Haze Removal Using the Difference-Structure-Preservation Prior

Linyuan He, Jizhong Zhao, *Member, IEEE*, Nanning Zheng, *Fellow, IEEE*, and Duyan Bi

***Abstract*— Fog cover is generally present in outdoor scenes, which limits the potential for efficient information extraction from images. In this paper, the goal of the developed algorithm is to obtain an optimal transmission map as well as to remove hazes from a single input image. To solve the problem, we meticulously analyze the optical model and recast the initial transmission map under an additional boundary prior. For better preservation of the results, the difference-structure-preservation dictionary could be learned, such that the local consistency features of the transmission map could be well preserved after coefficient shrinkage. Experimental results show that the method preserves the natural appearance of the image.**

***Index Terms*— Visibility restoration, contrast restoration, single image dehazing, difference-structure-preservation, local depth consistency.**

I. Introduction

OUTDOOR images are usually affected by the presence of clouds or haze, which obstruct image structural features such as edges. As shown in Fig. 1, the visibility, contrast, and vividness of the scene are dramatically degraded, which makes it difficult to distinguish objects [1], [2]. Hence, it is necessary to recover the original scene to improve human identification ability. In general, dehazing is a special case of image restoration. Following Koschmiedar's law, a hazy image is normally modeled by [3]–[5]:

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

Note that our goal is to recover the underlying scene radiance, where $c \in \{r, g, b\}$, $I^c(x) \in R^C$ is the observed intensity at pixel x, $J^c(x)$ is the scene radiance or haze-free image, A^c is the sky brightness for the whole image, and $t(x)$ is the scene transmission, which is correlated with the scene depth. Because A^c is usually determined empirically, the major difficulty of image dehazing is to calculate $t(x)$. To make the ill-posed problem tractable, many methods use suitable priors or additional information to estimate the transmission map and then obtain the haze-free image.

Recently, three types of prior have been widely used for the transmission map: global prior, local prior and similarity prior. The global prior approaches are split into two groups. The approach of measurements assumes that the natural image belongs to a continuous function space, and can define a norm or semi-norm to measure the global regularity [6]. Due to the strict mathematical theory, it has received widespread attention. However, a natural image is complex and mutative, so it is difficult to decide which smooth space an image is embedded into. In addition, the global regularity prior only focuses on the whole image property rather than the local structures, so it may lead to deviations in describing the details of the image. A typical dehazing algorithm uses total variation minimization [6]. The purpose of the other group of methods is to determine the specific characteristics based on the statistical distribution prior. Although it uses the likelihood method for many kinds of single scenes, it ignores the relative differences among these scenes, resulting in ever more deviations. For instance, Nishino [7] used a Bayesian statistical prior that jointly estimated the scene albedo and depth from a single degraded image, but it tended to over-enhance the contrast on some occasions. After that, Zhu *et al.* [8] proposed the global color-attenuation prior to help create a linear model for the scene depth of a hazy image. Unfortunately, the global color-attenuation prior can be violated when large depth jumps exist. As mentioned previously, pixels are considered a kind of distribution across the entire image, therefore capturing a global characteristic. However, they only construct a roughly estimated transmission map due to neglect of the local features.

In consideration of the limitations of global techniques, we turn to the local image prior for help, since it contains the more valid information that hides in the neighborhood framework. Through the analysis of every local structure, which can provide additional prior information for image dehazing, we can easily determine more local features to solve this ill-posed problem. For example, Tan's method [9] defined a prior based on the idea that images with enhanced visibility should have higher contrast than images plagued by bad weather, which improved the visibility considerably by maximizing the local contrast; however, the results looked unnatural because of over-enhancement. After that, Caraffa and Tarel [10] and Fattal [11] viewed the image as a relaxed Markov Random Field (MRF) to obtain better results using the planar constraint. Wang [12] presented a multi-scale

Manuscript received June 25, 2015; revised November 28, 2015, March 24, 2016, June 6, 2016, August 23, 2016, and November 15, 2016; accepted December 16, 2016. Date of publication December 21, 2016; date of current version January 20, 2017. This work was supported by the National Natural Science Foundation of China, under Grant 61372167 and Grant 61379104. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Karen Egiazarian.

L. He is with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Department of Aeronautics and Astronautics, Air Force Engineering University, Xi'an 710038, China (e-mail: hal1983@163.com).

J. Zhao and N. Zheng are with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zjz@mail.xjtu.edu.cn; nnzheng@mail.xjtu.edu.cn).

D. Bi is with the Department of Aeronautics and Astronautics, Air Force Engineering University, Xi'an 710038, China (e-mail: biduyan@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2644267

depth fusion method as a local energy minimization problem that incorporated spatial Markov dependence. The technique of Fattal [13] restored the scene by assuming that the colors in a local window lie on a single line in the RGB color space. He *et al.* [14] proposed an efficient haze removal and transmission map estimation method for outdoor colored RGB images. Among the methods listed above, the Dark Channel Prior received the most attention [14]. However, it may fail to account for higher-order geometric relationships among scene elements. Hence many follow-up methods were proposed to boost the performance of the initial rough transmission map using different filters [15]–[17]. Unfortunately, they seem unable to achieve significant improvements on the estimates of scene transmission. More importantly, because these methods still used DCP to obtain their rough transmission maps, this initial process may further compromise the accuracy of the final results in the original haze model. Meng *et al.* [18] and Lai *et al.* [19] leveraged a local prior to develop a depth propagation framework for situations in which the dark channel prior was misleading. Although the framework followed the local probabilistic principle, it still tended to obtain a rough estimate of the scene depth, owing to the hypotheses from the local statistical distribution or empirical observation.

If the global and local priors can be considered attributes of natural images, the similarity prior can be regarded as a typical expression of visual features that have been shown to have strong relevance based on the dictionary model. Normally, a dictionary model can be regarded as an oracle assumption [20], [21] and data-driven hypothesis [22]–[26], but a large number of experiments have shown that the latter is more effective than the former. Despite its popularity, the similarity prior is also inaccurate, since it implies an unrealistic extension of the geometric scene structure that can lead to severe reconstruction artifacts.

From the previous work discussed above, we can conclude that the key to settling this problem is to recover the transmission map, which is the major task in single-image dehazing. The general principle is to explore additional constraints or priors. Inspired by this idea, we begin this study by deriving an inherent boundary constraint on the pixel values on the scene transmission, which has a clear geometric interpretation. Compared to DCP [14], it is able to truly reflect the local consistency and the differences between the pixels. In addition, to preserve the consistency of the local structure, a neighboring-structure dictionary is defined that can efficiently characterize the local structure by combining the spatial similarity and diversity relationships among neighboring patches. In the choice of similar patches, our model can be viewed as comparable to BM3D [26], and similar to K-LLD [24]. More importantly, we concentrate on establishing the difference-structure-preservation, basis in similar local patches instead of the wavelet basis used in BM3D [26]. Different from the algorithm in K-LLD [24], which aims only to preserve the global Euclidean structures of degraded images, an improvement of the dictionary method can well preserve the local intrinsic structures. Although K-SVD [22] used the minimum squared error to update the dictionary, it did not consider the structures induced by the similarities among neighboring pixels. By contrast, we provide a different point of view. Similar to the work on LPG-PCA [27], we first model a pixel and its nearest neighbors as a vector variable whose training samples are selected from the larger local window using the cosine similarity method. This guarantees that only the sample blocks with similar contents are considered in the local statistical calculation. In particular, motivated by the linear assumption in TDNL [28], we also assume that each patch can be linearly represented by a dictionary, in order to maintain the maximum difference of local similarity. Hence, the model can avoid the "curse of dimensionality" as well as achieve perfect generalization performance. Experiments indicate that the difference-structure-preservation variation method performs as well as or better than the state-of-the-art dehazing methods.

The remainder of this paper is organized as follows. In Section II we briefly analyze an inherent boundary constraint on the scene transmission map so that it is formalized into an optimization problem to recover the unknown transmission. Section III presents the DSPV dehazing algorithm in detail. Additionally, in Section IV, we present experimental validation of our method and compare it, visually and quantitatively, to some recently proposed popular dehazing methods. Finally, we conclude the paper and describe future work in Section V.

II. Solution of Scene Transmission

In this section, compared to the dark channel constraint, we show that the estimate of the transmission can be effectively bounded via geometric analysis. We also present the derived bounds, which have been directly used as the initial transmission map. The illustration shows the superiority of the boundary constraint.

A. Dark Channel on Transmission

The atmospheric-scattering model, which is simplified as (1), has been demonstrated to be successful for single-image dehazing [6]–[19]. For this model, the most widely used algorithm is DCP. The dark channel prior is based on the key observation that a haze-free outdoor image exhibits at least one color channel with a very low intensity value in patches of the image that do not contain any sky parts [14]. According to the dark channel prior method, the minimum intensity value can be described as a function and represented as

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

where J^c is a color channel of the RGB image J, $\Omega(x)$ is a local patch around x, $\min_{C\in\{r,g,b\}}$ is the minimum value of the color channel and $\min_{y\in\Omega(x)}$ is the minimum filter. When the outdoor image lacks haze, the dark channel value of the haze-free image is close to zero and can be represented by $J^{dark}(x) \to 0$. In other words, if the local dark channel value is large in a region, this region is likely to exhibit haze. As such, we can estimate the amount of haze via (2).

Fig. 1. Effects of atmospheric scattering.

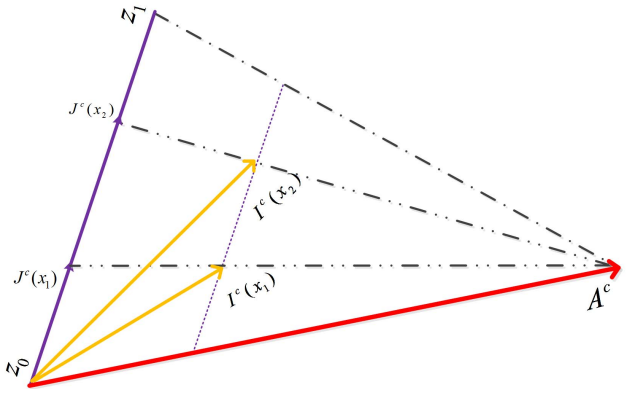


Fig. 2. The boundary constraint of Real Radiance. For each pixel, we require the extrapolation of $J^c(x)$ cannot cross over the boundary of the radiance which is corresponding to the constraint assumption.

According to the work in [14], the resulting estimation of the initial transmission map is

$$t(x) = 1 - \beta \min_{y \in \Omega(x)} \left(\min_{C \in \{r,g,b\}} \frac{I^c(y)}{A^c} \right) \tag{3}$$

where $\beta = 0.95$ retains a portion of the haze for distant objects. Moreover, Nishino [7] suggested that the optimal patch size Ω of the dark channel prior is 15×15. Once both $t(x)$ and A^c are estimated, the scene radiance can be recovered by

$$J^c(x) = \frac{I^c(x) - A^c}{t(x)} + A^c \tag{4}$$

Although DCP has had great success, it is quite possible that some areas of the image do not satisfy the dark channel prior. As a result, it may suffer from one or more of the following problems. First, it usually results in color oversaturation because it cannot agree with the limit requirement that the obtained dark channel should be no brighter than the minimum color channel. Second, coarse transmission estimation with a larger patch can be used to prevent the white-object problem and obtain more correct transmission information, but it simultaneously loses details and produces blocky and contouring effects on edge discontinuities. In contrast, when the patch size is small, the filtering result retains more rich details. Nevertheless, the detail scale contains more incorrect transmission information. Finally, the restored haze-free image usually looks dark after haze removal.

B. Geometric Bounds on Transmission

Obviously, He's algorithm encounters the problems described above due to the assumption of a local dark channel prior. To solve these problems, we make assumptions on two basic geometric properties in the haze model to obtain the optimal parameter $t(x)$:

1) The scene radiance $J^c(x)$ is nonnegative and it is related to the neighboring pixels.

2) Pixels in a local patch have similar depth values and consistent transmission maps.

From (1), $I^c(x)$ can be represented as a linear combination of $J^c(x)$ and A^c. Thus, we provide a geometric perspective that could explain the optical model properly, as illustrated in Fig. 2.

To view the whole image, the scene radiance $J^c(x)$ is a collection of all the clean pixels, and it is pushed towards the global atmospheric light A^c by the fog and haze. As a result, $J^c(x_i)$, $I^c(x_i)$ and A^c can be seen as having a linear relationship on each black dotted line. From the local patch perspective and following the above assumptions, all of the pixels in a local patch Ω share the same transmission map $t(x)$, as illustrated by the points on the blue and dotted blue lines in Fig. 2. Hence given any two points $J(x_1)$ and $J(x_2)$ in a local patch, must be located at the same depth and transmission. Moreover, A^c is a global constant [14], so we can obtain the relationship using

$$\begin{cases} I^c(x_1) = J^c(x_1)t(x) + A^c(1 - t(x)) \\ I^c(x_2) = J^c(x_2)t(x) + A^c(1 - t(x)) \end{cases} \tag{5}$$

From (5), we can deduce the transmission map as (6), while the inverse of the expression used in [14] is given by (7). By comparing the two different methods, we can conclude that the blue line is parallel to the dotted blue line from a geometric point of view.

$$t(x) = \frac{I^c(x_1) - I^c(x_2)}{J^c(x_1) - J^c(x_2)} \tag{6}$$

$$t(x) = \frac{A^c - I^c(x)}{A^c - J^c(x)} \tag{7}$$

More importantly, from Fig. 2, the radiance $J^c(x)$ is variable and bounded as $z_0 \leq J^c(x) \leq z_1$ in a local patch, where z_0 and z_1 are two fixed values that are relevant to the given local patch. In particular, when $J^c \to z_0 = 0$, (7) can be used directly to compute the transmission map by assuming the dark channel of $J^c(x)$ to be zero. However, it ignores some important characteristics such as local relativity. Compared to (7), it is worth noting that the additional boundary constraint of $J^c(x)$ is more influential in (6). In many cases, the local boundary constraint can reflect the relationship between neighboring pixels. More specifically, for any x, a general requirement is that the extrapolation of $J^c(x)$ must be located in the radiance range bounded by the radiances of neighboring pixels The a above requirements on $J^c(x)$ lead us to assume that the two functions $J^c(x)$ and $I^c(x)$ are statistically correlated on the adjacent pixels in each channel.

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

Fig. 3. Deriving de-hazed images under different transmission maps. (a) (d) Transmission/haze-free image derived by Dark Channel prior, (b) (e) Transmission/haze-free image derived by Guided filter and (c) (f) Transmission/haze-free image derived by our method.

Thus, a more efficient bound on $t(x)$ is imposed, which can be determined by (8).

$$0 \leq \hat{t}(x) \leq t(x) \leq 1 \tag{8}$$

where $\hat{t}(x) = \min\left(I^c(x_1) - I^c(x_2)/J^c(x_1) - J^c(x_2), 1\right)$. Now our goal is to approximate the transmission $\hat{t}(x)$. Rewriting (6), we have $\hat{t}(x) = \min(p(x)q(x), 1)$, where $p(x) = 1/(J^c(x_1) - J^c(x_2))$ and $q(x) = I^c(x_1) - I^c(x_2) \forall x_1, x_2 \in \Omega(x)$. For convenience of calculation, we assume that the transmission map is the same in each channel. Simultaneously, in order to better characterize the relationship in a local patch, we derive

$$J(\hat{t}(x)) = \sum_x \sum_{y \in \Omega(x)} \left(\left\| \hat{t}(y) - q(y)p(y) \right\|^2 + (\varepsilon p(y))^2 \right) \tag{9}$$

where ε is a regularization parameter and $\hat{t}(y)$, $q(y)$ denote the estimated transmission and neighboring relativity over a small window around each pixel x, respectively. Moreover, we assume $p(y)$ is locally consistent and $p(y) = p_x \forall y \in \Omega(x)$. Consequently, (9) can also be expressed as

$$\begin{aligned} J(\hat{t}(x)) &= \sum_x \sum_{y \in \Omega(x)} \left(\left\| \hat{t}(y) - q(y)p_x \right\|^2 + (\varepsilon p_x)^2 \right) \\ &= \sum_x \left\| \hat{\mathbf{t}}_x - \mathbf{q}_x^T \mathbf{p}_x \right\|^2 \end{aligned} \tag{10}$$

In addition, $\mathbf{q}_x$ and $\hat{\mathbf{t}}_x$ are vectors of size $(|\Omega(x)| + 1) \times 1$, with their last elements respectively equal to epsilon and 0. To solve (10), we assume that every patch $\hat{\mathbf{t}}_x$ is given and derive the optimal $\mathbf{p}_x$ by solving a least square problem [29]:

$$\hat{\mathbf{p}}_x = \arg\min \left\| \hat{\mathbf{t}}_x - \mathbf{q}_x^T \mathbf{p}_x \right\|^2 = (\mathbf{q}_x^T \mathbf{q}_x)^{-1} \mathbf{q}_x^T \hat{\mathbf{t}}_x \tag{11}$$

After substituting (11) into (10), the objective function only depends on the scene transmission. Let $\mathbf{\Lambda} = \mathbf{E} - \mathbf{q}_x(\mathbf{q}_x^T \mathbf{q}_x)^{-1} \mathbf{q}_x^T$, and then we have $J(\mathbf{t}) = \sum_x \hat{\mathbf{t}}_x^T \mathbf{\Lambda}^T \mathbf{\Lambda} \hat{\mathbf{t}}_x$. In addition, in order to avoid crossing the boundary, we use the minimum to limit the range:

$$\hat{\mathbf{t}} = \min(\sum_x \hat{\mathbf{t}}_x^T \mathbf{\Lambda}^T \mathbf{\Lambda} \hat{\mathbf{t}}_x, \mathrm{T}) \tag{12}$$

where $\mathbf{E}$ is a unit matrix and $\mathbf{T}$ is a matrix of ones. An illustration of the transmission estimation is shown in Fig. 3. It displays an example sharing a similar depth in a local patch. The derived input suffers from low visibility mainly in regions with dense haze, loss of contrast, and vividness conditions. The idea of a replacement technique was suggested previously by He in [14], where he considered that the center pixel can be replaced by the minimum value of neighboring pixels. This is because the low-intensity pixels existed throughout most of the local region, which is described in Fig. 3(a). However, halo artifacts can be clearly seen in the recovered result in Fig. 3(d). To overcome this limitation, another refined map, which is designed by guided filter [15], can be used to obtain better results in degraded regions as shown in Fig. 3(e). However, the more elaborate the transmission map is, the less accurate the dehazed image will be; for example, consider the whole brick wall in Fig. 3(e). This effect occurs because the whole brick wall has the same scene depth and transmission map, but the elaborate estimation could reduce the contrast in the dehazed image. Specifically, He *et al.* [15] believed that the pixel could be represented by the surrounding pixels. However, he ignored the constraints of local consistency. Therefore, determining how to balance the difference and local consistency of the transmission map becomes a key point for solving this problem. By referring to (12), we obtain the initial scene transmission in Fig. 3(c) and Fig. 3(f).

It is evident that (12) could suppress small variations through adjacent pixels. However, we may still obtain an inconsistent scene transmission for objects with large structural regions. To overcome this limitation and get a better transmission map, we follow the idea described in Li *et al.* [30] of minimizing the cost function to obtain the refined $t(x)$.

$$\inf_{\mathbf{t} \in R} E_{\mathbf{t}} = u \left\| \hat{\mathbf{t}} - \mathbf{t} \right\|^2 + \Re(\mathbf{t}) \tag{13}$$

where the first term is the data-fidelity term, which represents the fidelity between the observed degraded images and the restored image, $\Re(\mathbf{t})$ is the regularization, which gives a prior model of the image, and u is the regularization parameter, which controls the trade-off between the data-fidelity and prior items. Normally, images consist of various patterns and it is difficult to describe an image by one distribution. Fortunately, structural similarities are prevalent in natural images, which will be fully exploited to build the regularization $\Re(\mathbf{t})$.

III. Difference-Structure-Preservation Prior

Following the discussion above, determining how to investigate the regularization that can accurately reflect the local consistency becomes a critical topic in our solution. First, we provide a method to measure the local depth consistency. In addition, to use the locally consistent depth as a regularizer,

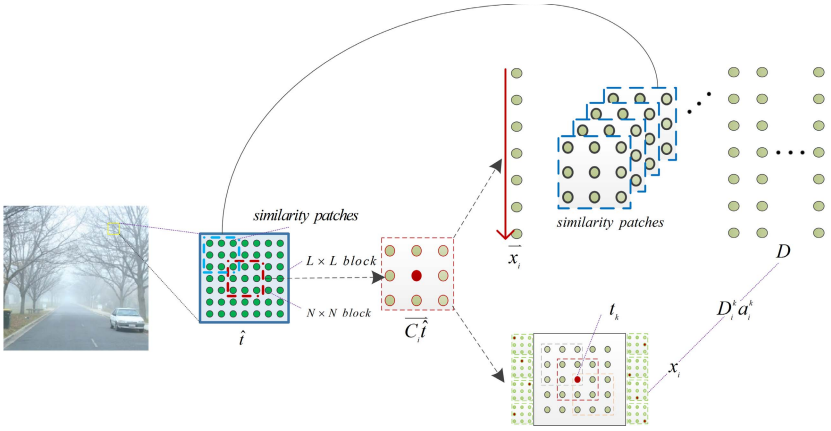


Fig. 4. The process of generating the training samples & Illustration of overlap patches.

we assume that a local patch can be approximated by a sparse linear combination of elements from a neighbor basis set. More importantly, through building the difference-structure - preservation dictionary, the results have demonstrated that our approach is effective at restoring images. To this end, we recast the problem as finding a solution to transmission under scene constraints in the feasible range. Through the selection of parameters, an optimal transmission can be proposed and the complete calculation process is summarized.

A. Spatially Consistent Depth

As in previous literature [12], the depth edge consistency is decided using neighboring pixels. Hence, whether pixels lie on the same depth is decided only by the locations of the pixels. Following this idea, we model a pixel and its nearest neighbors as a vector variable to measure the depth information. We call this principle the local depth consistency. To determine whether there are other pixels at the same depth, we need a set of neighboring training sample patches so that the transformation matrix can be calculated. For this purpose and following the LPG-PCA [27], we set a $N \times N$ window ω_i centered at the pixel i in the initial transmission map $\hat{\mathbf{t}}$. In addition, let $\mathbf{C}_i$ be the matrix of ones, and we assume that $\mathbf{C}_i\hat{\mathbf{t}} = \left[\hat{t}_1, \hat{t}_2 \ldots \hat{t}_s\right]$, where s is the number of pixels in the image. The vector contains all the components within the $N \times N$ window. Then, use a $L \times L(L > N)$ training block to find the training samples, as shown in Fig. 4. The simplest and most efficient way is to replace $\mathbf{C}_i\hat{\mathbf{t}}$ with the surrounding training samples in the $L \times L(L > N)$ training block. We denote $\mathbf{x}_i = \mathbf{C}_i\hat{\mathbf{t}}$ as the column sample vector containing the pixels in the red $N \times N$ block and denote $\mathbf{x}_j$, where $j = 1, 2 \ldots (L-N+1)^2 - 1$, as the sample vectors corresponding to the neighboring blocks shown in blue in Fig. 4. In this way, there is a matrix $\mathbf{X}_{Neighbour_i} = \left[\mathbf{x}_1, \ldots \mathbf{x}_j, \ldots \mathbf{x}_{(L-N+1)^2}\right]$ for each initial transmission patch $\mathbf{x}_i$. Although there are many neighboring patches, it may lead to inaccurate estimation vector $\mathbf{x}_i$ and then produce misty results with high residual fog because sub-patches may not have the same depth. If these patches have the same structure, they may be successful in preserving a consistent transmission over the same object and can better reflect the structure of the image. Thus, we adopt the cosine similarity method to correct for the above mistake. It can be easily calculated that

$$sim(\mathbf{x}_i, \mathbf{x}_j) = \cos\theta = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \cdot \|\mathbf{x}_j\|} \tag{14}$$

where $j = 1, 2 \ldots (L-N+1)^2 - 1$. If $\cos\theta < \phi = 0.2$, we select it as a similar sample vector under (14). For convenience, we select the most similar sample vectors of $\mathbf{x}_j$ and arrange them as a matrix $\mathbf{X}_{Neighbour_i} = \left[\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{m-1}\right]$, where m is the number of the most similar sample vectors. Therefore, we consider these similar patches with local depth consistency as a regular term.

Following the principle of BM3D [26], which implies that similar patches have similar sparse representations, let $\mathbf{a}$ be dictionary atoms representing $\mathbf{X}_{Neighbour_i} = \left[\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{m-1}\right]$ instead of averaging them with the similarity measure as the weights

$$\begin{aligned}(\mathbf{D}_i, \mathbf{a}_i) &= \arg\min \left\| \mathbf{C}_i\hat{\mathbf{t}} - \mathbf{D}_i\mathbf{a}_i \right\|^2 + \lambda \|\mathbf{a}_i\|_0 \\ &= \arg\min \|\mathbf{x}_i - \mathbf{D}_i\mathbf{a}_i\|^2 + \lambda \|\mathbf{a}_i\|_0 \end{aligned} \tag{15}$$

where $\mathbf{x}_i$ denotes the i^{th} patch of size $N \times N$, and $\mathbf{a}_i = [\mathbf{a}_0, \mathbf{a}_1 \cdots \mathbf{a}_{m-1}]^T$ is the vector of codes with respect to dictionary $\mathbf{D}_i$. Let λ be the regularization parameter and $j = 0, 1, \ldots m-1$. For a given patch, this procedure implicitly assumes that all patches are independent of one another, which is questionable since they overlap, as described in Fig. 4. However, this approximation makes the corresponding optimization tractable. Once the dictionary $\mathbf{D}_i$ and codes $\mathbf{a}_i$ have been learned, every pixel of the transmission map $N \times N$ is estimated, so its value can be computed by averaging:

$$t_k = \frac{1}{N^2} \sum_{i|k\in\omega_i} \mathbf{D}_i^k \mathbf{a}_i^k \tag{16}$$

where k is a pixel in the transmission map, $\mathbf{D}_i$ and $\mathbf{a}_i$ represent the corresponding dictionary and its coefficients for each patch overlap, respectively.

B. Difference-Structure-Preservation Dictionary

Obviously, $\mathbf{D}$ is an important indicator, which can be regarded as the dictionary. However, based on theoretical

analysis and experimental observation, most dictionaries [23]–[26] relying on the Euclidean distance would fail to measure the essential structure embedded in an original high-dimensional data space since an image is embedded into a high-dimensional space as a manifold from the mathematic aspect. More importantly, spatial geometric structure can be destroyed among nearby data by global Euclidean structure so that the local consistency may be weakened. An alternative approach is to use the manifold learning method, which employs an adjacency matrix to model the spatial relationship of data points and efficiently preserves the local structure of the data. Unfortunately, this kind of method can maintain the similar information but neglect the information differences, which also play a critical role in maintaining the information structure. In this paper, a novel method called difference-structure-preservation is proposed, and it uses the similar structures as fully as possible and maintains the differences between similar patches as much as possible. Based on the discussion above, the following points should be considered:

1) From a statistical point of view, if the two samples are very close or similar, i.e., the distance between them is small or the similarity is high, they provide little information; on the contrary, the difference information should be great.

2) On the premise of similarity, as far as it is possible to reflect the different state of similar samples, these samples provide many information models. In addition, since they are very sparse, the different samples can reflect more differences of information, and statistical attributes can also be estimated through these samples.

3) We can obtain patches from L nearest neighbors to measure the difference information through Euclidean distance, while patches located outside of the neighborhood can be ignored.

According to these assumptions and in order to maintain the differences of the original data and express the local spatial relationships, the elements of the coefficient matrix $\mathbf{b}_{ij}$ could measure the distinctive information conveyed by $\mathbf{x}_i$ and $\mathbf{x}_j$. For ease of description, it is defined as

$$\max \sum_{ij} (a_i - a_j)^2 b_{ij} \tag{17}$$

where $\mathbf{a}_i$ and $\mathbf{a}_j$ are the weights of the original similarity patches $\mathbf{x}_i$ and $\mathbf{x}_j$ on the specific dictionary, and $\mathbf{b}_{ij}$ is the divergence between the patches in high-dimensional space

$$\mathbf{b}_{ij} = \begin{cases} \exp(-\frac{\tau}{\|\mathbf{x}_i - \mathbf{x}_j\|^2}), & \mathbf{x}_j \in \mathbf{X}_{x_i} \\ 0 & others \end{cases} \tag{18}$$

here $\mathbf{X}_{x_i}$ denotes similar neighborhood patches around $\mathbf{x}_i$, $\tau = 0.01$. For instance, the symmetric weights $\mathbf{b}_{ij}$ impose a heavy penalty if two similar patches are far apart in the original image space but are mapped very close to each other. In that case, the function in (17) is maximized to guarantee the diversity when the data are represented by the corresponding dictionary. In other words, the step seeks to find the lower-dimensional compact subspace that efficiently preserves the diversity among nearby data points. Similarly, we note the function that can retain the hidden related information:

$$\min \sum_i (\mathbf{a}_i - \sum_j \mathbf{w}_{ij} \mathbf{a}_j)^2 \tag{19}$$

where $\mathbf{w}_{ij}$ is an affinity to measure the similarity between the different patches with $\sum_j \mathbf{w}_{ij} = 1 \forall i$. The particular parameter depends on the normalized results of the cosine similarity method. For instance, if two points are very close to each other in the observed initial transmission map, $\mathbf{w}_{ij}$ would impose a heavy penalty. Obeying the above assumptions, we define a constraint: each data patch $\mathbf{x}_i$ is connected only with its neighbors, so $\mathbf{w}_{ij} = 0$ if $\mathbf{x}_j$ does not belong to the set of neighbors of $\mathbf{x}_i$. Therefore, minimization could ensure much closer results on the dictionary basis as well as in the original space.

Because the purpose is to search for the optimal mapping direction that can efficiently retain the local similar structures as well as the difference information, we construct a desirable projection using the following criterion:

$$\mathbf{D} = \arg\max \frac{\sum_{ij} (\mathbf{a}_i - \mathbf{a}_j)^2 \mathbf{b}_{ij}}{\sum_i (\mathbf{a}_i - \sum_j \mathbf{w}_{ij} \mathbf{a}_j)^2} \tag{20}$$

We assume that $\mathbf{D}_i^{-1} = \mathbf{D}_i^T$, and then substitute $\mathbf{a}_i = \mathbf{D}_i^T \mathbf{x}_i$ and $\mathbf{a}_j = \mathbf{D}_i^T \mathbf{x}_j$ into the denominator of (20):

$$\begin{aligned} J(\mathbf{D}) &= \max \frac{\sum_{ij} (\mathbf{a}_i - \mathbf{a}_j)^2 \mathbf{b}_{ij}}{\sum_i (\mathbf{a}_i - \sum_j \mathbf{w}_{ij} \mathbf{a}_j)^2} \\ &= \frac{\sum_{ij} (\mathbf{D}_i^T \mathbf{x}_i - \mathbf{D}_i^T \mathbf{x}_j)^2 \mathbf{b}_{ij}}{\sum_i \mathbf{D}_i^T \mathbf{x}_i (\mathbf{E} - \sum_j \mathbf{w}_{ij})^T (\mathbf{E} - \sum_j \mathbf{w}_{ij}) \sum_i \mathbf{x}_i^T \mathbf{D}_i} \\ &= \frac{2\mathbf{D}^T (\sum_i \mathbf{x}_i \mathbf{h}_{ii} \mathbf{x}_i^T - \sum_{ij} \mathbf{x}_i \mathbf{b}_{ij} \mathbf{x}_j^T) \mathbf{D}}{\mathbf{D}^T \sum_i \mathbf{x}_i (\mathbf{E} - \mathbf{W})(\mathbf{E} - \mathbf{W})^T \sum_i \mathbf{x}_i{}^T \mathbf{D}} \\ &= \frac{2\mathbf{D}^T (\mathbf{X}\mathbf{H}\mathbf{X}^T - \mathbf{X}\mathbf{B}\mathbf{X}^T)\mathbf{D}}{\mathbf{D}^T \mathbf{X}\mathbf{M}\mathbf{X}^T \mathbf{D}} = \frac{2\mathbf{D}^T \mathbf{X}\mathbf{L}_h \mathbf{X}^T \mathbf{D}}{\mathbf{D}^T \mathbf{X}\mathbf{M}\mathbf{X}^T \mathbf{D}} \end{aligned} \tag{21}$$

where $\mathbf{E}$ is m order identity matrix and $\mathbf{X}$ is all the blocks with the pixel i as the center. Note that $\mathbf{L}_h = \mathbf{H} - \mathbf{B}$ is called the difference-structure-preservation matrix, $\mathbf{D}_{ii}$ is a diagonal matrix and $\mathbf{h}_{ii} = \sum_j \mathbf{b}_{ij}$, $\mathbf{M} = (\mathbf{E} - \mathbf{W})(\mathbf{E} - \mathbf{W})^T$ is called the Laplacian similarity matrix. Using the Lagrange multiplier for the solution, we have

$$\mathbf{X}\mathbf{L}_h \mathbf{X}^T \mathbf{D} = \delta \mathbf{X}\mathbf{M}\mathbf{X}^T \mathbf{D} \tag{22}$$

Then, using the Cholesky decomposition for $\mathbf{XMX}$, we obtain $\mathbf{XMX} = \mathbf{G}\mathbf{G}^T$, where $\mathbf{G}$ is a lower triangular matrix. Denoting $\mathbf{z} = \mathbf{G}^T \mathbf{D}$, we have$\mathbf{D} = (\mathbf{G}^{-1})^T \mathbf{z}$, which we substitute into the denominator of (21), where$\mathbf{S} = \mathbf{G}^{-1} \mathbf{X}\mathbf{L}_h \mathbf{X}^T (\mathbf{G}^{-1})^T$ is the symmetric matrix. In this way, we can easily obtain the solution.

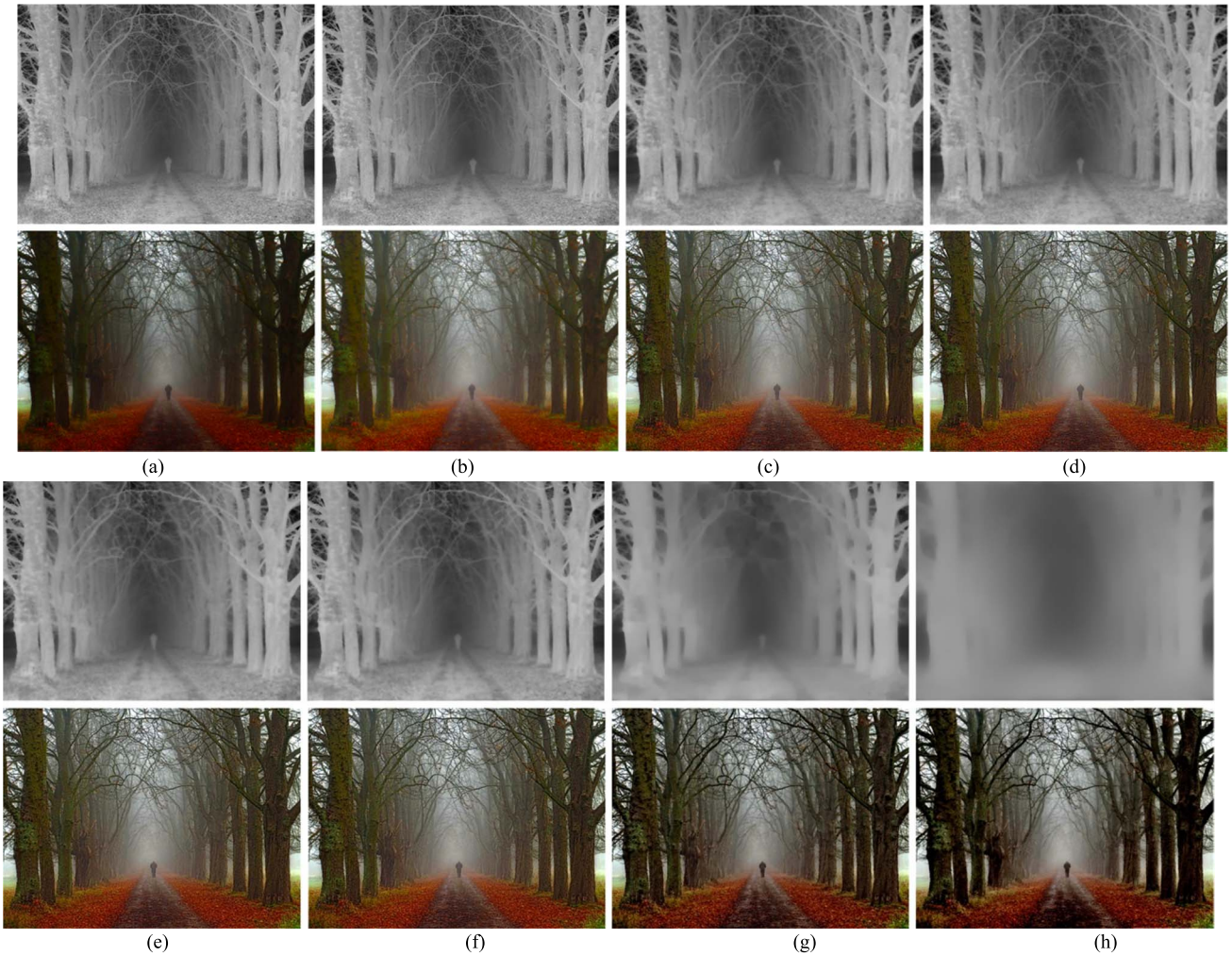


Fig. 5. The effect of the parameters, (a):initial transmission and results (b) $\lambda = 2 \times 10^{-3}$; (c)$\lambda = 2 \times 10^{-1}$; (d) $\lambda = 2 \times 10^{1}$ (e) $\mu = 3000$; (f) $\mu = 300$ (g) $\mu = 30$; (h) $\mu = 3$. The smaller the value λ is, the restored structure is closer to the input image. More importantly, some branches become under-exposed if the value of μ becomes larger. However, higher value can lead to the color drift. In contrast, darker radiance is caused by lower value such as roads.

C. The Step for DSPV Model

The fact that semantically meaningful image constructs are formed by spatially coherent contiguous patches suggests that the image is piecewise stationary. We propose a difference-structure-preservation variation model, which follows (13) and (15).

$$\begin{aligned}\inf_{t\in R} E_t &= \arg\min \mu \left\|\hat{\mathbf{t}}-\mathbf{t}\right\|^2 + \sum_i \|\mathbf{C_i t} - \mathbf{D_i a_i}\|^2 + \sum_i \lambda \|\mathbf{a_i}\|_0 \\ &= \arg\min \mu \left\|\hat{\mathbf{t}} - \mathbf{t}\right\|^2 + \left\|\hat{\mathbf{t}} - \frac{1}{N^2} \sum_{i|k\in\omega_i} \mathbf{D}_i^k \mathbf{a}_i^k\right\| \end{aligned} \tag{23}$$

The first term of (23) forces the proximity between the haze observation and the real scene image, and the parameter μ specifies the required degree of proximity, as illustrated in Fig. 5. We can see that the fidelity term function is close to the original value when μ is large. As such, the effect of the control is not significant and restored image changes are smaller than those of the hazed image when the iteration step is constant. This is also observed in our experiment with images in Fig. 5. It seems more appropriate to set $\mu = 30$. The second term integrates local similarity priors to form a regularizing force. Evidently, the proposed approach is closely related to K-LLD, K-SVD, and BM3D. However, it can be considered their extension under the framework of the local depth consistency assumption. Moreover, we choose a value that is not too high for the parameter $\lambda = 2 \times 10^{-1}$ to control sparse codes. Compared with the abovementioned algorithms, the most significant difference is the definition of a difference-structure-preservation dictionary, which is an adaptive similarity data matrix for indicating local characterization. In another aspect, it is also more consistent with the laws of physics, which suggests that the pixels have identical depth and will be degraded to the same degree.

As we mentioned in Section III, there are four steps to obtain the haze-free image, which are described in Fig. 6. First, the problem can be divided into the transmission map and atmospheric light issue. Similar to He's algorithm [14], we have adopted two steps to approximate the real transmission map. In step 2, we recast the initial transmission map under an additional boundary prior, which is described as (12). To produce more details for the results, following the variation framework [30], (13) can be proposed for refining

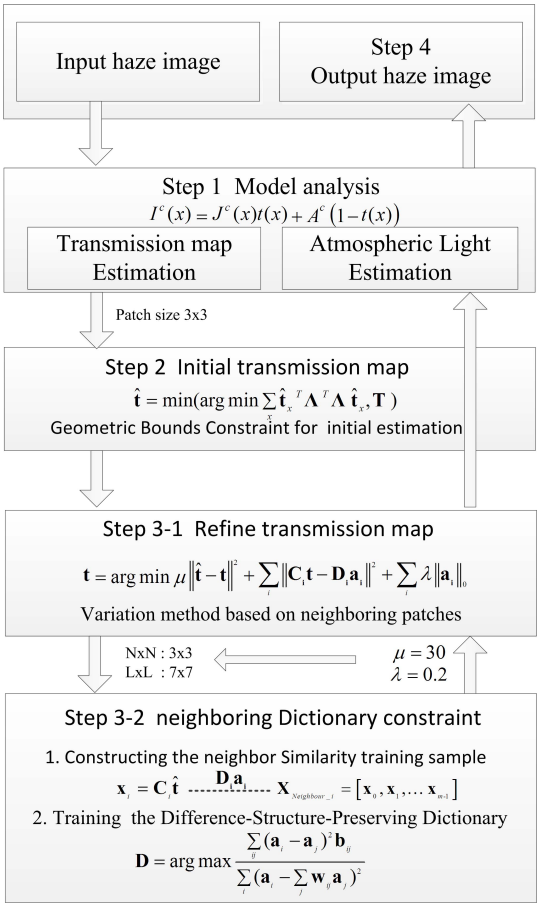


Fig. 6. DSPV flowchart.

the coarse transmission map in step 3-1. Hence, in 3-2, we draw on the idea of LPG-PCA [27] so that it can search for the local similarity patches as the constraint to express the small patch $\mathbf{C}_i\hat{\mathbf{t}} = \left[\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2 \ldots \hat{\mathbf{t}}_m\right]$. In addition, to guarantee veracity, the most similar neighboring patches $\mathbf{X}_{Neighbour_i} = \left[\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{m-1}\right]$ are selected to characterize i-centered patches by the cosine similarity method. More importantly, following the principle of BM3D [26], we organize the adaptive domain selection strategy to code each patch, which is called the difference-structure -preservation dictionary. Based on the initial dictionary in (21), using the cross-interactive method and setting the time to 5, we can obtain dictionary $\mathbf{D}_i$ and codes $\mathbf{a}_i$. It should be noted that this approach learns the dictionary on the set of overlapping patches so that sparse image models can handle such situations by exploiting the redundancy between overlapping patches. Finally, under the atmospheric light proposed by He *et al.* [14], an optimal transmission can be expressed as (23), and (4) is used for dehazing the image. In summary, the proposed method not only maintains the local consistency but also reflects the difference of each pixel.

IV. Visual Evaluation on Real Images

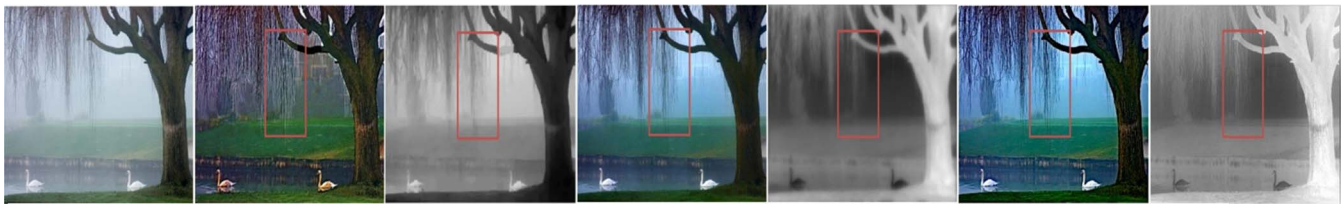
Assessing restored images is a very difficult task because real-world, haze-free reference images have not been validated for quantification of restored images. To demonstrate the effectiveness of our method, some real images of outdoor scenes and synthetic hazy images are tested. In particular, these synthetic images are taken from two Foggy Road Image Databases, which include 84 uniform-haze images [31], [32]. Our method outperforms the state of the art method from both quantitative and qualitative points of view. Our experiments were conducted on benchmark fog images collected from a variety of sources; for example: http://www.cs.huji.ac.il/∼raananf/projects/dehaze_cl/results/, http://perso.lcpc.fr/tarel.jean-philippe/.

A. Comparison for Different Haze Removal Algorithms

The method is compared with that of Tan [9] and DCP [14] in Fig. 7. As shown by the results, halo artifacts still exist in Tan's method, for instance, in the area between the tree trunks, which is been shown in the red rectangular box. Conversely, our method can remove halo artifacts effectively and retain more details in the tree trunks due to similarity in the local window for estimating the transmission map. In addition, the image dehazed by Tan appears over-saturated, which is caused by the underestimation of the transmission map. In comparison, our method can recover and preserve the structures without sacrificing the fidelity of the colors, such as with the swans in this image. More importantly, even though our result seems mistier than that obtained using Tan's method, our result seems visually more reliable in distant regions. Meanwhile, our algorithm can provide more detail information than DCP in the whole image and looks more pleasing. Most subjective dehazing methods are aimed at increasing contrast, but this may cause non-uniform enhancement and some regions to be too bright. They are very prone to color distortion or over-enhancement, as shown in Fig. 8. CLAHE and MSRCR do not consider the degeneration factor; they only focus on enhancing the image visibility. Even though the results of these methods have higher contrast, color distortion and uneven distribution exist in some areas, such as the sky and the road.

Fig. 9 shows the comparisons with Nishno [7], Fattal [11], Fattal [13], Gibson *et al.* [33], Yeh *et al.* [34] and Kratz and Nishino [35]. Their methods achieve comparable results in haze-free images. For instance, Gibson *et al.* [33] consider color ellipsoids that can be tied to depth cues within an image. Nishino utilizes FMRF fields to estimate the albedo scene. Yeh *et al.* [34] continue DCP theory and estimates haze density. Fattal [11], [13] created a raw estimation by using the statistically uncorrelated relationship, and he took advantage of MRF to acquire haze-free images in 2014. However, our transmission map contains more local details, so it truly reflects the underlying scene content by choosing the local similarity. As shown in Fig. 9, the results demonstrate that our dehazing method has outperformed the other methods in terms of visual contrast. This success comes from the accurate estimation of our transmission map. In contrast, plenty of algorithms depend on global priors or certain distributions, which may produce dense haze regions that will make them unreliable.

As seen in Fig. 10, our operator is able to yield comparable and even better results against the conventional

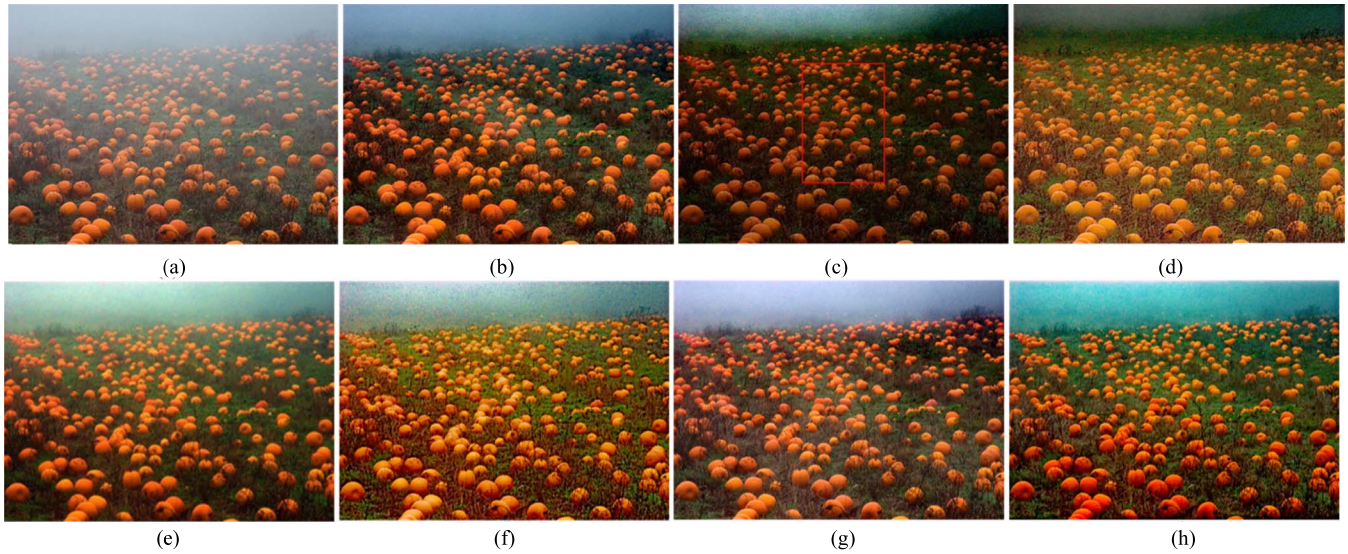


(a) (b) (c) (d) (e) (f) (g)

Fig. 7. Comparison of the proposed algorithm and other haze removal algorithms. (b) (c) Tan result, (d) (e) DCP by guided filter, and (f) (g) proposed method.

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

Fig. 8. Experimental result comparison. (a) Original image, (b) MSRCR result, (c) CLAHE result, (d) Our method.



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

(e) (f) (g) (h)

Fig. 9. Haze removal results for pumpkin image. (a) Input image; (b) result by [11]; (c) result by [35]; (d) result by [7]; (e) result by [34]; (f) result by [33]; (g) result by [13]; (h) result by the proposed algorithm.

dehazing techniques. Tan's method [9] generates many saturated pixels in Fig. 10(c) since it simply maximizes the contrast. The method of Caraffa and Tarel [10] is computationally less complicated, but it changes color tones and exhibits halo artifacts, as shown in Fig. 10(f). The method of Zhu *et al.* [8] only considers the darkest pixel value for dehazing, and it thus removes the shadow of the cloud as in Fig. 10(d). Although Nishino's method [7] can enhance the visibility effectively, it increases the noise, since it mainly relies on MRF formulation rather than fitting the haze model. The main novel contribution of Kopf *et al.* [3] consists of using a rough 3D map approximation of the scene that improves the quality of the scene. However, the limitation of this method is very small due to un-acquirement of the 3D map. The method of Fattal [13] yields more natural results, but it cannot sufficiently remove haze in some regions, for example, the buildings around the horizon and the clouds that are marked with yellow rectangle in Fig. 10(h). In contrast, our defog results achieve not only vivid colors and better perception but also the best objective measurements. This demonstrates the success in terms of depth accuracy and the effectiveness of constraining both the smoothness and construct preservation. Therefore, the proposed algorithm can suppress most of the artifacts that occur in the conventional dehazing algorithms.

We have seen that, visually, the proposed approach compared favorably with respect to the state of the art. To obtain a more objective evaluation, we propose to test the algorithms on some synthetic reference images that were manually selected from the FRIDA database [31], [32].

In Fig. 11, the four algorithms, including the proposed algorithm, are tested on synthetic images for which the ground-truth images are known. It is obvious that Tarel's

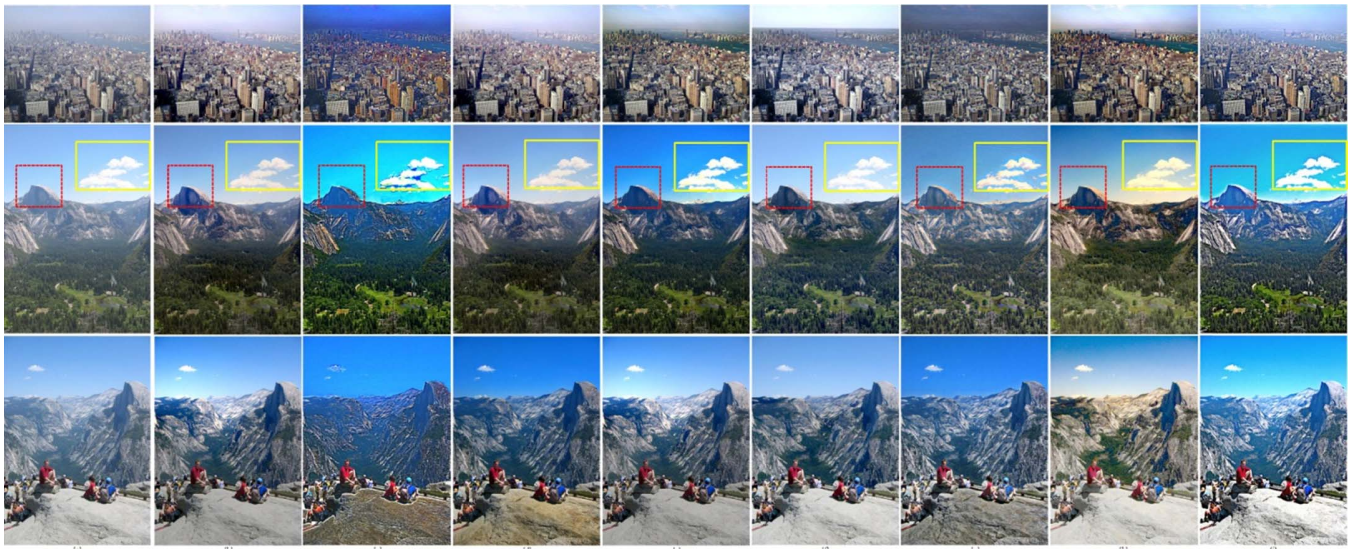


Fig. 10. Comparison of the mainly dehazing methods. Besides the initial hazy images (a) in figures are displayed, the results of (b)-(i) are obtained by Kopf *et al.* [3], Nishno [7], Tan [9], Fattal [11], He *et al.* [14], Caraffa and Tarel [10], Fattal [13] and proposed algorithm.

Fig. 11. Results on stereo images where the ground truth solutions are known. (a) The hazy images. (b) Tarel results (2009). (c) Tarel results (2010). (d) He results. (e) Our results. (f) Ground truth.

results are quite different from the ground-truth images, as the results are much brighter because the fog is not completely removed, such as in the backgrounds in Fig. 11(b) and Fig. 11(c). He's results are more similar to the ground-truth images but still show some inaccuracies. Note that the trees and buildings are still surrounded by fog in Fig. 11(d).

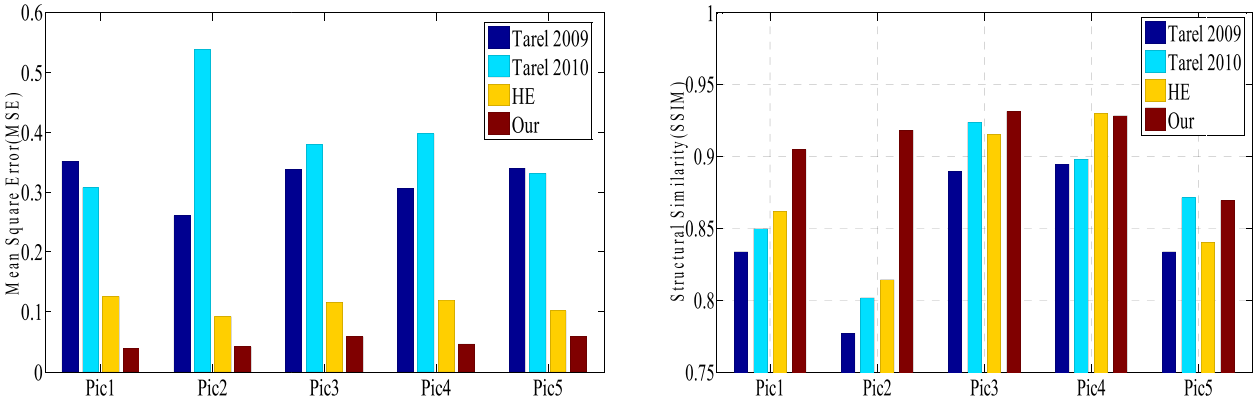


Fig. 12. Mean squared error (MSE) and Structural similarity (SSIM)of different algorithms.

TABLE I
QUALITATIVE COMPARISON FOR FIG. 10

		Fattal 08	Tan	HE	Kopf	Tarel	Nishno	Fattal 14	Our
Building	e	-0.06	-0.14	**0.06**	0.05	0.07	-0.01	**0.04**	**0.04**
	Σ	0.09	0.02	0.00	**0.00**	0.00	0.46	0.00	0.01
	r	1.32	**2.34**	1.42	1.42	1.88	1.81	1.86	1.73
Landscape	e	0.04	0.08	0.08	**0.09**	0.02	0.11	0.03	0.05
	Σ	0.02	0.01	0.01	**0.00**	0.00	0.71	0.31	0.15
	r	1.23	**2.28**	1.33	1.62	2.09	1.79	1.83	1.82
Person	e	0.05	-0.03	0.03	0.05	0.04	**0.12**	0.10	**0.11**
	Σ	0.03	0.02	0.01	**0.00**	0.01	0.41	0.25	0.27
	r	1.15	**2.32**	1.52	1.52	2.22	2.19	1.93	2.04

In contrast, our results can remove haze in the distance as well as maintain the original structure of the objects (see Fig. 11(e)).

B. Quantitative Comparison

Basically, there are two major categories of quantitative metrics: non-reference methods and reference methods. For this reason, we adopted well-known quantitative metrics to evaluate the restoration efficacy of each compared method for images captured in real-world scenarios. In Table 1, following the comparison experiments of Tarel and Hauti [36], we considered three the images that are described in Fig. 10, where the e indicator represents edges that are newly visible after restoration, r represents the mean ratio of the gradients at visible edges and Σ represents the percentage of pixels that become completely black or completely white after restoration.

Analyzing the results of Table I, in general, all of the demonstrated techniques have small values of the descriptor Σ. In addition, e shows that most of the algorithms, depending on the processed image, remove some of the visible edges. Nevertheless, the methods of Nishno [7], Tan [9], and Fattal [11] are characterized by negative values of e for some of the considered images.

Moreover, indicator r produces proper values for our results. One may observe that the appearance in our case is locally restored and the value of indicator r is kept close to the maximum value, while the local contrast was restored moderately. This feature is achieved also by the methods of Fattal [13], He *et al.* [14] and Kopf *et al.* [3]. The techniques of Tan [9] and Caraffa and Tarel [10] increase the local contrast too strongly, and as a result, these approaches have higher values of indicator r, which indicates the presence of spurious edges and artifacts. To conclude, the methods of Kopf *et al.* [3], Fattal [13], and He *et al.* [14], along with our method, obtained better results with respect to these indicators. However, the method by Kopf *et al.* [3] requires a semi-automatically aligned geo-referenced 3D terrain model to obtain the depth of the scene, which cannot be widely used in all scenarios. In Fattal's method [13], the application domain is more restrictive and was not applicable in some of the experiments.

To quantitatively assess and rate the algorithms, we used another assessment method [8] as a comparison index that is based on the mean-square error (MSE) and the structural similarity (SSIM) of the results in Fig. 12.

As can be seen, the results of Tarel's method produce the highest MSEs overall, which are mainly due to the large number of residual fogs in the dehazed images. The scores of the results of He's [14] method are better and are the second smallest out of all the approaches. Our method achieves the lowest MSEs in all cases.

The structural similarity (SSIM) image quality assessment index is introduced to evaluate the abilities of the compared algorithms to preserve the structural information. A high SSIM represents high similarity between the dehazed image and the ground-truth image, while a low SSIM conveys the opposite meaning. Most of Tarel's results are lower than the others, indicating that much structural information in the images has been lost. In general, the SSIMs of He *et al.* [14] are much higher in these images, but sometimes the results are not very stable and are occasionally even lower than Tarel's. Our results achieve the highest SSIMs due to focus on the structural consistency throughout the dynamic difference-structure-preservation process.

V. CONCLUSION

This paper proposed a dehazing algorithm based on the difference-structure-preservation prior, which can estimate

the optimal transmission map and restore the actual scene. To obtain the rough transmission map, we use two basic properties in the haze model to resolve the optimal parameter at the same depth. Afterwards, we assume that an image patch can be approximated by a sparse linear combination of elements from a neighbor basis set to obtain a more accurate transmission map that can better preserve the structures of images. Experimental test results were also used to verify that the method effectively achieves accurate and true representation. In the future, this method will be studied in global air-light to improve operational efficiency and target the problem of color error, and further applications in video dehazing will be explored.

REFERENCES

[1] K. B. Gibson, D. T. Vo, and T. Q. Nguyen, "An investigation of dehazing effects on image and video coding," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 662–673, Feb. 2012.

[2] Z. Chen, B. R. Abidi, D. L. Page, and M. A. Abidi, "Gray-level grouping (GLG): An automatic method for optimized image contrast enhancement—Part II: The variations," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2303–2314, Aug. 2006.

[3] J. Kopf *et al.*, "Deep photo: Model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 32–39, Dec. 2008.

[4] E. J. McCartney, "Scattering by molecules and particles," in *Optics of the Atmosphere*. New York, NY, USA: Wiley, 1976.

[5] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Learn.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[6] L. Li, W. Feng, and J. W. Zhang, "Contrast enhancement based single image dehazing via TV-L1 minimization," in *Proc. IEEE Int. Conf. Multimedia Expo*, Chengdu, China, Jul. 2014, pp. 1–6.

[7] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, Jul. 2012.

[8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[9] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[10] L. Caraffa and J.-P. Tarel, "Markov random field model for single image defogging," in *Proc. IEEE Intell. Vehicles Symp.*, Gold Coast, QLD, Australia, Jun. 2013, pp. 994–999.

[11] R. Fattal, "Single image dehazing," *ACM Trans. Graph*, vol. 27, no. 3, Aug. 2008, Art. no. 72.

[12] Y.-K. Wang and C.-T. Fan, "Single image defogging by multiscale depth fusion," *IEEE Trans. Image Process*, vol. 23, no. 11, pp. 4826–4837, Nov. 2014.

[13] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph*, vol. 34, no. 1, pp. 256–269, Nov. 2014.

[14] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[15] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, Crete, Greece, Sep. 2010, pp. 1–14.

[16] B. Xie, F. Guo, and Z. Cai, "Improved single image dehazing using dark channel prior and multi-scale Retinex," in *Proc. Int. Conf. Intell. Syst. Design Eng. Appl.*, Oct. 2010, pp. 848–851.

[17] H. Xu, J. Guo, Q. Liu, and L. Ye, "Fast image dehazing using improved dark channel prior," in *Proc. IEEE Int. Conf. Inf. Sci. Technol.*, Mar. 2012, pp. 663–667.

[18] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. ICCV*, Dec. 2013, pp. 617–624.

[19] Y.-S. Lai, Y.-L. Chen, and C.-T. Hsu, "Single image dehazing with optimal transmission map," in *Proc. Int Conf. Pattern Recognit.*, Tsukuba, Japan, Nov. 2012, pp. 388–391.

[20] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[21] E. J. Candés and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Commun. Pure Appl. Math*, vol. 57, no. 2, pp. 219–266, Feb. 2002.

[22] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm fordesigning of over complete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[23] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[24] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries (K-LLD)," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1438–1451, Jul. 2009.

[25] P. Chatterjee and P. Milanfar, "Image denoising using locally learned dictionaries," *Proc. SPIE*, vol. 7246, no. 3, p. 72460V, Feb. 2009.

[26] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[27] L. Zhanga, W. Donga, D. Zhanga, G. Shib, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, Apr. 2010.

[28] X. Zhang, X. Feng, and W. Wang, "Two-direction nonlocal model for image denoising," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 408–412, Jan. 2013.

[29] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.

[30] L. Li, W. Feng, and J. Zhang, "Contrast enhancement based single image dehazing VIA TV-l1 minimization," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Chengdu, China, Jul. 2014, pp. 1–6.

[31] J.-P. Tarel, N. Hautière, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *Proc. IEEE Intell. Vehicle Symp.*, San Diego, CA, USA, Jun. 2010, pp. 478–485.

[32] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intell. Transp. Syst. Mag.*, vol. 4, no. 2, pp. 6–20, 2012.

[33] K. Gibson, D. Võ, and T. Nguyen, "An investigation in dehazing compressed images and video," in *Proc. IEEE OCEANS Conf.*, Sydney, NSW, Australia, Sep. 2010, pp. 1–8.

[34] C. H. Yeh, L. W. Kang, M. S. Lee, and C. Y. Lin, "Haze effect removal from image via haze density estimation in optical model," *Opt. Exp.*, vol. 21, no. 22, pp. 27127–27141, Nov. 2013.

[35] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *Proc. ICCV*, Sep. 2009, pp. 1701–1708.

[36] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 2201–2208.

Linyuan He received the B.S. and M.S. degrees in electrical engineering from Air Force Engineering University, Xi'an, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with Xi'an Jiaotong University. He is currently a Lecturer with the Department of Aeronautics and Astronautics, Air Force Engineering University, Xi'an.

Jizhong Zhao (A'12) received the B.S. and M.S. degrees in mathematics and Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1992, 1995, and 2001, respectively.

He is currently a Professor with Computer Science and Technology Department, Xi'an Jiaotong University. His research interests include computer software, pervasive computing, distributed systems, and network security.

Dr. Zhao is a member of the Association for Computing Machinery.

Nanning Zheng (SM'93–F'06) received the Degree from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1975, the M.S. degree in information and control engineering from Xi'an Jiaotong University in 1981, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

He was with Xi'an Jiaotong University in 1975, where he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. His research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems.

Dr. Zheng became a member of the Chinese Academy of Engineering in 1999. He is the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as an Executive Deputy Editor of the Chinese Science Bulletin.

Duyan Bi received the Degree from the Department of Electrical Engineering, National University of Defense Technology, Changsha, China, in 1983, the M.S. degree in signal processing from the National University of Defense Technology in 1987, and the Ph.D. degree in electrical engineering from Tours University, Tours, France, in 1997.

He was with the Department of Aeronautics and Astronautics, Air Force Engineering University, in 1987, where he is currently a Professor and the Director of the Laboratory of Field Reconnaissance and Surveillance Technology. His research interests include computer vision, pattern recognition, and image processing.