Online Depth Estimation and Application to Underwater Image Dehazing

Younggun Cho, Young-Sik Shin and Ayoung Kim
Department of Civil and Environmental Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Korea 34141
Email: [yg.cho,youngsik,ayoungk]@kaist.ac.kr

Abstract—Underwater images captured in a turbid medium often suffer from significant degradation of visibility. Conventional dehazing approaches focus on dehazing a single image by using multiple channels for color restoration and rarely consider computational efficiency. This paper proposes an online dehazing method with sparse depth priors using an incremental Gaussian Process (iGP). The main contribution of this paper is developing a practically usable dehazing method for underwater robots using incoming sparse depth priors (range measurements) from any calibrated depth sensors. To deal with incoming depth priors efficiently, we adopt iGP for incremental depth map estimation and dehazing. Because the input vector of the iGP model is easily reconfigured, we can use the same update method for both color and gray images. Our method also estimates color-balanced veiling light to compensate for the color attenuation problem. For the evaluation, we first verify the proposed method on a open RGBD dataset and test it on real underwater color and gray images, comparing the results with those of previous methods.

I. INTRODUCTION

Underwater images often suffer from the poor visual conditions of water, such as haze effect, contrast loss and color distortion which increase the visual-perception problems of underwater robots. To extract useful visual information for use in underwater robotic applications such as SLAM (Simultaneous Localization and Mapping) [1, 2], an efficient dehazing method is required. Many dehazing methods have been proposed [3, 4, 5, 6, 7, 8, 9] because haze images are troubles not only for underwater images but also for outdoor images.

Fattal [3] predicts scene albedo and transmission using Independent Component Analysys (ICA). This method generates good results under the assumption that transmission and surface shading are uncorrelated. However, it requires high computation cost and fails to estimate transmission under thick haze conditions. Tarel [8] proposes a dehazing method for both color and gray images. This algorithm strengthens the global contrast assuming that the depth map is smooth except for the edges of objects. He [4] demonstrates single image dehazing using Dark Channel Prior (DCP). This idea is related to a strong prior that at least one of the channels among red, green, and blue of each pixel is low in haze-free images. DCP-based dehazing is effective and simple, but this prior is broken for sky and underwater backgrounds. Also, this method is hard to operate in real time because of its soft-matting technique, which requires high computation time. He [10] also proposes guided filters for time-effectiveness, but the method mispredicts transmission of underwater haze images. Zhu [5] proposes a fast dehazing method using Color Attenuation Prior (CAP). This method uses the effect of attenuation so that regions with more haze are related to high value and less saturation. The CAP method shows fast performance using simple training. However it also reveals limitations in a different environment, such as underwater.

In comparison with the previous four methods discussed [3, 8, 10, 5], Carlevaris-Bianco [6] proposes underwater single-image dehazing methods using wavelength-dependent attenuation of light in water. This method calculates the depth prior from the strong difference in attenuation between color channels; then the prior is processed to recover the transmission using Markov random field (MRF). Using MRF shows good dehazing performance, but it requires large processing time. To overcome this issue, Ancuti [7] proposes multi-scale fusion-based underwater image dehazing with two generated inputs. It is time effective; however, it cannot construct relative depth or a transmission map. Babaee [9] introduces hybrid dehazing methods by the fusion of optical images and acoustic images. With previous research of sensor fusion, Babaee [9] matches the partial depth of imaging sonar to optical images and applied MRF with the intensity of optic images and acoustic images to estimate dense depth maps. This can predict a real-scale depth map from sonar images and reconstruct dehazed images in strong haze optical images. However, the usage of imaging sonar is limited for underwater robots because of the substantially high cost of a sensor. In addition, the method requires a high computation time to estimate dense depth maps with very sparse data using MRF. From previous methods, we can summarize several objectives of dehazing for addressing robotic problems. First, the method should run in real time while learning data online. Second, the algorithm should work regardless of the number of channels. The last objective is the estimation of the normalized depths of images.

In this paper, we propose an online image dehazing method with partial and sparse depth data using iGP depth estimation. In comparison with [9], the proposed method requires only low-level fusion with any range sensors. Because a dense depth map is needed to reconstruct haze-free images in a strong turbid medium, we estimate a dense depth map using iGP regression and remove haze with color-corrected global (veiling) light. Also, we show that our method is applicable for both color and gray images.
For evaluation, we initially verify our method on a general aerial image with a true depth map. Then, we compare our algorithm with previous methods [3, 4, 5, 6, 8] on publicly open underwater color images containing various veiling lights and gray images. The proposed method is also evaluated via the computation time comparison.

II. ATMOSPHERIC SCATTERING MODEL

The image captured in the haze environment can be modeled as two components: the direct transmission of scene radiance from objects and the back-scattering effect by particles in a turbid medium. The mathematical model of a haze image, as introduced in [11], is written as

\[
I(u, v) = \underbrace{J(u, v)t(u, v)}_{\text{direct\-attenuation}} + \underbrace{A(1 - t(u, v))}_{\text{back\-scattering}}
\]

where \(I(u, v)\) is the haze image, \(J\) is the scene radiance (clear image), \(t(u, v)\) is the transmission and \((u, v)\) is the pixel point. The transmission \(t(u, v)\) is a function of scene depth and written as

\[
t(u, v) = \exp(-\beta d(u, v))
\]

where \(d(u, v)\) is the depth data of the pixel and \(\beta\) is the attenuation coefficient which presents thickness of haze. Physically, transmission \(t(x)\) refers to the color attenuation of scene radiance (original color of objects) due to the radiance traverses along the scattering medium.

In this paper, we slightly fix the haze model by using normalized depth \(\tilde{d}(u, v)\) instead of \(d(u, v)\) for accurate depth estimation and consistent scene radiance restoration.

III. INCREMENTAL GAUSSIAN PROCESS

A. Gaussian Process Regression

To predict the depth map of the haze image, we exploit the Gaussian Process (GP) regression method [12] using highly sparse depth data. The GP model is constructed with a training set \(D\) of \(n\) inputs (pixel points and values), \(D = \{(x_i, y_i) | i = 1, \cdots, n\}\), where \(x\) is the input vector, and \(y\) is an output (depth).

The inference for a test input \(x_*\) is the conditional Gaussian distribution given training inputs \(p(y_*|x_*, X, y) = \mathcal{N}(\mu_*, \Sigma_*)\), where \(y = \{y_{1:n}\}\) is a training output vector and \(X = \{x_1, \cdots, x_n\}\) is a \(D \times n\) design matrix with \(D\) dimensional input vector \(x\). The estimated Gaussian mean \(\mu_*\) and covariance \(\Sigma_*\) of the inference is written as follows

\[
\begin{align*}
\mu_* &= k(x_*, X)^T(K + \sigma^2 I)^{-1}y \\
\Sigma_* &= k(x_*, x_*) - k(x_*, X)^T(K + \sigma^2 I)^{-1}k(x_*, X)
\end{align*}
\]

where \(\sigma\) is the observation noise, \(k(\cdot, \cdot)\) is a kernel (covariance) function, and \(K = K(X, X)\) which is \(n \times n\) covariance matrix from calculated kernel function. As shown in (3), the prediction step includes the inverse operation \((K + \sigma^2 I)^{-1}\) which has computational complexity \(O(n^3)\). Because this complexity commonly causes large computation time, Ranganathan [13] proposes an online learning method to overcome this issue. We adopt an incremental approach in [13] that uses incremental update of the covariance matrix \(K\) with QR decomposition.

B. iGP Kernel Function

We find that the selection of the kernel function and hyperparameters affects GP prediction performance. We construct the kernel function using both stationary and non-stationary kernel functions. We use a SE kernel and NN kernel function for stationary and nonstationary kernel functions. The mathematical model of the kernel function is

\[
k(x, x') = \underbrace{\sigma^2_{SE} \exp\left(-\frac{\|x - x'\|^2}{2l^2_{SE}}\right)}_{\text{SE}} + \underbrace{\frac{2}{\pi} \sin^{-1}\left(\frac{2\tilde{x}^T\Sigma\tilde{x}'}{\sqrt{(1 + \tilde{x}^T\Sigma\tilde{x}')(1 + \tilde{x}'^T\Sigma^T\tilde{x})}}\right)}_{\text{NN}}
\]

where \(\sigma_{SE}\) is the variance of the SE kernel, \(l_{SE}\) is the length scale of SE, \(\Sigma = \text{diag}(\sigma_{SE}^2, \sigma^2)\) is the covariance of the NN kernel, and \(\tilde{x} = (1, x), x' = (1, x')\) are augmented input vectors of \(x\) and \(x'\).

The advantage of the mixture kernel functions is represented in Fig. 1. To evaluate the performance of kernels, SE, NN and mixture, we generated spatially disconnected synthetic data with a grid size of 400 × 400 as in Fig. 1(a). The ground truth
is composed of 16 Gaussian distributions with a monotonically increase mean and there is zero region in the interval $x_1 = [150, 200]$. To verify the interpolation and extrapolation abilities of the kernels, we randomly sample 50 training points from the set $S = \{(x_1, x_2) | x_1 \in [0, 300], x_2 \in [0, 400]\}$. As in Fig. 1, the effect of each kernel is significantly different. The SE kernel shows a better interpolation result than NN but fails to extrapolate (global tendency of mean increase). On the other hand, the NN kernel estimates global tendency of the training data but fails to capture detailed estimation result of interpolation. The mixture of kernel shows the best result, which includes the strengths of both the SE and NN kernels.

C. Covariance Matrix Update by QR Decomposition

To achieve online GP processing called iGP, it is necessary to incrementally update the covariance matrix for incoming training inputs (sparse depth priors). We adopt the covariance matrix updating method using QR decomposition and Givens rotation following [13]. The updating procedure is described in Fig. 2. As in phase 1, we first assume that we have decomposed matrices $Q_{K_{X+\sigma^2 I}}$ and $R_{K_{X+\sigma^2 I}}$ by applying QR decomposition to the matrix $K_{X+\sigma^2 I}$ from the initial training inputs $X$. Then covariance matrices $(K_{X+X'} = K(X', X))$ for newly added training inputs represented as a device matrix $X'$ are calculated from the kernel function. In phase 2 and 3, we apply Givens rotation matrices on the subset of initial inputs which are directly related to the new inputs, and the updated values are expressed as red blocks. The new $R_{updated}$ matrix is obtained directly from the procedure, and updates continue incrementally for next incoming inputs. After updating incoming inputs, the inverse term in (3) is easily calculated by simple matrix calculation as $(K_{updated} + \sigma^2 I)^{-1} = R_{K_{updated}}^*Q_{K_{updated}}$ where $^*$ refers to the back substitution operation.

D. iGP Regression Model with Sparse Depth Priors

iGP regression model is applied in order to estimate a depth map of a haze image. As we explained, iGP regression model is incrementally trained by incoming sparse depth points (priors). For training the model, the input vector of the model is constructed as $x = \{u, v, r, g, b\}$ where $(u, v)$ is a pixel position of a depth point and $(r, g, b)$ is a pixel value of a depth point, and vector elements are normalized as $I$. Also, the training output of the model, $y$ in (3), is $\hat{d}$, which is normalized by the maximum range of depth sensors. In the case of gray images, the input vector is modified to include intensity $i$ instead of $(r, g, b)$. The test inputs are all pixels in a haze image with the same format as training inputs $(x_a = \{u_a, v_a, r_a, g_a, b_a\})$. By using new notations, the regression model (3) is modified as

$$\hat{d}_a = K(X_a, X)^T(K + \sigma^2 I)^{-1}d$$

where $d = \{\hat{d}_1, \cdots, \hat{d}_n\}$ is training depth vector, $X_a$ is a device matrix of test inputs, $\hat{d}_a = \{\hat{d}_{a_1}, \cdots, \hat{d}_{a_m}\}$ test outputs (predicted depth points) and $m$ is the number of pixels of a haze image.

IV. DEHAZING

A. Haze Removal

Having predicted depth points, we can restore the haze-free image through estimating veiling light and transmission in (1). However, we use the additional veiling light term $\tilde{A}$ instead of direct calculation of $J$ from (1) and the equation is written as

$$J(u, v) = \frac{I(u, v) - A}{\max(l(I(u, v), t_0)) + \tilde{A}}$$

where $t_0$ is the minimum transmission (0.1). As we mentioned, $A$ represents underwater veiling light added by the haze effect. Because the color underwater veiling light $A$ is significantly biased by different color attenuation effects along the wavelengths of light, we propose color-balanced veiling light $\hat{A}$ to compensate color distortion.

Estimating $A$ and $\hat{A}$ is intuitive as follows. First, $A$ can be estimated from the depth map and an input image motivated from [5]. The dense depth map is already prepared; therefore, we pick pixels with 0.1% of the deepest points in the depth map and select the pixel color that has the maximum brightness. These pixel color values are used for $A$ the input image. $\hat{A}$ is color-balanced veiling light term achieved by $Lab$ color correction. Similar to coordinate (axis) shift of $a, b$ in [14], we calculate the reasonable origin of axis of $a, b$ in $Lab$ color space from uncasted color images gathered in Filcker dataset in [15]. By averaging origin position of axis, we set the origin axis of as $(a, b) = (6.35, 3.15)$ for $\hat{A}$. This procedure is described in (7) in $Lab$ color spaces,

$$\hat{A}_{Lab,i} = A_{Lab,i} - \bar{I}_{Lab,i} + \bar{I}_{Lab,i}$$

where subscription $Lab$ means $Lab$ space representation, $\hat{A}_{Lab,i}$ is the color-balanced veiling light, $A_{Lab,i}$ is a original value, $\bar{I}_{Lab,i}$ is an average of values in the test image and
Fig. 3. Dehazing under synthetic fog. (a) Original image of ICL-NUIM[16]. (b) Synthetic haze image by adding artificial fog. (c) True transmission (red: high, blue: low). (d) Randomly selected training points (red dots).

Fig. 4. Image dehazing and transmission map estimation under synthetic fog. Estimated transmission (top) and dehazing result (bottom). Absolute levels of transmission differ among the methods. The important fact is that relative transmission tendency of region in the test image. Transmission is getting high when color is red, and low when color is blue.

\( \tilde{I}_{Lab,i} \) is trained new origin from Flickr dataset [15]. The coordinate correction is applied in Lab space for the components \( i \in \{2,3\} \). Also we choose \( \beta \) between 0.5 and 1 from the estimated depth map. By averaging estimated depth values, we select the attenuation coefficient \( \beta \) to match in the interval of 0.5 to 1.

V. EXPERIMENTAL RESULTS

The performance of the proposed method is verified via testing on both terrestrial and underwater images. Our method is implemented in MATLAB for fair comparison with previous methods. Using (1), we first test the proposed method with images under synthetic fog for the comparison of transmission estimation. We then apply the method on real underwater color and gray images acquired from Google Photo. For depth priors, we randomly sample 10 depth points, and we assume that 1 point is added to the iGP model for 1 update.

A. Synthetic Fog

To generate synthetic fog we use a publicly available dataset, ICL-NUIM RGBD benchmark dataset [8]. The RGBD benchmark dataset offers a full depth map without the missing depth of all pixels so that we can synthetically corrupt original images with different fog level. We apply a synthetic haze on 640 × 480 haze-free images using a transmission map computed from a true depth map as in (1). The method is evaluated by comparing the result with the ground truth and four conventional approaches, [3], [6], [4] and [5].

Fig. 3 describes the depth estimation and dehazing for a sample image. While having a full depth map from the dataset, we only use randomly selected 10 training samples (red dots in Fig. 3(d)). The number of samples is chosen numerically by comparing dehazing results of synthetic fog images with those of previous methods. The of samples needed to get better results are differs for test images, but 10 samples result in enough performance of dehazing. Fig. 4 presents transmission estimation and dehazing performance compared to previous methods. By using 10 training inputs, we estimate a reliable transmission map and restore a haze-free image. Compared to other methods, the proposed approach shows the better result in terms of transmission map accuracy and color restoration, especially for very thick fog condition Fig. 3(b) compared to other methods (Fig. 4(b)) and Fig. 4(c)). The former method uses transmission and shadow models for dehazing and the estimate is limited under dense fog conditions. This result shows the initial possibility of the proposed method as online dehazing method.
B. Real Underwater Scene

We now evaluate the proposed method on real underwater images as described in Fig. 5. Four images in Fig. 5 are collected from Google to test the method in various environments (global veiling light color and objects). For examples, Fig. 5(e) and Fig. 5(g) have greenish veiling light, while Fig. 5(f) and Fig. 5(h) have bluish and yellowish veiling lights.

As mentioned, we assume sparse depth priors for dehazing. In the synthetic fog case, we randomly select 10 depth points from the true depth map. However, it is not available to get real depth data for underwater test images. For this reason, we manually build a rough depth map (white (far) and black (near) gradation) for each and we randomly sample 10 training points (red dots) as in the first row of Fig. 5. These depth maps contain only relative range information (near or far) without details.

Two figures, Fig. 6 and Fig. 7, show the dehazing results of test images in Fig. 5. We first describes dehazing results in Fig. 6 with estimated transmission map compared with previous methods [3], [6], [4] and [5]. Although the method use rough depth information which is not from true depth, the result of our method shows the reliable result as in Fig. 6. The first row of Fig. 6 presents predicted transmission map. The transmission map of our method follows 10 sampled points with edge details of objects and is estimated reasonably. From the results of previous methods, methods of Fattal [3] and Carlevaris-Bianco [6] show better transmission estimation results than He [4] and Zhu [5]. This is because of different color attenuation depends on the wavelength in underwater. The ray with longer wavelength (red color) is attenuated quickly than shorter (blue color). This phenomenon breaks the haze-relevant priors such as DCP and CAP.

Also, the proposed method result in contrast-enhanced and color-balanced images compared to previous methods. The

Dehazing results of test color images (Fig. 5(f), Fig. 5(g) and Fig. 5(h)) are described in Fig. 7. As a result, the proposed method shows the best performance of dehazing results only with sparse depth priors independent on background colors (bluish, greenish and yellowish). Because of color shifting effect of $Lab$ correction using Flickr dataset [15], near objects appears little reddish (first and second row) but contrasts are strengthened. The results of previous methods represent same trend as in Fig. 6. Fattal [3] and Carlevaris-Bianco [6] show relatively better dehazing result than He [4] and Zhu [5]. Fig. 8 represents performances on the underwater gray images. As we mentioned, the proposed method can be applied independent of the number of color channels. For evaluation, we compare the result with [8] and [4] which are applicable on gray images. The results of the proposed method (second column) show a contrast enhanced and smooth result. Tarel’s results (fourth...
column) shows an edge-enhanced result, but noise and halo effect occur. Contrary to this, He’s results (third column) is smoother than Tarel’s results, but contrast are degraded and images are getting dark.

We also verify the proposed method by comparing processing time. As we mentioned in introduction, the proposed method focusing on online dehazing performance for underwater robot navigation. Evaluation is done with the $1280 \times 960$ color image (Fig. 5(e)), $640 \times 480$ color images (Fig. 5(f), Fig. 5(g) and Fig. 5(h)) and $640 \times 480$ gray images (Fig. 8). Comparison result is represented in Table I. The computation times of $640 \times 480$ color and gray images are calculated as average. The computation times of our method of are more than two times faster than previous methods with better dehazing performance.

VI. CONCLUSION

This paper has focused on iGP based online haze removal with sparse partial depth cues for practical usage of underwater robots. Through incremental updates of GP model, we can efficiently deal with incoming depth priors and restore the haze-free image online. Having obtained a depth map, in addition, we estimate parameters $\hat{A}$ and $\hat{A}$ in a dehazing model to restore the color-balanced and dehazed image. For evaluation, we test out the method on synthetic fog and underwater images (both for color and gray images) and compare computation efficiency. We verify the proposed method with rough depth prior for underwater images, and we will evaluate the method on real underwater datasets for robot navigation and compare with previous methods.

ACKNOWLEDGMENT

This work is supported through a grant from the Korea Evaluation Institute of Industrial Technology (KEIT) (Award #N02150031) and NRF (Award #N01150984), and Ministry of Land Infrastructure and Transport’s U-city program.

REFERENCES


<table>
<thead>
<tr>
<th>Data</th>
<th>Proposed</th>
<th>Fattal</th>
<th>Carlevaris-Bianco</th>
<th>Tarel</th>
<th>He</th>
<th>Zhu</th>
</tr>
</thead>
<tbody>
<tr>
<td>color (640x480)</td>
<td>0.562</td>
<td>102.6</td>
<td>605.2</td>
<td>-</td>
<td>60.42</td>
<td>1.53</td>
</tr>
<tr>
<td>color (1280x960)</td>
<td>1.374</td>
<td>1721.58</td>
<td>11704.53</td>
<td>-</td>
<td>84.07</td>
<td>4.348</td>
</tr>
<tr>
<td>gray (640x480)</td>
<td>0.532</td>
<td>-</td>
<td>-</td>
<td>2.894</td>
<td>46.55</td>
<td>-</td>
</tr>
</tbody>
</table>