

# Image Visibility Clarification and Enhancement

Gourab Ghosh Roy, Mohammad Rami Koujan, Raabid Hussain, Alexander Belyaev, and Yvan Petillot

School of Engineering & Physical Sciences

Heriot-Watt University, Edinburgh, UK

{ gg23 , mk62, rh37, A.Belyaev, Y.R.Petillot } @hw.ac.uk

**Abstract**— Haze can cause poor visibility and loss of contrast in images and videos. In this article, we study the dehazing problem which can improve visibility and thus help in many computer vision applications. An extensive comparison of state of the art single image dehazing methods is done. One simple contrast enhancement method is used for dehazing. Structure-texture decomposition has been used in conjunction with this enhancement method to improve its performance in presence of synthetic noise. Methods which use a haze formation model and attempt at solving an ill-posed problem using computer vision priors are also investigated. The two priors studied are dark channel prior and the non-local prior. Both qualitative and quantitative comparisons for atmospheric and underwater images on all three methods provide a conclusive idea of which dehazing method performs better. All this knowledge has been extended to video dehazing. A video dehazing method which uses the spatial and temporal information in a video is studied in depth. An improved version of video dehazing is proposed in this article, which uses the spatial-temporal information fusion framework but does not suffer from some of its limitations. The new video dehazing method is shown to produce better results on test videos.

## I. INTRODUCTION

In many computer vision applications, reduced visibility due to different environmental conditions is a major problem. Applications like object detection, surveillance, autonomous vehicles, depth-mapping, resource exploitation, feature extraction, scene classification and analysis all work best when the different objects in the scene have distinct outlines. The performance of these systems is adversely affected by even slight degradation in the input image.

Unwanted media in the environment such as haze, smoke, fog and rain are common causes of such degradation in natural outdoor images and videos [1]. These turbid media in the atmosphere absorb and scatter the incoming light from its original course of propagation. In long distance photography, this process has a substantial effect. For example, the contrast is reduced and the surface colors become faint. These degraded images lack visual vividness and account for poor visibility of image content.

Haze is independent of scene radiance and accounts for scattering of light through two phenomena. It attenuates the signal and introduces an additive component termed as airlight or ambient light [2]. Airlight can be defined as the color of the scene locale at infinity. The amount of degradation follows a directly proportional relation to the

distance of the scene object from the camera. Hazy images can be thought of as a combination of a haze free image and airlight.

Haze removal techniques are highly desired in both photography and computer vision applications as they can correct the color shift caused by airlight making the images more appealing and also increase the contrast between different objects, thereby improving the robustness and performance of those computer vision systems. Also, since the amount of haze is dependent on the distance between the object and camera, a depth map of the scene can sometimes be generated as a by-product of the algorithms. Haze removal techniques can be classified into image enhancement and image restoration processes [3]. The former tries to enhance the contrast of the image without taking into account the physics involved in hazing. Histogram equalization and contrast stretching are popular choices in this category. Whereas, the latter tries to integrate the image degradation model [4] to enhance the image.

Early approaches to image dehazing involved the use multiple images of the scene to extract information about the haze. One popular approach was to take images of the same scene at different degrees of polarization by rotating the polarization filter attached to the camera [5], [6], [7]. Another popular approach was to take images of the scene at different weather conditions [8], [9], [10]. The main idea behind this approach was to exploit differences between images distorted by different participating media. The main drawback of these techniques is that they can not handle dynamic scenes as they require capturing multiple images of exactly the same scene but at different conditions which is not possible when fast-moving objects are present in the scene. Thus a need for single image dehazing has gained importance.

In this article, we explore three promising state of the art single image dehazing techniques: Screened Poisson Equation based contrast enhancement [11], Dark-channel prior based haze removal [12] and Non-local image dehazing [13], followed by a video dehazing method based on spatial-temporal information fusion [14]. Based on these techniques, we develop an efficient framework for underwater video dehazing.

The rest of the article is organized as follows. Firstly, previous work is discussed in the background section for both single image and video dehazing. Then we provide an extensive description of the state of the art dehazing methods relevant to this work. This is followed by a description of our proposed algorithm for video dehazing. Experimental results are presented next followed by concluding remarks.

## II. BACKGROUND

Over the years, numerous single image dehazing algorithms have been proposed. Since, for single image dehazing, the system becomes under-constrained, these techniques rely on prior assumptions relating to the physical models of the system. Most of them assume that the pixels in a local patch have same amount of haze. Although the results are not precise but they provide a reasonable estimate and are computationally efficient. However, all such methods fail when their assumptions become invalid.

Tan [15] observed that a dehazed image should have a higher contrast as compared to its hazy version, and thus haze can be removed by maximizing the local contrast of the dehazed image. His results are visually captivating but not physically valid. Zhang et al. [16] used an iterative scheme assuming that small variations in chromaticity are produced due to scene albedo while large variations are associated with transmission. Fattal [17] makes use of the fact that there is no correlation between medium transmission and surface shading to estimate the albedo of the scene. This approach is physically valid but its performance deteriorates with the increase in density of the haze. Kim et al. [18] enhances the contrast of the hazy image by estimating the airlight based on quad-tree subdivision. Tarel et al. [19] proposed a median filter based approach to restore the visibility in a dehazed image.

One of the most popular techniques proposed so far is by He et al. [12] which takes advantage of the fact that in most local patches of a haze-free image, there is always some pixels that have a very low intensity in at least one color channel. The assumptions proposed by He et al. have been shown to provide significant results, however sometimes the method underestimates the transmission and fails when the objects in the scene have similar appearance to haze. Since their publication, a number of articles have been published with modifications to this dark channel prior algorithm. Ullah et al. [20] proposed to extract the dark channel from the saturation and intensity channels rather than the usual RGB color channels. Chu et al. [21] perform a mean shift segmentation as a pre-processing step. Yu et al. [22] and Park et al. [23] use Weighted Least Square (WLS) based edge-preserving smoothing [24]. Yeh et al. [25] incorporated the notion of bright channel prior into the dark channel prior to achieve improved performance.

Berman et al. [13] used the concept of non-local prior for dehazing. The method is based on the assumption that any haze free image can be approximated with only a few distinct colors and hence pixels form tight clusters in RGB color space. The pixels in a cluster are often non-local and spread around the entire image plane. In hazy images, these clusters become haze lines, which are later used to estimate the transmission and dehaze the image. This linear algorithm uses a better prior and is more robust, however it fails when the airlight is significantly brighter than the scene radiance.

For video dehazing, a simple approach is to apply any image dehazing algorithm on each frame independently. Although this approach produces best performance if one considers the output of all the frames separately, however few problems appear when the consecutive frames are taken into account. The frames lack coherence among each other and the colors are often not consistent. Various factors like moving objects, wind or water flow, camera angles etc. are common sources of this problem. Also, the processing time is not suitable for real applications.

Several frameworks have been proposed in [26], [27], [28] that take temporal coherence into account to deal with the above mentioned problems. The notion behind spatial-temporal information fusion is that since the transmission maps of consecutive frames are similar, the dehazing algorithm does not need to be applied on each frame. Instead a simple linear translation-variant filtering process may be used as an alternate. In order to consider the effects of atmospheric phenomena, the airlight of the current frame is regularized by the airlight of the previous frame. Camera motion can be dealt by using point correspondence between consecutive frames.

The above algorithms can also be extended to underwater imaging. In underwater imaging, suspended particles such as sand, plankton and minerals are the main sources of hazy degradation. The main difference between outdoor and underwater imaging is that the light is absorbed differently in each color channel. Related approaches are presented in [29], [30], [31] respectively. Serikawa et al. [32] used joint trilateral filter [33] for underwater dehazing. Zhao et al. [34] considers inherent optical properties of water to generate a dehazing model. An extension of the dark channel prior haze removal method for underwater video dehazing is described in [14].

## III. STATE OF THE ART

In this section we present in detail the state of the art single image and video dehazing methods and concepts used in our work. One category of image dehazing methods presented here is that which does not use any particular haze model. An image processing technique to improve the said method is also described. Then we present dehazing

methods which use a haze model and describe how the ill-posed dehazing problem defined on the model can be solved utilizing different computer vision priors. Finally a video dehazing algorithm which just does not extend the image dehazing method to video in a frame by frame static way but uses the spatial-temporal information in a video is described.

#### A. Screened Poisson Equation for Image Contrast Enhancement

The Screened Poisson Equation based method proposed in [11] is a gradient domain method which does not use any haze model. It is a form of image sharpening where the method works like a high pass filter. The objective is to preserve the image gradients and at the same time remove the nonuniform illumination effects in the image. This is done by minimizing, with respect to the function  $u$ , the functional  $J(u)$  consisting of two such corresponding terms.  $J(u)$  is given by:

$$J(u) = \int_{\Omega} ||\nabla u - \nabla f||^2 dx + \lambda \int_{\Omega} (u - \bar{u})^2 dx \quad (1)$$

where  $\lambda$  is the trade-off parameter between the two terms.

The nonlinear Euler-Lagrange equation for this minimization problem can be linearized by fixing the mean value of  $u$  to be same as the mean value of  $f$ . Then the solution can be obtained from the screened Poisson equation:

$$\lambda(u - \bar{f}) - \Delta u + \Delta f = 0, \text{ over } \Omega \quad (2)$$

with homogeneous Neumann boundary condition:

$$\frac{\partial u}{\partial n} = 0, \text{ over } \partial\Omega$$

where  $n$  denotes the normal vector to the boundary.

It is shown in [11] that the solution is not affected by the mean value term  $\lambda\bar{f}$ . Using the two dimensional discrete Fourier transform, the solution then is obtained in the frequency domain. The Neumann boundary condition is imposed by extending the input image symmetrically across its sides, so that the new  $J \times L$  image which is four times in size becomes symmetric and periodic. The frequency domain solution is obtained by the following:

$$\hat{u}_{mn} = \frac{((\frac{2\pi m}{J})^2 + (\frac{2\pi n}{L})^2)}{(\lambda + ((\frac{2\pi m}{J})^2 + (\frac{2\pi n}{L})^2))} \hat{f}_{mn} \quad (3)$$

The discrete inverse Fourier transform gives the spatial domain solution  $u$ .

The screened Poisson equation is applied to each color channel of the image separately. Simplest color balance [35] is applied to the input image, as also to the solution obtained from the screened Poisson enhancement. With this color balance, a small percentage of pixels are saturated at 255 and 0.

#### B. Structure-Texture Decomposition

The cartoon+texture decomposition method in [36] extracts the cartoon and texture parts from an image. The cartoon or the structure part contains contrasted shapes along with strong edges while the texture part is composed of oscillating patterns present in the image. So the cartoon can contain all frequencies including high frequencies, whereas texture contains frequencies in the middle and high range. This method which was originally published in [37] is a fast approximation of the variational problem proposed in [38]. It uses a nonlinear lowpass-highpass filter pair.

For each pixel  $x$  in image  $f$ , the local total variation (LTV) is defined as:

$$LTV_{\sigma}(x)(f) = G_{\sigma} * |\nabla f|(x) \quad (4)$$

where  $G_{\sigma}$  is a Gaussian kernel with standard deviation  $\sigma$ .

The relative reduction rate of the local total variation is defined as:

$$\lambda(x) = \frac{LTV_{\sigma}(x)(f) - LTV_{\sigma}(x)(L_{\sigma} * f)}{LTV_{\sigma}(x)(f)} \quad (5)$$

with  $L_{\sigma}$  being a low pass filter. For RGB images, the LTV is computed as an average of the LTV values in each of the channels and then  $\lambda(x)$  is obtained.

Textural points are observed to have a strong decay of LTV value on convolution with a low pass filter, with  $\lambda(x)$  close to 1. Based on this, the cartoon image is constructed as a weighted average of  $f$  and  $L_{\sigma} * f$  and the texture is just the difference of the cartoon part from the original image. This decomposition is not unique and depends on the choice of the scale parameter  $\sigma$ . Fig. 1 shows one example of the decomposition for a scale value of  $\sigma = 2.0$ .

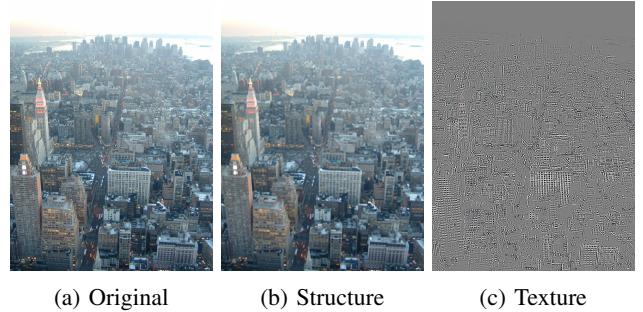


Fig. 1: Structure-Texture Decomposition

Imaging model based clarification using structure-texture decomposition [39] has been investigated in [40], where structure and texture components are processed separately. In one of our experiments related to the screened Poisson method described in the previous section, we also utilize structure-texture decomposition. First the image is decomposed into the corresponding components using

the algorithm described above and then the enhancement method is applied only to the structure component. The output is obtained by adding the unaltered texture back. The motivation like in [40] is to take noise into account. Here the noise is kept isolated in the texture part and not enhanced in the screened Poisson method which acts like a high pass filter.

### C. Dark-Channel Prior based Haze Removal

The dark channel prior (DCP) method [12] is a kind of statistics of the haze-free outdoor images. It relies on the observation that most local patches in haze-free outdoor images have some pixels with very low intensities in at least one color channel. This key observation has been validated in [12] on outdoor image sets from flickr.com and several other image search engines using 150 most popular tags annotated by the flickr users, where it has been shown that about 75% of the pixels in the dark channels have zero values, and the intensities of 90% of the pixels are below 25. Based on this observation, the haze degrading the image, airlight ( $A$ ) can be estimated along with the haze-free image.

Dark channel prior dictates that minimizing over local patches in outdoor images over three intensity channels (red, green, blue) gives very low intensity values. Mathematically, this minimization step is carried out as follows:

$$J^{dark}(x) = \min_{c \in \{r, g, b\}} \left( \min_{y \in \Omega(x)} (J^c(y)) \right) \quad (6)$$

where  $J^c$  is a color channel corresponding to true radiance  $J$  and  $\Omega(x)$  is a local patch centered at pixel  $x$ . In fact, there are three essential causes for this low intensity in outdoor scenes when minimizing over local patches:

- 1) Colorful object or surfaces
- 2) Shadows. e.g., the shadows of buildings, cars, etc.
- 3) Dark objects or surfaces. e.g., dark tree trunk and stone.

To eliminate haze, DCP method works in four main steps: a) Estimating the atmospheric light. b) Estimating the transmission  $t(x)$ . c) Soft matting. d) Recovering the scene radiance.

**Estimating the atmospheric light:** Airlight is estimated usually from the most haze-opaque (brightest) pixel in the degraded image. Nevertheless, this might be an inaccurate estimation in some cases where the brightest pixel of the image comes from a white object in the captured scene. Therefore, a rather better estimation can be obtained using the dark channel of the hazy image, which is used to approximate the denseness of the haze superimposed on the original image. Obtaining this estimation starts by selecting the top 0.1% brightest pixels in the dark channel. These pixels are the most haze-opaque compared to others. It is worth noting that these pixels may not be brightest in the whole image. The pixel with the highest intensity among these pixels or the average of these pixels is

selected as the atmospheric light in the corresponding image.

**Estimating the transmission:** The haze formation model followed in this method is [41]:

$$I(x) = t(x) \times J(x) + [1 - t(x)] \times A \quad (7)$$

where  $x$  denotes the pixel coordinates,  $I$  is the observed hazy image, and  $J$  is the true radiance of scene points. The term  $A$  represents the airlight and can be seen as the value of observed  $I$  in image areas where  $t = 0$ . The scene transmission  $t(x)$  is distance-dependent:

$$t(x) = e^{-\beta d(x)} \quad (8)$$

where  $d(x)$  is the distance from the scene point, captured at pixel  $x$ , to the camera lens, and  $\beta$  is the attenuation coefficient of the atmosphere. Although  $\beta$  is wavelength dependent and therefore  $t$  is different per color channel, this assumption is relaxed usually in single image dehazing methods as it is the case in DCP Dehazing method. It is also assumed that the transmission  $\tilde{t}(x)$  in a local patch is constant. Applying (6) on the general haze model (7) over the three channels gives the following:

$$\min_c \left( \min_{y \in \Omega(x)} \left( \frac{I^c(y)}{A^c} \right) \right) = \tilde{t}(x) \min_c \left( \min_{y \in \Omega(x)} \left( \frac{J^c(y)}{A^c} \right) \right) + (1 - \tilde{t}(x)) \quad (9)$$

According to the DCP, the dark channel  $J^{dark}$  of the haze-free radiance  $J$  should tend to be zero:

$$J^{dark}(x) = \min_c \left( \min_{y \in \Omega(x)} (J^c(y)) \right) = 0 \quad (10)$$

Since  $A^c$  is always positive, the first term on the right-hand side of (9) will be equal to zero. Eventually, this leads to the following form of (9):

$$\tilde{t}(x) = 1 - \min_c \left( \min_{y \in \Omega(x)} \left( \frac{I^c(y)}{A^c} \right) \right) \quad (11)$$

which is the dark channel of the normalized haze image  $\frac{I^c(y)}{A^c}$ . As the haze provides an indication for humans to perceive depth, removing it entirely might look unnatural and the feeling of depth may get lost. Hence, a small amount of haze should be kept in the image for distant objects. This can be accomplished as follows:

$$\tilde{t}(x) = 1 - \omega \min_c \left( \min_{y \in \Omega(x)} \left( \frac{I^c(y)}{A^c} \right) \right) \quad (12)$$

This modification has the advantage of adaptively keeping more haze for further objects. The value of  $\omega$  is indeed application-based.

**Soft Matting:** In this step, a soft matting is applied to refine the estimated transmission map, denoted as  $t$ , where the following cost function is minimized w.r.t  $t$ :

$$E(t) = t^T L t + \lambda (t - \tilde{t})^T (t - \tilde{t}) \quad (13)$$

given that  $L$  is the Matting Laplacian matrix proposed by Levin [42], and  $\lambda$  is a regularization parameter. The first term is the smooth term and the second term is the data term. The optimal  $t$  can be obtained by solving the following sparse linear system:

$$(L + \lambda U)t = \lambda \tilde{t} \quad (14)$$

where  $U$  is an identity matrix of the same size as  $L$ . As a result of solving the previous equation, a refined transmission map is obtained which manages to capture the sharp edge discontinuities and outline the profile of the objects.

**Recovering the Scene Radiance:** After obtaining the transmission map and airlight ( $A$ ), the original scene radiance  $J(x)$  can be recovered using (7). However, the direct attenuation term  $J(x)t(x)$  can be very close to zero when the transmission  $t(x)$  is around zero. Furthermore, the directly recovered scene radiance  $J$  is prone to noise. Therefore, it is necessary to restrict the transmission  $t(x)$  to a lower bound  $t_0$  preserving a small amount of haze in highly dense haze regions. Consequently,  $J(x)$  is recovered by the following equation:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A \quad (15)$$

#### D. Non-local Image Dehazing

One of the state of the art image dehazing methods that has been proposed recently is Non-Local Image Dehazing [13]. Non-Local dehazing is based on the supposition that colors of haze-free images are well approximated by a few hundred distinct colors that form tight clusters in RGB space. This prior has been validated and quantified on the Berkeley Segmentation Dataset (BSDS300), which is a diverse dataset of clear outdoor natural images representing an archetype of scenes that might be degraded by haze. Moreover, pixels in a given cluster are often non-local, i.e. those pixels are spread over the entire image plane at different distances from the camera. Due to the impact of haze on image pixels, each cluster in the undegraded image turns into a line in the RGB space in the hazy image. Those haze lines are used to recover the haze free image and depth map.

The haze formation model followed in this method is the same as the model used in DCP (7). Equation (7) has indeed three observations  $I(x)$ , given that  $A$  is known or estimated, and four unknowns:  $J(x)$  and  $t(x)$ . This results in an under-determined estimation problem, which necessitates introducing priors to regularize this system.

The non-local prior is demonstrated in Fig.2 by [13], where K-means is deployed to cluster a haze-free image to 500 clusters. Distinct color markers characterizing pixels

belonging to four of these clusters are shown in Fig. 2a and their RGB coordinates are plotted in Fig. 2b, proving tight clusters. It can be noticed that the clusters have pixels spread over the entire image with different distances from the camera. A synthetic hazy image was generated from the clear image (Fig. 2c) by [43]. The same pixels as in Fig. 2a are marked. However, because of the added haze, pixels that belonged to the same color cluster are not similar anymore. This is depicted in RGB space in Fig. 2d, where the color coordinates of these pixels are distributed along a haze-line spanned by the original color and the airlight.

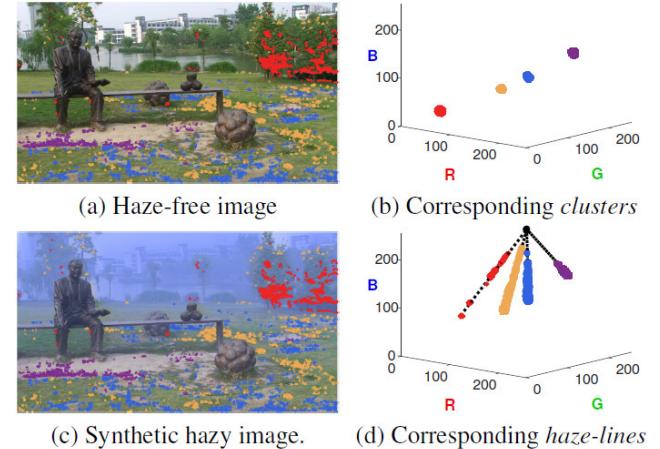


Fig. 2: Non-Local Image Dehazing. (a) Clustering of a haze free color image pixels using K-means. (b) The four color clusters are depicted in RGB space. (c) Synthetic haze is added to (a). (d) The hazy pixels plotted in RGB color space. Image Courtesy: [13]

Non-Local Dehazing algorithm consists of four chief steps: clustering the pixels into haze-lines, estimating an initial transmission map, regularization, and dehazing.

**Finding Haze-Lines:**  $I_A$  is defined as the hazy image representation in the 3D RGB space with the airlight at the origin of this system:

$$I_A = I(x) - A \quad (16)$$

Substituting the value of  $I(x)$  from (7) into 16 gives the following:

$$I_A(x) = t(x) \cdot [J(x) - A] \quad (17)$$

where the previous equation is represented in the Cartesian coordinates system given that  $J(x) = [j_1, j_2, j_3]$  and  $A = [a_1, a_2, a_3]$  are 3D vectors with  $x$ ,  $y$  and  $z$  coordinates and  $t(x)$  is a scalar.  $I_A$  can also be formulated in the spherical system,  $I(x) = [r(x), \theta, \phi]$ , using the equations below:

$$r(x) = \sqrt{x^2 + y^2 + z^2} = t(x) \times \|J(x) - A\| \quad (18)$$

$$\theta(x) = \arctan\left(\frac{y}{x}\right) = \arctan\left(\frac{(j2 - a2)}{(j1 - a1)}\right) \quad (19)$$

$$\phi(x) = \arctan\left(\frac{\sqrt{(j1 - a1)^2 + (j2 - a2)^2}}{(j3 - a3)}\right) \quad (20)$$

As it can be viewed from the previous three equations,  $r(x)$  depends on the value of  $t(x)$  for each pixel while  $\theta$  and  $\phi$  are only dependent on  $J(x)$  and  $A$ . This means that if two pixels in the original image have the same RGB value, those pixels will have the same  $\theta$  and  $\phi$  values but different  $r(x)$  when they are located at dissimilar distances from the camera lens:

$$J(x) \approx J(y) \Rightarrow \{\phi(x) \approx \phi(y), \theta(x) \approx \theta(y)\}, \forall t$$

In order to determine which pixels are on the same haze-line, pixels should be grouped according to their angles  $[\phi, \theta]$  with a 2-D histogram binning of  $\theta$  and  $\phi$  in the range  $[0, 2\pi] \times [0, \pi]$ . According to the analysis described before, several hundreds of haze-lines represent an image with a good approximation.

**Estimating Initial Transmission:** From (18), it is evident that in a given haze-line defined by  $A$  and  $J(x)$ , the maximum value of  $r(x)$  corresponds to  $t(x) = 1$ :

$$r_{max} = \|J - A\| \quad (21)$$

Using (18) and (21),  $t(x)$  is obtained as follows:

$$t(x) = \frac{r(x)}{r_{max}} \quad (22)$$

In order to obtain an estimation of  $r_{max}$ , it is assumed that the farthest pixel from the airlight is haze free, and such a pixel radius is used as an estimation of  $r_{max}$  obtained for a haze-line  $H$  as follows:

$$\hat{r}(x) = \max_{x \in H} \{r(x)\} \quad (23)$$

This assumption is not valid for all of the haze-lines in an image, however the regularization step, which is explained later, partially compensates for it. The per pixel estimation of the transmission is computed eventually using the next equation:

$$\tilde{t}(x) = \frac{r(x)}{\hat{r}_{max}} \quad (24)$$

**Regularization:** As  $J$  is positive, (7) puts a lower bound on the transmission:

$$t_{LB}(x) = 1 - \min_{c \in \{R, G, B\}} \left\{ \frac{I_c(x)}{A_c} \right\} \quad (25)$$

An additional bound is imposed on the estimated transmission, per-pixel:

$$\tilde{t}_{LB} = \max\{\tilde{t}(x), t_{LB}(x)\} \quad (26)$$

The estimation in (24) is performed per-pixel, without imposing spatial coherency. This estimation can be inaccurate if a small amount of pixels were mapped to a particular haze-line, or in very hazy areas, where  $r(x)$  is very small and noise can affect the angles significantly. The transmission map should be smooth, except for depth discontinuities, when the input image is smooth. Mathematically, the following function is minimized w.r.t.  $\hat{t}(x)$ :

$$\sum_x \frac{[\hat{t}(x) - \tilde{t}_{LB}(x)]^2}{\sigma(x)^2} + \lambda \sum_x \sum_{y \in N_x} \frac{[\hat{t}(x) - \hat{t}(y)]^2}{\|I(x) - I(y)\|^2} \quad (27)$$

where  $\lambda$  is a parameter for controlling the trade-off between the smoothness and data terms,  $\sigma(x)$  is the standard deviation of  $\tilde{t}_{LB}$ , which is calculated per haze-line, and  $N_x$  indicates the four nearest neighbors of  $x$  in the image plane [13].

**Dehazing:** After minimizing (27) and obtaining the value of  $\hat{t}(x)$ , the original image is reconstructed as follows:

$$\hat{J}(x) = \{I(x) - [1 - \hat{t}(x)]A\}/\hat{t}(x) \quad (28)$$

The method is summarized in Algorithm 1.

---

#### Algorithm 1 Non-local Dehazing

---

- 1: **Input:**  $I(x)$ ,  $A$
  - 2: **Output:**  $\hat{J}(x)$ ,  $\hat{t}(x)$
  - 3:  $I_A(x) = I(x) - A$
  - 4: Convert  $I_A$  to spherical coordinates to obtain  $[r(x), \theta(x), \phi(x)]$
  - 5: Cluster the pixels according to  $[\theta(x), \phi(x)]$ . Each cluster  $H$  is a haze-line
  - 6: **for** Each cluster  $H$  **do**
  - 7:     Estimate maximum radius:  $\hat{r}_{max}(x) = \max_{x \in H} r(x)$
  - 8: **end for**
  - 9: **for** Each pixel  $x$  **do**
  - 10:     Estimate transmission:  $\tilde{t}(x) = \frac{r(x)}{\hat{r}_{max}}$
  - 11: **end for**
  - 12: Perform regularization by calculating  $\hat{t}(x)$  that minimizes Eq. 24
  - 13: Calculate the dehazed image using Eq. 28
- 

#### E. Guided Filter

Guided filter [44] is a linear filter that filters the input image by considering a guidance image. Typically, the guidance image could be the input image itself or another different image. It has many applications in computer vision and computer graphics, including edge-aware smoothing, detail enhancement, HDR compression, image matting/feathering, joint upsampling, etc.

In the general linear translation-variant filtering process, the output at pixel  $i$  is represented as a weighted average of the input image  $p$ :

$$q_i = \sum_j W_{ij}(I)p_j \quad (29)$$

where  $i$  and  $j$  are pixel indexes, and  $W_{ij}$  is the filter kernel, which is a function of the guidance image  $I$  and independent of  $p$ . It is evident that the guided filter is linear function of the input image  $p$ . The bilateral filter is an archetype of such a filter.

Guided filter is based on a linear model between the guidance  $I$  and the filtering output  $q$ . The following equation presumes that  $q$  is a linear function of  $I$  in a window  $w_k$  centered at the pixel  $k$ :

$$q_i = a_k I_i + b_k, \forall i \in w_k \quad (30)$$

where  $(a_k, b_k)$  are some linear coefficients assumed to be constant in  $w_k$ . As this model is locally linear,  $I$  has an edge when  $q$  has the same,  $\nabla q = a \nabla I$ . The output  $q$  is also written as a function of the input image minus some unwanted components  $n$  like noise/textures:

$$q_i = p_i - n_i \quad (31)$$

The cost function shown below is minimized in order to obtain a solution that minimizes the difference between  $q$  and  $p$  while preserving the linear model (30).

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \epsilon a_k^2) \quad (32)$$

where  $\epsilon$  is a regularization parameter penalizing large  $a_k$ . The solution of the previous equation is the same as the one of linear ridge regression model [45] and is given by:

$$a_k = \frac{\frac{1}{|w|} \sum_{i \in w_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \epsilon} \quad (33)$$

$$b_k = \bar{p}_k - a_k \mu_k \quad (34)$$

where  $\mu_k$  and  $\sigma_k^2$  are the mean and variance of  $I$  in  $w_k$ ,  $|w|$  is the number of pixels in  $w_k$ , and  $\bar{p}_k = \frac{1}{|w|} \sum_{i \in w_k} p_i$  is the mean of  $p$  in  $w_k$ . Having obtained the linear coefficients  $(a_k, b_k)$ , the output  $q_i$  can be computed from (30) taking into account averaging all possible values of  $q_i$  as follows:

$$q_i = \bar{a}_i I_i + \bar{b}_i \quad (35)$$

where  $\bar{a}_i = \frac{1}{|w|} \sum_{k \in w_i} a_k$  and  $\bar{b}_i = \frac{1}{|w|} \sum_{k \in w_i} b_k$  are the average coefficients of all windows overlapping  $i$ . Since  $a$  and  $b$  are averaged in (35), their gradients are supposed to be much smaller than that of  $I$  near strong edges, indicating that sudden intensity changes in  $I$  can be quite preserved in the output image  $q$ . The previous three equations provide a definition of the guided filter.

One vital property of the Guided filter is called structure-transferring property. This property can be used to transfer the structure of the guidance image to the output image. One

application that exemplifies this property is the guided matting/feathering, which is used in some dehazing algorithms, e.g. DCP, relying on the fact that guided filter is closely related to the matting Laplacian matrix [44]. Therefore, guided filter can be used as an alternative to the soft matting step proposed initially in DCP with the added value of much higher computational speed and transferring more fine details to the refined map because of the structure-transferring property. Another prominent feature of this filter is that it is a fast and nonapproximate linear time,  $\mathcal{O}(N)$ , algorithm independently of the used kernel radius and the intensity range. This promotes the guided filter as a strong candidate in many applications that are computationally demanding such as video dehazing to speed up the computational time required to dehaze the entire video frames sequence.

#### F. Spatial-Temporal Information Fusion based Video Dehazing

Spatial-temporal information fusion framework for underwater video dehazing proposed in [14] estimates the transmission and background light separately to restore hazy frames in a video. It utilizes the same optical model for haze (7) used in Dark channel prior algorithm [12]. This model was originally proposed in [4]:

$$J_k(x) = \frac{I_k(x) - A_k(1 - t_k(x))}{t_k(x)}, k = 1 \dots N \quad (36)$$

where subscript  $k$  denotes the  $k^{th}$  frame and  $N$  is the total number of frames in the video.

The algorithm starts by estimating the transmission map of the first frame using dark channel prior followed by refinement with guided filter. The dark channel for the first frame can be expressed as:

$$J_1^{dark}(x) = \min_c (\min_{y \in \Omega(x)} J_1^c(y)) = 0 \quad (37)$$

The transmission in a local patch is assumed to be constant and is computed using:

$$t_1(x) = 1 - w \min_c (\min_{y \in \Omega(x)} \frac{I_1^c(y)}{A^c}) \quad (38)$$

where  $0 < w \leq 1$  is a constant parameter to keep a small amount of haze in the image so that the output looks natural. However, the assumption that the transmission is always constant in a patch is not always valid, so the transmission is refined using the guided filter:

$$\tilde{t}_1(x) = \sum_y W_{xy}(I_1) t_1(y) \quad (39)$$

where  $\tilde{t}_1(x)$  is the refined transmission,  $W_{xy}$  is filter kernel based on  $I_1$  which is set equal to the gray image of the frame.

To keep the colors consistent and reduce computational costs, the concept of spatial-temporal information fusion is

introduced which takes advantage of the fact that transmission maps of adjacent frames are similar. A linear translation-variant filter can be used to estimate the transmission map of the other frames. The guided filter is again utilized here:

$$t_k(x) = \sum_y W_{xy}(I_k) t_{k-1}(y) \quad (40)$$

Although the transmission maps in adjacent frames are similar, they are not exactly identical due to various factors like camera motion, water flow, moving objects etc. This is dealt with by adding an offset to the estimated transmission:

$$\tilde{t}_k(x) = t_k(x) + \Psi_k, k = 2 \dots N \quad (41)$$

where the offset is defined by minimizing the difference between the current and the first transmission using the least squares method over  $n$  tracking points. Tracking method proposed in [46] is used by the authors to automatically capture the same scene point in consecutive frames.

$$\Psi_k^* = \arg \min_{\Psi_k} \left( \sum_{i=1}^n [\tilde{t}_{ki}(x) - \tilde{t}_{1i}(y)]^2 \right) \quad (42)$$

While estimating the background light, if one assumes that airlight is constant, flickering artifacts start to appear in the video sequence because this assumption is not valid and the airlight is constantly changing among different frames. The background light, however, is not independent and does not change abruptly. This is resolved by associating the airlight of the previous frame with the current frame using an adjustment factor  $\alpha$ :

$$\tilde{A}_k = \alpha \tilde{A}_{k-1} + (1 - \alpha) A_k, k = 2 \dots N \quad (43)$$

The scene radiance is then recovered using the following equation:

$$J_k(x) = \frac{I_k(x) - \tilde{A}_k}{\max(\tilde{t}_k(x), t_0)} + \tilde{A}_k, k = 1 \dots N \quad (44)$$

where  $t_0$  is the lower bound on transmission. The authors in [14] proposed to divide the video into segments of  $N = 25$  and apply the fusion algorithm on each segment separately.

#### IV. PROPOSED ALGORITHM

In this section we present our video dehazing algorithm proposed in this article. The objective is to preserve consistency in terms of both transmission maps and airlight values across adjacent frames. It uses the same framework of spatial-temporal information fusion [14] as described in the previous section, but there are three major improvements proposed by us. Firstly, the method used to estimate the transmission map for the first frame is not the Dark Channel Prior method but the improved Non-local image dehazing method. The improvement is qualitatively and quantitatively analysed in the results section. The estimation of airlight from the Dark Channel method is still used in the process.

Secondly, here the video does not need to be divided into smaller videos of 25 frames as in [14], the algorithm can process a video of any size. In this method, an estimate of the transmission map for the second frame for instance is obtained by a linear interpolation of the transmission maps of the 1<sup>st</sup> frame and 11<sup>th</sup> frame, both of which are estimated using the Non-local dehazing method. This linear interpolation process for preserving consistency is repeated for all video frames in between two such frames at gaps of 10 which starts from the first frame. This estimate is then refined using the guided filter as in the spatial-temporal information fusion method, where the current frame acts as the guidance image. However, just this modification still suffers from the problem of flickering. This flickering is caused by the incoherent change in transmission values between the last frame of the previous set of 10 frames and the first frame of the next set of 10, where the estimate of the former frame is obtained from interpolation and that for the latter obtained from the direct application of Non-local dehazing. To circumvent this problem, the transmission estimate for the 11<sup>th</sup> frame for instance while dehazing that frame and moving forward is taken as an average of the transmission estimates of 11<sup>th</sup>, 1<sup>st</sup> and 21<sup>st</sup> frames. This is done for first frames in sets of 10, looking at one frame backward and forward at such a gap of 10. Then this transmission estimate is refined by the guided filter, similar to that for the intermediate frames as described. Here we are using the offline nature of the processing where frames at a gap of 10 are available. With this interpolation and averaging we are able to minimize flickering effects in the output.

Thirdly, this method does not use any addition of offset to the transmission values as in [14], which involves tracking of points across frames. In the approach of using transmission interpolation between first frames of sets of 10 frames, the effect of camera and object motion is captured. This eliminates the added complexity of tracking of scene points from the dehazing process.

The detailed pseudocode of the proposed algorithm is given in Algorithm 2.

#### V. EXPERIMENTAL RESULTS

##### A. Image Dehazing

This section presents the results of testing the image dehazing algorithms described in section III, namely, Screened Poisson enhancement (SP), Dark Channel Prior (DCP), and Non-Local Dehazing (NLD), on some hazy images affected by atmospheric or underwater haze, shown in Fig. 3, where comparisons are drawn emphasizing the pros and cons of each method based on subjective and objective measures.

As for the parameter settings, the experiments have been performed with the same set of parameter values for all images for fair comparison. For the SP method, the trade-off

---

**Algorithm 2** Proposed Video Dehazing Algorithm

---

```

1: Input Video with  $N$  frames
2: for  $i = 1$  to  $N$  do
3:   Read  $i^{th}$  frame
4:   if  $\text{modulo}(i, 10) == 1$  then
5:     if  $i == 1$  then
6:       Airlight  $A \leftarrow$  Dark Channel Method
7:       Transmission map of first frame in current set
      of 10 frames  $t_o \leftarrow$  Non-local dehazing method
8:       Transmission map of first frame in next set
       $t_n \leftarrow t_o$ 
9:       Transmission map of current frame  $t \leftarrow t_o$ 
10:      end if
11:
12:      if  $N - i \geq 10$  then
13:        Read  $(i + 10)^{th}$  frame
14:        Airlight  $nA$  for  $(i + 10)^{th}$  frame  $\leftarrow$  Dark
      Channel Method
15:         $nA \leftarrow 0.7A + 0.3nA$ 
16:        Transmission map of  $(i + 10)^{th}$  frame  $t_{nf} \leftarrow$ 
      Non-local dehazing method
17:      else
18:         $t_o \leftarrow t_n$ 
19:         $t_{nf} \leftarrow t_n$ 
20:      end if
21:
22:      if  $i > 1$  then
23:         $t \leftarrow \frac{1}{3}(t_n + t_o + t_{nf})$ 
24:         $t_o \leftarrow t$ 
25:      end if
26:
27:      if  $N - i \geq 10$  then
28:         $t_n \leftarrow t_{nf}$ 
29:      end if
30:
31:    else
32:       $tgap \leftarrow \text{modulo}(i, 10)$ 
33:      if  $tgap == 0$  then
34:         $tgap \leftarrow 10$ 
35:      end if
36:       $t \leftarrow t_o + \frac{tgap-1}{10}(t_n - t_o)$ 
37:    end if
38:
39:    if  $i > 1$  then
40:      Airlight  $A_l$  for  $i^{th}$  frame  $\leftarrow$  Dark Channel
      Method
41:       $A \leftarrow 0.7A + 0.3A_l$ 
42:    end if
43:
44:    Refined  $nt \leftarrow t$  using Guided filter with grayscale
       $i^{th}$  frame as guidance image
45:    Compute dehazed frame  $I_{out}$  from  $nt$  and  $A$ 
46:    Write frame  $I_{out}$  to output
47:  end for

```

---

TABLE I: Average Gradient of test images for different dehazing algorithms. (SP: Screened Poisson Enhancement, DCP: Dark Channel Prior, NLD: Non-Local Dehazing)

Image (a)	R	G	B	Average
Screened Poisson	3.6482	2.0700	<b>3.3902</b>	3.0361
Dark Channel Prior	2.7298	2.0125	2.1937	2.312
Non-Local Dehazing	<b>3.7276</b>	<b>2.4406</b>	3.2406	<b>3.1362</b>
Image (b)	R	G	B	Average
Screened Poisson	3.5708	2.9702	<b>3.8904</b>	3.4771
Dark Channel Prior	2.9163	2.2838	2.6644	2.6215
Non-Local Dehazing	<b>3.9455</b>	<b>3.0295</b>	3.5078	<b>3.4942</b>
Image (c)	R	G	B	Average
Screened Poisson	3.7111	3.5212	2.8741	3.3688
Dark Channel Prior	7.2691	5.7272	5.3061	6.1008
Non-Local Dehazing	<b>8.4363</b>	<b>7.0553</b>	<b>6.3048</b>	<b>7.2654</b>
Image (d)	R	G	B	Average
Screened Poisson	6.6755	6.8873	7.0667	6.8765
Dark Channel Prior	7.5385	7.2676	7.2183	7.3414
Non-Local Dehazing	<b>8.2923</b>	<b>9.6834</b>	<b>9.6253</b>	<b>9.2</b>

TABLE II: Computational time (in seconds) for different dehazing algorithms. (SP: Screened Poisson Enhancement, DCP: Dark Channel Prior, NLD: Non-Local Dehazing)

Image	Resolution	SP	DCP	NLD
Image (a)	1037x778	4.20	4.74	8.99
Image (b)	1024x768	3.67	4.65	8.60
Image (c)	400x600	1.08	1.56	2.77
Image (d)	576x768	1.95	2.32	4.31

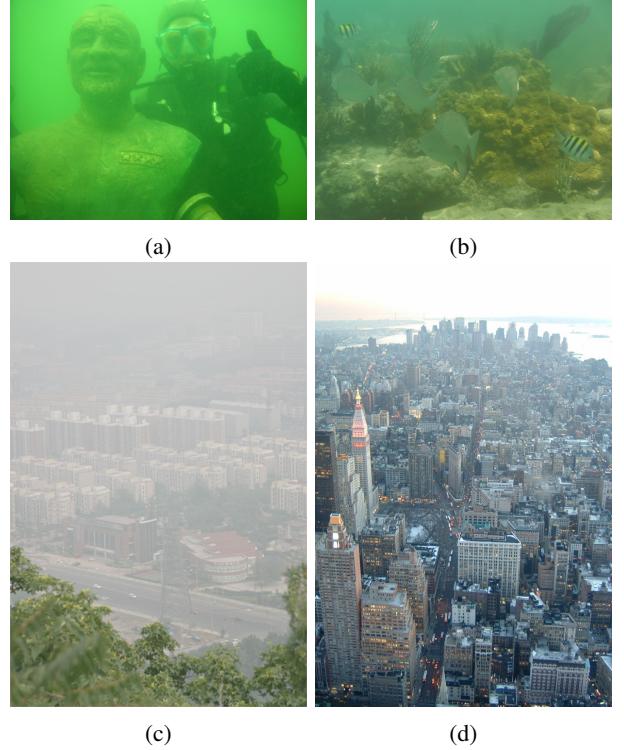


Fig. 3: Test hazy images

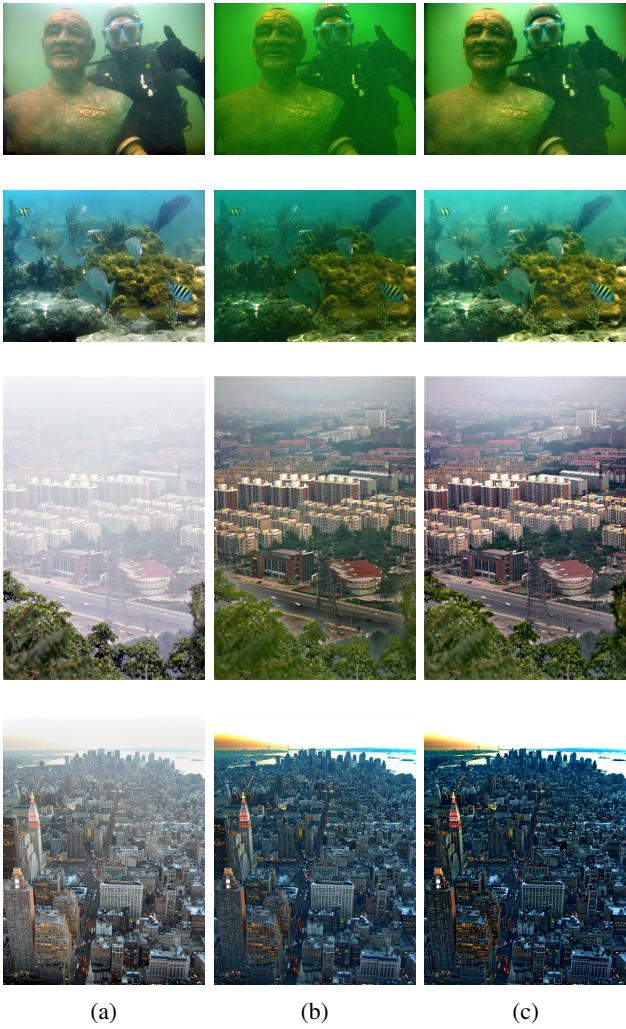


Fig. 4: Dehazed images using: (a) SP (b) DCP (c) NLD

parameter  $\lambda$  is set to 0.001, and the percentage of color saturation is 0.2. In the DCP method, the  $\omega$  parameter to preserve little haze is 0.95, whereas the lower bound on transmission  $t_0$  is 0.1. The size of the patch is 5 by 5. For guided filter refinement, the window size parameter is 15 (31 by 31 patch) and the  $\epsilon$  parameter is 0.001. No post-processing is performed on the output of DCP method to increase its exposure. For NLD, the parameter used for radiometric correction  $\gamma$  is 1. Most of these parameters are already optimized as provided by the authors of the algorithms.

In order to compare the results quantitatively, a gradient parameter average gradient  $G^c$  is computed for each dehazed image, where  $c$  denotes the color channel.  $G^c$  measures the local variance among the pixels of an image. Higher values of the metric denote better dehazing performance. It is defined as in [14], where we set  $N = 1$  for images:

$$G^c = \left( \frac{1}{(R-1)(L-1)} \times \sum_{i=1}^{R-1} \sum_{j=1}^{L-1} \sqrt{\frac{(I(i,j,c) - I(i+1,j,c))^2 + (I(i,j,c) - I(i,j+1,c))^2}{2}} \right)$$

Table I demonstrates the quantitative results, in terms of average gradient  $G^c$ , of applying SP, DCP, and NLD on the four test images shown in Fig. 3. It can be noticed from this table that NLD method always produces the highest value of  $G^c$  for all channels except in two images (a and b) where the highest value of  $G^c$  for the blue channel is obtained by SP method. In terms of the  $G^c$  averaged over all channels, NLD has the best result for all the four tested images revealing the superiority of this method in dehazing outdoor and underwater images over SP and DCP. This superiority can be attributed to the non-local prior as well as the regularization process which prevents any pixel that does not comply with model assumptions from contributing to the final result, both of which are explained in section III-D. However, the same method might fail when the airlight superimposed on the original image is significantly large rendering the process of detecting haze lines difficult.

Qualitatively, SP appears to over-emphasize the intensity values of the resultant image producing always brighter images compared to the other two methods. This comes from the fact that SP works as a high-pass filter and does not take into account any specific haze model. Although DCP tends to produce quite dim results, the resultant images seem more natural than the SP dehazed images. Finally, NLD indeed produces quite better results than SP and DCP. NLD dehazed images have the tendency to be sharper without being over- or under-emphasized, despite the fact that in some cases SP and NLD generate visually comparable results which makes it difficult to judge based on only subjective measures.

Table II shows the execution times of the different dehazing algorithms on the test images. The experiments are conducted in MATLAB on a standard Intel(R) Core(TM) 2.50 GHz processor. The run times show that SP method which does not use any haze model is the fastest, whereas the NLD method which has the best dehazing performance is the slowest.

In Fig. 5 the results of using structure-texture decomposition with Screened Poisson enhancement as described in Section III-B are shown for one image. Gaussian noise with a standard deviation of 10 has been added to the original underwater image. The results of just screened Poisson enhancement on the noisy image are compared with that of first decomposing the noisy image into structure and texture components and applying the enhancement only to the structure part. No smoothing is applied here. Quantitative analysis in terms of Peak Signal to Noise Ratio (PSNR) values is also possible. The enhanced image obtained from the original image in absence of noise is considered as ground truth. The higher PSNR in the output for the second case does show the usefulness of processing structure and texture components separately in presence of noise. However natural images do not contain synthetic noise like this case and from our experiments in absence of synthetic noise,

having this added step of structure-texture decomposition does not yield noticeable enhancement benefit.

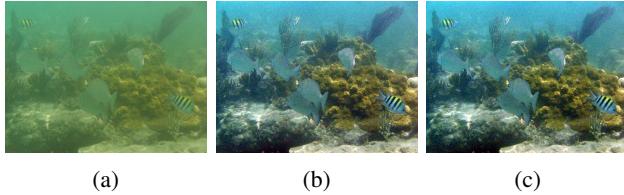


Fig. 5: Screened Poisson (SP) Enhancement with Structure-Texture (ST) Decomposition in presence of noise ( $\sigma = 10$ )  
(a): Original Image with noise (b): SP, PSNR = 29.86 dB  
(c): ST Decomposition + SP, PSNR = 33.28 dB

### B. Video Dehazing

In this section, results of video dehazing are presented. The proposed algorithm was tested on different underwater videos taken by various researchers and oceanographers. We compare our method with a modified version of the method presented in [14], where for the first frame in sets of 10 frames, the transmission map is estimated with the Dark Channel Prior method. The interval is 10 frames in both cases, but not 25 for the latter as used in [14]. Instead of finding the point correspondences for the latter method, we use the simple interpolation technique for transmission estimation as used in our method. As described in Section IV, for frames at gaps of 10 starting from the first frame, the transmission estimates are averaged with those of frames backward and forward at gaps of 10. This is done in our method to reduce flickering, which we do not add to the compared method. We believe this provides a ground of fair comparison for test videos without dividing them into smaller videos of 25 frames and without performing the computationally costly point tracking.

Results presented in this article are of videos with three different underwater scenes. Fig. 6, 7 and 8 display results of three frames for each of these videos. Visually, non-local based approach has the best results, where the contrast is more distinct, the colors are more natural and vibrant and the objects are clearer. In all of the three scenes, comparing the dehazed output of frame 10 for instance shows how our proposed method outperforms the dark channel prior based method. Also comparing the output of frame 10 with that of frame 11 for the proposed method, no incoherent color change is observed which minimizes the flickering artifact problem in the video.

Fig. 9 represents a typical frame in video 1. As one can observe, unlike in the case of non-local dehazing, there are halo effects around the scuba-diver in case of dark channel prior based method. The presence of halo-like effects around moving objects was observed throughout the video sequence. One of the main advantages of our non-local prior based approach is that we do not observe

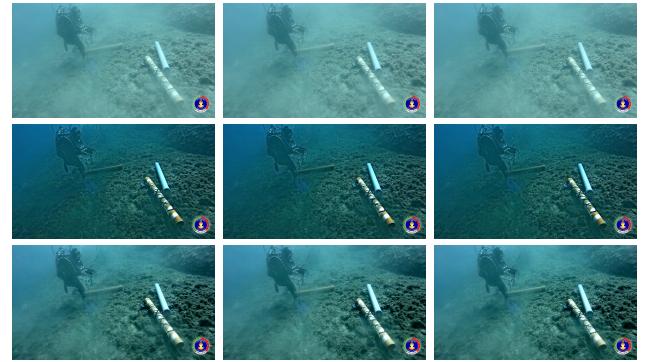


Fig. 6: Video 1: Output of the dehazing algorithms. Top to Bottom: Original video, Output of dark channel prior based approach, Output of non-local dehazing based approach. From Left to Right: Frame 1, 10, 11.



Fig. 7: Video 2: Output of the dehazing algorithms. Top to Bottom: Original video, Output of dark channel prior based approach, Output of non-local dehazing based approach. From Left to Right: Frame 1, 10, 11.

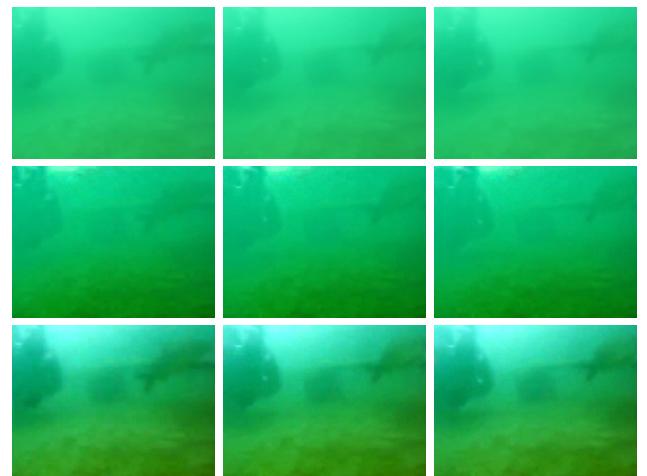


Fig. 8: Video 3: Output of the dehazing algorithms. Top to Bottom: Original video, Output of dark channel prior based approach, Output of non-local dehazing based approach. From Left to Right: Frame 1, 10, 11.

any such effects owing to a better prior.

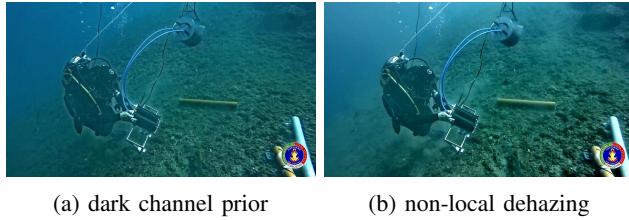


Fig. 9: Dehazed output of video 1 (frame 1600).

A quantitative analysis was also carried out for different dehazing approaches. Similar to the analysis carried out for images, average gradient was used to validate the results. Table III displays quantitative results. Average Gradient can be thought of a way of measuring the contrast between two objects sharing the same outline or boundary in the frame. Thus higher values can be related to higher contrast. On almost all the videos we tested, the proposed approach with non-local dehazing was giving the best results. In the results we present in this article, video 2 has been purposely included to show that sometimes, the algorithm performs worse than the dark channel based version. This is because the non-local dehazing method, although in general producing the best dehazed output of the methods we tested, is not a perfect technique. A few drawbacks of the technique where the non-local prior is not valid have been discussed in the previous sections.

TABLE III: Quantitative analysis (Average gradient) of Spatial-temporal fusion based video dehazing algorithm.

Video 1	R	G	B	Average
Original	1.1605	1.1265	1.1532	1.1467
Dark Channel Prior	1.8653	1.9138	1.9106	1.8965
Non-Local Dehazing	<b>1.9981</b>	<b>2.0700</b>	<b>2.0721</b>	<b>2.0467</b>
Video 2	R	G	B	Average
Original	0.3869	0.3710	0.4007	0.3862
Dark Channel Prior	<b>0.8287</b>	<b>0.7888</b>	<b>0.8234</b>	<b>0.8136</b>
Non-Local Dehazing	0.7668	0.6885	0.7610	0.7388
Video 3	R	G	B	Average
Original	0.1495	0.1052	0.1009	0.1185
Dark Channel Prior	0.2544	0.2009	0.2111	0.2221
Non-Local Dehazing	<b>0.3610</b>	<b>0.2779</b>	<b>0.3027</b>	<b>0.3139</b>

A comparison in terms of computational time between different dehazing approaches is shown in Table IV. Non-local dehazing and dark channel prior based algorithms were compared together for both the static dehazing approach in which all the frames are dehazed independent of each other and also the spatial-temporal fusion approach discussed earlier. The processing rates displayed are for experiments conducted in MATLAB on a standard Intel(R) Core(TM) 2.50 GHz processor. As one can observe, static dehazing approach is slower than the spatial-temporal based approach. Like in images, one drawback of using NLD algorithm is

that it is slower than the DCP algorithm, however the visual output is superior as discussed before. The computational efficiency can always be increased by GPU implementations.

TABLE IV: Computational time (Frame-rate) for various dehazing algorithms. (DCPS: Static Dark Channel Prior, NLDS: Static Non-Local Dehazing, DCPST: Spatial-Temporal fusion Dark Channel Prior, NLDST: Spatial-Temporal fusion Non-Local Dehazing (proposed approach))

Video Sequence	Frame Resolution	DCPS (fps)	DCPST (fps)	NLDS (fps)	NLDST (fps)
Video 1	1920x1080	0.074	0.109	0.037	0.065
Video 2	640x324	0.99	1.43	0.67	0.87
Video 3	640x480	0.63	0.92	0.45	0.59

## VI. CONCLUSION AND FUTURE WORK

In this article, an extensive comparison was carried out between different existing dehazing algorithms. Screened Poisson equation based contrast enhancement, dark channel prior based haze removal and non-local dehazing algorithms were studied in depth. Based on the analysis of the state of the art, an improved method for video dehazing utilizing the framework of spatial-temporal information fusion was also proposed . The method was tested on different underwater videos and was observed to produce better results.

Some possible future work may include integration of different water-based models as a replacement for the used optical model. Another topic of interest can be dehazing algorithms with complex models where different transmission coefficients for each color channel are considered. In this work the results were verified by visual analysis and a quantitative measure called average gradient. Alternative experiments can be carried out in which images of a color pattern are taken both in air and in hazy water. The images taken in air can be used as ground truth in order to compute mean squared errors and peak signal to noise ratios, which provide better quantitative evaluation of performance. The implementation of the algorithms was performed in MATLAB. Computational time can be improved by utilizing C++/Python and GPU implementations.

## VII. ACKNOWLEDGEMENTS

For the images used for testing as shown in Fig. 3, Images (a) and (b) have been used from [47]. Image (c) is from the authors of [13]. Image (d) is taken from the single image visibility restoration comparison database [19]. For the tests on video dehazing, video 1 related to the work in [48] and videos 2 and 3 related to work in [47] are used.

## REFERENCES

- [1] Srinivasa G. Narasimhan, Shree K. Nayar. "Vision and the atmosphere." *International Journal of Computer Vision*, Volume 48, Issue 3, pp. 233254, 2002.

- [2] M.C.W. van Rossum, Th M. Nieuwenhuizen. "Multiple scattering of classical waves: microscopy, mesoscopy, and diffusion." *Reviews of Modern Physics*, Volume 71, No. 1, pp. 313, 1999.
- [3] Shriya Sharma, Sakshi Bhalla. "Improved Haze Removal of Underwater Images using Particle Swarm Optimization." *International Journal of Computer Applications*, Volume 122, No.4, 2015.
- [4] H. Koschmieder, "Theorie der horizontalen Sichtweite: Kontrast und Sichtweite." Keim & Nemnich, 1925.
- [5] Sarit Schwartz, Einav Namer, Yoav Y. Schechner. "Blind haze separation." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2006*, Vol. 2, pp. 1984-1991.
- [6] J. N. Lythgoe. "Polarized light and underwater vision." *Nature*, Volume 213, pp. 893-894, 1967.
- [7] Yoav Y. Schechner, Srinivasa G. Narasimhan, Shree K. Nayar. "Instant dehazing of images using polarization." *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2001*, Volume 1, pp. I-325.
- [8] Yoav Y. Schechner, Nir Karpel. "Clear underwater vision." *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004*, Volume 1, pp. I-536.
- [9] Fabio Cozman, Eric Krotkov. "Depth from scattering." *Proceedings., 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 1997*, pp. 801-806.
- [10] S.G. Narasimhan and S.K. Nayar, "Contrast restoration of weather degraded images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713-724, 2003.
- [11] Jean-Michel Morel, Ana-Belen Petro, Catalina Sbert. "Screened Poisson equation for image contrast enhancement." *Image Processing On Line*, Volume 4, pp. 16-29, 2014.
- [12] K. He, J. Sun and X. Tang, "Single image haze removal using dark channel prior." *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341-2353, 2011.
- [13] D.Berman, T. Treibitz, S. Avidan. "Non-Local Image Dehazing." *The IEEE Conference on Computer Vision and Pattern Recognition CVPR 2016*, pp. 1674-1682.
- [14] C. Qing, F. Yu, X. Xu, W. Huang, J. Jin. "Underwater video dehazing based on spatialtemporal information fusion." *Multidimensional Systems and Signal Processing*, pp. 1-16, 2016.
- [15] R.T. Tan, "Visibility in bad weather from a single image." *IEEE IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [16] J. Zhang et al., "Local albedo-insensitive single image dehazing." *The Visual Computer*, vol. 26, no. 6-8, pp. 761-768, 2010.
- [17] R. Fattal, "Single image dehazing." *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 72, 2008.
- [18] J.H. Kim, J.Y. Sim, and C.S. Kim, "Single image dehazing based on contrast enhancement." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1273-1276.
- [19] J.P. Tarel and N. Hautiere. "Fast visibility restoration from a single color or gray level image." *IEEE 12th International Conference on Computer Vision*, 2009, pp. 2201-2208.
- [20] E. Ullah, R. Nawaz and J. Iqbal, "Single image haze removal using improved dark channel prior." *Proceedings of IEEE International Conference on Modelling, Identification & Control (ICMIC)*, 2013, pp. 245-248.
- [21] C.T. Chu and M.S. Lee, "A content-adaptive method for single image dehazing." *Pacific-Rim Conference on Multimedia*, Springer Berlin Heidelberg, pp. 350-361, 2010.
- [22] Jing Yu and Qingmin Liao, "Fast single image fog removal using edge-preserving smoothing." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1245-1248.
- [23] D. Park, D.K. Han, and H. Ko, "Single image haze removal with WLS-based edge-preserving smoothing filter." *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 2469-2473.
- [24] Z. Farbman et al., "Edge-preserving decompositions for multi-scale tone and detail manipulation." *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 67, 2008.
- [25] C.H. Yeh et al., "Efficient image/video dehazing through haze density analysis based on pixel-based dark channel prior." *IEEE International Conference on Information Security and Intelligence Control (ISIC)*, 2012, pp. 238-241.
- [26] J. Zhang et al., "Video dehazing with spatial and temporal coherence." *The Visual Computer* vol. 27, no. 6-8, pp. 749-757, 2011.
- [27] J.H. Kim et al., "Temporally x real-time video dehazing." *19th IEEE International Conference on Image Processing*, pp. 969-972, Sep 2012.
- [28] M. Wang et al., "Component-Based Distributed Framework for Coherent and Real-Time Video Dehazing." submitted for *Distributed, Parallel, and Cluster Computing*, Sep 2016.
- [29] J.Y. Chiang and Y.C. Chen, "Underwater image enhancement by wavelength compensation and dehazing." *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756-1769, Apr 2012.
- [30] N. Carlevaris-Bianco, A. Mohan, and R.M. Eustice, "Initial results in underwater single image dehazing." *OCEANS 2010 MTS/IEEE SEATTLE*, pp. 1-8, Sep 2010.
- [31] Z. Chen et al., "Region-specialized underwater image restoration in inhomogeneous optical environments." *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 9, pp. 2090-2098, 2014.
- [32] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter." *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 41-50, 2014.
- [33] K.N. Chaudhury, D. Sage and M. Unser, "Fast bilateral filtering using trigonometric range kernels." *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3376-3382, 2011.
- [34] X. Zhao, T. Jin and S. Qu, "Deriving inherent optical properties from background color and underwater image enhancement." *Ocean Engineering*, vol. 94, pp. 163-172, 2015.
- [35] N. Limare, J.L. Lisani, J.M. Morel, A.B. Petro, and C. Sbert. "Simplest Color Balance." *Image Processing On Line*, Volume 1, 2011.
- [36] Antoni Buades, Triet Le, Jean-Michel Morel, Luminita Vese. "Cartoon + Texture Image Decomposition." *Image Processing On Line*, Volume 1, pp. 1-8, 2011.
- [37] Antoni Buades, Triet Le, Jean-Michel Morel, Luminita Vese. "Fast cartoon + texture image filters." *IEEE Transactions on Image Processing*, Volume 19, Issue 8, pp. 1978-1986, 2010.
- [38] Y. Meyer. "Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures", *American Mathematical Society*, Volume 22, 2001.
- [39] L.I. Rudin, S. Osher, E. Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D: Nonlinear Phenomena*, Volume 60, Issue 1, pp. 259268, 1992.
- [40] M. Toda, K. Senzaki, M. Tsukada. "Image Clarification Method Based on Structure-Texture Decomposition with Texture Refinement." *International Conference on Image Analysis and Processing 2015*, Springer International Publishing, pp. 352-362.
- [41] W. E. K. Middleton. "Vision through the atmosphere." *Geophysik II/Geophysics II*, Springer Berlin Heidelberg, 1957. pp. 254-287.
- [42] A. Levin, D. Lischinski and Y. Weiss, "A closed form solution to natural image matting.", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, 2008.
- [43] R. Fattal, "Dehazing using color-lines." *ACM Transactions on Graphics (TOG)*, vol. 34, no. 1, pp. 13, 2014.
- [44] K. He, J. Sun and X. Tang, "Guided image filtering." *European conference on computer vision ECCV 2010*, Springer, pp. 1-14.
- [45] N. Draper and H. Smith, *Applied Regression Analysis*, 2<sup>nd</sup> ed., John Wiley, 1981.
- [46] K. Zhang et al., "Fast visual tracking via dense spatio-temporal context learning." *European Conference on Computer Vision*, Springer International Publishing, pp. 127-141, 2014.
- [47] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, "Enhancing underwater images and videos by fusion." *Proc. of IEEE Conference on Computer Vision and Pattern Recognition CVPR 2012*, pp. 8188.
- [48] F. Oleari, et al., "An Underwater Stereo Vision System: from Design to Deployment and Dataset Acquisition." *Proceedings of the IEEE/MTS OCEANS*, pp 1-6, May 2015.