The Next Best Underwater View

Mark Sheinin and Yoav Y. Schechner
Viterbi Faculty of Electrical Engineering
Technion - Israel Inst. of Technology, Haifa 32000, Israel
markshe@campus.technion.ac.il , yoav@ee.technion.ac.il

Abstract

To image in high resolution large and occlusion-prone scenes, a camera must move above and around. Degradation of visibility due to geometric occlusions and distances is exacerbated by scattering, when the scene is in a participating medium. Moreover, underwater and in other media, artificial lighting is needed. Overall, data quality depends on the observed surface, medium and the time-varying poses of the camera and light source (C&L). This work proposes to optimize C&L poses as they move, so that the surface is scanned efficiently and the descattered recovery has the highest quality. The work generalizes the next best view *concept of robot vision to scattering media and cooperative movable lighting. It also extends descattering to platforms that move optimally. The optimization criterion is* information gain*, taken from information theory. We exploit the existence of a prior rough 3D model, since underwater such a model is routinely obtained using sonar. We demonstrate this principle in a scaled-down setup.*

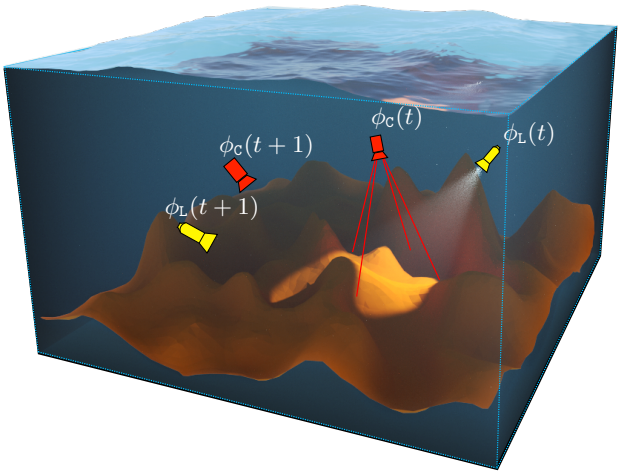

Figure 1. In a scattering medium at time t, the camera and light-source poses are represented by $\phi_{\mathrm{C}}(t)$, $\phi_{\mathrm{L}}(t)$ respectively. *The next best underwater view* task optimizes the poses $\phi_{\mathrm{C}}(t+1)$, $\phi_{\mathrm{L}}(t+1)$ for object reconstruction.

1. Introduction

Scattering media degrade images. Visibility enhancement often seeks single-image dehazing [13, 18] or relies on modulation of illumination properties, such as spatio-temporal structure [3, 8, 14, 19, 27, 34] and polarization [45, 47]. None of these methods exploit an important degree of freedom: the dynamic pose of the camera.

Pose dynamics is important, because most imaging platforms move *anyway*. Platform motion, however, needs to be efficient, covering the surface domain in the highest quality, in the shortest time. The camera needs to move, so that object regions that have not been well observed, will be efficiently recovered next. This is the *next best view* (NBV) concept, which has been extensively studied in the computer vision and robotics communities. There, viewpoint selection was driven by occlusions [30], geometric uncertainty in three dimensional (3D) scene reconstruction [9, 50] and active recognition [38]. Prior NBV designs, however, assumed no participating medium. A scattering medium may significantly disrupt visibility. This affects drones overflying wide hazy scenes, autonomous underwater vehicles scanning the sea floor to inspect infrastructure [10, 20] and fire-fighting rovers operating in smoke. Despite their motion and need to overcome scatter, existing systems conduct imaging paths [4] while ignoring scattering.

This work generalizes NBV to scattering media. We achieve 3D descattering in large areas and around occlusions, through sequential changes of pose. The obvious need to move the platform around large areas and occlusions is *exploited* for optimized dehazing, i.e, estimation of surface albedo. On the other hand, scattering by the medium influences the optimal changes of pose.

The challenge is exacerbated when lighting must be brought-in, in deep underwater operations, tissue and indoor smoky scenes. Scattering affects object irradiance and volumetric backscatter [16, 23] as a function of the *lighting pose*, not only the *camera pose* (Fig. 1). Usually both the

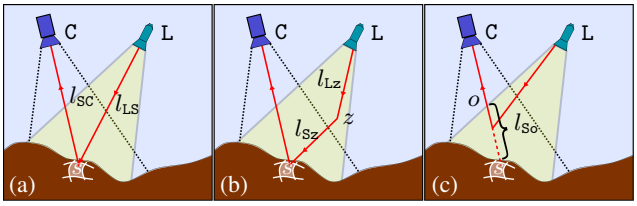

Figure 2. The sensed radiance of surface patch s is affected by three illumination components: (a) Direct illumination D_s, (b) Ambient illumination A_s, (c) Parasitic backscatter B_s.

camera and lighting (C&L) are mounted on the same rig. However, visibility can potentially be enhanced using separate platforms [24]. Therefore, the *next best underwater view* (NBUV), which is introduced in this work, optimizes the next joint poses of C&L.

The optimization criterion we use is *information gain*. We exploit a rough prior 3D model, since underwater such a model is routinely obtained using active sonar. We demonstrate the principle in scaled-down experiments.

2. Theoretical Background

2.1. Imaging in a Medium

Consider Figs. 1,2. At time t, the pose of light source L has a vector of location and orientation parameters, $\phi_{\mathrm{L}}(t)$. A source whose intensity is $\mathcal{C}_0$ irradiates a submerged surface patch s from distance l_{LS}. The irradiance [16] model at s is

$$\tilde{E}_s = D_s + A_s. \tag{1}$$

The component D_s is due to direct transmission from L to s, while A_s is ambient indirect surface illumination. The latter is mainly created by off-axis scattering of the illumination beam. The medium has extinction coefficient β. In a single scattering approximation [16, 31, 44],

$$D_s \propto \frac{\mathcal{C}_0 \exp[-\beta l_{\mathrm{LS}}]}{l_{\mathrm{LS}}^2}, \tag{2}$$

while A_s integrates all single scatter paths from L to s over the illuminated volume. Each path is of the form

$$\tilde{A}_s \propto \frac{\mathcal{C}_0 \exp[-\beta(l_{\mathrm{Lz}} + l_{\mathrm{Sz}})]}{(l_{\mathrm{Lz}} l_{\mathrm{Sz}})^2}, \tag{3}$$

where z is a point in the illuminated volume, and l_{Lz}, l_{Sz} are defined in Fig. 2. Monte-Carlo methods can render multiple scattering and complex shading effects [21, 25, 26].

At time t, the pose of C is represented by a vector of parameters $\phi_{\mathrm{C}}(t)$. The distance from s to camera C is l_{SC}. If the surface is Lambertian,[1] the *signal* measured by C is

[1] Surface reflectivity is described by the bidirectional reflection distribution function BRDF [2]. The validity of a Lambertian assumption increases underwater [40, 51].

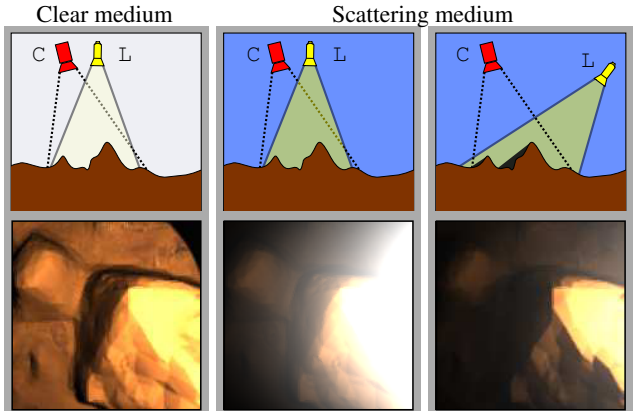

Figure 3. Using a small C&L distance either in a non-scattering [Left] or a scattering medium [Middle]. Backscatter significantly reduces image quality. [Right]: A large C&L separation reduces backscatter but exacerbates shadows and signal extinction.

$\rho_s E_s$, where ρ_s is the albedo at s and

$$E_s = \tilde{E}_s \exp(-\beta l_{\mathrm{SC}}). \tag{4}$$

The line of sight from C to patch s includes backscatter B_s, which increases [23, 41, 47] with l_{So} (see Fig. 2c). The measured radiance [7] is

$$I_s = \rho_s E_s + B_s + n_I, \tag{5}$$

where n_I is comprised of [37] photon[2] noise and read noise. The variance of photon noise is $\sigma^2_{\mathrm{PN}} = I_s$. Readout noise is assumed to be signal-independent, with variance σ^2_{RN}. The probability density function (PDF) of n_I is approximately Gaussian with variance:

$$\sigma^2_{I_s} = \sigma^2_{\mathrm{PN}} + \sigma^2_{\mathrm{RN}} = I_s + \sigma^2_{\mathrm{RN}}. \tag{6}$$

The signal-to-noise ratio (SNR) at s is therefore

$$\mathrm{SNR}_s \approx \frac{I_s}{\sqrt{I_s + \sigma^2_{\mathrm{RN}}}} \approx \frac{\rho_s E_s}{\sqrt{\rho_s E_s + B_s + \sigma^2_{\mathrm{RN}}}}. \tag{7}$$

Backscatter is negligible, $B_s \ll \sigma^2_{\mathrm{RN}}$, in a clear medium. Then from Eq. (7), under sufficient lighting $\mathrm{SNR}_s \sim \sqrt{\rho_s E_s}$. For the best SNR, E_s is maximized. This is achieved by avoiding shadows [39], i.e., placing L very close to C (Fig. 3).

Underwater, placing L very close to C results in significant backscatter B_s, which reduces SNR_s in (7). To reduce backscatter, L is usually separated from C. Such a separation may create shadows. In a shadow, $E_s \ll \sigma^2_{\mathrm{RN}}$, compounding light extinction by the medium (Eqs. 2-3). Thus optimal setting of L is non-trivial.

[2] In this paper, all the radiometric terms (I_s, E_s, n_I, etc..) are in photoelectron units [e].

2.2. Next Best View

The NBV task is generally formulated as follows. Let $\mathcal{O}$ represent a property of the object, e.g., the spatially varying albedo or topography. A computer vision system estimates this representation, $\hat{\mathcal{O}}$, using sequential measurements. By t, the camera has already accumulated image data $I(t')$, $\forall t' \leq t$. All preceding data is processed to yield $\hat{\mathcal{O}}(t)$. Let $\Phi_{\mathtt{C}}$ be the set of all possible camera poses. A *next view* is planned for time $t+1$, where the camera may be posed at $\phi_{\mathtt{C}}(t+1) \in \Phi_{\mathtt{C}}$, yielding new data. The new data helps getting an improved estimate $\hat{\mathcal{O}}(t+1)$. The NBV question is: out of all possible views in $\Phi_{\mathtt{C}}$, what is the *best* $\phi_{\mathtt{C}}(t+1)$, such that $\hat{\mathcal{O}}(t+1)$ has the best quality? Formulating this task mathematically depends on a quality criterion, prior knowledge about $\mathcal{O}$, and the type of camera; e.g., passive or active 3D scanner. Different studies have looked at different aspects of the NBV task [5, 36, 49]. Nevertheless, they were all designed for imaging in clear media.

2.3. Information Gain

Consider a random variable a. Let $f(a)$ be its PDF. The differential entropy [1] of a is then

$$H(a) = -\int f(a)\ln[f(a)]da. \tag{8}$$

At time t, the variable a has entropy $H_t(a)$. Then, at time $t+1$, new data decreases the uncertainty of a, consequently the PDF of a is narrowed and its differential entropy decreases, $H_{t+1}(a) < H_t(a)$. The *information gain* [33] due to the new data is then

$$\mathcal{I}_{t+1}(a) = H_t(a) - H_{t+1}(a). \tag{9}$$

Suppose a is normally distributed, with variances $\sigma_a^2(t)$ and $\sigma_a^2(t+1)$ at t and $t+1$, respectively. Then Eqs. (8,9) yield

$$H_t(a) = (1/2)\ln[2\pi e\sigma_a^2(t)], \tag{10}$$

$$\mathcal{I}_{t+1}(a) = (1/2)\ln[\sigma_a^2(t)\sigma_a^{-2}(t+1)]. \tag{11}$$

3. Least Noisy Descattered Reflectivity

Underwater bathymetry (depth mapping) is routinely done using sonar [4, 6, 11], which penetrates water to great distances. Hence, in relevant applications, the surface topography is roughly available [4, 22] before optical inspections. The C&L pose parameters are concatenated into a vector $\mathbf{v}(t) = [\phi_{\mathtt{C}}(t), \phi_{\mathtt{L}}(t)]$. This vector is approximately known during operation, using established localization methods [11, 12, 29, 32, 35]. Moreover, the water scattering and extinction characteristics are global parameters, that can be measured in-situ. Consequently, B_s and E_s can be pre-assessed for each $\phi_{\mathtt{C}} \in \Phi_{\mathtt{C}}$, $\phi_{\mathtt{L}} \in \Phi_{\mathtt{L}}$ and surface patch index s.

At close distance, optical imaging and descattering seek the spatial distribution of the surface albedo $\mathcal{O} = \bigcup_s \rho_s$, to notice sediments, defects in submerged pipes, parasitic colonies in various environments etc.[3] Beyond removal of bias by backscatter and attenuation, descattered results need to have low noise variance, so that fine details [46] can be detectable. This is our goal.

Using Eq. (5), descattering based on an image at t is

$$\hat{\rho}_s(t) = [I_s(t) - B_s(t)]/E_s(t). \tag{12}$$

From (12), the noise variance of $\hat{\rho}_s(t)$ is

$$\sigma_s^2(t) = \sigma_{I_s}^2/E_s^2(t)\ . \tag{13}$$

Note that $\sigma_s^2(t)$ is unknown, since Eqs. (5,6) depend on the unknown ρ_s. Nevertheless, it is possible to define an *operating point* value for ρ_s, by a typical value denoted $\bar{\rho}$. The reason is that, per application, the typical albedo encountered is familiar: typical soil in the known region, anti-corrosive paints in a known familiar bridge support, etc. The value of $\bar{\rho}$ is rough, but provides a guideline. Consequently

$$\sigma_{I_s}^2 \approx \bar{\sigma}_{I_s}^2 \equiv \bar{\rho}E_s(t) + B_s(t) + \sigma_{\text{RN}}^2, \tag{14}$$

$$\sigma_s^2(t) \approx \frac{\bar{\rho}_s E_s(t) + B_s(t) + \sigma_{\text{RN}}^2}{E_s^2(t)}. \tag{15}$$

Multi-Frame Most-Likely Descattering

As described in Sec. 2.2, by discrete time t, the system has already accumulated data $\{I_s(t')\}_{t'=0}^{t}$. The measurements have independent noise. Hence, the joint likelihood $L_s(t) \equiv L[\{I_s(t')\}_{t'=0}^{t}]$ of the data is equivalent to the product of probability densities $\forall t'$. Consequently, the log-likelihood is

$$\tilde{L}_s(t) = \ln L_s(t) \simeq \sum_{t'=0}^{t} \frac{[I_s(t') - B_s(t') - \rho_s E_s(t')]^2}{\bar{\sigma}_{I_s}^2(t')}. \tag{16}$$

Differentiating Eq. (16) with respect to ρ_s, the maximum likelihood (ML) estimator of the descattered ρ_s, using all accumulated data is

$$\hat{\rho}_s^{\text{ML}}(t) = \frac{\sum_{t'=0}^{t} \hat{\rho}_s(t')[\sigma_s(t')]^{-2}}{\sum_{t'=0}^{t}[\sigma_s(t')]^{-2}}, \tag{17}$$

where $\hat{\rho}_s(t'), \sigma_s(t')$ are derived in Eqs. (12,15). The variance of this estimator is

$$[\sigma_s^{\text{ML}}(t)]^2 = \left\{\sum_{t'=0}^{t}[\sigma_s(t')]^{-2}\right\}^{-1}. \tag{18}$$

Pre-calculate $\forall s, \mathbf{v}(t)$ a quality measure of s

$$q_s(t) \equiv 1/\sigma_s^2(t). \tag{19}$$

[3] Vision can further enhance the topography estimation [4].

From Eqs. (18,19), the quality of the ML descattered reflectivity $\hat{\rho}_s^{\mathrm{ML}}(t)$ is

$$Q_s^{\mathrm{ML}}(t) \equiv [\sigma_s^{\mathrm{ML}}(t)]^{-2} = \underbrace{\left[\sum_{t'=0}^{t-1} q_s(t')\right]}_{Q_s^{\mathrm{ML}}(t-1)} + q_s(t) \tag{20}$$

Eq. (20) shows how new datum updates $\sigma_s^{\mathrm{ML}}(t)$.

4. Next Best Underwater View

After time t, the next view $\mathbf{v}(t+1)$ yields information gain $\mathcal{I}_{(t+1)}(\mathcal{O})$. Let $\boldsymbol{\mathcal{V}}$ be the set of all possible (or permissible) camera-lighting poses for time $t+1$. The next underwater view and lighting poses are selected from $\boldsymbol{\mathcal{V}}$, to maximize the information gain measure $\mathcal{I}_{t+1}(\mathcal{O})$,

$$\hat{\mathbf{v}}(t+1) = \arg\max_{\mathbf{v}\in\boldsymbol{\mathcal{V}}} \mathcal{I}_{t+1}(\mathcal{O}). \tag{21}$$

We now derive $\mathcal{I}_{t+1}(\mathcal{O})$ in our case. Information is an additive quantity for independent measurements. Hence, information gained by enhanced estimation of ρ_s in N_s independent surface patches is

$$\mathcal{I}_{t+1}(\mathcal{O}) = \sum_{s=1}^{N_s} \mathcal{I}_{t+1}(\hat{\rho}_s^{\mathrm{ML}}). \tag{22}$$

From Eq. (11),

$$\mathcal{I}_{t+1}(\hat{\rho}_s^{\mathrm{ML}}) = \frac{1}{2}\ln\left\{[\sigma_s^{\mathrm{ML}}(t)]^2[\sigma_s^{\mathrm{ML}}(t+1)]^{-2}\right\}. \tag{23}$$

From Eqs. (20,23),

$$\mathcal{I}_{t+1}(\hat{\rho}_s^{\mathrm{ML}}) = \frac{1}{2}\ln\left[\frac{\sum_{t'=0}^{t+1} q_s(t')}{\sum_{t'=0}^{t} q_s(t')}\right] = \ln\left[1+\frac{q_s(t+1)}{Q_s^{\mathrm{ML}}(t)}\right]^{\frac{1}{2}} \tag{24}$$

Suppose prior to $t+1$, patch s has not been observed. Nevertheless, $0 \le \rho_s \le 1$. For a uniformly distributed random sample in this range, the variance is $\sigma_{\max}^2 = 1/12$. Therefore, if s is unobserved until $t+1$,

$$\mathcal{I}_{t+1}(\hat{\rho}_s^{\mathrm{ML}}) = \frac{1}{2}\ln\left[1+\frac{q_s(t+1)}{1/\sigma_{\max}^2}\right]. \tag{25}$$

5. Path Planning

Our formalism has focused on optimization of the next best view, underwater. What about the next best *sequence* of views? Indeed the formalism can be extended to path planning, beyond a single next view. The information gain from t to $t+1$ is given by Eqs. (22,24,25). Similarly, the information gain of patch s due to a path from t_1 to t_2 is

$$\mathcal{I}_{t_1\to t_2}(\hat{\rho}_s^{\mathrm{ML}}) = \frac{1}{2}\ln\left[\frac{\sum_{t'=0}^{t_2} q_s(t')}{\sum_{t'=0}^{t_1} q_s(t')}\right] = \frac{1}{2}\ln\left[\frac{Q_s^{\mathrm{ML}}(t_2)}{Q_s^{\mathrm{ML}}(t_1)}\right] \tag{26}$$

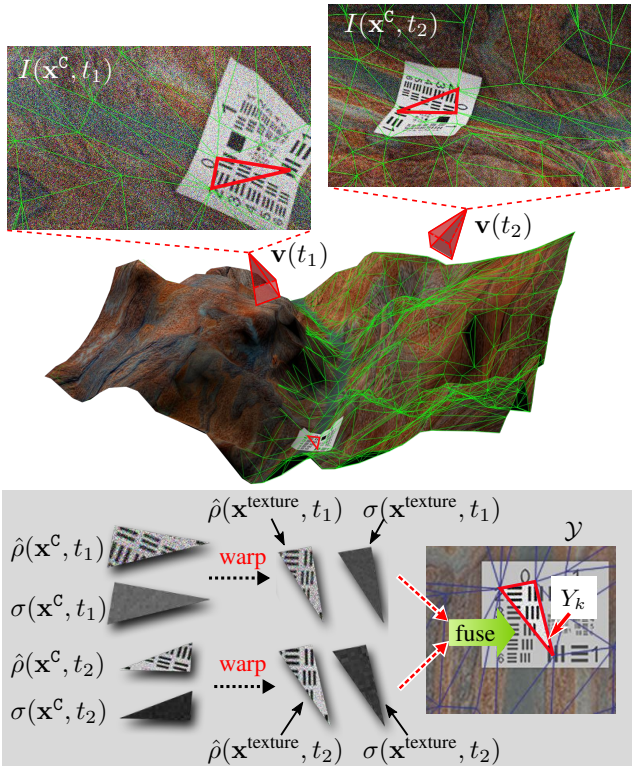

Figure 4. A surface mesh (green) is imaged from poses $\mathbf{v}(t_1)$ and $\mathbf{v}(t_2)$. Face T_k is a red-marked. Imaged faces are warped and scaled to match Y_k, amplifying $\sigma(\mathbf{x}^{\mathrm{texture}}, t)$ according to $|\mathcal{U}_s(t)|$.

Thus

$$\mathcal{I}_{t_1\to t_2}(\mathcal{O}) = \sum_{s=1}^{N_s} \frac{1}{2}\ln\left[\frac{Q_s^{\mathrm{ML}}(t_2)}{Q_s^{\mathrm{ML}}(t_1)}\right]. \tag{27}$$

A path of C&L is $\mathcal{L} \equiv [\mathbf{v}(t_1), \mathbf{v}(t_1+1)\ldots\mathbf{v}(t_2)]$. Then, in terms of information gain, an optimal path satisfies

$$\mathcal{L}_{\mathrm{best}} = \underset{\mathcal{L}}{\mathrm{argmax}}\,[\mathcal{I}_{t_1\to t_2}(\mathcal{O})]. \tag{28}$$

We implement Eq. (28) by perturbing an initial path $\mathcal{L}_0$.

6. Discrete Domain Expressions

In practice, both data and models are often not expressed in surface patches. Rather, data is given in *image pixels*, while a surface model is given by a *mesh* having a *texture map* [4]. This section describes how these representations affect the expressions.

Let the surface be modeled by a triangulated mesh[4] $\mathcal{M}$, comprising a set of mesh faces $\{T_k\}_{k=1}^{N_{\mathrm{m}}}$. Surface albedo is represented by a *texture map* domain $\mathcal{Y} = \{Y_k\}_{k=1}^{N_{\mathrm{m}}}$. Face T_k corresponds to a triangle Y_k in the map domain (Fig. 4d). A texture-map pixel $\mathbf{x}^{\mathrm{texture}} \in \mathcal{Y}$ has one-to-one correspondence with a specific surface patch: $s(\mathbf{x}^{\mathrm{texture}})$.

[4] Any polygonal representation of a mesh can be used here.

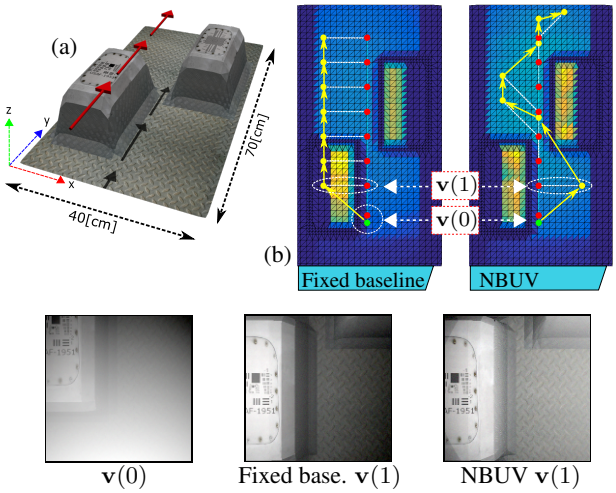

Figure 5. Simulation. (a) The scanned surface. The camera's scanning trajectory is along the red arrows. (b) Trajectories of fixed-baseline and NBUV. Face color represent $q_k(t=8)$. Note how NBUV avoids casting shadows from the bumps to the floor.

In the image-plane of camera $\mathtt{C}$, a spatial location is denoted $\mathbf{x}^{\mathtt{C}}$, in pixel units. At time t, the camera has a simulated projection operator $\mathcal{P}_t^{\mathtt{C}}$. Then, patch s projects to a spatial set $\mathbf{x}^{\mathtt{C}} \in \mathcal{U}_s(t)$:

$$\mathbf{x}^{\text{texture}} \leftrightarrow s \leftrightarrow \mathcal{U}_s(t). \tag{29}$$

The area of $\mathcal{U}_s(t)$ is $|\mathcal{U}_s(t)|$, in pixel units. Using this model, computer graphics renders components of Eqs. (12,15), $E(\mathbf{x}^{\mathtt{C}},t) = \mathcal{P}_t^{\mathtt{C}}(E_s)$, $B(\mathbf{x}^{\mathtt{C}},t) = \mathcal{P}_t^{\mathtt{C}}(B_s)$.

The corresponding image data is $I(\mathbf{x}^{\mathtt{C}},t)$. Thus, Eq. (12) is expressed as

$$\hat{\rho}(\mathbf{x}^{\mathtt{C}},t) = [I(\mathbf{x}^{\mathtt{C}},t) - B(\mathbf{x}^{\mathtt{C}},t)]/E(\mathbf{x}^{\mathtt{C}},t) \tag{30}$$

in the camera coordinates. This value is transfered to the *texture map* by a warping [15] operator (see Fig. 4)

$$\hat{\rho}(\mathbf{x}^{\text{texture}},t) = \text{WARP}\{\hat{\rho}(\mathbf{x}^{\mathtt{C}},t)\}_{\mathbf{x}^{\mathtt{C}}\in\mathcal{U}_s(t)} \ . \tag{31}$$

Similarly, Eq. (15) is expressed by

$$\sigma^2(\mathbf{x}^{\mathtt{C}},t) \approx \frac{\bar{\rho}(\mathbf{x}^{\mathtt{C}},t)E(\mathbf{x}^{\mathtt{C}},t) + B(\mathbf{x}^{\mathtt{C}},t) + \sigma_{\text{RN}}^2}{E^2(\mathbf{x}^{\mathtt{C}},t)} \tag{32}$$

in the camera coordinates. Note that when $|\mathcal{U}_s(t)| > 1$, warping by Eq. (31) implicitly involves spatial averaging. Thus, warping decreases the variance of the value in the texture map, in proportion to $|\mathcal{U}_s(t)|$. Hence, the result of Eq. (32) is generally transfered to the *texture map* by

$$\sigma^2(\mathbf{x}^{\text{texture}},t) = |\mathcal{U}_s(t)|^{-1}\text{WARP}\{\sigma^2(\mathbf{x}^{\mathtt{C}},t)\}_{\mathbf{x}^{\mathtt{C}}\in\mathcal{U}_s(t)} \ . \tag{33}$$

If $|\mathcal{U}_s(t)| < 1 \ \ \forall t$, then patch s is never observed at sufficient resolution. Per patch, the image sequence should strive to provide raw measurements $I(\mathbf{x}^{\mathtt{C}},t)$ at sufficient resolution, in which $|\mathcal{U}_s(t)| \geq 1$. Over a flat terrain, this requirement is easily met by keeping $\mathtt{C}$ under a specific altitude. In a complex terrain, the trajectory altitude and projected patch resolution vary. To keep our optimization unconstrained, we took the following step. When $|\mathcal{U}_s(t)| < 1$, the patch's variance is penalized as

$$\sigma^2(\mathbf{x}^{\text{texture}},t) = \exp(\eta\{|\mathcal{U}_s(t)|^{-1} - 1\}) \\ \text{WARP}\{\sigma^2(\mathbf{x}^{\mathtt{C}},t)\}_{\mathbf{x}^{\mathtt{C}}\in\mathcal{U}_s(t)} \tag{34}$$

where η is a constant parameter, which we set to 10. We found that this penalty keeps $\mathtt{C}$ from distancing from the surface, and provides good results.

Exploiting Eq. (29) to replace s by $\mathbf{x}^{\text{texture}}$, descattering is done by Eqs. (17,31,33,34). Similarly, Eq. (29) affects the representation of the fields $q_s(t)$, $Q_s^{\text{ML}}(t)$ and $\mathcal{I}$ in Eqs. (19,20,24,25).

To save memory, our implementation made a minor approximation. For a fine $\mathcal{M}$, the illumination and thus noise variance are rather uniform within each T_k. For each face k the average noise standard deviation is

$$\sigma_k(t) = \frac{1}{|Y_k|}\sum_{\mathbf{x}^{\text{texture}}\in Y_k} \sigma(\mathbf{x}^{\text{texture}},t) \ , \tag{35}$$

where $|Y_k|$ is the number of texture-map pixels in Y_k. Eq. (35) yields a per-face quality measures $q_k(t) = 1/\sigma_k^2(t)$ and $Q_k^{\text{ML}}(t) = \sum_{t'=0}^{t} q_k(t')$, in analogy to the terms $q_s(t)$ and $Q_s^{\text{ML}}(t)$. In this approximation, the information gain (24) for T_k is

$$\mathcal{I}_{t+1}(T_k) \approx \frac{|Y_k|}{2}\ln\left[1 + \frac{q_k(t+1)}{Q_k^{\text{ML}}(t)}\right], \tag{36}$$

The information gain is therefore:

$$\mathcal{I}_{t+1}(\mathcal{O}) \approx \sum_{k=1}^{N_m}\mathcal{I}_{t+1}(T_k). \tag{37}$$

7. Simulations

We set $\sigma_{\text{RN}} = 13.1$[e] and a full well of 24,000 [e] in a perspective camera $\mathtt{C}$, based on Canon 60D specs, while $\mathtt{L}$ is a spotlight with no lateral falloff. Fig. 5 illustrates a simple case study. The medium's parameters are $\beta = 5[1/\text{m}]$, while the Henyey-Greenstein phase anisotropy parameter is $g = 0.6$ [16, 17]. The scattering model of [16] renders the images (5).

Camera positions are manually set in a straight path 40cm above a surface. Eight views $\{\phi_{\mathtt{C}}(t)\}_{t=1}^{8}$ are spaced uniformly along the path. $\mathtt{C}\&\mathtt{L}$ start from $\mathbf{v}(0)$. The initial $\mathtt{LC}$ baseline is 2cm. Underwater this baseline results in significant backscatter. Hence, the baseline increases to

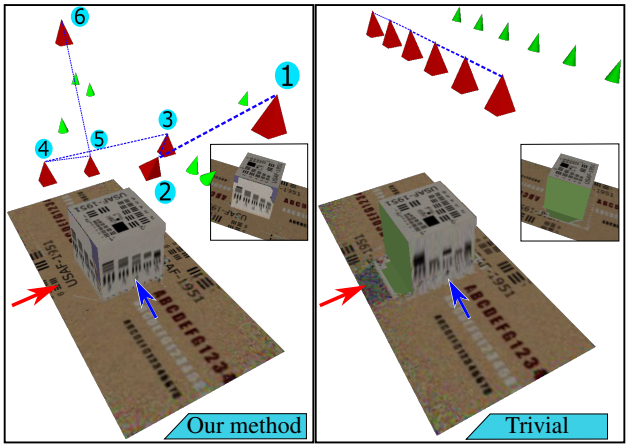

Figure 6. Path planning. Red cones - C. Green cones - L. In the trivial path, both left and right faces are occluded. Additionally, the box casts a shadow to the left of the plane (Red arrow). The NBUV path overcomes occlusion and shadowing, improving the estimation uncertainty by 30%.

12cm. A traditional path separates L from C by a fixed baseline whose length and orientation are fixed, $\vec{\mathtt{LC}} = 12\hat{x}$. L points to the center of C's field of view on the surface. To the best of our knowledge, prior descattering methods are oblivious to SNR variability in image sequences. To simulate this oblivion, we set $\sigma_s(t) = 1\ \forall t$ in Eq. (17) during the analysis of the traditional path.

Keeping the same $\{\phi_{\mathtt{C}}(t)\}_{t=1}^{8}$, NBUV optimized $\phi_{\mathtt{L}}(t+1), \forall t$. The state of L was selected out of a set of 32 possible locations at different distances around $\phi_{\mathtt{C}}(t+1)$. Moreover, in each location there are 9 orientations of $\phi_{\mathtt{L}}$, facing nadir, $\pm 10^\circ$ or $\pm 20^\circ$ off nadir, per lateral-coordinate. Hence, NBUV relies on $|\boldsymbol{\mathcal{V}}(t)| = 288$ options per t. We used exhaustive search.

Fig. 5b shows the two C&L trajectories. Clearly the illumination should face opposite the bumps, when the camera passes above them. This is evident in $\mathbf{v}(1)$, where NBUV yields a better lit image than an image of a traditional path.

Fig. 6 illustrates path planning by NBUV. A cube having 28cm edges is placed on a flat surface in a scattering medium ($\beta = 2.5[1/\mathrm{m}],\ g = 0.6$). An initial trivial scanning path for C is set 84cm above the surface. The initial path consists of 6 uniformly distributed views. To avoid backscatter, the baseline is $\vec{\mathtt{LC}} = 34\hat{x}$ cm. In the initial trivial path, the left and right faces (see Fig. 6 red arrow) are occluded as C passes over the cube. In addition, a shadow from the cube heavily degrades the left side of the surface.

Denote the scanning path by $\mathcal{L} = \{\mathbf{v}(1), \mathbf{v}(2)..\mathbf{v}(6)\}$. There are 60 degrees of freedom to L.[5] Optimization of L was initialized by the trivial path, followed by 20 iterations of Matlab's direct search function [28]. The resulting path moves C&L to cover the occluded regions (Fig. 6 views 3-4). Note that the front and back faces of the cube (blue arrow in Fig. 6) are better resolved. This is thanks to views 2 and 5. NBUV reduced the total estimation uncertainty by 30%, relative to the trivial path.

[5]Rotation about axis Z is degenerate, thus excluded for both L and C.

Scanned Topography

We built a model, described in Sec. 8, having an arbitrary non-trivial topography. Emulating a sonar scan, the surface model was scanned using a Kinect 2 time-of-flight camera in a clear medium, producing a 3D mesh. We tested what happens when the scattering medium is ignored. This is achieved by setting $\beta \to 0$ in the optimization. The resulting $\{\phi_{\mathtt{L}}(t)\}_{t=1}^{10}$ seeks to avoid shadows and maximize E_s over the surface. This mostly leads to a very small $|\vec{\mathtt{LC}}|$ and thus poor visibility (NBV in Fig. 7). NBUV reduces backscatter and shadowing (NBUV in Fig. 7).

Next, we tested NBUV robustness to deviations from our model's assumptions. We simulated recovery of a non-Lambertian surface ignoring the fact that specularities exist.[6] As seen in Fig. 7, relative to a fixed-baseline path, NBUV proved superior in practice. We believe that this is because backscatter and shadows are worse factors than non-Lambertian reflectance.

The 3D geometry prior can be coarse. This was tested by feeding the algorithm coarsened versions of the true terrain (Fig. 7). Optimized views were largely insensitive to the coarsened topography prior. We also induced errors in the view parameters, $\mathbf{v}(t)$ leading to small misalignments during texture mapping (Sec. 6). Misalignment is resolved by standard computer vision. For details see [43].

8. Experiments

In Exp1, the setup (Fig. 8) was submerged in water polluted by some milk, for scattering conditions consistent with our image formation model [31]. From Fig. 8c, the surface is somewhat shiny. A machine vision camera was submerged in a watertight housing. The intrinsic parameters of C and the illumination cone angle were calibrated underwater, to account for water refraction. A robotic 2D plotter moved C&L to $\{\mathbf{v}(t)\}_{t=0}^{10}$.

In Exp2, the setup (Fig. 9) was in fog created by a ATMOS FOG AF-1200 machine. The illumination was made by a LED (Mouser Electronics 'Warm White' 3000K). The extinction and phase-function parameters β,g were estimated in-situ:[7]

$$g, \beta = \underset{g,\beta}{\mathrm{argmin}} \|E(\mathbf{x}^{\mathtt{C}}, t) + B(\mathbf{x}^{\mathtt{C}}, t) - I(\mathbf{x}^{\mathtt{C}}, t)\|^2, \quad (38)$$

[6]Measurement were simulated using a Phong-model with parameters $K_\alpha = 0, K_d = 1, K_s = 0.5, \alpha = 20$.

[7]Underwater $\beta = 12[1/m],\ g = 0.6$. In the fog, we manually assessed $\beta \approx 2.5$ and $g = 0.6$.

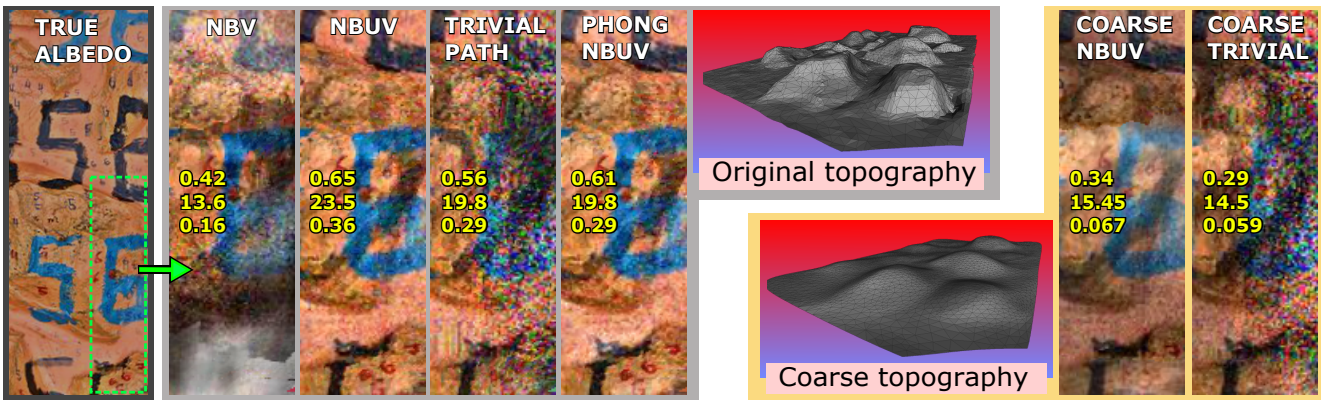

Figure 7. Simulations including coarsened geometry and Phong reflectance. Comparing to ground truth, quality measures are overlaid in yellow, top to bottom: SSIM [48], PSNR, VIF [42].

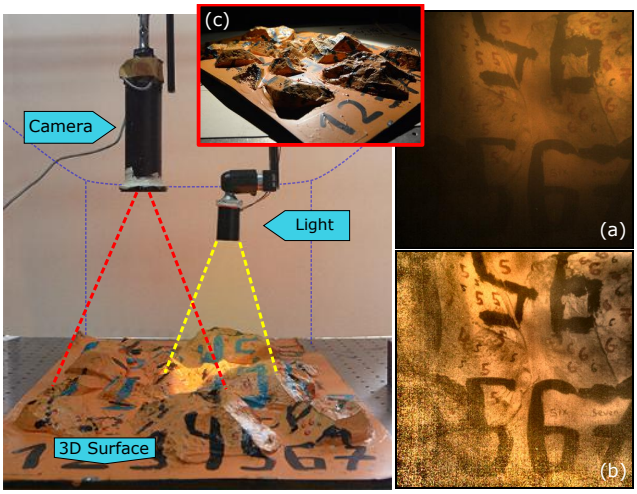

Figure 8. Experimental setup. A camera in a watertight housing images a LED-illuminated surface, submerged in a water tank (not shown). (a) Empirical image $I(\mathbf{x}^{\mathtt{C}}, t)$. (b) Conditioned recovery of $\hat{\rho}(\mathbf{x}^{\mathtt{C}}, t)$, using Eqs. (12,39). (c) Specularities.

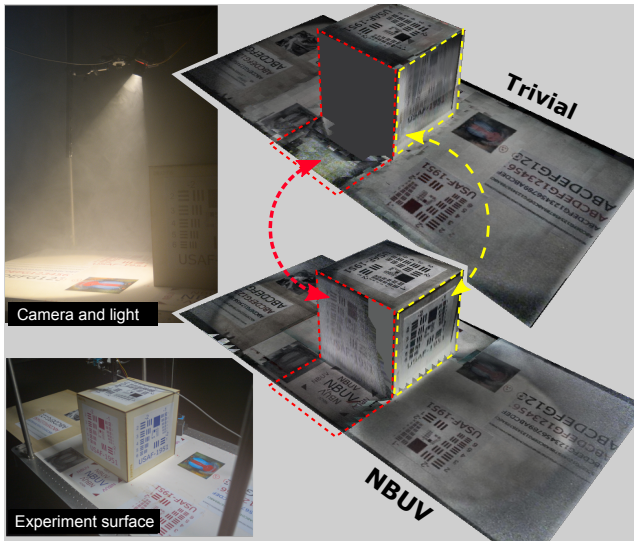

Figure 9. Path planning experiment. Red and yellow-marked regions show areas that significantly benefited from NBUV.

when C&L were placed in a known state above a flat white sheet ($\rho_s \approxeq 1$).

8.1. Numerical Conditioning

Eq. (12) may become unstable as $E_s \to 0$, due to shadows or multiple scattering. Therefore, $E(\mathbf{x}^{\mathtt{C}}, t)$ is stabilized by

$$\hat{E}(\mathbf{x}^{\mathtt{C}}, t) = \\ = \left(\left[E(\mathbf{x}^{\mathtt{C}}, t) * h_E\right](1-w) + \left[I(\mathbf{x}^{\mathtt{C}}, t) * h_I\right] w\right) * h_T. \tag{39}$$

Here h_E, h_I and h_T are Gaussian kernels, and w is an alpha mask. We set $w(\mathbf{x}) = 1$ whenever $\hat{\rho}(\mathbf{x}^{\mathtt{C}}, t) > 1$, which is an indicator of unstable estimation. An example of conditioned estimation is seen in Fig. 8b.

8.2. Results

In Expl, $\{\phi_{\mathtt{C}}(t)\}_{t=0}^{10}$ are manually set in uniformly-spaced locations starting from $\mathbf{v}(0)$ shown in Fig. 10a. Due to mechanical limitations, we allow C&L to move only horizontally, and oriented directly down. The elevation of C&L is 20cm above the lowest point of the surface.

In the fixed-baseline configuration, $\vec{\mathtt{LC}}$ is fixed so as to avoid overwhelming backscatter. By NBUV, per t, L can be placed in $|\boldsymbol{\mathcal{V}}(t)| = 40$ states around $\phi_{\mathtt{C}}(t)\ \forall t \in [1..10]$. The recovered albedo images are mapped to the 3D mesh of the surface (Fig. 10b-c). NBUV provides a better overall surface estimation, relative to data obtained by a fixed-baseline path. In our particular surface, shadows on the right and center of the surface are filled-in by NBUV. Fig. 10A-C show areas where the expected estimation noise is significantly lowered by NBUV. The total $\mathcal{I}(\mathcal{O})$ is lower, though

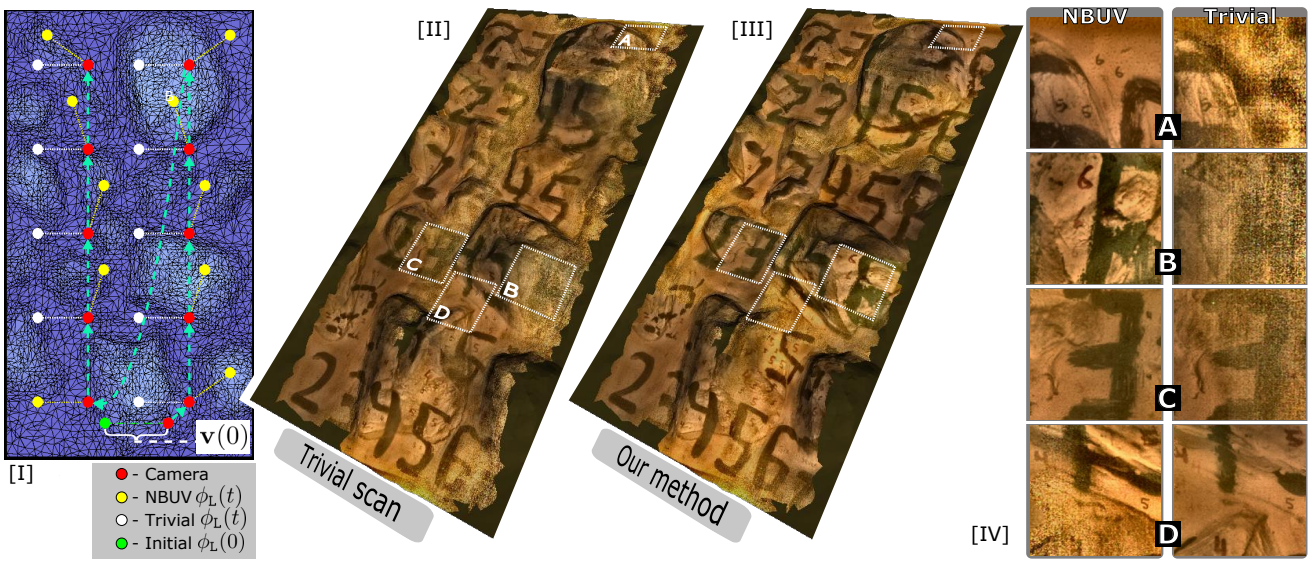

Figure 10. Experiment results. [I] Scanning path and light locations. [II-III]: Fixed-baseline and NBUV results. [IV]: Close up comparison. A-C show improved estimation quality while the area in D is better estimated using a fixed baseline.

some surface patches do not benefit from NBUV (Fig. 10D).

`Exp2` reproduces the setup and paths described in the cube simulation (Fig. 6). As in the simulation, the path planned by NBUV images all sides of the box (Fig. 9). The actual medium parameters deviate from the ones assumed. This discrepancy causes a bias in $\hat{\rho}(\mathbf{x}^{\mathrm{C}}, t)$, revealed in brightness variations in Fig. 6. As in `Exp1`, not all patches benefit from NBUV.

9. Discussion

The paper defines NBUV and path planning accounting for scattering effects. NBUV optimizes viewpoints so the descattered albedo is least noisy, allowing resolution of fine details. It generalizes dehazing to scanning multi-view platforms. We believe this approach can make drone imaging flights and underwater robotic imaging significantly more efficient when operating in poor visibility. Further work can use more comprehensive scattering models, image statistics priors and path-length penalties. Moreover, the principle we proposed can benefit from optimization algorithms that are more efficient, as the number of degrees of freedom increases. The principle can possibly be generalized to multiple cameras cooperatively scanning the scene.

Acknowledgments: We are grateful to Aviad Levis, Vadim Holodovsky, Amit Aides and Tali Treibitz for useful discussions. We thank Ina Talmon, Johanan Erez and Dani Yagodin for technical support. Y. Y. Schechner is a Landau Fellow - supported by the Taub Foundation. His research is supported by the Israeli Ministry of Science, Technology & Space (Grant 51349) and conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

Algorithm 1 Next Best Underwater View

procedure NBUV($\boldsymbol{\mathcal{V}}(t+1), \hat{\mathcal{O}}(t)$)
  for each view $\mathbf{v}(t+1) \in \boldsymbol{\mathcal{V}}(t+1)$ **do**
    Pre-compute $E(\mathbf{x}^{\mathrm{C}}, t+1), B(\mathbf{x}^{\mathrm{C}}, t+1)$.
    Pre-compute $\sigma^2(\mathbf{x}^{\mathrm{C}}, t+1)$ using Eq. (32).
    for each $s(\mathbf{x}^{\text{texture}})$ visible from $\mathbf{v}(t+1)$ **do**
      $\sigma^2(\mathbf{x}^{\text{texture}}, t+1) \leftarrow \sigma^2(\mathbf{x}^{\mathrm{C}}, t+1)$ using Eqs. (33,34).
      Compute $q_s(t+1)$ using Eq. (19).
    Compute $\mathcal{I}_{t+1}(\mathcal{O})$ using Eqs. (22,24,25).
  $\hat{\mathbf{v}}(t+1) = \arg\max_{\mathbf{v} \in \boldsymbol{\mathcal{V}}} \mathcal{I}_{t+1}(\mathcal{O})$.
  Set **NBUV** as $\mathbf{v}(t+1) = \hat{\mathbf{v}}(t+1)$.
  Take the step: $t \leftarrow t+1$
  $I(\mathbf{x}^{\mathrm{C}}, t) \leftarrow$ Take image from $\mathbf{v}(t)$
  Compute $\hat{\rho}(\mathbf{x}^{\mathrm{C}}, t)$ using Eq. (30).
  for each $s(\mathbf{x}^{\text{texture}})$ visible from $\mathbf{v}(t)$, **do**
    Warp $\hat{\rho}(\mathbf{x}^{\text{texture}}, t) \leftarrow \hat{\rho}(\mathbf{x}^{\mathrm{C}}, t)$ using Eq. (31)
    Fuse $\hat{\rho}(\mathbf{x}^{\text{texture}}, t)$ to $\mathcal{Y}$ using Eq. (17).
    Update $Q_s^{\mathrm{ML}}(t)$.
  return $\hat{\rho}^{\mathrm{ML}}$

References

[1] N. A. Ahmed and D. Gokhale. Entropy expressions and their estimators for multivariate distributions. *IEEE Trans. IT*, 35:688–692, 1989.

[2] M. Ashikhmin and P. Shirley. An anisotropic Phong BRDF model. *Graph. Tools*, 5:25–32, 2000.

[3] D. Berman, T. Treibitz and S. Avidan. Non-local image dehazing In *Proc. IEEE CVPR*, 2016.

[4] R. Campos, R. Garcia, P. Alliez, and M. Yvinec. A surface reconstruction method for in-detail underwater 3D optical mapping. *Int. J. Robot. Res.*, 34:64–89, 2014.

[5] S. Chen and Y. Li. Vision sensor planning for 3-D model acquisition. *IEEE Trans. Syst., Man, Cybern. Part B: Cybern.*, 35(5):894–904, 2005.

[6] E. Coiras and J. Groen. Simulation and 3D reconstruction of side-looking sonar images. In *Advances in Sonar Technology*, ed. S.R. Silva, 2009.

[7] C. K. Cowan, B. Modayur, and J. L. DeCurtins. Automatic light-source placement for detecting object features. In *Proc. SPIE Conf. Intelligent Robots and Computer Vision*, pages 397–408, 1992.

[8] F. R. Dalgleish, A. K. Vuorenkoski, and B. Ouyang. Extended-range undersea laser imaging: Current research status and a glimpse at future technologies. *Mar. Tech. Soc. J.*, 47(5):128–147, 2013.

[9] E. Dunn, J. van den Berg, and J.-M. Frahm. Developing visual sensing strategies through next best view planning. In *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, pages 4001–4008, 2009.

[10] B. Englot and F. S. Hover. Sampling-based coverage path planning for inspection of complex structures. *In Proc. AAAI Conf. Automated Planning and Scheduling,* 2012.

[11] B. Englot and F. S. Hover. Three-dimensional coverage planning for an underwater inspection robot. *The Int. J. Robotics Research*, 32(9-10):1048–1073, 2013.

[12] R. Eustice, O. Pizarro and H. Singh. Visually augmented navigation for autonomous underwater vehicles. *IEEE J. Oceanic Eng.*, 33(2):103-122, 2008.

[13] R. Fattal. Single image dehazing. In *ACM TOG*, 27(3):1–9, 2008.

[14] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale light transport decomposition using interferometry. *ACM TOG*, 34(4), Article 37, 2015.

[15] C. A. Glasbey and K. V. Mardia A review of image-warping methods *J. Applied Statistics*, 25(2):155–171, 1998.

[16] M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *Proc. IEEE CVPR*, 2008.

[17] V. I. Haltrin. One-parameter two-term Henyey-Greenstein phase function for light scattering in seawater. *Applied Optics*, 41(6):1022–1028, 2002.

[18] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Trans. PAMI*, 33(12):2341–2353, 2011.

[19] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics Express*, 22(21):26338–26350, 2014.

[20] G. A. Hollinger, B. Englot, F. Hover, U. Mitra, and G. S. Sukhatme. Uncertainty-driven view planning for underwater inspection. In *Proc. IEEE Int. Conf. Robot. and Automation*, pages 4884–4891, 2012.

[21] V. Holodovsky, Y. Y. Schechner, A. Levin, A. Levis, and A. Aides. In-situ multi-view multi-scattering stochastic tomography. In *Proc. IEEE ICCP,* 2016.

[22] E. Iscar and M. Johnson-Roberson. Autonomous surface vehicle 3D seafloor reconstruction from monocular images and sonar data. In *Proc. IEEE/MTS OCEANS*, 24(3):1040–1049, 2005.

[23] J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE J. Oceanic Eng.*, 15(2):101–111, 1990.

[24] J. S. Jaffe. Multi autonomous underwater vehicle optical imaging for extended performance. In *Proc. IEEE OCEANS*, 2007.

[25] W. Jarosz, M. Zwicker, and H. W. Jensen. The beam radiance estimate for volumetric photon mapping. In *ACM SIGGRAPH Classes*, 2008.

[26] H. W. Jensen and P. H. Christensen. Efficient simulation of light transport in scenes with participating media using photon maps. In *Proc. SIGGRAPH*, pages 311–320, 1998.

[27] D. M. Kocak, F. R. Dalgleish, F. M. Caimi and Y. Y. Schechner. A focus on recent developments and trends in underwater imaging. *Mar. Tech. Soc. J.*, 42(1):52-67, 2008.

[28] T. G. Kolda, R. M. Lewis, and V. Torczon. A generating set direct search augmented lagrangian algorithm for optimization with a combination of general and linear constraints. SAND2006-5315, *Sandia National Laboratories*, 2006.

[29] F. Maurelli, S. Krupiński, Y. Petillot, and J. Salvi. A particle filter approach for AUV localization. In *Proc. IEEE OCEANS*, pages 1–7, 2008.

[30] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Trans. PAMI,* 15(5):417–433, 1993.

[31] S. G. Narasimhan, M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen. Acquiring scattering properties of participating media by dilution. *ACM TOG*, 25(3):1003–1012, 2006.

[32] S. Negahdaripour, and H. Madjidi. Stereovision imaging on submersible platforms for 3-D mapping of benthic habitats and sea-floor structures. *IEEE J. Oceanic Eng.*, 28(4):625–650, 2003.

[33] K. H. Norwich. *Information, Sensation, and Perception*. Academic Press, San Diego, 1993.

[34] M. O'Toole, R. Raskar, and K. N. Kutulakos. Primal-dual coding to probe light transport. *ACM TOG*, 31(4):39, 2012.

[35] L. Paull, S. Saeedi, M. Seto, and H. Li. AUV navigation and localization: A review. *IEEE J. Oceanic Eng.*, 39(1):131–149, 2014.

[36] R. Pito. A solution to the next best view problem for automated surface acquisition. *IEEE Trans. PAMI*, 21(10):1016–1030, 1999.

[37] N. Ratner and Y. Y. Schechner. Illumination multiplexing within fundamental limits. In *Proc. IEEE CVPR*, 2007.

[38] S. D. Roy, S. Chaudhury, and S. Banerjee. Active recognition through next view planning: A survey. *Pattern Recognition*, 37(3):429–446, 2004.

[39] S. Sakane and T. Sato. Automatic planning of light source and camera placement for an active photometric stereo system. In *Proc. IEEE Robotics and Automation*, pages 1080–1087, 1991.

[40] Y. Y. Schechner, D. J. Diner, and J. V. Martonchik. Spaceborne underwater imaging. *Proc. IEEE ICCP*, 2011.

[41] A. Sedlazeck and R. Koch. Simulating deep sea underwater images using physical models for light attenuation, scattering, and refraction. In *Proc. of VMV Workshop*, 2011.

[42] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. IP.*, 15(2):430–444, 2006.

[43] M. Sheinin and Y. Y. Schechner. The next best underwater view: Supplementary material. In *Proc. IEEE CVPR*, 2016.

[44] B. Sun, R. Ramamoorthi, S. G. Narasimhan, and S. K. Nayar. A practical analytic single scattering model for real time rendering. *ACM Trans. TOG*, 24(3):1040–1049, 2005.

[45] T. Treibitz and Y. Y. Schechner. Active polarization descattering. *IEEE Trans. PAMI,* 31(3):385–399, 2009.

[46] T. Treibitz and Y. Y. Schechner. Recovery limits in pointwise degradation. *Proc. IEEE ICCP*, 2009.

[47] T. Treibitz and Y. Y. Schechner. Turbid scene enhancement using multi-directional illumination fusion. *IEEE Tran. IP*, 21(11):4662–4667, 2012.

[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. IP*, 13(4):600–612, 2004.

[49] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, and J. Denzler. An information theoretic approach for next best view planning in 3-D reconstruction. In *Proc. IEEE ICPR,* 1:103–106, 2006.

[50] P. Whaite and F. P. Ferrie. Autonomous exploration: Driven by uncertainty. *IEEE Trans. PAMI,* 19(3):193–205, 1997.

[51] H. Zhang and K. J. Voss. Bidirectional reflectance study on dry, wet, and submerged particulate layers: effects of pore liquid refractive index and translucent particle concentrations. *Applied Optics*, 45(34):8753–8763, 2006.