

An efficient Video Dehazing Algorithm Based on Spectral Clustering

Fan Zhao, Zao Yao, Xiaofang Song and Yi Yao

Department of Information Science, Xi'an University of Technology, Xi'an, China, 710048

[e-mail: vcu@xaut.edu.cn]

*Corresponding author: Fan Zhao

*Received January 18, 2016; revised April 22, 2017; accepted March 14, 2018;
published July 31, 2018*

Abstract

Image and video dehazing is a popular topic in the field of computer vision and digital image processing. A fast, optimized dehazing algorithm was recently proposed that enhances contrast and reduces flickering artifacts in a dehazed video sequence by minimizing a cost function that makes transmission values spatially and temporally coherent. However, its fixed-size block partitioning leads to block effects. The temporal cost function also suffers from the temporal non-coherence of newly appearing objects in a scene. Further, the weak edges in a hazy image are not addressed. Hence, a video dehazing algorithm based on well designed spectral clustering is proposed. To avoid block artifacts, the spectral clustering is customized to segment static scenes to ensure the same target has the same transmission value. Assuming that edge images dehazed with optimized transmission values have richer detail than before restoration, an edge intensity function is added to the spatial consistency cost model. Atmospheric light is estimated using a modified quadtree search. Different temporal transmission models are established for newly appearing objects, static backgrounds, and moving objects. The experimental results demonstrate that the new method provides higher dehazing quality and lower time complexity than the previous technique.

Keywords: Video dehazing, spectral clustering, transmission estimation, atmospheric light, spatial consistency, temporal consistency

This work is supported in part by the National Natural Science Foundation of China (NSFC) Project No. 61671376.

1. Introduction

With the development of industry, the warming of the climate, the accelerated process of urbanization and the increasing of population density, the haze weather have been produced in many parts of the world. Haze weather not only affects people's physical health, but also has become the invisible killer of video surveillance. Haze removing is the technology which eliminates the impact of haze from the image or video, and restores the clear image. Because the dehazed image contains more information and is more visually pleasing, restore the true image from the fog-impaired version has important academic and practical significance.

For the natural scene images, the fog density is closely related to the depth information of the scene. Early haze removal methods mainly rely on additional depth information or multiple observations of the same scene [1-7]. These methods are relatively stable by using additional information such as depth maps or optical transmission maps. But the depth information of the scene is always unknown, which needs to be manually obtained and thus brings a certain degree of difficulty and challenge to the dehazing algorithm. The visible light has a wavelength range of 400-700nm whereas, NIR has a wavelength ranging 700-1100nm. Because of the higher wavelength, NIR light is scattered less than visible light and can capture more details of distant objects, image dehazing using visible and near-infrared (NIR) image pair has seen growing interest in last decade for improving visibility in landscape photographs [8-12]. Usually these frameworks exploit the dissimilarity between visible and NIR images for atmospheric light or depth map estimation. However, with the constraints of algorithm complexity and the performance of digital camera modules, they are not widely used in real-time video dehazing applications.

In recent years, image dehazing using single visible image has been extensively studied to overcome the limitations of multi-image de-fogging methods. These algorithms take advantage of strong assumptions or limitations to remove fog in an image. By assuming that the surface shading of the objects and the scene albedo are locally statistically uncorrelated, Fattal et al. de-fog the hazy image based on the local statistics in [13]. The actual scene is restored by dividing the hazy image into regions of constant albedo, which result in sufficient color information is required. But when objects are far from the camera, there exist a certain degree of mist. By assuming that haze-free image has a higher contrast, Tan et al. enhance the hazy image by maximizing local contrast [14]. Although this method is successful in regions with very dense haziness, the color of the haze-free image is often oversaturated. By describing an image as a factorial Markov random field, in which the scene albedo and depth are two statistically independent latent layers, Kratz et al. implement a canonical expectation maximization algorithm to factorize the image and thus recover haze-free images with fine edge details [15]. The output images by Kratz et al.'s algorithm sometimes are over-enhanced. Tarel et al. propose a novel single image dehazing algorithm based on a new filter which preserves edges and corners with obtuse angle as an alternative to the median filter [16]. The main advantage of this algorithm is its

speed since its complexity is only a linear function of the input image size.

In recent years, an effective single visible image haze removal method using the dark channel prior has achieved great progress in theory and practice [17]. The dark channel prior is based on the following observation about outdoor haze-free images: in most of the non-sky patches at a given image, the pixel intensities are very low on at least one color channel, some even close to 0. Defined as the color of the most haze-opaque region, the atmospheric light A plays a very important role in the dehazing effect. He et al. are the first to estimate the atmospheric light based on dark channel prior. The top 0.1% brightest pixels in the dark channel, which are usually the most haze-opaque, are first selected. Among these pixels, the pixels with the highest intensity in the image are selected as the atmospheric light. Second, the transmission is estimated depending on the value of the atmospheric light. This atmospheric light estimation method has very low complexity. However, objects that are brighter than the atmospheric light may lead to an undesirable selection of the atmospheric light. Therefore, many researchers have improved atmospheric light estimation methods based on He et al.'s method. In order to obtain more accurate values of atmospheric light, Xiang et al. select the 0.2% brightest pixels in the dark channel of the input image [18], and Wang et al. propose a single image dehazing method that is based on a physical model and the dark channel prior principle [19]. The atmospheric light value is estimated based on a variogram, which slowly weakens areas in the image that do not conform to the dark channel prior. This method overcomes the deficiency of the dark channel prior, and reduces the influence of white objects or sky areas on the whole image. The atmospheric light is estimated based on the bright channel prior of an input image, which corresponds to the dark channel prior, in [20]. Morphological processing is performed on gray-scale images to eliminate the effect of white objects in [21], but the method tends to corrupt edges. Kumari et al. first remove larger areas of white objects using a gray-scale erosion and dilation operation to calculate the atmospheric light. The gray-scale opening operation smooths the gray-scale and the white target is eliminated, but it also blurs edges. Shin et al. propose an approach to calculate atmospheric light values based on adaptive temporal averaging between adjacent frames to avoid flickering artifacts in a video sequence [22].

Although many dark channel prior-based dehazing algorithms have achieved good performance, these dehazing algorithms still have some problems. The first problem is that an image dehazed using the dark channel prior may suffer color bias to a certain extent. One of the causes is that optical transmission is wavelength-dependent. Because of Rayleigh's law, scattering is more intense in the blue band, which makes the dehazed image appear to have deeper bluish hue. Second, when some objects are brighter than the atmospheric light, prior-based dehazing algorithms fail to estimate the transmission correctly. Because the white regions in the image do not meet the dark channel prior assumptions, the hue of these regions is a relatively deeper yellow hue after dehazing. Third, a series of algorithms based on the dark channel prior are very sensitive to adjustable parameters such as window size and transmission estimation threshold selection. The proper values of all these parameters must be set by experience to achieve a good dehazing effect. Fourth, to extend an image dehazing algorithm to video, it is necessary to address

challenges such as computational efficiency, spatial consistency, and temporal coherence, which is not a trivial task [23–30].

The amount of video data in a sequence is much larger than that of a single image, so the computational efficiency must be improved to ensure real-time video dehazing. To maintain spatial consistency, it is necessary to handle the intra-frame discontinuities in the video. Maintaining the temporal coherence is also an important element for video haze removal. If the continuity between the video frames is not considered and the image dehazing algorithm is just applied to the video frame-by-frame, it may break the temporal coherence and produce a restored video with severe flickering artifacts. In order to solve the problem of inter-frame temporal coherence, Zhang et al. [23] first extract the transmission map frame-by-frame using the dark channel prior model, then estimate the forward and backward optical flow between two neighboring frames to find matching pixels. Next, flow fields are used to build an MRF model to refine the transmission map, which is ensured to improve the spatial and temporal coherence. Xiang et al. [18] use the dark color prior to obtain the transmission map of each frame image, and then use a block matching technique to match the corresponding blocks between adjacent frames. They then use a weighted mean filter to refine the final transmission map. Optical flow fields or block matching technology can ensure the continuity of video frames to a certain extent, but their computational complexity is high.

Kim et al. propose a method based on a cost function to optimize video dehazing that can reduce the loss of information as much as possible while increasing the contrast of the image [24]. First, by exploiting the fact that the variance of pixel values is generally low in hazy regions, e.g., sky, they propose a hierarchical search method based on a quad-tree subdivision to choose the atmospheric light that is as bright as possible. Second, in order to maintain spatial consistency of the video frame, they use an intra-frame cost function to calculate transmission based on a fixed-size block segmentation. Finally, the similarity probability model between the adjacent frames of the pixels at the same location is established and solved to ensure temporal coherence between adjacent frames. Kim et al.'s algorithm is capable of removing haze effectively and restoring images faithfully, even in real time. However, there are still some problems yet to be resolved. The partition mode of fixed-size blocks may lead to the same target being separated into different blocks or different targets can be merged in the same block, which violates the assumption that the scene radiance of the same object is invariant between adjacent frames. Hence, the restored image resulted by the fixed block processing in the transmission estimation is not sufficiently natural. The temporal coherence of a newly appearing object or a quickly moving target is still not retrieved using the same position between adjacent frames, which leads to deviation in the transmission estimation.

In this work, we propose a fast video dehazing algorithm based on spectral clustering. The proposed algorithm is based on Kim et al.'s method. To avoid block artifacts and improve the subjective appearance, spectral clustering is customized to segment static scenes in order to ensure the same targets are partitioned into the correct category. The transmission rate of each category is calculated assuming that all rates are all equal. To alleviate blurring of the restored image caused by video dehazing, we add an edge intensity

function to Kim et al.'s total cost function because the edge is clearer after dehazing. Furthermore, for video dehazing, assuming that the scene radiance of an object is highly consistent along its motion trajectory, we use the corresponding position rather than the same position in the temporal coherence cost function. The experimental results demonstrate that the restoration performance of the proposed method yields higher quality images and is more robust than the approach of Kim et al.

This paper is organized as follows: Section 2 briefly reviews the dehazing principles and related technologies. Section 3 presents the proposed dehazing algorithm. The experiment results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

2.1 Dehazing principles

The haze image model is widely used in computer vision [13,14,17,24] and is described as follows:

$$J(x) = (J_r(x), J_g(x), J_b(x))^T \quad (1)$$

where $J(x) = (J_r(x), J_g(x), J_b(x))^T$ and $I(x) = (I_r(x), I_g(x), I_b(x))^T$ denote the original and the observed (r, g, b) colors at pixel position x , respectively, and $A = (A_r, A_g, A_b)$ is the global atmospheric light that represents the ambient light in the atmosphere. $t(x)$ is the transmission of the reflected light, which is between 0 and 1 and determined by the distance between the scene point and the camera. For deriving the original image $J(x)$, which is desired to be obtained by haze-free techniques, under $I(x)$ being known, the atmospheric light A and transmission $t(x)$ are needed to be properly estimated.

2.2 Related dehazing technology

The dehazing technologies, which only use a single foggy image instead of additional depth information or multiple observations of the same scene as input, have important significance in practical application. Two representative algorithms are introduced below in detail.

2.2.1 Dehazing algorithm based on dark channel prior

He et al. [17] first propose the dark channel prior by the statistics of the haze-free outdoor images. The rule can be described as: in most of the non-sky patches, at least one color channel has some pixels whose intensities are very low. For image $J(y)$, the dark prior can be describe as:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r, g, b\}} J^c(y)) \rightarrow 0 \quad (2)$$

where J^c is a color channel of $J(y)$ and $\Omega(x)$ is a local patch centered at x . J^{dark} is the dark channel map of image $J(y)$.

Defined as the color of the most haze-opaque region, the atmospheric light, A , plays a very important role in the dehazing effect. A number of atmospheric light estimation methods have been proposed based on dark channel prior [18-21]. After the atmospheric light A is estimated, the restored scene radiance $J(y)$ depends on the selection of the transmission t . He et al. used the atmospheric scattering model of Eq. (1) to estimate the transmission t .

First, the haze imaging equation (1) is first normalized by A :

$$\frac{I^c(x)}{A^c} = t(x) \frac{J^c(x)}{A^c} + 1 - t(x) \quad (3)$$

Note that each color channel is normalized independently.

On assuming that the transmission $\tilde{t}(x)$ in a local patch $\Omega(x)$ is constant, Eq. (4) is derived by applying the two minimum operators to both sides of Eq.(4).

$$\min_{y \in \Omega(x)} \left(\min_c \frac{I^c(y)}{A^c} \right) = \tilde{t}(x) \min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right) + 1 - \tilde{t}(x) \quad (4)$$

As a haze-free image, the dark channel of J is close to zero according to the dark channel prior:

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_c J^c(y) \right) = 0 \quad (5)$$

As A^c is always positive, this leads to

$$\min_{y \in \Omega(x)} \left(\min_c \frac{J^c(y)}{A^c} \right) = 0 \quad (6)$$

So the transmission $\hat{t}(x)$ is estimated simply by:

$$\hat{t}(x) = 1 - \min_{y \in \Omega(x)} \left(\min_c \frac{I^c(y)}{A^c} \right) \quad (7)$$

Dark channel prior theory can be widely applied to non-sky areas, but the required conditions do not always hold for areas of sky. But fortunately, the color of the sky in a hazy image I is usually very similar to the atmospheric light A . So, the prior of the sky region is proposed by Wang et al. [19] as:

$$\min_{y \in \Omega(x)} \left(\min_c \frac{I^c(y)}{A^c} \right) \rightarrow 1$$

The estimated \hat{t} by Eq. (7) for the sky regions is indeed close to zero, which conforms to the observation that the sky is infinitely distant and its transmission is indeed close to zero, so Eq. (7) gracefully handles both sky and non-sky regions and does not need to separate the sky regions beforehand.

In practice, even on clear days, the atmosphere is not absolutely free of particles. Hence, haze still exists when we look at distant objects. Moreover, the presence of haze is a fundamental depth scene cue for humans [25-26]. This phenomenon is called aerial perspective. For this reason, He et al. introduced a constant parameter ω ($0 < \omega < 1$) into Eq. (7):

$$\hat{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left(\min_c \frac{I^c(y)}{A^c} \right) \quad (8)$$

where the value of ω is generally set to 0.95.

The estimated transmission map $\hat{t}(x)$ is refined by applying the soft matting method [27]. Since the complexity of this algorithm is high, and then a variety of improved methods, such as the bilateral filtering algorithm and oriented filtering, are put forward in [17, 24, 28-29].

After the atmospheric light A and the transmission map $\hat{t}(x)$ are found, the final dehazed image is can be recovered according to Eq.(1) as:

$$J(x) = \frac{I(x) - A}{\hat{t}(x)} + A \quad (9)$$

When the transmission $\hat{t}(x)$ is close to zero, the directly recovered scene radiance J is prone to containing noise. Hence, Wang et al. [19] restricted the transmission by a lower bound t_0 to preserve a small amount of haze in the sky area. The final scene radiance $J(x)$ is recovered by

$$J(x) = \frac{I(x) - A}{\max(\hat{t}(x), t_0)} + A \quad (10)$$

where t_0 is normally set to 0.1.

2.2.2 Dehazing algorithm based on cost function

He et al.'s dark channel prior has been adopted and improved in many algorithms, but these algorithms suffer from huge computational complexity and flicker effects on restored videos if they are simply extended to video images [23,30]. To ensure time continuity between frames while achieving a high computation speed, a fast dehazing algorithm for images and videos based on the optimized contrast enhancement is proposed by Kim et al. [24]. Assuming that a haze-free image has a higher contrast ratio than the hazy image, Kim et al. aim to maximize the contrast meanwhile alleviating the information loss in static image. The main steps of this algorithm are described as follows. Firstly, by recursively dividing an image into four smaller regions and selecting the region with the brightest pixel

value, the atmospheric light is chosen within the optimal region. Secondly, the image is divided into non-overlapping 32×32 sized blocks. On assuming that scene depths are locally similar, a cost function E is designed for each block to maximize the contrast and alleviate the information loss simultaneously. A single transmission value for each block B is found by optimizing the designed cost function. The block-based transmission values are then refined into the pixel-based ones using an edge preserving filter. Finally, given the transmission map and the atmospheric light, the scene radiance is restored from the input hazy image. For video dehazing, Kim et al. further add the temporal coherence cost $E_{temporal}$ to E to ensure the time continuity between frames. The specific ideas of this algorithm are described as follows.

A. Static image dehazing

To not only enhance contrast but also reduce information loss, Kim et al. design a contrast cost function $E_{contrast}$ and an information loss cost function E_{loss} for each divided block B of an image, and then minimize the two cost functions simultaneously.

Contrast cost $E_{contrast}$ is defined as the negative sum of the mean square error (MSE) contrasts for three color channels of each block B as follows:

$$E_{contrast} = - \sum_{c \in \{r, g, b\}} \sum_{P \in B} \frac{\left(J_c(P) - \bar{J}_c \right)^2}{N_B} = - \sum_{c \in \{r, g, b\}} \sum_{P \in B} \frac{\left(I_c(P) - \bar{I}_c \right)^2}{t^2 N_B} \quad (11)$$

where \bar{J}_c and \bar{I}_c are the average values of $J_c(p)$ and $I_c(p)$ in B , respectively, and N_B is the number of the pixels in block B . In fact, minimizing $E_{contrast}$ equals maximizing the MSE contrast.

Information lost cost E_{loss} for block B is defined as the squared sum of truncated values.

$$E_{loss} = - \sum_{c \in \{r, g, b\}} \sum_{P \in B} \left\{ \left(\min\{0, J_c(P)\} \right)^2 + \left(\max\{0, J_c(P) - 255\} \right)^2 \right\} \quad (12)$$

Here, the terms $\min\{0, J_c(p)\}$ and $\max\{0, J_c(p) - 255\}$ denote the truncated values due to the underflow and the overflow, respectively.

For block B , the optimal transmission t^* is derived by minimizing the overall cost function E .

$$E = E_{contrast} + \lambda_{loss} E_{loss} \quad (13)$$

where λ_{loss} is a weighting parameter that controls the relative importance of the contrast cost and the information loss cost.

For estimating the transmission more reliably, a single constraint is added to the optimizing cost function:

$$t \geq \max \left\{ \min_{c \in \{r, g, b\}} \min_{p \in B} \left\{ \frac{I_c(p) - A_c}{-A_c} \right\}, \max_{c \in \{r, g, b\}} \max_{p \in B} \left\{ \frac{I_c(p) - A_c}{255 - A_c} \right\} \right\} \quad (14)$$

The first term in Eq.(14) is the same constraint that is employed as the dark channel prior in the He et al.'s algorithm [17]. However, when some objects are brighter than the atmospheric light, He et al.'s algorithm fails to estimate the transmission correctly. The second constraint in Eq.(14) is to prevent the overflow of restored pixel values.

B. Video dehazing

By minimizing the cost function in Eq.(13), the optimal transmission t_0^* for each block in the first frame is found. Assuming that the scene radiance of an object point is invariant between adjacent frames, a temporal coherence cost $E_{temporal}$ is added to the total cost function E in Eq. (13). For subsequent frames, the optimal transmission t_k^* of each block is obtained by minimizing the augmented cost function

$$E = E_{contrast} + \lambda_{loss} E_{loss} + \lambda_T E_{temporal} \quad (15)$$

where λ_T is a weighting parameter which is determined by considering the trade-off between flickering artifacts and the qualities of individual frames. The temporal coherence cost $E_{temporal}$ is defined as follows.

First, based on the differential image between the two frames, a simple probability model $\omega_k(p)$ is introduced to represent the likelihood that two pixels match.

$$\omega_k(p) = \exp \left(- \frac{\left(I_Y^k(p) - I_Y^{k-1}(p) \right)^2}{\sigma^2} \right) \quad (16)$$

where $I_Y^k(p)$ is the pixel value of Y component at p point in k th frame. σ controls the variance of the probability model and is empirically set to 10.

Second, assuming that a linear relationship exists between the transmission $t_k(p)$ in the current frame and the transmission $t_{k-1}(p)$ in the previous frame, the temporal coherence factor $\tau_k(p)$ is given by

$$\tau_k(p) = \frac{I_Y^k(p) - A_Y}{I_Y^{k-1}(p) - A_Y} \quad (17)$$

Then, the temporal coherence factor $\bar{\tau}_k$ for block B is then defined as

$$\overline{\tau}_k(p) = \frac{\sum_{p \in B} \omega_k(p) \tau_k(p)}{\sum_{p \in B} \omega_k(p)} \quad (18)$$

The block similarity between the two frames is defined as

$$\overline{\omega}_k = \frac{1}{N_B} \sum_{p \in B} \omega_k(p) \quad (19)$$

Finally, the temporal coherence cost is defined as

$$E_{temporal} = \overline{\omega}_k \left(t_k - \overline{\tau}_k t_{k-1} \right)^2 \quad (20)$$

An additional weight $\overline{\omega}_k$ is introduced into the squared difference term in the $E_{temporal}$ to overcome the problem that the squared difference cannot exactly reflect the similarity between the corresponding blocks in the current and previous frames when a scene change occurs or a new object appears. By minimizing total cost function E , the optimal transmission is obtained, which not only reduces the flickering artifacts in a dehazed video sequence by making transmission values temporally coherent but also enhances the contrast. In addition, based on the fact that the variance of pixel values is generally low in hazy regions, e.g. sky, Kim et al. estimate an atmospheric light for a video by a quad-tree subdivision hierarchical searching method, which can avoid the influence of the brighter object on the selection of atmospheric light. Experimental results demonstrated that Kim et al.'s algorithm can effectively removes haze and is sufficiently fast for real-time dehazing applications.

Kim et al.'s algorithm still has some problems: the fixed-size block partitioning mode means that the same target can be divided into different blocks or different targets are grouped in the same block. However, in reality, the transmissions of the same target in a frame or at the two consecutive frames are same or similar. If fixed-size blocks are used, the estimated transmission of the same target will inevitably become biased, resulting in a restored image is not sufficiently real. In addition, the linear relationship of corresponding block transmissions between consecutive frames does not hold when a scene change occurs or a new object appears. Hence, the temporal coherence function is not suitable in this case. We instead adopt some mechanisms to improve the accuracy of the subsequent transmission estimation. First, to ensure that the same target has the same or similar transmission, a reliable segmentation of the static scene is done by spectral clustering. Second, when a scene change occurs or new object appears, the temporal coherence cost is omitted from the total cost function to avoid unnecessary calculations because the premise does not hold at this time. Kim et al.'s temporal coherence function is further simplified to adapt to the case when temporal coherence exists between consecutive frames. Third, assuming that a haze-free image has stronger edges than a hazy one, we add an edge enhancement term to the total cost function for image deblurring. The specific algorithm is described in detail below.

3. The Framework of Dehazing Algorithm Based on Spectral Clustering

3.1 The Framework of Dehazing Algorithm

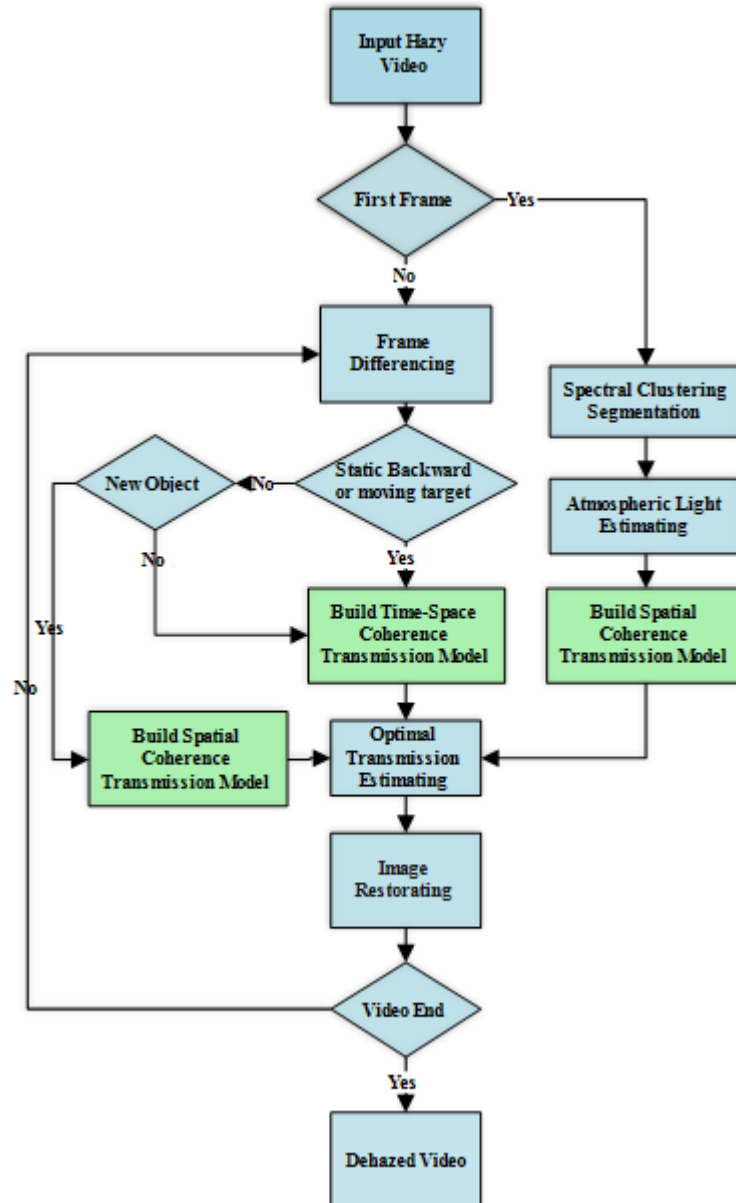


Fig. 1. The diagram of the proposed video dehazing

Fig. 1 shows the block diagram of the proposed dehazing algorithm. Using spectral clustering, the first frame of the input hazy video is segmented into different classes. The spectral clustering results are used to estimate atmospheric light A first and then establish the spatial coherence transmission model. Because the proposed dehazing algorithm is based on a static background, the spectral clustering segmentation results and atmospheric light obtained in the first frame can be applied to later frames. Afterwards, for subsequent images, the moving targets are detected using the inter-frame difference. The spatial consistency transmittance model is built for newly appearing targets without references in the previous frames. There exists a counterpart in the previous frame, the transmission model with time-space coherence is created for static background or a moving target that appears continuously. Assuming that scene depths are similar within a category, the class-based optimal transmissions are obtained from solving their respective models. These transmission values are further refined into the pixel-based ones by a guided filter [31]. Given the transmission map and the atmospheric light, the dehazed video can be restored from the input hazy images. The key technologies are described below.

3.2 Class Segmentation Based on the Spectral Clustering

The dehazing results of Kim et al.'s transmission estimation based on fixed block partitioning is shown in **Fig. 2**. The upper left red rectangle in the restored image indicates where a building and green belt below it are grouped into the same block, which results in a silvery mist over them. The lower right red rectangle shows an undesired block artifact resulting from the river being divided into different blocks. Transmission estimation based on fixed size block division destroys the spatial consistency of the image and adds pixel bias to the dehazed image. Hence, to improve the transmission estimation accuracy, a more accurate image segmentation method is necessary. Image segmentation essentially involves finding clusters in the high-dimensional space of the pixels. Compared to traditional algorithms such as K-means or single linkage, spectral clustering has many fundamental advantages [32]. It is very simple to implement and can be solved efficiently by standard linear algebra methods. A multiple time spectral clustering algorithm is used by Qin et al. to construct a basis matrix, and then regression-based foreground detection is executed using the background basis [33]. In this paper, we attempt to use a spectral clustering well designed to separate the static scene into different classes and then to estimate class-based transmissions. The principle of spectral clustering algorithm and its application of image segmentation are described below.



Fig. 2. Dehazy effect by using of Kim et al's transmission estimation based on fixed block partition

A. Spectral Clustering

Input : the segmentating image $I_{w \times h}$, number m of clusters to construct.

Step1 : Construct a weighted undirected graph $\mathcal{G} = (V, E)$ mapped from $I_{w \times h}$.

Step2 : Let $W_{n \times n}$ be weighted adjacency matrix of G , $n = w \times h$.

Step3 : Compute the normalized Laplacian matrix $L_{n \times n}$.

Step4 : Compute the first k eigenvectors v_1, v_2, \dots, v_k of $L_{n \times n}$.

Step5 : Let $H_{n \times k}$ be the matrix containing the vectors v_1, v_2, \dots, v_k as columns.

Step6 : For $i=1,2,\dots,n$, let $y_i \in R^k$ be the vector corresponding to the i -th row of H .

Step7 : Cluster the points $(y_i)_{i=1,2,\dots,n} \in R^k$ with the k -means algorithm into m clusters C_1, \dots, C_m .

Output : Clusters A_1, \dots, A_m with $A_i = \{j \mid y_j \in C_i\}$,

determine which class each pixel belongs to according to A_1, \dots, A_m .

Fig. 3. The flowchart of spectral clustering algorithm

The spectral clustering algorithm flowchart is listed in **Fig. 3**. The main steps can be summarized as follows.

Step1 : For the first frame $I_{w \times h}$ (w and h denote the width and height of the image, respectively), let $G = (V, E)$ be an undirected graph where the vertex set $V = \{v_1, v_2, \dots, v_n\}$, $n = w \times h$ stands for all pixels in $I_{w \times h}$ and the edge set E denotes the spatially neighboring pixel pair set.

Step2 : The weighted adjacency matrix $W = (w_{ij})_{i,j=1,2,\dots,n}$ of the graph G is constructed using a designed similarity function. The element w_{ij} represents the similarity between two vertices v_i and v_j . The similarity function is chosen based on the fact that the pixel differences in both position and luminance are very small within an object. To be more suitable for scene segmentation requirements, both the distance and luminance differences between two vertices v_i and v_j are considered simultaneously as follows:

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|_2^2}{\sigma_1^2}} \times \begin{cases} e^{-\frac{\|X(i)-X(j)\|_2^2}{\sigma_2^2}} & i \neq j \\ 0 & i = j \end{cases} \quad (21)$$

where $F(i)$ and $X(i)$ represent luminance and position of the i th pixel in image $I_{w \times h}$, respectively. σ_1 and σ_2 adjust the proportion of the two terms in Eq. (21), here they are set as $\sigma_1 = 0.3$, $\sigma_2 = 0.7$.

Step3: After the symmetric and non-negative matrix W is formulated, a diagonal matrix D is formed whose (i, i) th element is the sum of the W 's i th row element denoted as d_i .

$$D = \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & d_n \end{bmatrix} \quad (22)$$

The unnormalized graph Laplacian matrix L is defined as $L = D - W$, and the normalized Laplacian matrix L is derived by Eq.(23).

$$L = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}} (D - W) D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (23)$$

Step4: The objective of clustering is to separate pixels of image $I_{w \times h}$ into different groups according to their similarities. Given a graph G and W , the simplest and most direct way to construct a partition, formed of sets A_1, A_2, \dots, A_k , is as follows. If $A_i \cap A_j = \emptyset, i \neq j$ and $A_1 \cup A_2 \cup \dots \cup A_k = V$, then the pixels within the same cluster are similar to each other (which means that within-cluster similarity should be minimized) and pixels in different clusters are dissimilar from each other (which means that between-cluster similarity should be maximized). The normalized cut (Ncut) was proposed by Shi and Malik to solve the min-cut problem [34], which is NP-hard and highly complex. To implement the partition in practice, the first k eigenvectors of the normalized Laplacian matrix L are computed, which denote as v_1, v_2, \dots, v_k .

Step5: Let $H_{n \times k}$ be the matrix containing the vectors v_1, v_2, \dots, v_k as columns.

Step6: For $i = 1, 2, \dots, n$, let y_i be the vector which corresponds to the i th row element of H .

Step7: Each row of matrix y_i represents a sample that corresponds to a pixel in the input image. According to the number m of artificially set classes, K-means algorithm is used to cluster $\{y_1, y_2, \dots, y_n\}$ to obtain cluster centers denoted as C_1, C_2, \dots, C_m .

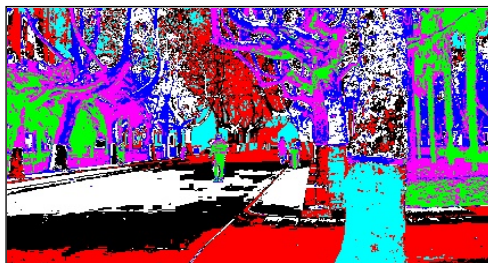
After all the samples are associated with one of the cluster centers C_1, C_2, \dots, C_m , the clustered classes A_1, A_2, \dots, A_m are derived with $A_i = \left\{ j \mid y_j \in C_i \right\}$. Because a pixel corresponds to a row element of the matrix H , this decides the category to which each pixel belongs.

B. Image Clustering Segmentation Results

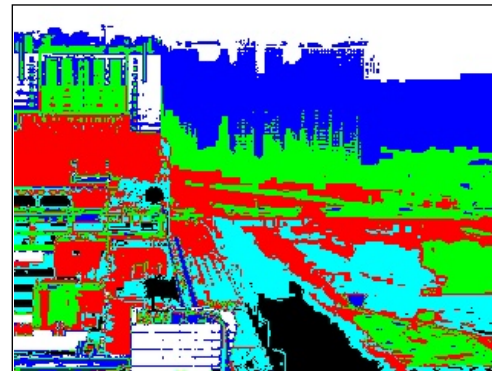
To reduce the complexity, the first frame of the video, namely $I_{w \times h}$, is down-sampled to $I_{w_- \times h_-}$, its width and height are $w_- = w/s_1$ and $h_- = h/s_2$, respectively. s_1 and s_2 are down-sampling rate, here $s_1 = s_2 = 2$. The Y components of $I_{w_- \times h_-}$ in YUV color space, denoted by $I_{w_- \times h_-}^Y$, are extracted and used as the input image for spectral clustering. The number of clusters are initialized to m , which can be adjusted according to the scene content. The spectral clustering for the image $I_{w_- \times h_-}^Y$ is performed as follows:



(a)



(b)



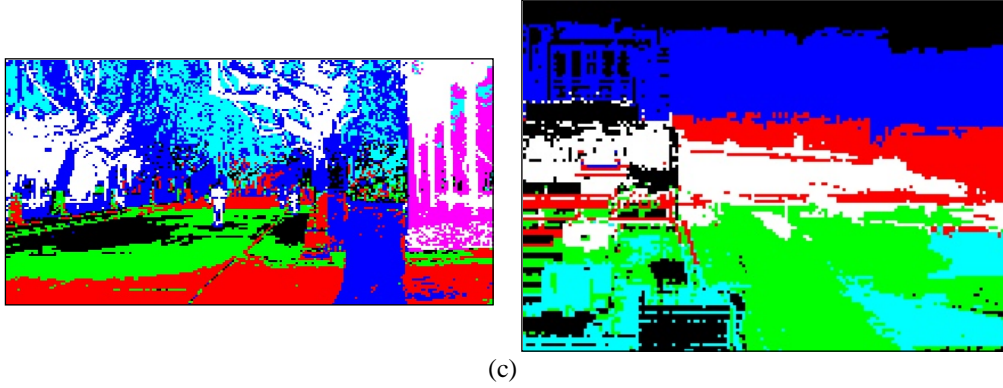


Fig. 4. The hazy images and the corresponding clustering segmentation maps, (a) Hazy images for "Shade Avenue"(left column) and "Riverside"(right column), (b) K-means clustering segmentation maps and (c) Spectral clustering segmentation maps.

First, The image $I_{w_- \times h_-}^y$ is mapped to a weighted undirected graph $G(V, W)$. Each vertex v_i ($i = 1, 2, \dots, n, n = w_- \times h_-$) corresponds to a pixel p_i of $I_{w_- \times h_-}^y$, and an edge element w_{ij} in adjacency matrix $W_{n \times n}$ represents the similarity between a pixel pair (p_i, p_j) . Diagonal matrix D and Laplacian matrix L are obtained from the adjacency matrix $W_{n \times n}$. After L is normalized to L_{nor} , the singular value decomposition (SVD) is done for L_{nor} and the eigenvalue matrix $H_{n \times k}$ is constructed by containing the first eigenvectors v_1, v_2, \dots, v_k of L_{nor} as columns. Each row element of H is qualified according to the cluster centers, and the class that each sample belongs to is obtained. Because each pixel corresponds to a row element, this determines the category of each pixel. Spectral clustering segmentation matrix $S_{w_- \times h_-}^m$ consists of the categories of all the pixels of $I_{w_- \times h_-}^y$. $S_{w_- \times h_-}^m$ is upsampled to $S_{w \times h}^m$, which indicates the classification results of the input hazy image.

In order to evaluate the effect of the spectral clustering algorithm, K-means clustering is also performed for the segmentation of the down-sampled image $I_{w_- \times h_-}^y$, the segmentation effect maps of the two clustering algorithms are figured as follows. **Fig. 4** shows the hazy image and the corresponding segmentation maps of "Shade Avenue" image and "Riverside" image in left column and right column, respectively. **Fig. 4(a)** shows the input hazy images. **Fig. 4(b)** shows the segmentation maps resulted from K-means clustering and **Fig. 4(c)** shows the segmentation maps resulted from the proposed spectral clustering. Here, different colors are used to highlight different classes. Given the scene content, the numbers of segmentation categories are seven and six for "Shade Avenue" and "Riverside" respectively. Experimental results show that the spectral clustering method has

a more satisfying segmentation performance in subjective visual effect. Take "Shade Avenue" image, for example, as shown in the left column of Fig. 4(b), even though the distances are different, some trunks and ground with similar pixel values are grouped into the same category by K-means clustering, which are marked with white color; some of them are divided into another category, which are marked with red color. Through considering of difference of both distance and brightness of pixels simultaneously, the spectral clustering improves the performance of image segmentation. As shown in the left column of Fig. 4(c), the same targets are basically divided into the same category. For example, the walking people is not only distinguished clearly from the surrounding environment, but also grouped with the green trees at the same distance, as marked with white color. Take "Riverside" image for example, K-means algorithm clusters the regions with different distances into a category, which are marked as red color in the right column in Fig. 4(b). In the right column in Fig. 4(c), sky, building, green vegetation, and small lake of the "Riverside" image are all fairly well separated from each other, and the regions with similar distances most likely to be classified into one same category.

The next step is to build the transmission estimation model by using these segmentation results and calculate the transmittance map of each frame

3.3 Atmospheric Light Estimation Based on Spectral Clustering Results

Atmospheric light A is often considered as the brightest color in an image, as a large amount of haze causes a bright color. However, objects, which are brighter than the atmospheric light, may lead to undesirable selection of the atmospheric light. By exploiting the basic fact that the variance of pixel values is generally low in hazy regions, e.g. in sky, Kim et al. propose a hierarchical search method based on a quad-tree subdivision to choose the atmospheric light that is as bright as possible. The method is consistent with our general understanding of atmospheric light. The computational complexity linearly increases with the number of recursive quad-tree subdivisions. In order to further reduce the computational complexity, the optimal class is selected from the spectral segmentation maps $S_{w \times h}^m$ and atmospheric light within it is estimated. The specific steps of the algorithm are given below. Adopting the idea from Kim et al., we first define the score of each class as the average pixel value subtracted by the standard deviation within the category. We then select the class with the highest score as the optimal class. Because the shape of each class may be irregular, the recursive quad-tree subdivision criterion is not applicable here. A certain number rectangular blocks with their area larger than a certain threshold are selected randomly from the optimal class and the block with the highest score is selected as the candidate block. Within the candidate block, we can divide it further into four smaller regions and repeat this process until the size of the selected region is smaller than a pre-specified threshold. Finally, we choose the color vector that minimizes the distance $\|(I_r(x), I_g(x), I_b(x)) - (255, 255, 255)\|$ as the atmospheric light. As compared with Kim et al.'s method, the proposed method can save 18% time for a 640×480 sized color image.

3.4 Transmission Estimation Based on Spectral Clustering Results

Because the proposed dehazing algorithm is based on a static background, only the first frame of the input hazy video is segmented into different classes using spectral clustering. Assuming that each class has the same transmission rate, class-based transmission estimation is done for the first frame and subsequent frames of the input hazy video by solving different transmission models. Without previous frame or the references in the previous frames, the spatial consistency transmittance model is built for all the categories in the first frame and newly appearing objects. Because the counterpart exists in the previous frame, the time-space coherent transmission model is established for the static background or moving objects that appear continuously. The transmission values are then obtained by optimizing the corresponding model. The specific transmission models are presented below.

A. Spatial Consistency Transmittance Model

When the dehazing is processed, Kim et al. consider to enhance the contrast of the image and to reduce the information loss, but they ignore the fact that the contours and shapes of a fog-free image are clearer than those of a hazy-image. The effect of different transmittance on edge strength is illustrated below.

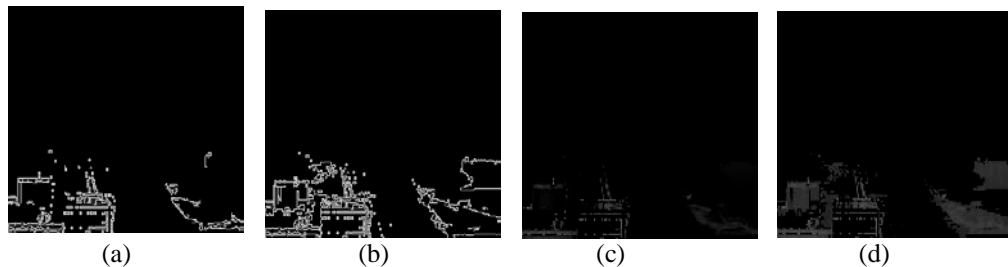


Fig. 5. Edge images and gray images within a specific category of "Riverside" image at different transmission rates: (a) Edge images at the transmission rate is initialized to 0.3, (b) Edge images when the transmission parameter is the optimum value, (c) gray images correspond to (a) and (d) gray images correspond to (b).

Fig. 5 shows the effect of different values of transmission t on the edge and gray images within a specific category. **Fig. 5(a)** and **Fig. 5(b)** show the edge images when the transmission t is initialized to 0.3 and is optimized, respectively. **Fig. 5(c)** and **Fig. 5(d)** are the corresponding gray image of **Fig. 5(a)** and **Fig. 5(b)**, respectively. These images show that there are great differences between the edge images restored by different transmission values for the hazy image within a particular category. Compared to the edge and gray images which derived by the original transmission value, the edge image restored with the optimized value is richer in detail and the corresponding gray image is clearer. Hence, we conclude that the richness of the edge can reflect the effect of dehazing. That is, the more detail there is in an edge image, the more accurate the transmission value is. Therefore, we modify Kim et al.'s optimal cost model to form a new spatial consistency

transmittance model. We estimate the optimal transmission t so that the dehazed category has the maximum contrast, minimum information loss, and strongest edge. In the implementation, it was found that the effect of fog on chrominance components U and V is much smaller than on the Y component, and the dehazing effect using the Y component is basically the same as by using RGB components. For simplicity, we define the cost function based on a haze model only for the Y component, which is expressed as:

$$J_k^m(p) = \frac{1}{t} (I_k^m(p) - A) + A \quad (24)$$

where $J_k^m(p)$ and $I_k^m(p)$ denote the original and the observed Y components within category m at pixel position p of k th frame, respectively. A is the global atmospheric light.

● **Edge cost function:**

Both sides of Eq. (24) are filtered through canny edge detector to pick up the image details. The edges are summed together to form the edge cost function E_{edge} , which is given by

$$E_{edge} = -\text{mean}\left(\text{edge}\left(J_k^m(p)\right)\right) = -\text{mean}\left(\text{edge}\left(\frac{1}{t}(I_k^m(p) - A) + A\right)\right) \quad (25)$$

where $\text{mean}(\cdot)$ is mean function, $\text{edge}(\cdot)$ is edge filtering function. Canny edge detector is used here for filtering the class-based pixel set $I_k^m(p)$.

● **Spatial consistency cost function:**

The new cost function E^m consists of three parts: the contrast cost function $E_{contrast}$, the information loss cost function E_{loss} and the edge cost function E_{edge} .

$$E^m = E_{contrast} + \lambda_L E_{loss} + \lambda_E E_{edge} \quad (26)$$

The first two terms in Eq.(26) are the functions proposed by Kim et al., but only the Y components of the pixels within the class m are involved in minimizing the overall cost function. λ_L and λ_E are the weighting parameters that control the relative importance of the contrast cost, the information loss cost and edge cost.

● **Optimal transmission estimation:**

The transmission map obtained by the proposed model is richer in detail than that of Kim et al.'s model and provides a higher image quality.

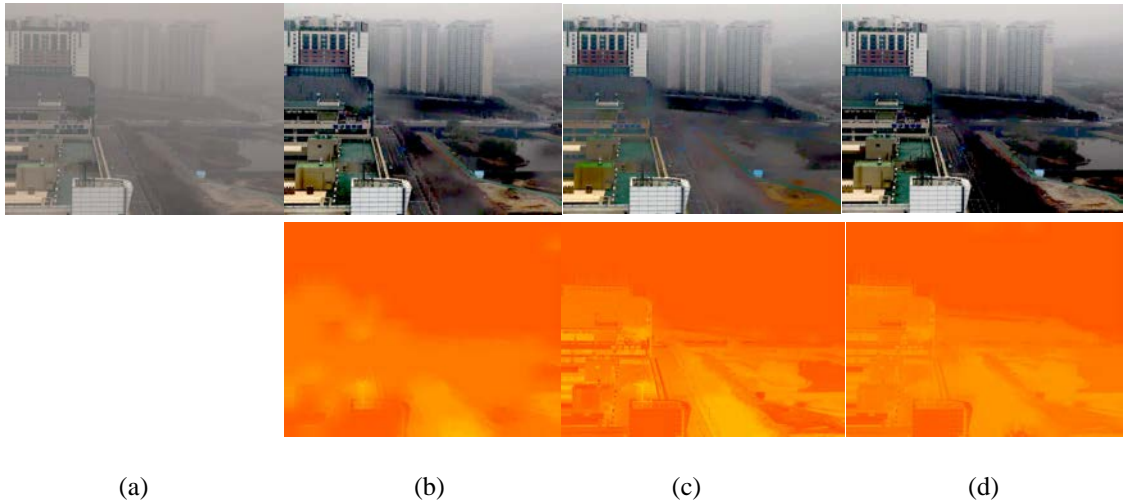


Fig. 6. Comparative results of dehazed images and the transmission maps on a "Riverside" image, (a) the hazy images, The dehazed images and the corresponding transmission maps obtained by employing (b) Kim et al.'s algorithm, (c) K-Means clustering segmentation and (d) our algorithm.

Top: the original input image, the dehazed image by Kim et al.'s method, K-Means clustering segmentation and by our model, respectively. Bottom: the transmission maps of Kim et al.'s method, K-Means clustering segmentation and our method, respectively.

After the above dehazing total cost function is established, for each class k , the optimal transmission t_k^* is found by minimizing the overall cost function. We evaluate the performance of the proposed spatial consistency transmission model comparing to Kim et al.'s model on hazy "Riverside" image in **Fig. 6(a)**. **Fig. 6(b)** shows the restored dehazed image and the estimated transmission map by using Kim et al.'s cost model, where yellow and red pixels represent near and far scene points, respectively. **Fig. 6(c)** shows the restored dehazed image and the estimated transmission map by using K-means clustering segmentation. **Fig. 6(d)** shows the restored dehazed image and the estimated transmission map by the proposed spatial consistency transmission model. The transmission map obtained by the proposed model is not only richer in detail than that of Kim et al.'s model, but also has a stronger sense of distance than K-means clustering. Compared to Kim et al.'s model and K-means clustering, the proposed model has a higher dehazed image quality.

B. Time-Space Coherence Transmission Model

In the dehazing processing, not only the spatial consistency of the video frame should be considered, but the temporal continuity between video frames should also not be ignored. If the transmission similarity between the successive frames could not be guaranteed, flicker will appear in the video and thus affect the visual results. However, this similarity assumption does not hold when a scene change occurs or a new object appears. In view of such situations, we take the following approach. When a scene change is detected by a simple scene change detection algorithm, e.g., color histogram difference, the atmospheric

light and spectral-clustering separation are estimated again. For an unchanged scene over a period of time, moving objects are first detected by threshold processing of frame differences. For a newly appearing object, only the spatial consistency model is applied to estimate the transmission value because there is no reference in the previous frame. However, for the moving objects that appear continuously, the corresponding position of each moving object in the previous frame is obtained by motion estimation. For the same scene point or points at the trajectory of a moving object, a Time-Space coherent transmission model is built along the corresponding positions in previous and current frames. After the optimal transmission is obtained by solving the cost function, the video image is reconstructed. For the same scene point or points in the trajectory of a moving object, a linear relationship exists between the transmission $t_k(p)$ in the k th frame and the transmission $t_{k-1}(p)$ in the $(k-1)$ th frame, so we build the following temporal coherence model:

$$E_{temporal} = (t_k(p) - \tau_k t_{k-1}(p))^2 \quad (27)$$

where p is the corresponding position, and τ_k is the temporal coherence factor, which is denoted as Eq.(17).

To dehaze a moving object and static scene point, we reformulate the overall cost in Eq. (26) by adding the temporal coherence cost of Eq. (27) as follows:

$$E^m = E_{contrast} + \lambda_L E_{loss} + \lambda_E E_{edge} + \lambda_T E_{temporal} \quad (28)$$

where λ_T is a weighting parameter which strikes a compromise between flickering artifacts and the qualities of individual frames.

4. Experimental Classification Results and Analysis

Although the spectral clustering algorithm has a high computational complexity, if the scene does not change, we only perform spectral clustering on the first frame of a specific scene to eliminate the complexity brought about by clustering. The video dehazing process is described below. First, for a specific scene, the first frame of the video is divided into different classes according the spectral clustering algorithm. Second, an iterative method is used to minimize the designed cost function for each class of pixel sets. Further, the corresponding transmission value is obtained for each class. Finally, the transmittance map is optimized using fast-guided filtering, and each pixel in the hazy image is restored accordingly. In order to evaluate the performance of the proposed video dehazing algorithm, this paper uses OpenCV 2.3.1 image processing functions in a Visual Studio 2010 environment to implement the proposed method. The proposed algorithm is verified using both subjective visual and objective image quality assessment. The experimental results are given below.

A. Subjective Visual Effects Evaluation

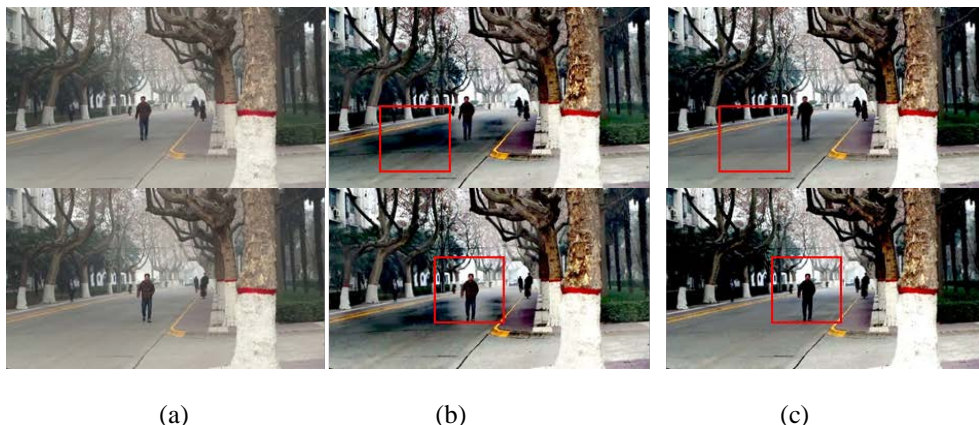


Fig. 7. Comparative results of the proposed algorithm and Kim et al.'s algorithm on the "Shade Avenue" images: (a) the hazy images. The dehazed images obtained by (b) Kim et al.'s algorithm and (c) our proposed algorithm.

We first compare the performance of the proposed algorithm with Kim et al.'s algorithms on the "Shade Avenue" images in **Fig. 7**. **Fig. 7(a)** shows the hazy images, **Fig. 7(b)** shows the dehazed images restored by Kim et al.'s algorithm and **Fig. 7(c)** shows the dehazed images restored by our method. From top to bottom, the two rows are the 50th and 150th frames of the video, respectively. The results show that Kim et al.'s algorithm has two drawbacks: First, it destroys the intra-frame spatial consistency because the transmission estimation is based on fixed-size block segmentation. Take the red rectangle box of the first row as an example; it has a greater color difference than in the results of our method. Second, because of the excessive reliance on the information of the same position in the previous frame, a white mist appears around the person where he walks along, which is marked with the red rectangle in the second row. We see in **Fig. 7** that the visual results of our algorithm are relatively more natural, taking the road and moving target as examples. The subjective results show that the proposed algorithm performs better in terms of spatial consistency and time continuity than Kim et al.'s method.

We also compare the performance of the proposed algorithm with Kim et al.'s algorithm on the "Riverside" images in **Fig. 8**. From top to bottom, the two rows are the 20th and 40th frames of the video, respectively. **Fig. 8(b)** shows that Kim et al.'s algorithm produces halo artifacts on the green belt below the tall building, which is marked by a red rectangle box on the top row of **Fig. 8(b)**. The central portion of the river has a significant blocking artifact, which is marked by a yellow rectangle box on second row of **Fig. 8(b)**. The fog on part of road does not dissipate, which is marked with a green rectangle box on the second row of **Fig. 8(b)**. **Fig. 8(c)** shows that the subjective visual effects of the restored objects in a frame (such as river water, green belt, road and others) by our method are much better than those of Kim et al.'s algorithm.

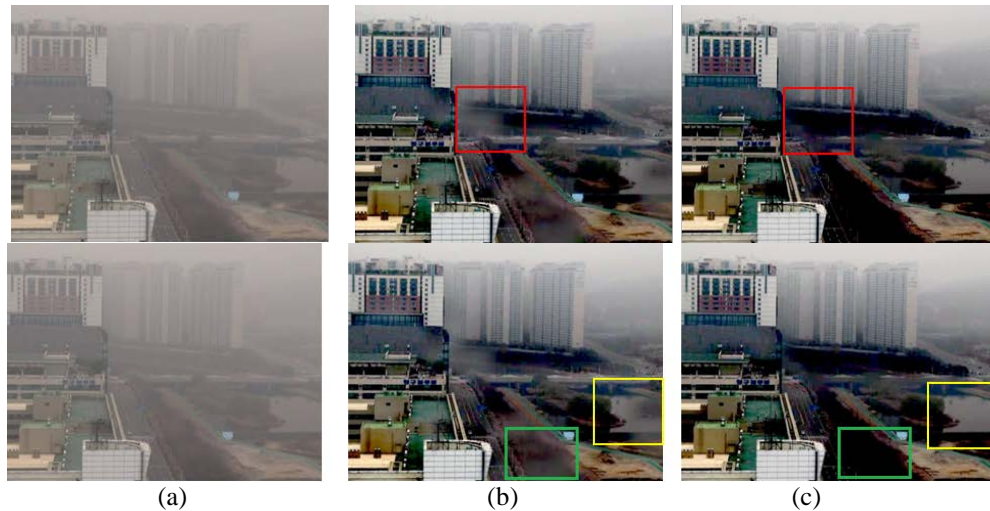
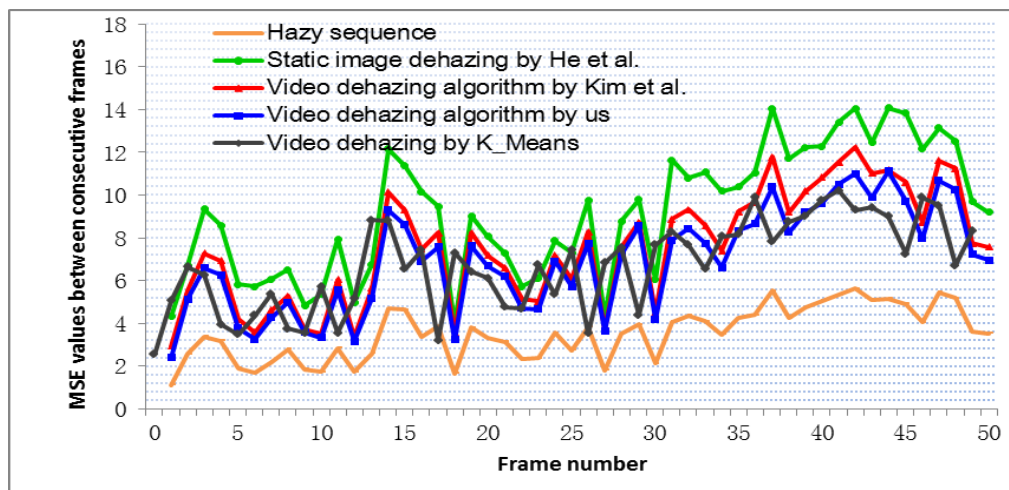


Fig. 8. Comparative results of the proposed algorithm and Kim et al.'s algorithm on the "Riverside" images: (a) the hazy images. The dehazed images obtained by (b) Kim et al.'s algorithm and (c) our proposed algorithm.

B. Objective Quality Evaluation

The intra-frame spatial consistency can be observed by its subjective effect. Discontinuity between frames is reflected by the occurrences of flicker, which can be objectively measured by the MSE (Mean Square Error) between adjacent frames. The computational efficiency is also needs to be evaluated because the complexity of the algorithm is important for application and execution. Below, we give the objective evaluation results for MSE and real-time performance.



(a)

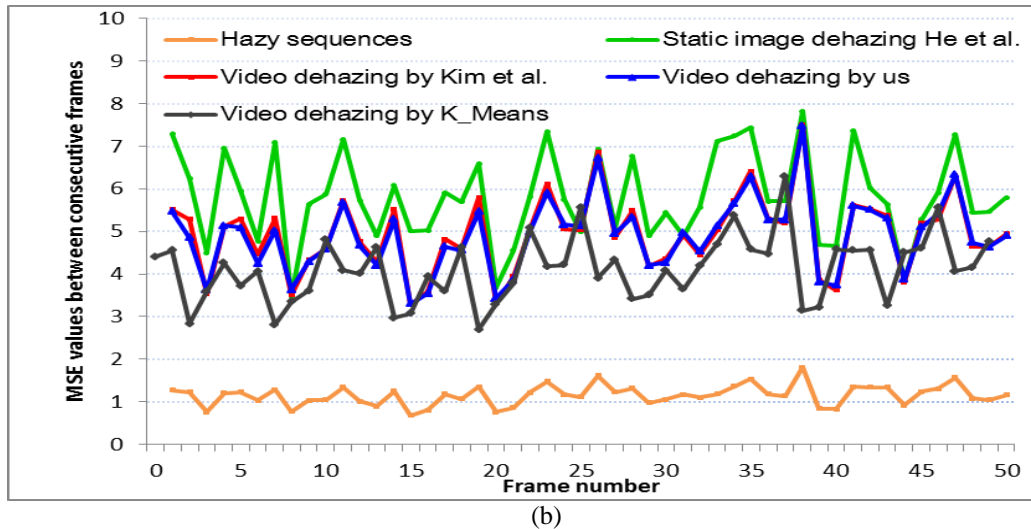


Fig. 9. Comparison of the MSE values between consecutive frames. (a) "Shade avenue" sequence and (b) "Riverside" sequence.

We quantitatively show how the proposed video dehazing algorithm suppresses flickering artifacts by employing the temporal coherence cost for moving objects and same scene points in the optimization function. **Fig. 9** plots the MSE between consecutive frames for the "Shade Avenue" and "Riverside" sequences. The orange line represents the MSE values between adjacent frames of the input hazy sequences. The red, green, gray and blue lines represent the MSE values between consecutive frames of the dehazed sequences by the algorithm of Kim et al., He et al., K-means clustering and our method, respectively. These results show that the MSE values of the input hazy video are low and increase significantly after dehazing. This indicates that dehazing objectively enhances image contrast. As shown by the green lines, when the static image dehazing algorithm is independently applied to each frame of the video, the MSE experiences relatively large fluctuations compared with the input hazy sequences. Although the image contrast is significantly enhanced by He et al.'s algorithm, the curve is not smoother than that of the proposed video dehazing method. This also proves that the single-frame image dehazing algorithm cannot be directly applied to video dehazing. For "Shade Avenue" sequence, the average MSE obtained by the proposed algorithm, K-means clustering and Kim et al.'s method is 6.95, 6.73 and 7.56, respectively. For "Riverside" sequence, the average MSE obtained by the proposed algorithm, K-means clustering and Kim et al.'s method is 4.90, 4.10 and 4.94, respectively. The MSE curve obtained by the proposed algorithm is further smoother than Kim et al.'s method, which shows that our algorithm can obtain better continuity between the video frames and reduces the flickering artifacts more efficiently. Compared to K-means clustering segmentation, the spectral clustering segmentation achieves smoother and higher MSE values.

C. Complexity Evaluation

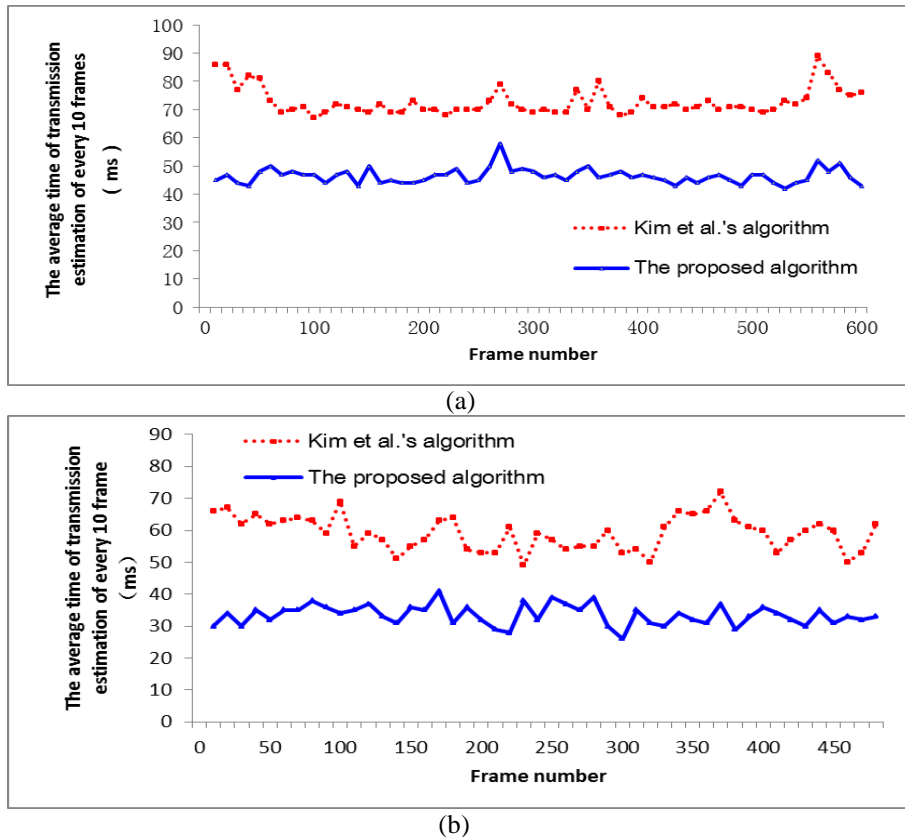


Fig. 10. Comparison of the transmission estimation time of the proposed method with Kim et al.'s work. (a) "Shade avenue" sequence and (b) "Riverside" sequence.

The spectral clustering algorithm has high complexity, for example, it takes about 12 seconds for a 640×480 sized image. But fortunately, spectral clustering is done only for the first frame and the segmentation results can be used for the subsequent atmospheric light estimation and the transmission estimation. In the video dehazing process, the transmission estimation is slow, so we compared the real time implementation results of our method with those of Kim et al.'s algorithm in Fig. 10, in which the transmission estimation time for every 10 frames is averaged. For a video file with a resolution of 640×480 , our approach can process about 26 frames per second for transmission estimation. For the "Shade Avenue" sequence, the processing time on a PC with a 3.10 GHz Intel Core i5-2400 CPU is about 46.48 ms while Kim et al.'s method needs about 72.7 ms, showing that our method saves 35.85% time for transmission estimation. For the "Riverside" sequence, the processing time of the proposed method is about 33.48 ms while Kim et al.'s method takes about 59.15 ms, which saves 43.40% of the time needed for transmission estimation. This figure shows that the time complexity of transmission estimation in the

proposed algorithm reduced by 0.39 times compared to Kim et al.'s method. Therefore, the proposed video dehazing algorithm based on spectral clustering in this paper not only solves the problem of spatial consistency and inter-frame time continuity, but also reduces the time complexity and improves computational efficiency.

5. Conclusion

In this work, we proposed a video dehazing algorithm based on spectral clustering. The proposed algorithm first segments the first frame of the video using a well designed spectral clustering algorithm that simultaneously considers the brightness and position distances of pixels in the image. The proposed algorithm then selects the atmospheric light in the category with brightest and smallest gray value variance. Next, the proposed algorithm adds an edge cost function to Kim et al.'s optimized spatial consistency dehazing model to enhance weak restored image edges in addition to increasing the contrast and reducing information loss. When a new object appears, the block similarity between the two consecutive frames at the same position is small. Hence, we cancel Kim et al.'s temporal coherence cost and just use spatial consistency in this case. Because there is a reference, the simplified temporal coherence cost is used for the same scene points or moving objects detected by thresholding the frame difference. This algorithm is verified with respect to subjective visual effects, objective image quality assessment, and time complexity. The results show that it obtains a better enhancement effect and less processing time than Kim et al.'s algorithm.

References

- [1] ZhangNarasimhan S G, Nayar S K, "Vision and the Atmosphere," *International Journal of Computer Vision*, vol.48, no.3, pp. 233–254, 2002. [Article \(CrossRef Link\)](#).
- [2] Hautiere N, Tarel J P, Aubert D, "Towards Fog-Free In-Vehicle Vision Systems through Contrast Restoration," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June, 2007. [Article \(CrossRef Link\)](#).
- [3] S. NarasimhanS, . Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713-724, June, 2003. [Article \(CrossRef Link\)](#).
- [4] S. Shwartz, E. Namer, Y. Schechner, "Blind haze separation," in *Proc. of IEEE CVPR'06*, June, 2006. [Article \(CrossRef Link\)](#).
- [5] Y. Schechner, S. Narasimhan, S. Nayar, "Instant dehazing of images using polarization," in *Proc. of IEEE CVPR 2001*, pp. 325–332, Dec. 2001. [Article \(CrossRef Link\)](#).
- [6] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, "Deep photo: model-based photograph enhancement and viewing," *ACM Trans. Graph*, vol. 27, no. 5, pp.1-10, Dec. 2008. [Article \(CrossRef Link\)](#).
- [7] Di Wu, Qionghai Dai, "Data-driven visibility enhancement using multi-camera system," in *Proc. of SPIE Enhanced and Synthetic Vision*, May, 2010. [Article \(CrossRef Link\)](#).

- [8] Schaul L, Fredembach C, Susstrunk S. "Color Image Dehazing using the Near-Infrared," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 1629-1632, Nov. 2009. [Article \(CrossRef Link\)](#).
- [9] Ashish V. Vanmali, Samrudha G. Kelkar, Vikram M. Gadre. "A novel approach for image dehazing combining visible-NIR images," in *Proc. of 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Dec. 2015. [Article \(CrossRef Link\)](#).
- [10] C. Feng, S. Zhuo, X. Zhang, L. Shen, and S. Susstrunk, "Near-infrared guided color image dehazing," in *Proc. of Image Processing, ICIP 2013. 20th IEEE International Conference on*, pp. 2363–2367, Sep. 2013. [Article \(CrossRef Link\)](#).
- [11] X. Zhang, T. Sim, and X. Miao, "Enhancing photographs with near infrared images," in *Proc. of Computer Vision and Pattern Recognition, CVPR 2008. IEEE Conference on*, pp. 1–8, June, 2008. [Article \(CrossRef Link\)](#).
- [12] S. Zhuo, X. Zhang, X. Miao, and T. Sim, "Enhancing low light images using near infrared flash images," in *Proc. of Image Processing, 2010. ICIP 17th IEEE International Conference on*, pp. 2537–2540, Sept. 2010. [Article \(CrossRef Link\)](#).
- [13] Fattal R. "Single image dehazing," *ACM Transactions on Graphics*, 2008, vol. 27, no. 3, pp. 1-9, Aug. 2008. [Article \(CrossRef Link\)](#).
- [14] Tan R T. "Visibility in bad weather from a single image," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, pp. 1-8, June.2008. [Article \(CrossRef Link\)](#).
- [15] Kratz L, Nishino K. "Factorizing Scene Albedo and Depth from a Single Foggy Image," in *Proc. of IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 1701-1708, 2009. [Article \(CrossRef Link\)](#).
- [16] Tarel J P, Hautière N. "Fast visibility restoration from a single color or gray level image," in *Proc. of IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 2201-2208, 2009. [Article \(CrossRef Link\)](#).
- [17] He K, Sun J, Tang X. "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011. [Article \(CrossRef Link\)](#).
- [18] Xiang X, Cheng Y, Tang J. "A Novel Video Dehazing Method Based on Temporal Visual Coherence," in *Proc. of 7th International Conference on Internet Multimedia Computing and Service*, Aug. 2015. [Article \(CrossRef Link\)](#).
- [19] Wang J B, He N, Zhang L L, et al. "Single image dehazing with a physical model and dark channel prior," *Neuro computing*, vol. 149, Part B, pp. 718-728, Feb. 2015. [Article \(CrossRef Link\)](#).
- [20] Yeh C H, Kang L W, Lin C Y, et al. "Efficient image/video dehazing through haze density analysis based on pixel-based dark channel prior," in *Proc. of International Conference on Information Security and Intelligence Control (ISIC)*, pp. 238-241, 2012. [Article \(CrossRef Link\)](#).
- [21] Kumari A, Sahdev S, Sahoo S K. "Improved Single Image and Video Dehazing Using Morphological Operation," in *Proc. of International Conference on VLSI Systems, Architecture, Technology and Applications (VLSI-SATA)*, pp. 1-5, Jan. 2015. [Article \(CrossRef Link\)](#).
- [22] Shin D K, Yong M K, Park K T, et al. "Video dehazing without flicker artifacts using adaptive temporal average," in *Proc. of The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, pp. 1-2, June. 2014. [Article \(CrossRef Link\)](#).

- [23] Zhang J, Li L, Zhang Y, et al. "Video dehazing with spatial and temporal coherence," *Visual Computer*, vol. 27, no. 6-8, pp. 749-757, June, 2011. [Article \(CrossRef Link\)](#).
- [24] Kim J H, Jang W D, Sim J Y, et al. "Optimized contrast enhancement for real-time image and video dehazing," *Journal of Visual Communication & Image Representation*, vol. 24, no. 3, pp. 410-425, April, 2013. [Article \(CrossRef Link\)](#).
- [25] Goldstein E B. "Sensation and Perception," in *Proc. of 31st International Congress of Psychology*, pp.24–29, July 2016. [Article \(CrossRef Link\)](#).
- [26] Preetham A J, Shirley P, Smits B. "A practical analytic model for daylight," in *Proc. of the 26th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 1999:91-100*. [Article \(CrossRef Link\)](#).
- [27] A. Levin, D. Lischinski, and Y. Weiss. "A closed form solution to natural image matting," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June, 2006. [Article \(CrossRef Link\)](#).
- [28] Lv X, Chen W, Shen I F. "Real-Time Dehazing for Image and Video," in *Proc. of IEEE Computer Society Conference on Computer Graphics & Applications*, pp. 62-69, Sept. 2010. [Article \(CrossRef Link\)](#).
- [29] Gujral A, Gupta S, Bhushan B. "A Novel Defogging Technique for Dehazing Images," *International Journal of Hybrid Information Technology*, vol. 7, no. 4, pp. 235–248, July. 2014. [Article \(CrossRef Link\)](#).
- [30] J. Oakley, H. Bu, "Correction of simple contrast loss in color images," *IEEE Trans. Image Process*, vol. 16, no. 2, pp. 511–522, Feb. 2007. [Article \(CrossRef Link\)](#).
- [31] K. He, J. Sun, X. Tang, "Guided image filtering," in *Proc. of ECCV, 2010*, Part 1, LNCS 6311, pp. 1–14, 2010. [Article \(CrossRef Link\)](#).
- [32] Ulrike von Luxburg. "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, issue. 4, pp 395–416, Dec. 2007. [Article \(CrossRef Link\)](#).
- [33] Ming Qin, Yao Lu, Huijun Di, and Wei Huang. "A Background Basis Selection-Based Foreground Detection Method," *IEEE Transactions on Multimedia*, vol. 18, no. 7, pp. 1283–1296, July, 2016. [Article \(CrossRef Link\)](#).
- [34] Shi J, Malik J. "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 22, issue. 8, pp. 888-905, Aug.2000. [Article \(CrossRef Link\)](#).



Fan Zhao received a Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, Xi'an, China, in 2009. She worked as a postdoctoral fellow in the Department of Computer Science and Engineering, Xi'an Jiaotong University from July 2010 to April 2012. She is now an associate professor in the Department of Information Science at Xi'an University of Technology, Xi'an, China. Her research interests include image processing, object tracking, and pattern recognition. She is CCF professional member and IEEE CS member.



Zao Yao received the BE degree from Xi'an University of Technology, Xi'an, China, in 2013. She is a postgraduate at the Department of Information Science at Xi'an University of Technology, Xi'an, China. She is mainly engaged in the research of image/video dehazing.



Xiaofang Song received the BE degree from Xi'an University of Technology, Xi'an, China, in 2014. She is a postgraduate at the Department of Information Science at Xi'an University of Technology, Xi'an, China. She is mainly engaged in the research of video dehazing and object tracking.



Yi Yao received the BE degree from Xi'an University of Technology, Xi'an, China, in 2014. She is a postgraduate at the Department of Information Science at Xi'an University of Technology, Xi'an, China. She is mainly engaged in the research of video dehazing and object tracking.