

Robust Image and Video Dehazing with Visual Artifact Suppression via Gradient Residual Minimization

Chen Chen^{1(✉)}, Minh N. Do¹, and Jue Wang²

¹ University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
{cchen156,minhdo}@illinois.edu

² Adobe Research, Seattle, WA 98103, USA
juewang@adobe.com

Abstract. Most existing image dehazing methods tend to boost local image contrast for regions with heavy haze. Without special treatment, these methods may significantly amplify existing image artifacts such as noise, color aliasing and blocking, which are mostly invisible in the input images but are visually intruding in the results. This is especially the case for low quality cellphone shots or compressed video frames. The recent work of Li *et al.* (2014) addresses blocking artifacts for dehazing, but is insufficient to handle other artifacts. In this paper, we propose a new method for reliable suppression of different types of visual artifacts in image and video dehazing. Our method makes contributions in both the haze estimation step and the image recovery step. Firstly, an image-guided, depth-edge-aware smoothing algorithm is proposed to refine the initial atmosphere transmission map generated by local priors. In the image recovery process, we propose Gradient Residual Minimization (GRM) for jointly recovering the haze-free image while explicitly minimizing possible visual artifacts in it. Our evaluation suggests that the proposed method can generate results with much less visual artifacts than previous approaches for lower quality inputs such as compressed video clips.

Keywords: Video dehazing · Image dehazing · Contrast enhancement · Artifact suppression

1 Introduction

Due to atmospheric absorption and scattering, outdoor images and videos are often degraded to have low contrast and visibility. In addition to the deterioration of visual quality, heavy haze also makes many computer vision tasks more

Most work was done when Chen was an intern at Adobe.

Electronic supplementary material The online version of this chapter (doi:[10.1007/978-3-319-46475-6_36](https://doi.org/10.1007/978-3-319-46475-6_36)) contains supplementary material, which is available to authorized users.

difficult, such as stereo estimation, object tracking and detection *etc.* Therefore, removing haze from images and video becomes an important component in a post-processing pipeline. Conventional global contrast enhancement methods often do not perform well because the degradation is spatially-varying. In general, accurate haze estimation and removal from a single image is a challenging task due to its ill-posed nature.

Haze removal has been extensively studied in the literature. Early approaches focus on using multiple images or extra information [12, 20, 23, 24] for dehazing. Recently dehazing from a single image has gained considerable attention, and they can be broadly classified into two groups: methods based on transmission estimation [7, 10, 19, 26] and ones based on adaptive contrast enhancement [6, 9, 25]. Techniques in the latter group do not rely on any physical haze model, thus often suffer from visual artifacts such as strong color shift. The state-of-the-art methods often depend on a physical haze model for more accurate haze removal. They first estimate the atmosphere transmission map along with the haze color based on local image priors such as the dark channel prior [10] and the color-line prior [7]. The latent, haze-free image is then computed by directly removing the haze component in each pixel's color. Some methods are proposed to deal with special cases. For example, planar constraints can be utilized in road images [27]. Li *et al.* proposed a method to dehaze videos when the coarse depth maps can be estimated by multi-view stereo [17].

The state-of-the-art methods usually can generate satisfactory results on high quality input images. For lower quality inputs, such as images captured and processed by mobile phones, or compressed video clips, most existing dehazing methods will significantly amplify image artifacts that are visual unnoticeable in the input, especially in heavy haze regions. An example is shown in Fig. 1, where the input image is one video frame extracted from a sequence captured by a cellphone camera. After dehazing using previous methods [10, 16], strong visible artifacts appear in the sky region of the results. These artifacts cannot

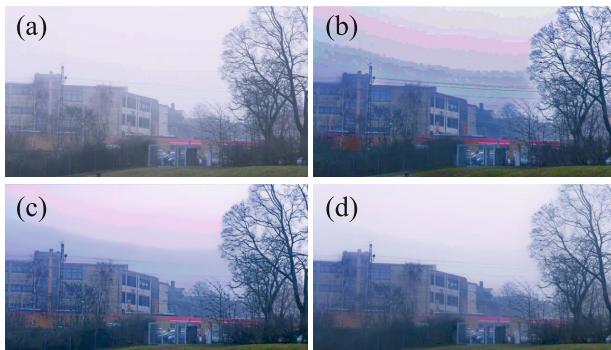


Fig. 1. Dehaze one video frame. (a) Input image. (b) Result of He *et al.* [10]. (c) Result of Li *et al.* [16]. (d) Ours. Note the strong banding and color shifting artifacts in the sky region in (b) and (c). (Color figure online)

be easily removed using post-processing filters without hampering the image content of other regions. Similarly, removing the original artifacts completely without destroying useful image details is also non-trivial as a pre-processing step.

Li *et al.* [16] were the first to consider the problem of artifact suppression in dehazing. Their approach is designed to remove only the blocking artifacts that are usually caused by compression. In this method, the input image is first decomposed into a structure layer and a texture layer, and dehazing is performed on the structure layer and deblocking is applied on the texture layer. The final output image is produced by re-combining the two layers. This method however often does not work well for other artifacts that commonly co-exist in lower quality inputs, *e.g.*, the color banding artifact in Fig. 1 and color aliasing in later examples. In addition, their final results tend to be over-smoothed with missing fine image details, as we will show in our experimental results. This suggests that independent dehazing and deblocking on two separate layers is sub-optimal.

In this work, we propose a new method for image and video dehazing with an emphasis on preventing different types of visual artifacts in the output. Our method follows the general two-step framework and makes contributions in each step: estimating atmosphere transmission map first, then recover the latent image. In the first step, after initializing the transmission map using existing local priors such as the dark channel prior [10], we refine it using a global method based on image guided Total Generalized Variation (TGV) [3] regularization. Compared with other commonly used refinement approaches, our method tends to produce transmission maps that are physically more correct: it produces very smooth regions within the surfaces/objects, while generates strong edges at depth discontinuities. Observing that the boosted visual artifacts by existing methods are often not visible in the input image, in the second stage, we propose a novel way to recover the latent image by minimizing the gradient residual between the output and input images. It suppresses new edges which does not exist in the input image (often are artifacts), but has little effects on the edges that already exist, which are ideal properties for the dehazing task. Considering the existence of artifacts, the linear haze model may not hold on every pixel. We then explicitly introduce an “error” layer in the optimization, which could separate out the large artifacts that violate the linear haze model. Both quantitative and qualitative experimental results show that our method generates more accurate and more natural-looking results than the state-of-the-art methods on compressed inputs. In particular, our method shows significant improvement on video dehazing, which can suppress both spatial and temporal artifacts.

2 Overview of Transmission Map Initialization

The transmission map in our framework is required to be initialized by existing local priors, *e.g.*, the widely used dark channel prior [10]. Here we provide a quick overview of the basic image formation model and this method. Note

that our main contributions, transmission map refinement and image recovery, are orthogonal to the specific method that one could choose for initializing the transmission map.

Koschmieder *et al.* [13] proposed a physical haze model as:

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

where \mathbf{I} is the hazy image, \mathbf{J} is the scene radiance, \mathbf{A} is the atmospheric light and assumed to be constant over the whole image, t is the medium transmission and x denotes the image coordinates. The transmission describes the portion of the light reaches to the camera without scattered. The task of dehazing is to estimate \mathbf{J} (with \mathbf{A} and t as by-products) from the input image \mathbf{I} , which is a severely ill-posed problem.

The dark channel prior, proposed by He *et al.* [10], is a simple yet efficient local image prior for estimating a coarse transmission map. The dark channel is defined as:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} J^c(y)), \quad (2)$$

where c denotes the color channel and $\Omega(x)$ is a local patch around x . Natural image statistics show that J^{dark} tends to be zero. We can rewrite Eq. (1) and take the minimum operations on both sides to get:

$$\min_{y \in \Omega(x)} (\min_c \frac{I^c(y)}{A^c}) = \min_{y \in \Omega(x)} (\min_c \frac{J(y)}{A^c} t(x)) + 1 - t(x). \quad (3)$$

By assuming the transmission map is constant in each small local patch, we can eliminate J^{dark} to obtain the coarse transmission map:

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} \frac{I^c(y)}{A^c}), \quad (4)$$

where the atmospheric light color \mathbf{A} can be estimated as the brightest pixel color in the dark channel. This coarse transmission map is computed locally, thus often need to be refined. In practice it is often refined by soft matting [14] or guided image filtering [11]. Finally, the scene radiance is recovered by:

$$\mathbf{J}(x) = (\mathbf{I}(x) - \mathbf{A})/\tilde{t}(x) + \mathbf{A}. \quad (5)$$

The dark channel prior described above is an elegant solution and often achieves high quality results for high quality images. However, as observed by Li *et al.* [16], image artifacts, such as noise or blocking, can affect both dark channel computation and transmission map smoothing. The original dark channel approach often cannot generate high quality results for images with artifacts.

3 TGV-Based Transmission Refinement

In He *et al.*'s method, the transmission map is refined by soft matting [14] or guided image filtering [11]. Both methods are edge-aware operations. They work

well with objects that have flat appearances. However, for objects/regions with strong textures, the refined transmission map using these methods tend to have false variations that are correlated with such textures. This is contradictory to the haze model, as the amount of haze in each pixel is only related to its depth, not its texture or color. Therefore, we expect the refined transmission map to be smooth inside the same object/surface, and only has discontinuities along depth edges. We thus propose a new transmission refinement method to try to achieve this goal without recovering the 3D scene.

We formulate the transmission refinement as a global optimization problem, consisting of a data fidelity term and regularization terms. Note that the transmission values of white objects are often underestimated by the dark channel method. We need a model that is robust to such outliers or errors. Instead of the commonly-used ℓ_2 norm data term, we use the ℓ_1 norm to somewhat tolerate outliers and errors. The second-order Total Generalized Variation (TGV) [3, 8, 21, 28] with a guided image is adopted for regularization. Compared with conventional Total Variation (TV) regularization that encourages *piecewise constant* images and often suffers from undesired staircasing artifacts, TGV prefers *piecewise smooth* images. This is a desired property for the transmission, as we may have a slanted plane (e.g., road, bridge) whose transmission varies smoothly along with the change of depth.

Given the initial transmission \tilde{t} and a guided image I , the optimization problem with TGV regularization is:

$$\min_{t,w} \{\alpha_1 \int |D^{1/2}(\nabla t - w)| dx + \alpha_0 \int |\nabla w| dx + \int |t - \tilde{t}| dx\}, \quad (6)$$

where $D^{1/2}$ is the anisotropic diffusion tensor [28] defined as:

$$D^{1/2} = \exp(-\gamma|\nabla I|^\beta) nn^T + n^\perp n^{\perp T}, \quad (7)$$

where n is the direction of the gradient of the guided image $n = \frac{\nabla I}{|\nabla I|}$ and n^\perp is the perpendicular direction, γ, β are parameters to adjust the sharpness and magnitude of the tensor, w is an auxiliary variable. Our experiments show that the sharp depth edges cannot be preserved without the guided image when using the TGV regularization. Unlike the previous local refinement methods, TGV performs globally and is less sensitive to the local textures.

To solve this problem, we apply the prime-dual minimization algorithm [4] with the Legendre Fenchel transform. The transformed primal-dual problem is given by:

$$\min_{t,w} \max_{p \in P, q \in Q} \{\alpha_1 \langle D^{1/2}(\nabla t - w), p \rangle + \alpha_0 \langle \nabla w, q \rangle + \int |t - \tilde{t}| dx\}, \quad (8)$$

where p, q are dual variables and their feasible sets are:

$$\begin{aligned} P &= \{p \in R^{2MN}, \|p\|_\infty \leq 1\}, \\ Q &= \{q \in R^{4MN}, \|q\|_\infty \leq 1\}. \end{aligned} \quad (9)$$

The algorithm for transmission refinement is formally summarized in Algorithm 1.

Algorithm 1. Transmission map refinement by Guided TGV

```

Initialization:  $t^0 = \tilde{t}$ ,  $w^0, \bar{t}^0, \bar{w}^0, p^0, q^0 = 0$ ,  $\sigma_p > 0$ ,  $\sigma_q > 0$ ,  $\tau_t > 0$ ,  $\tau_w > 0$ 
for  $k = 0$  to  $Maxiteration$  do
     $p^{k+1} = \mathcal{P}[p^k + \sigma_p \alpha_1 (D^{1/2} (\nabla \bar{t}^k - \bar{w}^k))]$ 
     $q^{k+1} = \mathcal{P}[q^k + \sigma_q \alpha_0 \nabla \bar{w}^k]$ 
     $t^{k+1} = thresholding_{\tau}(t^k + \tau_u \alpha_1 \nabla^T D^{1/2} p^{k+1})$ 
     $w^{k+1} = w^k + \tau_w (\alpha_0 \nabla^T q^{k+1} + \alpha_1 D^{1/2} p^{k+1})$ 
     $\bar{t}^{k+1} = t^{k+1} + \theta(t^{k+1} - t^k)$ 
     $\bar{w}^{k+1} = w^{k+1} + \theta(w^{k+1} - \bar{w}^k)$ 
end for

```

In the algorithm, $\sigma_p > 0$, $\sigma_q > 0$, $\tau_t > 0$, $\tau_w > 0$ are step sizes and k is the iteration counter. The element-wise projection operators \mathcal{P} is defined:

$$\mathcal{P}[x] = \frac{x}{\max\{1, |x|\}}. \quad (10)$$

The $thresholding_{\tau}()$ denotes the soft-thresholding operation:

$$thresholding_{\tau}(x) = \max(|x| - \tau, 0)\text{sign}(x). \quad (11)$$

θ is updated in every iteration as suggested by [4]. The divergence and gradient operators in the optimization are approximated using standard finite differences. Please refer to [4] for more details of this optimization method.

Figure 2 shows the transmission maps estimated by guided filter, matting followed by bilateral filter and TGV refinement. Compared with guided image

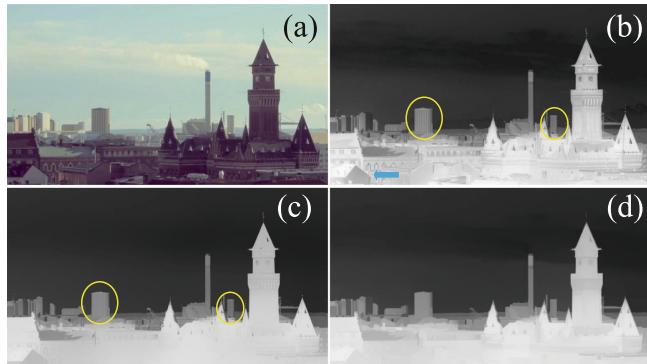


Fig. 2. Comparisons of transmission refinement methods. (a) Input image. (b) Result of guided image filtering [11]. (c) Result of matting followed by bilateral filtering [10]. (d) Ours. (Color figure online)

filtering or bilateral smoothing, our method is aware of the depth edges while producing smooth surface within each objects (see the buildings indicated by the yellow circles). In addition, our optimization scheme does not exactly trust the initialization and it can somewhat tolerate the errors (see the house indicated by the blue arrow).

4 Robust Latent Image Recovery by Gradient Residual Minimization

After the transmission map is refined, our next goal is to recovery the scene radiance \mathbf{J} . Many existing methods obtain it by directly solving the linear haze model (5), where the artifacts are treated equally as the true pixels. As a result, the artifacts will be also enhanced after dehazing.

Without any prior information, it is impossible to extract or suppress the artifacts from the input image. We have observed that in practice, the visual artifacts are usually invisible in the input image. After dehazing, they pop up as their gradients are amplified, introduce new image edges that are not consistent with the underlying image content, such as the color bands in Fig. 1(b,c). Based on this observation, we propose a novel way to constrain the image edges to be structurally consistent before and after dehazing. This motivates us to minimize the residual of the gradients between the input and output images under the sparse-inducing norm. We call it Gradient Residual Minimization (GRM). Combined with the linear haze model, our optimization problem becomes:

$$\min_{\mathbf{J}} \left\{ \frac{1}{2} \int \|\mathbf{J}t - (\mathbf{I} - \mathbf{A} + \mathbf{At})\|_2^2 dx + \eta \int \|\nabla \mathbf{J} - \nabla \mathbf{I}\|_0 dx \right\}, \quad (12)$$

where the ℓ_0 norm counts the number of non-zero elements and η is a weighting parameter. It is important to note that the above spares-inducing norm only encourages the non-zero gradients of \mathbf{J} to be at the same positions of the gradients of \mathbf{I} . However, their magnitudes do not have to be the same. This good property of the edge-preserving term is very crucial in dehazing, as the contrast of the overall image will be increased after dehazing. With the proposed GRM, new edges (often caused by artifacts) that do not exist in the input image will be penalized but the original strong image edges will be kept.

Due to the existence of the artifacts, it is very possible that the linear haze model does not hold on every corrupted pixel. Unlike previous approaches, we assume there may exist some artifacts or large errors \mathbf{E} in the input image, which violates the linear composition model in Eq. (1) locally. Furthermore, we assume \mathbf{E} is sparse. This is reasonable as operations such as compression do not damage image content uniformly: they often cause more errors in high frequency image content than flat regions. With above assumptions, to recover the latent image, we solve the following optimization problem:

$$\min_{\mathbf{J}, \mathbf{E}} \left\{ \frac{1}{2} \int \|\mathbf{J}t - (\mathbf{I} - \mathbf{E} - \mathbf{A} + \mathbf{At})\|_2^2 dx + \lambda \int \|\mathbf{E}\|_0 dx + \eta \int \|\nabla \mathbf{J} - \nabla \mathbf{I}\|_0 dx \right\}, \quad (13)$$

where λ is a regularization parameter. Intuitively, the first term says that after subtracting \mathbf{E} from the input image \mathbf{I} , the remaining component $\mathbf{I} - \mathbf{E}$, together with the latent image \mathbf{J} and the transmission map \mathbf{A} , satisfy the haze model in Eq. (1). The second term \mathbf{E} represents large artifacts while the last term encodes our observations on image edges.

However, the ℓ_0 minimization problem is generally difficult to solve. Therefore in practice, we replace it with the closest convex relaxation – ℓ_1 norms [5, 15]:

$$\min_{\mathbf{J}, \mathbf{E}} \left\{ \frac{1}{2} \int \|\mathbf{J}t - (\mathbf{I} - \mathbf{E} - \mathbf{A} + \mathbf{A}t)\|_2^2 dx + \lambda \int \|\mathbf{E}\|_1 dx + \eta \int \|\nabla \mathbf{J} - \nabla \mathbf{I}\|_1 dx \right\}. \quad (14)$$

We alternately solve this new problem by minimizing the energy function with respect to \mathbf{J} and \mathbf{E} , respectively. Let $\mathbf{Z} = \mathbf{J} - \mathbf{I}$, and the \mathbf{J} subproblem can be rewritten as:

$$\min_{\mathbf{Z}} \left\{ \frac{1}{2} \int \|(\mathbf{Z} + \mathbf{I})t - (\mathbf{I} - \mathbf{E} - \mathbf{A} + \mathbf{A}t)\|_2^2 dx + \eta \int \|\nabla \mathbf{Z}\|_1 dx \right\}, \quad (15)$$

which is a TV minimization problem. We can apply an existing TV solver [1] for this subproblem. After \mathbf{Z} is solved, \mathbf{J} can be recovered by $\mathbf{J} = \mathbf{Z} + \mathbf{I}$. For the \mathbf{E} subproblem:

$$\min_{\mathbf{E}} \left\{ \frac{1}{2} \int \|\mathbf{J}t - (\mathbf{I} - \mathbf{E} - \mathbf{A} + \mathbf{A}t)\|_2^2 dx + \lambda \int \|\mathbf{E}\|_1 dx \right\}, \quad (16)$$

it has a closed-form solution by soft-thresholding. The overall algorithm for latent image recovery is summarized in Algorithm 2.

Algorithm 2. Robust Image Dehazing

```

Initialization:  $E^0 = 0$ ,  $J^0 = \frac{\mathbf{I}-\mathbf{A}}{t} + \mathbf{A}$ 
for  $k = 0$  to  $Maxiteration$  do
     $Z_b = \mathbf{I} - E^k - \mathbf{A} + At - It$ 
     $Z = \arg \min_Z \left\{ \frac{1}{2} \int \|Zt - Z_b\|_2^2 dx + \eta \int \|\nabla Z\|_1 dx \right\}$ 
     $J^{k+1} = \mathbf{I} + Z$ 
     $E^{k+1} = thresholding_\lambda(\mathbf{I} - J^{k+1}t - (1-t)\mathbf{A})$ 
end for

```

The convergence of Algorithm 2 is shown in Fig. 3. We initialize \mathbf{J} with the least squares solution without GRM and a zero image \mathbf{E} . As we could see, the object function in Eq. (14) decreased monotonically and our method gradually converged. From the intermediate results, it can be observed that the initial \mathbf{J} has visible artifacts in the sky region, which is gradually eliminated during the optimization. One may notice that \mathbf{E} converged to large values on the tower and building edges. As we will show later, these are the aliasing artifacts caused by compression. And our method can successfully separate out these artifacts.

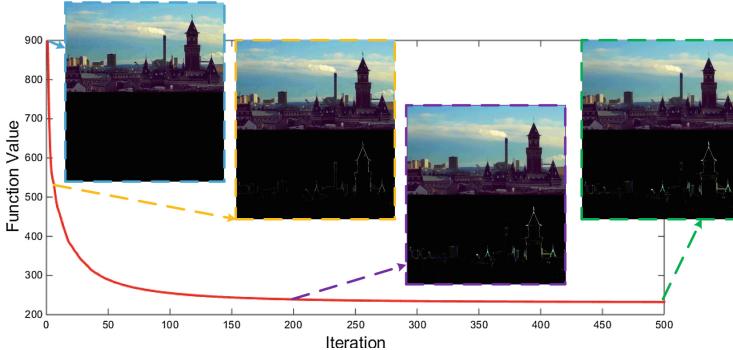


Fig. 3. The convergence of proposed method. The object function in Eq. (14) is monotonically decreasing. The intermediate results of \mathbf{J} and $10 \times \mathbf{E}$ at iteration 1, 5, 200 and 500 are shown.

5 Experiments

More high resolution image and video results are in the supplementary material. For quality comparisons, all the images should be viewed on screen instead of printed version.

5.1 Implementation Details

In our implementation, the tensor parameters are set as $\beta = 9$, $\gamma = 0.85$. The regularization parameters are $\alpha_0 = 0.5$, $\alpha_1 = 0.05$, $\lambda = 0.01$ and $\eta = 0.1$. We found our method is not sensitive to these parameters. The same set of parameters are used for all experiments in this paper. We terminate Algorithm 1 after 300 iterations and Algorithm 2 after 200 iterations.

We use the same method in He *et al.*'s approach to estimate the atmospheric light \mathbf{A} . For video inputs, we simply use the \mathbf{A} computed from the first frame for all other frames. We found that fixing \mathbf{A} for all frames is generally sufficient to get temporally coherent results by our model.

Using our MATLAB implementation on a laptop computer with a i7-4800 CPU and 16 GB RAM, it takes around 20 s to dehaze a 480×270 image. In comparison, 10 min per frame is reported in [17] on the same video frames. Same as many previous works [7], we apply a global gamma correction on images that become too dark after dehazing, just for better displaying.

5.2 Evaluation on Synthetic Data

We first quantitatively evaluate the performance of the proposed transmission estimation method using a synthetic dataset. Similar to previous practices [26], we synthesize hazy images from stereo pairs [18, 22] with known disparity maps. The transmission maps are simulated in the same way as in [26]. Since our

method is tailored towards suppressing artifacts, we prepare two test sets: one with high quality input images, the other with noise and compression corrupted images. To synthesize corruption, we first add 1% of Gaussian noise to the hazy images. These images are then compressed using the JPEG codec in Photoshop, with the compression quality 8 out of 12.

In Tables 1 and 2 we show the MSE of the haze map and the recovered image by different methods, on the clean and the corrupted datasets, respectively. The results show that our method achieves more accurate haze map and latent image than previous methods in most cases. One may find that the errors for corrupted inputs sometimes are lower than those of noise-free ones. It is because the dark channel based methods underestimated the transmission on these bright indoor scenes. The transmission may be slightly preciser when noise makes the images more colorful. Comparing the results of the two tables, the improvement by our method is more significant on the second set, which demonstrates its ability to suppress artifacts.

Table 1. Quantitative comparisons on the clean synthetic dataset. Table reports the MSE (10^{-3}) of the transmission map (left) and the output image (right).

	Aloe	Barn	Cones	Dolls	Moebius	Monopoly	Teddy	Rocks
He et al. [10]	5.6/17.4	0.9/ 7.9	8.6/13.7	8.0/14.3	7.5/18.6	11.2/30.1	10.3/20.1	4.6/11.4
Li et al. [16]	4.5/13.2	1.6/13.9	5.8/8.9	4.1 /7.0	5.7/ 12.7	8.9/23.8	4.0/6.6	3.7/9.5
Ours	4.4/10.4	0.8/9.0	5.6/8.0	3.8 /6.8	5.4/12.7	7.9/20.7	6.6/10.7	3.2/8.0

Table 2. Quantitative comparison on the noise and compression corrupted synthetic dataset. Table reports the MSE (10^{-3}) of the transmission map (left) and the output image (right).

	Aloe	Barn	Cones	Dolls	Moebius	Monopoly	Teddy	Rocks
He et al. [10]	5.3/17.0	1.0/11.2	8.1/12.9	7.6/13.8	7.2/17.4	10.7/27.9	10.0/19.3	4.1/10.3
Li et al. [16]	4.4/13.2	1.5/14.2	5.6/8.9	4.0/7.1	5.7/12.7	8.8/22.6	4.0/6.7	3.6/9.3
Ours	3.9/9.9	1.0/12.8	5.2/7.1	3.5/6.2	5.1/11.4	7.5/18.6	6.5/10.1	2.8/6.8

5.3 Real-World Images and Videos

We compare our method with some recent works [9, 16, 19] on a real video frame in Fig. 4. The compression artifacts and image noise become severe after dehazing by Meng *et al.*'s method and the Dehaze feature in Adobe Photoshop. Galdran *et al.*'s result suffers from large color distortion. He *et al.* have pointed out the similar phenomenon of Tan *et al.*'s method [25], which is also based on contrast enhancement. Li *et al.*'s method [16] is designed for blocking artifact suppression.

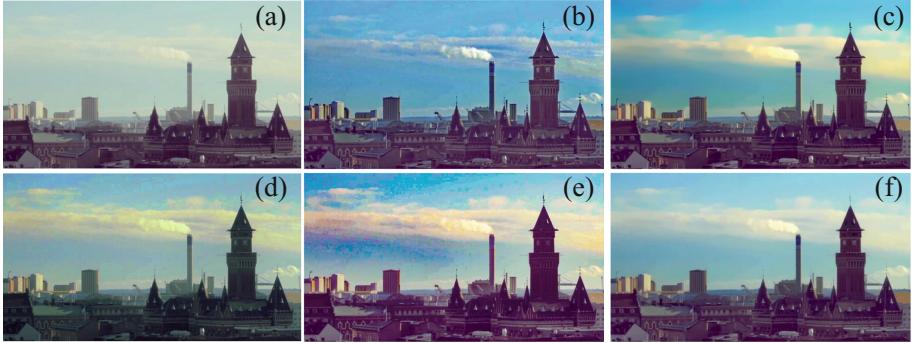


Fig. 4. Dehazing results of different methods. (a) Input image. (b) Meng *et al.*'s result [19]. (c) Li *et al.*'s result [16]. (d) Galdran *et al.*'s result [9]. (e) Photoshop 2015 dehazing result. (f) Our result.

Although their result does not contain such artifacts, the sky region is quite over-smoothed. Our result maintains subtle image features while at the same time successfully avoids boosting these artifacts.

Our method can especially suppress halo and color aliasing artifacts around depth edges that are common for previous methods, as shown in the zoomed-in region of the tower in Fig. 5. Except the result by our method, all other methods produce severe halo and color aliasing artifacts around the sharp tower boundary. Pay special attention to the flag on the top of the tower: the flag is dilated by all other methods except ours. Figure 5(h) visualizes the artifact map \mathbf{E} in Eq. (14), it suggests that our image recovery method pays special attention to the boundary pixels to avoid introducing aliasing by dehazing. We also include our result without the proposed GRM in Fig. 5(f). The blocky artifacts and color aliasing around the tower boundary can not be reduced on this result, which demonstrates the effectiveness of the proposed model.

In Fig. 6, we compare our method with two variational methods [9, 17] proposed recently on a video frame. Galdran *et al.*'s method [9] converged in a few iterations on this image, but the result still contains haze. The method in

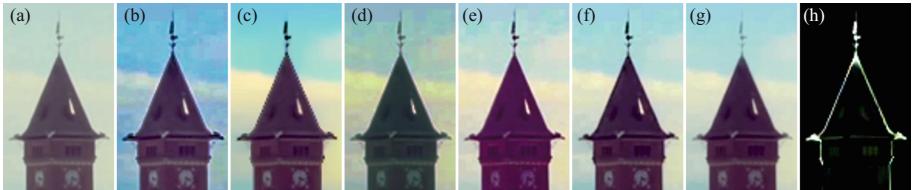


Fig. 5. Zoomed-in region of Fig. 4. (a) Input image. (b) Meng *et al.*'s result [19]. (c) Li *et al.*'s result [16]. (d) Galdran *et al.*'s result [9]. (e) Photoshop 2015 result. (f) Our result without the proposed GRM. (g) Our result. (h) Our $\mathbf{E} \times 10$.

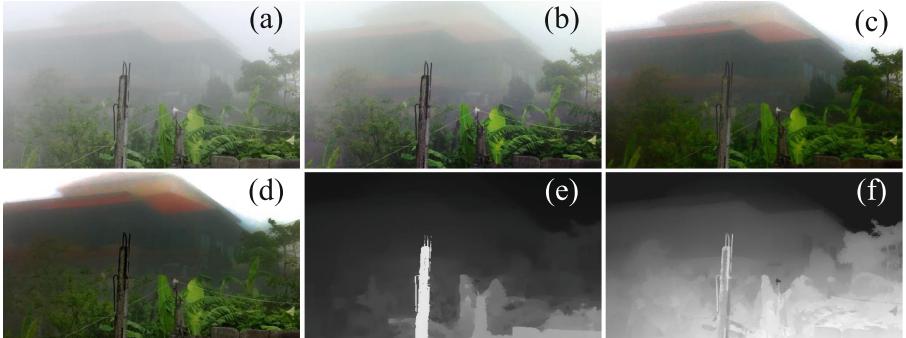


Fig. 6. Comparison with some recent methods. (a) Input video frame. (b) Galdran *et al.*'s result [9]. (c) Li *et al.*'s result [17]. (d) Our result. (e) Li's depth [17] (computed using the whole video). (f) Our transmission map.

[17] performs simultaneously dehazing and stereo reconstruction, thus it only works when structure-from-motion can be calculated. For general videos contain dynamic scenes or a single image, it cannot be applied. From the results, our method is comparable to that in [17], or even better. For example, our method can remove more haze on the building. This is clearer on Li's depth map, where the shape of the building can be hardly found.

We further compare our method with the deblocking based method [16] on more video sequences in Fig. 7. Li *et al.*'s method generates various artifacts in these examples, such as the over-sharpened and over-saturated sea region in the first example, the color distortion in the sky regions of the second, and the halos around the buildings and the color banding in the third example. In the bottom example, there is strong halo near the intersection of the sky and sea. Another drawback of Li *et al.*'s method is that fine image details are often lost, such as the sea region in the last example. In contrast, our results contain much less visual artifacts and appear to be more natural.

For videos, the flickering artifacts widely exist on the previous frame-by-frame dehazing methods. It is often caused by the artifacts and the change of overall color in the input video. Recently, Bonneel *et al.* proposed a new method to remove the flickering by enforcing temporal consistency using optical flow [2]. Although their method can successfully remove the temporal artifacts, it does not work for the spatial artifacts on each frame. Figure 8 shows one example frame of a video, where their result inherits all the structured artifacts from the existing method. Although we only perform frame-by-frame dehazing, the result shows that our method is able to suppress temporal artifacts as well. This is because the input frames already have good temporal consistency. Such temporal consistency is transferred into our result frame-by-frame by the proposed GRM.

We recruited 34 volunteers through the Adobe mail list for a user study of result quality, which contained researches, interns, managers, photographers *etc.* For each example, we presented three different results anonymously (always

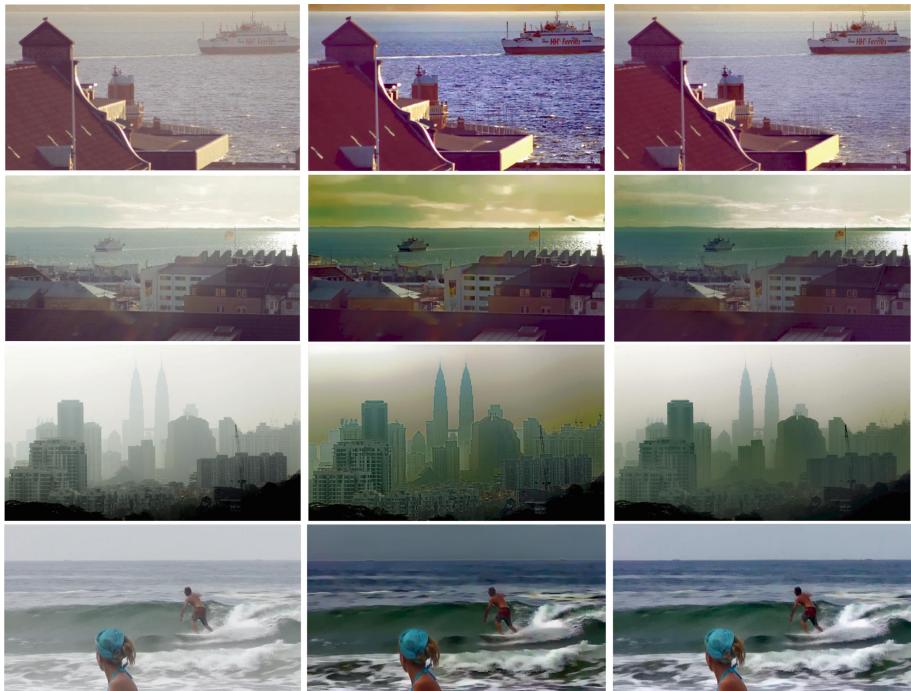


Fig. 7. Comparison with Li *et al.*'s method. First column: input video frame. Second column: Li *et al.*'s result [16]. Third column: our result.

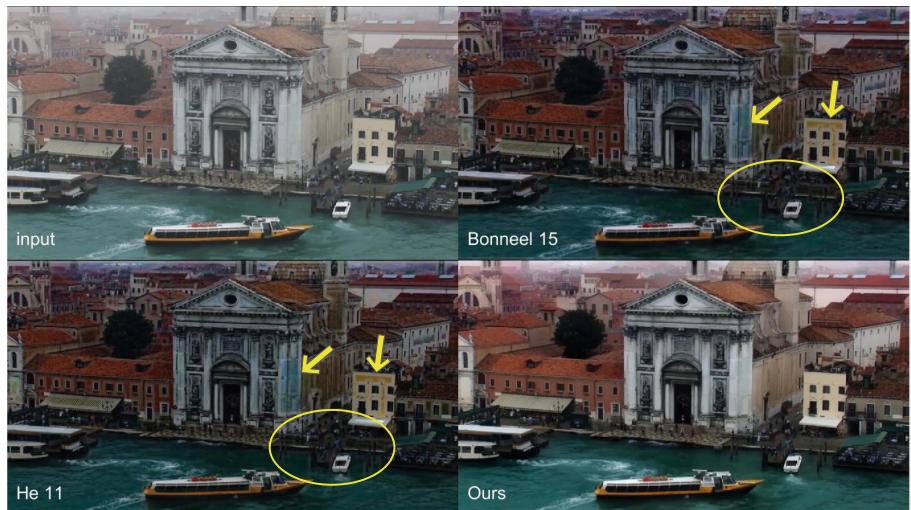


Fig. 8. A frame of video dehazing results. The full video is in the supplementary material. The halos around the pillars and structured artifacts are indicated by the yellow circle and arrows. (Color figure online)

including ours) in random orders, and asked them to pick the best dehazing result, based on realism, dehazing quality, artifacts *etc.* 52.9 % subjects preferred our “bali” result in Fig. 6, 47.1 % preferred the result in [17] and 0 % for He *et al.*’s [10]. We have mentioned above that [17] requires external structure-from-motion information, while ours does not and can be applied to more general dehazing. For Fig. 8, 91.2 % preffered our results over He *et al.*’s [10] and Bonneau *et al.*’s [2]. For the rest of examples in this paper, our results were the preferred ones also (by 73.5 %–91.2 % people), where overall 80.0 % picked our results over Li *et al.*’s [16] (14.7 %) and He *et al.*’s [10] (5.3 %).

5.4 Discussion

One may argue there are simpler alternatives to handle artifacts in the dehazing pipeline. One way is to explicitly remove the image artifacts before dehazing, such as Li *et al.*’s method. However, accurately removing all image artifacts itself is a difficult task. If not done perfectly, the quality of the final image will be compromised, as shown in various examples in this paper. Another alternative is to simply reduce the amount of haze to be removed. However, it will significantly decrease the power of dehazing, as we show in the tower example in the supplementary material. Our method is a more principle way to achieve a good balance between dehazing and minimizing visual artifacts.

Despite its effectiveness, our method still has some limitations. Firstly, our method inherits the limitations of the dark channel prior. It may over-estimate the amount of haze for white objects that are close to the camera. In addition, for very far away objects, our method can not significantly increase their contrast, which is due to the ambiguity between the artifacts and true objects covered by very thick haze. It is even difficult for human eyes to distinguish them without image context. Previous methods also have poor performance on such challenging

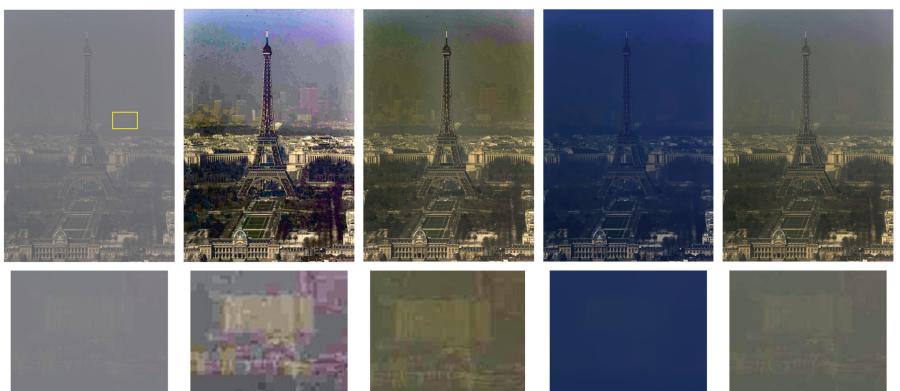


Fig. 9. Dehazing a low quality JPEG image. From left to right are the input image and the results by: Fattal *et al.* [7], He *et al.* [10], Li *et al.* [16] and ours. The bottom row shows the zoomed-in areas corresponding to the yellow box. (Color figure online)

tasks: they either directly amplify all the artifacts or mistakenly remove the distant objects to produce over-smoothed results.

Figure 9 shows one such example that contains some far-away buildings surrounded by JPEG artifacts. Both Fattal *et al.*'s result and He *et al.*'s have serve JPEG artifacts after dehazing. On the contrary, in Li *et al.*'s result, the distant buildings are mistakenly removed by their deblocking filter, and become much less visible. Although our method cannot solve the ambiguity mentioned above to greatly enhance the far-away buildings, it can automatically take care of the artifacts and generate a more realistic result.

6 Conclusion

We have proposed a new method to suppress visual artifacts in image and video dehazing. By introducing a gradient residual and error layer into the image recovery process, our method is able to remove various artifacts without explicitly modeling each one. A new transmission refinement method is introduced in this work, which contributes to improving the overall accuracy of our results. We have conducted extensive evaluation on both synthetic datasets and real-world examples, and validated the superior performance of our method over the state-of-the-arts for lower quality inputs. While our method works well on the dehazing task, it can be potentially extended to other image enhancement applications, due to the similar artifacts-amplification nature of them.

References

1. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**(11), 2419–2434 (2009)
2. Bonneel, N., Tompkin, J., Sunkavalli, K., Sun, D., Paris, S., Pfister, H.: Blind video temporal consistency. *ACM Trans. Graph.* **34**(6), 196 (2015)
3. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010)
4. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
5. Chen, C., Li, Y., Liu, W., Huang, J.: Image fusion with local spectral consistency and dynamic gradient sparsity. In: CVPR, pp. 2760–2765 (2014)
6. Fattal, R.: Single image dehazing. *ACM Trans. Graph.* **27**, 72 (2008)
7. Fattal, R.: Dehazing using color-lines. *ACM Trans. Graph.* **34**(1), 13 (2014)
8. Ferstl, D., Reinbacher, C., Ranftl, R., Rüther, M., Bischof, H.: Image guided depth upsampling using anisotropic total generalized variation. In: ICCV, pp. 993–1000 (2013)
9. Galdran, A., Vazquez-Corral, J., Pardo, D., Bertalmío, M.: Enhanced variational image dehazing. *SIAM J. Imaging Sci.* **8**(3), 1519–1546 (2015)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011)
11. He, K., Sun, J., Tang, X.: Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013)

12. Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., Uyttendaele, M., Lischinski, D.: Deep photo: model-based photograph enhancement and viewing. *ACM Trans. Graph.* **27**, 116 (2008)
13. Koschmieder, H.: Theorie der horizontalen Sichtweite. In: *Beitrge zur Physik der freien Atmosphre* (1924)
14. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 228–242 (2008)
15. Li, Y., Chen, C., Yang, F., Huang, J.: Deep sparse representation for robust image registration. In: *CVPR*, pp. 4894–4901 (2015)
16. Li, Y., Guo, F., Tan, R.T., Brown, M.S.: A contrast enhancement framework with JPEG artifacts suppression. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014, Part II*. LNCS, vol. 8690, pp. 174–188. Springer, Heidelberg (2014)
17. Li, Z., Tan, P., Tan, R.T., Zou, D., Zhiying Zhou, S., Cheong, L.F.: Simultaneous video defogging and stereo reconstruction. In: *CVPR*, pp. 4988–4997 (2015)
18. Lu, S., Ren, X., Liu, F.: Depth enhancement via low-rank matrix completion. In: *CVPR*, pp. 3390–3397 (2014)
19. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: *ICCV*, pp. 617–624 (2013)
20. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(6), 713–724 (2003)
21. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In: *IEEE Intelligent Vehicles Symposium*, pp. 401–407 (2012)
22. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision* **47**(1–3), 7–42 (2002)
23. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Instant dehazing of images using polarization. In: *CVPR*, vol. 1, pp. I–325 (2001)
24. Shwartz, S., Namer, E., Schechner, Y.Y.: Blind haze separation. In: *CVPR*, vol. 2, pp. 1984–1991 (2006)
25. Tan, R.T.: Visibility in bad weather from a single image. In: *CVPR*, pp. 1–8 (2008)
26. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: *CVPR*, pp. 2995–3002 (2014)
27. Tarel, J.P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H., Gruyer, D.: Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intell. Transp. Syst. Mag.* **4**(2), 6–20 (2012)
28. Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., Bischof, H.: Anisotropic Huber-L1 optical flow. In: *BMVC*, vol. 1, p. 3 (2009)