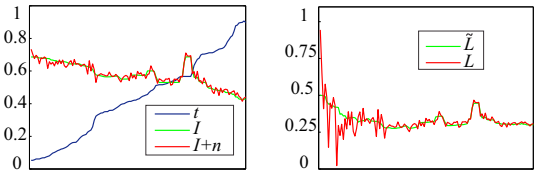


Figure 2: *Noise magnification. 1D signals I and $I+n$ in (a) are with very small difference due to noise (x-axis: position; y-axis: value). However for the part with small transmission t (left side of (a)-(b)), the finally computed L in (b) is dissimilar to the ground truth $\tilde{L}$ owing to noise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

major problem when dealing with dense media, where ubiquitous floating impurities can be notably intensified.

Regularized inversion employed in [Schechner and Averbuch 2007] however cannot handle strong image noise and compression artifacts. Taral et al. [2009] used depth dependent median filtering. In dense media, the sizes are hard to determine accurately. Joshi et al. [2010] 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 noise suppression, transmission estimation, and computation efficiency. A unified framework is developed for enhancing pictures taken in challenging underwater environment or in meteorological phenomena.

1.2 The Problem

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

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

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. $t(\mathrm{x})$ 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 the 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 of the latent image $L(\mathrm{x})$. Based on the estimation of t and B, the latent image L can be recovered as

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

The simple inversion works well for general thin scattering media [Schechner et al. 2003; Fattal 2008; He et al. 2009]. It however invokes problem for pixels with small transmission t, which happens when the object is distant or the medium is dense. Fig. 2 illustrates the noise magnification problem for small t. Considering the inevitable camera noise, even though it is unnoticeable within the fog layer, its saliency could be notably raised after fog removal.

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 $t(\mathrm{x}) = e^{D(\mathrm{x})}$. As 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, but not on the transmission variables that form different distributions.

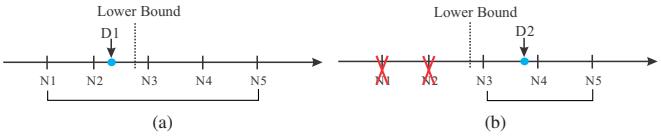


Figure 3: *Illustration of neighbor selection. (a) Taking all neighbors $N1$-$N5$ in regularization yields result $D1$, violating the transmission lower-bound condition. (b) Our method explicitly removes $N1$ and $N2$ in regularization as they are smaller than the lower bound, and obtain the result $D2$, naturally satisfying the condition.*

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

$$D(\mathrm{x}) = \ln(|B - I(\mathrm{x})|) - \ln(|B - L(\mathrm{x})|). \tag{3}$$

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

$$E_D(D) = \sum_{\mathrm{x}} \sum_{c} |D(\mathrm{x}) - (\bar{i}^c(\mathrm{x}) - \bar{l}^c(\mathrm{x}))|^2. \tag{4}$$

To model the smoothness property of the scene depth and preserve discontinuities, we resort to a non-local total variation regularizer, which is expressed as

$$E_S(D) = \sum_{\mathrm{x}} \sum_{\mathrm{y} \in W(\mathrm{x})} w_d(\mathrm{x},\mathrm{y}) |D(\mathrm{x}) - D(\mathrm{y})|, \tag{5}$$

where $W(\mathrm{x})$ is a patch centered at x. Weight $w_d(\mathrm{x},\mathrm{y})$ requires $D(\mathrm{x})$ and $D(\mathrm{y})$ to be close when pixels x and y are similar in appearance. To handle rich details in natural images, we resort to a structure map S [Xu et al. 2012] to define $w_d(\mathrm{x},\mathrm{y}) = \exp(-|S(\mathrm{x}) - S(\mathrm{y})|^2/(2\sigma_s^2))$. The structure map S is useful to remove excessive details while still preserving large-scale structures.

Constrained Model There is a natural lower-bound for pixel-wise transmission, based on the non-negativity of scene luminance. Given $L(\mathrm{x}) \geq 0, B > 0$, it always holds that

$$t(\mathrm{x}) = \frac{B - I(\mathrm{x})}{B - L(\mathrm{x})} \geq 1 - \frac{I(\mathrm{x})}{B} = \max\left(1 - \min_{c \in \{r,g,b\}} \frac{I^c(\mathrm{x})}{B^c}, 0\right). \tag{6}$$

Now, given the terms defined in Eqs. (4) and (5) and the constraint introduced in Eq. (6), the final objective function to estimate D (and correspondingly t) is written as

$$\min \;\; E_D + \lambda E_S \;\; \text{s. t.} \;\; D(\mathrm{x}) \geq v(\mathrm{x}), \tag{7}$$

where $v(\mathrm{x}) = \ln\left(\max\left(1 - \min_c \frac{I^c(\mathrm{x})}{B^c}, \epsilon\right)\right)$. ϵ is a small positive number to avoid $\ln 0$. Eq. (7) is a constrained non-linear optimization problem and is difficult to solve due to the non-local regularizer and the pixel-wise constraint.

Optimization To solve this problem, 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. (5) for regularization, based on their current transmission values, so that the constraint $D(\mathrm{x}) > v(\mathrm{x})$ can be explicitly satisfied. Intuitively, if one pixel y in Eq. (5) possibly pulls x out of

the required range, we discard it in regularization, as demonstrated in Fig. 3. In detail, we modify the weight as

$$w_n(\mathrm{x},\mathrm{y}) = \begin{cases} g(|S(\mathrm{x}) - S(\mathrm{y})|, \sigma_s), & \text{if } D(\mathrm{y}) \geq v(\mathrm{x}); \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

The new object function with the modified weight is therefore an unconstrained one. According to [Li and Osher 2009], it can be efficiently solved by iteratively applying median filter. 3 iterations are generally enough. Initially, we set $l(\mathrm{x})$ to make $\bar{i}^c(\mathrm{x}) - \bar{l}^c(\mathrm{x}) = v(\mathrm{x})$. Fig. 4 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

$$E_d(L^c) = \sum_{\mathrm{x}} t(\mathrm{x})^2 |L^c(\mathrm{x}) - L_0^c(\mathrm{x})|^2, \quad (9)$$

where c indexes color channels. L_0^c is the result intuitively computed using Eq. (2). The data energy term in Eq. (9) suggests that the optimized latent image should be similar to L_0^c weighted by $t(\mathrm{x})^2$. When t is large – that is, the object is not distant – we should trust L_0^c 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 non-local total variation:

$$E_s(L) = \sum_{\mathrm{x}} \sum_{\mathrm{y} \in W(\mathrm{x})} \bar{m}(\mathrm{x},\mathrm{y}) |L(\mathrm{x}) - L(\mathrm{y})|. \quad (10)$$

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

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

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.

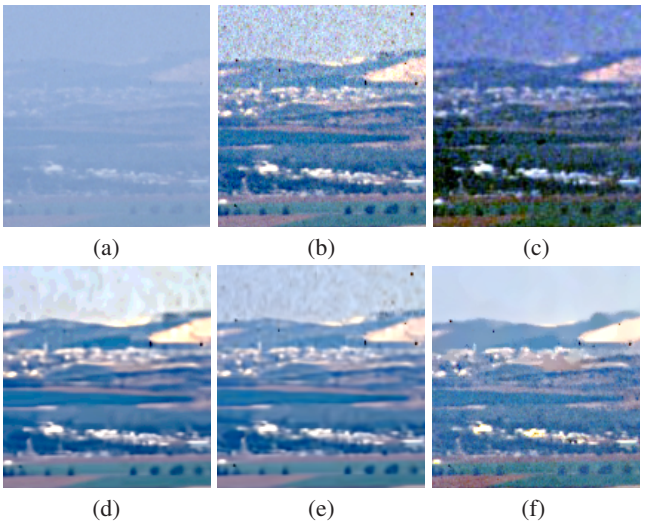
The final energy for estimating L is therefore given by

$$E(L) = E_d(L) + \lambda_L E_s(L). \quad (11)$$

Our solver is similar to that for D estimation. Two or three iterations are enough to produce the results.

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.



(a) (b) (c) (d) (e) (f)

Figure 6: *A dehazing example. (a) Input. (b) and (c) Result of [He et al. 2009] and [Schechner and Averbuch 2007]. (d) and (e) Results by applying denoise algorithm before and after dehazing. (f) Our Result. It not only removes intensive noise, also retains a great amount of details as well.*

(a) Input (b) Our result

Figure 7: *Underwater and dust storm picture restoration.*

3.1 Comparison on Transmission Map

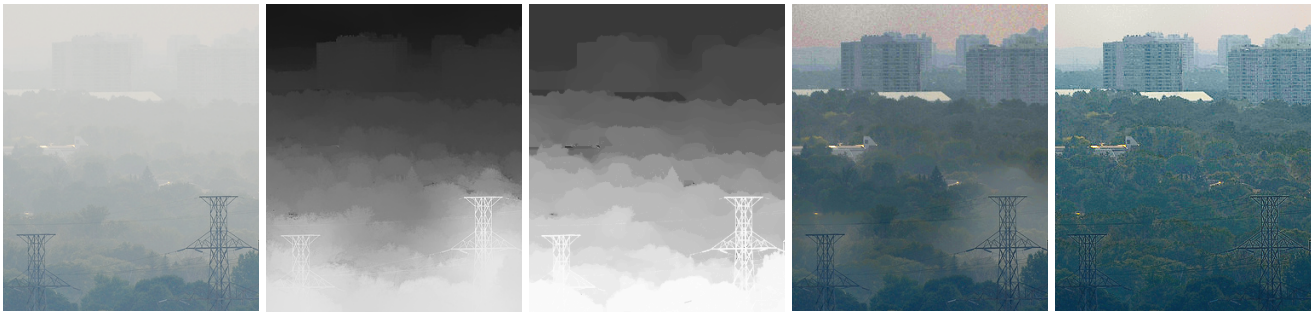
Our non-local total variation smoothness term, working in concert with our point-wise data fidelity, is able to preserve thin structures comparing to the patch-based prior define in [He et al. 2009]. One example is shown in Fig. 5. (b) is the transmission map of [He et al. 2009], where inaccurate structure boundaries exist. They induce halos to the result shown in (d). Our method, shown in (c), preserves the structural edges. The corresponding restored image is in (e). We note that fine structures are very common in natural scenes. Their restoration is thus vital in scattering media removal.

3.2 Comparison on Noise Reduction

We exhibit our advantageous ability in handling significant noise. We compare our result with a previous regularized restoration approach [Schechner and Averbuch 2007], as well as strategies applying denoising before and after layer removal. One example is shown in Fig. 6. In the result of [He et al. 2009] (b) unnoticeable noise in the original image is greatly enhanced. The result shown in

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

Figure 4: *Depth maps (i.e., negative D maps) estimated in two iterations. Depth map D^0 is initialized with the lower bound measure v.*



(a) (b) (c) (d) (e)

Figure 5: *Fog image. (a) Input image. (b) and (d) Transmission map t and corresponding image L according to Eq. 2 by [He et al. 2009], which does not consider dense scattering media. (c) and (e) Our results.*

(d), 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 (e), even state-of-the-art BM3D denoising method destroys many latent image details, while considerable noise in the sky is left over. (c) is the result of [Schechner and Averbuch 2007] with a local TV regularizer. (f) is our final result. Noise is suppressed while underlaying structures are preserved.

3.3 General Scattering Medium Removal

Our restoration method makes no assumption on the scattering media, making it applicable to fog, dust storm, and underwater pictures. Fig. 7 shows examples in underwater and dust scene, in which 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.

References

DABOV, K., FOI, A., KATKOVNIK, V., AND EGIAZARIAN, K. 2008. Image restoration by sparse 3d transform-domain collaborative filtering. In *SPIE Electronic Imaging*, vol. 6812.

DAY, G., AND SCHOEMAKER, P. 2004. Driving through the fog: managing at the edge. *Long Range Planning 37*, 2, 127–142.

FATTAL, R. 2008. Single image dehazing. *ACM Trans. Graph. 27*, 3.

HE, K., SUN, J., AND TANG, X. 2009. Single image haze removal using dark channel prior. In *Computer Vision and Pattern Recognition*, 1956–1963.

JOSHI, N., AND COHEN, M. F. 2010. Seeing mt. rainier: Lucky imaging for multi-image denoising. In *International Conference on Computational Photography*.

KOPF, J., NEUBERT, B., CHEN, B., COHEN, M. F., COHEN-OR, D., DEUSSEN, O., UYTTENDAELE, M., AND LISCHINSKI, D. 2008. Deep photo: model-based photograph enhancement and viewing. *ACM Trans. Graph. 27*, 5, 116.

LI, Y., AND OSHER, S. 2009. A new median formula with applications to pde based denoising. *Communications in Mathematical Sciences 7*, 3, 741–753.

NARASIMHAN, S. G., AND NAYAR, S. K. 2003. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell. 25*, 6, 713–724.

NAYAR, S. K., AND NARASIMHAN, S. G. 1999. Vision in bad weather. In *International Conference on Computer Vision*, 820–827.

SCHECHNER, Y. Y., AND AVERBUCH, Y. 2007. Regularized image recovery in scattering media. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 9, 1655–1660.

SCHECHNER, Y. Y., AND KARPEL, N. 2004. Clear underwater vision. In *Computer Vision and Pattern Recognition*, 536–543.

SCHECHNER, Y. Y., NARASIMHAN, S. G., AND NAYAR, S. K. 2003. Polarization-based vision through haze. *Applied Optics 42*, 3.

TAN, R. T. 2008. Visibility in bad weather from a single image. In *Computer Vision and Pattern Recognition*, 1–8.

TAREL, J.-P., AND HAUTIÈRE, N. 2009. Fast visibility restoration from a single color or gray level image. In *International Conference on Computer Vision*, 2201–2208.

XU, L., YAN, Q., XIA, Y., AND JIA, J. 2012. Structure extraction from texture via natural variation measure. *ACM Trans. Graph. 31*, 6.