IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 October 2024, accepted 26 December 2024, date of publication 30 December 2024, date of current version 14 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3524316

RESEARCH ARTICLE

Drone-View Haze Removal via Regional Saturation-Value Translation and Soft Segmentation

TRUONG-DONG DO[1,2], (Graduate Student Member, IEEE), LE-ANH TRAN[3], (Member, IEEE), SEOKYONG MOON[3], JIO CHUNG[3], NGOC-PHI NGUYEN[4], AND SUNG KYUNG HONG[2,5]

[1]Department of Aerospace Systems Engineering, Sejong University, Seoul 05006, South Korea
[2]Department of Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, South Korea
[3]Department of Research and Development, Mindintech Inc., Seoul 05854, South Korea
[4]Institute of Mechanical and Electrical Engineering, Center for Industrial Mechanics, University of Southern Denmark, 6400 Sønderborg, Denmark
[5]Faculty of Mechanical and Aerospace Engineering, Sejong University, Seoul 05006, South Korea

Corresponding authors: Ngoc-Phi Nguyen (npnguyen@sdu.dk) and Sung Kyung Hong (skhong@sejong.ac.kr)

This work was supported in part by the Future Space Navigation and Satellite Research Center through the National Research Foundation funded by the Ministry of Science and ICT (MSIT), Republic of Korea, under Grant 2022M1A3C2074404; in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2020R1A6A1A03038540; and in part by MSIT through the Information Technology Research Center (ITRC) Support Program, Supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2024-RS-2024-00437494.

ABSTRACT This paper proposes an innovative single image dehazing framework, termed Regional Saturation-Value Translation (RSVT), to address the color distortion problems commonly encountered in bright regions by conventional dehazing approaches. The proposed RSVT framework is developed based on two key insights derived from the HSV color space: first, the hue component shows negligible variation between corresponding hazy and haze-free points; and second, in the 2D saturation-value coordinate system, the majority of lines connecting hazy-clean point pairs tend to converge near the atmospheric light coordinates. Consequently, haze removal can be achieved through appropriate translations within the saturation-value coordinates. Additionally, a robust soft segmentation method that employs a morphological min-max channel is integrated into the framework. By combining the soft segmentation mask with the RSVT prior, a comprehensive single image dehazing framework is established. Experimental evaluations across various datasets demonstrate that the proposed approach effectively mitigates color distortion and successfully restores visually appealing images. Moreover, a case study involving actual flight test demonstrates the feasibility and effectiveness of the proposed approach in real-world scenarios. The code is available at https://github.com/tranleanh/rsvt.

INDEX TERMS Dehazing prior, haze removal, image defogging, image dehazing, image restoration.

I. INTRODUCTION

Unmanned aerial vehicles (UAVs), also known as drones, have emerged as powerful tools that provide high-resolution imaging data across various remote areas that are difficult to access, enabling researchers to monitor changes in ecosystems, track deforestation, and assess biodiversity. However, the rise of industrialization has led to air pollution all over the world, resulting in dust, smoke, and haze in the natural atmosphere. As a result, aerial images captured by drones often suffer from reduced visibility with a loss in contrast and low color fidelity. This issue has posed challenges for various vision tasks that involve UAV imagery, where accurate scene interpretation is critical, such as object detection, target navigation, disaster management, and environmental monitoring. Therefore, drone-view image dehazing is a crucial preprocessing step to ensure that vision-based algorithms in aerial scenarios can operate effectively. Generally,

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

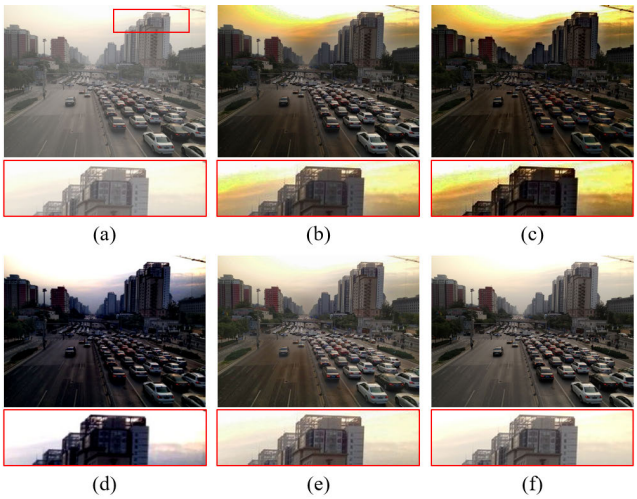


FIGURE 1. Dehazing results of various approaches on hazy images containing bright regions: (a) input, (b) DCP [2], (c) CEP [3], (d) NLID [4], (e) RSVT (ours), and (f) clean image.

haze removal algorithms can be classified into two main categories: *prior-based* and *deep learning-based* methods, each genre has its own strengths and weaknesses. Prior-based methods are effective for restoring visibility but they often result in over-saturation and artifacts in the output, whereas deep learning-based approaches have the potential to improve the realism of the restored images yet heavily rely on training datasets that necessitate hazy-clean image pairs from the same scene which can be costly for data preparation [1].

This paper is primarily targeted toward addressing a common issue faced by conventional prior-based dehazing algorithms when dealing with hazy images captured by aerial vehicles that usually comprise high-intensity regions such as the sky or bright objects. Typically, such images consist of two main parts: *foreground* and *background*. The background typically consists of sky regions while the foreground includes objects that are closer to the observer. While most prevailing prior-based dehazing algorithms perform quite well in restoring the visibility of the foreground regions, they often struggle with color distortion in the background areas. To address this issue, several approaches were proposed which tried to separately process the sky regions [5], [6], [7], [8]. However, it is very difficult or unattainable to segment images captured under dense haze circumstances, and in fact, hard segmentation may never yield accurate results, particularly in dim areas where there is no distinct boundary between the foreground and background regions. Even when the sky region is not visible, color distortion can still occur in the output due to bright objects like water surfaces or whitish buildings, as can be observed in Fig. 1b-1d. From those observations, we have analyzed the relationship between hazy points and their respective haze-free points within the high-intensity regions in the HSV color space. This analysis has yielded two significant findings. First, the difference in terms of hue component for a pair of hazy-clean points is marginal, as shown in Fig. 2c. This suggests that the impact of haze on the hue channel is insignificant, and the pixel value variations caused by haze may mainly occur in the saturation and value spaces. Second, when considering the 2D coordinate system formed by the saturation and value components, it is observed that the majority of lines connecting corresponding pairs of hazy-clean points, referred to as *S-V lines*, tend to converge near the atmospheric light coordinates, as illustrated in Fig. 2d. Based on these two observations, we can assume that the haze-free pixel can be estimated by shifting the hazy point along the corresponding S-V line by an appropriate amount. In order to validate this hypothesis, various statistical analyses have been conducted. Consequently, an innovative image dehazing prior called *Regional Saturation-Value Translation* (RSVT) is proposed to address the above-mentioned limitations of existing dehazing methods. On the other hand, dark channel prior (DCP) [2] is a widely recognized method due to its capacity to yield favorable restoration outcomes in both close-range and non-sky scenes. However, DCP struggles with bright regions such as the sky or white objects, as these areas naturally exhibit high-intensity pixels even in haze-free scenes, thereby violating the assumption of DCP. To address this limitation, an innovative dehazing framework is proposed which combines the potency of the proposed RSVT prior in dealing with the bright regions with the ability of DCP in handling foreground areas.

In contrast to other approaches that adopt hard segmentation to decompose a hazy image into background and foreground regions [5], [6], [8], the proposed method segregates the input image into three main components: *hard foreground* (regions that are completely classified as foreground), *hard background* (bright areas such as the daytime sky or whitish objects), and *middle ground* (dim areas that cannot be definitively classified as either background or foreground). To this end, we adopt a soft segmentation process using a morphological min-max channel. In our proposed framework, the RSVT prior and DCP methods handle the hard background and hard foreground regions, respectively, while the restoration for the middle ground is the weighted average solution of these two algorithms. The proposed approach's effectiveness is typically illustrated through the result shown in Fig. 1e, which shows that the proposed scheme significantly reduces color distortion and successfully recovers visually appealing images when compared with the other algorithms.

Part of this work was presented in [9] where the preliminary idea of the RSVT prior was introduced. In this paper, we improve the image decomposition process and conduct further experiments to prove the effectiveness of the proposed dehazing scheme as well as it applicability to real-world scenarios through a case study. The remainder of this paper is organized as follows: Section II presents the background materials including the haze imaging model, a brief review of single image dehazing research, and the guided filtering algorithm. The assumptions along with

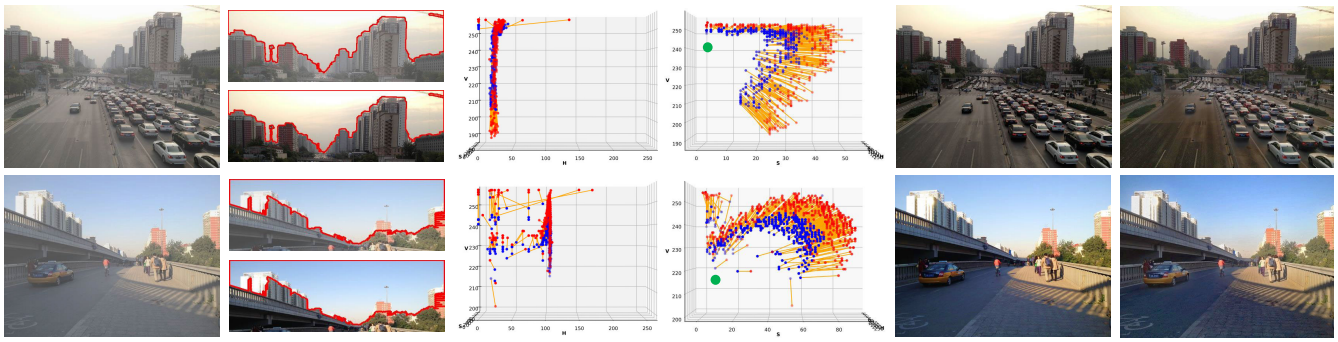


FIGURE 2. Regional Saturation-Value Transition: (a) hazy image, (b) sky regions in hazy (top) and clean (bottom) images bounded in red color, (c,d) Hue-Value and Saturation-Value perspectives, respectively: hazy and clean points are correspondingly indicated in blue and red, orange segments connect respective hazy-clean pairs, (e) clean image, and (f) dehazed image by the proposed method.

statistical analyses for the verification of the proposed prior are presented in Section III. In Section IV, a soft segmentation method for decomposing an input image is introduced, afterward a comprehensive single image dehazing framework is proposed. Section V presents the experimental results and comparisons. Section VI concludes the paper.

II. PRELIMINARIES

A. HAZE IMAGING MODEL

Generally, a hazy image can be modeled as a per-pixel convex combination between the actual scene radiance and the global atmospheric light [10], this model is known as the scattering atmospheric model or Koschmieder's law [11]:

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

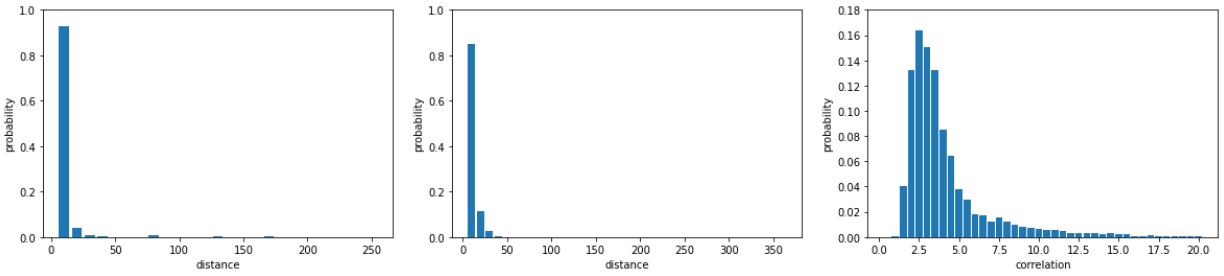
where $I(x)$, $J(x)$, A, and $t(x)$ denote the observed intensity, the scene radiance, the global atmospheric light, and the transmission map, respectively. From Eq. (1), most of the prior-based methods try to estimate the actual scene radiance $J(x)$ via the following function:

$$J(x) = \frac{I(x) - A}{t(x)} + A. \tag{2}$$

Relying on this recovering function, restored images can be obtained by solving the unknowns $t(x)$ and A. The value of A is generally assumed to be a high and constant value, while that of $t(x)$ for the bright and distant regions is usually very small and is typically set to 0.1 [2]. Consequently, a small difference in the magnitudes of $I(x)$ and A, for example, $|I(x) - A| = 10$, can cause a shift of 100 intensity levels in the restored image which yields noise and distorted color. Hence, this hazy imaging model may be invalid for performing dehazing on images that contain bright and sky areas.

B. BRIEF REVIEW ON IMAGE DEHAZING

As mentioned in Section I, haze removal algorithms can be categorized into two groups: prior-based and deep learning-based methods. Prior-based methods have been developed based on strong prior knowledge or assumptions. Meng et al. [12] introduced a boundary constraint and contextual regularization (BCCR) method which enforces constraints on the image boundaries and applies contextual regularization to enhance the sharpness of dehazed images. Zhu et al. [13] proposed a color-attenuation prior (CAP) which models the scene depth of a hazy image and learns the model parameters through supervised learning. Berman et al. [4] introduced a non-local image dehazing (NLID) prior based on the observation that hazy pixels in RGB color space are distributed along lines passing through the atmospheric light coordinates. Notably, dark channel prior (DCP) [2], which is developed based on a key observation that most local patches in haze-free outdoor images contain some pixels with very low intensity in at least one color channel, can directly estimate the haze thickness and recover a haze-free image via Eq. (2) with certain predetermined constraints. Although prior-based dehazing algorithms can achieve acceptable results in various circumstances, they are often incompetent in processing high-intensity regions such as daytime sky and bright objects which are commonly involved in high-altitude images captured by UAVs. Accordingly, drone-view image dehazing has emerged as a niche research area that has also gained considerable attention. Numerous studies have explored this research topic by adopting a fundamental approach of decomposing a hazy image into sky and non-sky regions. Wang et al. [6] proposed to utilize the quad-tree method to identify a seed point for the sky regions and employed region-growing segmentation with a Gaussian filter applied afterward for image smoothing. This approach, however, encounters difficulties when the sky region is obstructed by structures such as trees or buildings. Sebastiàn et al. [7] adopted a similar procedure but replaced the quad-tree method with the local Shannon entropy calculation. Nevertheless, growing-based techniques necessitate accurate estimation of seed points and careful control of the growth process, which can be time-consuming in order to prevent under and/or over-segmentation [14]. In addition to conventional methods, several studies [15], [16] have also incorporated neural networks for sky segmentation.



(a) (b) (c)

FIGURE 3. Statistical analyses: (a) differences between the hue components of hazy and haze-free point pairs within the sky regions (each bin stands for 10 distance levels), (b) distances between the intersections of S-V lines and the global atmospheric light coordinates in the Saturation-Value coordinate system (each bin stands for 10 distance levels), and (c) distribution of the correlation between the S-V ratio and the transmission.

However, it should be acknowledged that deep learning models are more computationally expensive to apply to sky segmentation which is not the primary objective of a dehazing system.

On the other extreme, deep learning models have demonstrated their remarkable ability to learn non-linear mappings in image translation tasks. Various learning-based dehazing approaches, such as [10], [17], [18], [19], [20], [21], [22], [23], and [24], have been proposed that can directly estimate $t(x)$ or $J(x)$ from a given input hazy image. These approaches commonly adopt convolutional neural networks (CNNs) as the primary backbone. However, the training process of CNNs requires extensive datasets with pairs of hazy and haze-free images captured in identical scenes which are too difficult and/or unachievable to prepare under real-world circumstances. A common approach to this issue is to synthesize hazy image data using the haze imaging model, but there always exists a certain gap between the synthetic and real-world data.

Even though learning-based approaches are dominant in the present era of deep learning, we believe that innovative priors also hold significant importance owing to their consistent statistical nature and strong underlying assumptions. Moreover, data-driven and learning-based approaches often entail expensive computational costs, which limit their applicability in real-time tasks. In contrast, prior-based methods offer greater simplicity and flexibility in deployment, making them well-suited for implementation on edge devices, such as companion computers integrated with drones. This has inspired us to come up with a novel and robust prior in an attempt to introduce a comprehensive image enhancement framework for high-quality haze-free image restoration.

C. GUIDED FILTERING

Guided filtering [25] is an image processing technique for edge-preserving image smoothing by utilizing a bilateral filter to eliminate noise and unexpected texture artifacts. The process takes a noisy image and a guidance image as input signals, and yields a corrected image. Given a guidance image G, an input image p, and an output image q, the fundamental assumption of the guided filter is the existence of a local linear model between the guidance G and the filtering output q which can be determined as follows:

$$q_i = a_k G_i + b_k, \forall i \in \omega_k, \tag{3}$$

where i is the pixel index, k is the index of a local patch ω, and (a_k, b_k) are linear coefficients assumed to be constant in ω_k. By minimizing the following cost function:

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k G_i + b_k - p_i)^2 + \epsilon a_k^2), \tag{4}$$

where ϵ is a regularization parameter that controls the degree of smoothness, the solution for (a_k, b_k) is given by:

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i \in \omega_k} G_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \epsilon}, \tag{5}$$

$$b_k = \bar{p}_k - a_k \mu_k, \tag{6}$$

where μ_k and σ_k represent the mean and variance of G in ω_k, respectively, while $|\omega|$ denotes the number of pixels in ω_k. Ultimately, the computation of the filtering output is carried out as follows:

$$q_i = GuidedFilter(p_i, G_i) = \bar{a}_i G_i + \bar{b}_i, \tag{7}$$

where $\bar{a}_i$ and $\bar{b}_i$ correspondingly are the average of a and b on the window ω_i centered at i.

III. REGIONAL SATURATION-VALUE VARIABILITY

The proposed prior is formulated based on two key insights derived from the bright areas of hazy-clean image pairs in the HSV color space. First, the impact of haze on the hue channel is found to be insignificant. Second, in the 2D saturation-value coordinate system, it is observed that the majority of S-V lines tend to converge near the atmospheric light coordinates. To verify these findings, a selection of 1,000 hazy-clean image pairs containing high-intensity pixels has been extracted from the RESIDE database [26] for comprehensive statistical analysis. The high-intensity regions are identified by applying image thresholding segmentation utilizing a high threshold, followed by manual verification.

These procedures lead to the formulation of three hypotheses aimed at substantiating the correctness of the proposed prior.

A. THE 1st ASSUMPTION

The first assumption states that: the hue component of a pixel in a bright region of a hazy image is relatively similar to that of the corresponding pixel in a haze-free image, as depicted in Fig. 2c. This assumption suggests that the changes in pixel intensity induced by haze primarily occur in the saturation and value spaces, as illustrated in Fig. 2d, and the dehazing process for bright hazy regions may be executed without considering the hue channel.

To validate this assumption, the absolute differences between the hue components of hazy and haze-free point pairs within the bright areas are measured. This measurement yields a probability distribution as shown in Fig. 3, which reveals that approximately 93% of the distances between the hue components are below 10, thus confirming the validity of our first assumption.

B. THE 2nd ASSUMPTION

The second assumption states that: when disregarding the impact of haze on the hue space, it can be observed that the majority of S-V lines, which represent the lines connecting each pair of hazy-clean points, are likely to intersect near the atmospheric light coordinates which are typically found within the brightest pixels of a hazy image.

In order to verify this assumption, the following procedure for each hazy-clean image pair is performed. First, the intersection point, denoted as E, of all the S-V lines is calculated. Subsequently, the Euclidean distance between E and the estimated global atmospheric light A is measured, where the estimation of A is achieved by randomly applying DCP [2] and NLID [4] methods:

$$||E - A|| = \sqrt{(S_E - S_A)^2 + (V_E - V_A)^2}, \tag{8}$$

where S and V denote the saturation and value components, respectively. Note that $S, V \in [0, 255]$, thus the maximum distance is $\sqrt{255^2 + 255^2} \approx 360$. Note also that all the S-V lines do not intersect perfectly at one single point, hence, we collect every intersection of each pair of S-V lines and E is computed as the average of all the intersections.

Fig. 3b illustrates the likelihood of the measured distances between E and A. It is evident the majority of distances are in proximity to small values, suggesting that the intersection is typically near the atmospheric light. This statistical analysis provides robust evidence in favor of the third assumption presented in the next section.

C. THE 3rd ASSUMPTION

In the analysis presented in Fig. 2d, it can be observed that in the HSV color space, a clean point $c(x)$ is derived by shifting the corresponding hazy point $h(x)$ away from the intersection E (indicated as a green point) by a magnitude that is directly proportional to the distance between the hazy point and the intersection, denoted as $||h(x) - E||$. The challenge at hand is to establish the relationship between $||h(x) - E||$ and any potential indicators associated with the distribution of haze in the input hazy image. Interestingly, upon revisiting the haze imaging model, the transmission in the RGB color space is represented as [2]:

$$t(x) = \frac{||I(x) - A||}{||J(x) - A||}, \tag{9}$$

which implies that the proportion between the distance from the hazy point to the atmospheric light point and the distance from the haze-free point to the atmospheric light point in the RGB color space directly represents the transmission. This leads us to the exploration of a similar type of proportion in the HSV color space. To this end, we initially eliminate two types of lines: one consists of all the S-V lines where the hazy or haze-free points have a value or saturation component that is close to 0 or 255, as these points' magnitudes may have been truncated due to the 8-bit image property; and the other category includes excessively short lines which can be regarded as "outliers" when calculating the distribution. Subsequently, a specific proportion, referred to as the *S-V ratio* $r(x)$ for each S-V line within the saturation-value space, is computed:

$$r(x) = \frac{||h(x) - E||}{||c(x) - E||}, \tag{10}$$

and we examine the likelihood of the correlation between the S-V ratio and the transmission, denoted as $R(x)$:

$$R(x) = \frac{r(x)}{t_b(x)}, \tag{11}$$

where $t_b(x)$ represents the transmission of the background region. In order to compute $t_b(x)$, we refer to the method described in [27] which utilizes the dark channel with a

FIGURE 4. Morphological min-max channel: (a) hazy image, (b) min channel with a large radius, (d) min channel with a small radius, (f) max channel with a large radius, and (c,e,g) refined images of (b,d,f), respectively.

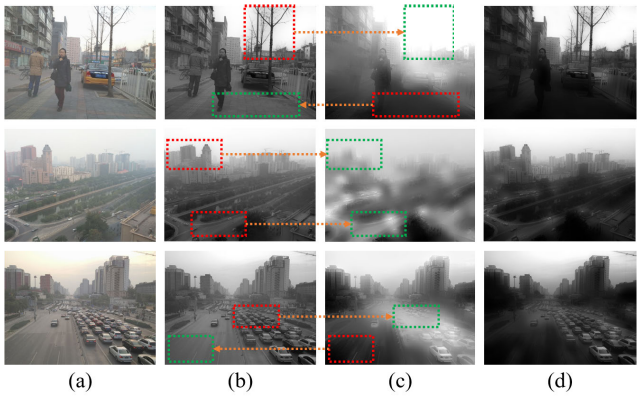


FIGURE 5. Fused channel: (a) hazy image, (b) refined min channel, (c) refined max channel, and (d) fused channel of (b) and (c) through spatial-wise multiplication. The red and green boxes indicate oppositely low-intensity and high-intensity image patches from two channels at the same location, respectively.

small window size, i.e., 3×3, to initially estimate a coarse transmission $\tilde{t_b}(x)$:

$$\tilde{t_b}(x) = 1 - \omega \min_{y \in p_{3\times3}(x)} (\min_{c \in (r,g,b)} \frac{I_c(y)}{A_c}), \tag{12}$$

where ω ($0 < \omega \leq 1$) denotes a predefined parameter to optionally keep a small amount of haze for the distant objects (ω is typically set to 0.95 [2]), afterward $t_b(x)$ can be obtained by refining $\tilde{t_b}(x)$ using guided filtering:

$$t_b(x) = GuidedFilter(\tilde{t_b}(x), I_{gray}), \tag{13}$$

where I_{gray} denotes the gray image of the input.

Fig. 3c illustrates the obtained distribution of $R(x)$, which shows that the highest frequencies for $R(x)$ occur within the interval [2.0, 5.0] with approximately 85-90% of cases. As a rough assumption, we consider setting $R(x)$ as a constant for any input image such that the intensities of the restored image are minimally affected when $R(x)$ varies within the range of its highest frequencies (this assumption is verified in Section V). However, it is worth noting that this assumption may not hold true in all circumstances, and in fact, $R(x)$ can be fine-tuned for each image to obtain an optimal result. In order to partially compensate for this limitation, a justification is presented in the following which explains the applicable range of $R(x)$.

By examining Fig. 2d, it can be observed that each clean point $c(x)$ is derived by shifting the corresponding hazy point $h(x)$ far away from the intersection E (marked as a green point). That is, $h(x)$ tends to be closer to E when compared with $c(x)$. Consequently, it can be concluded that $||h(x) - E|| \leq ||c(x) - E||$, indicating that the S-V ratio $r(x)$ is roughly less than 1:

$$r(x) \leq 1. \tag{14}$$

On the other hand, $t_b(x)$ is used to indicate the transmission for the background regions. Typically, the minimum value for $t_b(x)$ is set to 0.1 in most scenarios to represent the transmission of sky regions: $t_b(x) \geq 0.1$ [2]. However, in our situation, $t_b(x)$ not only represents the transmission for the sky but also for certain nearby bright objects. Therefore, the lower bound for $t_b(x)$ in our case can be adjusted slightly higher, i.e., 0.2:

$$t_b(x) \geq 0.2 \text{ or } 1/t_b(x) \leq 5. \tag{15}$$

From Ep. (14) and Ep. (15), we have:

$$R(x) = r(x)/t_b(x) \leq 5. \tag{16}$$

This constraint is intriguingly justifiable with respect to the distribution shown in Fig. 3c, where the value of $R(x)$ is generally below 5. According to this constraint and the distribution, we hypothesize that the potential values for $R(x)$ lie within the range of [2.0, 5.0] and we examine multiple configurations of $R(x)$ within this interval in Section V.

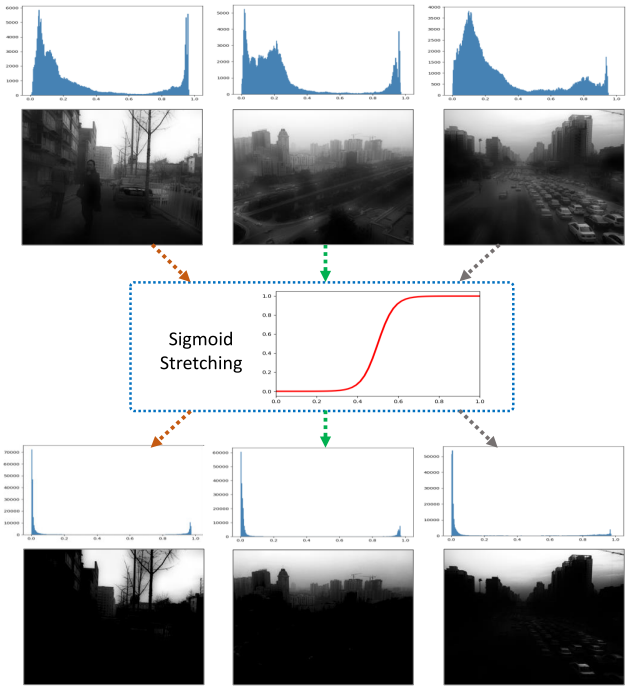


FIGURE 6. Sigmoid stretching: the coarse soft segmentation masks and their respective histogram before (top) and after (bottom) stretching process.

IV. THE PROPOSED DEHAZING METHOD

In this section, we first present a soft segmentation method utilizing a morphological min-max channel to effectively separate an input hazy image into three components: hard foreground, hard background, and middle ground. Subsequently, we propose a robust framework for single image dehazing that integrates the proposed RSVT prior with DCP.

A. MORPHOLOGICAL MIN-MAX CHANNEL

In this section, a straightforward yet efficient soft segmentation strategy, called morphological min-max channel, which is inspired by dark/bright channel concepts [2], [29] is

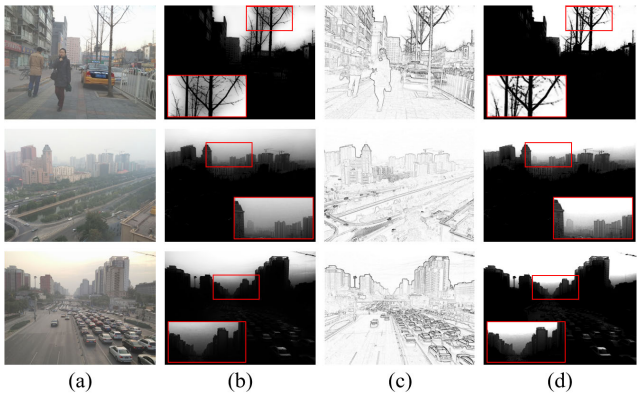


FIGURE 7. Refinement of fused channel: (a) hazy image, (b) coarse fused channel, (c) inverse edge image, and (d) refined fused channel.

proposed. Specifically, two channels, namely the foreground channel and the background channel, are extracted from a given hazy image. These channels are computed to emphasize the textures of their respective foreground/background regions. Subsequently, the two channels are combined to generate a unified mask that can effectually segment the input image.

Fig. 4a shows an example of a hazy image that can be roughly segmented into two main regions: sky and non-sky. The non-sky region, which is considered the foreground, typically consists of various types of object textures and shows higher contrast compared to the sky region. Consequently, the majority of non-sky regions should be connected with the top lowest-intensity pixels. Moreover, the foreground region usually contains floating bright pixels. To address this, a pixel-level min operation followed by a morphological erosion process is adopted to highlight the foreground. The purpose of the morphological erosion is to eliminate the floating bright pixels, thereby retaining only the significant foreground object textures. The computation of the coarse foreground channel $\tilde{I}_{fore}(x)$ for an input hazy image is as follows:

$$\tilde{I}_{fore}(x) = \min_{c\in RGB} I_c(x) \ominus \Omega_f(x), \tag{17}$$

where "$\ominus$" represents the morphological erosion operator while $\Omega_f(x)$ denotes a circle-shaped structuring element centered at x for the erosion operation. Afterward the refined foreground channel $I_{fore}(x)$ is obtained by using guided filtering method:

$$I_{fore} = GuidedFilter(\tilde{I}_{fore}, I_{gray}). \tag{18}$$

In addition, to preserve the edge details and prevent the occurrence of shadow artifacts between the background and foreground areas, the radius of $\Omega_f(x)$ is typically set to a small value. In order to illustrate this effect of radius on the outcome, an example is given in Fig. 4b and Fig. 4d which show the coarse foreground channels with radius values of 30 and 3, respectively, while Fig. 4c and Fig. 4e depict the refined foreground channels of Fig. 4b and Fig. 4d,

Algorithm 1 The Proposed Image Dehazing Framework

Input: Hazy image $I(x)$, correlation value $R(x)$
Output: Haze-free image $J(x)$

 # Soft-segmentation Mask Extraction:
$I_{gray} = \text{rgb2gray}(I)$ ▷ RGB to gray
$I_{edge} = 1 - \text{sobel}(I_{gray})$ ▷ inverse edge
$\tilde{I}_{fore}(x) = \min_{c\in RGB} I_c(x) \ominus \Omega_f(x)$
$I_{fore} = \text{guidedfilter}(\tilde{I}_{fore}, I_{gray})$
$\tilde{I}_{back}(x) = \max_{c\in RGB} I_c(x) \oplus \Omega_b(x)$
$I_{back} = \text{guidedfilter}(\tilde{I}_{back}, I_{gray})$
$I_{fuse} = \text{sigmoid}(I_{fore} I_{back})$ ▷ stretching
$I_{fuse}(x) = \begin{cases} I_{fuse}(x), & \text{if } I_{fuse}(x) \geq \tau \\ 0, & \text{if } I_{fuse}(x) < \tau \end{cases}$
$M = \min(\lambda I_{edge} I_{fuse}, 1)$ ▷ soft-seg mask

 # Saturation-Value Translation:
$A = \text{average}(\text{top } 0.1\%(\forall M(x) | M(x) \neq 1))$
$\tilde{t_b}(x) = 1 - \omega \min_{y\in p_{3\times 3}(x)}(\min_{c\in(r,g,b)} \frac{I_c(y)}{A_c})$
$t_b(x) = \text{guidedfilter}(\tilde{t_b}(x), I_{gray})$
$h = \text{rgb2hsv}(I)$ ▷ RGB to HSV
$\delta_{S,V}(x) = \frac{1 - R(x)t_b(x)}{R(x)t_b(x)} ||h_{S,V}(x) - A_{S,V}||$
$H_c(x) = H_h(x)$
$S_c(x) = \begin{cases} S_h(x) + \delta_S(x), & \text{if } S_h(x) \geq S_A \\ S_h(x) - \delta_S(x), & \text{if } S_h(x) < S_A \end{cases}$
$V_c(x) = \begin{cases} V_h(x) + \delta_V(x), & \text{if } V_h(x) \geq V_A \\ V_h(x) - \delta_V(x), & \text{if } V_h(x) < V_A \end{cases}$
$D = \text{hsv2rgb}(\{H_c, S_c, V_c\})$ ▷ HSV to RGB

 # Restoration:
$\tilde{t}(x) = 1 - \omega \min_{y\in p_{15\times 15}(x)}(\min_{c\in(r,g,b)} \frac{I_c(y)}{A_c})$
$t(x) = \text{guidedfilter}(\tilde{t}(x), I_{gray})$
$J(x) = \frac{I(x) - A(1 - t(x))}{max(t_0, t(x))}(1 - M(x)) + D(x)M(x)$

respectively. It can be observed clearly that a small radius can keep the details of objects while a large one can create shadow artifacts around the edges and result in a loss in object details.

On the other hand, the majority of bright regions ought to be associated with the top highest-intensity pixels which can be obtained by adopting a pixel-wise max operation. Furthermore, these bright regions typically lack object textures and are spread out across a local region of an image. Hence, morphological dilation with a substantial radius can enhance the background characteristics in this context. As a result, the coarse background channel, denoted as $\tilde{I}_{back}(x)$, is accordingly computed as:

$$\tilde{I}_{back}(x) = \max_{c\in RGB} I_c(x) \oplus \Omega_b(x), \tag{19}$$

where "$\oplus$" denotes the morphological dilation operator while $\Omega_b(x)$ represents a circle-shaped structuring element centered at x for the dilation operation. Then the refined

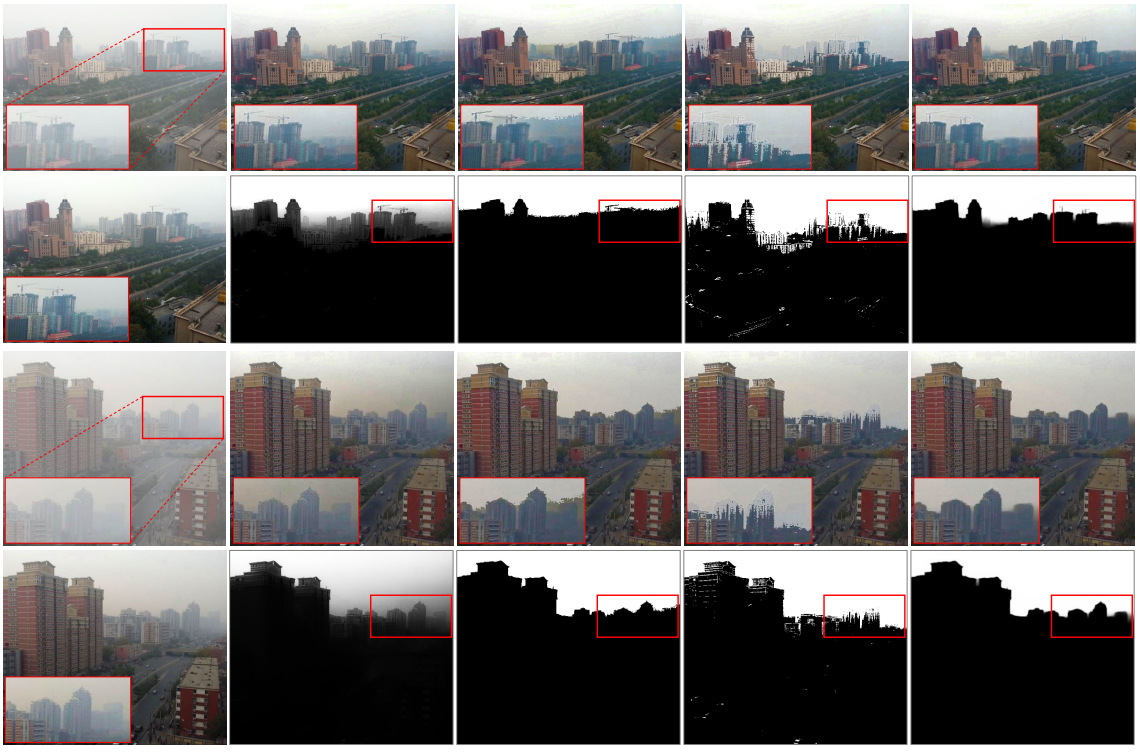


FIGURE 8. Dehazing results when RSVT is combined with various segmentation methods: (a) hazy (top) and clean (bottom) images, (b) the proposed morphological min-max channel method, (c) region growing-based method adopted in [6] and [7], (d) GMM-based method used in [8], and (e) deep learning-based method using U^2-Net [28]. The respective segmentation mask is shown below each output.

background channel can be attained as:

$$I_{back} = GuidedFilter(\tilde{I}_{back}, I_{gray}). \tag{20}$$

Fig. 4f and Fig. 4g illustrate the coarse and refined background channels, respectively. As can be observed from Fig. 4e, the refined foreground channel can capture fine details of object textures, but still contains some non-dark areas. Interestingly, those corresponding regions in the refined background channel, depicted in Fig. 4g, can be leveraged in order to cover the uneven bright areas in the refined foreground channel. Therefore, these two channels are fused through spatial-wise multiplication, resulting in a coarse soft segmentation mask that roughly separates the sky and non-sky regions into bright and dark regions, respectively, as illustrated in Fig. 5. However, the intervals of the output values may be unsatisfactory due to the spatial-wise multiplication step. Therefore, a sigmoid model can be employed to stretch the values of the coarse mask:

$$I_{fuse} = Sigmoid(I_{fore} I_{back}), \tag{21}$$

where I_{fuse} denotes the fusion outcome. Here the sigmoid model is adopted to stretch the foreground and background regions to have their weight values closer to 0 and 1, respectively. An illustration of this stretching process is presented in Fig. 6.

B. MASK REFINEMENT

As can be seen from Fig. 6, the stretched mask can separate an input hazy image into two parts of background and foreground to a certain extent. However, it still contains residual shadow halos at some regions between background and foreground, as shown in red boxes in Fig. 7b. To address this issue, a refinement process is carried out to attain an improved and refined mask. Typically, the background regions are spread out across a local area of the image with a low variation of pixel intensities. To take advantage of this property, the Sobel edge detection operator [30] is adopted to compute the pixel gradients and extract the basic edge image of the input, followed by an image inverting step to bring the background intensities closer to 1:

$$I_{edge} = 1 - Sobel(I_{gray}), \tag{22}$$

where I_{edge} denotes the obtained inverse edge image, as shown in Fig. 7c.

On the other hand, from the stretched mask shown in Fig. 7b, we observe that the regions having intensities closer to 0 have a higher probability of being in the foreground. Therefore, these regions with very low intensities can be

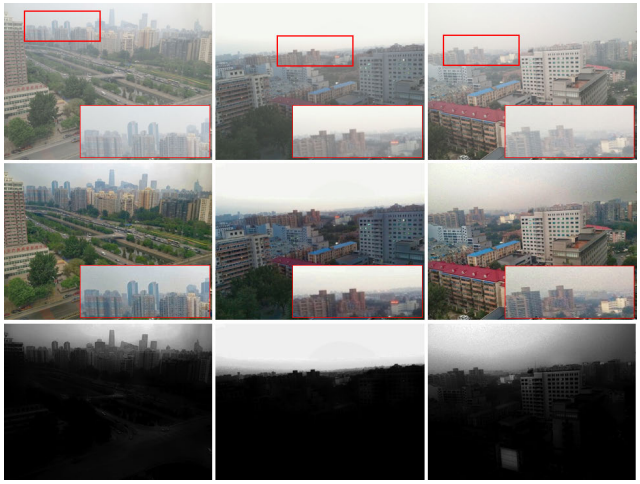


FIGURE 9. Dehazing results by the proposed method (top: input image, middle: restored image, bottom: soft segmentation mask).

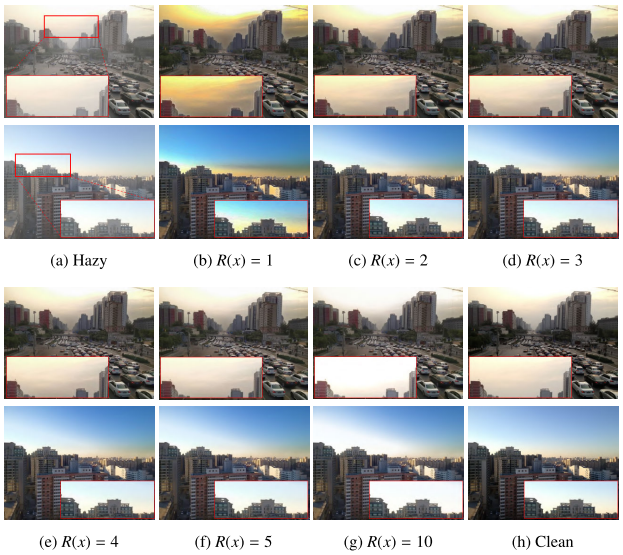


FIGURE 10. Dehazing results with various configurations of $R(x)$.

considered *hard foreground* with a constraint:

$$I_{fuse}(x) = \begin{cases} I_{fuse}(x), & \text{if } I_{fuse}(x) \geq \tau \\ 0, & \text{if } I_{fuse}(x) < \tau \end{cases} \tag{23}$$

where τ is a very low threshold and is typically set to 0.01. Subsequently, a spatial-wise multiplication is applied to the stretched mask and the inverse edge image. In so doing, the high intensity of the background in the inverse edge image can amplify the background region while the lower intensity of the stretched mask can down-weight the foreground region. This particular procedure can be referred to as a type of spatially varying detail enhancement [31] where the detail layer and the amplification factor are represented by the inverse edge image I_{edge} and the stretched mask I_{fuse}, respectively. In practice, an outdoor input image may or may not contain sky regions, and the certainty of a pixel belonging to the foreground/background region can be considered random. Therefore, from the probability perspective, we assume that the stretched mask can be regarded as a random distribution, with element values ranging from 0 to 1 and a mean value $\phi = 0.5$. Accordingly, the pixel-wise product of the stretched mask and the inverse edge image yields a darker image, as both input signals have values lower than 1, and the overall brightness of the detail layer could be degraded by ϕ. To handle this, we adopt a compensation coefficient λ, with $\lambda = 1/\phi = 2$, to counterbalance the brightness degradation in the background. With the effect of λ, the outcome may contain values exceeding 1, which are then classified as *hard background*. To obtain a refined and normalized soft segmentation mask, these pixel values are limited to a maximum of 1. The remaining pixels in the refined mask, which do not belong to either the hard foreground or hard background, are referred to as the *middle ground*. Ultimately, the refined mask M is obtained as:

$$M = min(\lambda I_{edge} I_{fuse}, 1). \tag{24}$$

C. ATMOSPHERIC LIGHT ESTIMATION

In the DCP method, the estimation of atmospheric light is carried out by selecting the pixels with the highest intensity respective to the top 0.1% brightest pixels in the dark channel. However, this strategy is not always valid and can lead to color shift in the final output. In order to overcome this limitation and achieve a more accurate selection of global atmospheric light, we propose referring to the top 0.1% brightest pixels in the middle ground. This approach can avoid the selection of pixels that fall within excessively bright regions which may be associated with bright or white objects. The estimated atmospheric light is then computed as the average of these selected pixels.

D. REGIONAL SATURATION-VALUE TRANSLATION

As stated in Section III-C, the haze-free point can be estimated by translating the hazy point far away from the atmospheric light by an amount directly proportional to its distance to the atmospheric light. To determine the precise amount of translation required for estimating the haze-free point, from Eq. (10) we have:

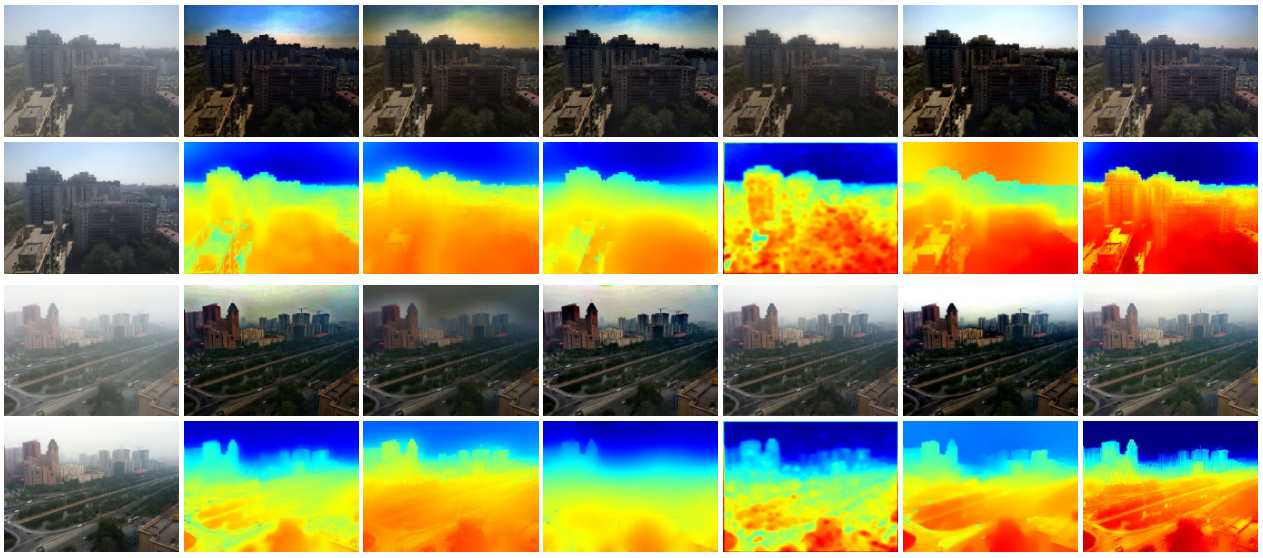
$$\frac{||c(x) - A||}{||h(x) - A||} = \frac{1}{r(x)}, \tag{25}$$

note that the intersection E is replaced with the estimated atmospheric light A based on the second assumption presented in Section III-B. Combining with Eq. (11), Eq. (25) is equivalent to:

$$\frac{||c(x) - h(x)||}{||h(x) - A||} + 1 = \frac{1}{R(x)t_s(x)}, \tag{26}$$

and the translation quantity $\delta(x)$ can be calculated as:

$$\delta(x) = ||c(x) - h(x)|| = \frac{1 - R(x)t_s(x)}{R(x)t_s(x)} ||h(x) - A||. \tag{27}$$



(a) Hazy/clean (b) CEP (c) DCP (d) BCCR (e) CAP (f) NLID (g) RSVT (ours)

FIGURE 11. Typical dehazing results of various approaches on SOTS-Outdoor dataset: (a) Hazy (top) and clean (bottom) images, (b-g) results by CEP [3], DCP [2], BCCR [12], CAP [13], NLID [4], and the proposed RSVT method, respectively. The respective recovered transmission map for each method is shown below every output.

As a result, the haze-free image is restored by the following adjustments in the HSV color space:

$$H_c(x) = H_h(x), \tag{28}$$

$$S_c(x) = \begin{cases} S_h(x) + \delta_S(x), & \text{if } S_h(x) \geq S_A \\ S_h(x) - \delta_S(x), & \text{if } S_h(x) < S_A \end{cases} \tag{29}$$

$$V_c(x) = \begin{cases} V_h(x) + \delta_V(x), & \text{if } V_h(x) \geq V_A \\ V_h(x) - \delta_V(x), & \text{if } V_h(x) < V_A \end{cases} \tag{30}$$

where (H_c, S_c, V_c), (H_h, S_h, V_h) denote three channels of the restored image and the hazy image in the HSV color space, respectively.

E. HAZE-FREE IMAGE RECOVERY

After acquiring the segmentation mask, the proposed RSVT prior and the DCP approach process the hard background and hard foreground regions, respectively. Meanwhile, the restoration of the middle ground is achieved through a weighted average solution derived from these two algorithms. The proposed dehazing scheme can be expressed as follows:

$$J(x) = \frac{I(x) - A(1 - t(x))}{max(t_0, t(x))}(1 - M(x)) + D(x)M(x), \tag{31}$$

where $t(x)$ is estimated by DCP and t_0 (typically set to 0.1) is a lower bound of $t(x)$ [2], while $D(x)$ is the dehazed image produced by RSVT formatted in the RGB color space. The Python-like pseudocode of the proposed dehazing framework is illustrated in Algorithm 1.

V. EXPERIMENTS AND DISCUSSION

In this section, the effectiveness of the proposed framework is examined in various experimental analyses. We first present the experimental settings including the datasets used in experiments as well as listing the methods utilized for comparisons. The performance of scene segmentation stage and the selection of $R(x)$ are then discussed. Subsequently, the performance of the proposed dehazing method is compared with other prevalent techniques in terms of both dehazing effectiveness and computational efficiency. Lastly, a case study is conducted to evaluate the practical applicability of the proposed dehazing framework in real-world scenarios.

TABLE 1. Ablation study on the values of $R(x)$. The best and second-best results are indicated in bold and blue, respectively.

Dataset	Metric	$R(x)$					
		1	2	3	4	5	10
SOTS Outdoor	PSNR	17.68	21.09	**22.31**	21.98	21.06	18.35
	SSIM	0.7894	0.8835	**0.9031**	0.8668	0.8691	0.8115
HSTS Synthetic	PSNR	16.59	21.22	**21.53**	21.06	20.45	16.82
	SSIM	0.7740	0.8798	**0.8868**	0.8785	0.8325	0.7853

A. EXPERIMENTAL SETTINGS

For the sake of fair comparison, the proposed method has been compared with three main types of other methods: ***prior-based approaches*** (CEP [3], DCP [2], CAP [13], NLID [4], CLAHE [32], BCCR [12]), ***lightweight CNNs*** (AOD-Net [17], MSCNN [18], DehazeNet [19], GFN [33], GCANet [20], D4+ [23], SDDN-T [24]), and ***upsupervised CNNs*** (CycleDehaze [34], YOLY [21], RefineDNet [1]). The experiments were conducted on four data sets selected from the RESIDE benchmark [26], including two synthetic datasets, SOTS-Outdoor and HSTS-Synthetic, and two realistic image sets, HSTS-Realistic and RESIDE-Natural.

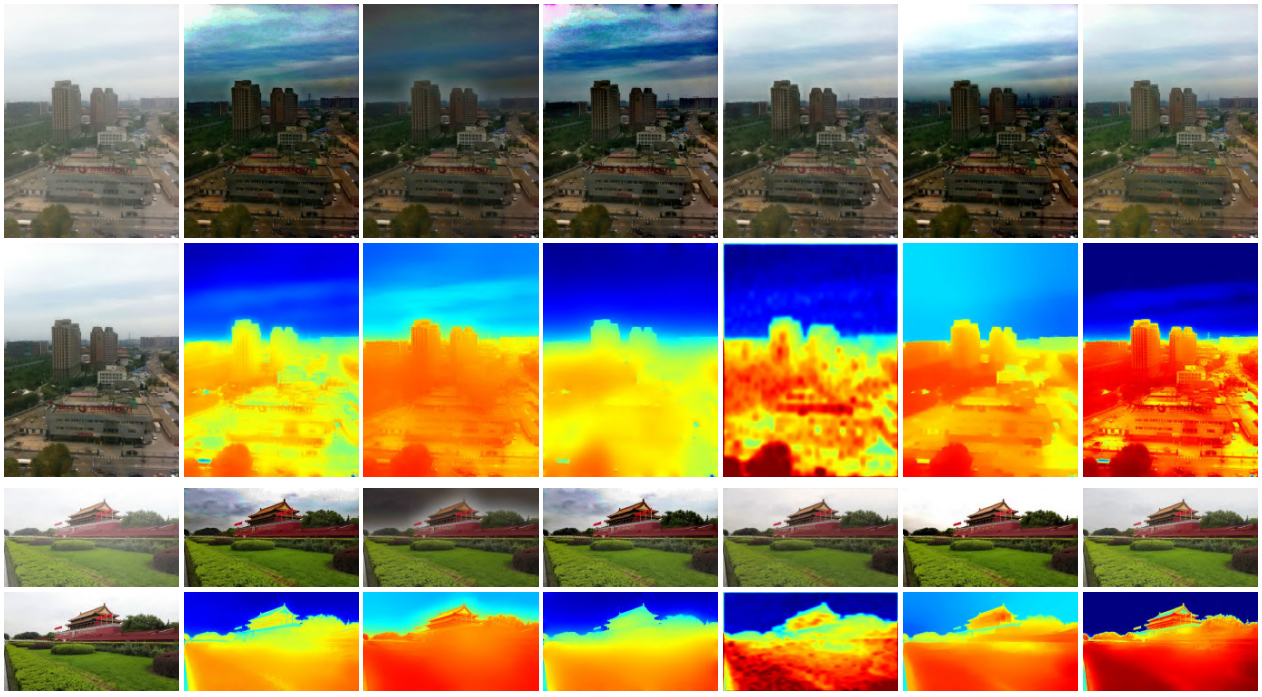


FIGURE 12. Typical dehazing results of various approaches on HSTS-Synthetic dataset: (a) Hazy (top) and clean (bottom) images, (b-g) results by CEP [3], DCP [2], BCCR [12], CAP [13], NLID [4], and the proposed RSVT method, respectively. The respective recovered transmission map for each method is shown below every output.

The SOTS-Outdoor and HSTS-Synthetic datasets consist of 500 pairs and 10 pairs of hazy-clean outdoor images, respectively, while the HSTS-Realistic dataset includes 10 real-world hazy images, and the RESIDE-Natural dataset contains 20 randomly sampled images with natural haze. These data sets exhibit significant scene variability and serve as de facto benchmarks for dehazing algorithm evaluation due to their widespread adoption in various studies. Quantitative performance on synthetic data was measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), while dehazing performance on natural images was evaluated using Natural Image Quality Evaluator (NIQE) [35] and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [36]. The experiments were conducted on an Intel(R) Core(TM) i5-8600K CPU @ 3.60GHz.

B. EVALUATION OF SCENE SEGMENTATION

In order to validate the effectiveness of the proposed segmentation approach, a qualitative analysis is conducted to compare the segmentation masks generated by our approach and those obtained from other segmentation methods, including the region-growing method [37], GMM-based segmentation [38], and U^2-Net [28]. Note that the region-growing and GMM-based approaches have been adopted in several existing image decomposition-based dehazing schemes like [6], [7], and [8], while U^2-Net represents a recent robust neural network-based image enhancement method for images with the appearance of the sky [15], [16]. Based on the visual comparisons presented in Fig. 8, it is evident that our proposed method is capable of generating smooth soft segmentation masks with a natural transition from foreground to background and can avoid generating undesirable halo artifacts at the boundary regions. Notably,

TABLE 2. Quantitative performances of various dehazing methods on SOTS-Outdoor and HSTS-Synthetic datasets. The best and second-best results are indicated in bold and blue, respectively.

Type	Method	SOTS-Outdoor		HSTS-Synthetic	
		PSNR	SSIM	PSNR	SSIM
Prior-based	CEP [3]	13.44	0.7053	14.02	0.7270
	DCP [2]	16.62	0.8179	17.01	0.8030
	CLAHE [32]	16.38	0.8072	15.55	0.7886
	BCCR [12]	16.88	0.7913	15.21	0.7470
	NLID [4]	17.29	0.7489	17.62	0.7980
	CAP [13]	19.05	0.8364	21.57	0.8730
CNN-based	AOD-Net [17]	20.29	0.8770	20.55	0.8970
	MSCNN [18]	19.48	0.8390	18.64	0.8170
	DehazeNet [19]	22.46	0.8510	**24.48**	0.9150
	GFN [33]	21.55	0.8440	22.06	0.8470
	GCANet [20]	21.66	0.8670	21.37	0.8740
	CycleDehaze [34]	19.95	0.8858	20.02	0.8780
	YOLY [21]	20.39	0.8890	21.02	0.9050
	RefineDNet [1]	20.81	0.8765	20.66	0.8897
	D^4+ [23]	**26.30**	**0.9600**	22.93	**0.9266**
	SDDN-T [24]	24.39	0.9107	23.36	0.8964
Prior-based	RSVT (ours)	22.31	0.9031	21.53	0.8868

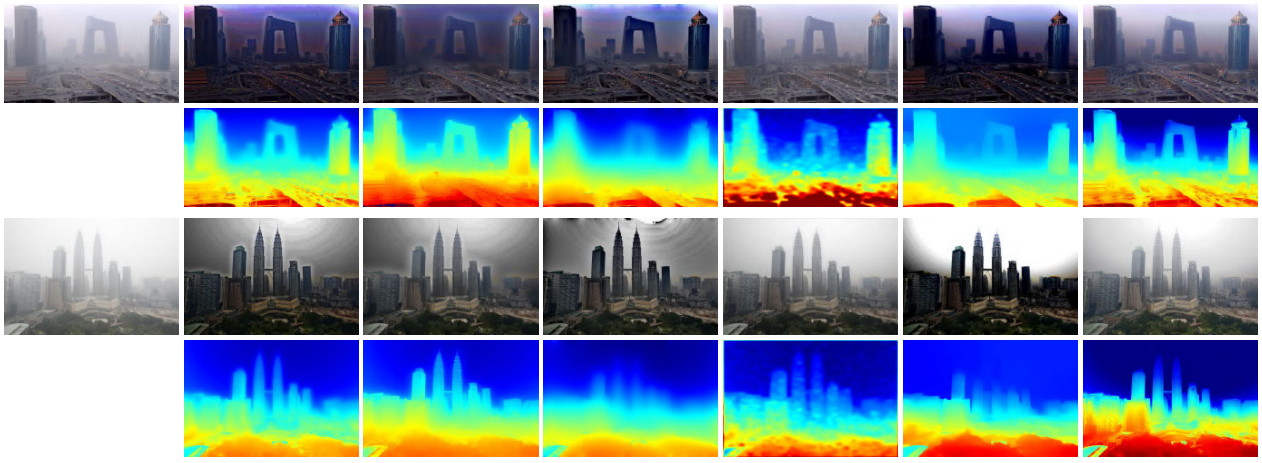


FIGURE 13. Typical dehazing results of various approaches on HSTS-Realistic dataset: (a) Hazy image, (b-g) results by CEP [3], DCP [2], BCCR [12], CAP [13], NLID [4], and the proposed RSVT method, respectively. The respective recovered transmission map for each method is shown below every output.

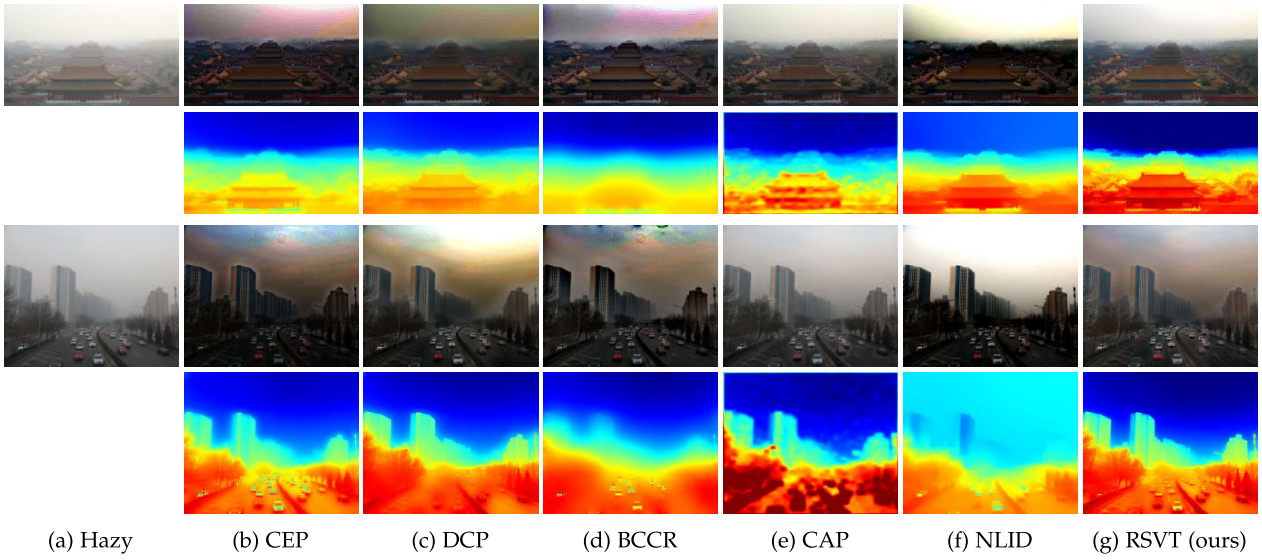


FIGURE 14. Typical dehazing results of various approaches on RESIDE-Natural dataset: (a) Hazy image, (b-g) results by CEP [3], DCP [2], BCCR [12], CAP [13], NLID [4], and the proposed RSVT method, respectively. The respective recovered transmission map for each method is shown below every output.

our approach outperforms both the region-growing and GMM-based segmentation methods, and can be considered competitive when compared with the deep learning method, U^2-Net, while it is worth mentioning that our approach does not necessitate a training process.

Additionally, the dehazing results obtained by integrating our proposed RSVT prior with some other segmentation methods discussed earlier are also examined, as depicted in Fig. 8. The analysis reveals that the proposed approach produces visually compelling restored images, particularly in challenging regions, where other methods struggle. Specifically, in distant and dim areas where the boundary between foreground and background regions is unclear, the other techniques under comparison fail to deliver satisfactory dehazing outputs. This limitation is particularly evident in scenes with low contrast and minimal visual cues. In contrast, our method effectively restores these difficult regions, highlighting its robustness against the other algorithms. Further visual dehazing results generated by the proposed framework, along with their corresponding segmentation masks, are presented in Fig. 9. These additional examples provide further compelling evidence supporting the advantages of our proposed approach in handling complex environments, illustrating the consistency and reliability of the RSVT prior

FIGURE 15. Dehazing results of the proposed RSVT method in comparison to those of some deep learning-based methods on synthetic hazy images.

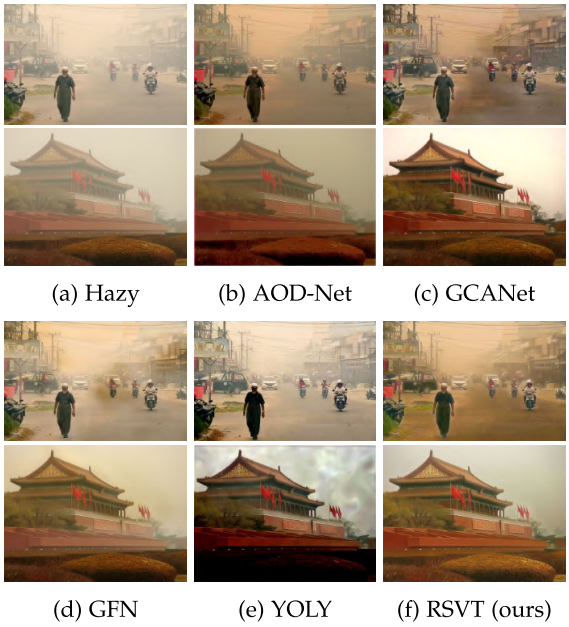


FIGURE 16. Dehazing results of the proposed RSVT method in comparison to those of some deep learning-based methods on natural hazy images.

TABLE 3. Quantitative comparisons of various dehazing methods on HSTS-Realistic and RESIDE-Natural datasets. The best and second-best results are indicated in **bold** and blue, respectively.

Type	Method	HSTS-Realistic		RESIDE-Natural	
		NIQE	BRISQUE	NIQE	BRISQUE
Prior-based	CEP [3]	13.583	49.785	12.260	30.419
	DCP [2]	12.318	52.234	14.181	30.947
	CLAHE [32]	13.450	50.272	13.446	23.911
	BCCR [12]	13.713	49.261	14.052	34.072
	NLID [4]	14.146	62.126	13.493	41.265
	CAP [13]	13.899	38.305	12.900	33.001
CNN-based	AOD-Net [17]	12.554	41.200	11.971	24.658
	GFN [33]	14.612	48.178	12.970	24.983
	GCANet [20]	15.303	58.829	12.406	23.360
	YOLY [21]	14.542	46.535	12.854	24.235
	D^4+ [23]	13.568	43.354	12.533	25.397
	SDDN-T [24]	**12.357**	43.683	12.006	22.358
Prior-based	RSVT (ours)	13.414	**36.711**	**11.780**	**20.659**

across diverse scene types, and underscoring its robustness in accurately processing and interpreting intricate visual data.

C. SELECTION OF R(X)

According to the constraint presented in Section III-C, it is assumed that $R(x)$ could be set as a constant for any input image such that the intensities of the restored image are minimally affected when $R(x)$ varies within the range of its highest frequencies, [2.0, 5.0]. To verify this assumption, we examine the impact of different values of $R(x)$ on the visibility of the outputs when $R(x)$ is set to in-range and out-of-range values, $R(x) = \{1, 2, 3, 4, 5, 10\}$, where $R(x) = 1$ and $R(x) = 10$ are typical out-of-range values.

The results of the quantitative evaluation are summarized in Table 1, which shows that the best performance is achieved when $R(x)$ is set closer to the highest-frequency value, specifically $R(x) = 3$. Also, the other in-range values of $R(x)$ such as $R(x) = \{2, 4, 5\}$ also yield competitive results despite of slight drops in performance. On the contrary, the use of out-of-range values, $R(x) = \{1, 10\}$, results in a degradation in the dehazing performance. To visually verify these measurements, qualitative restoration results are demonstrated in Fig. 10. As can be seen Fig. 10, $R(x) \in [2.0, 5.0]$ can result in competitive visual dehazing outcomes whereas $R(x) = \{1, 10\}$ may lead to color distortion or over-saturation. Based on the distribution obtained, as shown in Fig. 3c, alongside the comprehensive quantitative evaluation detailed in Table 1, we have determined that $R(x) = 3$ should be established as the primary configuration in the proposed framework.

D. EVALUATION ON BENCHMARK DATASETS

Table 2 summarizes the quantitative results of various dehazing approaches on the SOTS-Outdoor and HSTS-Synthetic datasets. As reported in Table 2, the proposed RSVT framework outperforms all the prior-based methods under comparison by a significant margin in both PSNR and SSIM measures. Moreover, the proposed approach achieves competitive performance in SSIM when compared against some deep learning-based models such as GCANet and YOLY. On the other hand, further quantitative comparisons on natural hazy image datasets, HSTS-Realistic and RESIDE-Natural, are summarized in Table 3. As reported in Table 3, the proposed method also shows competitive outcomes

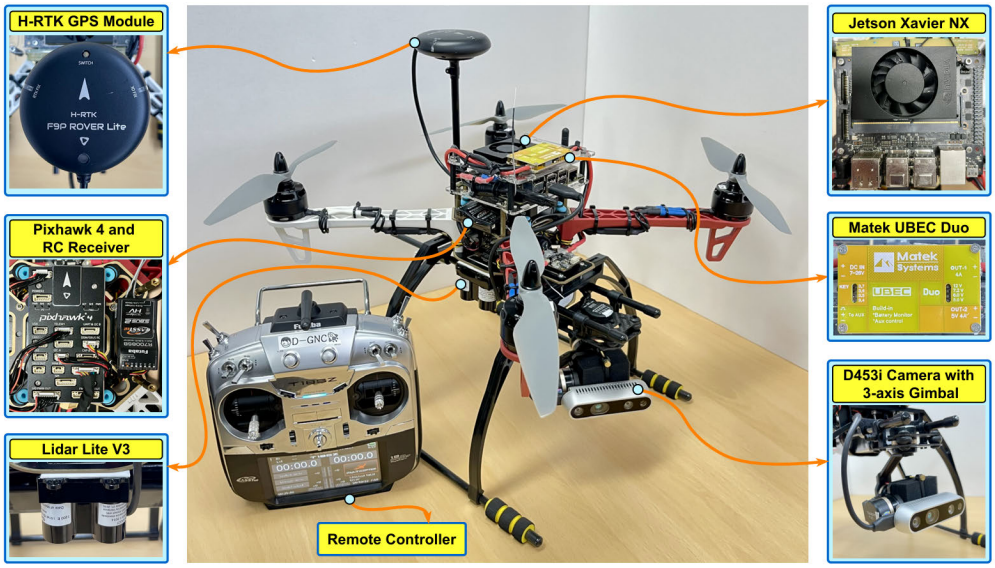


FIGURE 17. The UAV platform used in this study.

against some of the leading methods in both prior-based and deep learning-based categories. Deep learning-based methods excel in accuracy but require extensive training data and high computational costs, while prior-based approaches offer better interpretability and efficiency despite lower accuracy in complex scenarios. As demonstrated by experimental results, the proposed RSVT prior can surpass existing traditional methods in performance while maintaining computational efficiency.

Additionally, the comparisons in terms of visibility obtained by our proposed framework and the other prior-based methods on the considered datasets are also given in Fig.s 11-14. These visual results demonstrate that our proposed method is capable of producing visually appealing haze-free images without halo artifacts or color distortion. Notably, the sky regions restored by our proposed method can be considered significantly improved when compared to those produced by the other approaches. It implies that our proposed prior effectively addresses the color distortion problem in the bright regions, which is a common issue in most other prior-based algorithms. Moreover, the process of RSVT, which is independent of the haze imaging model, allows the proposed method to eliminate the influence of transmission lower bound for the sky regions. As a result, the soft segmentation mask can be utilized as weight to emphasize the transmission in the sky areas. The restored transmission maps shown in Fig.s 11-14 indicate that the depth information provided by our proposed method, particularly in the sky regions, is more accurate when compared against those given by the other prior-based methods under consideration. In addition, typical comparisons between the outputs yielded by our method and various deep learning-based models are also depicted in Fig. 15 and Fig. 16. These comparisons clearly show that the proposed approach achieves competitive visual effects when compared with deep learning-based methods for both synthetic and natural hazy scenes.

TABLE 4. Average processing time of various dehazing methods (in seconds).

Method	Platform	Runtime (sec.)	
		600×450	1080×720
CEP [3]	Python (CPU)	**0.101**	**0.182**
DCP [2]	Python (CPU)	0.112	0.204
BCCR [12]	Python (CPU)	0.445	0.865
NLID [4]	Python (CPU)	1.675	3.565
CAP [13]	Python (CPU)	0.689	1.243
AOD-Net [17]	PyTorch (CPU)	0.127	0.242
GCANet [20]	PyTorch (CPU)	0.209	0.385
YOLY [21]	PyTorch (CPU)	5.126	8.542
D^4+ [23]	PyTorch (CPU)	0.266	0.835
SDDN-T [24]	Tensorflow (CPU)	0.132	0.642
RSVT (ours)	Python (CPU)	0.231	0.364

E. EXECUTION SPEED

Efficiency is also considered a crucial factor when evaluating an image enhancement method. It partially reflects the suitability of the algorithm for real-world applications such as drone's autonomous navigation or target detection and tracking. In order to assess the efficiency of our method, a comparative analysis with several typical dehazing approaches in terms of average execution time has been conducted, with the results reported in Table 4. As summarized in Table 4, the proposed dehazing framework takes approximately 0.231 and 0.364 seconds to process a 600 × 450 image and a 1080 × 720 image, respectively, with an Intel(R) Core(TM) i5-8600K CPU @ 3.60GHz. As compared against the other prior-based algorithms, even though the proposed approach

FIGURE 18. Visual dehazing results of various approaches on real-world collected image data.

may not be considered the fastest method, it can manage to achieve a more desirable trade-off between effectiveness and efficiency. Particularly, when compared with DCP or CEP, our method achieves notable improvement in effectiveness while only slightly increasing the processing time. In addition, our method can also be compared competitively with other lightweight deep learning-based approaches such as GCANet while noting that our proposed method does not require an extensive training process on any database. In conclusion, the proposed algorithm shows remarkable efficiency, making it a feasible choice for real-world applications.

F. CASE STUDY

To further assess the applicability of the proposed dehazing framework in real-world scenarios, a case study was conducted using a surveillance UAV in a realistic scenario. This study aimed to assess both the effectiveness and efficiency of the framework in a localized area of Seoul, South Korea, during foggy conditions.

The information of the UAV platform used in this study is presented in Fig. 17. The surveillance quadcopter was operated by an onboard flight computer unit (FCU) Pixhawk 4. This system measures altitude using the Lidar Lite V3 laser-ranging sensor, while positioning is done using the F9P Rover Lite real-time kinematic positioning GPS receiver. Additionally, the power supply system consists of a 4-cell, 14.8 V, 10,000 mAh Li-Po battery coupled with a Matek UBEC Duo power regulator. The remote piloting capability is enabled using a radio frequency transmitter and receiver set. Data collection occurred at altitudes between 80 and 120 meters, allowing for sufficient resolution across diverse scenes such as roads, buildings, and vehicles. The qualitative performance of various dehazing algorithms was initially analyzed using a set of realistic hazy images. As illustrated in Fig. 18, the proposed framework's visual results were compared to those of several other methods. The outcomes indicate that the proposed approach delivers visually compelling results in comparison to other schemes. Specifically, methods like DCP and BCCR struggled to process sky regions, GCANet and D4+ exhibited color distortion issues, whereas the proposed framework demonstrated competitive performance with the recent SDDN-T model. It is important to note that the proposed framework does not aim to completely eliminate haze in distant scenes but intentionally retains a slight amount. This is not a shortcoming but is intended to preserve the natural appearance of the restored image.

Additionally, the efficiency of the dehazing framework within the UAV system was examined. The framework was implemented on the companion computer, and its processing speed was measured. The proposed RSVT framework achieved an average processing time of 0.519 seconds per 1080×720 image, corresponding to a frame rate of approximately 2 FPS, which can be considered adequate for UAV-based surveillance tasks.

VI. CONCLUSION

In this paper, an innovative prior for image dehazing, called Regional Saturation-Value Transition (RSVT), is proposed. The prior is derived based on two key observations made in the sky and bright areas of hazy-clean image pairs. First, the difference in hue components of a pair of hazy and haze-free points is trivial, indicating that haze has a negligible impact on the hue channel. Second, in the 2D coordinate system formed by the saturation and value channels, most of the lines passing through corresponding pairs of hazy-clean points intersect

around the atmospheric light coordinates. This suggests that haze removal can be performed by appropriately translating hazy points in the saturation-value coordinate system. The proposed prior is combined with image decomposition and dark channel prior method to form a unified dehazing framework. A novel morphological min-max channel is introduced for image decomposition and estimation of the global atmospheric light. Experimental results show that the proposed framework can effectively handle diverse hazy scenes, restoring the natural appearance of the sky, with a case study highlighting its practicality in real-world scenarios. Future work will focus on enhancing the framework's performance and applicability by addressing more complex scenarios.

ACKNOWLEDGMENT

(Truong-Dong Do and Le-Anh Tran contributed equally to this work.)

REFERENCES

[1] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "RefineDNet: A weakly supervised refinement framework for single image dehazing," *IEEE Trans. Image Process.*, vol. 30, pp. 3391–3404, 2021.

[2] 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.

[3] T. M. Bui and W. Kim, "Single image dehazing using color ellipsoid prior," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 999–1009, Feb. 2018.

[4] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.

[5] Z. Shi, J. Long, W. Tang, and C. Zhang, "Single image dehazing in inhomogeneous atmosphere," *Optik*, vol. 125, no. 15, pp. 3868–3875, Aug. 2014.

[6] W. Wang, X. Yuan, X. Wu, and Y. Liu, "Dehazing for images with large sky region," *Neurocomputing*, vol. 238, pp. 365–376, May 2017.

[7] S. Salazar-Colores, E. U. Moya-Sánchez, J.-M. Ramos-Arreguín, E. Cabal-Yépez, G. Flores, and U. Cortés, "Fast single image defogging with robust sky detection," *IEEE Access*, vol. 8, pp. 149176–149189, 2020.

[8] F. Yu, C. Qing, X. Xu, and B. Cai, "Image and video dehazing using view-based cluster segmentation," in *Proc. Vis. Commun. Image Process. (VCIP)*, Nov. 2016, pp. 1–4.

[9] L.-A. Tran, D.-H. Kwon, and D.-C. Park, "Single image dehazing via regional saturation-value translation," in *Proc. Int. Conf. Ind. Sci. Comput. Sci. Innov.*, vol. 237, Jan. 2024, pp. 517–524.

[10] L.-A. Tran, S. Moon, and D.-C. Park, "A novel encoder–decoder network with guided transmission map for single image dehazing," *Proc. Comput. Sci.*, vol. 204, pp. 682–689, Jun. 2022.

[11] W. Middleton and V. Twersky, "Vision through the atmosphere," *Phys. Today*, vol. 7, no. 3, p. 21, 1954.

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

[13] 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.

[14] A. Nguyen and B. Le, "3D point cloud segmentation: A survey," in *Proc. 6th IEEE Conf. Robot., Autom. Mechatronics (RAM)*, Nov. 2013, pp. 225–230.

[15] Y. Liu, H. Li, and M. Wang, "Single image dehazing via large sky region segmentation and multiscale opening dark channel model," *IEEE Access*, vol. 5, pp. 8890–8903, 2017.

[16] O. Liba, Y. Movshovitz-Attias, L. Cai, Y. Pritch, Y.-T. Tsai, H. Chen, E. Eban, and J. T. Barron, "Sky optimization: Semantically aware image processing of skies in low-light photography," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2230–2238.

[17] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4770–4778.

[18] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Jan. 2016, pp. 154–169.

[19] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[20] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1375–1383.

[21] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1754–1767, May 2021.

[22] L.-A. Tran and D.-C. Park, "Encoder–decoder networks with guided transmission map for effective image dehazing," *Vis. Comput.*, vol. 2024, pp. 1–24, Apr. 2024.

[23] Y. Yang, C. Wang, X. Guo, and D. Tao, "Robust unpaired image dehazing via density and depth decomposition," *Int. J. Comput. Vis.*, vol. 132, no. 5, pp. 1557–1577, May 2024.

[24] L.-A. Tran and D.-C. Park, "Lightweight image dehazing networks based on soft knowledge distillation," *Vis. Comput.*, vol. 2024, pp. 1–20, Sep. 2024.

[25] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[26] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[27] S.-C. Huang, D.-W. Jaw, W. Li, Z. Lu, S.-Y. Kuo, B. C. M. Fung, B.-H. Chen, and T. Numnonda, "Image dehazing in disproportionate haze distributions," *IEEE Access*, vol. 9, pp. 44599–44609, 2021.

[28] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U^2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.

[29] L. Tao, C. Zhu, J. Song, T. Lu, H. Jia, and X. Xie, "Low-light image enhancement using CNN and bright channel prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3215–3219.

[30] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE J. Solid-State Circuits*, vol. 23, no. 2, pp. 358–367, Apr. 1988.

[31] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 120–129, Jan. 2015.

[32] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, Sep. 1987.

[33] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

[34] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 825–833.

[35] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[36] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[37] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993.

[38] D. M. Titterington, A. F. Smith, and U. E. Makov, "Statistical analysis of finite mixture distributions," 1985.

TRUONG-DONG DO (Graduate Student Member, IEEE) received the B.Eng. degree in automation and control engineering from the HCMC University of Technology and Education, Vietnam, in 2018, and the M.S. degree in computer science from the Department of Artificial Intelligence Convergence, Chonnam National University, South Korea, in 2021. He is currently pursuing the Ph.D. degree in aerospace engineering with Sejong University, Seoul, South Korea. His research interests include deep learning, computer vision, robust control, robotics, and autonomous vehicles.

LE-ANH TRAN (Member, IEEE) received the Ph.D. degree in electronics engineering from Myongji University, South Korea, in 2024. From 2018 to 2019, he was an AI Engineer at the Department of AI Transformation, FPT Software Company Ltd., Ho Chi Minh City, Vietnam. Since 2020, he has been a Computer Vision Scientist and a Research Staff Member at Mindintech Inc., Seoul, South Korea. His research interests include computer vision (autonomous driving), neural networks, pattern recognition, cluster analysis, and image enhancement.

SEOKYONG MOON received the B.S. degree in electrical engineering from Yonsei University, Seoul, South Korea, in 1986, the M.S. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 1988, and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, WA, USA, in 1995. In 1995, he joined the Speech Research Department, AT&T Bell Laboratories, Murray Hill, NL, USA, where he was involved in research work on fast speech recognition. He was a Senior Research Engineer at the CDMA Network Department, Samsung Electronics, Seoul. In 1997, he joined as the Founder and the CEO of Infinity Telecom Company Ltd., where he developed and commercialized the world's smallest voice recognition engine for Samsung mobile phone, in 2002. He was selected for best partner and first vendor at Samsung Electronics, in 2004. Since 2018, he has been the CTO at MindInTech Inc., specializing in deep learning technology.

JIO CHUNG received the B.E. and M.E. degrees from the Department of Electronics and Control Engineering, Myongji University, Yongin, South Korea, in 1997 and 1999, respectively. From 1999 to 2018, he was with the Intelligence Communication Network (ICN) Laboratory, Infinity Telecom Company Ltd., Seoul, South Korea. Since 2018, he has been with the Research Center, Mindintech Company Ltd., Seoul. He is currently with the AI Group. His research interests include deep learning algorithms in object detection, lane detection, and pattern recognition.

NGOC-PHI NGUYEN received the B.Sc. and M.Sc. degrees in mechatronics engineering from the HCMC University of Technology and Education, Vietnam, in 2012, the M.Sc. degree from Vietnamese–German University, Vietnam, in 2015, and the Ph.D. degree in aerospace engineering from Sejong University, Seoul, South Korea, in 2020. He was an Assistant Professor with the Department of Aerospace Engineering, from March 2020 to August 2024. He is currently a Postdoctoral Researcher with the Drone Mechatronics Team to develop robust control algorithms for single or multiple drones. His research interests include fault-tolerant control, nonlinear control, intelligent control, and formation control.

SUNG KYUNG HONG received the B.S. and M.S. degrees in mechanical engineering from Yonsei University, Seoul, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 1998. From 1989 to 2000, he was with the Flight Dynamics and Control Laboratory and the Unmanned Aerial Vehicle System Division, Agency for Defense Development, South Korea. He is currently a Full Professor with the Department of Aerospace Engineering, Sejong University, South Korea. His research interests include fuzzy logic controls, inertial sensor applications, and flight control systems.