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.

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
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, $\phi_L(t)$. A source whose intensity is $C_0$ irradiates a submerged surface patch $s$ from distance $l_{LS}$. The irradiance [16] model at $s$ is

$$\tilde{E}_s = D_s + A_s. \quad (1)$$

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 $\beta$. In a single scattering approximation [16, 31, 44],

$$D_s \propto C_0 \exp[-\beta l_{LS}] \frac{1}{l_{LS}^2}, \quad (2)$$

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

$$A_s \propto C_0 \exp[-\beta (l_{Lz} + l_{Sz})] \frac{1}{(l_{Lz}l_{Sz})^2}, \quad (3)$$

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 $\phi_C(t)$. The distance from $s$ to camera $C$ is $l_{SC}$. If the surface is Lambertian, the signal measured by $C$ is

$$\rho_s E_s, \quad \text{where } \rho_s \text{ is the albedo at } s \text{ and}$$

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

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

$$I_s = \rho_s E_s + B_s + n_I, \quad (5)$$

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

$$\sigma_{I_s}^2 = \sigma_{PN}^2 + \sigma_{RN}^2 = I_s + \sigma_{RN}^2. \quad (6)$$

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

$$\text{SNR}_s \approx \frac{I_s}{\sqrt{\sigma_{I_s}^2}} \approx \frac{\rho_s E_s}{\sqrt{\rho_s E_s + B_s + \sigma_{RN}^2}} \quad (7)$$

Backscatter is negligible, $B_s \ll \sigma_{RN}^2$, in a clear medium. Then from Eq. (7), under sufficient lighting $\text{SNR}_s \sim \sqrt{\rho_s E_s}$. 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 $\text{SNR}_s$ in (7). To reduce backscatter, $L$ is usually separated from $C$. Such a separation may create shadows. In a shadow, $E_s \ll \sigma_{RN}^2$, compounding light extinction by the medium (Eqs. 2-3). Thus optimal setting of $L$ is non-trivial.

---

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

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 $O$ represent a property of the object, e.g., the spatially varying albedo or topography. A computer vision system estimates this representation, $\hat{O}$, using sequential measurements. By time $t$, the camera has already accumulated image data $I(t')$, for all $t' \leq t$. All preceding data is processed to yield $\hat{O}(t)$. Let $\Phi_C$ be the set of all possible camera poses. A next view is planned for time $t + 1$, where the camera may be posed at $\phi_C(t + 1) \in \Phi_C$, yielding new data. The new data helps getting an improved estimate $\hat{O}(t + 1)$. The NBV question is: out of all possible views in $\Phi_C$, what is the best $\phi_C(t + 1)$, such that $\hat{O}(t + 1)$ has the best quality? Formulating this task mathematically depends on a quality criterion, prior knowledge about $O$, 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

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

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

$$ I_{t+1}(a) = H_t(a) - H_{t+1}(a). \quad (9) $$

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

$$ H_t(a) = (1/2) \ln[2\pi e \sigma^2_a(t)], \quad (10) $$

$$ I_{t+1}(a) = (1/2) \ln[\sigma^2_a(t) / \sigma^2_a(t + 1)]. \quad (11) $$

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 $\nu(t) = [\phi_C(t), \phi_L(t)]$. 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 $\phi_C \in \Phi_C$, $\phi_L \in \Phi_L$ and surface patch index $s$.

At close distance, optical imaging and descattering seek the spatial distribution of the surface albedo $O = \bigcup_s \rho_s$, to notice sediments, defects in submerged pipes, parasitic colonies in various environments etc. 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

$$ \hat{\rho}_{\nu}(t) = (I_s(t) - B_s(t)) / E_s(t). \quad (12) $$

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

$$ \sigma^2_{\nu}(t) = \sigma^2_s / E^2_s(t). \quad (13) $$

Note that $\sigma^2_s(t)$ is unknown, since Eqs. (5,6) depend on the unknown $\rho_s$. Nevertheless, it is possible to define an operating point value for $\rho_s$, 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

$$ \sigma^2_{\nu}(t) \approx \bar{\sigma}^2_s + \bar{\rho} E_s(t) + B_s(t) + \sigma^2_{2RN}. \quad (14) $$

$$ \sigma^2_{\nu}(t) \approx \bar{\rho} E_s(t) + B_s(t) + \sigma^2_{2RN}. \quad (15) $$

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

$$ \tilde{L}_s(t) = \ln L_s(t) \approx \sum_{t'=0}^{t} \left( I_s(t') - B_s(t') - \rho_s E_s(t') \right)^2 / \sigma^2_{\nu}(t'). \quad (16) $$

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

$$ \hat{\rho}_{sML}(t) = \frac{\sum_{t'=0}^{t} \hat{\rho}_{\nu}(t') [\sigma_s(t')]^{-2}}{\sum_{t'=0}^{t} [\sigma_s(t')]^{-2}}, \quad (17) $$

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

$$ [\sigma^2_{sML}(t)]^{-2} = \left\{ \sum_{t'=0}^{t} [\sigma_s(t')]^{-2} \right\}^{-1}. \quad (18) $$

Pre-calculate $\forall s, \nu(t)$ a quality measure of $s$

$$ q_s(t) \equiv 1 / \sigma^2_{\nu}(t). \quad (19) $$

\footnote{Vision can further enhance the topography estimation [4].}
From Eqs. (18,19), the quality of the ML descattered reflectivity $\hat{\rho}_s^{ML}(t)$ is

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

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

4. Next Best Underwater View

After time $t$, the next view $v(t+1)$ yields information gain $I_{t+1}(O)$. Let $\mathcal{V}$ 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 $\mathcal{V}$, to maximize the information gain measure $I_{t+1}(O)$,

$$\hat{v}(t+1) = \arg \max_{v \in \mathcal{V}} I_{t+1}(O). \quad (21)$$

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

$$I_{t+1}(O) = \sum_{s=1}^{N_s} I_{t+1}(\hat{\rho}_s^{ML}). \quad (22)$$

From Eq. (11),

$$I_{t+1}(\hat{\rho}_s^{ML}) = \frac{1}{2} \ln \left[ \frac{[\hat{\sigma}_s^{ML}(t)]^2 [\sigma_s^{ML}(t+1)]^{-2}}{Q_s^{ML}(t)} \right]. \quad (23)$$

From Eqs. (20,23),

$$I_{t+1}(\hat{\rho}_s^{ML}) = \frac{1}{2} \ln \left[ \frac{\sum_{t'=0}^{t-1} q_s(t')}{\sum_{t'=0}^{t-1} q_s(t')} \right] = \ln \left[ 1 + \frac{q_s(t+1)}{Q_s^{ML}(t)} \right] \frac{1}{2} \quad (24)$$

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

$$I_{t+1}(\hat{\rho}_s^{ML}) = \frac{1}{2} \ln \left[ 1 + \frac{q_s(t+1)}{1/\sigma_{\rho_s}^2} \right]. \quad (25)$$

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

$$I_{t_1 \rightarrow t_2}(\hat{\rho}_s^{ML}) = \frac{1}{2} \ln \left[ \frac{\sum_{t_1}^{t_2} q_s(t')}{\sum_{t_1}^{t_2} q_s(t')} \right] = \frac{1}{2} \ln \left[ \frac{Q_s^{ML}(t_2)}{Q_s^{ML}(t_1)} \right] \quad (26)$$

Thus

$$I_{t_1 \rightarrow t_2}(O) = \sum_{s=1}^{N_s} I_{t_1 \rightarrow t_2}(\hat{\rho}_s^{ML}). \quad (27)$$

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

$$L_{\text{best}} = \arg \max_{L} [I_{t_1 \rightarrow t_2}(O)] . \quad (28)$$

We implement Eq. (28) by perturbing an initial path $L_0$.

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, comprising a set of mesh faces $\{ T_k \}_{k=1}^{N_m}$. Surface albedo is represented by a texture map domain $Y = \{ Y_k \}_{k=1}^{N_m}$. Face $T_k$ corresponds to a triangle $Y_k$ in the map domain (Fig. 4d). A texture-map pixel $x_{\text{texture}} \in Y$ has one-to-one correspondence with a specific surface patch: $s(x_{\text{texture}})$.

---

4Any polygonal representation of a mesh can be used here.
In the image-plane of camera $C$, a spatial location is denoted $x^C$, in pixel units. At time $t$, the camera has a simulated projection operator $P^C_t$. Then, patch $s$ projects to a spatial set $x^C \in U_s(t)$:

$$x^\text{texture} \leftrightarrow s \leftrightarrow U_s(t).$$  \hfill (29)

The area of $U_s(t)$ is $|U_s(t)|$, in pixel units. Using this model, computer graphics renders components of Eqs. (12,15), $E(x^C, t) = P^C_t(E_s)$, $B(x^C, t) = P^C_t(B_s)$.

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

$$\hat{\rho}(x^C, t) = |I(x^C, t) - B(x^C, t)|/E(x^C, t)$$  \hfill (30)

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

$$\hat{\rho}(x^\text{texture}, t) = \text{WARP} \{\hat{\rho}(x^C, t)\}_{x^C \in U_s(t)}.  \hfill (31)$$

Similarly, Eq. (15) is expressed by

$$\sigma^2(x^C, t) \approx \hat{\rho}(x^C, t) E(x^C, t) + B(x^C, t) + \sigma^2_{\text{RN}}$$  \hfill (32)

in the camera coordinates. Note that when $|U_s(t)| > 1$, warping by Eq. (31) implicitly involves spatial averaging. Thus, warping decreases the variance of the value in the texture map, in proportion to $|U_s(t)|$. Hence, the result of Eq. (32) is generally transferred to the texture map by

$$\sigma^2(x^\text{texture}, t) = |U_s(t)|^{-1} \text{WARP} \{\sigma^2(x^C, t)\}_{x^C \in U_s(t)}.  \hfill (33)$$

If $|U_s(t)| < 1 \ \forall t$, then patch $s$ is never observed at sufficient resolution. Per patch, the image sequence should strive to provide raw measurements $I(x^C, t)$ at sufficient resolution, in which $|U_s(t)| \geq 1$. Over a flat terrain, this requirement is easily met by keeping $C$ 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 $|U_s(t)| < 1$, the patch’s variance is penalized as

$$\sigma^2(x^\text{texture}, t) = \exp(\eta(|U_s(t)|^{-1} - 1))$$  \hfill (34)

WARP$\{\sigma^2(x^C, t)\}_{x^C \in U_s(t)}$

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

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

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

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

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

$$I_{t+1}(T_k) \approx \frac{|Y_k|}{2} \ln \left[ 1 + \frac{q_k(t+1)}{Q_k^ML(t)} \right].  \hfill (36)$$

The information gain is therefore:

$$I_{t+1}(\mathcal{O}) \approx \sum_{k=1}^{N_{m}} I_{t+1}(T_k).  \hfill (37)$$

7. Simulations

We set $\sigma_{\text{RN}} = 13.1$ [e] and a full well of 24,000 [e] in a perspective camera $C$, based on Canon 60D specs, while $L$ is a spotlight with no lateral falloff. Fig. 5 illustrates a simple case study. The medium’s parameters are $\beta = 5/1$ [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 $\{\phi_k(t)\}_{t=1}^{8}$ are spaced uniformly along the path. $C$ and $L$ start from $v(0)$. The initial $LC$ baseline is 2cm. Underwater this baseline results in significant backscatter. Hence, the baseline increases to
12 cm. A traditional path separates L from C by a fixed baseline whose length and orientation are fixed, \( L = 12\hat{x} \). 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 what happens when the scattering medium is ignored. This is achieved by setting \( \beta \to 0 \) in the optimization. The resulting \( \{ \phi_L(t) \}_{t=1}^{10} \) seeks to avoid shadows and maximize \( E_s \) over the surface. This mostly leads to a very small \( |L\hat{C}| \) 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. 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 \( \mathbb{C} \cup \mathbb{L} \) to \( \{ \mathbf{v}(t) \}_{t=0}^{10} \).

In Exp2, the setup (Fig. 9) was in fog created by a AT-MOS FOG AF-1200 machine. The illumination was made by a LED (Mouser Electronics ‘Warm White’ 3000K). The extinction and phase-function parameters \( \beta, g \) were estimated in-situ:

\[
g, \beta = \arg\min_{g, \beta} \| E(x^C, t) + B(x^C, t) - I(x^C, t) \|^2,
\]

\( ^5 \)Rotation about axis \( Z \) is degenerate, thus excluded for both \( \mathbb{L} \) and \( \mathbb{C} \).

\( ^6 \)Measurement were simulated using a Phong-model with parameters \( K_s = 0, K_d = 1, K_a = 0.5, \alpha = 20 \).

\( ^7 \)Underwater \( \beta = 12[1/m], \quad g = 0.6 \). In the fog, we manually assessed \( \beta \approx 2.5 \) and \( g = 0.6 \).
when C&L were placed in a known state above a flat white sheet ($\rho_s \approx 1$).

### 8.1. Numerical Conditioning

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

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

(39)

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

### 8.2. Results

In Exp1, $\{\phi_C(t)\}_{t=0}^{10}$ are manually set in uniformly-spaced locations starting from $v(0)$ 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, $\vec{L}$C is fixed so as to avoid overwhelming backscatter. By NBUV, per t, L can be placed in $|\nabla(t)| = 40$ states around $\phi_C(t)$ $\forall t \in [1..10]$. 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 $I(O)$ is lower, though
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 $\hat{\rho}(x^c, t)$, 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

```
1: procedure NBUV($V(t+1), O(t)$)
2:     for each view $v(t+1) \in V(t+1)$ do
3:         Pre-compute $E(x^c, t + 1), B(x^c, t + 1)$.
4:     for each $s(x^{\text{texture}})$ visible from $v(t+1)$ do
5:         $\sigma^2(x^{\text{texture}}, t + 1) \leftarrow \sigma^2(x^c, t + 1)$ using Eqs. (33,34).
6:         $\rho(x^{\text{texture}}, t + 1) \leftarrow \rho(x^c, t + 1)$ using Eq. (17).
7:     return $\hat{\rho}_V^\text{ML}(t)$.
```

References


