ABSTRACT

We present in this paper a fast single image defogging method that uses a novel approach to refining the estimate of amount of fog in an image with the Locally Adaptive Wiener Filter. We provide a solution for estimating noise parameters for the filter when the observation and noise are correlated by decorrelating with a naively estimated defogged image. We demonstrate our method is 50 to 100 times faster than existing fast single image defogging methods and that our proposed method subjectively performs as well as the Spectral Matting smoothed Dark Channel Prior method.

Index Terms— Defogging, Visibility, Image Enhancement, Bilateral Filter

1. INTRODUCTION

Contrast enhancement is often needed for image and video capture applications possibly due to lens configurations, haze, or fog. Consider the possible scenarios where contrast enhancements are needed for a vehicle driving down a foggy road or a surveillance camera viewing through a fog layer near the ocean. To obtain higher contrast imagery immediately, each image or frame from a video stream (un-compressed) is needed to be enhanced in near-real time.

A first attempt to providing a solution to the problem would be to employ Histogram Equalization (HE), a statistical contrast enhancement method where the probability distribution function (PDF) of the image is transformed to be more uniform. Another similar approach is to enforce Gray-Level-Grouping [1] which also achieves spreading the original PDF while preserving dark and/or light pixels.

A common problem with fog in scenes is that the amount of contrast degradation is spatially varying. This can be seen in everyday life by observing the amount of haze present in front of two distant mountains. The farthest mountain would appear more hazy. This effect has been an important technique by artists for conveying scene depth on a two dimensional canvas [2].

A reasonable approach to improving the spatially varying contrast is to design the contrast enhancement to be spatially varying as well. A few examples are the Adaptive HE [3], Contrast Limited Adaptive HE [4], Adaptive GLG [1].

The statistical methods attempt to achieve a desired PDF structure for an image without knowledge of the scene’s physical parameters. If there is prior knowledge of the scene then it makes sense to take advantage of this knowledge. For this paper, indeed the prior knowledge is that the scenes contain fog and we wish to apply a Single Image Defogging method.

Single Image Defogging attempts to estimate the amount of fog that exists in an image and then removes the fog which results in a higher contrast image. The byproduct from defogging is the scene depth map is also acquired since the amount of fog observed is a function of scene depth. There is a growing number of methods that enhance image contrast via Single Image Defogging [5, 6, 7, 8, 9].

In this paper, we address scenes that are affected by fog that need to be enhanced in near-real time for the purpose of video processing and computer vision applications. We present a new contrast enhancement that is based on the physics model of the scene to obtain an arbitrary to scale image of scene depth and an enhanced image. The speed of our method is achieved by employing separable filters that approximate the FIR Wiener filter for refining the image of scene depth estimates.

2. BACKGROUND

2.1. Atmospheric Dichromatic Model

Removing fog from a single image is an under constrained problem that requires an inference method or prior knowledge of the scene. The amount of fog observed in an image is dependent on the distance of the object to the camera, wavelength of the light, and the size of the scattering particles in the atmosphere [10]. Given a foggy image at pixel location \( i \), \( y_i \in \mathbb{R}^3 \), the “defogged” version \( x_i \in \mathbb{R}^3 \) is related to the foggy image with the atmospheric dichromic model [11],

\[
y_i = t_i x_i + (1 - t_i) a,
\]

where the airlight is \( a \in \mathbb{R}^3 \) and transmission is \( t_i \in \mathbb{R} \). The right hand side of (1) is considered the veiling [8, 12],

\[
v = (1 - t_i) a. \tag{2}
\]

For our proposed method, we make the assumption that the application is for scenes with fog or at least wavelength independent atmospheric scattering. This assumption allows the transmission to be a scalar quantity \([11]\) which validates the model in (1). The transmission is then dependent on a homogenous (spatially-invariant) \( \beta \) term and scene depth, \( r_i \),

\[
t_i = e^{-\beta r_i}. \tag{3}
\]

Assuming we know all of the desired parameters, then the defogged image is obtained with

\[
\hat{x}_i = \frac{y_i - a}{\min(t_i, t_0)} + a. \tag{4}
\]

where \( t_0 \) is chosen for numerical soundness \((t_0 = 0.01)\) [7].

Single Image Defogging methods estimate the veiling \( \nu = \nu a \) which is composed of the transmission \( t \) and airlight \( a \)(see (2)). We begin with the assumption that the airlight is known or estimated. We refer the reader to [5, 11, 7] for methods of estimating the airlight. The next step is to estimate the scalar veiling component.
If we make an assumption that one or more of the color components of \( x \) is dark, then we may take the minimum of all of the color components from \( y \) and measure the veiling with

\[
v = d_i a, \tag{5}
\]

and

\[
d_i = \min_{c \in (r, g, b)} y_i(c), \tag{6}
\]

where the \( c^{th} \) color component of \( y \) is \( y(c) \). This method of measuring a “dark prior” or “whiteness” of the image, \( d \), is commonly used in single image defogging models as the first step in estimating the transmission (or veiling) [7, 8, 9].

3. PROPOSED METHOD

3.1. Estimation Refinement with Adaptive Wiener Filter

The observation \( d \) (6) used as an estimate for veiling contains significant noise and needs refinement. This noise is due to many factors. Although statistically supported natural scenes [7], the assumption that at least one color component of a radiant object is zero is not necessarily valid. An example is when a bright colored paved road appears to be the same color as the horizon. There is ambiguity in distinguishing the range of the road with respect to the horizon.

Another factor that contributes to the noise component is the texture of the scene. As modeled in (2) and (3), we expect the veiling to be representative of the scene depth. The depth variation in scenes is typically piece-wise smooth and void of any texture components. The texture is not desired and needs to be removed to have an accurate veiling estimation. Otherwise, false colors are introduced when enhancing with (4).

Existing methods employ statistical smoothing operators [5, 8, 9] and Spectral Matting [7] to smooth the veiling estimate. In this paper, we employ the Locally Adaptive Wiener Filter for the purpose of refining our initial veiling estimate. Note that our approach significantly differs from Shuai et al. [13] because their method applies the Wiener filter as noise removal for the enhanced image and not as a refinement for the veiling estimate.

We follow the development of the Locally Adaptive Wiener Filter and refer the reader to Lim [14] and Jain [15] for a more complete development of the filter. We only present in this paper the necessary terms needed for describing the proposed method.

We wish to estimate \( v \) at pixel location \( i \) given the observation model

\[
d_i = v_i + n_i, \tag{7}
\]

where we characterize our observation as the desired veiling and additive noise (texture, color ambiguity). The random observation \( d_i \) is locally stationary within a sample window, \( \Omega_i \), both centered at pixel location \( i \). The signals outside the window are uncorrelated.

The veiling is considered a random variable with a mean and variance, \( v = (\mu_v + \sigma^2_v w_i) \) and \( w_i \sim N(0, 1) \). The noise is also zero mean but with a variance of \( \sigma^2_n, n \sim N(0, \sigma^2_n) \). The veiling and noise are expected to be independent, \( E[vn] = E[v]E[n] \).

Given (7), the Locally Adaptive Wiener filter within each sample window \( \Omega_i \) [14, 15] that estimates the atmospheric veil is

\[
\hat{v}_i = \mu_{v,i} + \frac{\sigma^2_{v,i} \sigma^2_n}{\sigma^2_{v,i} - \sigma^2_n} (d_i - \mu_{v,i}), \tag{8}
\]

The strength in this approach is that although simple to implement, the filter can adapt for scene depth discontinuities. It applies a weight, \( \frac{\sigma^2_{v,i} - \sigma^2_n}{\sigma^2_{v,i}} \), where if it is low then the smoothed sample is chosen whereas if the weight is high then the original signal is preserved. This bilateral filtering technique is effective for our purpose in smoothing textures but preserving depth discontinuities.

The difficulty with (8) is that the mean and variances are not known. An approach by [14] is to use the signal itself to estimate the local mean and variances,

\[
\hat{\mu}_v = \frac{1}{|\Omega_i|} \sum_{j \in \Omega_i} d_i, \tag{9}
\]

\[
\hat{\sigma}^2_v = \frac{1}{|\Omega_i|} \sum_{j \in \Omega_i} d^2_i - \hat{\mu}^2_v, \tag{10}
\]

where \( |\Omega_i| \) is the number of pixels within the local sample window. The general algorithm for the Adaptive Wiener Filter is outlined in Fig. 1.

We will address the estimation of the noise variance, \( \sigma^2_n \), in the following sections.

3.2. Naive Estimation of Noise Variance

The first step in our proposed method in estimating the noise variance is to take a naive approach and assume that (7) is valid; \( v \) and \( n \) are not correlated and the mean of \( n \) is zero. The variance of our observation in (7) becomes

\[
\text{var}[d_i] = \sigma^2_{d,i} = \sigma^2_{v,i} + \sigma^2_n. \tag{11}
\]

Similar to Tan [6] and Fattal [5], we expect the transmission (and hence veiling) to be correlated for large sample windows and exhibit low signal variance \( \sigma^2_v \). We naively assume \( \sigma^2_v << \sigma^2_n \) throughout the image and approximate the noise variance as a global average of the observation variance,

\[
\hat{\sigma}^2_n = \frac{1}{M} \sum_{j=0}^{M-1} \sigma^2_{d,j}, \tag{12}
\]

where \( \hat{\sigma}^2_{v,j} = \sigma^2_{d,j} \) which is defined in (10) and \( M \) is the total number of pixels in the entire image. Here we indicate a naive estimate with the superscript \( \hat{\cdot} \).

3.3. Noise Estimation Correction

The problem with with (12) is that the noise variance is dependent on the amount of fog. The texture of scene objects is attenuated by the fog and reduces the observation local variance [16]. The depth discontinuities exhibiting high variance in the veiling \( v \) will also correlate with edges in the image observation \( d \). This noise dependence invalidates the model in (7) because the noise and signal must be uncorrelated in order to develop the filter in (8).

We find a way to estimate an uncorrelated noise by examining the non-foggy scene component which can be considered a composition of a low-resolution and hi-resolution (texture, noise) component

\[
x_i = x_{L,i} + x_{H,i}. \tag{13}
\]
1: procedure WIENER DEFOG(y, Ω, t₀, w, steps)  ▷ First Step
2:  
3: a is estimated from [5, 11, 7]  
4: d′ ← from (6)  
5: ... The results are in Fig. 3. For timing measurement consistency, the sample window sizes were selected to be 32 × 32 instead of 1000 for digital images.  

3.4. Summary of Proposed Method

We denote the One-Step Wiener and Two-Step Defog methods with

\[ \hat{x}_{w_1}, \hat{t}_{w_1} \leftarrow WienerDefog(y, Ω, t₀, w, 1) \] (17)

\[ \hat{x}_{w_2}, \hat{t}_{w_2} \leftarrow WienerDefog(y, Ω, t₀, w, 2), \] (18)

respectively where the detailed algorithm is in Fig. 2. The foggy image is y. For all our experiments, we set w = 0.9 and t₀ = 0.01. The Two-Step Wiener Defog method is chosen when the defogged image has edges that appear like “halos” or exhibit “burn-in” appearances. The One-Step method can be used for most foggy scenes with less severe depth discontinuities.

The Adaptive Wiener Filter requires the local statistics to be estimated. These local statistics are measured within a square shaped moving sample window Ω with a size K × K. In [15], the size of the window was suggested to be at least 32 × 32 for digital images. Therefore a reasonable approach would be to select the sample window Ω to have a size of 32 × 32. For our experiments, the sample window size is adjusted for each image to improve subjective performance. The sizes varied from 22 × 22 to 45 × 45.

4. RESULTS AND DISCUSSION

4.1. Speed

The focus in this paper is on fast defogging methods therefore we investigate a subset of single image defogging methods that aim to improve the contrast under a time budget constraint. We compare the Visibresto method from Tarel and Hautière [8], Gaussian smoothed DCP (Gauss. DCP) from Long et al. [20], Median Dark Channel Prior (MDCP) from Gibson et al. [9], our One-Step Wiener Defog method, and the Two-Step Wiener Defog method (see Fig. 2). All of said methods are designed for near real-time speeds for videos with resolution near 720 × 480 resolution. The images were processed on a 64-bit 2.70GHZ processor. The software was written in Matlab without parallel processing enabled. There were 21 images processed with resolutions ranging from 465 × 384 to 3072 × 2304.

For all images, the diagonal length in pixels was recorded vs. time to process. The results are in Fig. 3. For timing measurement consistency, the sample window sizes were selected to be 32 × 32.
Fig. 4: (a) Original foggy image from [7]. (b) Transmission from One-step Wiener Defog. (c) One-step Wiener Defogged image. (d) Two-step Wiener Defogged image.

Fig. 5: Subjective comparison between DCP and our proposed method. For each column: (a) Original foggy image. (b) Defogged image from [7]. (c) One-Step Wiener Defogged image. (d) Transmission from [7]. (e) Estimated transmission from One-step Wiener Defog.

pixels in size for all methods. For all image sizes, the slowest to fastest methods were Visibresto, MDCP, Two-Step Wiener Defog, and a close tie between Gauss, DCP and One-Step Wiener Defog. The One-Step Wiener method is on average 100 times faster than the Visibresto method and almost 50 times faster than MDCP method. Although the method by [20] is as fast as our One-Step Wiener method, it does not account for depth discontinuities and suffers from significant halo artifacts. The Visibresto method and MDCP requires median filters to account for depth discontinuities which are significantly slower than the Wiener Defog method because the filter uses separable linear filters to estimate the local statistics in (9) and (10).

4.2. One-Step vs. Two-Step Wiener Defog

The estimation of the noise variance is important for preserving scene depth discontinuities in the veiling (hence transmission) estimate. In Fig. 4 we demonstrate the importance of updating the noise variance estimate. The transmission estimates, kw1 and kw2, are the One-Step and Two-Step Wiener Defog methods respectively. Notice the building tops in Fig. 4(b) are blurred. This is due to the local estimate of σv′ being either equal to or lower than the estimated global noise variance σv′ (see Fig. 1.2). After estimating the global noise variance with a naive defogged image, the tops of the buildings (edges and corners) are preserved in Fig. 4(c). Most importantly, the defogged image is improved with an updated noise variance estimate. The blurred depth discontinuities in the transmission estimate causes the defogged image to appear to have “burned” edges as seen on the building edges in 4(d). This “burn in” is removed in 4(d) after updating the noise variance.

4.3. Subjective Results

Due to the subjective nature of contrast enhancements, there has not been an established state-of-the-art method for single image defogging. We choose to compare our method to the Dark Channel Prior (DCP) method by He et al. [7] since it is the most common method compared to in single image defogging methods and its ability to generate a smooth transmission estimate. The DCP method, however, is complex and takes several seconds to minutes to process and is not considered a fast algorithm. We show in Fig. 5 that our Wiener Defogging method not only is extremely fast, but is subjectively comparable to the DCP method. More results can be viewed on our website at http://videoprocessing.ucsd.edu/˜kgibson.

5. CONCLUSION

We introduce in this paper the Wiener Defogging method that adapts to scene depth discontinuities. The goal of this method is to achieve faster processing speed for the purpose of near real-time uncompressed video processing. We demonstrate that our method is 50 to 100 times faster than existing methods because of the simplicity of our algorithm which estimates local statistics with a linear filter. We also demonstrate that our method is subjectively comparable to the DCP method by He et al. [7]. However instead of seconds to minutes to process the images, our method takes a fraction of a second.
6. REFERENCES


